www.koreaai.org

제6회 한국 인공지능 학술대회

The 6th Korea Artificial Intelligence Conference

Collaborative Intelligence: Human-AI Synergy in the Real World

Program

일자 2025년 9월 24일(수) ~ 9월 27일(토)

장소 서귀포 칼 호텔

주최 한국통신학회

주관 한국통신학회 인공지능 소사이어티

후원 ICT Convergence Korea 포럼, ETRI 한국전자통신연구원, KETI 한국전자기술연구원, TTA, Hanwha Aerospace, CITAC, 경북대학교 ICT·자동차융합연구센터, 국민대 AI Mobility 연구소, 한국기술교육대학교 첨단기술연구소, 국민대 FSO/OCC 국방기술특화연구실

KICS 한국통신학회

준비위원

▣ 자문위원

강충구(고려대)
김영한(숭실대)
노종선(서울대)
박세웅(서울대)
방승찬(ETRI)
신요안(숭실대)
신희동(KETI)
정성호(한국외국어대)
정일영(한국외국어대)
조용수(중앙대)
조유제(경북대)
조현훈(키엘연구원)
허 준(고려대)
홍인기(경희대)

▣ 조정위원장

장영민(국민대)

조정위원

김대중(TTA)
김승환(ETRI)
박종빈(키엘연구원)
신제민(한국보건의료연구원)
여 현(순천대)
이인규(고려대)
유원필(ETRI)
조성래(중앙대)
최병호(KETI)
황승훈(동국대)

▣ 운영위원장

한동석(경북대)

▣ 운영부위원장

김상철(국민대)
신석주(조선대)
이규만(건양대)
남해운(한양대)
김도현(제주대)

운영위원

고준호(TTA)
김덕경(인하대)
김동균(경북대)
김동성(금오공과대)
김정곤(한국공학대)
김중헌(고려대)
김홍국(광주과학기술원)
백명선(세종대)
석준희(고려대)
이예훈(서울과기대)
이현우(ETRI)
정연호(부경대)
최계원(성균관대)
최윤호(부산대)
최지웅(DGIST)
한종기(세종대)
황인태(전남대)

준비위원

▣ 프로그램위원장

손인수(동국대) 한연희(한국기술교육대) 박현희(명지대)

▣ 프로그램부위원장

김중헌(고려대) 김평수(한국공학대) 신수용(금오공대)

프로그램위원

고남석(ETRI) 고석주(경북대) 고정길(연세대)
고한얼(경희대) 권민혜(숭실대) 권태수(서울과기대)
길준민(제주대) 김광주(ETRI) 김덕수(한국기술교대)
김민석(상명대학교) 김수민(한국공학대) 김원태(한국기술교대)
김은경(한밭대) 김정현(세종대) 김호원(부산대)
김효수(중앙대) 문상미(나사렛대) 박대진(경북대)
박승현(한성대) 박준구(경북대) 박형곤(이화여대)
박혜영(경북대) 배정숙(ETRI) 서익수(국방과학연구소)
서효운(성균관대) 송윤정(ETRI) 신오순(숭실대)
심병효(서울대) 안희진(KAIST) 양현종(서울대)
오지용(ETRI) 오흥선(한국기술교대) 유준혁(대구대)
이승형(광운대) 이정훈(한국외대) 이재호(덕성여대)
이재민(금오공대) 이종혁(세종대) 이주현(한양대)
이해윤(한국기술교대) 임길택(ETRI) 임완수(금오공대)
장혜령(동국대) 전강욱(한국기술교대) 전상운(한양대)
전찬준(조선대) 정성아(서울시립대) 정준영(ETRI)
조봉균(KIAPI) 조오현(충북대) 조정훈(경북대)
최민석(경희대) 최우열(중앙대학교) 최현호(한경대학교)
홍준표(홍익대)

▣ 출판위원장

김은경(한밭대)

▣ 등록위원장

이규만(건양대)

▣ 현장위원장

이규만(건양대) 백명선(세종대)

진행시간표

2025년 9월 24일 Wednesday

<table>
<tr><th rowspan="2">Time</th><th colspan="4">Session information</th></tr>
<tr><th>Room 1 다이아몬드룸</th><th>Room 2 에메랄드룸</th><th>Room 3 사파이어룸</th><th>Room 4 로즈룸</th></tr>
<tr><td>12:00–12:20
(20")</td><td colspan="4">Registration</td></tr>
<tr><td>12:20–13:20
(60")</td><td>일반세션 A-1
대규모언어생성AI 1
최우열 교수 (중앙대)</td><td>일반세션 A-11
영상신호AI 1
최계원 교수 (성균관대)</td><td>일반세션 A-18
지능형산업응용 1
김광주 실장 (ETRI)</td><td>학부생세션 B-1
대규모언어생성AI 1
이정훈 교수 (외국어대)</td></tr>
<tr><td>13:20–13:30
(10")</td><td colspan="4">Session Break</td></tr>
<tr><td>13:30–14:00
(30")</td><td rowspan="2">Top Conference I & II
좌장: 한동석 교수 (경북대)
장소: 다이아몬드룸</td><td colspan="3">Top Conference I
Discovering Features with Synergistic Interactions in Multiple Views (ICML 2024)
이창희 교수 (고려대)</td></tr>
<tr><td>14:00–14:30
(60")</td><td colspan="3">Top Conference II
LLM-Guided Regional Identification for Instruction-based Image Editing (ACM MM 2025)
박민호 연구원 (ETRI)</td></tr>
<tr><td>14:30–14:40
(10")</td><td colspan="4">Session Break</td></tr>
<tr><td>14:40–15:20
(40")</td><td rowspan="2">Tutorial I & II
좌장: 신석주 교수 (조선대)
장소: 다이아몬드룸</td><td colspan="3">Tutorial I
스마트시티의 AI기반 도심인프라 관리 기술
김광주 실장 (ETRI)</td></tr>
<tr><td>15:20–16:00
(40")</td><td colspan="3">Tutorial II
Physical AI based on Decentralized Federated Learning
김도현 교수 (제주대)</td></tr>
<tr><td>16:00–17:00
(60")</td><td>일반세션 A-19
지능형산업응용 2
김도현 교수 (제주대)</td><td rowspan="2">[특별세션] S-7
ETRI 피지컬 AI (Physical AI)
이승환 본부장 (ETRI)</td><td>[특별세션] S-4
ICT혁신인재4.0 (금오공대)
신수용 교수 (금오공대)</td><td>일반세션 A-13
인공지능응용 1
이규만 교수 (건양대)</td></tr>
<tr><td>17:00–18:00
(60")</td><td>일반세션 A-10
시스템최적화
손인수 교수 (동국대)</td><td>[특별세션] S-5
ICT혁신인재4.0 (금오공대)
신수용 교수 (금오공대)</td><td>일반세션 A-8
사이버보안AI 1
박승현 교수 (한성대)</td></tr>
</table>

진행시간표

2025년 9월 25일 Thursday

Time	Session information			
	Room 1 다이아몬드룸	Room 2 에메랄드룸	Room 3 사파이어룸	Room 4 로즈룸
08:30-09:00 (30")	Registration			
09:00-10:00 (60")	**일반세션 A-20** 지능형산업응용 3 정한민 교수 (한국과학기술정보연구원/과학기술연합대학원대학교)	**일반세션 A-5** 멀티모달AI 1 박현희 교수 (명지대학교)	**일반세션 A-17** 자율로보틱스 노송 교수 (인천대)	**학부생세션 B-8** 지능형산업응용 1 백명선 교수 (세종대)
10:00-10:10 (10")	Session Break			
10:10-10:20 (10")	**개회식** 좌장: 손인수 교수 (동국대)			
10:20-11:00 (40")	**Keynote I** 제조 피지컬 AI와 SDF 장영재 교수 (KAIST)			
11:00-11:40 (40")	**Keynote II** Towards Reliable and Efficient Multimodal AI 윤상두 소장 (NAVER AI Lab.)			
11:40-12:20 (40")	**Keynote III** Interdisciplinary R&D and Business between Interactive Machine Learning and Information Communication technology Prof. Kohno (IEEE & IEICE Life Fellow)			
12:20-14:00 (100")	Lunch			
14:00-14:40 (40")	**Tutorial III & IV** 좌장: 신석주 교수 (조선대학교) 장소: 다이아몬드룸	**Tutorial III** ML-Based Wireless Network Optimization and Applications Prof. Mikio Hasegawa		
14:40-15:20 (40")		**Tutorial IV** Indoor Position Estimation for Smart Factories and its Application to Radio Environment Map Prof. Takeo Fujii		
15:20-15:30 (10")	Session Break			
15:30-16:30 (60")	**학부생세션 B-2** 대규모언어생성AI 2 박천음 교수 (한밭대)	**[특별세션] S-1** 사이버C.O.P. 특별세션 I 이용태 교수 (한림대)	**[특별세션] S-6** FSO/OCC특화연구실 장영민 교수 (국민대)	**[특별세션] S-9** 인공지능 전환 (NIA) 김동희 팀장 (한국지능정보사회진흥원)
16:30-17:30 (60")	**일반세션 A-21** 지능형산업응용 4 김대년 박사 (경북IT융합산업기술원)	**[특별세션] S-2** 사이버C.O.P. 특별세션 II 김지온 교수 (한림대)	**일반세션 A-14** 인공지능응용2 이용민 박사 (ETRI)	**학부생세션 B-3** 인공지능응용1 길준민 교수 (제주대)
18:30-20:00	Banquet / 사회: 김상철 교수 (국민대)			

진행시간표

2025년 9월 26일 Friday

Time	Session information			
	Room 1 다이아몬드룸	Room 2 에메랄드룸	Room 3 사파이어룸	Room 4 로즈룸
08:30-09:00 (30")	Registration			
09:00-10:00 (60")	**일반세션 A-2** 대규모언어생성AI 2 김정곤 교수 (한국공학대)	**일반세션 A-15** 인공지능응용 3 김영선 박사 (한국전기연구원)	**[특별세션] S-10** 금오공대 MOSA 김동성 교수 (금오공대)	**학부생세션 B-9** 지능형산업응용 2 김봉완 박사 (ETRI)
10:00-10:10 (10")	Session Break			
10:10-11:10 (60")	**일반세션 A-12** 영상신호AI 2 윤상민 교수 (국민대)	**일반세션 A-22** 지능형산업응용 5 김용강 교수 (공주대)	**[특별세션] S-3** 경북대-ETRI 미래 디지털 융합기술 송윤정 박사 (ETRI)	**학부생세션 B-4** 인공지능응용 2 김상철 교수 (국민대)
11:10-12:10 (60")	**일반세션 A-7** 바이오AI 김평수 교수 (한국공학대)	**일반세션 A-9** 사이버보안AI 2 백명선 교수 (세종대)	**학부생세션 B-5** 인공지능응용3 한연희 교수 (한국기술교육대)	**일반세션 A-25** AI안전및정책 윤주상 교수 (동의대)
12:10-13:30 (80")	Lunch			
13:30-14:00 (30")	**Top Conference III** 좌장: 김평수 교수 (한국공학대학교) 장소: 다이아몬드룸	**Top Conference III** Episodic Future Thinking Mechanism for Multi-agent Reinforcement Learning (NeurIPS 2024) 권민혜 교수 (숭실대)		
14:00-15:00 (60")	**일반세션 A-6** 멀티모달AI 2 김민석 교수 (상명대)	**일반세션 A-23** 지능형산업응용 6 김은경 교수 (한밭대)	**일반세션 A-3** 대규모언어생성AI 3 김동완 교수 (동아대)	**학부생세션 B-6** 인공지능응용 4 정민영 교수 (안산대)
15:00-16:00 (60")	**일반세션 A-4** 대규모언어생성AI 4 황인태 교수 (전남대)	**일반세션 A-24** 지능형산업응용 7 김은경 교수 (한밭대)	**일반세션 A-16** 인공지능응용 4 김민석 교수 (상명대)	**학부생세션 B-7** 인공지능응용 5 장혜령 교수 (동국대)

논문 발표 세부 안내

2025년 09월 24일(수)

논문발표 A-1. 대규모언어생성AI 1

09월24일(수) 12:20-13:20(60분) | 좌장: 최우열 교수 (중앙대)

A-1-1 래더링 인터뷰 기법을 활용한 대규모 언어 모델(LLM)의 정치적 성향 평가 및 분석 프레임워크 연구 1
장교철(서울대학교), 최재훈(서울대학교), 신민정(서울대학교), 이정(서울대학교), 서봉원(서울대학교)

A-1-2 대규모 언어모델의 자기성찰 능력 향상을 위한 피드백 최적화 기법 4
김준석(서울대학교), 정교민(서울대학교)

A-1-3 Chain-of-Function: Inducing Sequential Plans Over Decomposed Code Indexes for Code Generation 6
김종윤(서울대학교), 황승원(서울대학교)

A-1-4 작업별 중요도 기반 프루닝을 활용한 LoRA 어댑터 병합 기법 8
윤도경(한양대학교), 서지원(서울대학교), 조동현(한양대학교)

A-1-5 한국어 추론에서 Zero-shot 프롬프트 유형별 성능 분석 연구 10
박선우(성균관대학교), 문준렬(성균관대학교), 김윤하(성균관대학교), 임완수(성균관대학교)

논문발표 A-11. 영상신호AI 1

09월24일(수) 12:20-13:20(60분) | 좌장: 최계원 교수 (성균관대)

A-11-1 다양한 조명 환경에서 보행자 탐지를 위한 퍼지 논리 기반 전처리 프레임워크 12
이원일(롯데이노베이트), 김태형(Lotteinnovate), 정성균(롯데이노베이트), 유연승(롯데정보통신),
이민호(롯데이노베이트), 이수빈(롯데정보통신), 최흥준(롯데정보통신)

A-11-2 가우시안 어둠 분포 학습을 통한 잠재 확산 기반 저조도 이미지 밝기 복원 기법에 관한 연구 14
송순용(한국전자통신연구원)

A-11-3 STOI 메트릭 기반의 다중 레벨 음성 향상 및 데이터 증강 기법 16
신도경(LIG넥스원), 김영대(엘아이지넥스원(주))

A-11-4 3D 모델 합성 데이터셋을 이용한 항공 영상 객체 탐지 연구 18
이도현(엘아이지넥스원(주)), 김성용(엘아이지넥스원(주)), 조규태(엘아이지넥스원(주))

A-11-5 최신 비전 트랜스포머 변형 모델 동향에 관한 연구 20
지훈(한양대학교), 임한비(한양대학교), 이연준(한양대학교)

논문 발표 세부 안내

논문발표 A-18. 지능형산업응용 1

09월 24일(수) 12:20-13:20(60분) | 좌장: 김광주 실장 (ETRI)

A-18-1 조건 기반 배경 이미지 생성과 희귀 객체 합성에 관한 연구 22
박상영(엘아이지넥스원(주))

A-18-2 업종별 데이터 변동성에 따른 최적 머신 러닝 모델 선택 전략 : 서울시 상권 예측 사례 24
이하나(Korea University Graduate School of SW·AI convergence)

A-18-3 분산전력 시스템을 위한 강화학습 기반 CVR 제어 및 에너지 절감 효과 연구 26
이주승(한국전기연구원), 이순우(한국전기연구원), 강지명(한국전기연구원)

A-18-4 Two-Stage Stratified Machine Learning Approach for Identifying Root-Cause Factors In Manufacturing Processes 28
이호진(포스코), 정의홍(포스코), 조창권(포스코), 조수정(포스코), 안성수(포스코)

A-18-5 다중 위성 영상 데이터를 활용한 연합 학습 기반의 객체 탐지 연구 30
김성철(엘아이지넥스원(주)), 김성용(엘아이지넥스원(주)), 조규태(엘아이지넥스원(주))

A-18-6 CFD 데이터셋과 적대적 공격 기법을 활용한 차세대 스텔스 향상 설계 방법에 관한 연구 32
박지성(스마트엠투엠)

학부생 논문발표 B-1. 대규모언어생성AI1

09월 24일(수) 12:20-13:20(60분) | 좌장: 이정훈 교수 (외국어대)

B-1-1 감성 극성 분류 모듈과 다중 에이전트 구조를 활용한 개선된 다중 문서 요약 기술 연구 34
조병웅(성균관대학교), 봉한성(성균관대학교), 홍주화(성균관대학교)

B-1-2 생성형 인공지능과 검색 증강 생성을 활용한 학과 정보 제공 시스템 개발 37
천민우(성결대학교), 정세준(성결대학교), 신혜민(성결대학교), 김시현(성결대학교), 최정열(Sungkyul University)

B-1-3 온프레미스(On-premise) 환경에서금융투자정보 분석을 위한 MCP(Model Context Protocol), A2A(Agent to Agent) 하이브리드 기반 멀티에이전트 시스템 설계 및 구현 39
박시형(동서울대학교), 김두홍(동서울대학교), 이재희(동서울대학교)

논문 발표 세부 안내

논문발표 A-19. 지능형산업응용 2

09월 24일(수) 16:00-17:00(60분) | 좌장: 김도현 교수 (제주대)

A-19-1 KV Cache Storage Offloading에서 유사 Prompt의 Caching 방법에 관한 연구 41
이청솔(삼성전자)

A-19-2 Scene Graph 생성에서 환각 완화를 위한 Negative Data Augmentation 기법 43
백종우(대구경북과학기술원), 소진현(대구경북과학기술원)

A-19-3 SHAP 기반 Subpopulation 분석과 머신러닝을 활용한 OLED 증착 공정 PPA 편차 원인 인자 도출 방법론 46
양상식(삼성디스플레이(SDC)), 서동주(삼성디스플레이(SDC))

A-19-4 BLE fingerprinting 을 활용한 DRSS-DNN 기반 실내 위치 추정에 관한 연구 48
공연정(충남대학교), 권예은(충남대학교), 진형준(충남대학교), 강주호(충남대학교), 김민희(충남대학교), 김민석(충남대학교)

A-19-5 코바늘 객체 인식 기반 기호도 자동 생성 연구 50
오장훈(성균관대학교), 이송희(건국대학교), 김문기(동국대학교), 전휘빈(상지대학교), 김유현(University College London), 유진희(안양대학교), 권민규(서울대학교)

A-19-6 초심층 잔차 네트워크의 그래디언트 흐름 분석과 격자 기반 분화-병합 패러다임 52
박병준(경북대학교), 한동석(경북대학교)

특별세션 S-4. ICT혁신인재4.0

09월 24일(수) 16:00-17:00(60분) | 좌장: 신수용 교수 (금오공대)

S-4-1 화재 진압용 드론 시스템을 위한 카메라-LiDAR 센서 융합 기반의 건물 유리창 감지 및 파손 위치 추정 기술 연구 ... 54
이신우(국립금오공과대학교), 신수용(국립금오공과대학교)

S-4-2 Asynchronous Reputation Management Framework For V2X Networks Under Intermittent Connectivity 56
NAKAYIZA HOPE LETICIA(국립금오공과대학교), Love Allen Chijioke Ahakonye(국립금오공과대학교), 김동성(국립금오공과대학교), 이재민(국립금오공과대학교)

S-4-3 가우시안 혼합 모델을 활용한 방수 궤적의 확률적 추론 58
임건호(국립금오공과대학교), 이영준(국립금오공과대학교), 김승혁(국립금오공과대학교), 주백석(국립금오공과대학교)

S-4-4 대규모 언어 모델에서 차별적 계층 식별 방법 60
곽지호(국립금오공과대학교), 정유철(국립금오공과대학교)

논문 발표 세부 안내

특별세션 S-5. ICT혁신인재4.0

09월 24일(수) 17:00-18:00(60분) | 좌장: 신수용 교수 (금오공대)

S-5-1 온디바이스 AI 기반 웨어러블 스토킹 방지 시스템 63
jehyeok ahn(국립금오공과대학교), 이재민(국립금오공과대학교), 김동성(국립금오공과대학교)

S-5-2 자연어 기반 자율이동을 위한 Dual-LLM 접근법 65
김인곤(국립금오공과대학교), 신수용(국립금오공과대학교)

S-5-3 경량화된 행동 인식 분류 알고리즘을 위한 시계열 데이터 변환 및 특징 추출 시스템 제안 67
PARK JAE HAN(국립금오공과대학교), 이만희(국립금오공과대학교), 신수용(국립금오공과대학교)

논문발표 A-13. 인공지능응용 1

09월 24일(수) 16:00-17:00(60분) | 좌장: 이규만 교수 (건양대)

A-13-1 Transformer의 Feed-Forward Network 가속화 연구 69
이창근(성균관대학교), 신새빈(성균관대학교), 임완수(성균관대학교)

A-13-2 Impact of Sampling Rate on Deep Learning-based Human Activity Recognition Using Wearable Sensors 71
응우옌티호아이투(경북대학교), 한동석(경북대학교)

A-13-3 엣지-디바이스 협업 환경에서 지연 시간 최소화를 위한 ResNet 동적 분할 전략 연구 73
정준호(동의대학교), 박진효(동의대학교), 윤주상(동의대학교)

A-13-4 베이지안 GMM 기반 탐지를 활용한 노이즈 견고 연합학습 기법 75
조예지(세종대학교), 김정현(세종대학교), 송홍엽(연세대학교)

A-13-5 Predicting Deepfake Virality on Social Networks 77
정서윤(성균관대학교), 김지은(성균관대학교), S. Shyam Sundar(성균관대학교), 한진영(성균관대학교)

A-13-6 논문 서지 데이터를 활용한 AI 분야의 한·미·중 연구자 이동성 분석 79
권태훈(한국과학기술정보연구원)

논문발표 A-10. 시스템최적화

09월 24일(수) 17:00-18:00(60분) | 좌장: 손인수 교수 (동국대)

A-10-1 실시간 공항 운영을 위한 PPO 기반 자원 배정 최적화 시스템 설계 81
김준규(케파스), 천민범(케파스), 안광모(케파스)

A-10-2 연산 스토리지 기반 전처리 오프로딩을 적용한 AI 컨테이너 오케스트레이션 아키텍처 설계에 관한 연구 ... 83
길주현(한국전자기술연구원), 안재훈(Korea Electronics Technology Institute),
김영환(Korea Electronics Technology Institute)

A-10-3 다중 테넌트 환경에서 GPU 커널 스케줄링 기반 적응형 자원 할당 시스템 연구 85
권용진(한국전자기술연구원), 안재훈(Korea Electronics Technology Institute),
김영환(Korea Electronics Technology Institute)

A-10-4 AI 반도체 자원 풀링을 위한 시공간 통합 관리 아키텍처에 관한 연구 87
안성모(한국전자기술연구원), 안재훈(Korea Electronics Technology Institute),
김영환(Korea Electronics Technology Institute)

A-10-5 멀티 클러스터 환경에서 AI 워크로드 GPU,NPU 자원 할당과 운용 비용 최적화 90
조성민(한국전자기술연구원), 안재훈(Korea Electronics Technology Institute),
김영환(Korea Electronics Technology Institute)

A-10-6 Kubernetes 마이크로서비스의 리소스 효율성을위한 예측 기반 오토스케일링 기법 92
정현수(제주대학교), 길준민(제주대학교)

논문발표 A-8. 사이버보안AI 1

09월 24일(수) 17:00-18:00(60분) | 좌장: 박승현 교수 (한성대)

A-8-1 강화학습 기반 GNFS 분석을 활용한 RSA 보안성 평가 연구 94
이영도(NSHC), 김수리(성신여자대학교), 윤기순(NSHC)

A-8-2 보이스피싱 탐지 AI 강화를 위한 ATT&CK -RAG 프레임워크 96
설동명(한국전자통신연구원), 김지용(한국전자통신연구원)

A-8-3 RF Fingerprinting과 DNN을 이용한 NFC 태그 인증 시스템 98
이은규(고려대학교), 유호근(고려대학교), 허경무(고려대학교), 이재훈(고려대학교)

A-8-4 노이즈에 강인한 아크탐지 알고리즘 100
이순우(한국전기연구원), 강지명(한국전기연구원), 이주승(한국전기연구원), 최성수(한국전기연구원)

A-8-5 심층 강화학습을 활용한 비동기식 Self-Play 사이버 공방 연구 102
김정현(상명대학교), 정재혁(상명대학교), 장준원(상명대학교), 김민석(상명대학교)

논문 발표 세부 안내

2025년 09월 25일(목)

논문발표 A-20. 지능형산업응용 3

09월 25일(목) 09:00-10:00(60분) | 좌장: 정한민 교수 (한국과학기술정보연구원/과학기술연합대학원대학교)

A-20-1 비학습 기반 하이브리드 기법을 이용한 LiDAR 차선 검출 105
강대근(경북대학교), 은승우(경북대학교), 한동석(경북대학교)

A-20-2 비지도 학습 기반의 TEM 이미지 결손 원자 탐지 알고리즘 107
김민성(한국기초과학지원연구원), 이상철(한국기초과학지원연구원)

A-20-3 현장 신뢰성 향상을 위한 인과 분석 • 심각도 추정 통합 기반 웨이퍼 결함 분석 및 자연어 리포트 생성 109
김서현(명지대학교), 박현희(명지대학교)

A-20-4 주파수 대역 분할 및 그룹 컨볼루션을 이용한 개선된 고속 푸리에 컨볼루션 111
허재호(SK텔레콤(주)), 배주한(SK텔레콤(주)), 주현철(SK텔레콤(주)), 나태영(SK텔레콤(주))

A-20-5 다목적 강화학습 기반 최적 산업구조 탐색에 관한 연구 113
홍윤기(한국과학기술연구원), 이상철(한국과학기술연구원), 조한얼(세종대학교), 김찬수(한국과학기술연구원)

A-20-6 Trend-Seasonal 분해를 적용한 Transformer 기반 전기차 배터리 SoH 예측 115
천유림(성균관대학교), 윤유정(성균관대학교), 변예린(성균관대학교), 임완수(성균관대학교)

논문발표 A-5. 멀티모달비전AI 1

09월 25일(목) 09:00-10:00(60분) | 좌장: 박현희 교수 (명지대학교)

A-5-1 전자동 생성형 이미지 증강 시스템에 대한 연구 117
정성윤(엘아이지넥스원(주)), 정남훈(엘아이지넥스원(주)), 조규태(엘아이지넥스원)

A-5-2 Synthetic-driven Text-to-Image Generation with Preference-based Reinforcement Learning ... 119
최범규(엘아이지넥스원(주)), 정남훈(엘아이지넥스원(주)), 조규태(엘아이지넥스원(주))

A-5-3 도메인 일반화를 위한 프롬프트 앙상블 기반 Zero-shot 화재 탐지 기법 121
정성균(롯데이노베이트), 최흥준(롯데정보통신), 이민호(롯데이노베이트), 이원일(롯데이노베이트),
이수빈(롯데정보통신), 유연승(롯데정보통신), 김태형(Lotteinnovate)

A-5-4 한국어 기반 Text-to-Video 프롬프트 엔지니어링 시스템 연구 : CogVideoX를 중점으로 123
임영기(피씨엔), 최지희(피씨엔), 엄태선(피씨엔), 이우성(피씨엔), 구영현(세종대학교)

A-5-5 다중모달 신체·행동 지표를 이용한 머신러닝 기반 경도인지장애 단계 예측에 관한 연구 125
이은주(한림대학교), Yeo Jin Kim(강동성심병원)

논문 발표 세부 안내

논문발표 A-17 자율로보틱스

09월 25일(목) 09:00-10:00(60분) | 좌장: 노송 교수 (인천대)

A-17-1 시각 정보를 이용한 3DGS 기반 안전 경로 계획 127
함형찬(한국과학기술원), 안희진(한국과학기술원)

A-17-2 DQN 기반 적응형 전송 기법을 통한 NR-V2X 혼합 트래픽의 환경의 성능 개선 129
강태구(한밭대학교), 황경호(한밭대학교)

A-17-3 자율주행 환경에서 상황 인지 기반 적응형 시맨틱 통신 기법 131
오성현(한국공학대학교), 임영훈(한국공학대학교), 김정곤(한국공학대학교)

A-17-4 모달리티 불균형 완화를 위한 개별 헤드 기반 예측 융합 133
윤영우(경북대학교), 최원준(경북대학교), 이동규(경북대학교)

A-17-5 모빌리티 환경을 위한 AI 모델 분할 기반 지능형 컨테이너 오케스트레이션 135
박기철(한국전자기술연구원), 안재훈(Korea Electronics Technology Institute),
김영환(Korea Electronics Technology Institute)

학부생논문발표 B-8. 지능형산업응용 1

09월 25일(목) 09:00-10:00(60분) | 좌장: 백명선 교수 (세종대)

B-8-1 지진파 신호 노이즈 제거를 위한 9개 딥러닝 모델 성능 비교 연구 137
황선준(연세대학교), 박세희(연세대학교 미래캠퍼스), 고강민(연세대학교 미래캠퍼스),
백지윤(연세대학교 미래캠퍼스)

B-8-2 감성-VMD-기술지표 융합 cAR-GAN 합성데이터와 주가 로그수익률 예측 성능 검증 139
이지훈(순천향대학교), 김재윤(순천향대학교)

B-8-3 마스크 조건TimeGAN 기반 비정규 센싱 데이터의 정규 시계열 복원 141
김현정(국립한밭대학교), 이연서(국립한밭대학교), 김파란하늘(국립한밭대학교), 현장훈(국립한밭대학교)

B-8-4 연속학습의 자율주행 엣지 스트리밍 적응성 분석 143
이지은(숙명여자대학교), 홍서희(숙명여자대학교), 김지수(숙명여자대학교), 김윤희(숙명여자대학교)

B-8-5 ChromaTag: Delta E 색차 제어 기반의 투명 마커 포즈 추정 145
박유진(국립한밭대학교), 박천음(국립한밭대학교), 김은경(국립한밭대학교)

B-8-6 이미지 텍스처 특징에 기반한 센싱 데이터 이상탐지 연구 147
최용훈(상명대학교), 김민서(상명대학교), 김민석(상명대학교)

논문 발표 세부 안내

학부생 논문발표 B-2. 대규모언어생성AI 2

09월 25일(목) 15:30-16:30(60분) | 좌장: 박천음 교수 (한밭대)

B-2-1 인공지능 기반 외국인 노동자 근로 계약서 번역 및 법률 정보 제공 앱 149
정민성(국립한밭대학교), 황경호(한밭대학교), 옥주용(한밭대학교), 육종범(한밭대학교),
이수현(국립한밭대학교), 강윤서(국립한밭대학교)

B-2-2 온 프레미스(On-Premise) 환경에서의 MCP(Model Context Protocol)를 활용한 범용 Text-to-SQL
에이전트 설계 및 구현에 관한 연구..... 151
김지섭(동서울대학교), 김동우(동서울대학교), 이재희(동서울대학교)

B-2-3 RAG-LLM을 활용한 고위험 환경 시계열 이상 감지: 프롬프트 최적화와 의사 이상 데이터 생성 153
최하진(상명대학교), 박지원(상명대학교), 김민석(상명대학교)

B-2-4 음성 지원 대화형 AI 서비스를 활용한 정서 케어 지원 155
유영호(서울시립대학교), 김세현(서울시립대학교), 박도현(서울시립대학교), 홍수지(서울시립대학교),
이병정(서울시립대학교)

특별세션 S-1. 사이버 C.O.P 1

09월 25일(목) 15:30-16:30(60분) | 좌장: 이용태 교수 (한림대)

S-1-1 LLM 기반 합성 범죄수사 데이터 생성 시스템의 설계..... 157
김희두(고려대학교), 우병관(한림대학교), 문병훈(성균관대학교), 김지온(한림대학교)

S-1-2 수사 데이터 중복 식별을 위한 알고리즘 기반 처리 프레임워크 연구 160
문병훈(성균관대학교)

S-1-3 COVID-19 감염경로 분석을 위한 역학조사서 전처리 방법론에 관한 연구..... 163
김시원(한림대학교), 김지온(한림대학교)

S-1-4 AI 기반 Web Agent를 활용한 불법 데이터 자동 탐지 및 수집 연구 166
이수정(연세대학교), 이요셉(숭실대학교), 조용성(한국전자통신연구원), 송경우(연세대학교)

S-1-5 통신수사에서의 통화내역 분석과 SNA 기반 대포폰 탐지 알고리즘 연구 168
우병관(한림대학교), 김지온(한림대학교)

S-1-6 사이버범죄 수사 지원을 위한 온프레미스 LLM 기반 비정형 텍스트 핵심 정보 추출 170
윤병휘(한림대학교), 박현호(한국전자통신연구원), 김지온(한림대학교)

논문 발표 세부 안내

특별세션 FSO/OCC 특화연구실(국민대)

09월 25일(목) 15:30-16:30(60분) | 좌장: 장영민 교수 (국민대)

S-6-1 Call for Applications for IEEE 802.15 Interest Group on Next-Generation Optical Wireless Communication (IG NG-OWC) 172
Huy Nguyen, Yeong Min Jang (Kookmin University)

S-6-2 AI 기반 실시간 UAV Motor Error Detection 시스템 174
윤희창, 김상현, 유승호(부경대학교)

S-6-3 연합학습 기반 UAV 이상탐지 및 VC인증·ChaCha20 기반 보안 세션 아키텍처 176
박성수, 김기형 (아주대학교)

S-6-4 Non-IID 데이터 기반 연합학습에서 채널별 대칭 양자화에 따른 통신 효율성 및 정확도 분석 178
이동욱, 이웅희 (한성대학교)

S-6-5 K-공간 MRI에서 어텐션 메커니즘을 이용한 알츠하이머병 검출의 개념적 접근법 180
Belausava Anhelina, 김상철(국민대학교)

논문발표 A-21. 지능형산업응용AI 4

09월 25일(목) 16:30-17:30(60분) | 좌장: 김대년 박사 (경북IT융합산업기술원)

A-21-1 VAE-GMM 기반 생체모방 통신용 휘슬 시퀀스 생성 방법 182
박근호(국방과학연구소), 안종민(국방과학연구소), 김완진(국방과학연구소), 김인수(국방과학연구소), 이동훈(국방과학연구소)

A-21-2 국방 객체 데이터 부족 문제 해결을 위한 합성데이터 생성 184
김병현(엘아이지넥스원(주))

A-21-3 무선 영상 전송 위한 이중 대역 Deep JSCC 적용 기법 186
이승현(한국공학대학교), 오성현(한국공학대학교), 김정곤(한국공학대학교)

A-21-4 Understanding the Role of CutMix in ETF During Neural Collapse 188
Saqib Nazmus(제주대학교), 길준민(제주대학교)

A-21-5 I/Q-FFT 이중 표현을 활용한 ParallelCNN 기반 광대역 스펙트럼 센싱 모델 성능 분석 190
권순영(부산대학교), 김형남(부산대학교)

논문 발표 세부 안내

특별세션 S-2. 사이버 C.O.P 2

09월 25일(목) 16:30-17:30(60분) | 좌장: 김지온 교수 (한림대)

S-2-1 사이버범죄 수사지원 특화 대규모언어모델 개발을 위한 벤치마크 구축 및 보상체계 설계 192
계효선(한국전자기술연구원), 이미라(한국전자기술연구원), 김지형(한국전자기술연구원)

S-2-2 사이버범죄 수사 추론 에이전트를 위한 MCP 기반 도구 연동 플랫폼 개발 194
김민수(한국전자기술연구원), 김성섭(한국전자기술연구원), 권영민(한국전자기술연구원),
이승우(한국전자기술연구원)

S-2-3 이종 네트워크 연계 기반 HGT모델을 활용한 범죄 사건 내 피의자 후보군 판별 기술 연구 196
한현규(세종대학교), 백명선(세종대학교)

S-2-4 비지도 학습 및 그래프 신경망을 활용한 인물 계층 추론 방법 연구 198
구예리(동심우), 김지온(한림대학교), 우병관(한림대학교)

S-2-5 간첩통신에 활용된 스테가노그라피 특징과 수사기관의 법적쟁점 201
오소정(성균관대학교), 장지혜(성균관대학교), 김기범(성균관대학교)

논문발표 A-14. 인공지능응용 2

09월 25일(목) 16:30-17:30(60분) | 좌장: 이용민 박사 (ETRI)

A-14-1 저전력/고성능 Edge AI 위성 탑재체 개념 연구 203
이용민(한국전자통신연구원), 장동필(한국전자통신연구원)

A-14-2 Correlation Analysis of Readability Assessment Using EEG and Eye-Tracking Signals 205
윤세휘(모두의연구소), 송명주(모두의연구소), 황세미(연세대학교 원주연세의료원)

A-14-3 시맨틱 통신과 생성형 인공지능 융합에 관한 최신 연구 동향 208
이승찬(중앙대학교), 이동현(중앙대학교), 송치현(중앙대학교), 김재민(중앙대학교), 조성래(중앙대학교)

A-14-4 두피 이미지 분석을 위한 다중 라벨, 다중 클래스 분류 모델에 대한 예비 연구 210
유현선(연세대학교), 고동환(연세대학교), 이세영(연세대학교 미래캠퍼스), 김관(연세대학교 미래캠퍼스)

논문 발표 세부 안내

학부생 논문발표 B-3. 인공지능응용 1

09월 25일(목) 16:30-17:30(60분) |좌장: 길준민 교수 (제주대)

B-3-1 CLIP 기반 의미론적 지도를 활용한 SAR-RGB 이미지 변환 기법 연구 212
전지훈(한국항공대학교), 길태영(한국항공대학교), 김경태(한국항공대학교), 김선옥(한국항공대학교), 정재훈(한국항공대학교)

B-3-2 증상 검출 및 질병 분류 작업의 결합을 통한 효율적인 넙치 상태 진단을 위한 다중작업학습 접근법 214
김수영(충남대학교), 이명철(한국전자통신연구원)

B-3-3 시계열 이상탐지 성능 향상을 위한 마할라노비스 기반 점수화 기법 제안 216
김파란하늘(국립한밭대학교), 한재웅(한밭대학교), 김현정(국립한밭대학교), 이연서(국립한밭대학교), 현장훈(국립한밭대학교)

B-3-4 스펙트럼-정합 확산 샘플링을 통한이미지 초해상도 연구 218
전승재(한밭대학교), 권진경(한밭대학교), 전윤호(국립한밭대학교), 박천음(국립한밭대학교)

B-3-5 AI 기반 맞춤형 아동 성장 관리 서비스 설계 및 구현 220
KounJaeYoung((주)신라시스템), 배민중((주)신라시스템), 윤수현((주)신라시스템), 백성하((주)심플한)

B-3-6 AI 기반 보이스피싱 시뮬레이션을 활용한 개인 맞춤형 사이버보안 교육 방법론 연구 222
홍윤이(한림대학교), 최다온(한림대학교), 양윤모(한림대학교), 박지원(한림대학교), 박성미(한림대학교)

2025년 09월 26일(금)

논문발표 A-2. 대규모언어생성AI 2

09월 26일(금) 09:00-10:00(60분) | 좌장: 김정곤 교수 (한국공학대)

A-2-1 Memory-aware RAG 기반 접근을 통한 언어모델의 장기 기억 강화 225
장효석(과학기술연합대학원대학교), 이상철(한국과학기술연구원), 조한얼(세종대학교), 김찬수(한국과학기술연구원)

A-2-2 비전 LLM을 활용한 차세대 X-ray 보안 검색 시스템 개발 227
김민종(한국원자력연구원), 박형달(AIRISS CO., Ltd.)

A-2-3 An Analysis of the Value Bias of Large Language Models based on the PERMA Theory 229
Chaewon Kim(연세대학교), Keummin Ka(연세대학교), Jinhong Jeong(연세대학교), Seungbeen Lee(연세대학교), Youngjae Yu(서울대학교)

A-2-4 스타일 특징 추출 기법 융합을 통한 미술 화풍 분류 RAG-LLM 시스템 232
오현찬(한국전자기술연구원), 김용화(한국전자기술연구원), 윤상필(한국전자기술연구원), 김영민(한국전자기술연구원)

A-2-5 도메인 특화 LLM 지식 내재화 평가 프레임워크 234
양동헌(한국과학기술정보연구원), 임찬욱(한국과학기술정보연구원)

A-2-6 Reasoning LLM 내재적 추론 응답 품질 평가 프레임워크 236
이재용(한국과학기술정보연구원), 양동헌(한국과학기술정보연구원), 임찬욱(한국과학기술정보연구원)

논문 발표 세부 안내

논문발표 A-15. 인공지능응용 3

09월 26일(금) 09:00-10:00(60분) | 좌장: 김영선 박사 (한국전기연구원)

A-15-1 디지털트윈 컨소시엄을 중심으로 한 북미지역의 AI 기술 적용 동향에 관한 연구 238
김영선(한국전기연구원), 오휘명(한국전기연구원), 손상우(한국전기연구원)

A-15-2 과학기술 도메인 특화 언어모델의 Instruction Tuning 기반 성능 향상 240
정창후(한국과학기술정보연구원), 장광선(한국과학기술정보연구원), 양동헌(한국과학기술정보연구원),
임찬욱(한국과학기술정보연구원)

A-15-3 디지털 트윈 기반 네트워크 최적화 연구 동향 241
김주영(중앙대학교), 김재민(중앙대학교), 김가현(중앙대학교), 최성진(중앙대학교), 조성래(중앙대학교)

A-15-4 이미지 검색 애플리케이션을 위한 차원 축소 기능을 갖춘 EfficientNet 모델 243
Ghazal Saadloonia(조선대학교), 아흐마드 이자즈(고려대학교), 신석주(조선대학교)

A-15-5 NISQ 시대에 인공지능을 위한 양자-고전 하이브리드 컴퓨팅 소프트웨어 스택 분석 및 발전방향 245
김남규(한국과학기술정보연구원), 전인호(한국과학기술정보연구원)

A-15-6 안무 동영상의 출력 품질 향상을 위한 Attention 구조에 관한 연구 247
박노갑(SK텔레콤(주))

특별세션 S-10. 금오공대 MOSA

09월 26일(금) 09:00-10:00(60분) | 좌장: 김동성 교수 (금오공대)

S-10-1 MOSA기반 전기적 인터페이스 표준화에 대한 전략적 접근 249
권기협(국립금오공과대학교), 김재우(국립금오공과대학교), 김동성(국립금오공과대학교)

S-10-2 MOSA 기반 무인체계용 엣지 AI 통합 프레임워크 설계 251
김재우(국립금오공과대학교), 권익현(국립금오공과대학교), 이재민(국립금오공과대학교),
김동성(국립금오공과대학교)

S-10-3 K-PART 기반 K-MOSA 평가체계와 AI 활용 평가 소프트웨어 개발의 정책적 필요성 253
안재광(ICT Convergence Research Center), 김동성(국립금오공과대학교), 이재민(국립금오공과대학교),
김정은(국립금오공과대학교)

S-10-4 국방 AI 자율 소프트웨어 정책에 관한 고찰: 사례 기반 분석과 정책과제 255
이다정(국립금오공과대학교), 권익현(국립금오공과대학교), 김동성(국립금오공과대학교)

S-10-5 MOSA 표준군에서의 AI 통합 가능성 비교: 인터페이스·데이터·모듈성 257
류원재(ICT-CRC), 이재민(국립금오공과대학교), 김동성(국립금오공과대학교)

논문 발표 세부 안내

학부생 논문세션 B-9. 지능형산업응용2

09월 26일(금) 09:00-10:00(60분) | 좌장: 김봉완 박사 (ETRI)

B-9-1 VLM을 활용한 자율주행 차량의 장애물 개척 의사결정 시스템 259
서원준(울산과학기술원), 최승찬(울산과학기술원), 김휘순(울산과학기술원), 조성현(포항공과대학교), 배준혁(육군사관학교), 이용준(육군사관학교), 함형찬(한국과학기술원), 이은재(한국과학기술원), 김종환(육군사관학교), 안희진(한국과학기술원)

B-9-2 Reddit 소셜미디어 데이터를 활용한 밈주식의 Contrarian Effect 분석 261
최서빈(서강대학교), 김건희(서강대학교), 황유진(서강대학교)

B-9-3 실시간 ESS 화재 예측을 위한 Convolutional Autoencoder 263
이연서(국립한밭대학교), 김현정(국립한밭대학교), 김파란하늘(국립한밭대학교), 현장훈(국립한밭대학교)

B-9-4 히트맵 기반 적응형 이미지 슬라이싱을 통한 소형 객체 탐지 : 건설현장 적용 사례 265
한일주(충남대학교), 장다연(성결대학교), 김광수(한국전자통신연구원), 김봉완(한국전자통신연구원)

B-9-5 엣지 컴퓨팅 환경인 Raspberry Pi 5에서 가역 신경망 HiNet 이미지 은닉모델의 구현에 관한 연구..... 267
김정훈(동서울대학교), 김세빈(동서울대학교), 정준(동서울대학교)

논문발표 A-12. 영상신호AI 2

09월 26일(금) 10:10-11:10(60분) | 좌장: 윤상민 교수 (국민대)

A-12-1 Minimal-Step Diffusion 기반 초경량 TinyUNet의 해상도 복원 성능 분석..... 269
황병오(서울대학교), 이호성(서울대학교), 김다산(서울대학교), 강명주(서울대학교)

A-12-2 주사투과전자현미경 이미지 원자 위치의 딥러닝 기반 회귀 분석 271
이승환(한국기초과학지원연구원), 이상철(한국기초과학지원연구원)

A-12-3 투과전자현미경 원자분해능 이미지의 딥러닝 기반 정량 분석을 위한 AI 학습 데이터 생성 시뮬레이터 설계 ... 273
이태영(한국기초과학지원연구원), 이상철(한국기초과학지원연구원)

A-12-4 광 카메라 통신 기반 위치 추정을 위한 딥러닝 객체 탐지 알고리즘 비교: YOLOv5와 YOLOv8 275
김주현(영남대학교), 황나영(영남대학교), 정성윤(영남대학교)

A-12-5 Speech Pause Detection Using CTC Forced Alignment 277
구다연(광주과학기술원), 박화영(광주과학기술원), Hong Kook Kim(광주과학기술원)

A-12-6 이미지 복원을 위한 구조 및 의미 일관성 딥러닝 모델 279
이동규(경북대학교), 한동석(경북대학교)

논문 발표 세부 안내

논문발표 지능형산업응용 5

09월 26일(금) 10:10-11:10(60분) | 좌장: 김용강 교수 (공주대)

A-22-1 자연 발화 환경에서의 통합 표현 융합 기반 음성 감정 인식 프레임워크 282
전준일(건국대학교), 전효진(건국대학교), 김은이(건국대학교)

A-22-2 Feasibility of Reweighted A* as SSSP for Efficient MAPF via Sorting Barrier Breakthrough 284
김진호(성균관대학교), 안예찬(성균관대학교), 송용현(성균관대학교), 국태용(성균관대학교)

A-22-3 복합 시각 특성을 활용한 AI 차선 손상률 예측 모델 286
한준규(과학기술연합대학원대학교), 김민석(과학기술연합대학원대학교), 류승기(한국건설기술연구원)

A-22-4 Massive MIMO 시스템의 Fixed Point Network 알고리즘 연구 동향 288
송치현(중앙대학교), 김재민(중앙대학교), 조성래(중앙대학교)

A-22-5 O-RAN에서의 트래픽 관리 및 자원 할당에 대한 조사 290
최성진(중앙대학교), 원동욱(중앙대학교), 김재민(중앙대학교), 김주영(중앙대학교), 조성래(중앙대학교)

A-22-6 동적 그래프 학습 기반 통신 트래픽 예측 연구 동향 조사 292
김재민(중앙대학교), 오준석(중앙대학교), 김가현(중앙대학교), 송치현(중앙대학교), 조성래(중앙대학교)

특별세션 S-3. 경북대-ETRI 미래 디지털 융합 기술

09월 26일(금) 10:10-11:10(60분) | 좌장: 송윤정 박사 (ETRI)

S-3-1 분해된 보행 패턴을 활용한 뱀 로봇 강화학습의 탐색 공간 축소 294
송봉섭(한국전자통신연구원), 권우경(한국전자통신연구원), 남승우(한국전자통신연구원),
정윤수(한국전자통신연구원)

S-3-2 로봇 매니퓰레이터의 동작 계획을 위한 모델 예측 경로 적분 제어의 적응형 샘플링 접근 296
유지민(경북대학교), 강동엽(한국전자통신연구원), 동지연(한국전자통신연구원), 박찬은(경북대학교)

S-3-3 다중 호흡 신호 및 딥러닝 기반 비침습 운동 유발 지방 대사 평가 알고리즘 298
이준영(한국전자통신연구원), 이병주(한국과학기술원), 노효웅(한국전자통신연구원), 한영기(경북대학교),
이대식(한국전자통신연구원)

S-3-4 MobileNetV2를활용한궤양성대장염중증도예측: 온디바이스적용을위한경량화접근 300
김범휘(한국전자통신연구원), 우동원(경북대학교병원), 정성문(경북대학교)

S-3-5 추종로봇을 위한 UWB-RGB 센서 융합 기반의 사용자 위치 추정 방법 302
최낙원(한국전자통신연구원), 이준구(한국전자통신연구원), 오지용(한국전자통신연구원)

S-3-6 소프트웨어 정의 카메라 통합관제시스템 개발 304
김진홍(한국전자통신연구원), 최윤원(한국전자통신연구원), 백장운(한국전자통신연구원),
임길택(한국전자통신연구원), 김광주(한국전자통신연구원)

S-3-7 Sub-data Reloader for Weather Removal 306
김정환(경북대학교), 한동석(경북대학교)

S-3-8 데이터마이닝 기반 과불화화합물 노출 간암세포주(HepG2)의 차등 발현 유전자 및 생물학적 경로 분석 연구 308
이현정(한국전자통신연구원), 양소영(한국전자통신연구원), 임길택(한국전자통신연구원),
송윤정(한국전자통신연구원), 최은창(한국전자통신연구원)

S-3-9 초보자 접근성 향상을 위한 모바일 로봇 티치 펜던트 UX 디자인 프레임워크 연구 310
안지훈(한국전자통신연구원), 동지연(한국전자통신연구원)

학부생 논문발표 B-4. 인공지능응용 2

09월 26일(금) 10:10-11:10(60분) | 좌장: 김상철 교수 (국민대)

B-4-1 특징 공간 기반 중심 제약 손실 함수를 활용한 GAN의 품질 및 다양성 향상 312
서상원(국립한국교통대학교)

B-4-2 엣지 컴퓨팅 환경에서 FastSAM-s 모델의 구현 및 성능 비교에 관한 연구 314
김동우(동서울대학교), 황준석(동서울대학교), 강인경(동서울대학교)

B-4-3 엣지 컴퓨터 환경(Raspberry Pi5)에서 BitNet-b1.58 모델과 Gemma3 모델의 구현 및 성능비교에 관한 연구 ... 316
이승환(동서울대학교), 박성준(동서울대학교), 김정훈(동서울대학교)

B-4-4 녹내장 진단을 위한 ViT 기반 모델의 해석 가능성 비교: CNN과의 시각적 해석력 차이 분석 및 Occlusion, Grad-CAM 기법 활용 318
유호명(서강대학교), 하정욱(서강대학교)

B-4-5 FiLM 기반 CNN을 활용한 Website Fingerprinting 공격 모델 연구 321
장채은(한밭대학교), 김은경(국립한밭대학교)

B-4-6 YOLOv8-pose 기반의 실시간 스쿼트 자세 분석 및 교정 시스템 구현 323
정재성(국립한밭대학교), 신은호(국립한밭대학교), 전윤호(국립한밭대학교)

논문발표 A-7. 바이오AI

09월 26일(금) 11:10-12:10(60분) | 좌장: 김평수 교수 (한국공학대)

A-7-1 다계층 네트워크 분석 기반의 갑상선기능항진증 대응 후보 한약 및 핵심기전 도출 325
한예원(한국한의약진흥원), 이원융(원광대학교), 한석훈(원광대학교), 박찬림(에스에스엘), 김상진(한국한의약진흥원)

A-7-2 딥러닝 기반 노이즈 제거를 통한 장기 혈당 예측 안정성 향상 327
김수진(영남대학교), Sungho Kim(영남대학교), 정성윤(영남대학교)

A-7-3 AI와 인간의 신뢰 관계: 심리학적 관점에서의 분석 329
정주영(무소속)

A-7-4 중환자실 환자의 약물 처방 패턴을 반영한 장기 인공호흡 예측 모델 개발 333
배지원(건양대학교병원), 이승희(건양대학교병원)

A-7-5 심탄도 신호를 이용한 심전도/광용적맥파 복원: 손실 함수 설계와 성능 분석 335
박웅(한림대학교), 김진원(한림대학교), 이성한(한림대학교), 정인철(한림대학교)

논문발표 A-9. 사이버보안AI 2

09월 26일(금) 11:10-12:10(60분) | 좌장: 백명선 교수 (세종대)

A-9-1 A study on cyberattack targeted for semantic communication systems 337
Mehwish Ali Naqvi(동국대학교), 손인수(동국대학교)

A-9-2 FMCW 레이더를 이용한 걸음걸이 기반 신원 식별 및 침입자 탐지 시스템 339
조희재(부산대학교), 박상윤(부산대학교), 이정훈(부산대학교), 전우진(부산대학교), 전민욱(부산대학교), 김형남(부산대학교)

A-9-3 Vision-Language Pre-training 모델에서의 전이 가능성 기반 적대적 공격 연구 동향 341
김용재(한양대학교 에리카), 송민혁(한양대학교), 이연준(한양대학교)

A-9-4 Knowledge Graph 기반 RAG를 활용한 LLM 취약점 탐지 방법론 343
황시준(한양대학교), 이연준(한양대학교)

A-9-5 엣지 기반 정확도 검증을 통한 계층적 연합학습의 악의적 클라이언트 방어 기법 345
김건호(제주대학교), 정현수(제주대학교), 길준민(제주대학교)

학부생 논문발표 B-5. 인공지능응용 3

09월 26일(금) 11:10-12:10(60분) | 좌장: 한연희 교수 (한국기술교육대)

B-5-1 범용 GPU 환경에서 YOLOv8 모델 크기별 양자화 기반 최적화 성능 분석 347
박경택(경북대학교), 한동석(경북대학교)

B-5-2 실내 위치 추정을 위한 우도 기반 동적 데이터 가지치기 349
이서인(충남대학교), 문하은(충남대학교), 한승우(충남대학교), 김민석(충남대학교)

B-5-3 제지공정 변동성 반영을 위한 Neural SDE 기반 강화학습 연구 351
유동건(국립한밭대학교), 이상금(한밭대학교), 도윤미(한국전자통신연구원), 최진영(한국전자통신연구)

B-5-4 가지치기-양자화 결합을 통한 온디바이스 경량화 실증 연구 353
최상아(이화여자대학교), 권소현(이화여자대학교), 이주홍(이화여자대학교), 박형곤(이화여자대학교)

B-5-5 조건부 Diffusion Model 기반 전처리 및 증강을 통한 수중 객체 탐지 향상 연구 355
순동현(대구경북과학기술원), 이태복(국민대학교)

B-5-6 OptimABS: 절댓값 기반 적응적 옵티마이저의 연산 효율성 개선 358
이건수(한국항공대학교), 김정훈(한국항공대학교), 김지홍(한국항공대학교), 정재훈(한국항공대학교)

논문발표 A-25. AI안전및정책

09월 26일(금) 11:10-12:10(60분) | 좌장: 윤주상 교수 (동의대)

A-25-1 정렬 원칙 기반 범용 AI 위험 평가의 실증적 타당성 검증 360
신예진(한국정보통신기술협회), 강상연(한국정보통신기술협회)

A-25-2 변화하는 디지털 전장: 인공지능 적용의 확대 362
차용준(한국과학기술연구원), 조한얼(세종대학교), 이상철(한국과학기술연구원), 김찬수(한국과학기술연구원)

A-25-3 산불 발화점 추정을 위한 디지털 트윈 기반 감시 프레임워크 366
김동영(한국전자통신연구원), 김상원(한국전자통신연구원), 장인수(한국전자통신연구원),
김광주(한국전자통신연구원), 이경오(한국전자통신연구원)

A-25-4 소수자 게임의 집단 효용 극대화: 강화학습의 보상 균형화 기반 신속 안정화 정책 368
Seung-ho Baek(한국과학기술연구원), 이상철(한국과학기술연구원), 조한얼(세종대학교),
김찬수(한국과학기술연구원)

A-25-5 공공 AI 사업 상호호환성 확보를 위한 요구사항에 관한 연구 370
박눈솔(한국지능정보사회진흥원), 최한경(한국지능정보사회진흥원), 양진규(한국지능정보사회진흥원),
김동희(한국지능정보사회진흥원)

논문 발표 세부 안내

논문발표 A-6. 멀티모달AI 2

09월 26일(금) 14:00-15:00(60분) | 좌장: 김민석 교수 (상명대)

A-6-1 스마트 플랜트 가상환경 구축을 위한 모바일-서버 연동형 AI 기반 메타정보 자동 생성 프레임워크.....372
조윤호((주)올포랜드), 김세현((주)올포랜드), 최인하((주)올포랜드), 김승엽((주)올포랜드)

A-6-2 오류 누적 완화를 위한 의사결정 트랜스포머-흐름 매칭 통합 구조.....374
김민준(한국기술교육대학교), 지창훈(한국기술교육대학교), 최요한(한국기술교육대학교),
Nurlanbek kyzy Asel(한국기술교육대학교), 한연희(한국기술교육대학교)

A-6-3 자산가격 모델링에서 효율적 트랜스포머 구조와 데이터 정규화에 대한 연구.....376
김규진(Korea University Graduate School of SW·AI convergence), 김동주(고려대학교)

A-6-4 실시간IPTV방송스트림 기반 멀티모달 상품정보 추론 및 스코어링 시스템 개발.....378
배주한(SK텔레콤(주)), 허재호(SK텔레콤(주)), 주현철(SK텔레콤(주)), 나태영(SK텔레콤(주))

A-6-5 문맥 인지 적응형 마스크 선택을 통한 일관된 참조 비디오 객체 분할.....381
서준혁(경북대학교), 윤영우(경북대학교), 최원준(경북대학교), 이동규(경북대학교)

논문발표 A-23. 지능형산업응용AI 6

09월 26일(금) 14:00-15:00(60분) | 좌장: 김은경 교수 (한밭대)

A-23-1 스포츠 중계 스트림에서 효과적인 중간 광고 구간 검출 기법.....383
주현철(SK텔레콤(주)), 허재호(SK텔레콤(주)), 배주한(SK텔레콤(주)), 나태영(SK텔레콤(주))

A-23-2 다축 평면 투영 기반 3D 객체 탐지를 활용한 ROI 보존형 포인트 클라우드 다운샘플링.....385
임채윤(명지대학교), 박현희(명지대학교)

A-23-3 물체 분류를 위한 엣지 AI 기반 매니퓰레이터 동작 제어 연구.....387
김민경(동의대학교), 정준호(동의대학교), 주우현(동의대학교), 윤주상(동의대학교)

A-23-4 엣지 디바이스 인터워킹을 활용한 SDV 서비스 재구성 기법 연구.....389
권태현(동의대학교), 성은산(동의대학교), 윤주상(동의대학교

A-23-6 교통약자의 보행 안전을 위한 경사도 기반 보행 취약지역 분석 및 최적 경로 탐색에 관한 연구.....391
임윤정(차세대융합기술연구원), 송규원(차세대융합기술연구원), 박호건(성균관대학교)

논문발표 A-3. 대규모언어생성AI 3

09월 26일(금) 14:00-15:00(60분) | 좌장: 김동완 교수 (동아대)

A-3-1 다중 LLM 상호 검증을 통한 선행 연구 검색 및 분석에서의 AI 환각 완화에 관한 연구 393
정한민(한국과학기술정보연구원)

A-3-2 ZCD: LVLM 환각을 완화하는 Zero-Image Contrastive Decoding 기법 395
하지우(대구경북과학기술원), 소진현(대구경북과학기술원)

A-3-3 LLM의 만진법 수치 처리 향상을 위한 전처리 방법론(Preprocessing Methodology for Enhancing LLM Numeric Processing in Ten-Thousand System) 398
김혜련(KODATA), 목경서(KODATA), 이호진(KODATA), 이래형(KODATA)

A-3-4 테스트 출력 예측을 위한 LLM 친화적 의사코드 400
김재진(서울대학교), 황승원(서울대학교)

A-3-5 Multi-hop 질의응답을 위한 Five WS 기반 RAG Retrieval 기법 403
정일균(고려대학교)

A-3-6 LLM 기반 최적화 모델 자동 생성 파이프라인 406
강윤환(한국과학기술연구원), 김찬수(한국과학기술연구원), 조한얼(세종대학교), 이상철(한국과학기술연구원)

학부생 논문발표 B-6. 인공지능응용 4

09월 26일(금) 14:00-15:00(60분) | 좌장: 정민영 교수 (안산대)

B-6-1 SISEON: AIoT 인체공학 디스플레이 자동 제어 시스템 408
배제언(삼성 청년 SW·AI 아카데미), 차성현(삼성 청년 SW·AI 아카데미), 유형탁(삼성 청년 SW·AI 아카데미), 최인혁(삼성 청년 SW·AI 아카데미), 최혁진(삼성 청년 SW·AI 아카데미), 김형석(삼성 청년 SW·AI 아카데미), 이상현(삼성 청년 SW·AI 아카데미)

B-6-2 터널 내 딥러닝 기반 BLE RSSI 차량 측위 410
정기원(충남대학교), 김민석(충남대학교)

B-6-3 SEA-Unet v2: 초음파 영상에서의 스펙클 노이즈 제거 412
이채령(건양대학교), 박지희(건양대학교), 전소연(건양대학교), 최동혁(건양대학교)

B-6-4 우수 관망 시뮬레이션을 활용한 맨홀 수위 예측 모델 연구 414
주우현(동의대학교), 성은산(동의대학교), 이태진((주) 브로드웨이브), 윤주상(동의대학교)

B-6-5 차트 패턴 형성 전조의 비지도 유형화와 확률적 매핑을 통한 중장기 주가 방향 예측에 관한 연구 416
김진주(순천향대학교), 정민찬(순천향대학교), 김재윤(순천향대학교)

B-6-6 사용자 피드백을 반영한 스마트 조명 목표 디밍 레벨 예측 모델 418
김서연(국립한밭대학교), 이준희(한국전자통신연구원), 김대호(한국전자통신연구원), 성정식(한국전자통신연구원), 전윤호(국립한밭대학교)

논문발표 A-4. 대규모언어생성AI 4

09월 26일(금) 15:00-16:00(60분) | 좌장: 황인태 교수 (전남대)

A-4-1 생성형 AI 저작권 침해 대응을 위한 LangGraph 기반 LLM 에이전트 설계 420
정현우(아주대학교), 최윤서(아주대학교), 최동훈(아주대학교), 손경아(아주대학교)

A-4-2 sLLM 양자화의 성능-효율성 트레이드오프에 대한 실증적 분석 연구 422
김강우(경북대학교), 함승재(경북대학교), 박상재(경북대학교), 성민규(경북대학교), 강재모(경북대학교)

A-4-3 A Multiple LLM RAG Platform Using Answer Templates and Matching-Based Integration 424
Athiruj Poositaporn(한국과학기술정보연구원), 김재현(성균관대학교), 정한민(한국과학기술정보연구원)

A-4-4 전략적 의사결정에서 LLM의 편향과 유연성 탐색: 시나리오 기반 벤치마크 분석 426
최덕진(삼성전자)

A-4-5 LoRA 기반 LLM 미세 조정 기법을 통합한 연구 논문 분류에 관한 연구 429
Biswas Dipto(대구가톨릭대학교), 변태영(대구가톨릭대학교), 길준민(제주대학교), 조재춘(제주대학교)

A-4-6 Time-LLM 프레임워크를 통한 시계열 예측 성능에 대한 LLM의 영향 분석 432
공무경(경북대학교), 임수환(경북대학교), 방대호(경북대학교), 박민수(경북대학교), 강재모(경북대학교)

논문발표 A-24. 지능형산업응용 7

09월 26일(금) 15:00-16:00(60분) | 좌장: 김은경 교수 (한밭대)

A-24-1 Quantum-Hybrid Generative Flow Networks 434
이호성(서울대학교), 김다산(서울대학교), 황병오(서울대학교), 강명주(서울대학교)

A-24-2 온라인 콘텐츠의 부분 이미지 도용 탐지를 위한 특징점 기반의 효율적인 시스템 436
김다산(서울대학교), 황병오(서울대학교), 이호성(서울대학교), 강명주(서울대학교)

A-24-3 PPO를 활용한 APF 알고리즘 인력/척력 계수의 환경 적응형 자동 최적화 모델 438
이종훈(빅앤딥), 이세비(빅앤딥), 유현석(빅앤딥)

A-24-4 Plug-and-Play Bi-CoT : 설명가능한 Multi-Hop QA를 위한 자기인식 정방향 추론과 역방향 검증 440
KangJiHeon(연세대학교), 정수환(부경대학교)

A-24-5 제어 장벽 함수를 이용한 샘플링 기반 안전 필터 연구 442
박준영(한국과학기술원), 성현태(한국과학기술원), 안희진(한국과학기술원)

A-24-6 라이다 객체 검출을 위한 몬테카를로 드롭아웃 효과에 관한 연구 444
이종록(한국전자기술연구원), 박민철(한국전자기술연구원)

논문발표 A-16. 인공지능응용 4

09월 26일(금) 15:00-16:00(60분) | 좌장: 김민석 교수 (상명대)

A-16-1 실시간 복합음향 잡음 저감을 위한 WaveNet-VNNs 기반 ANC 시스템 설계 및 구현 446
이상민(영남대학교), 최승렬(인천대학교), 김태리(경북대학교), 이예지(상명대학교), 신애리(고려대학교 세종캠퍼스), 김우현(고려대학교), 유길상(고려대학교)

A-16-2 통신 음영 지역의 에너지 효율 향상을 위한 강화학습 기반 지능형 반사 표면 제어 기법 연구 448
김지형(한국전자통신연구원), 김정현(세종대학교)

A-16-3 DDQN을 활용한 NR-V2X 모드2의 동적 리소스 예약 조절 기법 450
김성현(경북대학교), 김채현(경북대학교), 양영준(경북대학교), 김동균(경북대학교)

A-16-4 이기종 가속기 기반 엣지 AI 시스템의 시간-공간 하이브리드 스케줄러 설계 452
김한결(한국전자기술연구원), 안재훈(Korea Electronics Technology Institute), 김영환(Korea Electronics Technology Institute)

A-16-5 EEG 신호를 활용한 그룹 대조 학습 기반 신경적응형 CBT 반응 생성 454
이진권(성균관대학교), 김효빈(성균관대학교), 황세미(연세대학교 원주연세의료원), 김지윤(모두의연구소), 송명주(모두의연구소), 오하영(성균관대학교)

학부생 논문발표 B-7. 인공지능응용 5

09월 26일(금) 15:00-16:00(60분) | 좌장: 장혜령 교수 (동국대)

B-7-1 MobileNetV2 기반의 딥러닝 모델을 활용한 넙치 다중 질병 분류 연구 456
이채령(건양대학교), 이명철(한국전자통신연구원)

B-7-2 최근 정보통신기술 One-Step Image-to-Image Translation의 원격주행용 Diffusion Model 기반 Semantic Compression 적용에 관한 연구 458
윤성빈(성균관대학교), 김예준(성균관대학교), 강민우(성균관대학교), 서효운(성균관대학교)

B-7-3 어안렌즈 카메라 캘리브레이션 전역 품질 지표에 관한 연구 460
김한구(경북대학교), 강신재(경북대학교), 차대웅(오토아이티(주)), 한동석(경북대학교)

B-7-4 Wi-Fi CSI 데이터를 활용한 재실 인원 감지 462
오태린(충남대학교), 김범석(충남대학교), 정기원(충남대학교), 김민석(충남대학교)

B-7-5 CSI 기반 실내 위치 추정을 위한 구역 선택형 전문가 혼합(RSMoE) 464
한승우(충남대학교), 문하은(충남대학교), 이서인(충남대학교), 김민석(충남대학교)

B-7-6 영상광용적맥파 기반 혈압 추정: 관심 영역 지정 여부에 따른 딥러닝 성능 평가 466
김연준(한림대학교), 김진원(한림대학교), 이성한(한림대학교), 정인철(한림대학교)

제6회 한국 인공지능 학술대회

The 6th Korea Artificial Intelligence Conference

Proceedings

래더링 인터뷰 기법을 활용한 대규모 언어 모델(LLM)의 정치적 성향 평가 및 분석 프레임워크 연구

장교철, 최재훈, 신민정, 이정, 서봉원
서울대학교

kyochul@snu.ac.kr, hoon95@snu.ac.kr, shinmj1024@snu.ac.kr, leejungp2@snu.ac.kr, bongwon@snu.ac.kr*

A Framework for Analyzing Political Knowledge Injection of LLMs(Large Language Models) Using the Laddering Interview Technique

Jang Kyochul, Shin Minjeong, Lee Jung, Choi Jaehoon, Suh Bongwon
Seoul National University

요 약

대규모 언어 모델(LLM: Large Language Model)은 사회 전반에 막대한 영향을 미치며, 특히 정치적 사안에 대한 편향된 답변은 여론에 큰 영향을 줄 수 있다. 기존의 단순한 질문 응답 방식으로는 모델이 내포한 복잡한 가치관과 정치적 성향을 심층적으로 평가하기 어렵기 때문에, 본 연구에서는 사다리 인터뷰(Laddering Interview) 기법을 도입하여 좌(진보)/우(보수)로 파인튜닝된 LLM 의 정치적 성향을 체계적이고 정량적으로 분석하고 평가하는 프레임워크 개발을 목표로 한다.

Ⅰ. 서 론

대규모 언어 모델(LLM: Large Language Model)은 사회 전반에 걸쳐 광범위한 영향을 미치고 있으며, 특히 정치적 편향이 담긴 답변을 통해 사용자의 정치적 인식과 여론 형성에 직접적인 영향을 줄 수 있다. [4] 그러나 기존의 단순 질의응답이나 키워드 분석 방식만으로는 LLM 이 내재한 복잡한 정치적 가치관과 판단 기준을 정확히 평가하는 데 한계가 있다.

이에 본 연구는 심리학 및 질적 연구 방법론에서 활용되는 사다리 인터뷰(Laddering Interview) 기법을 도입한다. 특히 좌(진보)/우(보수)로 다르게 파인튜닝된 LLM 을 대상으로 경제, 사회, 정부 역할, 외교안보, 환경 그리고 이민과 같은 주요 정치적 이슈에 대해 가치 충돌과 우선순위를 탐색하는 질문 은행(Question Bank)을 구축하고 이를 활용하여 분석을 수행한다. 본 연구는 미국의 좌우 성향을 기준으로 한다.

본 연구의 궁극적 목적은 LLM 의 정치적 성향을 명확하고 정량적으로 평가할 수 있는 분석 프레임워크를 개발하고, 이를 통해 사용자들이 LLM 에게 주입된 정치적 편향을 이해할 수 있게 하는데 있다.

Ⅱ. 본론

LLM 은 뉴스 생성, 소셜 미디어 콘텐츠 추천 등에서 활용되며, 사용자의 정치적 인식과 여론 형성에 영향을 미친다. LLM 이 좌/우 입장을 일관되게 취하면 사용자의 이념이 편향될 수 있다. 기존 연구는 질의응답이나 키워드 분석에 의존해 LLM 의 복잡한 가치 체계를 충분히 파악하지 못한다. [5] 이에 본 연구는 사다리 인터뷰 기법을 도입해 LLM 의 정치적 성향을 정교히 평가한다. [2] 사다리 인터뷰 기법을 통한 평가 방법은 다음과 같이 세가지로 나뉜다. **1 단계 (기초 질문):** 각 분야의 기본 입장을 묻는 6 개 질문을 제시한다. 예: "정부가 의료 및 에너지 산업에서 규제와 자유 시장을 어떻게 조정해야 한다고 보는가?" **2 단계 (상위 이유 탐색):** 1 단계 답변을 좌/중/우로 분류하고, 분류에

따라서 18(6*3)개 유형의 후속 질문 중 하나의 질문을 다시 LLM 에게 추론한다: "그 입장이 사회에 어떤 영향을 미친다고 보는가?" **3 단계 (가치 충돌/우선순위)**: 54(18*3)개 유형의 질문을 통해 가치 충돌 상황에서 우선순위를 묻는다. 예: "경제 성장과 공정성 간의 트레이드오프에서 무엇을 우선시하는가?".

질문 은행은 중립성을 유지하며, 기존 설문 자료를 참고한다. [3] 또한, 총 78 개 질문은 저자들과 협력하여 특정 기준에 따라 개선하였다. 개선 기준으로는 질문의 중립성(특정 성향을 유도하지 않는가?)과 명확성(모호하지 않은가?)을 적용한다.

이렇게 구성된 사다리 인터뷰 기법이 적용된 인터뷰 데이터셋에 기본 모델, exaone-7.8b, llama3-8b, minitron-8b, qwen3-8b 을 선행연구에서 제안한 좌우 성향 데이터로 파인튜닝하여 각 모델별로 좌/우 성향을 지니고 있는 총 8 개 모델을 생성하였다. [1] 이 모델들은 개발된 78 개 질문 데이터셋에 적용되어 추론을 수행하였다. 인터뷰는 1 단계부터 3 단계 질문을 순차적으로 적용해 6 개 인터뷰 스레드를 생성한다. 각 답변은 matous-volf/political-leaning-politics BERT 모델을 사용하여 좌(-1), 중(0), 우(1)로 분류하고, 가중치(w1=2, w2=1.5, w3=1)를 사다리 인터뷰 기법의 단계별로 적용해 최종 점수를 계산하여 모델을 정치적 성향에 따라 음수는 좌성향, 양수는 우성향으로 분류하였다.

<표 1> 사다리 인터뷰 기법을 기반으로 한 본 연구의 분류 결과와, 미국의 좌우 정치 스펙트럼을 기준으로 인간이 분류한 결과를 비교한 표. Ours 의 점수는 음수일수록 좌파, 양수일수록 우파 성향을 의미한다.

모델명	*훈련 성향*	*Ours*	*인간 평가*
Minitron-8b	Left	-2.92	L: 16, R: 1, C: 1
Minitron-8b	Right	4.50	L: 0, R: 18, C: 0
Exaone-7.8b	Left	-0.58	L: 12, R: 6, C: 0
Exaone-7.8b	Right	4.08	L: 0, R: 17, C: 1
LlaMA3-8b	Left	-3.92	L: 18, R: 0, C: 0
LlaMA3-8b	Right	3.83	L: 0, R: 18, C: 0
Qwen3-8b	Left	-4.50	L: 18, R: 0, C: 0
Qwen3-8b	Right	2.00	L: 4, R: 13, C: 1

제안한 모델 분류 기준에 따라 훈련된 모델의 추론 결과와 인간 평가 결과를 <표 1>에 요약하였다. 인간 평가 결과와 훈련 데이터의 성향이 완전히 일치하여, 모델이 해당 성향을 성공적으로 학습했음을 확인할 수 있다. 또한, 제안한 분류 기준을 적용하였을 때, 분류 결과 정확도는 100% 인 것을 <표 1>을 통해 확인할 수 있다. 앞서 설명한 바와 같이, 음수 값은 진보(좌) 성향, 양수 값은 보수(우) 성향을 의미하며, 절댓값은 해당 성향의 강도를 나타낸다. 우리가 개발한 분류 방법을 사용하여 모델에 주입된 정치적 성향을 정확하게 확인할 수 있을 뿐만 아니라, 그 성향이 주입된 정도까지 확인할 수 있다.

Ⅲ. 결론

본 연구는 사다리 인터뷰 기법을 활용하여 LLM 의 정치적 성향을 평가하고 분석하는 프레임워크를 제안한다. 경제, 사회, 정부의 역할 등 미국에서 주요 논쟁이 되는 6 개 분야에 대해 각 분야별로 각 성향별로(좌/중/우) 3 개의 질문을 포함한 총 78 개의 인터뷰 질문을 구성하고, 인간 검증을 통해 데이터셋을 정제하였다. 이후 미국의 좌우 성향 데이터셋을 활용하여 네 종류의 기본 모델을 파인튜닝함으로써 정치 성향이 반영된 편향 모델을 구축하였다. 인간 평가를 통해 훈련된 모델의 타당성을 검증하였으며, 본 논문에서 제안하는 평가 방식을 통해 모델 분류 및 분석 방법의 유효성을 추가적으로 입증하였다.

ACKNOWLEDGMENT

이 논문은 *2025*년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 **[NO.RS-2021-II211343,인공지능대학원지원(서울대학교)]**

참 고 문 헌

[1] Lin, Z., Wang, Y., & Zhang, H. (2024). *PoliTune: Analyzing and mitigating political bias in large language models through fine-tuning.* arXiv preprint arXiv:2404.08699. https://arxiv.org/abs/2404.08699

[2] Landfield, A. W. (1971). *Personal construct systems in psychotherapy*. Springfield, IL: Charles C. Thomas.

[3] Pew Research Center. (2014). *Political polarization in the American public.* https://www.pewresearch.org/politics/2014/06/12/appendix-a-the-ideological-consistency-scale/

[4] Motoki, F., Pilditch, T. D., & Savadori, L. (2023). *More human than human: Measuring ChatGPT political bias.* Public Choice, 198(1-2), 3-23. https://doi.org/10.1007/s11127-023-01097-2

[5] Jiang, H., Beeferman, D., Roy, B., & McKeown, K. (2022). *CommunityLM: Probing partisan worldviews from language models*. Proceedings of the 29th International Conference on Computational Linguistics, 6818-6826.

<부록 1>미국의 정치적 도메인별 좌파와 우파의 주요 특징 비교

도메인 (DOMAIN)	좌파 / 진보 (LIBERAL / PROGRESSIVE)	우파 / 보수 (CONSERVATIVE)
경제 (ECONOMY)	정부 개입 선호: 규제 강화, 부유층 고세율, 사회복지 확대. 경제 평등 및 노동자 보호 중시 (예: 민주당의 부자 증세, 노동조합 강화).	정부 개입 최소화: 규제 완화, 저세율, 자유시장 지지. 개인 책임 및 민간 번영 강조 (예: 공화당의 감세, 시장 중심 정책).
사회 및 문화 (SOCIETY & CULTURE)	사회적 진보, 다양성, 소수자 권리 중시. 성소수자, 인종/젠더 평등 정책 강조 (예: 다문화주의, 포용성).	전통적 가치 중시. 가족, 종교, 국가 정체성 보존 강조 (예: 애국심, 기존 규범).
정부 역할 (ROLE OF GOVERNMENT)	큰 정부 지지: 복지, 교육, 의료 등 정부 개입 확대. (예: 민주당 지지자 78%가 정부 확대 지지).	작은 정부 선호: 민간 중심 해법. 과도한 관료주의 비판 (예: 공화당 지지자 71%가 정부 축소 지지).
외교 및 안보 (FOREIGN POLICY & SECURITY)	동맹, 외교, 국제 협력 중시. 군사력보다 외교적 해법 선호.	국방력, 주권 중시. "아메리카 퍼스트" 강조, 강한 군사력 및 국경 안보 우선.
환경 및 기후 (ENVIRONMENT & CLIMATE)	기후변화 대응, 환경 규제 지지. 재생에너지, 국제 협력 중시 (예: 파리협정).	경제 성장 우선. 환경 규제 회의적, 경제 부담 우려.
이민 (IMMIGRATION)	합법적 이민 확대, 불법 이민자 보호 정책 지지.	불법 이민 억제, 국경 통제 중시. 법 집행 및 추방 정책 지지.

대규모 언어모델의 자기성찰 능력 향상을 위한 피드백 최적화 기법

김준석, 정교민
서울대학교
kim.junseok@snu.ac.kr, kjung@snu.ac.kr

Optimizing Feedback to Improve Self-Reflection in Large Language Models

Kim Jun Seok, Jung Kyo Min
Interdisciplinary Program in Artificial Intelligence, Seoul National Univ.

요 약

본 논문은 대규모 언어모델의 낮은 자기성찰(self-reflection) 성능의 원인을 피드백 품질 부족에서 찾고, 이를 개선하여 성능을 향상시키는 방법을 제안한다. LLM을 활용해 초기 답변, 피드백, 그리고 피드백 이후의 답변을 생성하고, 답변의 정답률을 기반으로 피드백 품질을 정량화한 뒤, 이를 바탕으로 선호도 데이터를 구축하여 DPO 방식으로 LLM을 학습시켰다. 학습된 LLM은 자기성찰 후 성능 향상에 기여했으며, 피드백 품질이 성찰 기반 문제 해결 성능 향상에 중요한 역할을 기여함을 확인하였다. 또한 트리 구조를 활용한 데이터 생성 방식으로 특정 도메인에 한정되지 않고, 다양한 과제에 적용 가능한 일반화된 자기성찰 향상 기법을 제시한다.

Ⅰ. 서 론

본 논문에서는 자기성찰(self-reflection) 능력을 향상시키기 위한 새로운 접근 방식을 제안한다. 자기성찰이란 문제를 푸는 과정에서 자신의 답변이 정답인지 여부를 스스로 평가하고, 그 결과에 따라 답변을 수정하는 과정을 의미한다. [1] 이는 인간에게 있어 자연스러운 사고 과정이지만, 기존의 대규모 언어모델(Large Language Models, LLMs)은 이러한 능력을 본질적으로 갖추고 있지 않다. 실제로 LLM이 자기성찰을 수행할 경우, 오히려 성능이 저하된다는 사실이 여러 선행 연구를 통해 밝혀져 왔다. [2]

이러한 문제를 해결하기 위한 시도로, 최근에는 LLM 간의 찬반 토론을 통해 자기성찰을 유도하는 방식이 제안되었다. [3] 그러나 이 방식은 높은 연산 비용을 요구하며, 반박을 생성하는 모델의 품질이 낮을 경우 오히려 잘못된 방향으로 자기성찰이 이루어질 수 있다는 한계가 있다.

이에 본 연구에서는 자기성찰 과정에서 생성되는 피드백의 품질을 높이는 데 초점을 맞추었다. 구체적으로는 LLM을 이용하여 초기 답변, 해당 답변에 대한 피드백, 그리고 피드백을 반영한 후속 답변을 순차적으로 생성함으로써 피드백 품질과 관련된 데이터셋을 구축하였다. 이후, 이 데이터셋을 기반으로 강화학습을 적용하여 초기 답변에 대해 보다 정확하고 유익한 피드백을 생성할 수 있도록 LLM을 학습시켰다.

양질의 피드백은 자기성찰의 효과를 극대화하는 핵심 요소이며, 본 연구는 이러한 피드백의 질적 향상을 통해 LLM의 자기성찰 기반 문제 해결 능력을 보다 효율적으로 강화하는 것을 목표로 한다.

Ⅱ. 본론

본 논문에서 제안하는 방법론은 크게 두 단계로 나뉜다. 첫 번째 단계에서는 피드백의 품질을 정량적으로 측정할 수 있는 데이터셋을 생성하고, 두 번째 단계에서는 이 데이터를 활용하여 LLM이 더 나은 피드백을 생성할 수 있도록 학습시킨다.

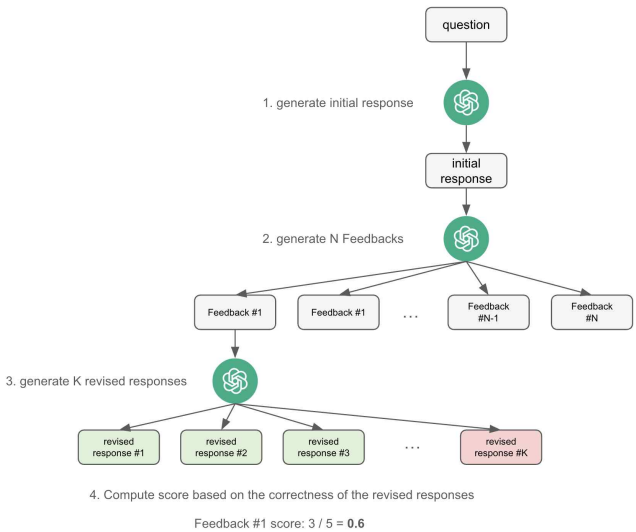


그림 1 피드백 품질 관련된 데이터 생성 방식

그림 1은 피드백 품질에 관련된 데이터를 생성하는 방식에 대해서 요약하고 있다. 먼저, 주어진 문제에 대해 LLM의 temperature 값을 1로 설정하여 초기 답변에 대한 다양한 피드백과 추후 답변을 얻어낼 수 있다. temperature 값을 높게 설정함으로써 보다 다

양한 방식의 답변을 유도할 수 있으며, 이는 후속 피드백과 자기성찰 답변의 다양성을 확보하는 데 도움이 된다. 생성된 초기 답변에 대해서, 동일한 LLM을 사용하여 n개의 피드백을 생성한다. 이 피드백은 각 초기 답변의 오류를 지적하거나 개선 방향을 제시하는 텍스트로 구성된다.

그 다음 단계에서는, 생성된 각 피드백에 대해 다시 LLM을 활용하여 n개의 자기성찰 이후 k개의 답변을 생성한다. 즉, 하나의 피드백이 있을 때, 이를 반영한 후속 답변을 여러 개 생성하여 피드백의 효과를 측정하는 방식이다. 이렇게 생성된 자기성찰 답변 중에서 정답을 맞춘 답변의 개수를 세고, 이를 k로 나눈 값을 해당 피드백의 품질 점수로 정의한다. 예를 들어, 하나의 피드백으로부터 생성된 5개의 자기성찰 답변 중 3개가 정답일 경우, 이 피드백의 점수는 0.6이 된다.

이러한 방식으로, 문제와 초기 답변, 피드백, 그리고 해당 피드백의 정답 유도율에 기반한 점수로 구성된 데이터셋을 구축할 수 있다. 이후 이 데이터셋에서 피드백 간의 점수를 비교함으로써 어떤 피드백이 더 나은지를 판단할 수 있으며, 이 비교 결과를 활용하여 선호도에 따른 데이터 쌍을 구성한다. 예를 들어, 두 개의 피드백 중 하나의 점수가 더 높다면, 그 피드백을 우선적으로 선호하는 데이터 쌍을 만들 수 있다.

이렇게 생성된 데이터를 바탕으로, LLM을 Direct Preference Optimization(DPO) 방식으로 학습시킨다. [4] DPO는 모델이 더 나은 출력을 생성하도록 학습하는 방식으로, 본 연구에서는 초기 답변과 피드백을 입력으로 받아 더 우수한 피드백을 출력하는 LLM을 만드는 데 사용된다. 학습된 모델은 동일한 초기 답변에 대해 더 높은 품질의 피드백을 생성할 수 있으며, 이는 자기성찰을 거친 최종 답변의 정확도를 향상시키는 데 기여한다.

모델	데이터셋	자가성찰 방법론	
		일반 모델	학습된 모델
llama-3.2-3B-Instruct	GSM8K	53.29	71.57 **(+18.28)**
	CSQA	36.12	64.54 **(+28.42)**
	CoinFlip	27.35	38 **(+10.65)**

표 1 피드백 생성 시 학습된 모델의 성능 비교

제안한 방법론의 유효성을 검증하기 위해, 기존 연구에서 널리 사용된 세 가지 추론 벤치마크에 대해 실험을 수행하였다. 특히 벤치마크 선정 기준으로는 수학, 상식, 기호적 추론 능력을 다루는 다양한 도메인에 대해서 평가하고자 하였다. 실험 모델로는 보편적으로 일반 컴퓨터에서도 사용할 수 있는 3B 크기의 LLM을 사용하였다. [5] 학습된 LLM을 사용하여 피드백을 생성하고, 이를 자기성찰에 반영한 후 최종 답변의 성능을 평가하였다. 실험 결과는 표 1에 제시되어 있으며, 학습 이전의 기존 LLM과 비교했을 때 학습된 LLM을 사용할 경우 피드백 이후의 성능이 일관되게 향상됨을 확인할 수 있었다. 이는 피드백의 품질이 자기성찰 성능에 직결되며, 본 연구에서 제안한 피드백 품질 향상 기법이 실제로 문제 해결 능력 향상에 효과적이라는 것을 보여준다.

Ⅲ. 결론

본 연구에서는 기존 LLM의 자기성찰 성능이 낮은 원인 중 하나로 피드백의 품질 부족에 주목하고, 이를 개선함으로써 자기성찰 이후의 최종 성능을 유의미하게 향상시킬 수 있음을 보였다. 단순히 자기성찰을 수행하는 것만으로는 성능 저하가 발생할 수 있으나, 본 논문에서 제안한 방식처럼 피드백의 품질을 정량적으로 평가하고 이를 학습에 반영할 경우, 성찰 과정 자체가 실질적인 도움이 될 수 있다는 점을 실험을 통해 입증하였다. 또한, 기존의 자기성찰 능력 향상 기법들이 주로 수학적 추론과 같은 특정 도메인에 집중되어 있었던 반면, 본 연구는 트리 구조를 활용한 데이터 생성 방식을 도입함으로써 도메인에 종속되지 않는 자기성찰 향상 방법론을 제안하였다. 이를 통해 수학 문제뿐만 아니라 보다 일반적인 자연어 처리 과제에서도 자기성찰 기반 성능 향상이 가능함을 확인하였다. 향후에는 본 연구에서 제안한 방법론을 다양한 태스크와 모델 아키텍처에 확장하여, LLM의 보편적인 자기성찰 능력 향상에 기여할 수 있을 것으로 기대한다.

ACKNOWLEDGMENT

이 논문은 *2025*년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 **[NO.RS-2021-II211343,인공지능대학원지원(서울대학교)]**

참 고 문 헌

[1] Madaan, A., Tandon, N., Gupta, P., Hallinan, S., Gao, L., Wiegreffe, S., ... & Clark, P. (2023). Self-refine: Iterative refinement with self-feedback.*Advances in Neural Information Processing Systems,36*, 46534-46594.

[2] Huang, J., Chen, X., Mishra, S., Zheng, H. S., Yu, A. W., Song, X., & Zhou, D. (2023). Large language models cannot self-correct reasoning yet.*arXiv preprint arXiv:2310.01798*.

[3] Liang, T., He, Z., Jiao, W., Wang, X., Wang, Y., Wang, R., ... & Tu, Z. (2024, November). Encouraging Divergent Thinking in Large Language Models through Multi-Agent Debate. In*Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*(pp. 17889-17904).

[4] Rafailov, R., Sharma, A., Mitchell, E., Manning, C. D., Ermon, S., & Finn, C. (2023). Direct preference optimization: Your language model is secretly a reward model.*Advances in Neural Information Processing Systems,36*, 53728-53741.

[5] Grattafiori, A., Dubey, A., Jauhri, A., Pandey, A., Kadian, A., Al-Dahle, A., ... & Vasic, P. (2024). The llama 3 herd of models.*arXiv preprint arXiv:2407.21783*.

Chain-of-Function: Inducing Sequential Plans Over Decomposed Code Indexes for Code Generation

Jongyoon Kim, Seung-won Hwang*
Seoul National University, Seoul, Republic of Korea

{ john.jongyoon.kim, seungwonh }@snu.ac.kr

Abstract

Large Language Models (LLMs) have shown performance gains in code generation tasks via Retrieval Augmented Generation (RAG). However, these methods, known as Code-RAG, implicitly rely on retrieving ground-truth code, so their gain mostly comes from this. This reflects the issue of the Code-RAG pipeline itself, which has not been designed for reusability. When the gold snippet is absent, LLMs easily fail to leverage the retrieved codes. To overcome this, we propose Chain-of-Function (CoF): treat code as a composition of reusable functions. We decompose the retrieval pool into individual functions and encode them as pseudocode for a modality match between code and problem description. At inference, we generate a pseudocode plan based on the problem description, decompose it into required functions, and retrieve each independently, enabling function reuse without a gold snippet. On LiveCodeBench, our method achieves up to a +2% pass rate over PERC and improves the reuse rate, the number of retrieved codes referred on generated code increases to 2.3% from 0.69%, with shorter retrieved context compared to PERC.

Ⅰ. Introduction

Retrieval-augmented generation (RAG) for code aims to improve large language models (LLMs) by supplying relevant external code during inference. While effective in some cases, recent benchmarks such as CodeRAG-Bench show that substantial gains occur mainly when the retrieved snippet is a *gold snippet*, a complete, ground-truth solution. Without such a snippet, models often ignore the retrieved content.

A key challenge highlighted by recent empirical studies is the **low reusability of retrieved code** [1,2]. Current pipelines index code at coarse granularity, such as entire code file or large code blocks, collapsing multiple uncrelated functions into a single emberdding. Even wehn retrievers find partially relevant snippets, models often struggle to disentagle useful components from the surrounding irrelevant code blocks. Real codebases are compositional: a solution can be expressed as $C = f_1 \circ f_2 \circ \ldots \circ f_n$, where each is a distinct, reusable function. When retrieval ignores this **chain-of-function** structure, relevant functions are buried in noise, making partial reuse difficult.

Our **chain-of-function retrieval** framework redefines the retrieval pool by decomposing code into function-level units, which are then represented as pseudocode and indexed independently. This enables partial reuse without relying on a gold snippet, improving both accuracy and reuse rate. The query process mirrors this decomposition: we generate a high-level pseudocode plan for the problem, break it down into functional steps, and use each step as a precise query to retrieve a corresponding function. These retrieved functions are then aggregated and supplied as context to an LLM to generate the final solution.

II. Related Works

Prior work in Code-RAG has advanced along two primary axes: retrieval indexing and query formation. Initial approaches to **indexing** typically treated entire files or large code blocks as single documents for retrieval [3], which often dilutes the semantic signal of individual functions. Recognizing this limitation, more recent work has explored finer-grained indexing strategies, such as retrieving at the function level [4] or leveraging structured representations like programming knowledge graphs [5].

Similarly, **query formation** techniques have grown more sophisticated, moving from simple similarity search [3] toward structured approaches like PERC (plan-as-query) [6] or iterative refinement [7,8]. However, despite these parallel advancements, a fundamental disconnect persists: the granularity of the retrieved items often fails to match the compositional nature of the code generation task. Our work directly addresses this gap by synchronizing both the retrieval and query processes at the function level, ensuring that each piece of retrieved information is a targeted, reusable component for the final code assembly.

Ⅲ. Methodology

A. Problem Formulation

In a standard Code-RAG setting, the goal is to generate a solution code C_s for a given problem description p, using a set of retrieved code contexts E. The generation is conditioned on a prompt that includes the task query (q) and the retrieved examples: $c_s = LLM(p, E;\ Prompt)$ The retrieved contexts E are selected from a large retrieval pool D by scoring each candidate $c \in D$ based on its similarity to a query q:

$$E = Top_{c \in D} sim(\psi(q), \psi(c))$$

where ψ is an encoding function. This formulation highlights the dependency on retrieving a highly relevant code snippet$c$.

B. Stage 1: Function-Level Retrieval Pool Construction

Our approach begins by reformulating the retrieval pool D. Instead of indexing entire code solutions, we use an LLM to refactor each code snippet (c) into a set of single-responsibility functions $\{ f_1, f_2, \ldots, f_n \}$. Each function is then converted into natural language pseudocode. This process builds a retrieval pool that focuses on the functional essence of the code rather than its syntax, enhancing retrieval accuracy.

* Corresponding author

B. Stage 2: Plan-Guided Function Induction and Retrieval

At inference, we first generate a high-level pseudocode plan for the given problem description (p). This plan is then decomposed into a list of required functions, $\{g_1, g_2, \ldots, g_n\}$. For each required function, we generate a detailed pseudocode description, which serves as a fine-grained query. Unlike PERC, that uses the entire plan as one query, our function-specific queries allow for the precise retrieval of relevant, reusable components.

IV. EXPERIMENTS AND RESULTS

A. Experimental Setup

- **Dataset:** LiveCodeBench [7] v5 (880 problems) with the programming-solutions corpus from CodeRAG-bench [8].
- **Models:** nomic-embed-text for retrieval; llama3.1-8B, 70B [9], and gpt-4o-mini [10] for generation.
- **Metric:** Pass@1, which measures the percentage of test cases passed by the generated solution.

B. Results

Table 1 shows that CoF consistently improves the Pass@1 score by +1.6% to 2% over the state-of-the-art PERC model across various LLMs. This indicates that CoF retrieves more relevant functions that are useful for solution generation.

	llama3.1-8b	llama3.1-70b	gpt-4o-mini
No Retrieval	23.64	36.05	52.61
Problem-as-query	22.84	36.7	53.52
PERC [6]	24.2	35.57	52.84
CoF (ours)	**25.8**	**37.62**	**54.02**

Table 1: Pass@1 score of LiveCodeBench [3] problem with programming-solution corpus from CodeRAG-Bench [2].

C. Analysis

Table 2 highlights CoF's superior reusability. CoF reused 2.3% of retrieved functions for solving 37 questions, compared to only 0.69% for PERC for 11 questions. As CoF retrieves more functions while shortening the length of the retrieved context, we can validate that the retrieved context from CoF enables LLMs to leverage the context more effectively.

	PERC [6]	CoF (ours)
Average Number of codes retrieved	3	**4.71**
Average length of retrieved codes	857.56	**677.83**
Number of retrieved codes referred	0.69% (11 questions)	**2.3% (37 questions)**

Table 2: Reusability Analysis between retrieved codes and generated codes.

V. Conclusion

We addressed the limitations of Code-RAG by decomposing the retrieval pool into pure, single-responsibility functions and aligning queries via pseudocode decomposition. Our CoF approach significantly enhances retrieval precision and solves more problems on a Pass@1 basis. Future work includes exploring the dynamic assembly of retrieved units via an agentic approach to close remaining performance gaps.

ACKNOWLEDGMENT

This work was partly supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) [NO.RS-2021-II211343, Artificial Intelligence Graduate School Program (Seoul National University)] and also supported by Electronics and Telecommunications Research Institute (ETRI) grant funded by ICT R&D program of MSIT/IITP (2022-0-00995, Automated reliable source code generation from natural language descriptions).

References

[1] W. Gu *et al.*, "What to Retrieve for Effective Retrieval-Augmented Code Generation? An Empirical Study and Beyond," *arXiv preprint arXiv:2503.20589*, 2025.

[2] Z. Z. Wang *et al.*, "Coderag-Bench: Can Retrieval Augment Code Generation?," *arXiv preprint arXiv:2406.14497*, 2024.

[3] F. Zhang *et al.*, "Repocoder: Repository-Level Code Completion Through Iterative Retrieval and Generation," *arXiv preprint arXiv:2303.12570*, 2023.

[4] Y. Ding *et al.*, "Function-Level Code Retrieval with Multi-Faceted Representations," in *Proc. 45th Int. Conf. Softw. Eng. (ICSE)*, 2023.

[5] I. Saberi and F. Fard, "Context-Augmented Code Generation Using Programming Knowledge Graphs," *arXiv preprint arXiv:2410.18251*, 2024.

[6] J. Yoo, H. Han, Y. Lee, J. Kim, and S.-W. Hwang, "PERC: Plan-As-Query Example Retrieval for Underrepresented Code Generation," *arXiv preprint arXiv:2412.12447*, 2024.

[7] D. Zan *et al.*, "When Language Models Meet Code: A Survey on Retrieval-Augmented Code Generation," *arXiv preprint arXiv:2312.17581*, 2023.

[8] W. Gu *et al.*, "What to Retrieve for Effective Retrieval-Augmented Code Generation? An Empirical Study and Beyond," *arXiv preprint arXiv:2503.20589*, 2025.

[7] N. Jain *et al.*, "Livecodebench: Holistic and Contamination Free Evaluation of Large Language Models for Code," *arXiv preprint arXiv:2403.07974*, 2024.

[9] AI@Meta, "Llama 3 Model Card," 2024. [Online]. Available: https://huggingface.co/docs/hub/en/model-cards

[10] J. Achiam *et al.*, "GPT-4 Technical Report," *arXiv preprint arXiv:2303.08774*, 2023.

작업별 중요도 기반 프루닝을 활용한 LoRA 어댑터 병합 기법

윤도경 [1], 서지원 [2,3], 조동현 [1*]
[1] 한양대학교, [2] 서울대학교, [3] 서울대학교 자동화연구소

sb0636@hanyang.ac.kr, seojiwon@snu.ac.kr, *doncho@hanyang.ac.kr

TIP: Task Importance-based Pruning for LoRA Adapter Merging

Dogyeong Yun[1], Jiwon Seo[2,3], Donghyeon Cho[1*]
[1]Hanyang University, [2]Seoul National University, [3]Seoul National University ASRI

요 약

본 논문은 다중 작업(multi-task) 환경에서 LoRA(Low-Rank Adaptation) 어댑터를 하나의 어댑터로 효율적으로 병합하기 위한 새로운 기법 TIP(Task Importance-based Pruning)을 제안한다. 기존 병합 방식은 정적인 sparsity 설정이나 단순 선형 결합에 의존함으로써, 레이어 간 중요도 차이를 반영하지 못하고 성능 저하를 유발하는 한계가 있다. 제안 기법은 입력 샘플 기반의 중요도 분석을 통해 레이어별 연산 기여도를 정량화한 뒤, 이에 따라 동적인 sparsity를 적용하고, WANDA[3] 기반 어댑터 pruning을 수행하여 병합을 진행한다. LLaMA 2 7B 모델과 총 7개 벤치마크 태스크에 대한 실험 결과, 제안한 방법은 기존 방식 대비 평균 +17.6% 높은 정확도를 기록하였으며, 특히 CoQA에서는 기존 연구 대비 최대 +43.8%, HellaSwag에서는 +31.2%의 상대적 성능 향상을 달성하였다.

Ⅰ. 서 론

대규모 언어 모델(LLM)은 다양한 자연어 처리 과제에서 뛰어난 성능을 보여주고 있다. 그러나 이러한 모델을 완전 미세조정(full fine-tuning)하는 데에는 막대한 계산 자원과 시간이 요구된다. 이를 해결하기 위한 대안으로, 효율적인 파라미터 튜닝 기법인 LoRA(Low-Rank Adaptation)[1]가 널리 활용되고 있다. 하지만 multi-task 환경에서 LoRA 를 적용할 경우, task 별로 개별 어댑터를 유지·로드해야 하므로, 추론 시 불필요한 오버헤드가 발생하고 시스템 효율성이 저하된다. (Fig. 1)

본 연구에서는 이러한 비효율성을 해소하기 위해, 서로 다른 task 에 특화된 LoRA 어댑터들을 하나의 어댑터로 병합하여 multi-task 를 단일 batch 내에서 효율적으로 처리할 수 있게 하는 방법인 TIP(Task Importance-based Pruning)을 제안한다. 특히, 기존의 단순 선형 결합이나 정적 pruning 기반 병합 방식과 달리, 본 방법은 task 별로 중요한 레이어를 식별한 뒤, 각 레이어에 대해 동적으로 sparsity 를 조정하는 방식을 통해 보다 정밀하고 효과적인 LoRA 병합을 가능하게 한다.

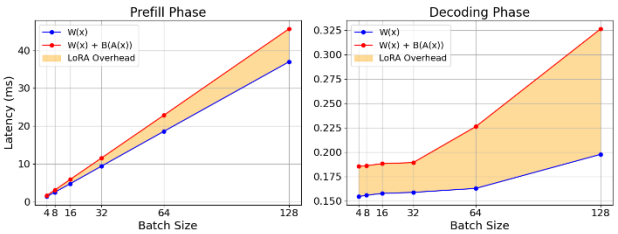


Figure 1: 배치 별 추가적인 LoRA 연산 오버헤드

Ⅱ. 관련 연구

Model Merge

기존 LoRA 어댑터 병합 방식 중 가장 단순한 접근은 선형 결합(linear combination)이다. 그러나 이 방식은 어댑터 간 파라미터 간섭(interference)을 유발할 수 있으며, 특히 task 간 feature 분포가 상이한 경우 병합 후 성능 저하가 발생할 수 있다.

DARE [2]는 병합 시 각 어댑터의 weight 를 무작위로 pruning 하고, 남은 weight 를 sparsity 비율에 따라 rescale 하여 병합함으로써 파라미터 간 간섭을 확률적으로 희석한다. 그러나 무작위 방식은 연산에 중요한 요소까지 제거할 수 있는 위험이 있으며, 모든 레이어에 동일한 sparsity 를 정적으로 적용하기 때문에 작업 간 특성이나 레이어별 중요도를 반영하지 못한다는 한계가 있다.

Ⅲ. 본론

3.1. Methods

본 연구에서는 task 간 레이어 중요도의 차이를 반영하기 위해, 각 task 의 train split 에서 무작위로 추출한 128 개 샘플을 활용하여 각 레이어의 상대적 중요도를 동적으로 측정하고, 이를 기반으로 pruning 비율을 조정하는 방식을 채택한다.

레이어 중요도 측정 과정은 3.1.1 절에서, 이를 활용한 Pruning 및 LoRA 어댑터 병합 과정은 3.1.2 절에서 상세히 설명한다.

3.1.1 레이어 별 중요도 측정

레이어 중요도 측정 과정은 두 단계로 구성된다. 먼저, intra-layer 분석을 통해 각 레이어 l의 가중치 행렬 (W^l)와 입력 샘플(X^l)의 통계적 특성을 바탕으로 평균 활성화 값(A^l)을 산출한다:

$$A_{ij}^l = |W_{ij}^l| \odot \|X_j^l\|_2 \quad (1)$$

여기서 $\|X_j^l\|_2$는 레이어 l에 입력된 샘플들에 대해 j번째 입력 차원의 평균 L2 norm 이다. 이후, A^l의 평균을 취해 전체 레이어의 중요도를 하나의 스칼라 값 I_l로 나타낸다:

$$I_l = mean(A_{ij}^l) \quad (2)$$

레이어 간 상대적 중요도를 비교할 수 있도록 정규화 과정을 수행한다. 이를 위해 $\{I_1, I_2, \dots, I_L\}$의 최소값과 최대

값을 기준으로 각 레이어에 대해 다음과 같은 정규화 스코어($\tilde{I}_l$)를 계산한다:

$$\tilde{I}_l = \left(\frac{I_l - \min(I)}{\max(I) - \min(I)} \right) \quad (3)$$

정규화된 $\tilde{I}_l$ 은 이후 중심 정렬을 통해, 각 레이어의 sparsity 조정치가 평균 0 을 중심으로 분포되도록 한다:

$$\Delta S_l = \tilde{I}_l - mean(\tilde{I}_l) \quad (4)$$

이렇게 조정된 ΔS_l은 전체 pruning 비율 S_{org}와 결합되어, 각 레이어의 최종 sparsity 비율로 사용된다:

$$S_l = S_{org} + \Delta S_l \quad (5)$$

아래의 그림 2 는 본 기법을 적용했을 때, 두 가지 태스크(arc_challenge, hellaswag)에 대해 레이어별로 계산된 최종 sparsity importance 차이를 시각화한 것이다. 이를 통해 태스크별로 중요하게 작동하는 레이어가 어떻게 달라지는지를 확인할 수 있다.

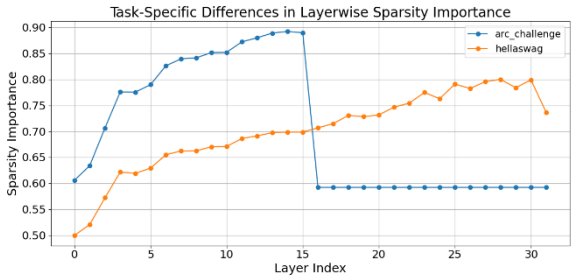


그림 2: layer 별 layer-wise sparsity 비율 ($S_{org} = 0.7$)

3.1.2 TIP 을 활용한 LoRA 어댑터 병합

레이어별 중요도에 따라 정해진 sparsity 비율을 기반으로, 본 연구는 LoRA 어댑터에 WANDA[3] 기법을 적용하여 pruning 을 수행한다. WANDA 는 weight 자체의 크기뿐만 아니라, 해당 weight 에 대응되는 입력값의 정도까지 함께 고려함으로써, 보다 정교한 출력 활성화 기반 pruning 을 가능하게 한다. WANDA 기법은 수식 (1)과 같으며 레이어별로 사전에 결정된 S_l에 따라 중요도가 낮은 weight 부터 순차적으로 제거한다.

이 과정을 모든 LoRA 어댑터에 개별적으로 적용한 뒤, element-wise summation 방식으로 병합함으로써, 여러 task 에 최적화된 통합 어댑터를 구성한다.

아래의 그림은 위 절차를 시각적으로 설명한 것이다. 그림에서 t는 서로 다른 task 를 의미하며, 각각의 $W_{ij}^{l(t)}$와 $\|X_j^{l(t)}\|$는 해당 task 의 LoRA weight 와 입력 샘플을 나타낸다.

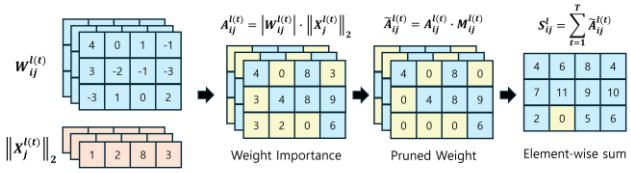


그림 3: Pruning 을 활용한 병합 과정 개요도

3.2. Evaluation

본 실험은 LLaMA 2 7B 모델을 기반으로 하여 총 7 가지 대표적인 벤치마크 태스크(ARC-Challenge, BoolQ, CoQA, HellaSwag, OpenBookQA, PIQA, RTE)를 대상으로 평가하였다. 모든 태스크에 LoRA rank 8 설정을 적용하였으며, 사용된 어댑터는 HuggingFace[4]에서 공개된 사전학습(pre-trained) 어댑터를 그대로 활용하였다. 평가는 각 태스크의 공식적으로 제공된 평가 split(test 또는 validation)을 사용하였으며, 성능 지표는 accuracy 를 기준으로 산출하였다. 모든 실험은 NVIDIA A40 GPU (48GB) 단일 장비에서 수행되었다.

표 1 은 병합하지 않은 단일 LoRA(Single), 단순 선형 병합(Linear), DARE[2], 그리고 본 논문에서 제안한 TIP 병합(TIP)의 정확도를 비교한 결과이다. Linear 와 DARE 는 병합 후 성능 저하가 발생하는 반면, TIP 은 성능을 유지하거나 오히려 기존보다 향상시켜, 일종의 앙상블 효과와 유사한 성능 개선을 달성했음을 보여준다.

특히 CoQA 와 HellaSwag 와 같이 문맥 의존성이 높고 reasoning 이 요구되는 태스크에서 TIP 은 두드러진 성능 향상을 보였다. CoQA 에서는 DARE 대비 약 +43.9%, HellaSwag 에서는 +31.3%의 상대적인 정확도 향상이 관측되었다. 한편, RTE 태스크에서는 단일 LoRA 대비 성능 하락이 있었으나, TIP 은 여전히 Linear 나 DARE 보다 높은 정확도를 기록하여 정보 손실 없이 안정적인 통합 성능을 유지함을 입증하였다.

	Single	Linear	DARE	TIP
ARC	0.4625	0.3370	0.3413	**0.4983**
Boolq	0.7774	0.7275	0.7174	**0.8138**
Coqa	0.6388	0.2260	0.2257	**0.6647**
Hellaswag	0.7600	0.4609	0.4547	**0.7675**
Openbookqa	**0.4420**	0.3840	0.3720	**0.4420**
Piqa	0.7905	0.6627	0.6540	**0.7933**
RTE	**0.6282**	0.5343	0.5379	0.5560

표 1: 태스크 별 정확도

Ⅳ. 결론

본 연구는 multi-task 환경에서 LoRA 어댑터 병합 시 발생하는 연산 비효율성과 파라미터 간섭 문제를 해결하기 위해, 입력 기반 레이어 중요도 분석과 동적 sparsity 조정을 결합한 새로운 병합 기법(TIP)을 제안하였다. 제안된 방법은 기존의 선형 결합 및 정적 sparsity 방식보다 일관되게 높은 정확도를 달성하였으며, 일부 복잡한 태스크에서는 병합 자체가 오히려 성능 향상을 유도하는 효과도 관찰되었다.

ACKNOWLEDGMENT

이 논문은 2025 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 **(No.RS-2020-II201373, 인공지능대학원지원(한양대학교); NO.RS-2021-II211343, 인공지능대학원지원(서울대학교))**

참 고 문 헌

[1] HU, Edward J., et al. Lora: Low-rank adaptation of large language models. *ICLR*, 2022, 1.2: 3.

[2] YU, Le, et al. Language models are super mario: Absorbing abilities from homologous models as a free lunch. In: *Forty-first International Conference on Machine Learning*. 2024.

[3] SUN, Mingjie, et al. A simple and effective pruning approach for large language models. *arXiv preprint arXiv:2306.11695*, 2023.

[4] https://huggingface.co/Styxxxx

한국어 추론에서 Zero-shot 프롬프트 유형별 성능 분석 연구

박선우, 김윤하, 문준렬, 임완수*
성균관대학교
*wansu.lim@skku.edu

A Study on the Performance of Zero-shot Prompting Strategies for Korean Language Inference

Park Sun Woo, Kim Younha, Moon Jun Ryeol, Lim Wansu*
Sungkyunkwan University

요 약

본 연구는 한국어 환경에서 프롬프트 템플릿의 차이가 대규모 언어모델(Large Language Model, LLM)의zero-shot 추론 성능에 미치는 영향을 체계적으로 분석한다. 표준성이 높은 Korean_SAT_MATH 데이터셋과 GPT-4를 사용하여 (i) 성능 유도형(instructive), (ii) 판단 교란형 (misleading), (iii) 과업 무관형(irrelevant), (iv) 프롬프트 미제공(zero-shot baseline) 등 네 가지 유형과 총 8개 프롬프트 조건을 비교한다. 모든 실험은 모델·데이터셋·생성 설정을 고정하고 프롬프트만 변경하여 수행하였으며, 정확도(Accuracy)로 성능을 평가하였다. 분석 결과 논리적 사고를 유도하는instructive 프롬프트는 zero-shot baseline 대비 최대 2.5% 성능 향상을 보였으며, misleading과 irrelevant 프롬프트는 성능을 크게 저하시켰다. 이러한 결과는 한국어 LLM 활용 시 프롬프트 설계가 성능과 직결됨을 실증하며, 실무 적용을 위한 설계 지침의 필요성을 뒷받침한다. 본 연구는 프롬프트 설계의 순수 효과를 정량적으로 규명하고, 향후 다양한 한국어 추론 과제 확장, 다중 샘플링 기반 분석, 역프롬프트에 대한 강건성 평가 등으로의 발전 가능성을 제시한다.

Ⅰ. 서 론

대규모 언어모델(LLM)은 프롬프트만으로 다양한 작업을 수행하는 zero-shot 능력을 보이며, 프롬프트 구성의 미세한 차이가 성능을 좌우한다 [1,2]. 한국어 환경에서도 이런 현상이 관찰되지만, 한국어 데이터·모델·프롬프트가 결합될 때의 성능 변동을 체계적으로 정량화한 연구는 부족하다. 한국어적 특성이 추론 판단에 미치는 영향이 프롬프트 설계에 의해 증폭되거나 상쇄될 수 있음에도, 이를 뒷받침하는 실증 근거는 부족하다. 더불어 실사용 현장에서는 미세조정 없이 즉시 적용 가능한 프롬프트 지침이 요구되지만, 한국어에서 프롬프트 설계의 순수 효과를 모델·데이터 변경 없이 분리해 보여준 사례가 드물다 [2].

이러한 한계를 보완하기 위해 본 연구는 단일 데이터셋과 단일 모델을 고정한 조건에서 zero-shot 프롬프트 템플릿의 정확도를 비교·정량화하고, 한국어 추론 과제에 적용 가능한 간결한 프롬프트 설계 지침을 제시한다. 아울러 혼선 유발·무관 문맥 프롬프트의 정확도까지 비교하여 효율적인 LLM 프롬프트 설계 가이드라인을 제시한다.

Ⅱ. 본론

2. 1 이론적 배경 및 선행연구

LLM은 방대한 말뭉치에 대한 자기회귀(next-token prediction) 사전학습을 통해 이전 토큰이 모두 주어졌을 때 다음 토큰이 나타날 확률 분포를 근사한다 [3]. 사전학습 이후 특정 자연어 태스크를 수행하는 방법은(i) 파라미터를 갱신하는 방식과 (ii) 파라미터를 고정한 채 입력을 설계하는 방식으로 구분된다. 전자는 특정 태스크별 미세조정(fine-tuning)과 자연어 지시를 모아 수행력을 높이는 지시튜닝(instruction tuning)이 대표적이며, 후자는 예시 질문 및 답변을 고정된 형식으로 프롬프트에 포함하여 형식적인 답변을 유도하는 few-shot(in-context) 학습, 중간 추론을 언어로 노출시키는 Chain-of-Thought(CoT), 그리고 어떠한 예시 없이 태스크의 정의와 출력 형식만으로 답을 유도하는 zero-shot이 핵심이다 [4,5]. 이 중 zero-shot은 추가 학습·데이터 준비·연산 비용이 들지 않아 비용 효율성 면에서 유리하다.

다만 zero-shot의 성능은 프롬프트 설정에 크게 좌우된다. [1]은 동일한 모델과 입력에서도 템플릿만 달리할 때 정확도의 변화가 크다고 보고한다. 특히 “Let’s think step by step.”과 같은 사고 절차 유도형은 기준선 대비 성능 향상을 보이는 반면, 의도적으로 오도하거나(misleading) 무관한(irrelevant) 문맥을 주입하면 언어 모델의 성능에 부정적인 영향을 미친다. 따라서 zero-shot을 실무에 적용하기 위해서는, 언어·과업 맥락에 맞는 템플릿을 정교하게 선택·설계하는 것이 필수적이다.

2. 2 실험 설정

본 연구는 이러한 관찰을 한국어 환경에서 검증하기 위해, zero-shot 프롬프트 유형과 문구에 따른 추론 성능을 정량 비교한다. 데이터셋은 표준성이 높은 Korean_SAT_MATH 문제 세트로 고정하고, 모델은 OpenAI의 GPT-4를 단일 설정으로 사용한다 [6]. 추가 학습이나 미세조정은 수행하지 않으며 데이터셋의 질문과 함께 입력되는 프롬프트만 변경하여 응답을 수집한다.

프롬프트 조건은 총 8개로 구성한다: 성능 유도형 instructive 2개, 판단 교란형 misleading 2개, 과업 무관형 irrelevant 2개, 그리고 zero-shot 기준선 3개. 모든 조건은 동일한 질문을 사용하며, 출력은 숫자와 latex 함수

로 출력된다. 평가는 Accuracy(정확도)를 사용한다. 디코딩은 temperature=0, top-p=1의 결정적 설정과 최대 256 토큰 생성 한도를 적용한다.

2. 3 실험 결과 및 분석

본 연구는 프롬프트 유형에 따른 언어 모델의 성능 변화를 정성적으로 평가하고, 정량적으로 분석한다. 표 1은 프롬프트의 설계 방식이 동일한 실험 환경에서도 정확도에 유의미한 영향을 미친다는 점을 보인다. 아래는 Instructive, Misleading, Irrelevant의 세 유형에 대한 언어 모델의 성능을 정량적으로 분석한 결과이다.

먼저, Instructive 유형은 전반적으로 Zero-shot baseline 대비 성능 유지 또는 향상을 보였다. 특히 논리적 사고 과정을 명시적으로 유도하는 1번 프롬프트에서 zero-shot baseline 대비 2.5%의 정확도 향상을 보였다. 이는 모델이 문제 해결 과정에서 단계적이고 체계적인 추론 경로를 따를 수 있도록 유도하는 프롬프트가 성능 개선에 기여함을 시사한다. 반면 2번 프롬프트는 baseline과 유사한 성능을 나타내어 모든 지시 유형의 프롬프트를 사용할 때, 동일한 효과를 발휘하는 것이 아님을 보인다.

Misleading 유형은 잘못된 해결 전략을 포함하여 모델 성능에 뚜렷한 영향을 미쳤다. 3번 프롬프트의 경우 모델이 스스로 reasoning 단계를 거치지 못하도록 “생각하지 말고”라는 문구를 포함하였다. 특히 4번 프롬프트는 모델에게 제시된 문제를 해결하지 못하도록 다른 문제를 프롬프트를 제시함으로써 5%의 저조한 성능을 기록한다. 이에 따라 Zero-shot baseline 유형의 6번 프롬프트 대비 20%, Instructive 유형의 1번 프롬프트 대비 22.5%까지 성능이 하락하였다. 이러한 결과는 모델이 프롬프트 내의 지시를 높은 비율로 수용하며 그 지시가 부정확하거나 문제 해결과 무관할 경우 추론 성능이 심각하게 저하됨을 의미한다.

Irrelevant 유형은 문제와 전혀 관련 없는 발화를 포함하여 성능 저하를 초래하였다. 불필요한 맥락 삽입이 모델의 self-attention을 방해하여 문제 이해 및 정답 산출 과정을 저해한다고 분석된다. 하지만 Misleading 유형의 4번 프롬프트 대비 성능 향상이 있는 이유는 언어 모델에게 제시된 문제와 입력된 프롬프트에 대한 답변 모두를 출력하기 때문이라고 분석된다.

Zero-shot 조건에서 6번은 단순 지시문으로 안정적인 성능을 보인 반면, 불필요한 지시를 제거하여, 요구되는 정답 형식을 제시하지 않은 7번 프롬프트는 성능이 소폭 하락하였다. 8번 프롬프트의 경우 최종 답만 작성하라는 지시 때문에 언어 모델이 스스로 reasoning 단계를 거치지 못하여 성능이 급감한 것으로 보인다.

유형	No.	템플릿	Acc(%)
Instructive	1	정답은 latex 함수를 사용해서 작성해. 논리적으로 풀이해보자.	27.5
	2	정답은 latex 함수를 사용해서 작성해. 문제를 확실히 파악한 후 정답을 도출하자.	24.17
Misleading	3	생각하지 말고 감대로 작성해보자.	21.67
	4	먼저 질문에 들어있는 ‘o’의 개수를 세어보자.	5.0
Irrelevant	5	내일 아침 샐러드 어때?	21.67
Zero-shot baseline	6	정답은 latex 함수를 사용해서 작성해.	25.0
	7	-	22.5
	8	최종 답만 작성해보자.	11.67

표 1 프롬프트 템플릿에 따른 LLM 추론 결과

Ⅲ. 결론

본 연구는 단일 데이터셋과 모델을 활용하여 프롬프트만을 조작해 zero-shot 성능 변화를 정량화함으로써, Korean_SAT_MATH 데이터셋 문제에서 프롬프트 설계의 순수 효과를 명확히 드러냈다. 그 결과, instructive 유형의 프롬프트 설계를 체계적이고 논리적으로 진행하면 성능 향상이 발생함을 확인했으며, 반대로 질문과 관련 없거나 reasoning에 방해가 되는 단어는 성능에 큰 영향을 미치는 것을 보였다. 이러한 관찰을 실제 적용이 가능하도록 정리하여 사용자에게 효율적인 가이드라인을 제시한다. 향후 연구로는 (i) 한국어의 다른 추론 과제(상식·상황·과학 문항)로의 확장, (ii) Self-Consistency 등 다중 샘플링 기반의 비용/성능 곡선 분석, (iii) 반사실적/역프롬프트 공격에 대한 강건성 평가를 진행하여, 프롬프트 설계 지침의 범용성과 안전성을 한층 보강할 예정이다.

ACKNOWLEDGMENT

본 연구는 산업통상자원부(MOTIE)와 한국에너지기술평가원(KETEP)의 지원을 받아 수행한 연구 과제입니다. (No. RS-2022-KP002701)

본 연구는 보건복지부의 재원으로 한국보건산업진흥원의 보건의료기술연구개발사업 지원에 의하여 이루어진 것임 (No. RS-2025-02223417)

참 고 문 헌

[1] T. Kojima, S. S. Gu, M. Reid, Y. Matsuo, and Y. Iwasawa, “Large language models are zero-shot reasoners,”in Advances in Neural Information Processing Systems (NeurIPS), vol. 36, Curran Associates, Red Hook, NY, USA, 2022.

[2] Sanh, V., Webson, A., Raffel, C., Bach, S. H., Sutawika, L., Alyafeai, Z., et al., "Multitask Prompted Training Enables Zero-Shot Task Generalization"

InternationalConferenceonLearningRepresentations(ICLR),2022.

[3] Y. Bengio, R. Ducharme, and P. Vincent, “A neural probabilistic language model,”in Advances in Neural Information Processing Systems(NeurIPS), vol. 13, 2000.

[4] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, et al., “Language models are few-shot learners,”in Advances in Neural Information Processing Systems (NeurIPS), vol. 34, Curran Associates, Red Hook, NY, USA, 2020.

[5] J. Wei, X. Wang, D. Schuurmans, M. Bosma, F. Xia, E. Chi, Q. Le, and D. Zhou, “Chain-of-thought prompting elicits reasoning in large language models,”in Advances in Neural Information Processing Systems (NeurIPS), vol. 36, Curran Associates, Red Hook, NY, USA, 2022.

[6] OpenAI, J. Achiam, S. Adler, et al., “GPT-4 technical report,”arXiv preprint arXiv:2303.08774, 2023.

다양한 조명 환경에서 보행자 탐지를 위한 퍼지 논리 기반 전처리 프레임워크

이원일, 유연승, 이민호, 이수빈, 정성균, 최흥준, 김태형*
롯데이노베이트㈜

wonil.lee@lotte.net, yys4000@lotte.net, minho_lee@lotte.net, leesubin@lotte.net,
seonggyun.Jeong@lotte.net, heungjunchoi@lotte.net, thkim07@lotte.net*

A Fuzzy Logic-based Preprocessing Framework for Pedestrian Detection in Various Lighting Conditions

Lee Won Il, Yu Yeon Seung, Lee Min Ho, Lee Su Bin,
Jeong Seong Gyun, Choi Heung Jun, Kim Tae Hyung*
LOTTE INNOVATE Co., Ltd.

요 약

영상 기반 보행자 탐지는 지능형 감시의 핵심 기술이지만, 안개, 야간의 저조도 환경에서 성능이 급격히 저하되는 문제가 있다. 이러한 문제를 해결하기 위해, 본 논문에서는 영상의 시각적 특징을 수치화하고 퍼지 논리 시스템을 통해 날씨 및 조도 환경을 판별하여 최적의 전처리 기법을 동적으로 적용하는 적응형 프레임워크를 제안한다. 제안하는 방법은 별도의 인공신경망 학습 과정 없이 동작하며, 부족한 광원 환경에서도 보행자 탐지 시스템의 정확도를 향상시킬 수 있다. 제안하는 프레임워크를 cctv 데이터셋으로 평가한 결과, 보행자 탐지에서 성능 향상 보인다.

Ⅰ. 서 론

영상 기반 객체 탐지 기술의 발전은 보행자 탐지 및 추적 시스템의 성능을 향상시켜왔다. 하지만 이러한 시스템의 성능은 입력 영상의 품질에 크게 의존한다. 안개, 야간, 역광 등 실제 도로 환경에서 빈번하게 발생하는 저하 요인들은 영상의 대비와 선명도를 감소시켜 탐지율을 급격히 떨어뜨리는 주된 원인이 된다. 기존의 연구에서 활용된 전처리 방식은 이미지의 대비(Contrast) 값에만 초점을 맞추어, 변화하는 환경에 능동적으로 대처하기 어렵다. 따라서 본 논문에서는 이미지의 밝기(Brightness), 선명도(Sharpness)를 확인한 후 그 수치에 따라 감마 보정(Gamma Correction), CLAHE(Contrast Limited Adaptive Histogram Equalization)[1] 등의 전처리 기법을 동적으로 선택하는 퍼지 논리(Fuzzy Logic) 기반의 적응형 프레임워크를 제안한다. 제안 방법의 효과를 확인하기 위해, 주간, 야간, 적외선 이미지가 포함된 데이터셋에 적용하여 보행자 탐지 성능을 평가한다. 실험 결과, 제안하는 퍼지 논리 기반 프레임워크는 전처리를 적용하지 보행자 탐지 방법에 비해 성능을 향상을 보인다.

Ⅱ. 본론

2.1. 보행자 탐지 및 추적

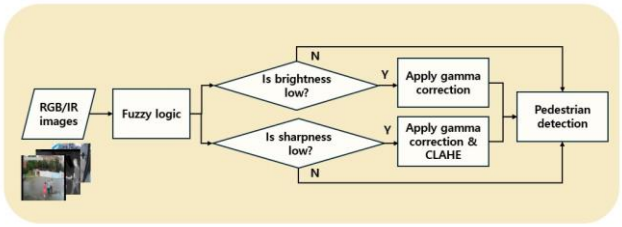
보행자 탐지 기술은 악천후 및 야간 환경에서 이미지 품질 저하로 인해 성능이 급격히 감소하는 명확한 한계를 가진다[2]. 기존의 관련 연구들은 CLAHE 를 각기 다른 방법으로 적용시켜왔다[3]. 예를 들어, [4]의 연구는 실시간 처리에 효율적인 적응형 파라미터 최적화를 제안했으며, [5]에서는 Guided Image Filtering 을 결합한 CLAHE-GIF 를 제안하여 안개 속 노이즈 억제력을 높였다. [6]의 연구에서는 Dark Channel Prior 를 동적 CLAHE 와 통합한 방법을 설계했다. 이들은 대부분 불균일한 조명을 해결하기 위한 대비 향상에 초점을 맞추고 있다.

2.2. 제안 방법

본 논문은 다양한 조명 환경에 대응하기 위해, 퍼지 논리를 활용하여 이미지 프로세싱을 동적으로 적용하는 프레임워크를 제안한다. 입력 이미지에서 밝기, 선명도를 수치화한 뒤, 그에 따라 'Night', 'Blurry', 'Clear' 등과 같이 현재 상태를 판정한 후 CLAHE 뿐만 아니라 Gamma correction 등 이미지 프로세싱을 적용한다.

이 시스템은 특징 추출, 퍼지 추론, 전처리 적용의 세 단계로 구성된다. 첫 단계에서는 픽셀 값의 평균으로 밝기를, 라플라시안 필터의 분산으로 선명도를 산출한다. 두 번째 단계인 퍼지 추론 엔진은 이 지표들을 입력받아 규칙 기반의 판단을 수행하며, 마지막 단계에서는 판단 결과에 따라 최적의 전처리를 선택한다.

예를 들어, Night 상태는 감마 보정을 통해 밝기를 높이고, Blurry 상태는 CLAHE 로 대비를 향상시키며, Clear 상태는 원본을 유지해 불필요한 왜곡을 방지한다. 이러한 흐름을 통해 수치로는 표현하기 어려운 다양한 촬영 환경을 유연하게 판별하고 적절한 보정을 수행할 수 있다.



[그림 1] 프레임워크 순서도

2.3. 실험

2.3.1 데이터

본 연구에서는 Roboflow cctv 데이터셋[7]을 사용하여 제안 방법을 평가한다. 평가에는 3,045 장의 데이터를 활용하였으며, 일반적인 컬러(RGB) CCTV 이미지뿐만 아니라 적외선(IR) 카메라로 촬영된 야간 영상도 포함되어 있다. 해당 데이터셋은 'person' 단일 클래스를 대상으로 한 객체 탐지용으로 구성되어 있으며, 주간과 야간 등 다양한 조명 조건에서의 보행자 탐지 모델을 평가하여 제안한 전처리 방법의 효과를 확인한다.

2.3.2 평가

[표 1]은 전처리를 적용하지 않은 방식과 CLAHE, 퍼지 논리 기반 제안 방식에 따른 보행자 탐지 성능을 비교한 결과를 나타낸다. CLAHE 적용 시 성능은 F1-score 0.6453 으로, baseline 대비 성능이 하락하였다. 이는 대비가 충분한 영상에서 CLAHE 적용 시 노이즈와 색 왜곡이 증가하여 탐지 성능을 저해할 수 있음을 시사한다. 반면, 제안 방법은 F1-score 0.6515 의 성능을 내며 원본 대비 향상된 결과를 보였으며, CLAHE 적용보다 우수한 결과를 얻었다. [그림 2]와 같이 기존에 탐지되지 않던 IR 이미지 내의 사람을 성공적으로 탐지하였다. 이러한 결과는 단일 필터를 모든 영상에 일괄 적용하는 정적 방식이 다양한 환경 변화에 대응하는 데 한계가 있음을 보여주며, 상황 적응형 전처리 접근의 필요성을 보인다.

[표 1] 전처리 방법 별 성능 비교표

Method	*Precision*	*Recall*	*F1-Score*
Baseline	0.8615	0.5180	0.6470
CLAHE	0.8515	0.5195	0.6453
Proposed	**0.8609**	**0.5241**	**0.6515**

[그림 2] 전처리 적용 전의 미탐 이미지(좌),
CLAHE 적용 후의 미탐 이미지(중),
제안 방법 적용 후 탐지된 사람 이미지(우)

Ⅲ. 결론

본 논문에서는 흐린 이미지나 저조도 환경에서 보행자 탐지를 위해 퍼지 논리를 기반으로 영상의 상태를 실시간으로 진단하고 최적의 전처리를 수행하는 적응형 프레임워크를 제안했다. 이 방법은 입력 이미지 밝기와 선명도를 분석하여, 감마 보정과 CLAHE 와 같은 전처리 기법들을 동적으로 선택하고 적용한다. Roboflow 의 cctv 데이터셋을 이용한 실험 결과, 제안된 프레임워크는 전처리를 적용하지 않거나 CLAHE 만을 사용한 방식에 비해 탐지 성능을 향상시키는 것을 확인했다. 본 방법은 다양한 환경이나 실시간 지능형 감시 시스템에서의 탐지 신뢰도를 높이는 데 적용할 수 있을 것이다.

참 고 문 헌

[1] Zuiderveld, Karel. "Contrast limited adaptive histogram equalization." *Graphics gems* IV. pp. 474-485, 1994.

[2] Chen, Wei, et al. "Occlusion and multi-scale pedestrian detection A review." *Array*, vol. 19, 2023

[3] Song, Yifei, and Yanfeng Lu. "A Review of Unmanned Visual Target Detection in Adverse Weather." *Electronics* vol. 14, no. 13, 2025.

[4] Yuan, Zhikang, et al. "CLAHE-based low-light image enhancement for robust object detection in overhead power transmission system." *IEEE Transactions on Power Delivery*, vol. 38, no. 3, pp. 2240-2243, 2023.

[5] Zhang, Yi, et al. "Artificial intelligence-aided thermal model considering cross-coupling effects." *IEEE Transactions on Power Electronics,* vol. 35, no.10, pp. 9998-10002, 2020.

[6] Lashkov, Igor, Runze Yuan, and Guohui Zhang. "Edge-computing-facilitated nighttime vehicle detection investigations with CLAHE-enhanced images." *IEEE Transactions on Intelligent Transportation Systems,* vol. 24, no. 11, pp. 13370-13383, 2023.

[7] https://universe.roboflow.com/dataset-uutxr/cctv-naxyo

가우시안 어둠 분포 학습을 통한 잠재 확산 기반 저조도 이미지 밝기 복원 기법에 관한 연구

송순용
한국전자통신연구원

soony@etri.re.kr

A Study on Gaussian Darkness Modeling in Latent Diffusion for Low-Light Image Enhancement

Soonyong Song
Electronics and Telecommunications Research Institute (ETRI)

요 약

본 논문은 잠재 확산 기반의 저조도 개선 기법인 LD-Gauss 를 제안하였다. 저조도 이미지를 안정적으로 밝게 만들기 위해, 잠재 공간에서 어두움을 가우시안 분포로 근사하고 그 분포의 평균과 분산을 조건부로 추정한 뒤, 확산 모델의 역과정에서 이를 이용해 원래의 광량을 복원하는 것이 핵심 아이디어이다. 제안한 모델은 사전학습된 Latent Diffusion 의 pretrained weight 를 freeze 하고, 저조도 조건으로부터 잠재 복원량을 예측하는 추정 네트워크와 잠재 U-Net 만 학습한다. 실험 결과, PSNR 은 9.66dB 에서 16.45dB 로 약 6.79dB 향상되었고, SSIM 은 0.351 에서 0.478 로 증가하였다. 반면 LPIPS 는 0.328 에서 0.369 로 열화되어 밝기가 올라가면서 세부 질감과 자연스러움의 손실이 발생됨을 시사하였다. 시각적인 평가 역시 전반적인 노출 회복에는 성공하였지만, 부분적 디테일 붕괴와 과도한 밝기 보정으로 인한 인공적인 느낌이 관찰되었다. 이 결과는 PSNR 과 SSIM 지표가 픽셀의 일치를 선호하는 반면, LPIPS 가 사람의 지각과 상관성이 높다는 기존의 결과와 일치한다.

Ⅰ. 서 론

저조도 영상은 촬영 단계에서 밝기가 부족하고 센서 잡음이 상대적으로 커지기 때문에, 단순히 톤을 조정하는 것 만으로는 충분한 밝기 회복과 디테일 유지를 동시에 달성하기 어렵다. 특히, 노이즈 특성은 광 신호의 세기에 따라 달리지기 때문에 푸아송 및 가우시안 잡음 모델을 통해 물리적 현실성을 반영한 저조도 합성이 필요하다. [1] 전통적으로, 저조도 이미지 개선 분야에서 GAN 과 같은 딥러닝 기법이 널리 사용되었다. 하지만, 실제 어두운 환경에서는 학습 이미지의 부족이나 과도 보정, 질감 손실 같은 한계가 존재하였다. 최근에는 Diffsion 기반 방법들이 이미지 품질 회복과 구조 보존 측면에서 유망한 결과를 보여주고 있다. Latent Diffusion 과 같은 확산 모델은 고해상도 이미지 합성 및 복원 작업에서 성공적으로 활용되는 모델로, VAE 를 활용하여 잠재 공간에서 확산 과정을 수행함으로써 픽셀 공간 기반보다 계산 효율성과 시각 품질 면에서 모두 우수한 결과를 보여준다. [2]

본 연구에서는 저조도에서의 어둠을 잠재 공간에서 가우시안 분포로 모델링하고, 이를 확산 복원 과정에 통합하는 새로운 접근법인 LD-Gauss 기법을 제안한다. 이 방법은 사전학습 가중치를 고정한 상태에서, 밝기 손실 복원량(평균 및 분산)을 조건적으로 추정하는 중간 네트워크와 잠재 U-Net 기반 확산 역과정을 학습함으로써, 잠재공간에서의 통계적인 분포 근사화를 통해 노출 부족 문제와 잡음으로 무너진 구조를 동시에 복원하는 것을 목표로 한다.

Ⅱ. 본론

본 연구의 목표는 잠재 공간에서 노출 손실의 확률 구조를 모델링하여 저조도 입력을 원래 노출로 복원하는 것이다. 구체적으로는 세 단계로 요약할 수 있다. 첫째, 원본 이미지를 sRGB 에서 선형 광도로 바꾼 뒤, 노출 스톱 감소, 푸아송 잡음 추가, 가우시안 잡음 추가, sRGB 역변환을 거쳐 저조도 데이터를 합성한다. 둘째, 사전학습된 Latent Diffusion 의 사전학습 가중치를 그대로 고정하고, 고정된 scaling factor 규칙에 따라 입력과 저조도의 잠재 특성을 얻는다. 셋째, 저조도 영상에서 얻은 밝기, 엣지, 주저파 조명 신호와 추정 잡음 세기 맵을 조건으로 삼아 잠재 복원량의 평균과 분산을 예측하는 네트워크와 확산 역과정의 노이즈 예측 U-Net 을 함께 학습한다. 확산 손실과 함께 평균 및 분산이 실제 잠재 차이와 들어맞도록 Gaussian NLL 을 적용하고, Latent Diffusion 복호화 후의 사이클 일치 손실을 보조로 합산하였다. 이런 구성은 물리 합성 단계에서의 푸아송-가우시안 가정과, Latent Diffusion

비용 효율성을 동시에 달성한다. 데이터셋은 STL-10 을 사용하였다. 이 데이터셋은 96x96 해상도의 컬러 이미지로 구성되고, 지도 학습에 쓸 수 있는 라벨 수는 적지만 비지도 학습을 위한 대량의 비라벨 데이터를 포함한다. 본 연구에서는 회귀성 복원 과제이므로 라벨 정보는 사용하지 않았다. 저해상도이긴 하지만, 잠재 해상도가 8 배 축소되는 Latent Diffusion 구성을 시험하기에는 충분한 해상도를 갖는다. 정량 평가에서는 PSNR, SSIM, LPIPS 세 가지 지표를 사용하였다. PSNR 은 평균 제곱 오차 기반이기 때문에 밝기나 색이 일치할수록 값이 커지고, SSIM 은 구조적 유사성을 통해 대비, 명암, 구조의 보존을 측정한다. LPIPS 는 사전학습 신경망의 특성 공간에서 거리로 유사도를 측정한 지표로 사람의 지각과 더 높은 상관을 보이는 것으로 알려져 있다. 이 세 가지 지표를 통해 복원된 이미지의 품질을 균형적으로 평가할 수 있다. [3] 실험 결과, 저조도 입력 대비 복원 전과 복원 후를 비교하였을 때, PSNR 은 9.66dB 에서 16.45dB 로 6.79dB 상승했고, SSIM 은 0.351dptj 0.478 로 0.127 증가하여 약 36% 향상이 있음을 확인하였다. 반면 LPIPS 는 0.328 에서 0.369 로 0.041 증가하여 특성의 거리 측면에서 손실이 있음을 확인하였다. 이러한 결과는 노출과 전반적인 구조 회복에는 성공하였지만, 자연스러운 질감은 충분히 복원되지 못하였음을 의미한다. 그림 1 에는 학습된 모델로 복원한 결과를 나타내었다. 시각적으로 전체 밝기는 target 이미지와 비슷하게 올라가고 전반적인 실루엣은 유사하게 나타났지만, 세부적인 특징인 무늬, 경계면 등에서 세부 디테일이 무너지는 것으로 판단되며, 일부 장면에서는 과다 보정으로 인한 인공적인 결과가 출력되었다. LPIPS 성능의 열화는 이러한 결과와 일치한다. 이러한 결과는 세 가지 원인에 의한 것으로 분석할 수 있다. 첫째, 잠재 영역에서 해상도의 축소 때문으로, Latent Diffusion 기본 구성처럼 축소 후 복원 절차를 진행하면 고주파 표현 능력의 저하가 발생한다. 둘째, 본 모델은 노출과 잡음의 통계적 복원에 초점을 둔다. 밝기를 올리고 노이즈를 정규화하는 방향의 손실이 강하면 텍스쳐를 복원하기 위해 필요한 미세 특징이 희생될 수 있다. 셋째, 학습 데이터셋의 해상도가 낮은 것과 이미지의 다양성 부족이 문제일 수 있다. 실사용에 가까운 질감을 확보하기 위해서는 큰 해상도로 다양한 장면들의 학습이 필요하다. 이번 결과에서 PSNR 과 SSIM 이 유의미하게 좋아진 것은 제안한 방법론은 합리적으로 선택된 것을 의미한다. 제안한 방법의 절차가 밝기와 구조를 복원하기에는 효과적인 것을 보여준다. 반면 LPIPS 가 열화된 사실은 텍스쳐 손실 항이 더 필요한 것을 시사한다.

Ⅲ. 결론

본 논문은 잠재 공간에서 어둠을 가우시안으로 근사해 밝기를 복원하는 LD-Gauss 를 제안하였다. 제안 방법은 PSNR 상승과 SSIM 개선으로 밝기와 구조 회복이 가능함을 보였고, 동시에 LPIPS 의 열화와 시각적 평가에서의 디테일 붕괴와 과보정이라는 한계를 드러냈다. 그럼에도 본 결과는 잠재공간에서 노출 손실의 확률적인 분포를 학습하여 밝기를 복원하려는 접근은 유효함을 보여주며, 고해상도 데이터를 활용한 학습, 톤 제어, 저손실 VAE 채택 등의 방법을 적용하여 단점을 보완해 나간다면 실제 응용으로 이어질 가능성이 충분한 것으로 판단된다.

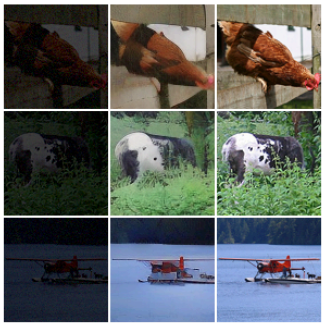


그림 1. 저조도 이미지 밝기 복원 결과: 입력 이미지 (좌), 복원 결과 (중), 목표 이미지(우)

ACKNOWLEDGMENT

본 연구는 한국전자통신연구원 연구운영비지원사업의 일환으로 수행되었음. [25ZR1100, 자율적으로 연결·제어·진화하는 초연결 지능화 기술 연구]

참 고 문 헌

[1] Rombach, Robin, et al. "High-resolution image synthesis with latent diffusion models." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022.

[2] Jiang, Hai, et al. "Lightendiffusion: Unsupervised low-light image enhancement with latent-retinex diffusion models." European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2024.

[3] Zhang, Richard, et al. "The unreasonable effectiveness of deep features as a perceptual metric." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

STOI 메트릭 기반의 다중 레벨 음성 향상 및 데이터 증강 기법

신도경†, 김영대
LIG 넥스원
dokyung.shin@lignex1.com†, youngdae.kim@lignex1.com

STOI Metric-Based Multi-Level Speech Enhancement and Data Augmentation

Do Kyuing Shin†, Young Dae Kim
LIG Nex1

요 약

최근 음성 신호 처리 분야에서 데이터 부족 문제 해결을 위한 음성 데이터 증강 기술이 주목받고 있다. 기존 GAN 기반 음성 향상 기법은 전통적인 손실 함수를 사용하여 인간의 지각적 품질과의 상관성이 낮고, 고품질 데이터에만 초점을 맞춰 일반화 능력이 부족하다는 한계를 갖는다. 본 논문에서는 STOI(Short-Time Objective Intelligibility) 메트릭을 활용한 다중 레벨 음성 향상 및 데이터 증강 모델을 제안한다. 제안된 모델은 MetricGAN 기반으로 5개의 품질 레벨(STOI: 0.55~0.95)에서 순차 학습을 수행하며, 생성자는 L1 손실과 메트릭 손실을 결합한 손실 함수로, 판별자는 STOI 점수를 예측하는 회귀 모델로 동작한다. Emo 데이터셋 실험 결과, 각 품질 레벨별로 목표 STOI에 근사한 음성 생성이 가능함을 확인하였으며, 스펙트로그램 분석을 통해 고품질 레벨에서 원본과 유사한 패턴 생성을 입증하였다. 제안된 방법은 다양한 품질의 데이터를 의도적으로 생성하여 모델의 견고성과 일반화 능력을 향상시키는 새로운 패러다임을 제시한다.

Ⅰ. 서 론

최근 음성 데이터 증강(Speech Data Augmentation)은 음성인식, 화자인식, 음성 합성 등 다양한 음성 신호 처리 분야에서 데이터 부족 문제를 해결하기 위한 핵심 기술로 주목받고 있다. 음성 데이터 증강 연구는 크게 신호 변형 기반(signal-level based), 특징 변환 기반(feature-level based), 생성 모델 기반(generative model-based)의 세 가지 접근 방식으로 분류된다. 생성적 적대 신경망(GAN, Generative Adversarial Networks)은 오디오 신호 처리 분야에서 광범위하게 활용되고 있으며, 특히 음성 신호 처리에서 상당한 성과를 달성하고 있다. 그러나 기존의 GAN 기반 음성 향상 기법들은 평균 제곱 오차(MSE, Mean Squared Error), STFT 차이, SNR과 같은 전통적인 신호 기반 손실 함수를 사용함으로써 실제 인간의 지각적 품질(perceptual quality)과의 상관성이 낮다는 근본적인 한계를 갖는다. 이러한 한계점을 해결하기 위해 Fu 등[1]은 MetricGAN을 제안하였다. MetricGAN은 음성 처리를 위한 새로운 GAN 기반 패러다임으로, 지각적으로 의미 있는 음성 품질 메트릭을 직접 최적화하는 방법을 적용한다. 기존 GAN의 판별자가 실제 샘플과 생성된 샘플을 구분하는 이진 분류를 수행하는 것과 달리, MetricGAN은 PESQ나 STOI와 같은 비미분 가능한 품질 지표를 근사하는 대리 모델(surrogate model)을 학습하여 생성자 훈련에 활용한다. 본 논문에서는 MetricGAN의 핵심 메커니즘을 기반으로 하여 품질 메트릭을 직접 손실 함수로 활용하는 음성 데이터 증강 모델을 제안한다. 품질 메트릭으로는 STOI(Short-Time Objective Intelligibility)를 사용하였으며, 이는 단시간 단위에서 깨끗한 음성과 향상된 음성 간의 상관관계를 통해 음성의 명료도를 측정하는 지표이다. STOI 점수는 0과 1 사이의 값을 가지며, 높을수록 더 나은 품질을 의미한다. 본 연구의 핵심 목표는 품질 메트릭을 활용하여 5단계 품질별 데이터 증강을 통해 다양한 품질의 데이터를 생성하는 것이다. 일반적으로 고품질 데이터만으로 학습된 모델은 과적합(overfitting)에 취약하며, 실제 환경의 다양성을 충분히 반영하지 못한다. 기존 연구들이 주로 잡음 제거 및 품질 향상에 초점을 맞춘 반면, 본 연구에서는 의도적으로 다양한 품질의 데이터를 생성하여 모델의 견고성과 일반화 능력을 향상시키는 새로운 패러다임을 제안한다.

Ⅱ. 제안한 방법

본 논문에서 제안하는 다중 레벨 품질 기반 오디오 데이터 증강 모델은 생성된 오디오의 품질 레벨에 따라 음성 향상 모델을 단계적으로 훈련하는 프레임워크를 채택한다. 학습 과정은 Algorithm 1에 제시된 바와 같이, N개의 서로 다른 STOI(Short-Time Objective Intelligibility) 목표값에 대해 순차적으로 모델을 훈련한다. 본 실험에서는 표 1과 같이 5개의 품질 레벨을 설정하였으며, 각 레벨의 STOI 목표값은 0.55, 0.65, 0.75, 0.85, 0.95로 구성된다. 제안된 모델의 생성자(Generator)는 잡음이 포함된 음성 신호를 입력으로 받아 향상된 음성을 생성한다. 생성자의 손실 함수는 수식 (1)과 같이 L_1 손실과 메트릭 손실의 결합으로 구성된다.

$$L_G = L_{1_}loss + \lambda \cdot MSE(stoi_pred, target_stoi) \quad (1)$$

수식 (1)에서 L_1 손실은 생성된 음성과 실제 깨끗한 음성 간의 차이를 최소화하며, 메트릭 손실은 판별자가 예측한 STOI 점수와 현재 레벨의 목표 STOI 값 간의 평균제곱오차(MSE)를 나타낸다. 이러한 결합 손실 함수를 통해 생성자는 단순한 음성 복원을 넘어서 특정 품질 수준에 맞는 음성을 생성하도록 학습된다. 제안 모델의 판별자(Discriminator)는 기존 GAN의 이진 분류 방식과 달리 STOI 점수를 예측하는 회귀 모델로 동작한다. 판별자는 입력된 음성 신호에 대해 STOI를 예측하며, 다음 손실 함수를 통해 훈련된다: 수식 (2)에서 $D(x)$는 판별자의 예측값이고, stoi_real은 깨끗한 음성과 향상된 음성 간에 계산된 실제 STOI 값이다. 이를 통해 판별자는 음성의 객관적 품질을 정확하게 평가할 수 있는 능력을 획득하며, 동시에 생성자에게 품질 개선을 위한 지도 신호를 제공한다.

표 1 Emo 데이터 셋의 각 클래스 별 5단계 품질 레벨에 따른 목표 STOI 대비 증강 데이터의 평균 STOI 비교

Lv.	Target STOI	Aug. STOI	STOI of class-specific augmented data						
			A	E	F	L	N	T	W
1	**0.55**	0.65	0.67	0.67	0.62	0.66	0.76	0.57	0.60
2	**0.65**	0.70	0.74	0.74	0.64	0.74	0.80	0.60	0.63
3	**0.75**	0.75	0.77	0.79	0.75	0.76	0.82	0.63	0.70
4	**0.85**	0.80	0.80	0.81	0.83	0.80	0.84	0.70	0.83
5	**0.95**	0.86	0.88	0.87	0.88	0.85	0.89	0.75	0.89

Algorithm 1. Multi-Level MetricGAN Training

```
For each target_STOI in [0.30, 0.46, 0.63, 0.79, 0.95]:
    Initialize G, D networks
    While not converged:
        # Generator training
        x_enhanced = G(x_noisy)
        stoi_pred = D(x_enhanced)
        stoi_real = STOI(x_clean, x_enhanced)
        L_G = L1_loss + λ * MSE(stoi_pred, target_STOI)

        # Discriminator training
        L_D = MSE(D(x), stoi_real)

        Update G, D networks
    Save model for current quality level
```

$$L_D = MSE(D(x), stoi_real) \quad (2)$$

두 네트워크는 상호보완적 학습을 통해 성능을 향상시킨다. 생성자는 판별자의 피드백을 활용하여 목표 품질 수준에 부합하는 음성 생성 능력을 학습하고, 판별자는 생성된 오디오의 품질을 평가함으로써 보다 정확하고 다양한 레벨의 품질 예측 능력을 발전시킨다. 각 품질 레벨에서 목표 품질에 수렴할 때까지 이 과정을 반복하여, 최종적으로 다양한 품질 수준의 음성 데이터를 생성할 수 있는 모델을 구축한다.

Ⅲ. 실험 결과

본 연구에서는 음성 발화 증강 실험을 위해 Emo 데이터셋을 사용하였다. 해당 데이터셋은 불안(A), 역겨움(E), 행복(F), 지루함(L), 중립(N), 슬픔(T), 분노(W)의 7가지 감정 클래스로 구성되어 있으며, 총 353개의 발화를 포함한다. 모든 입력 오디오는 16kHz 샘플링 레이트로 표준화하고 3초 단위로 분할하여 학습에 사용하였다. 증강된 데이터 또한 3초 길이로 생성하였으며, 각 원본 클래스 데이터 개수의 3배수로 증강을 수행하였다. 표 1은 각 클래스별 5개 레벨에 대한 증강 데이터의 STOI 평균값을 비교한 결과를 보여준다. 목표 STOI 품질 대비 증강된 데이터의 평균 STOI를 분석한 결과, 다음과 같은 특성을 확인하였다. 레벨 1과 레벨 2에서는 목표값 대비 과향상(over-enhancement) 경향이 나타났으며, 이는 저품질 목표에서 모델이 보다 적극적인 품질 개선을 수행하기 때문으로 분석된다. 레벨 3에서는 목표값과 증강 결과가 정확히 일치하여 가장 안정적인 수렴을 보였다. 반면 레벨 4와 레벨 5에서는 목표 대비 미달성(under-achievement) 결과를 나타냈으며, 이는 고품질 목표값에 대한 학습의 한계를 시사한다. 이러한 결과를 바탕으로 향후 고품질 구간에서의 성능 개선을 위해 적응적 손실 가중치 조정 및 계층적 학습 전략의 도입이 필요할 것으로 판단된다. 그림 1은 원본 발화 오디오와 5개 품질 레벨로 생성된 오디오의 스펙트로그램을 비교한 결과이다. 그림 1(a)는 원본 오디오, 그림 1(b)는 잡음 데이터를 나타낸다. 그림 1(c)-(g)는 각각 레벨 1부터 레벨 5까지의 증강 데이터 결과를 보여준다. 스펙트로그램 분석 결과, 고품질 레벨로 갈수록 증강된 데이터가 원본 데이터의 스펙트럼 패턴과 유사하게 생성되는 것을 확인할 수 있다. 이는 제안된 모델이 목표 품질 수준에 따라 적절한 품질의 음성 데이터를 생성할 수 있음을 시각적으로 입증한다.

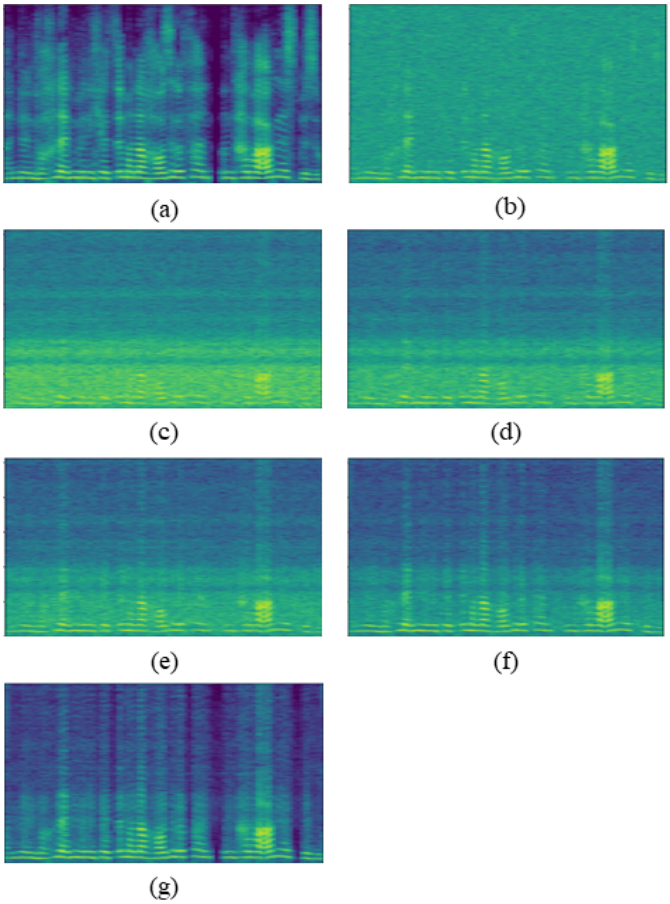


그림 1 'A 클래스의 원본 음성과 다중 레벨 품질로 생성된 증강 데이터의 스펙트로그램 비교: (a) 원본 음성, (b) 잡음 음성, (c) 레벨 1 (STOI=0.538), (d) 레벨 2 (STOI=0.642), (e) 레벨 3 (STOI=0.751), (f) 레벨 4 (STOI=0.839), (g) 레벨 5 (STOI=0.890)

참 고 문 헌

[1] Fu S. W., Liao C. F., Tsao Y., Lin S. D.,"MetricGAN: Generative Adversarial Networks based Black-box Metric Scores Optimization for Speech Enhancement," International Conference on Machine Learning (ICML), arXiv:1905.04874, 2019.

3D 모델 합성 데이터셋을 이용한 항공 영상 객체 탐지 연구

이도현, 김성용, 조규태

LIG 넥스원

dohyun.lee1@lignex1.com, sungyong.kim@lignex1.com, kyutae.cho2@lignex1.com

A Study on Aerial Image Object Detection Using a 3D Model Synthetic Dataset

Lee Do Hyun, Kim Sung Yong, Cho Kyu Tae

LIG NEX1

요 약

본 연구에서는 실데이터 확보와 라벨링의 어려움을 해결하기 위해 3D 기반 이미지 합성 기법을 활용한 객체 탐지 학습 데이터 자동 생성 방법을 제안한다. Blender로 실제 드론 촬영과 유사한 시점에서 렌더링한 객체와 해당 마스크를 이용해 배경 이미지에 자연스럽게 합성하고, 실제 객체 크기와 유사한 범위 내에서 크기와 위치를 무작위로 설정하여 강건성을 확보하였다. 합성과 동시에 바운딩 박스를 자동 기록하여 대규모 학습 데이터를 효율적으로 생성할 수 있으며, 합성 데이터만으로 학습한 YOLOX-s 모델이 테스트셋 기준 mAP@[0.50:0.95] 0.563을 달성하였다. 향후 실제 데이터도 3D 모델로 제작해 동일 방식으로 학습하면 성능 향상을 기대할 수 있다.

Ⅰ. 서 론

최근 항공 영상 기반의 객체 탐지 기술은 국방, 재난 대응, 정찰 등의 분야에서 높은 활용 가치를 보이며 활발히 연구되고 있다 [1]. 특히 드론을 활용한 고공 촬영 영상은 넓은 지역을 신속하게 확보할 수 있는 장점이 있어, 다양한 인공지능 기반 컴퓨터 비전 알고리즘의 학습에 효과적으로 활용되고 있다. 그러나 실제 항공 영상 데이터를 수집하고 학습에 적합한 형태로 라벨링하는 과정은 많은 비용과 시간이 요구되며, 다양한 환경과 시나리오를 충분히 반영하기 어렵다는 한계가 존재한다.

기존의 항공 시점 객체 데이터는 대부분 고정된 객체를 기준으로 드론이 일정한 경로를 따라 이동하며 촬영한 이미지로 구성되며, 배경이나 조명, 위치 조건의 다양성이 부족한 경우가 많다. 이로 인해 실제 환경에서 발생할 수 있는 복잡한 상황을 충분히 반영하지 못하며, 탐지 모델의 일반화 성능에도 제약을 준다.

이에 본 연구에서는 이러한 한계를 극복하기 위해 3D 모델 기반의 합성 데이터셋 제작 방식을 활용하였다. 학습 데이터는 관심 객체의 3D 모델을 Blender[2]에서 드론 촬영 시점과 유사하게 렌더링한 뒤, VBS(Virtual Battlespace) 시뮬레이션 환경에서 추출한 항공 배경 이미지에 무작위로 합성하였다. 이를 통해 다양한 각도, 크기, 배경 조건이 반영된 학습용 데이터를 구성하였다.

반면, test 데이터는 VBS 시뮬레이션 환경 내에 3D 객체를 직접 배치하고, 드론 시점에서 실제로 촬영하듯 데이터를 수집함으로써, 실제 운용 환경에 가까운 구성을 실현하였다.

본 연구에서 생성한 데이터셋만을 활용하여 객체 탐지 모델을 학습한 결과, 테스트셋 기준으로 mAP@[0.50:0.95] 0.563을 달성하였다. 이는 실제 데이터 수집 없이도 자동 라벨링과 다양한 시나리오가 반영된 고품질 학습 데이터를 통해 높은 수준의 성능을 확보할 수 있음을 정량적으로 입증하는 결과이며, 향후 실존 객체를 기반으로 한 pretraining이나 도메인 확장 연구에도 활용 가능성이 크다.

Ⅱ. 제안 방법

본 연구에서는 실제 항공 데이터를 수집하지 않고도 다양한 시나리오를 반영한 객체 탐지용 학습 데이터를 구축하기 위해, 3D 모델 렌더링과 시뮬레이션 환경 기반의 이미지 합성 파이프라인을 설계하였다. 이 파이프라인은 ① VBS 배경 이미지 확보, ② 3D 객체 렌더링 및 마스크 이미지 생성, ③ 이미지 합성과 자동 라벨링의 세 단계로 구성된다.

VBS 배경 이미지 확보

객체 합성에 사용되는 배경 이미지는 VBS 시뮬레이션 환경에서 수집한 항공 시점 이미지들 중 무작위로 선택하여 활용하였다. VBS는 다양한 지형과 조명 조건, 시점 구성이 가능하여 실제 드론 촬영을 대체할 수 있는 현실적인 배경을 제공한다. 본 연구에서는 객체가 배치될 환경적 다양성을 확보하기 위해 수집된 배경 이미지 중 랜덤하게 선택하여 합성에 사용하였다.

3D 객체 렌더링 및 마스크 생성

학습용 객체 이미지는 Blender를 이용해 렌더링하였다. 3D 모델은 일정 거리에서 시작하여 위쪽으로 각도를 변화시키며, 실제 드론이 상공에서 다양한 시점으로 촬영한 것처럼 보이도록 이미지를 생성하였다. 이때 객체와 배경을 분리하기 위한 마스크 이미지도 그림1과 같이 생성된다. 마스크 이미지는 이후 합성 과정에서 활용되며, 배경을 제거하고 객체만 배경 이미지 위에 삽입되도록 하는 역할을 수행한다. 이를 통해 객체 외의 배경 요소는 완전히 제거된 상태로 합성이 이루어지며, 자연스러운 외곽 표현이 가능하다. 또한, 객체 이미지는 경계에 맞게 자동으로 크롭되어 저장되므로, 불필요한 여백 없이 정확한 위치 정렬이 가능하다.

이미지 합성과 자동 라벨링

합성 단계에서는 렌더링된 객체 이미지와 해당 객체의 마스크를 활용하

여, 객체만을 배경 이미지 위에 합성한다(그림 2). 이때, 마스크를 이용하여 객체 외곽의 불필요한 영역을 제거하고, 배경과의 경계가 자연스럽게 어우러지도록 합성을 수행한다. 또한, 실제 객체 크기와 유사한 범위 내에서 객체의 크기를 무작위로 설정하고, 배경 이미지 내 위치 역시 무작위로 배치하여 다양한 시나리오에 대응할 수 있는 강건성을 높였다. 더불어, 합성과 동시에 객체의 바운딩 박스를 자동으로 계산하여 라벨링 정보를 기록한다. 이러한 과정은 대규모 학습 데이터를 수작업 라벨링 없이 생성할 수 있도록 하며, 실제 데이터 수집 없이도 효율적인 객체 탐지 모델 학습을 가능하게 한다.

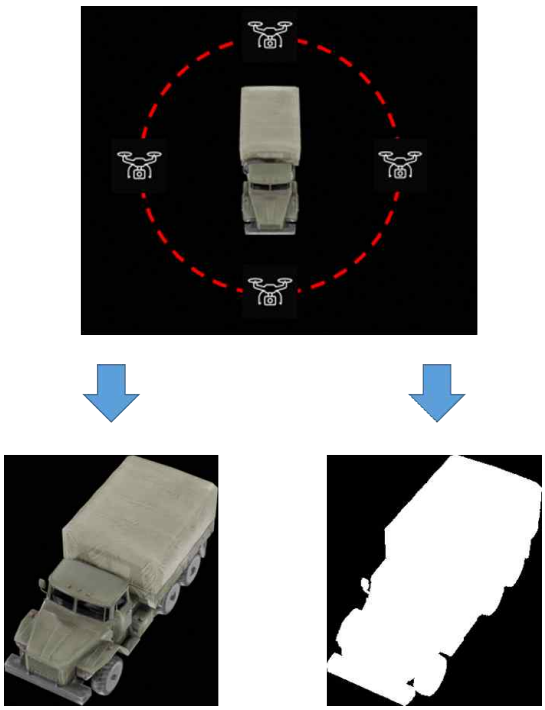


그림 1 3D 객체 렌더링 및 마스크 생성 이미지

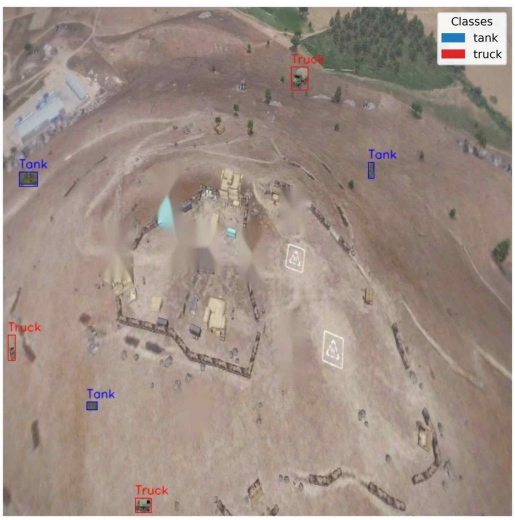


그림 2 합성 이미지 예시

Ⅲ. 실험 결과

본 연구에서는 구축한 합성 데이터셋을 활용하여 객체 탐지 모델의 학습과 평가를 수행하였다. 학습에는 매 epoch마다 새롭게 생성된 10,000장의 합성 이미지와 자동 생성된 라벨을 사용하였으며, 테스트는 VBS 시뮬레이션 환경에서 3D 객체를 직접 배치한 뒤 드론 시점으로 촬영한 총 2,000장으로 구성하였다. 이 중 400장은 검증(Validation)용으로 분할해 학습 중 체크포인트 선택에 사용하고, 나머지 1,600장으로 최종 테스트를 수행하였다. 객체 탐지기는 YOLOX-s(Megvii-BaseDetection, commit 6ddff48; 0.3.0 + 38 commits, 2025-06-08), 입력 640×640을 사용하였다[3]. 합성 데이터만으로 학습한 결과, 테스트셋 기준 mAP@[0.50:0.95] = 0.563을 기록하였다. 학습은 최대 300 epochs로 진행했으며, 검증 성능이 가장 높은 체크포인트를 테스트에 사용하였다 (표 1). 이는 실제 촬영 데이터를 사용하지 않고도 시뮬레이션 배경과 드론 시점 렌더링 기반의 3D 객체 합성만으로도 의미 있는 탐지 성능을 달성할 수 있음을 보여준다. 추후 실제 영상 수집과 실데이터 기반 3D 객체 제작이 이루어지면, 제안 방법을 재평가할 계획이다.

표 1 데이터 통계 및 학습 설정 요약

항목	내용
학습 데이터	합성 10,000장/epoch
테스트 데이터	1,600장
검증 데이터	400장
입력 해상도	640x640
모델	YOLOX-s
학습 epoch	300 (best-val ckpt 활용)
평가 지표	mAP@[0.50:0.95]

Ⅳ. 결론

본 연구는 실제 항공 데이터 수집의 제약을 극복하고 다양한 시나리오를 반영한 객체 탐지 학습 데이터 구축을 위해 3D 기반 이미지 합성 방식을 제안하였다. Blender 렌더링과 VBS 시뮬레이션 배경을 결합하여 자동 라벨링이 가능한 고품질 합성 데이터를 생성하였으며, 합성 데이터만으로도 mAP@[0.50:0.95] 0.563의 성능을 달성하였다. 비록 합성 데이터는 실제 데이터와의 도메인 차이로 인해 실환경 적용 시 한계가 존재하지만, 본 연구의 방법은 실제 데이터를 기반으로 정밀한 3D 모델을 제작하여 동일한 합성 절차를 적용함으로써 이러한 한계를 완화하고 성능을 향상시킬 수 있는 잠재력을 지닌다. 마스크 기반의 배경 제거와 자동 라벨링 구조는 다양한 객체·환경에 적용 가능하며, 향후 정밀 3D 모델링, 합성·실제 데이터의 혼합 학습, domain adaptation, 다양한 기상·시점 변화를 반영한 시뮬레이션 고도화를 통해 성능 향상을 기대할 수 있다. 종합적으로, 본 연구는 자동화된 합성 데이터셋 구축이 객체 탐지 분야에서 데이터 확보와 모델의 일반화 성능 향상의 실질적 대안이 될 수 있음을 제시한다.

참 고 문 헌

[1] Zhuang, S., Hou, Y., & Wang, D. (2025). Towards Efficient Object Detection in Large-Scale UAV Aerial Imagery via Multi-Task Classification. Drones, 9(1), 29.

[2] Blender Online Community. (2024). Blender - a 3D modelling and rendering package. Stichting Blender Foundation. Retrieved from https://www.blender.org

[3] Ge, Z., Liu, S., Wang, F., Li, Z., & Sun, J. (2021). Yolox: Exceeding yolo series in 2021. arXiv preprint arXiv:2107.08430.

최신 비전 트랜스포머 변형 모델 동향에 관한 연구

지훈*, 임한비**, 이연준***
*한양대학교 컴퓨터공학과(대학원생), **한양대학교 정보보호학과(대학원생), ***한양대학교 컴퓨터공학과(교수)

*greenpea0819@hanyang.ac.kr, **hanbi1@hanyang.ac.kr, ***yeonjoonlee@hanyang.ac.kr

A Survey on Recent Trends in Vision Transformer Variants

Hoon Ji*, Hanbi Yim**, Yeonjoon Lee***
*Hanyang University, Computer Science and Engineering (Master Student)
**Hanyang University, Department of Computer and Information Security (Master Student)
***Hanyang University, Computer Science and Engineering (Associate Professor)

요 약

Vision Transformer (ViT)는 기존의 합성곱 신경망(CNN) 중심의 패러다임에 새로운 방향을 제시하며 컴퓨터 비전 분야에 혁신을 가져왔다. 그러나, 입력 크기에 따른 2 차 연산 복잡도와 대규모 데이터셋에 대한 의존성이라는 명확한 한계를 지니고 있기에, ViT 의 아키텍처를 변형하여 이를 극복하려는 연구들이 활발히 진행되고 있다. 본 논문은 이러한 변형 ViT 모델들을 아키텍처 설계 전략에 따라 분류하고, 이에 속한 연구들을 분석한다. 본 논문에서는 변형 ViT 모델들을 접근법에 따라, (1) 어텐션 구조 효율화, (2) CNN-Tranformer 하이브리드 아키텍처, 그리고 (3) 토큰 처리 효율화라는 세 가지 기준으로 분류하고 최신 연구 동향을 조망한다. 각 분류에 속하는 주요 모델들의 핵심 아이디어와 기술적 차별점에 대해 분석함으로써 ViT 분야의 주요 해결 과제와 발전 방향에 대한 통찰을 제공하는 것을 목표로 한다. 나아가, ViT 가 목표로 삼아야 하는 주요 과제에 대해 제시함으로써 추후 연구에 관한 방향성을 제시한다.

Ⅰ. 서 론

Vision Transformer (ViT)는 자연어 처리에 특화된 트랜스포머 구조를 비전 태스크에 적용하여 [1], 기존의 합성곱 신경망 (CNN) 중심의 패러다임에 새로운 방향을 제시하였다. ViT 는 셀프 어텐션 매커니즘을 통해 이미지의 전역적인 문맥을 효과적으로 학습할 수 있음을 보여주었으나, 입력 크기에 따른 2 차 연산의 복잡도와 대규모 데이터셋 의존성이라는 명확한 한계점을 지니고 있으며, 이를 해결하기 위한 후속 연구들이 활발히 수행되고 있다 [2]. 따라서, 수많은 후속 연구들을 적절히 분류하고 분석할 수 있는 기준을 마련하는 것은 필수적이다.

본 논문은 최신 ViT 변형 모델들의 동향을 조망하고, 이들의 핵심적인 아키텍처 설계 전략에 대해 분석한다. 이를 위해, 우리는 ViT 의 한계점을 해결하는 전략에 따라 후속 연구들을 크게 (1) 어텐션 구조 효율화, (2) CNN-Transformer 하이브리드 아키텍처, 그리고 (3) 토큰 처리 효율화로 분류한다. 우리는 각 분류에 속하는 주요 모델들의 핵심 아이디어와 기술적 차별점에 대해 분석함으로써, ViT 분야의 동향 및 후속 연구의 방향에 대한 통찰을 제공하는 것을 목표로 한다.

Ⅱ. 본론

본론에서는 앞서 제시한 세 가지 분류 기준에 따라 최신 변형 ViT 관련 연구들을 분류하고, 각 연구들의 핵심적인 아이디어와 접근 방식을 살펴본다.

1. 어텐션 구조 효율화

초기 ViT 의 셀프 어텐션은 모든 토큰 간의 관계를 계산하기에, 입력 이미지의 크기가 커짐에 따라 연산량이 기하급수적으로 증가하는 2 차 복잡도 문제를 내재하고 있다. 이를 해결하기 위해, 어텐션의 계산 범위를 전체가 아닌 특정 지역으로 한정하거나 계층적 구조를 도입하여 연산 효율을 극대화하려는 연구들이 수행되었다

Liu et al. [3]은 이미지를 겹치지 않는 윈도우로 분할하고 그 내부에서만 어텐션을 수행하며, 연속된 블록 간에는 윈도우를 이동시켜 정보의 연결성을 확보하였다. 이러한 계층적 설계는 연산량을 선형 복잡도로 감소시켰을 뿐만 아니라, 다양한 비전 태스크의 핵심 백본으로 자리매김하였다. 이와 유사하게, Chu et al. [4]은 각 블록을 지역 그룹 어텐션과 다운샘플링된 토큰을 활용한 전역 어텐션으로 순차 구성했으며, Zhanget al. [5]은 NLP 의 Longformer 구조를 차용하여 밀집된 지역 어텐션과 희소한 전역 어텐션을 결합했다. 이러한 지역-전역 어텐션 분리 전략은 모델이 세밀한 지역 패턴과 넓은 문맥을 동시에 효율적으로 포착할 수 있게 한다. 한편 몇몇 연구들은 다중 해상도 정보를 병렬적으로 처리하는 구조를 제시했다. Yuan et al. [6]은 모델 전체에 걸쳐 여러 해상도의 스트림을 유지하며 지속적으로 정보를 교환해 고해상도 특징을 보존했으며, Lee et al. [7]은 각 단계에서 여러 스케일의 특징을 동시에 추출하는 다중 경로 구조를 통해 더욱 풍부한 공간 정보를 학습하도록 설계했다.

2. CNN-Transformer 하이브리드 아키텍처

기존 ViT 는 지역적 귀납 편향이 부족하여 데이터 의존성이 높다는 한계를 가지므로, 풍부한 귀납 편향을

통해 데이터 효율성을 확보한 CNN 의 장점을 결합하는 하이브리드 아키텍처가 활발히 연구되었다. 이러한 접근법은 CNN 으로 추출한 특징 맵을 ViT 의 입력으로 사용하는 직렬 방식과, 두 아키텍처에서 독립적으로 추출한 특징을 융합하는 병렬 방식으로 크게 나뉜다.

Guo et al. [8]은 모델 초반부에 CNN 으로 지역적인 특징을 추출한 후, 후반부에서 Transformer 를 통해 전역적인 관계를 학습하는 직렬 융합 모델을 제시한다. 반면, 병렬 융합 방식의 연구들인 Xu et al. [9]은 토큰화 및 다운샘플링 과정에 컨볼루션 연산을 통합하여 자연스러운 계층 구조를 형성했으며, Xia et al. [10]은 트랜스포머 블록 내에 다중 스케일 컨볼루션 분기를 병렬로 연결하여 지역 및 전역 정보의 상호작용을 극대화하는 데 집중했다. 추가적으로, Fan et al. [11]은 별도의 컨볼루션 모듈과 어텐션 모듈이 양방향으로 정보를 교환하는 방식으로 두 방식의 장점을 극대화하는 구조를 제시하였다.

3. 토큰 처리 효율화

앞서 말한 두 방식과 다르게, 모델의 연산량을 근본적으로 줄이기 위해 이미지의 내용에 따라 입력 토큰의 수를 동적으로 조절하는 연구들 또한 등장하였다. Yin et al. [12]은 어텐션 스코어를 기반으로 중요도가 낮은 토큰을 다음 레이어의 연산에서 제외시키는 '토큰 가지치기' 방식을 제안하였다. 이 접근법은 이미지의 배경과 같이 정보량이 적은 영역의 계산을 효율적으로 생략한다. 반면, Norouzi et al. [13]은 유사한 토큰들을 점진적으로 병합하는 구조를 제시한다. 초기 레이어에서는 인접 토큰을 합쳐 지역적 특징을 응축하고, 후기 레이어에서는 의미적으로 유사한 토큰들을 찾아 병합함으로써 정보 손실을 최소화하며 토큰 수를 줄이는 방식을 사용한다.

이처럼 입력에 따라 계산량을 조절하는 동적 처리 방식은 ViT 의 연산 패러다임을 고정적 구조에서 적응형 구조로 전환하여 모델의 실용성을 크게 높였다.

Ⅲ. 결론

본 논문은 초기 ViT 가 지닌 2 차 연산 복잡도와 데이터 의존성 문제를 해결하기 위한 ViT 변형 모델에 관한 최신 연구 동향을 세 가지 분류를 통해 분석하였다. 어텐션 구조 효율화는 연산 범위를 제한하여 계산 효율을 높이는 방식을 제안하였고, CNN-Transformer 하이브리드 아키텍처는 CNN 의 귀납 편향을 Transformer 에 결합하여 데이터의 효율성을 개선하였다. 마지막으로 토큰 처리 효율화는 입력에 따라 동적으로 계산량을 조절하는 새로운 패러다임을 제시하였다. 이러한 연구 흐름들은 ViT 가 단순 이미지 분류를 넘어 비전 태스크의 핵심 아키텍처로 발전하는 데 기여하였다. 향후 연구는 이러한 개별적인 효율화 기법들을 유기적으로 통합하여 시너지를 극대화하고, 제한된 자원을 가진 엣지 디바이스에서도 효과적으로 구동될 수 있는 초경량 모델을 개발하는 방향으로 나아가야 할 것이다.

ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원-대학 ICT 연구센터(ITRC)의 지원(IITP-2025-RS-2024-00438056, 50%)과 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(RS-2024-00341722, 지능형 서비스 로봇의 사이버 레질리언스 확보를 위한 보안기술 개발, 50%)

참 고 문 헌

[1] Alexey Dosovitskiy et al., "An Image is Worth 16x16Words: Transformers for Image Recognition at Scale",arXiv preprint arXiv:2010.11929, 2020

[2] K. Han et al., "A Survey on Vision Transformer,"IEEETransactions on Pattern Analysis and MachineIntelligence, vol. 45, no. 1, pp. 87-108, 2023

[3] Ze Liu, et al. "Swin transformer: Hierarchical visiontransformer using shifted windows."Proceedings of theIEEE/CVF international conference on computer vision,2021

[4] Xiangxiang Chu et al. "Twins: Revisiting the design ofspatial attention in vision transformers."Advances inneural information processing systems 3, 9355-9366,2021

[5] Pengchuan Zhang et al. "Multi-scale vision longformer:A new vision transformer for high-resolution imageencoding."Proceedings of the IEEE/CVF internationalconference on computer vision, 2021

[6] Yuhui Yuan et al. "Hrformer: High-resolution visiontransformer for dense predict."Advances in neuralinformation processing systems 34, 7281-7293, 2021

[7] Youngwan Lee et al. "Mpvit: Multi-path visiontransformer for dense prediction."Proceedings of theIEEE/CVF conference on computer vision and patternrecognition, 2022

[8] Jianyuan Guo et al. "Cmt: Convolutional neural networksmeet vision transformers."Proceedings of the IEEE/CVFconference on computer vision and pattern recognition,2022

[9] Yufei Xu et al. "Vitae: Vision transformer advanced byexploring intrinsic inductive bias."Advances in neuralinformation processing systems 34, 28522-28535, 2021

[10] Chunlong Xia et al. "Vit-comer: Vision transformerwith convolutional multi-scale feature interaction fordense predictions."Proceedings of the IEEE/CVFconference on computer vision and pattern recognition,2024

[11] Qihang Fan et al. "Lightweight vision transformer withbidirectional interaction."Advances in Neural InformationProcessing Systems 36, 15234-15251, 2023

[12] Hongxu Yin et al. "A-vit: Adaptive tokens for efficientvision transformer."Proceedings of the IEEE/CVFconference on computer vision and pattern recognition,2022

[13] Narges Norouzi et al. "Algm: Adaptive local-then-global token merging for efficient semantic segmentationwith plain vision transformers."Proceedings of theIEEE/CVF Conference on Computer Vision and PatternRecognition, 2024

조건 기반 배경 이미지 생성과 희귀 객체 합성에 관한 연구

박상영*, 정성윤, 김병현, 정남훈, 조규태
LIG넥스원
{sangyoung.park*, seongyun.jeong, byunghyun.kin, namhoon.jung, kyutae.cho2}@lignex1.com

A Study for Rare Object Synthesis with Conditioned Background Image Generation

Sangyoung Park*, Seongyun Jeong, Byunghyun Kim, Namhoon Jung, Kyutae Cho
LIG Nex1

요 약

본 연구는 다양한 환경에서의 희귀 객체를 포함한 2-stage 이미지 생성 방법을 제안한다. 기존 딥러닝 기반 이미지 생성 모델들은 대부분 일반적인 객체에 대한 학습에 집중되어 있어, 희귀 객체에 대해서는 생성 성능이 떨어지는 한계가 있다. 또한, 이러한 객체에 대해 인식 및 분류 등의 성능을 향상 시키기 위해서는 다양한 환경에서 해당 객체가 포함 되어있는 많은 이미지가 필요하며, 목적에 따라 특정 환경에서의 객체 이미지가 필요한 경우도 존재한다. 본 연구는 이러한 문제를 해결하기 위해 사용자가 지정한 위치에 생성할 물체를 text로 입력하여 이미지를 생성하는 생성 모델을 기반으로 배경 이미지를 생성한 뒤, 희귀 객체를 합성하는 방식으로 이미지를 생성한다. 객체 삽입 후 diffusion 모델을 이용해 객체의 구조는 보존하되 디테일의 다양성을 추가함으로써 희귀 객체에 대한 학습 데이터를 효율적으로 생성할 수 있으며, 보다 강건하고 일반화된 인식 및 분류 모델 학습에 기여할 수 있다.

Ⅰ. 서 론

딥러닝 기반 이미지 생성 기술은 최근 다양한 분야에서 활용되고 있으며, 특히 Generative Adversarial Networks (GAN)와 Diffusion 모델은 고해상도의 사실적인 이미지를 생성하는 데 있어 뛰어난 성능을 보이고 있다. 하지만 이러한 생성 모델은 일반적인 객체에 비해, 더 많은 세부 구조와 디테일을 요구하는 희귀 객체를 생성하는 데 한계가 존재한다. 이는 실제 응용 분야에서 희귀 객체의 인식 및 분류가 시스템의 성능과 안정성에 중대한 영향을 미친다는 점에서 중요한 문제로 작용한다.

더불어, 희귀 객체에 대한 데이터는 수집 자체가 어렵고, 시간, 날씨, 배경 등 다양한 환경 조건을 반영한 이미지를 확보하는 것은 더욱 큰 과제로 남아 있다. 이와 같은 한계점을 극복하기 위해 생성 모델을 통해 희귀 객체 이미지를 생성하는 다양한 연구 [1], [2]가 진행되고 있지만, 아직 근본적인 문제를 완전히 해결하지는 못한 상황이다. 특히 텍스트 기반 생성 모델의 경우, 단순한 프롬프트만으로 객체의 정확한 위치를 지정하는 데 어려움이 있어 실제 응용 가능한 정제된 데이터 생성을 제한하는 요인이 된다.

이러한 한계를 극복하기 위해, 본 연구에서는 사용자가 원하는 객체와 위치를 직접 지정할 수 있도록 지원하는 GLIGEN [3] 기반의 2-stage 이미지 합성 방법을 제안한다.

첫 번째 단계에서는 GLIGEN을 활용해 사용자가 지정한 위치에 생성할 물체 및 option prompt를 정해 배경 이미지를 생성하여 다양한 배경에서의 희귀 객체 이미지를 얻을 수 있도록 한다. 두 번째 단계에서는 희귀 객체를 해당 위치에 삽입한 뒤, 삽입된 객체의 patch를 diffusion 모델에 통과시켜 객체의 구조적 특성을 유지하면서 디테일을 자연스럽게 보완할 수 있도록 한다. 이와 같은 2-stage 이미지 합성 방식을 통해 학습 모델의 도메인 차이를 줄여 높은 일반화 성능을 기대할 수 있다.

Ⅱ. 본 론

본 논문에서는 비어있는 이미지 내에 bounding box를 지정하고 그 부분에 생성할 물체 및 option prompt를 입력하여 배경 이미지를 생성하는 “Background Generation” 및 사용자가 가지고 있는 Object 이미지를 생성한 배경 이미지 내 원하는 부분에 합성하는 “Synthesis Process” 과정을 통해 이미지 합성을 진행한다. 전체적인 생성 과정을 [그림 1]를 통해 나타내었다.

1-stage: Background Generation

해당 과정에서는 GLIGEN을 활용하여 희귀 객체가 배치될 배경 이미지를 만든다. 사용자는 배경 이미지 내의 특정 위치를 지정하여 해당 위치에 배치될 물체 및 option(환경)을 직접 설정할 수 있고 물체가 복수인 경우에도 적용 가능하다. 다만 위치를 겹치게 설정하는 경우 생성 품질이 떨어질 수 있다.

바다, 숲속, 도시 등과 같아 특정 환경 내에 존재하는 객체 이미지가 필요한 경우 유용하게 사용할 수 있으며, 조건에 맞게 랜덤하게 생성되는 배경 이미지들을 활용하여 다양한 환경에 희귀 객체를 합성할 수 있도록 배경 이미지를 생성한다.

Option의 경우 물체로 표현되기 어려운 형용사 및 배경 이미지의 스타일을 prompt 형태로 적어 생성하고자 하는 이미지에 디테일을 추가한다. 그 후, 생성된 배경 이미지 내에 Region Of Interest (ROI)를 지정하여 희귀 객체가 합성될 위치를 특정한다.

2-stage: Synthesis Process

해당 과정에서는 만들어진 배경 이미지에 희귀 객체를 합성하여 최종 이미지를 생성한다. 우선, 희귀 객체 이미지의 배경을 제거해 객체만 남긴 후 ROI 내에 해당 객체를 붙여 넣어 patch를 생성한다.

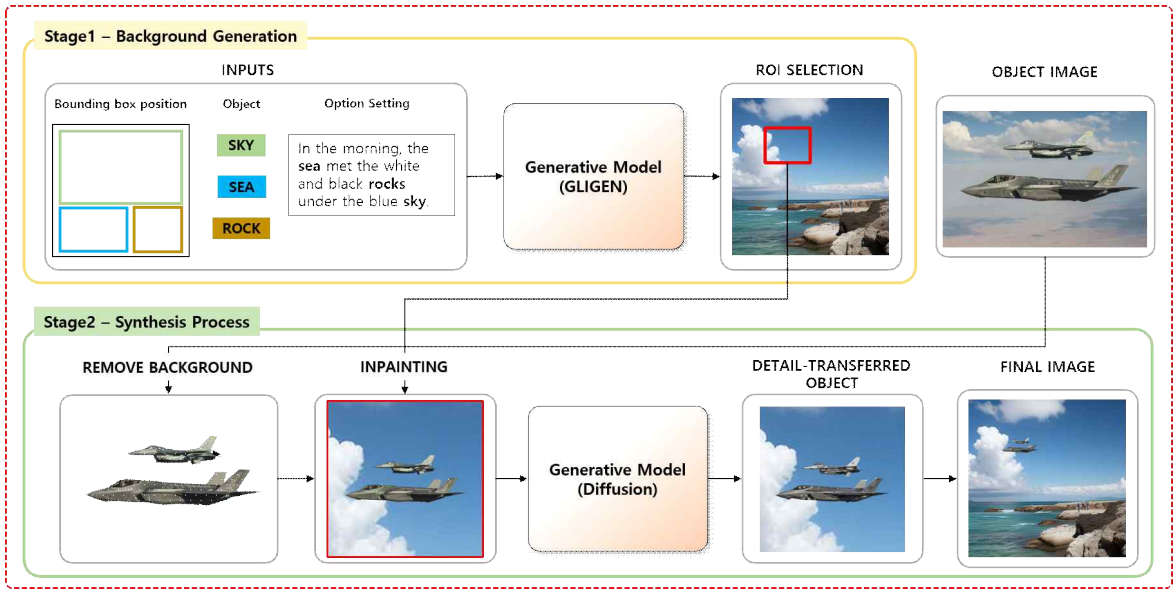


그림 1 전체 이미지 생성 과정

그 후, 해당 patch에 diffusion을 적용하여 객체의 특성을 유지하면서도 디테일이 변경될 수 있도록 한다. 이를 위해 원본 patch를 VAE에 통과시켜 latent space로 인코딩한 뒤, diffusion 과정을 적용할 때 활용한 후, 디코딩을 진행한다. [그림 2]에 원본 patch와 함께 diffusion 및 디코딩을 완료한 sample patch 2개의 확대 이미지를 나타내었다. 객체의 구조는 보존한 체 무늬, 색과 같은 디테일이 변경된 것을 확인할 수 있다.

마지막으로 해당 patch를 배경 그림의 ROI 위치에 붙여 넣어 최종 이미지를 생성한다.

그림 2 원본 patch 및 diffusion 과정 진행 후의 patch (확대)

Diffusion 과정 중 노이즈 제거량을 너무 높이게 되면 객체의 구조 자체가 변형되어 객체의 구조 보존의 의미가 사라지고 너무 낮춰 생성하면 객체의 디테일 변화가 거의 없어지기 때문에 적절한 hyperparameter를 설정하는 것이 중요하다. 또한, 과정 중에 사용되는 positive prompt 및 negative prompt는 공란으로 설정하였다.

해당 과정을 통해 매번 동일한 object를 붙여 넣는 경우와 비교해 모델이 다양한 representation을 학습할 수 있도록 하여 높은 일반화 성능을 기대할 수 있다.

Ⅲ. 실험 세팅

이미지 합성에는 다음과 같은 환경을 사용하였다.

Background Generation

diffusion model: endlessreality_v11

VAE: vae-ft-mse-840000-ema-pruned

Sampling: steps-12, cfg-12, euler sampler, normal echeduler

LoRA: add-detail-xl

Synthesis Process

diffusion model: DreamShaperXL_Lightning

VAE: DreamShaperXL_Lightning base VAE

Sampling: steps-5, cfg-2, euler ddpm_sde, karras scheduler

LoRA: add-detail-xl

Ⅳ. 결 론

본 논문에서는 2-stage 이미지 생성 및 희귀 객체 합성 방법에 대해 기술 하였다. Bounding box와 그 안에 생성될 물체, option prompt를 설정하여 배경 이미지를 쉽게 생성할 수 있어 원하는 환경에서의 객체 이미지를 획득할 수 있으며, 배경 이미지에 객체를 합성하는 경우에도 별도의 diffusion process를 거쳐 학습모델이 객체의 다양한 representation을 학습할 수 있도록 구성하였다.

응용 분야의 이미지가 일반적이고 획득이 어렵지 않은 경우라면 사진을 촬영 하거나 [4]와 같이 생성 모델을 이용하는 등의 방법을 통해 추가적인 성능 향상을 기대할 수 있지만, 군사·의료 등과 같이 한정된 domain이나 fine-grained class에 대한 학습이 필요한 경우와 같이 해당 방법이 제한되는 경우에 사용할 수 있을 것으로 기대한다.

참 고 문 헌

[1] Pan, Z., et al., “FineDiffusion: scaling up diffusion models for fine-grained image generation with 10,000 classes,” Applied Intelligence, vol. 55, no. 5, p. 309, 2025.

[2] Monsefi, A. K., et al., “TaxaDiffusion: Progressively Trained Diffusion Model for Fine-Grained Species Generation,” arXiv preprint arXiv:2506.01923, 2025.

[3] Li, Y., et al., “Gligen: Open-set grounded text-to-image generation,” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 22511 - 22521, 2023.

[4] K. Islam et al., “DiffuseMix: Label-Preserving Data Augmentation with Diffusion Models,” arXiv preprint arXiv:2405.14881, 2024.

업종별 데이터 변동성에 따른 최적 머신 러닝 모델 선택 전략 : 서울시 상권 예측 사례

이하나
고려대학교 SW·AI 융합대학원
hanalee11@korea.ac.kr

Optimal Machine Learning Model Selection Strategy Based on Industry-Specific Data Volatility: A Case Study of Seoul Commercial District Prediction

Lee Ha Na
Korea University Graduate School of SW·AI convergence

요 약

본 연구는 업종별 데이터 변동성에 따른 최적 머신러닝 모델 선택 전략을 제안한다. 서울시 상권분석 데이터를 음식점업, 소매업, 개인서비스업으로 세분화하고, AutoML 기법을 활용하여 각 업종별로 8 개의 머신러닝 알고리즘 중 최적 모델을 탐색하였다. 2023 년부터 2024 년까지의 실험 결과, 데이터 변동성에 따라 최적 알고리즘이 체계적으로 달라짐을 확인하였다. 저변동성 업종(개인서비스업, σ=0.879)에서는 ExtraTrees 가, 중변동성 업종(음식점업, σ=2.253)에서는 VotingRegressor 가, 고변동성 업종(소매업, σ=4.622)에서는 Ridge 가 각각 최적 성능을 보였다. 이를 통해 **변동성-복잡도 역상관 법칙과 유사업종 변화 지배 법칙**을 발견하였으며, 실무에서 활용 가능한 모델 선택 가이드라인을 제시하였다.

I. 서론

본 논문에서는 도시 상권의 변화 예측은 창업자의 사업 계획 수립, 임대업자의 투자 결정, 그리고 지방자치단체의 정책 수립에 중요한 정보를 제공한다. 특히 서울시와 같은 대도시에서는 상권의 급속한 변화와 업종별 상이한 특성으로 인해 정확한 예측이 더욱 중요해지고 있다.

기존 연구들은 주로 전체 상권을 통합적으로 분석하는 접근법을 사용해왔으나, 이는 업종별 고유한 특성과 영향 요인의 차이를 충분히 반영하지 못하는 한계가 있다. 예를 들어, 음식점업은 식사 시간대 유동인구와 밀접한 관련이 있는 반면, 개인서비스업은 거주 인구의 안정성에 더 영향을 받는다.

본 연구의 목적은 AutoML 기법을 활용하여 업종별 데이터 특성에 따른 최적 머신러닝 모델 선택 전략을 제시하는 것이다. 특히 데이터 변동성과 최적 알고리즘 간의 관계를 체계적으로 분석하여, 실무에서 활용 가능한 모델 선택 가이드라인을 개발하고자 한다.

II. 본론

1. 관련 연구

1.1 상권 분석 연구

강민경(2022)은 서울시 상권변화 유형도출 및 예측모델 연구에서 상권의 시계열적 변화 패턴을 분석하였으나, 업종별 특화 접근보다는 전체적인 변화 유형 분류에 중점을 두었다.[1] 김지연 등(2024)은 딥러닝 기반 서울시 행정동별 외식업종 상권 변화 예측 연구를 통해 외식업종에 특화된 딥러닝 모델을 제안하였으나, 단일 업종에 국한된 접근이었다.[2] 기존 상권 분석 연구들은 대부분 전체 업종을 통합하거나 단일 업종에만 집중하는 접근법을 취해왔다. 이러한 접근법들은 각각의 장점이 있으나, 업종 간 특성 차이를 고려한 선택적 모델링 전략은 제한적으로만 시도되었다.

1.2 머신러닝 기반 도시 예측

도시 데이터 분석 분야에서 Random Forest 와 XGBoost 는 높은 예측 성능으로 널리 활용되고 있다. 특히 공간 정보와 시계열 특성이 복합된 도시 데이터에서 앙상블 기법의 효과가 입증되었다.

1.3 업종별 특화 모델링

소매업 분야에서는 이미 고객 세분화 기반 개별 모델링의 효과가 검증되었으나, 상권 예측 분야에서는 업종별 최적 알고리즘 탐색이 제한적으로만 시도되었다. 특히 AutoML 을 활용한 체계적 알고리즘 비교를 통해 데이터 특성과 모델 성능의 관계를 분석한 연구는 부족한 실정이다. 본 연구는 이러한 연구 공백을 채우고자 한다.

2. 연구 방법론

2.1 데이터 수집 및 전처리

본 연구는 서울시에서 제공하는 상권분석 공공데이터를 활용하였다. 2023 년부터 2024 년까지 2 년간의 분기별 데이터를 수집하여 총 8 개 시점을 분석하였다.

업종 분류:

- 음식점업: 한식, 중식, 일식, 서양식 음식점 (4 개 하위 업종)
- 소매업: 편의점, 슈퍼마켓, 의복소매, 화장품소매 (4 개 하위 업종)
- 개인서비스업: 미용업, 세탁업, 수선업, 애완동물서비스 (4 개 하위 업종)

2.2 업종별 특화 변수 설계

각 업종의 특성을 반영한 맞춤형 변수를 다음과 같이 설계하였다:

공통 특화 변수:

- 프랜차이즈비율 = 프랜차이즈점포수 / 전체점포수
- 회전율 = (개업점포수 + 폐업점포수) / 전체점포수
- 업종집중도 = 해당업종점포수 / 행정동내전체점포수
- 시계열변화량 = 현재분기 - 이전분기 점포수 변화

유사업종 변화 변수의 정의와 중요성:

유사업종 변화는 본 연구의 핵심 발견 중 하나로, 해당 행정동 내에서 동일한 업종군에 속하는 모든 점포의 분기별 변화량을 의미한다. 예를 들어, 특정 동의 음식점업 예측 시, 해당 동 내 모든 음식점(한식, 중식, 일식, 서양식)의 총 점포수가 이전 분기 대비 얼마나 증감했는지를 나타낸다.

이 변수가 중요한 이유는 다음과 같다.

1. **상권 생태계 효과**: 동일 업종 점포들은 고객층과 입지 조건을 공유하므로, 한 지역의 업종 전체 트렌드가 개별 점포에 직접적 영향을 미침
2. **시장 포화도 신호**: 유사업종 증가는 시장 기회를, 감소는 시장 위축을 신호함
3. **정책/환경 변화 반영**: 지역 개발, 규제 변화 등이 업종 전체에 동시에 영향을 미치는 양상을 포착

2.3 AutoML 기반 모델 설계

2.3.1 포괄적 알고리즘 탐색 및 입출력 구성 각 업종별로 8 개의 머신러닝 알고리즘을 체계적으로 비교 평가하였다. 모든 알고리즘은 동일한 입출력 구성을 갖는다:

공통 입력 구성(X):

- 10 개 특성 변수: [점포수, 유사 업종 점포수, 프랜차이즈점포수, 프랜차이즈_비율, 개업비율, 폐업비율, 회전율, 업종집중도, 점포수 변화, 유사업종 변화]
- 데이터 형태: 연속성 수치 데이터 (정규화 미적용)
- 입력 차원: (n_samples, 10) - 각 업종별 샘플 수는 약 13,000 개

공통 출력 구성(Y):

- 목표 변수: 순개업점포수 (연속형 회귀 문제)
- 값 범위: 음식점업(-86~43), 소매업(-268~131), 개인서비스업(-11~13)

2.3.2 업종별 최적 모델 선택 각 업종의 데이터 특성에 맞는 최적 알고리즘을 자동으로 탐색하고, 3-fold 시계열 교차검증을 통해 성능을

검증하였다. 모든 알고리즘은 동일한 평가기준(RMSE)으로 비교되며, 최저 RMSE 를 달성한 모델을 해당 업종의 최적 모델로 선정하였다.

2.3.3 데이터 변동성 기반 모델 선택 전략 업종별 목표변수의 표준편차를 기준으로 다음 가설을 검증하였다:

- 저변동성($\sigma < 2.0$): 복잡한 비선형 모델이 효과적
- 중변동성($2.0 \leq \sigma < 4.0$): 앙상블 기법이 적합
- 고변동성($\sigma \geq 4.0$): 정규화된 단순 모델이 안정적

3. 실험 결과

3.1 포괄적 AutoML 기반 최적 모델 탐색 결과

8 개의 최신 머신러닝 알고리즘을 포함한 포괄적 AutoML 탐색을 실시하였다. 다중 평가지표를 통해 모델 성능을 종합적으로 평가하였다.

표 1. 업종별 포괄적 AutoML 결과 (다중 지표)

업종	최적 모델	RMSE	MAE	R^2	방향성 정확도
음식점업	VotingRegressor	1.147	0.117	0.845	0.829
소매업	Ridge	0.735	0.214	0.908	0.896
개인서비스업	ExtraTrees	0.200	0.020	0.962	0.987

특히 개인서비스업에서 가장 높은 예측 성능을 달성하였으며(R^2 = 0.962), 방향성 예측 정확도도 98.7%로 실무적으로 매우 유용한 수준을 보였다. 반면 변동성이 높은 소매업에서도 Ridge 모델을 통해 89.6%의 방향성 정확도를 확보하였다.

3.2 데이터 변동성 구간의 객관적 설정

K-means 클러스터링 분석을 통해 업종별 데이터 변동성을 3 개 구간으로 객관적 분류하였다.

표 2. 클러스터링 기반 변동성 구간 설정

클러스터 ID	중심점(σ)	포함 업종	클러스터 특성
클러스터 2	0.879	개인서비스업	저변동성 그룹
클러스터 0	2.253	음식점업	중변동성 그룹
클러스터 1	4.622	소매업	고변동성 그룹

K-means 클러스터링 결과, 각 업종이 서로 다른 클러스터에 명확히 분류되어 **변동성 기반 구간 설정의 통계적 근거**를 확보하였다. 이는 경험적 구간 설정이 아닌 데이터 기반 객관적 분류임을 입증한다.

3.3 최적 모델별 변수 중요도 분석

최적 알고리즘으로 선정된 모델들의 feature importance 를 분석한 결과, **모든 업종에서 "유사업종_변화"가 최상위 중요도를 보이는 공통 패턴**을 발견하였다.

표 3. 업종별 최적 모델의 상위 5 개 중요변수

순위	음식점업 (VotingRegressor)	소매업 (Ridge)	개인서비스업 (ExtraTrees)
1	유사업종_변화 (0.751)	유사업종_변화 (0.916)	유사업종_변화 (0.469)
2	점포수_변화 (0.143)	폐업비율 (0.564)	점포수_변화 (0.323)
3	폐업비율 (0.031)	개업비율 (0.549)	개업비율 (0.085)
4	개업비율 (0.024)	업종집중도 (0.249)	폐업비율 (0.070)
5	프랜차이즈_점포_수 (0.023)	점포수_변화 (0.101)	회전율 (0.024)

핵심 발견인 "유사업종_변화 지배 법칙"의 실무적 의미: 모든 업종에서 유사업종_변화가 압도적 1 위를 차지하는 현상은 상권 예측에서 다음과 같은 중요한 시사점을 제공한다.

1. **생태계 관점의 예측**: 개별 점포의 성공/실패는 해당 점포의 특성보다 그 지역 업종 전체의 성장/쇠퇴 패턴에 더 크게 좌우된다. 즉, "숲을 보고 나무를 예측하는" 접근이 효과적이다.
2. **선행 지표 역할**: 유사업종의 변화는 시장 트렌드, 소비자 선호 변화, 정책 효과 등을 종합적으로 반영하는 선행 지표로 작용한다.
3. **업종별 민감도 차이**: 소매업(91.6%)이 가장 높은 중요도를 보이는 것은 소매업이 경기 변동, 온라인 쇼핑 트렌드 등 외부 환경 변화에 가장 민감하게 반응함을 의미한다.

이러한 발견은 창업자나 투자자들이 개별 점포의 내부 요인보다 **해당 지역 내 동종 업계의 전반적 동향**을 우선적으로 모니터링해야 함을 시사한다.

4. 토론 및 활용 방안

4.1 연구 결과의 의미

본 연구 결과는 다음과 같은 중요한 의미를 갖는다:

1. **업종별 최적 알고리즘의 다양성**: AutoML 탐색을 통해 업종별로 최적 알고리즘이 현저히 다름을 확인하였다. 특히 앙상블 기법(VotingRegressor), 극도 랜덤화 기법(ExtraTrees), 정규화 기법(Ridge)이 각각 다른 변동성 특성에 최적화되어 있음을 발견하였다.
2. **변동성-복잡도 역상관 법칙**: 데이터 변동성이 증가할수록 최적 모델의 복잡도가 감소하는 명확한 패턴을 발견하였다. 이는 고변동성 데이터에서 과적합 방지가 성능 향상의 핵심임을 시사한다.
3. **유사업종 변화의 지배적 영향**: 모든 업종에서 유사업종_변화가 최상위 중요도를 보여, 시계열 변화 패턴이 AI 모델 성능에 결정적 영향을 미침을 입증하였다.

4.2 업종별 모델 적용 전략

데이터 변동성에 따른 체계적 모델 선택 가이드라인을 다음과 같이 제시한다.

표 4. 실무용 모델 선택 가이드라인 (클러스터링 검증 기반)

변동성 분류	표준편차 범위	권장 모델	예상 성능	방향성 정확도	적용 사례
저변동성	$\sigma \approx 0.9$	ExtraTrees	R^2>0.95	>98%	안정적 서비스업
중변동성	$\sigma \approx 2.3$	Voting Regressor	R^2>0.84	>82%	일반 상업 시설
고변동성	$\sigma \approx 4.6$	Ridge	R^2>0.90	>89%	변동성 큰 소매업

본 가이드라인은 사전 데이터 분석(표준편차 계산)만으로도 최적 모델 유형을 예측할 수 있어, 실무에서 효율적인 AutoML 모델 선택이 가능하다.

4.3 AI 모델 선택 방법론의 확장 가능성

- **다른 시계열 데이터:** 본 연구의 변동성 기반 모델 선택 전략은 시계열 특성을 가진 다른 데이터에서도 활용 가능성이 있음
- **AutoML 도구 개선:** 데이터 변동성을 자동 분석하여 최적 알고리즘을 추천하는 지능형 AutoML 시스템 구축에 기여
- **적응형 모델링:** 데이터 특성 변화에 따라 동적으로 모델을 전환하는 적응형 시스템 개발의 이론적 기반 제공

4.4 한계점 및 향후 연구 방향

본 연구의 한계점은 서울시 상권 데이터라는 단일 도메인에 국한되어 일반화 가능성이 제한적이라는 점이다. 향후 연구에서는 다양한 도메인의 데이터를 활용하여 변동성-복잡도 역상관 법칙의 범용성을 검증하고, 더 많은 최신 알고리즘을 포함한 확장된 AutoML 프레임워크를 개발할 예정이다.

III. 결론

본 연구는 AutoML 기법을 활용하여 업종별 데이터 변동성에 따른 최적 머신러닝 모델 선택 전략을 제시하였다. 8 개의 머신러닝 알고리즘을 체계적으로 비교한 결과, 주요 발견은 다음과 같다:

핵심 기여:

1. **변동성-복잡도 역상관 법칙**: 저변동성 업종(ExtraTrees, R^2=0.962) → 중변동성 업종(VotingRegressor, R^2=0.845) → 고변동성 업종(Ridge, R^2=0.908)으로 데이터 변동성이 증가할수록 최적 모델의 복잡도가 체계적으로 감소함을 발견
2. **유사업종 변화 지배 법칙**: 모든 업종에서 유사업종_변화가 최상위 중요도(46.9~91.6%)를 보여, 시계열 변화 패턴이 AI 모델 성능에 결정적 영향을 미침을 입증
3. **포괄적 AutoML 검증**: 전통적 방법론부터 최신 앙상블 기법까지 8 개 알고리즘 비교를 통해 업종별 최적 모델이 현저히 다름을 체계적으로 검증

이는 기존의 "단일 알고리즘 적용" 접근법의 한계를 극복하고, **데이터 특성 기반 알고리즘 선택**의 중요성을 입증한다. 특히 K-means 클러스터링으로 검증된 변동성 구간을 통해 최적 모델 유형을 사전 예측할 수 있는 실무적 가이드라인을 제공하여, AutoML 적용의 효율성을 크게 향상시켰다.

향후 연구에서는 더 다양한 시계열 데이터로 확장하여 변동성 기반 모델 선택 이론의 일반화 가능성을 검증할 계획이다.

참 고 문 헌

[1] 강민경. (2022). "서울시 상권변화 유형도출 및 예측모델을 활용한 변화예측." 석사학위논문, 한양대학교.

[2] 김지연, 오수민, 박민서. (2024). "딥러닝 기반 서울시 행정동별 외식업종 상권 변화 예측 연구." 국제문화기술진흥원, 10(2), 459-463.

[3] 서울특별시. (2024). "서울시 상권분석 서비스 데이터." Retrieved from https://data.seoul.go.kr/

분산전력 시스템을 위한 강화학습 기반 CVR 제어 및 에너지 절감 효과 연구

이주승*, 이순우, 강지명
한국전기연구원
jseunglee@keri.re.kr, rheesw@keri.re.kr, jmkang@keri.re.kr

A Study on Reinforcement Learning-Based CVR Control and Energy Saving for Distributed Energy Systems

Jooseung Lee, Soonwoo Lee, Jimyung Kang
Korea Electrotechnology Research Institute

요 약

현재 탄소중립과 환경 보호 정책의 확산에 따라 재생에너지 기반의 분산전력 시스템이 활발히 도입되고 있다. 이러한 분산전력 시스템에서는 다양한 제어요소를 활용하여 전압 수준을 효율적으로 조절함으로써 부하 전력을 절감하는 보존 전압 강하 기법이 주목받고 있다. 본 논문에서는 강화학습을 이용하여 변압기의 tap을 일정 수준의 저전압 상태로 고정한 상황에서도 에너지 저장장치를 적절히 제어함으로써 전압 강하에 따른 부하 전력 절감을 실현하는 방안을 제안한다. 제안된 방법을 통해 다양한 부하 조건에서 저전압 상황을 방지하면서 동시에 에너지 소비를 줄일 수 있으며, 향후 분산전력 시스템 내에서 실시간 지능형 전압 제어 및 보존 전압 강하 수행을 위한 기반 기술로 활용될 수 있을 것으로 기대된다.

Ⅰ. 서 론

최근 탄소중립과 환경 보호를 위한 정책이 강화됨에 따라, 기존의 중앙집중형 전력 시스템에서 벗어나 분산전력 시스템의 도입이 활발히 이루어지고 있다 [1]. 특히 태양광, 풍력 등 재생에너지의 확산으로 인해, 지역 단위의 전력 품질 관리와 에너지 효율 향상이 중요한 과제가 되었다. 이러한 환경에서 보존 전압 강하(Conservation Voltage Reduction, CVR) [2] 기법은 부하의 전압 민감도 특성을 활용하여 소비 전력을 줄이는 효과적인 방법으로 주목받고 있다. 일반적으로 부하의 전압-전력 관계는 ZIP 부하 모델을 통해 표현되며 [3], ZIP 모델에 따르면 일정 범위 내에서 전압을 낮출 경우 유효전력 소비가 줄어들 수 있지만, 변압기의 tap 위치를 무분별하게 하강시킬 경우 일부 구간에서 저전압이 발생하여 부하 오동작이나 전력 품질 저하가 발생할 수 있다.

이에 따라 본 논문에서는 변압기의 tap을 일정 수준 이하로 고정한 상태에서, 에너지 저장장치(Energy Storage System, ESS)를 능동적으로 제어함으로써 저전압 문제를 방지하고, 동시에 CVR을 수행하는 방안을 제안한다. ESS는 부하 전력을 보조하는 기능을 통해 이러한 상황에서 효과적으로 작동할 수 있으며, 이와 같은 제어를 위해 최근 강화학습 기반의 제어 기법이 활발히 연구되고 있다 [4]. 이 중에서도 Deep Q-Network (DQN)은 비교적 단순한 구조에도 불구하고 우수한 학습 안정성과 제어 성능을 보여주며 다양한 에너지 응용 분야에 적용되고 있다 [5][6]. 본 연구에서는 DQN을 기반으로 ESS 제어 정책을 학습하고, 분산전력 시스템 환경에서의 효과적인 CVR 수행 가능성을 실험을 통해 검증한다.

Ⅱ. 시스템 구성 및 제어 목표

본 연구에서 실험에 사용된 분산전력 시스템의 구조는 그림 1과 같다. 해당 시스템은 그리드에 연결된 변압기, 두 개의 부하, 태양광 발전원(Photovoltaic, PV), 그리고 ESS로 구성되어 있으며, ESS는 부하1과 직접 연결되어 있다. 각 부하의 전력 소비 특성은 ZIP 부하 모델을 기반으로 하며, 부하1의 ZIP 계수는 40:30:30, 부하2는 20:30:50으로 설정하였다.

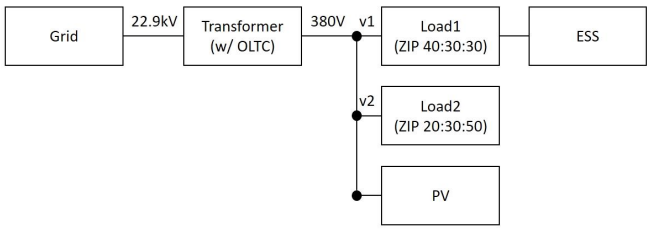


그림 1 Simplified Microgrid Circuit.

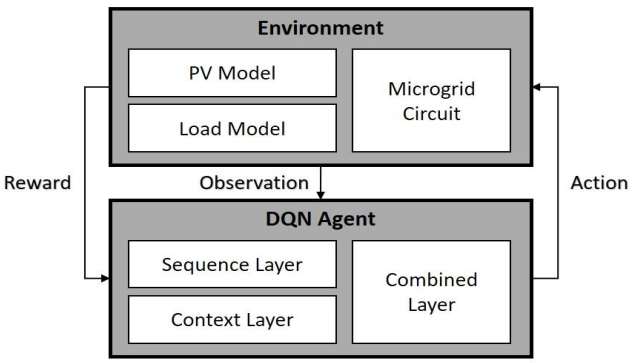


그림 2 Reinforcement Model Architecture.

제어 주기는 15분 간격으로 설정하였으며, 제어 목적은 각 부하 노드의 전압이 0.93 p.u. 이상을 유지하는 조건 내에서, 가장 낮은 전압 상태를 유도하여 부하를 절감하는 것이다. 변압기는 neutral tap 포지션 13을 기준으로 설정되어 있으며, 각 tap 포지션은 약 0.52%의 전압 조절 범위를 가진다. ESS 제어 없이 0.93 p.u. 이상의 전압을 유지하려면 tap 포지션을 17로 설정해야 하며, 본 실험에서는 이를 레퍼런스 상태로 정의하였다. 최종적으로 본 연구에서는 tap 포지션을 14로 고정한 상태에서 ESS를 제어하여 전압 저하를 방지하면서 동시에 부하 절감을 수행한다. 전력 시뮬레이션은 Python 기반의 pandapower 라이브러리를 사용하여 수행되었다.

Ⅲ. 강화학습 기반 제어 모델 구조

본 논문에서는 위에서 정의한 시스템을 기반으로, DQN 구조를 활용한 강화학습 기반의 제어 모델을 설계하였다. 전체 모델은 그림 2와 같이 environment와 agent로 구성되며, agent는 policy network와 target network로 분리된 표준 DQN 구조를 따른다. Action space는 ESS의 충/방전 전력을 의미하며, -50kW부터 +50kW까지 10kW 간격으로 총 11개의 불연속 action으로 구성된다.

Agent의 정책 결정은 미래의 예측 sequence data와 현재 상태 context data를 입력으로 받아 수행된다. Sequence data는 미래의 Load1, 2, PV의 출력값이며, context data는 현재의 시간 정보, ESS 상태, ESS 제어가 없는 경우의 각 노드별 전력 및 전압값으로 구성된다. Sequence data는 DQN에서 Long Short-Term Memory (LSTM) layer를 거쳐 압축되며, context data는 Multi-Layer Perceptron (MLP) layer를 통해 처리되고, 두 처리된 중간 출력은 MLP로 구현된 combined layer로 입력되어 각 action에 대한 value값을 출력한다.

한편 보상 함수는 다음의 요소들을 종합하여 정의하였다:

1. 각 노드의 전압이 0.93 p.u. 미만으로 하락했는지 여부
2. 6시간 및 24시간 이동 평균 전력 소비량
3. Episode 종료 시 누적 전력 소비량
4. 종료 시점 ESS의 상태와 초기 ESS 상태의 차이

Ⅳ. 실험 및 결과 분석

DQN 기반 제어 모델의 효과를 평가하기 위해, 총 5일치 data를 1개 episode로 설정하여 학습을 수행하고, 성능을 평가하였다. 시뮬레이션 결과 중 대표적인 하루의 전압 변화에 따른 부하를 그림 3에 나타내었다. 해당 그림에서는 ESS 제어가 없는 경우와 DQN-ESS 제어가 활성화된 경우를 비교하였으며, CVR 제어에 따른 부하 절감 효과를 시간대 별로 확인할 수 있다. DQN agent는 부하가 적은 구간에서 ESS를 충전해두고, 전압 저하가 예상되는 부하가 큰 시점에 방전하여 노드 전력을 보상함으로써, 전체 소비 전력을 낮추면서 동시에 전압 위반을 방지하는 형태로 동작하도록 학습된 것을 알 수 있다. 이와 같은 제어 전략을 통해 전압 범위 위반 없이 부하 절감 효과를 얻을 수 있었으며, 실험을 통해 CVR의 실제 적용 가능성을 확인할 수 있었다.

평가 data 중 앞의 4일 간의 data에서 레퍼런스 상태 대비 하루 단위 소비 전력 절감률은 날짜 별 부하 상태에 따라 1일차에 1.2%, 2일차에 1.0%, 3일차에 1.0%, 4일차에 0.7%로 나타났으며, 전체 평균으로 1.0%의 소비 전력 절감이 전압 조건 위반 없이 이루어졌다. 한편 학습 과정에서 보상의 변화 추이는 그림 4와 같이 점진적으로 개선되는 형태를 보였다.

Ⅴ. 결론

본 연구에서는 분산전력 시스템 환경에서 CVR을 통한 에너지 절감을 달성하기 위해, 강화학습 기반의 ESS 제어 기법을 제시하고 실험적으로 그 성능을 검증하였다. 변압기의 tap을 낮은 위치에 고정한 상태에서 DQN agent를 이용해 ESS를 제어함으로써, 전압이 0.93 p.u. 이상 유지되도록 하면서 동시에 소비 전력을 최소화하는 제어 정책 학습이 가능하였고, 정량적으로는 약 1% 수준의 전력 절감 효과를 안정적으로 달성하였다.

향후에는 본 기법을 보다 복잡한 전력 시스템에 확장 적용하고, Double DQN이나 Actor-Critic 계열 등 최신 강화학습 모델을 활용하여 제어 성능을 향상시키는 시스템을 구현할 계획이다.

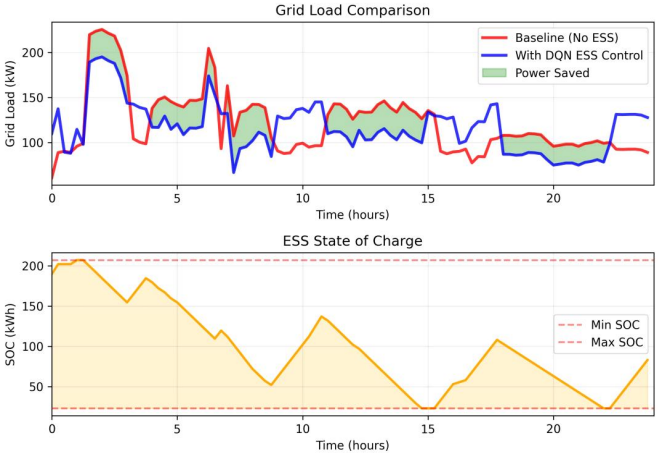


그림 3 Impact of DQN-Based ESS Control

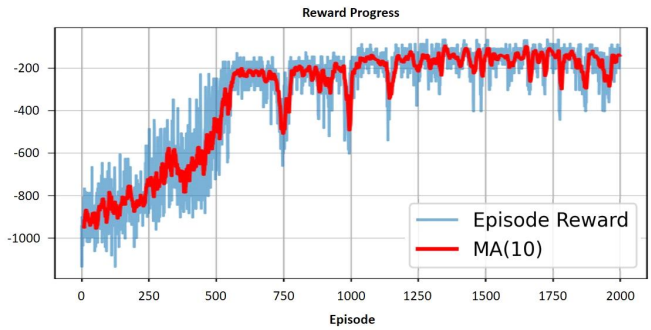


그림 4 Reward Progress on Training

ACKNOWLEDGMENT

본 연구는 산업통상자원부(MOTIE)와 한국에너지기술평가원(KETEP)의 지원을 받아 수행한 연구 과제입니다. (No. 202202090010A)

참 고 문 헌

[1] X. Fang, S. Misra, G. Xue, and D. Yang, "Smart Grid—The New and Improved Power Grid: A Survey," *IEEE Commun. Surveys Tuts.*, vol. 14, no. 4, pp. 944 - 980, 2012.

[2] K. P. Schneider, B. A. Mather, B. C. Pal, C. H. Warren, D. P. Chassin, et al., "Evaluation of Conservation Voltage Reduction (CVR) on a National Level," *Pacific Northwest National Laboratory (PNNL) Technical Report PNNL-19596*, July 2010.

[3] H. Renmu, M. Jin, and D. J. Hill, "Composite Load Modeling via Measurement Approach," *IEEE Trans. Power Syst.*, vol. 21, no. 2, pp. 663 - 676, May 2006.

[4] Y. Zhang, X. Wang, J. Wang, and Y. Zhang, "Deep reinforcement learning based Volt-VAR optimization in smart distribution systems," *IEEE Trans. Smart Grid*, vol. 12, no. 1, pp. 361 - 371, Jan. 2021.

[5] V. Mnih et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, pp. 529 - 533, Feb. 2015.

[6] H. van Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with Double Q Learning," in *Proc. AAAI Conf. Artif. Intell.*, vol. 30, no. 1, pp. 2094 - 2100, 2016.

Two-Stage 머신러닝 기반 제조업 불량 핵심인자 추출 정합성 향상 기법

이호진 [1], 정의홍 [1], 조창권 [1], 조수정 [2], 안성수 [1*]
[1]POSCO 품질기술부, [2]POSCO DX 전략실

hoho-z@posco.com, je1436@posco.com, birang99@posco.com, soojung@posco.com, virtuor@posco.com

Two-Stage Stratified Machine-Learning Approach for Identifying Root-Cause Factors In Manufacturing Processes

Hojin Lee[1], Uihong Jeong[1], Changkwon Cho[1], Soojung Cho[2], Seongsoo Ahn[1*]
[1]Department of Quality System, [2]DX Strategy Office, POSCO, Republic of Korea

요 약

본 논문은 단일 단계 접근 시 층별 요인의 특성이 변수 중요도를 왜곡하는 층별 편향 문제를 다룬다. 이를 해결하기 위해, 층별화 요인(S)으로 대표 층을 선정하고 해당 층에서 조업 변수(O)만으로 학습하는 Stratified Two-Stage ML 을 제안한다. 11 개 불량 사례 실험에서 제안법은 Hit@5=1.00 을 기록하여 기존 방법(RF=0.30, Drop Column ΔF1=0.10) 대비 우수하였다. 본 연구는 층별 편향 제거와 해석 가능성 강화에 기여한다.

Ⅰ. 서 론

Industry 4.0 이후 제조 현장에는 대규모 센서 데이터가 축적되었고, 머신러닝은 품질 검사와 공정 모니터링에 폭넓게 도입되고 있다[1]. 그러나 전체 시스템의 AI 전환은 비용과 공정 연계성 문제로 한계가 있어, 국소적 원인 규명(root cause)으로 사용하여 기존 시스템과 도메인 지식을 활용한 개선이 주로 이루어지고 있다.

기존 연구에서는 변수 중요도(feature importance) 기법이 원인 규명에 적용되었으며[2], Rotter et al.[3]은 전통적 중요도와 딥러닝 기반 방법을 실제 데이터에서 비교하였다. 다만 해당 연구는 상대적으로 단순한 공정, 제품군을 대상으로 했기 때문에 층별화 요인에 따른 편향 문제가 크게 드러나지 않았다. 반면 실제 제조 현장은 제품과 공정이 다양하여, 단일 단계 접근 시 층별 편향이 불가피하게 발생한다.

따라서 복잡한 제조 데이터에서 층별 편향을 제거하고 핵심 조업 인자를 안정적으로 규명할 수 있는 새로운 접근이 필요하다.

Ⅱ. 본론

Ⅱ-1. 데이터 및 설정

본 연구는 2022-2025 년 제조 현장의 불량 사례 중 당사 엔지니어가 핵심 영향인자를 확인한 11 건을 정답셋으로 사용하였다. 모든 사례는 동일 입력 데이터셋으로 학습되었으며, 전처리 과정은 (1) 수치형 결측치는 0 으로 대치, (2) 범주형 결측치는 “Missing”으로 처리 후 원-핫 인코딩, (3) 식별자 열 제거로 통일하였다. 데이터 전반에 클래스 불균형이 존재하나, 본 연구는 방법 간 상대 성능 비교를 목적으로 하므로 별도의 보정은 수행하지 않았다.

Ⅱ-2. 제안 방법 : Stratified Two-Stage ML

제조 데이터는 층별화 요인(S)에 따라 조업 변수(O)의 효과가 달라진다. 그러나 단일 단계 학습은 이를 고려하지 않으므로 조업 변수의 효과가 희석되거나 층별 요인 자체가 원인처럼 높은 중요도로 산출되는 층별 편향이 발생한다.

변수 중요도의 이상적 정의는 아래와 같이, 각 층 조건부 효과의 가중 평균이다.

$$I^*(o) = \sum_{s \in S} P(S = s) I(o|S = s),\ o \in O$$

하지만 단일 단계 학습은 조업 변수와 층의 특성 효과의 상관으로 편향 항 B(o)이 발생하여 핵심인자가 희석되거나 층별 특성 요인이 원인처럼 과대 평가된다.

$$I^{단일단계}(o) = I^*(o) + B(o)$$

본 연구에서는 이를 해소하기 위하여 대표 층을 먼저 선정하고, 해당 층에서 조업 변수만을 평가하는 Stratified Two-Stage ML 을 제안한다.

Stage1(층 선정), 층별화 요인(S)만으로 Random Forest 및 Rulefit 으로 학습하여 후보층을 생성하고 WRAcc 로 평가 및 대표 층(z^*)을 선택한다.

$$WRAcc(z) = p(z)(p(y = 1|z) - p(y = 1))$$

Stage2(조건부 중요도 산출) : 선택된 층(z^*)에서 조업 변수(O)만을 입력하여 Random Forest 를 학습하고 변수 중요도를 산출한다. 이를 통해 층 조건부 효과를 반영할 수 있다.

Ⅱ-3. 비교 및 평가 방법

아래와 같이 총 세 가지 방법을 비교하고 평가하였다.

1) RF : Random Forest 모델 학습 및 Gini 중요도 산출
2) DC : Rotter et al.[3] Drop Column 방식으로 기준 F1 대비, 변수 제거 후 5-fold 재학습 성능 감소량을 중요도로 정의
3) 제안법 : Stage1 에서 층을 선정한 뒤 해당 층에서만 Random Forest Gini 중요도 산출하였다.

각 사례에 대해서 당사 엔지니어 정답인자의 순위(1=최상위)를 기록하였다. 평가지표는 사례별 순위와 Hit@5 이며, 동점 발생 시 평균 순위를 기록하였다.

Ⅱ-4. 실험결과 및 분석

공정	불량유형	영향인자 순위(r_c)		
		(1)RF	(2)DC	(3)TS
A 공정	재질	3	79	1
	형상	14	63	1
	표면	6	37	1
	표면	111	49	2
B 공정	재질	125	51	4
C 공정	표면	81	37	3
D 공정	재질	4	28	1
	재질	31	24	3
E 공정	표면	5	67	2
	표면	7	5	1
F 공정	재질	24	15	1
Hit@5		0.3	0.1	1.0

[표 1. 세 방법에 따른 사례별 정답인자 순위 및 Hit@5]

표 1 은 세 방법에 대한 비교 결과를 요약하였다. 제안 방법은 모든 사례에서 정답 인자를 Top-5 내에 포함시켜 Hit@5=1.00 을 달성하였다. 반면 RF(Gini)는 0.30, Drop-Column(ΔF1)은 0.10 에 그쳤다.

제안 방법은 층별 특성 효과를 제거하여 핵심 원인인자를 안정적으로 상위에 배치되었으며, RF(Gini)는 층 특성 효과가 상위에 배치되는 현상을 보였다. Drop Column 방법의 경우에는 다중공선성이 심한 제조 데이터 환경에서 단일 변수 제거 효과가 미미해 과소추정 되었다.

Ⅲ. 결론

본 연구는 층별 특성 요인으로 인한 변수 중요도 왜곡 문제를 해결하기 위해 Stratified Two-Stage ML 을 제안하였다. 11 개 실제 불량 사례 실험에서 제안법은 모든 사례에서 Hit@5=1.00 을 기록하여 기존 방법(RF, Drop-Column) 대비 현저히 우수함을 보였다.

이 결과는 본 연구가 단순한 알고리즘 개선을 넘어, 실제 제조 현장에서 현장 엔지니어가 핵심 인자를 신뢰성 있게 확인할 수 있는 절차를 제공한다는 점에서 의의가 크다. 특히 규칙 기반 층 선정과 조건부 중요도 산출의 조합은 설명 가능성을 높여 즉시 개선 지침으로 활용될 수 있다.

다만 본 연구는 단일 대표 층만 고려했으며, 다수의 층을 동시 평가 다양한 ML 기법 비교 및 실제 라인 적용 검증은 추후 과제로 남아 있다.

향후 연구에서는 다층적 확장과 최신 기법과의 비교를 통해 일반화 가능성을 높이고, 실시간 분석 환경에 접목함으로써 지속적인 품질 개선 플랫폼으로 발전시킬 예정이다.

ACKNOWLEDGE

본 연구는 당사 DX 전략실의 지원으로 수행되었으며, 특히 IT 인프라 구축과 시스템 협조에 도움을 주신 관계자분들께 감사드립니다.

참 고 문 헌

[1] Bonomi, N., et al.: Smart quality control powered by machine learning algorithms. In: 2021 IEEE CASE, pp. 764–770 (2021)

[2] Carletti, M., Masiero, C., Beghi, A., Susto, G.A.: Explainable machine learning in industry 4.0: evaluating feature importance in anomaly detection to enable root cause analysis. In: 2019 IEEE SMC, pp. 21–26 (2019)

[3] Rotter, D., Liebgott, F., Kessler, D., Liebgott, A., & Yang, B. (2023). Machine Learning-Based Identification of Root Causes for Defective Units in Manufacturing Processes. In Advances in Automotive Production Technology – Towards Software-Defined Manufacturing and Resilient Supply Chains (SCAP 2022) (pp. 245–258). Springer.

[4] Breiman, L., Friedman, J. H., Olshen, R. A., Stone, C. J.: Classification and Regression Trees. Wadsworth, Belmont (1984)

[5] Montgomery, D. C.: Introduction to Statistical Quality Control, 8th ed. Wiley, Hoboken (2019)

[6] Juran, J. M., De Feo, J. A.: Juran's Quality Handbook: The Complete Guide to Performance Excellence, 7th ed. McGraw-Hill, New York (2017)

다중 위성 영상 데이터를 활용한 연합 학습 기반의 객체 탐지 연구

김성철, 김성용, 조규태
LIG 넥스원

sungcheol.kim@lignex1.com, sungyong.kim@lignex1.com, kyutae.cho2@lignex1.com

Federated Learning for Fine-Grained Vehicle Detection Across Multi-Platform Satellite Imagery

Kim Sung Cheol, Kim Sung Yong, Cho Kyu Tae
LIG NEX1

요 약

위성 영상 기반 객체 탐지 기술은 국방 및 다양한 산업 분야에서 핵심적인 역할을 한다. 그러나 플랫폼별 데이터 수집의 제약과 데이터 분산 문제로 인해 성능 향상에 어려움이 있다. 본 연구는 항공, 지상, 해상 플랫폼에서 수집된 데이터를 통합한 새로운 다중 플랫폼 데이터 셋을 제안하고, 폐쇄망 환경에서도 적용 가능한 연합 학습 기반 객체 탐지 프레임워크를 설계하였다. 실험 결과, 폐쇄망 환경에서의 학습 안정성을 확인할 수 있었고 향후 위성 체계 분야에 도움이 될 것이라 기대한다.

Ⅰ. 서 론

최근 몇 년간 위성 체계 및 무인 시스템을 활용한 다양한 산업 분야에서 객체 탐지 기술 연구가 활발히 진행되고 있다. 객체 탐지는 이미지 혹은 비디오 내에서 특정 물체를 탐지하고 분류하는 기술로, 경계 감시, 표적 식별, 전장 상황 인식 등 다양한 군사 작전에서 핵심적인 역할을 한다. 대표적인 딥러닝 기반 탐지 모델로 YOLO 가 있다.

위성 영상 분석은 항공, 지상, 해상 등 다양한 플랫폼에서 환경에서 적용되고 있다. 그러나 현실적으로 각 플랫폼에서 수집되는 데이터의 수는 제한적이다. 또한 위성 영상 데이터는 여러 기관 및 국가에 분산되어 저장되는 경우가 많아 데이터 접근성과 통합에 어려움이 존재한다. 이에 따라 연합 방식 기반의 연구가 제안되었다. 연합 학습이란, 원본 데이터를 직접 공유하지 않고 모델을 학습하는 방식으로 데이터 프라이버시를 보호할 수 있는 방법이다.

본 연구는 항공, 지상, 해상 플랫폼으로부터 수집된 다양한 센서 데이터를 활용하여 새로운 다중 플랫폼 데이터 셋을 제안한다. 또한 폐쇄망 환경에서도 적용 가능한 연합 학습 기반 객체 탐지 프레임워크를 설계하였다. 제안한 프레임워크는 CNN 및 Transformer 기반 탐지 모델을 포함하여 비교 실험을 수행하였다. 실험 결과, 제안된 연합 학습 모델이 개별 로컬 학습 모델에 비해 효과적임을 확인하였다.

Ⅱ. 본론

2.1 데이터 구성

본 연구는 기존의 공개 데이터 셋(Mar20 [1], ShipRSImageNet [2], DroneVehicle [3])을 통합하여 새로운 다중 플랫폼 객체 탐지 데이터 셋인 MetaFleet 을 구축하였다. 이때, 연합 학습 시나리오를 고려하여, 클래스 별 인스턴스 수가 700 개 미만인 이미지는 제외하여 데이터를 재구성하였다. MetaFleet 은 항공(aircraft), 해상(ship), 지상(ground) 세 가지 플랫폼을 포함하여, 총 20 개의 세분화된(fine-grained) 클래스로 구성되어 있다. 본 연구에서 제안한 데이터 셋은 Table 1.과 같다.

Datasets	Platform	Images (Train / Val)	Instances (Train / Val)
Mar20 [1]	Aircraft	2,689/ 769	15,577 / 4,527
ShipRSImageNet [2]	Ship	2,200/ 545	10,860 / 3,103
DroneVehicle [3]	Ground	17,948 / 1,469	297,716 / 22,462
MetaFleet	Aircraft, Ship, Ground	22,835 / 2,788	302,898 / 28,478

Table 1. 제안 데이터 셋 (MetaFleet)

제안 데이터 셋은 단일 플랫폼(항공, 해상, 지상) 데이터를 통합하여 학습 데이터의 도메인 다양성을 확보하였다. 또한 세분화된 클래스 데이터를 활용함으로써, 기존의 대분류(coarse-grained) 데이터 셋에 비해 더 정밀한 성능 분석이 가능하도록 구성하였다.

2.2 연합 학습 프레임워크

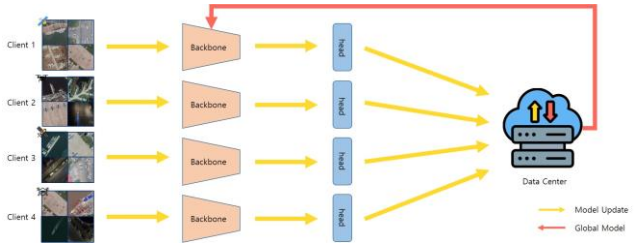


Figure 1. 연합 학습 프레임워크

Figure 1.은 폐쇄망 환경에서의 연합 학습 기반의 객체 탐지 프레임워크를 나타낸다. 제안한 데이터 셋을 네 명의 클라이언트에게 분배하였으며, 각 클라이언트는 항공, 지상, 해상 플랫폼에서 수집된 데이터를 활용하여 로컬 모델을 독립적으로 학습한다. 이때, 데이터는 비독립적이며 동일 분포를 따르지 않는 Non-IID 특성을 가진다. 본 연구는 CNN 기반의 YOLOv5, YOLOv8, YOLOv11 및 Transformer 기반의 D-fine 모델을 활용하여 학습을 수행하였다. 학습 후, 각 클라이언트는 로컬 모델의 업데이트를 데이터 센터(Data Center)로 전송하며, 데이터 센터에서는 FedAvg 알고리즘을 통해 글로벌 모델의 가중치를 업데이트한다. 이후 업데이트 된 글로벌 모델의 가중치는 각 클라이언트의 로컬 모델로 전달된다. 본 프레임워크는 폐쇄망 환경에서 설계되었으며, 모든 과정은 내부 네트워크에서 수행하였다.

Ⅲ. 실험

3.1 실험 세팅

본 연구에서는 연합 학습 환경을 구성하여 총 4 개의 클라이언트(client)를 설정하였으며, 각 클라이언트가 보유한 데이터는 Non-IID 분포를 따른다. 모든 클라이언트가 매 통신 라운드에 참여하도록 참여 비율을 1.0 으로 설정하였다. 각 클라이언트가 소유한 로컬 모델의 학습 에폭 수는 10, 20, 40 으로 설정하였다. 파라미터 집계 방식은 FedAvg(Federated Averaging) 알고리즘을 적용하였다.

객체 탐지 실험 세팅의 경우, YOLOv5, YOLOv8, YOLOv11, D-fine 아키텍처를 활용하였다. 이미지 크기는 1024x1024 로 고정하였으며, optimizer 는 Adamw 를 사용하였다. 학습률을 3e-4 로 설정하였고, 배치 크기는 8 로 고정하였다. 모델 성능 평가는 mAP50, mAP50-95 지표를 활용하였다.

3.2 실험 결과

Datasets	Client	Model	Centralized (%)	Federated (%)			
				Method	E10	E20	E40
MetaFleet	Ⅲ	YOLOv5n	81.7	FedAvg	72.6	76.2	76.9
		YOLOv8n	83.4		73.2	78.0	78.6
		YOLOv11n	83.8		73.5	78.1	79.0
		D-fine-n	85.9		83.5	83.9	84.7

Table 2. Local Epoch(E)에 따른 객체 탐지 성능 비교

Table 2.는 연합 학습 환경에서 로컬 에폭(Local Epoch, E)의 수에 따른 알고리즘 별 객체 탐지 성능 비교 결과를 나타낸다. 실험 결과, 모든 모델에서 로컬 에폭 수의 증가에 따라 연합 학습 성능이 점진적으로 향상되었으며, 특히 학습 초기 단계에서 나타나는 성능 불안정성이 완화하는 경향을 보였다. 이는 로컬 에폭 수의 증가가 클라이언트 내에서 반복적인 모델 업데이트를 수행함으로써, 각 통신 라운드 시 글로벌 모델이 보다 안정적이고 일관된 방향으로 수렴하도록 돕는다는 점을 시사한다.

Training	mAP50 / mAP50-95 (%)			
Model	YOLOv5n	YOLOv8n	YOLOv11n	D-fine-n
Centralized	81.7 / 61.6	83.4 / 63.7	83.8 / 64.1	85.9 / 67.1
Client 1	75.8 / 55.4	76.9 / 56.4	78.1 / 57.6	78.9 / 59.3
Client 2	75.7 / 55.2	76.8 / 56.7	77.5 / 57.1	78.6 / 59.5
Client 3	75.1 / 54.6	77.6 / 57.2	77.9 / 57.4	78.8 / 59.2
Client 4	74.9 / 54.8	76.8 / 56.6	76.8 / 56.8	77.7 / 58.6
Client Avg	75.4 / 55.0	77.1 / 56.8	77.6 / 57.3	78.5 / 59.2
FedAvg	76.9 / 56.1	78.6 / 58.2	79.0 / 68.8	84.7 / 65.3

Table 3. 로컬 클라이언트 및 연합 학습 기반의 객체 탐지 따른 객체 탐지 성능 비교

Table 3.은 로컬 클라이언트에서 개별적으로 학습한 모델(Local)과 클라이언트 간 파라미터만을 공유하는 연합 학습(FedAvg)를 비교 결과를 의미한다. 실험 결과, 연합 학습 방식으로 학습된 모델이 로컬 학습 대비 효과적임을 확인하였다. 특히 D-fine-n 의 경우, 로컬 모델과 6.2%의 성능 차를 보였다.

Ⅳ. 결론

본 연구에서는 항공, 해상, 지상 플랫폼에서 수집된 데이터를 통합하여 세분화된 데이터셋을 제안하고, 폐쇄망 환경에서의 연합 학습 기반 객체 탐지 프레임워크를 설계하였다. 실험 결과 데이터의 수가 적은 경우에는 로컬 에폭의 수를 길게 설정하여 로컬 모델의 가중치가 수렴하도록 하는 방식이 연합 학습에 유의미한 결과를 가져오는 것으로 확인하였다. 또한 연합 학습 방식으로 학습된 모델이 로컬 학습 대비 효과적임을 실험적으로 확인하였다. 향후 폐쇄망 환경에서도 이러한 접근 방식이 위성 영상 기반 객체 탐지 기술 발전에 기여할 수 있을 것으로 기대한다.

참 고 문 헌

[1] WENQI, Y. U., et al. MAR20: A benchmark for military aircraft recognition in remote sensing images. National Remote Sensing Bulletin, 2024, 27.12: 2688-2696.

[2] ZHANG, Zhengning, et al. ShipRSImageNet: A large-scale fine-grained dataset for ship detection in high-resolution optical remote sensing images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14: 8458-8472.

[3] SUN, Yiming, et al. Drone-based RGB-infrared cross-modality vehicle detection via uncertainty-aware learning. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32.10: 6700-6713.

CFD 데이터셋과 적대적 공격 기법을 활용한 차세대 스텔스 형상 설계 방법에 관한 연구

박지성, 최창우, 김호원*

㈜스마트엠투엠, ㈜스마트엠투엠, *㈜스마트엠투엠
부산대학교, 부산대학교, *부산대학교

parkjiseung@smartm2m.co.kr, changwoo@smartm2m.co.kr, *howonkim@gmail.com

A Study on a Next-Generation Stealth Geometry Design Using a CFD Dataset and Adversarial Attack Techniques

Park Ji Seong, Choi Chang Woo, Kim Ho Won*

SmartM2M Inc., SmartM2M Inc., *SmartM2M Inc.
Pusan Univ., Pusan Univ., *Pusan Univ.

요 약

본 논문은 3차원 포인트 클라우드(3D Point Cloud, 이하 3D 점군) 기반 딥러닝 모델과 전산 유체 역학(CFD) 데이터셋을 활용하여, AI 기반 탐지를 회피하면서도 우수한 공력(공기역학적 물리력, 항력과 양력) 성능을 갖춘 차세대 스텔스 형상 설계 전략을 제시한다. 이를 위해 ModelNet40 데이터셋으로 학습한 PointNet 분류 모델과, DrivAerNet++ 데이터셋으로 학습한 PointNet 회귀 모델(양력 C_l 및 항력 C_d 예측)을 구성하였다. 이후 세 가지 적대적 공격 기법 적용 방식을 통해 차량 형상의 점군 데이터를 최적화하였다: (1) 분류 모델의 탐지를 회피하는 공격, (2) 회귀 모델의 항력/양력 출력을 최소화하는 공격, (3) 두 목표를 동시에 달성하는 다중 목표 최적화 공격. 실험 결과, 제안한 방법으로 생성된 형상은 분류 모델의 탐지를 성공적으로 회피하면서도 항력/양력 계수를 효과적으로 감소시키는 것을 확인하였다. 이러한 AI-기반 탐지 회피와 공력 성능 향상의 동시 달성은 향후 차세대 스텔스 형상 설계 방법에 새로운 방향을 제시할 것으로 기대된다.

Ⅰ. 서 론

스텔스 형상 설계는 자율주행, 군수, 항공우주 분야 전반에서 탐지 회피 성능을 향상시키기 위한 핵심 과제이다. 기존에는 주로 레이저 반사 단면적(RCS)이나 시각적 노출 최소화에 집중했지만, 최근에는 AI 기반 객체 탐지 기술의 발전으로 단순한 형상 정보만으로도 탐지가 가능한 환경이 조성되었다. 예를 들어 3D 라이다/점군 데이터를 입력받는 PointNet[2]과 같은 딥러닝 모델은 차량, 보행자 등의 물체를 높은 정확도로 탐지할 수 있다. 한편 스텔스 형상 설계에서는 탐지 회피뿐만 아니라 공기역학적 성능도 중요한 고려 요소이다. 항력을 줄이는 형상은 단순히 연비와 주행 성능 향상뿐 아니라, 군수나 항공기 영역에서도 효율성과 기동성을 높이는 핵심 설계 요건이 된다. 따라서 탐지 회피와 공력 최적화를 동시에 고려한 다중목표 설계가 요구된다. 본 연구는 AI 탐지 회피와 공력 최적화라는 두 가지 문제를 통합적으로 다루어, 다중목표 스텔스 형상 설계의 가능성을 탐색한다. 딥러닝 기반 분류/회귀 모델을 활용하여 형상을 서로 다른 목표에 따라 최적화하면, 사람에게는 미세한 형상으로 인식되는 변화로도 AI의 인식을 교란하면서 항력까지 감소시키는 설계안을 얻을 수 있을 것으로 가정했다. 이러한 접근은 적대적 공격(Adversarial Attack) 개념에 기반하며, 기존에 2D 이미지나 3D 점군에서 모델 혼동을 유발하기 위해 연구된 적이 있다.[3][4][5] 특히 3D 점군 분야에서 딥러닝 모델이 적대적 변형에 취약함이 보고되었고, 거의 보이지 않는 미세한 점 이동으로도 분류를 오인시키는 공격이 가능하다. 본 연구는 이 개념을 역으로 활용하여, AI를 속이는 스텔스 형상을 설계함과 동시에 공력 성능을 향상시키는 새로운 방법론을 제안한다. 구체적으로, ModelNet40 CAD 모델 데이터셋으로 3D 물체 분류용 PointNet 모델을 학습하고, DrivAerNet++ 대규모 차량 공력 데이터셋으로부터 항력·양력 예측 모델을 학습하였다.[1] 이후 학습된 모델의 가중치를 활용하여 세 가지 시나리오별로 점군 형상 최적화를 수행하였다. 본 연구에서는 각 시나리오(분류 회피, 공력 최적화, 다중목표)에 대한 실험을 통해 제안 기법의 효과를 입증하고, 차세대 스텔스 형상 설계 전략으로서의 가능성을 논의한다.

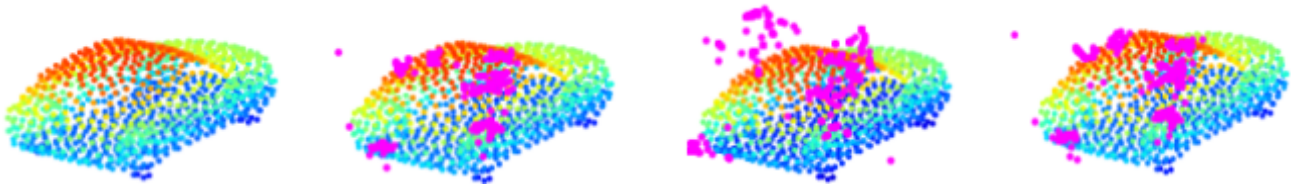
Ⅱ. 본론

본 논문에서는 학습된 분류/회귀 모델을 기반으로 손실함수를 설계하고, 세 가지 시나리오(분류 회피, 공력 최적화, 다중목표)에 따라 3D 점군 형상을 미세 조정하여 AI 탐지 회피와 공력 성능 향상을 동시에 달성할 수 있는지를 검증한다. 절차는 모델 및 데이터셋 구성, 공격 손실 설계, 시나리오별 최적화, 공격 성능 평가의 순서로 진행된다.

1) 모델 및 데이터셋

3D 점군 분류에는 PointNet 구조를 사용하였다. PointNet은 점집합을 직접 입력으로 받아 순서에 불변인(Point permutation invariant) 특성을 이용함으로써, ModelNet40 등 벤치마크에서 뛰어난 성능을 보인 바 있다. 본 연구에서는 ModelNet40 데이터셋으로 PointNet을 학습하여 3D 점군 데이터 기반 차량 형상에 대한 분류 능력을 확보할 수 있도록 하였다.

공력 계수 예측에는 PointNet 구조를 회귀용으로 변형한 PointNetReg 모델을 구성하였다. 출력 노드는 항력 계수 C_d와 양력 계수 C_l의 2차원



(a) 원본 (b) 분류 회피 최적화 (c) 공력 성능 최적화 (d) 다중목표 최적화 (분류+공력)

그림 1 적대적 공격 기법 기반 최적화 목표별 3D 차량 형상 변화 시각화

표 1 최적화 목표에 따른 PointNet 분류/회귀 모델 예측 결과

최적화 목표	자동차 분류 확률 (%)	항력 계수 C_d	예측 양력 C_l
(a) 원본	98.7	0.263	-0.036
(b) 분류 회피 최적화	**8.0**	0.258	-0.033
(c) 공력 성능 최적화	20.8	**0.237**	**-0.049**
(d) 다중목표 최적화 (분류+공력)	10.0	0.254	-0.034

록 연속값으로 하여, 입력 3D 점군 형상의 공기역학 계수를 직접 예측하도록 하였다. 학습 데이터는 다양한 차체 형태와 언더바디, 휠 구성 조합으로 생성된 8,000개 차량 3D 모델의 고정밀 CFD 시뮬레이션 결과가 담겨 있는 DrivAerNet++ 데이터셋을 활용하였다. 본 연구에서는 해당 데이터셋을 기반으로 회귀모델인 PointNetReg를 학습하였다. 학습된 회귀 모델은 복잡한 차량 형상에 대해서도 항력/양력 계수를 합리적인 정확도로 추정해주며, 설계 공간 전반에 대한 Surrogate 모델 역할을 수행한다.

2) 적대적 공격 기법

본 연구에서는 학습된 3D 점군 분류 모델과 회귀 모델을 기반으로, 차량 형상의 3D 점군(1,024개)의 일부를 단계적으로 변형하여 분류 결과와 공력 성능을 의도한 방향으로 최적화하는 적대적 공격 기법 기반 3D 점군 형상 최적화 방법을 설계하였다. 공격은 White-box 환경에서 수행되며, 모델 출력에 대한 점의 기울기를 계산해 좌표를 미세 조정하는 방식이다. 적대적 공격을 위한 손실 함수는 크게 분류 회피 손실, 회귀 최적화 손실, 형상 보존 손실 세 부분으로 구성된다. 분류 회피 손실은 분류 모델이 차량 클래스로 인식하지 않도록 확률을 낮추고, 출력 분포의 엔트로피를 높여 특정 클래스 편향을 줄이도록 설계하였다. 회귀 최적화 손실은 회귀 모델이 예측하는 항력(C_d)과 양력(C_l)을 최소화하여 공력 성능을 개선하도록 설계하였다. 형상 보존 손실은 최적화 과정에서 물리적으로 타당한 형상을 유지하도록 이방성 부착(soft-attachment), 라플라시안 평활화, 커버리지 보조 손실 등이 적용되도록 설계하였다. 최종 손실은 세 손실의 가중합으로 구성된다. 그림 1은 최적화 목표에 따른 3D 점군 변화를 나타낸다. (a)는 원본 형상, (b)는 분류 회피 최적화 결과, (c)는 항력·양력 최소화를 위한 회귀 최적화 결과, (d)는 두 목표를 함께 적용한 다중목표 최적화 결과로, 각 방법에 따른 형상 변화와 성능 차이를 시각적으로 확인할 수 있다.

3) 실험 및 결과

그림 1과 표 1은 최적화 목표별 3D 점군 형상 변화와 PointNet 예측 결과를 보여준다. 분류 회피 최적화는 분류 정확도를 92.0% 감소시켰다. 공력 성능 최적화는 항력 0.021, 양력 0.013 감소로 공기역학적 성능을 향상시켰고, 3D 점군 변화로 인해 분류 정확도도 79.2% 감소했다. 다중목표 최적화는 분류 정확도 90.0%의 감소와 함께 항력 0.009 감소를 달성하였다. 다중목표 최적화는 두 목표 간 균형점을 찾아, 외형과 성능 모두에서 절충된 결과를 가져왔다.

Ⅲ. 결론

본 연구는 CFD 데이터와 적대적 공격 기법을 결합하여, AI 기반 탐지를 회피하면서도 공력 성능을 개선하는 차세대 스텔스 형상 설계 방법을 제안하였다. PointNet 기반 3D 분류모델과 CFD 기반 회귀모델을 활용하여 형상을 최적화함으로써 AI 탐지 회피성과 공기 역학 성능을 동시에 향상시켰다. 제안한 방법은 기존 레이더 회피 중심의 스텔스 설계와 달리, AI 탐지 회피와 공기역학 최적화를 하나의 프레임워크에서 동시 달성할 수 있도록 설계되었다. 이를 통해 상충 관계에 있는 두 목표 간의 균형을 정량적으로 조율할 수 있는 새로운 접근 방식을 마련하였다. 다만 CFD 시뮬레이터의 Surrogate 모델의 예측 오차, 라이다 기반 실험 부재, 적대적 방어 기법 모델에 대한 검증 부족 등이 한계로 남는다. 향후 연구에서 CFD-in-the-loop 검증, 제조 제약을 고려한 다목적 최적화, 다양한 센서 위협모델 기반 평가, 방어 기법 대응 실험을 통해 실세계 적용성을 높인다면, 차세대 스텔스 형상 설계 표준 절차에 기여할 수 있을 것으로 기대한다.

ACKNOWLEDGMENT

이 논문은 2025년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No.2022-0-00431, (3세부) 지능형 디지털 트윈 연합 지원 개방형 서비스 플랫폼 및 저작도구 기술 개발)

참 고 문 헌

[1] Elrefaie, M., et al. “DrivAerNet++: A Large-Scale Multimodal Car Dataset with CFD Simulations and Deep Learning Benchmarks”, Proc. NeurIPS, 2024.

[2] Qi, C. R., Su, H., Mo, K., and Guibas, L. J., “PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation”, Proc. CVPR, pp. 652-660, 2017.

[3] Xiang, C., Qi, C. R., and Li, B., “Generating 3D Adversarial Point Clouds”, Proc. CVPR, pp. 9136-9144, 2019.

[4] Zhang, X., et al., “Evaluating the Robustness of LiDAR Point Cloud Tracking Against Adversarial Attack”, arXiv preprint, 2024.

[5] Bo Yang, et al. “Generating 3D Adversarial Point Clouds under the Principle of LiDARs”, NDSS AutoSec Workshop, 2022.

감성 극성 분류 모듈과 다중 에이전트 구조를 활용한 개선된 다중 문서 요약 기술 연구

조병웅, 봉한성, 홍주화*
성균관대학교

Whale1510@g.skku.edu, peterpen0110@g.skku.edu, j.hong@skku.edu*

multi-document summarization pipeline utilizing a sentiment polarity classification module and a multi-agent architecture.

Cho Byeong Ung, Han Sung Bong, Joo Wha Hong*
SungKyunKwan Univ.

요 약

하나의 문서만을 요약하는 단일 문서 요약과 달리, 다중 문서 요약은 상품 리뷰처럼 동일한 주제에 대한 수천 개의 문서 집합을 요약하는 과제이다. 이러한 다중 문서 요약은 요약 대상의 방대한 수와, 정보의 중복과 충돌 등의 요인으로 인해 단일 문서 요약 방법론의 확장으로는 해결할 수 없다. 그 때문에 ML 기반의 방법론을 적용하여 주요 정보를 추출하는 추출 요약과, 사전학습 Transformer 모델이나 LLM 을 활용한 추상 요약을 결합하여 다중 문서 요약을 수행하는 하이브리드 방법론이 등장하였다. 하지만 그러한 방법론은 추출 요약에서 발생하는 정보 누락 문제와 추상 요약에서 발생하는 환각 문제를 그대로 상속받으며, 리뷰에서 발생하는 긍정, 부정 의견 두 가지 극성을 제대로 반영하지 못한다. 이에 본 논문에서는 기존 방법론 구조에, 감성 극성 분류 모듈과 QA 기반의 feedback 을 수행하는 다중 에이전트 구조를 추가한 확장된 로직을 제안한다. 해당 로직을 적용했을 때, 기존 방법론보다 더 높은 성능을 달성함을 확인하였다.

Ⅰ. 서 론

온라인 상거래의 급성장과 함께 리뷰는 가격이나 상품의 매력에 버금가는 핵심 구매 결정 요소로 부상했으나, 방대한 리뷰를 소비자가 직접 모두 읽는 것은 현실적으로 불가능하다. 이에 따라 다수의 리뷰를 하나로 요약하는 Multi-Document Summarization(MDS) 기술이 주목되어 연구되어 왔다.

최근에는 Transformer 기반 모델이나 대형 언어 모델(LLM)을 활용한 추상 요약 방법론과 기존 ML 기반의 추출 요약 방식을 결합한 하이브리드 접근법[1][2]이 단순 LLM 기반 방법론에서의 토큰 제한 문제를 극복하고 리뷰 요약에 활용되고 있다. 그러나 이 방식은 여전히 LLM 의 환각, ML 의 정보 누락 문제를 상속하며 사용자 선호(양극성)를 반영하지 못하고 평균적인 요약문을 생성한다는 한계가 있다.

본 연구에서는 기존 하이브리드 방법론에서 (1) 긍/부정 리뷰를 분리하여 요약할 수 있도록 감성 분류 모듈을 추가하고, (2) Reflection 기법을 본 MDS 과제에 적용하고자 반복적 QA 기반 feedback 을 수행하는 논리적으로 분리된 다중 에이전트 구조를 추가하여 확장하는 기법을 제안한다.

Ⅱ. 관련 연구

Multi-Document Summarization

최근 LLM 의 등장은 자연어 처리 분야에서 큰 혁신을 가져왔고, 요약 분야에서도 기존 방법론들보다 LLM 을 활용한 요약이 기존 방법론들보다 더 뛰어나다는 것이 증명되었다.[3] 그러나 프롬프트에 원문을 탑재해야 하는 LLM 기반의 요약 방법론은 토큰 제한 문제로 인하여 수백~수천 개의 리뷰를 요약해야 하는 MDS 에서 적용하기 어렵다는 문제를 가지고 있다.

이를 극복하기 위해서 ML 기반의 추출 요약과 LLM 기반의 추상 요약을 결합하여 토큰 제약을 해결하려는 방법론[1][2]이 제시되었다. 그러나 이러한 방법론은 사용자 선호도를 반영하지 못하며 환각 및 정보 누락이 발생한다는 문제가 여전히 존재한다. 실제 MDS 의 환각 문제를 추적한 논문에서는 평균적으로 45%~75% 정도의 환각이 발생한다고 보고했다.[4]

Reasoning

LLM 의 환각 문제를 해결하기 위해 추론 능력을 활용하고자 하는 여러 연구가 진행되었다. LLM 추론을 활용하여 복잡한 문제를 해결하고자 하는 Reasoning 분야가 개척되었고, 이러한 발전 속에서 자가 피드백을 통해 출력을 개선하는 Reflection 방법론[5]이 새롭게 등장하면서 환각을 줄일 해결책이 등장하였다.

요약 과제에서도 그러한 Reflection 기법을 도입하여 환각과 정보 누락을 해결하고자 하는 연구[6]가 제안되었지만, 단일 문서에만 적용되며 단일 구조로 인한 자기 편향 문제가 발생한다는 한계가 있었다.

Ⅲ. 제안 기법

본 연구에서는 분류, 추출, 생성, 평가 단계로 구성된 다중 문서 요약 기법을 제안한다. 제안 기법은 분류

단계를 통해 긍정, 부정 리뷰 분류를 각각 수행하게 되며, 해당 리뷰들의 긍정, 부정 비율을 함께 결과로 제공한다.

이때 생성 단계와 평가 단계에서는 각각 단일 LLM API 인스턴스를 사용하되, '생성자'와 '평가자'로 프롬프트 구조를 명확하게 구분하여 논리적으로 독립된 역할을 갖고 상호작용하는 다중 에이전트 구조를 구축하여 환각과 정보 누락을 개선하면서도 자기 편향 문제를 해결하는 구조로 설계하였다.

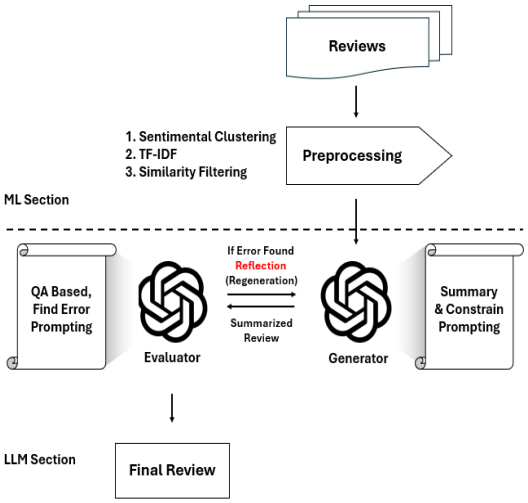


그림 1 Workflow Diagram

가. 분류 단계

분류 단계에서는 사전 훈련된 DistillBERT 를 활용한 감정 분류 모듈을 통해 리뷰 데이터셋을 긍정/부정 리뷰들로 분류하여 그룹화를 수행한다.

나. 추출 단계

추출 스테이지는 앞서 분류된 긍/부정 그룹에서 중요한 정보 추출을 수행한다. 각 그룹에서 TF-IDF 기법을 통해 주요 키워드를 추출하고, 해당 단어가 포함된 문장들을 선별하여 유사 문장 클러스터링을 진행한 후 각 중요 문장을 추출한다.

다. 생성 단계

추출 단계에서 추출한 그룹별 중요 문장들을 생성자에게 전달하여 자연스러운 요약문으로 생성하고 평가자에게 전달된다. 생성자는 GPT-4 모델을 사용하며, 제약조건이 주어진 프롬프트로 구성된다.

라. 평가 단계

평가 단계에서는 전달된 요약문을 평가자가 QA 기반 평가를 통해 환각 여부와 정보 누락 여부를 측정한다. 정밀도(Precision) 관점에서, 평가자는 요약문에서 뽑아낸 질의를 원문으로 검증하여, 요약문에 포함한 정보가 원문에서 사실적으로 근거하는지를 평가한다. 또한 재현율(Recall) 관점에서, 원문에서 생성된 질의를 요약문으로 검증하여, 원문에 포함된 필수 정보가 요약문에 누락되지는 않았는지를 평가한다. 이러한 QA 질의를 통해 MDS 의 원문과 요약문의 비교 평가를 간접적으로 수행하여 토큰 제약 문제를 해결하며, 평가자의 평가 결과에 따라 문제가 없다면 긍정, 부정 요약문을 각각 출력한다. 반면 평가자가 문제점을 확인한다면 이를 feedback 으로 제공하여 다시 생성자에게 재생성을 요청한다.

Ⅳ. 실 험

가. 데이터셋

본 연구에서는 리뷰와 관련된 다중 문서 후기 요약 데이터로 Amasum 벤치마크 데이터셋[7]과 Space 데이터셋[8]을 활용하였다. 해당 데이터셋들은 각 상품에 대해 주석자들이 작성한 정답 요약문과, 원문 리뷰 등으로 구성되어 있으며, Amasum 데이터의 정답 요약문은 Verdict(총평 요약), Pros(장점 요약), Cons(단점 요약)으로 개별화 되어있다.

나. 실험 구성

본 연구의 baseline 으로는 감정 분류와 평가 스테이지가 포함되지 않는 기존 연구의 하이브리드 방법론 구조를 차용하되, 기존 연구에서 사용한 BART 혹은 Transfomer 모델이 아닌 Openai 의 GPT-4 모델을 생성 요약 모델로 변경하였다. 이를 통해 baseline 과 제안 기법의 생성 모델을 같게 함으로써 생성단계에서 단순 LLM 의 추론 능력 차이에 따른 성능 차이를 배제하였다.

제안 기법 또한 baseline 과 마찬가지로 GPT-4 를 요약 모델로 사용하였으며, 평가 단계에서도 동일한 모델을 사용하였다. 평가 단계의 경우 최대 3 회까지 재생성이 가능하도록 제한하고 성능을 평가하였다.

다. 요약 성능 평가

공정한 비교를 위해 정답 요약을 Verdict·Pros·Cons 를 순서대로 연결한 단일 참조로 구성하였다. Baseline 의 단일 요약과, 본 연구 기법의 장점 요약+단점 요약을 연결한 요약을 각각 동일 참조에 대해 ROUGE 스코어로 평가·비교하였다. 각 데이터들은 Amasum 평가 데이터셋에서 50 개를 랜덤 샘플링하여 진행하였다.

.

	Rouge-1	Rouge-2	Rouge-L
Baseline	23.55	2.48	12.83
Ours	**27.30**	**3.17**	**13.74**

표 1 Amasum 데이터셋 요약 성능 평가 결과

Baseline 과의 비교에서, 모든 스코어에서 성능 향상이 발생하였다. 추상 요약의 특성상 빅그램(n-gram)에서의 성능이 낮게 측정되어 Rouge-2 에서는 큰 향상을 보이진 못했지만, Uni-gram 기반의 Rouge-1 에서는 두드러진 성능 향상이 나타났다. 이를 통해 기존 Baseline 보다 본 연구 기법이 더 개선된 요약을 생성함을 확인하였다.

라. 성분 분석 실험(Ablation Study)

QA 기반 피드백 평가 단계가 실제 요약 품질 향상에 기여하는지를 검증하고자, 감성 분류 단계만 제외하고 평가 단계는 유지한 변형 구조와, 평가 단계까지 모두 제거한 변형 구조 두가지를 구성하여 Space 데이터셋으로 실험을 진행하였다.

	Rogue-1	Rogue-2	Rogue-L
Non-Evaluator	32.73	7.62	20.6
Evaluator	**37.22**	**8.82**	**22.01**

표 2 평가 단계 효용성 검증 결과

실험결과, 평가 단계를 붙인 변형 구조의 결과가 더 성능이 향상되었으며, 본 기법에서 제안된 방식의 효과성을 입증하였다.

Ⅴ. 결론

본 연구에서는 기존 하이브리드 방법론의 한계점을 지적하고, 이를 해결하기 위해 긍정, 부정 그룹별 요약을 위한 감성 분류 모듈과 QA 기반의 피드백을 수행하는 다중 에이전트 구조를 활용하여 다중 문서 요약을 수행하였다. 이러한 기법을 통하여 기존 방법에서 발생하는 환각과 정보 누락 문제들을 개선할 수 있음을 실험을 통해 입증하였으며, 사용자의 선호도를 반영한 요약 결과에서도 더 성능이 향상됨을 확인하였다. 이러한 개선된 기법을 통하여, 리뷰 시스템이 적용되는 여러 분야에 개선된 리뷰 요약을 제공할 수 있을 것을 기대하며, 그동안 단일한 요약만을 제공해왔던 것을 넘어 사용자들에게 다양한 선호도를 대표하는 요약을 제공하여 리뷰 서비스에 있어 사용자 경험(UX)도 향상시킬 수 있을 것이라 기대한다. 향후 연구에서는 정보 누락과 환각을 식별·검증하기 어려운 ROUGE 스코어의 한계를 보완하기 위하여, QA 기반 다중 문서 요약 벤치마크 데이터셋 연구를 진행할 예정이다.

참 고 문 헌

[1] 이필원, 황윤영, 최종석, 신용태, "워드 임베딩 클러스터링을 활용한 리뷰 다중문서 요약 기법". 정보처리학회 논문지, 제 10 권, 제 11 호(통권 110 호), pp. 535-540, 2021 년.

[2] Bhaskar, A., Fabbri, A. R., and Durrett, G., "Prompted Opinion Summarization with GPT-3.5," in *Findings of the Association for Computational Linguistics: ACL 2023*, pp. 9282- 9300, July 2023

[3] Pu, X., Gao, M., and Wan, X., "Summarization is (Almost) Dead," arXiv preprint arXiv:2309.09558, Sep. 2023.

[4] C. G. Belém, P. Pezeshkpour, H. Iso, S. Maekawa, N. Bhutani, and E. Hruschka, “How LLMs Hallucinate in Multi-Document Summarization,” in Findings of the Association for Computational Linguistics: NAACL 2025, 2025, doi:10.18653/v1/2025.findings-naacl.293. ACL Anthology

[5] N. Shinn, F. Cassano, E. Berman, A. Gopinath, K. Narasimhan, and S. Yao, “Reflexion: Language Agents with Verbal Reinforcement Learning,” in Advances in Neural Information Processing Systems 36 (NeurIPS 2023), 2023. NeurIPS Papers

[6] H. Zhang, X. Liu, and J. Zhang, “SummIt: Iterative Text Summarization via ChatGPT,” in Findings of the Association for Computational Linguistics: EMNLP 2023, pp. 10644- 10657, 2023, doi:10.18653/v1/2023.findings-emnlp.714. ACL Anthology+ 1

[7] Braž inskas, A., Lapata, M., and Titov, I., “Learning Opinion Summarizers by Selecting Informative Reviews,” arXiv preprint arXiv:2109.04325, Sep. 2021.

[8] Angelidis, S., Amplayo, R. K., Suhara, Y., Wang, X., and Lapata, M., “Extractive Opinion Summarization in Quantized Transformer Spaces,” Trans. of the Assoc. for Computational Linguistics (TACL), 2021.

생성형 인공지능과 검색 증강 생성을 활용한 학과 정보 제공 시스템 개발

천민우, 정세준, 신혜민, 김시현, 최정열*

성결대학교

{minwoce, jungsj0524, shm7696, huen2003, passjay}@sungkyul.ac.kr

Development of an Academic Information Provision System Using Generative AI and Retrieval-Augmented Generation

Chun Min Woo, Jung Se Jun, Shin Hye Min, Kim Si Hyun, Choi Jung Yul*

Sungkyul Univ.

요 약

본 논문은 전공자율선택제에서 신입생의 전공 탐색 지원 및 교육과정 소개를 위해 생성형 인공지능과 검색 증강 생성에 기반한 학과 정보 제공 시스템 개발 내역을 소개한다. 교육과정 문서 내용을 바탕으로 청킹과 임베딩을 수행하고 데이터베이스에 저장하여 의미 기반 검색이 가능한 파이프라인을 구축하였다. 사용자 질의를 바탕으로 관련 정보를 검색한 결과를 생성형 인공지능에 전달하여 개인화된 응답을 생성하도록 구성하였다. 제안하는 시스템은 다양한 전공 체계에도 확장할 수 있는 학습 지원 도구로서의 가능성을 제시한다

Ⅰ. 서 론

대학 내 전공자율선택제의 확대로 신입생은 관심과 적성에 맞춰 전공을 선택할 수 있으나 교과목 간 선후수 관계나 진로 연계성에 대한 명확한 안내 부족으로 인해 수강 계획 수립이 어려운 경우가 많다. 이러한 상황은 학습 동기 저하와 전공 이수의 비효율성을 초래할 수 있어 신입생이 보다 직관적이고 개인화된 정보를 바탕으로 전공을 탐색할 수 있는 지원 도구가 필요하다.

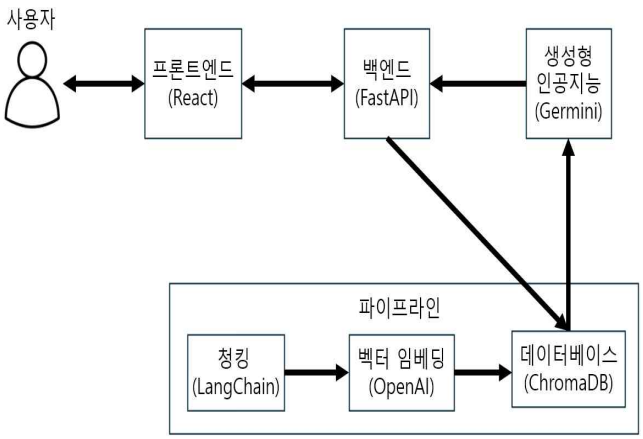
본 논문에서는 학생들의 전공 및 수강 과목 선택 과정에서 발생하는 정보 탐색의 어려움을 해소하고 개인화된 학습 경로 추천을 지원하기 위해 생성형 인공지능 기반의 학과 정보 제공 시스템을 제안한다. 기존의 정적 안내 웹페이지나 수동적 상담 방식과 달리 사용자의 다양한 질문 의도와 맥락을 실시간으로 파악하고 맞춤 과목 구성, 이수 순서, 진로 연계 추천 등에 대해 개인화된 답변을 제공할 수 있다는 점에서 차별성을 가진다.

Ⅱ. 본론

2.1 시스템 구조

본 논문에서 제안하는 맞춤형 학과 정보 안내 시스템은 생성형 인공지능과 검색 증강 생성 기술을 기반으로 한다 [1]. 제안하는 시스템은 랭체인(LangChain) 프레임워크와 ChromaDB를 활용한 지식 검색 및 응답 생성 파이프라인을 이용하여 구축하였다. 파이프라인은 교육과정 문서를 의미 단위로 청킹(Chunking)한 뒤, OpenAI API를 활용하여 임베딩(Embedding)하고, 그 결과 생성된 벡터 데이터를 데이터베이스에 저장하는 과정으로 진행된다.

사용자는 리액트(React) 기반의 프론트엔드 인터페이스를 활용하여 자연어 질의를 진행한다. 질의는 FastAPI 기반의 백엔드 서버로 전달되며, 사전에 구축한 파이프라인을 활용하여 의미적 유사도 검색을 실행한 뒤, 상위 검색 결과를 기반으로 적절한 컨텍스트(Context)를 구성하여 전달한다. 이후 생성된 컨텍스트는 프론트엔드로 반환되어 사용자의 질의에 대한 답변을 제공한다. 다음 절에서는 시스템 주요 구성 요소별 기능 및 과정을 상세히 설명한다. <그림 1>은 시스템의 주요 흐름을 요약한다.



<그림 1> 시스템 아키텍처

2.2 데이터베이스 설계 및 구축

데이터베이스는 컴퓨터공학과 교육과정 문서를 기반으로 구축하였다. 교육과정 문서는 PDF 형식으로 제공되었으며, 랭체인 라이브러리를 활용하여 문단 단위로 텍스트를 추출하고, 의미 단위별로 청킹 과정을 수행하였다. 청킹 데이터는 OpenAI의 text-embedding-3-small 모델을 이용해 벡터 임베딩 처리를 거친 후, ChromaDB에 컬렉션 형태로 저장하였다 [2]. 이때 각 청킹 블록에는 과목명, 페이지 번호, 교과목명, 전공역량과 같은 메타데이터가 함께 저장되어 검색 시 필터링 및 세부 검색이 가능하도록 설계하였다. 이러한 데이터 구축 과정은 사용자의 질의에 대해 보다 정확하고 개인화된 정보를 검색하고 제공하기 위한 기반이 된다.

2.3 검색 증강 생성(RAG) 파이프라인

본 시스템은 검색 증강 생성을 기반으로 구현하였다. 외부 지식베이스에서 관련 정보를 검색한 후, 이를 생성형 인공지능과 결합하여 정확하고 풍부한 응답을 생성할 수 있도록. 이를 통해 기존의 단순 생성형 언어 모델이 가지는 정적 지식 한계와 최신성 부족 문제를 보완할 수 있다. 사용자

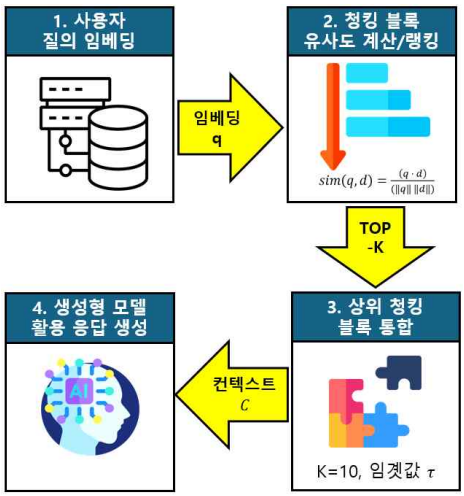
가 자연어 질의를 입력하면 백엔드 서버는 ChromaDB 내부의 컬렉션을 활용한 의미적 유사도 검색을 수행한다. 이때 의미 유사도는 벡터 임베딩 간의 코사인 유사도를 기준으로 계산되며, 그 수식은 다음과 같다.

$$sim(q,d) = \frac{(q \cdot d)}{(\|q\| \, \|d\|)} \quad (1)$$

q는 사용자 질의의 임베딩 벡터, d는 청킹 블록의 임베딩 벡터, $\| \cdot \|$는 벡터의 크기(norm), $\cdot$은 벡터의 점곱(dot product)을 의미한다. 상위 유사도를 가지는 k개의 청킹 블록은 컨텍스트를 구성하는 청킹 블록 집합 C로 정의되며 그 조건은 다음과 같다.

$$C = \{ d^1, d^2, \ldots, d_k \} \; such \; that \; sim(q, d_i) \geq \tau \quad (2)$$

여기서 τ는 유사도 임곗값이며 C는 최종적으로 생성형 인공지능에 전달되어 응답 생성을 위한 컨텍스트로 활용된다. 현재 시스템은 컬렉션 내에서 상위 유사도를 보이는 10개의 청킹 블록을 검사 후 컨텍스트로 통합한다. 이러한 파이프라인 구조를 통해 사용자는 자신의 질문에 가장 적합한 정보를 기반으로 한 응답을 실시간으로 제공받을 수 있으며, 기존의 키워드 기반 검색이나 단순 생성 방식에 비해 높은 정확도와 개인화된 정보 전달이 가능하다. <그림 2>는 이러한 과정을 네 단계로 시각화한 것이다. 먼저 사용자 질의를 바탕으로 임베딩하여 특정 조건(학년, 관심분야 등)에 따라 해당하는 컬렉션을 호출·선택한다. 이후 선택된 컬렉션 내부의 모든 청킹 블록을 대상으로 질의 벡터와의 코사인 유사도를 계산하여 관련성이 높은 순으로 정렬한다. 다음으로 상위 10개의 청킹 블록을 통합하여 컨텍스트를 구성한다. 마지막으로 컨텍스트를 생성형 인공지능 모델에 전달하여 최종 응답을 생성한다.



<그림 2> 파이프라인 동작 과정

2.4 페르소나 설정 및 웹 인터페이스 구성

RAG 파이프라인을 통해 만들어진 컨텍스트를 생성형 인공지능에 전달하여 사용자에게 적합한 답변을 만들어낼 수 있다. 생성형 인공지능 모델로 gemini-1.5-pro를 사용하였으며 '컴퓨터공학과 멘토'라는 페르소나(Persona)를 설정하였다 [3]. 설정된 페르소나는 친근하고 신뢰감을 주는 조언자의 역할을 수행하며 과목 선택과 학습 경로 구성을 돕는 안내자의 입장에서 응답하도록 설계하였다. 대화 방식은 자연스러운 구어체를 채택하였으며, 기술 용어의 사용을 최소화하여 비전공자나 신입생도 쉽게 이해할 수 있도록 하였다. 특히 과목 이수 순서 안내 시에는 학습 경로를 명확하고 단계적으로 제시하도록 하였다.

프론트엔드는 React를 기반으로 웹 인터페이스를 구현하였으며, 직관적인 질문 입력창과 답변 표시창을 제공하여 사용자가 질의를 입력하고 응답을 쉽게 확인할 수 있도록 구성하였다. 이후 FastAPI 기반 백엔드와 REST API를 통해 통신하며, CORS(Cross-Origin Resource Sharing) 설정을 통해 프론트엔드-백엔드 간 연동이 원활하게 이루어지도록 하였다. 질의 요청은 /ask 엔드 포인트로 전달되며, 백엔드는 생성형 인공지능을 활용해 완성한 답변 결과를 프론트엔드로 다시 반환한다. 아래의 <그림 3>은 1학년 자율전공학부 학생이 인공지능 개발자와 관련된 교과목을 추천받는 상황을 가정하여 질의 후 응답 예시이다.

<그림 3> 시스템 질의 응답 예시

Ⅲ. 결론

본 연구에서는 생성형 인공지능과 검색 증강 생성을 기반으로 학과 정보 제공 시스템을 설계·구현하였다. 이어서 컴퓨터공학과 교육과정 문서를 청킹하여 임베딩 처리를 거쳐 데이터베이스를 완성하였다. 사용자가 질의를 하면 파이프라인을 통해 의미 기반 청킹 블록 검색을 수행하고 그 결과를 바탕으로 컨텍스트를 구성하여 생성형 인공지능에 전달함으로써 응답을 생성하도록 하였다. 생성형 인공지능에는 '컴퓨터공학과 멘토'라는 페르소나를 부여하여 신입생이 친근하고 이해하기 쉬운 방식으로 안내를 받을 수 있도록 하였고, 프론트엔드와 백엔드는 React와 FastAPI를 통해 연동되도록 구현하였다.

본 시스템은 컴퓨터공학과에 대한 정보만을 포함하고 있으나 향후 대학 내 모든 전공에 대한 정보를 포함하도록 시스템을 확장함으로써 융합전공, 복수전공, 부전공 등 유연한 학사제도에서 학생들이 활용할 수 있을 것이며 이를 통해 맞춤형 학습 지원 도구로서의 활용될 것으로 기대한다.

참 고 문 헌

[1] P. Lewis, E. Perez, A. Karpukhin, N. Piktus, F. Petroni, V. Karpukhin, P. Stenetorp, S. Riedel, "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," Advances in Neural Information Processing Systems, Vol. 33, pp. 9459–9474, 2020.

[2] Q. Liu, M. J. Kusner, and P. Blunsom, "A Survey on Contextual Embeddings," arXiv preprint, arXiv:2003.07278, 2020.

[3] Gemini Team, A. Vyas, H. Xia, R. Urtasun, et al., "Gemini 1.5: Unlocking Multimodal Understanding across Millions of Tokens of Context," arXiv preprint, arXiv:2403.05530, 2024.

온프레미스(On-premise) 환경에서금융투자정보 분석을 위한 MCP(Model Context Protocol), A2A(Agent to Agent) 하이브리드 기반 멀티에이전트 시스템 설계 및 구현

박시형[1], 김두홍[1], 이재희[2]
[1]동서울대학교 컴퓨터 소프트웨어과 학부생
[2]동서울대학교 컴퓨터 소프트웨어과 교수
qkrtlgud0620@naver.com, dh0504k@gmail.com, lih7314@du.ac.kr

Design and Implementation of a MCP(Model Context Protocol), A2A(Agent to Agent) Hybrid-Based Multi-Agent System for Financial Investment Information Analysis in an On-premise Environment

Si-Hyung Park[1], Doo-Hong Kim[1], Jae-Hee Lee[2]
[1] [2]Dept. of Computer Software, Dong-Seoul University

요 약

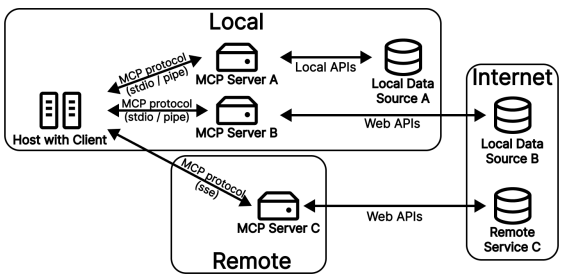
본 연구는 온프레미스 환경에서 금융투자 정보분석 자동화를 위해 MCP(Model Context Protocol)와 A2A(Agent-to-Agent)를 결합한 하이브리드 멀티에이전트 시스템을 설계·구현하였다. 제안 시스템은 생성, 검색, 검증 3개의 에이전트로 구성되며, MCP를 통해 금융 데이터를 수집하고 로컬에 저장된 문서는 RAG(Retrieval-Augmented Generation) 기법으로 검색·요약한 뒤, 최종적으로 MCP를 호출하여 금융투자정보 분석 보고서를 출력한다. 에이전트 간 A2A 프로토콜을 통해 조율되어 상호 협력하며 내부에서는 MCP기반으로 동작한다. 본 연구의 결과로 금융 보고서 작성의 자동화 가능성을 확인하였고, 멀티에이전트 기술의 확장 가능성을 검증하였다.

1. 서론

최근 금융투자 분야에서는 실시간 시장 데이터, 기업 재무정보, 뉴스 등 방대한 자료를 바탕으로 신속하고 정확한 의사결정이 요구되고 있다. 그러나 기존 보고서 생성 과정은 분석가의 수작업에 의존하여 데이터 최신성 확보가 어렵고, 누락이나 불일치가 발생하기 쉽다. 이를 보완하기 위해 RAG(Retrieval-Augmented Generation) 기반 자동화가 시도되고 있었다[1]. RAG시스템은 복잡한 질의에 대한 다단계 추론 과정에서 오류가 누적되거나 검색 결과의 진위 여부를 검증하기 어렵다는 한계가 여전히 존재한다. 따라서 금융투자 보고서와 같은 고신뢰 환경에서는 데이터 수집, 검증, 생성 전 과정을 통합적으로 관리하며 최신성과 신뢰성을 보장할 수 있는 새로운 자동화기법을 필요로 한다.

2. MCP(Model Context Protocol)

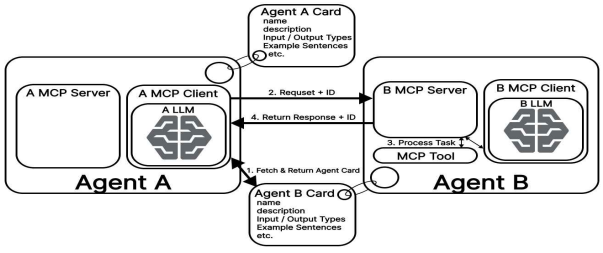
MCP는 인공지능 모델과 외부 데이터 소스 및 도구 간의 상호작용을 표준화하는 프로토콜이다[2]. 기존 RAG 기반 시스템은 검색된 문서를 단순히 모델 입력에 포함시키는 방식에 머물러, 실시간 데이터 반영이나 외부 자원에 대한 직접적 접근에는 한계가 있었다. MCP는 이러한 한계를 보완하여 모델이 실행 중 필요한 경우 특정 툴(Tool)을 호출하고, 이를 통해 외부 데이터베이스나 API와 같은 자원에도 직접 접근할 수 있도록 설계되었다. 따라서 다양한 금융 데이터를 일관된 방식으로 통합·처리할 수 있다. 그림 1은 본 연구에서 활용하는 MCP의 구조와 데이터 흐름을 보여준다.



[그림 1] MCP 구조도

3. A2A(Agent-to-Agent)

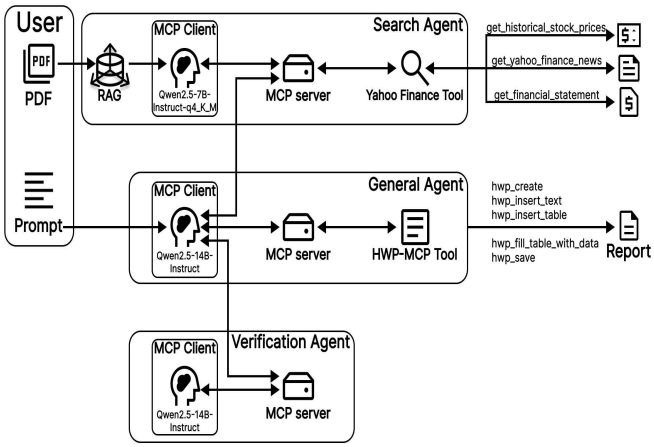
A2A는 여러 지능형 에이전트가 역할을 분담하고 상호 협력하여 복잡한 문제를 해결하는 프로토콜이다[3]. 단일 모델 기반 접근은 모든 작업을 하나의 모델이 수행하기 때문에 특정 업무에 최적화되기 어렵고, 다단계 추론 과정에서 오류가 누적될 수 있다. 이에 비해 A2A는 작업을 기능별로 세분화하고, 에이전트 간 메시지 교환을 통해 협력적으로 문제를 해결함으로써 이러한 한계를 보완한다. 그림 2는 A2A의 구성도를 나타낸다.



[그림 2] A2A 구조도

4. MCP, A2A 기반 금융투자정보 분석 자동화

본 연구에서는 금융투자정보 분석 자동화를 위해 MCP 기반 데이터 수집과 A2A 기반 멀티에이전트 협업 구조를 결합하였다. 제안 시스템은 생성, 검색, 검증의 세 에이전트가 상호 협력하여 동작한다. 생성 에이전트는 사용자 프롬프트를 입력받아 필요한 정보를 판별하고 검색 에이전트에 요청한다. 검색 에이전트는 MCP의 툴을 통해 주가, 재무, 뉴스와 같은 외부 금융 데이터를 수집하며, 로컬에 저장된 문서에 대해서는 RAG 기법을 적용해 관련 내용을 검색·요약한 후 생성 에이전트에 전달한다. 검증 에이전트는 생성된 보고서 초안을 일관성, 응집성, 인용 기준으로 평가하고, 필요한 경우 수정 사항을 생성 에이전트에 반환한다. 이 과정은 반복적으로 수행되며, 검증 에이전트가 추가 수정이 필요 없다고 판정할 때까지 보고서 품질은 점진적으로 향상된다. 마지막으로 생성 에이전트는 문서 작성 MCP를 호출하여 최종 보고서를 아래한글형식의 파일로 자동 생성한다. 그림 3은 제안 시스템의 전체 구조를 보여준다.



[그림 3] 제안한 방식의 구조도

5.실험 및 구현환경

본 연구에서 구축한 실험 및 구현환경은 표 1과 같다.

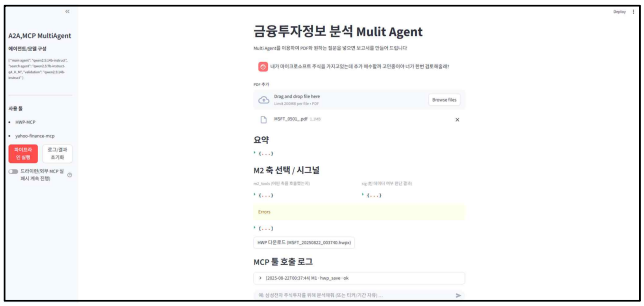
<표 1> 구축한 실험 및 구현 환경

구분	항목		규격
Work station	H/W	CPU	Intel Core i5-12600K
		RAM	32 GB
		GPU	NVIDIA RTX A5000 (24GB VRAM)
	S/W	OS	Windows 11
Frame work	Python		3.11
	LangGraph		0.6.4
	LangChain		0.3.26
	ollama		0.5.3
LLM model	생성 LLM		Qwen2.5-14B-Instruct
	검색 LLM		Qwen2.5-7B-Instruct-q4_K_M
	검증 LLM		Qwen2.5-14B-Instruct
MCP Tool	금융 데이터 수집		Yahoo Finance MCP
	보고서 작성		HWP-MCP

제안된 시스템을 이용하여 미국의 나스닥에 상장된 마이크로소프트(MSFT) 주식에 대한 금융투자정보 분석을 수행하였다. 그림 4는 금융투자 정보분석 과정중 MCP서버를 통한 툴(Tool)호출 결과화면이다.

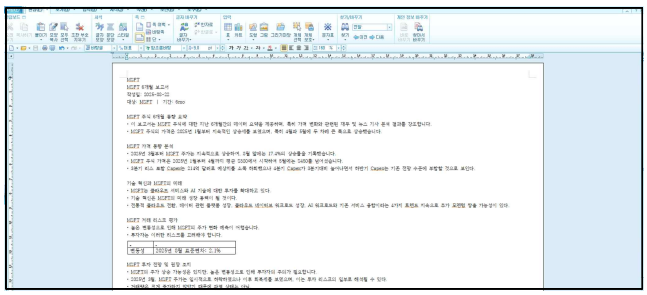
[그림 4] MCP서버를 통한 툴 호출

사용자 질의에 따라 보고서가 자동으로 생성되는 과정은 다음 그림 5와 같다.



[그림 5] 사용자 질의에 대한 보고서 생성

자동 생성된 금융투자 보고서는 다음 그림6과 같다.



[그림 6] 최종 출력 보고서

6.결론

본 연구에서는 금융투자정보 분석을 위해 온프레미스 환경에서 MCP, A2A 하이브리드 기반 멀티에이전트 협업 구조를 결합한 자동 분석 보고서 생성 시스템을 설계하고 구현하였다. 제안된 시스템은 다양한 금융 데이터를 수집하고, 생성, 검색, 검증 에이전트의 협력을 통해 보고서를 자동으로 생성할 수 있음을 확인하였다. 그러나 보고서의 질과 내용적 완성도를 높이기 위해 전이학습 된 맞춤형 LLM 모델 사용을 통한 성능 향상 연구가 필요하다고 판단된다. 본 연구는 금융투자 보고서 자동화의 가능성을 실증하였으며, 향후 연구를 통해 더욱 정교하고 신뢰성 있는 보고서 생성으로 발전할 수 있음을 검증하였다.

참고문헌

[1] Van-Duc Le, Auto-Generating Earnings Report Analysis via a Financial-Augmented LLM, arXiv:2412.08179

[2] Xinyi Hou et al., "Model Context Protocol (MCP): Landscape, Security Threats, and Future Research Directions", arXiv:2503.23278

[3] Awid Vaziry et al., Towards Multi-Agent Economies: Enhancing the A2A Protocol with Ledger-Anchored Identities and x402 Micropayments for AI Agents, arXiv:2507.19550

KV 캐시 스토리지 오프로딩에서 유사 프롬프트의 캐싱 방법에 관한 연구

이청솔
삼성전자

cs89.lee@samsung.com

A Study on the Caching Strategies for Semantically Similar Prompts in KV Cache Storage Offloading

Lee Choungsol

Samsung Electronics

요 약

스토리지 오프로딩 기반 LLM 캐시는 일반적으로 입력 토큰의 정합 일치(exact match)에 대해서만 KV 캐시 재사용을 허용해 실제 환경의 다양한 표현 변형에는 한계가 있다. 본 연구는 이러한 한계를 극복하기 위해 의미적으로 유사한 프롬프트까지 재사용 범위를 확장하는 유사도 기반 KV 캐시 재사용 기법을 제안한다. 공개 데이터셋을 활용한 실험에서 제안 기법은 응답 품질을 유지하면서도 TTFT 를 단축시켰으며, 특히 입력이 길수록 효과가 커져 TTFT 가 최대 80.4% 감소하였다. 본 기법은 추가 학습 없이 기존 서빙 스택과 호환되어 실서비스 환경에서 실용적 이득을 제공한다.

Ⅰ. 서 론

대규모 언어 모델(LLM)에서 KV 캐시를 활용한 성능 최적화가 중요한 연구 주제로 떠오르고 있다. LLM 추론의 총 지연시간(latency)은 프리필(입력 토큰 처리) 단계와 디코드(출력 토큰 생성) 단계로 구성되며, 특히 긴 컨텍스트를 처리할 경우 TTFT(Time To First Token)가 체감 성능을 좌우한다.

일반적으로 프리필 단계에서 KV 캐시를 GPU 메모리에서 찾아 사용하는데, 근래 연구에서는 GPU 메모리의 크기 한계로 인해 CPU 메모리나 스토리지로 오프로딩 하는 연구가 이루어지고 있다. 현재 대표적인 스토리지 오프로딩 사례로는 Deepseek 의 context caching[1]이 있으며, 캐싱 전략은 exact match 를 기반으로 한 KV 캐시 재사용이다. 그러나 원격 저장소 환경에서 다수의 사용자가 KV 캐시를 공유하는 상황이나, 단일 환경이라도 유사 프롬프트를 빈번하게 사용하는 등 실제 응용 상황을 고려할 때 exact match 방식은 한계가 있다. 본 논문에서는 문장 간 유사도가 높은 경우에도 캐시를 매칭하여 KV 캐시를 재사용하는 방법을 제안하며, 이를 통해 프리필 단계의 응답 지연 감소와 시스템 처리 효율을 향상시키는 것을 목표로 한다.

이를 해결하기 위해 본 연구는 (1) 유사문장이 입력되면 스토리지에 저장된 프롬프트 임베딩을 벡터 검색으로 찾아 해당 프롬프트에 매칭된 재사용할 KV 캐시를 선택하고, (2) 벡터 검색으로 찾은 원본 프롬프트의 토큰열을 그대로 사용하도록 입력을 대체하여, 최종적으로 모델에 주입되는 토큰열이 캐시 생성 시점과 정확히 동일해지도록 함으로써 exact match 기반 KV 캐시 재사용을 달성한다. 논문의 핵심 질문은 다음과 같다.

Q1. 스토리지에 저장된 KV 캐시를 재사용하면 TTFT 가 감소하는가?

Q2. TTFT 감소효과는 모델 및 입력길이에 따라 달라지는가?

Q3. 유사 프롬프트를 원본문장으로 정규화하여 KV 캐시를 재사용할 경우, 직접 인퍼런스를 수행한 결과와 의미적으로 동등한가?

Ⅱ. 본론

2.1 제안 기법 개요 본 연구에서는 문장 간 의미 유사도에 기반한 KV 캐시 매칭 알고리즘을 제안한다. 실제 서비스 환경에서 사용자가 생성하는 다양한 프롬프트를 최대한 대체할 수 있도록 여러 데이터셋을 활용하였다. 데이터셋의 문장을 패러프레이즈하여 유사문장을 생성한 뒤 실험을 진행하였으며, 검색 결과의 유사도가 일정 기준 이상인 경우 해당 원본문장의 KV 캐시를 스토리지에서 불러와 프리필 단계에서 재사용하였다.

2.2 실험 방법 먼저, vLLM 서버에 LMCache[2]를 연동하여 각 데이터셋의 문장을 입력하고 프리필을 수행하여 KV 캐시를 생성한 뒤 디스크에 저장하였다. 이어서 동일 문장의 임베딩을 계산하여 FAISS[3] 인덱스에 등록하였다. 이후 원본문장을 대규모 언어 모델로 패러프레이징 하여 단어 및 구문을 변형함으로써 exact match 가 성립하지 않도록 하였다. 패러프레이징 시 문장 생성 temperature 는 0.1 부터 1.0 까지 검증한

뒤, 의미 보존과 다양성의 균형을 고려하여 0.3 을 최종 설정하였다. 이렇게 생성된 유사문장은 임베딩 후 벡터 검색을 수행하여 Top-1 후보를 도출하고, 검색된 원본문장과 질의 간의 코사인 유사도가 사전에 설정한 임계치 이상일 경우 해당 원본문장의 KV 캐시를 디스크에서 불러와 프리필 단계에서 재사용하였다. 마지막으로, 정확도 평가는 각 문단에 대한 질문을 모델에 제시하여 출력 응답과 데이터셋의 정답을 비교하고, F1 및 EM 지표를 계산함으로써 수행하였다.

그림 1. 실험 개요

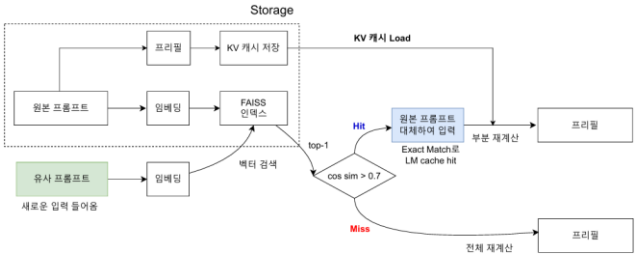


그림 1 은 제안 기법의 전체 실험 파이프라인 개요이다. 유사 프롬프트 입력 시 벡터 검색과 코사인 유사도 임계치(0.7)를 통해 원본 프롬프트로 정규화하고, 해당 KV 캐시를 로드하여 재사용한다.

실험에는 NVIDIA H100 80GB GPU, Intel(R) Xeon(R) Platinum 8480+ CPU, 삼성전자 서버용 DDR5 RAM, 삼성전자 Gen5 SSD 를 사용하였다.

표 1. Llama 3.1-8B 데이터셋별 F1/EM

데이터셋	특징	토큰	F1	EM
SQuAD	Wiki	150	0.81	0.63
QuAC	Wiki	400	0.72	0.6
NewsQA	CNN	800	0.61	0.37
TriviaQA	Wiki	10k	0.54	0.2

표 1 은 유사문장 벡터서치 및 KV 캐시를 재사용한 출력과 데이터셋의 정답을 비교한 것이다. 높은 F1 점수로 미루어 보아, 유사문장 입력시 원본문장과 매칭이 잘 이루어지고 있음을 알 수 있다. 토큰수가 적은 데이터셋일 수록 정확한 정답을 얻을 수 있었다.

그림 2. 데이터셋 별 TTFT 감소 효과 비교

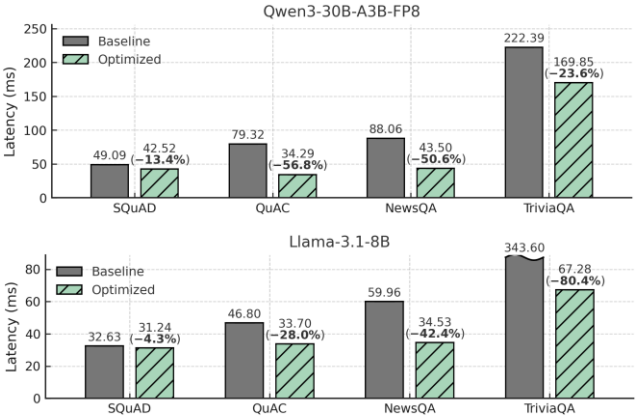


그림 2 에서 보이듯 TTFT 감소는 데이터셋의 입력 토큰 수가 증가할수록 더욱 뚜렷하게 나타났다. 이는 입력이 길어질수록 프리필 비용이 커지기 때문에, 캐시 재사용에 따른 상대적 이득이 확대되었기 때문이다. 다만 Qwen3-30B 모델은 가장 긴 입력에서 TTFT 감소율이 낮게 관찰되었다. 이는 해당 모델이 유사문장 생성 시 표현 다양성이 높아 miss 발생 빈도가 늘어났고, miss 가 발생할 경우 전체 프리필을 다시 수행해야 하므로 평균 지연시간 개선폭이 줄어든 것으로 추정된다.

Ⅲ. 결론

본 논문에서는 기존 exact match 기반 KV 캐시 재사용의 한계를 극복하기 위해 문장 간 의미 유사도 기반의 KV 캐시 매칭 기법을 제안하였다. KV 캐시를 디스크에 저장하고 불러오는 과정에서 TTFT 이득을 얻었고, 벡터검색과 코사인 유사도 검증으로 잘못된 캐시 로드를 방지해 신뢰성을 높였다. 서론에서 이야기한 논문의 핵심 질문에 대한 대답은 다음과 같다.

Q1. 스토리지에 저장된 KV 캐시를 재사용하면 TTFT 가 감소하는가? 모든 데이터셋에서 TTFT 가 유의미하게 감소하였다.

Q2. TTFT 감소 효과는 모델 및 입력 길이에 따라 달라지는가? TTFT 절감은 입력 길이에 따라 뚜렷하게 달랐다. 짧은 입력은 4% 수준에 그쳤으나, 10k 토큰 입력에서는 최대 80.4%까지 감소하였다. 다만 Qwen3-30B 모델은 miss 발생률이 높아 대규모 입력에서 개선 폭이 제한적이었다.

Q3, 유사 프롬프트를 원본문장으로 정규화하여 KV 캐시를 재사용할 경우, 직접 인퍼런스를 수행한 결과와 의미적으로 동등한가? 그렇다. 유사문장을 입력해 의미적으로 같은 원본문장을 찾아 대체해 KV 캐시를 재사용 한 경우 원래 데이터셋의 정답과 F1/EM 상 높은 일치도를 보였다.

ACKNOWLEDGMENT

This work was supported by the Samsung Electronics Samsung Memory Research Center (SMRC). The authors would like to express their gratitude to the SMRC research staff for their valuable technical assistance and insightful discussions throughout the course of this study.

참 고 문 헌

[1] DeepSeek, “Context Caching.” Available: https://api-docs.deepseek.com/guides/kv_cache

[2] OpenLM Research, “LMCache: Efficient KV Cache Storage Offloading for LLM Inference,” 2024. Available: https://lmcache.ai/

[3] J. Johnson, M. Douze, and H. Jegou, “FAISS: A library for efficient similarity search and clustering of dense vectors,” 2019. Available: https://github.com/facebookresearch/faiss

Scene Graph 생성에서 환각 완화를 위한 Negative Data Augmentation 기법

백종우, 소진현*
대구경북과학기술원

jaden9420@dgist.ac.kr, *jinhyun@dgist.ac.kr

Negative Data Augmentation for Mitigating Hallucination in Scene Graph Generation

Jongwoo Baek, Jinhyun So*
DGIST

요 약

Scene Graph는 이미지 내 객체와 그들 사이의 관계 및 속성을 구조적으로 표현함으로써 정밀한 시각적 이해를 가능케 한다. 그러나 현존하는 Scene Graph 데이터셋은 클래스 편향과 관계의 다양성 부족 문제를 내포하고 있어, 학습된 모델이 존재하지 않는 정보를 생성하는 환각(hallucination) 문제를 유발한다. 본 연구에서는 이러한 문제를 해결하기 위해 부정 데이터 증강(Negative Data Augmentation, NDA) 기법을 제안한다. NDA는 모델이 빈번히 혼동하는 관계 클래스를 기반으로 부정 후보를 생성하고 이를 학습에 활용하는 전략이다. 본 기법은 추가적인 데이터 수집 없이도 학습 데이터의 분포적 한계를 보완하여 모델 일반화 성능을 향상시킨다. 실험 결과, NDA를 Scene Graph 생성 모델에 적용하였을 때 환각 현상을 효과적으로 줄이고 다양한 관계 인식 성능을 개선함을 확인하였다.

Ⅰ. 서 론

Scene Graph는 이미지 내 객체(object)뿐만 아니라 객체 간의 관계(relation)를 구조화된 그래프 형태로 표현하는 자료구조이다. 이 구조는 객체 간의 상호작용을 명시적으로 기술할 수 있어, 이미지 캡셔닝, 비디오 이해, 질의응답 등 다양한 시각-언어 응용 분야에서 평가 벤치마크 등으로 많이 사용된다.

Scene Graph 생성 모델은 이미지로부터 객체, 관계 클래스, 속성을 예측하여 Scene Graph를 구성하는 모델이다. 이러한 모델의 학습을 위해서는 이미지-그래프 쌍으로 구성된 데이터셋이 필요하다. 그러나 Public Scene Graph 데이터셋들은 수량과 품질 모두에서 한계를 지니고 있으며, 그림 1,2 에서 볼 수 있듯, 관계 클래스의 편향이 심각하다. 특히 그림 2 의 Open Image[1], GQA[2] 데이터셋의 경우 상위 1~2 개의 관계 클래스가 전체 관계의 90% 이상을 차지한다. 이러한 데이터 편향은 모델이 학습 중 특정 패턴에 과적합되어 존재하지 않는 객체나 관계를 생성하는 환각(hallucination) 문제를 야기한다[3].

본 연구는 이러한 환각 문제를 완화하기 위한 부정 데이터 증강(Negative Data Augmentation, NDA) 기법을 제안한다. NDA는 모델이 자주 혼동하거나 오답을 생성하는 관계 클래스의 특징을 학습에 반영하여, 학습 중 잘못된 일반화를 방지하고 모델의 강건성을 높인다. 본 기법은 추가적인 데이터 수집 없이도 데이터의 표현력을 향상시키는 데이터 중심 해결책으로 작용한다[4].

Ⅱ. 배경 지식

Scene Graph 데이터셋 D는 이미지-그래프 쌍의 집합으로 정의된다.

$$\mathrm{D} = (I_n, G_n) | n = 1, \ldots, N$$

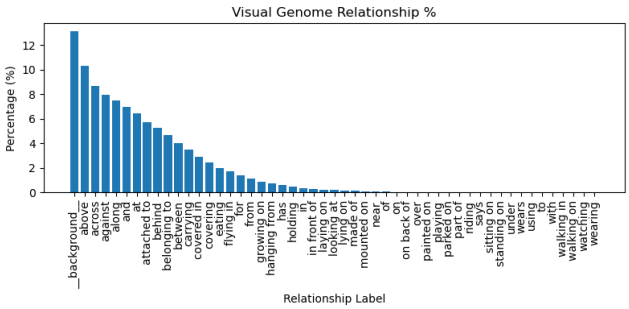


그림 1 Visual Genome 데이터셋의 관계 클래스 분포

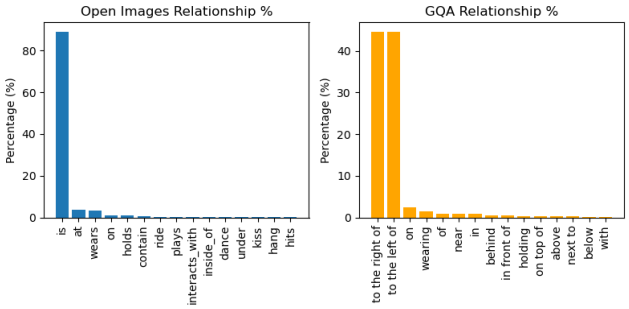


그림 2 Open Image, GQA 데이터셋의 관계 클래스 분포

여기서 I_n은 이미지, G_n은 이에 상응하는 그래프를 의미한다. 각 그래프 G_n은 삼중항(triplet) T의 집합이며, 각 triplet $\mathrm{T_i}$은 다음과 같이 정의된다:

$$\mathrm{T_i} = (o_i, r_i, s_i, b_{o_i}, b_{s_i})$$

여기서 $\mathrm{o_i}$는 주체(subject), $\mathrm{r_i}$는 관계(relation), $\mathrm{s_i}$는 객체(object), b_{o_i}와 b_{s_i}는 주체와 객체의 바운딩 박스(bounding box)이다. 바운딩 박스는 객체의 위치를 명확히 정의하기 위해 사용되며, 최소 및 최대 좌표 $(x_{min}, y_{min}, x_{max}, y_{max})$ 형태로 표현된다.

Ⅲ. 제안 기법: Negative Data Augmentation

본 연구에서는 사전 학습된 모델(Pre-trained Model)을 기반으로 오분류(class confusion)에 민감하게 반응하는 NDA 기법을 제안한다. 본 기법은 fine-tuning 단계에서 클래스 간 혼동(confusion) 정보를 활용하여, 모델의 일반화 성능을 향상시키고 드문 클래스의 예측 성능을 보완하는 것을 목표로 한다.

우선, 사전 학습된 모델을 이용하여 validation set 전체에 대한 예측 결과를 수집하고, 이로부터 정답 관계 클래스 c 에 대하여 모델이 오분류한 클래스들 간의 confusion matrix $\boldsymbol{M} \in \mathrm{R}^{C\times C}$ 를 구축한다. 여기서 C 는 전체 클래스 수를 나타낸다. 각 행 $\boldsymbol{M}_{c,:}$는 정답 c에 대해 오답으로 예측된 클래스들의 수를 나타내며, 다음과 같이 정규화되어 확률 분포로 사용된다.

$$\widetilde{\boldsymbol{M}}_{(c,j)} = \frac{\boldsymbol{M}_{(c,j)}}{\sum_{j'\neq c}\boldsymbol{M}_{(c,j')}} \ (for\ j \neq c)$$

이를 기반으로 데이터셋의 각 정답 클래스 $c^{(i)}$ 에 대하여, 정규화된 confusion 분포 $\widetilde{\boldsymbol{M}}_{c^{(i)},:}$ 로부터 샘플링을 수행한다. 이를 통해 $c^{(i)}$ 마다 크기 k 의 negative label 집합 $N_{c^{(i)}} \subset \{1,\dots,C\}$ 가 정의된다.

fine-tuning 과정에서는 기존 손실 함수에 샘플링된 negative label 의 예측을 억제하는 negative loss 항을 추가하여 진행된다. 각 입력에 대해 예측된 softmax 확률 벡터 $\mathrm{p} \in R^{C}$에서 negative label $\mathrm{j} \in N_c$에 해당하는 항만을 추출하여, 다음과 같이 손실을 정의한다:

$$\mathrm{L}_{\mathrm{neg}}^{(\mathrm{i})} = \gamma \cdot \frac{1}{\left|\mathrm{N}_{\mathrm{c}}^{(\mathrm{i})}\right|} \sum_{\mathrm{j}\in \mathrm{N}_{c(i)}} -\log\left(1-\mathrm{p}_{\mathrm{j}}^{(\mathrm{i})}\right)$$

여기서 γ는 negative loss 항의 기여 정도를 조절하는 하이퍼 파라미터이며, p_j는 클래스 j에 대한 예측 확률을 의미한다.

최종적으로 사용되는 loss 함수는 모델이 예측한 N_t 개의 클래스에 대해여 기존 모델의 loss 함수 L_{main}에 L_{neg}를 추가하여 다음과 같이 정의된다.

$$L_{total} = L_{main} + L_{\mathrm{neg}}$$

$$= L_{main} + \sum_{i=1}^{N_t} \gamma \cdot \frac{1}{\left|N_c^{(i)}\right|} \sum_{j\in N_{c(i)}} -\log\left(1-p_j^{(i)}\right)$$

이와 같이 손실 함수를 설계하여 모델이 자주 혼동하던 클래스에 대한 예측 확률이 억제되도록 한다.

Ⅳ. 실 험

본 연구는 제안하는 NDA 기법의 유효성을 정량적으로 검증하기 위해 SOTA Scene Graph 생성 모델 RelTR[5]을 기반으로 Visual Genome[6], Open Images[1] 데이터셋에서 실험을 수행하였다.

모든 실험은 동일하게 150 epoch 으로 사전 학습(pre-trained) 모델을 사용하였으며, NDA fine-tuning 10 epoch 을 추가로 진행하였다. 클래스 당 부정 샘플 개수 $\mathrm{k}=1$, negative loss 의 하이퍼 파라미터 $\gamma=0.5$ 를 사용하였다.

성능 평가는 이미지 입력만으로 클래스 레이블, 위치 정보를 추론하는 SGDET(Scene Graph Detection) 환경에서 상위 K 개의 클래스 평균에 대한 관계 예측 정밀도(top-K mean Precision, mP@K)와 관계 예측 재현율(top-K mean Recall, mR@K)을 활용하였다.

Metric	mP@20	mP@50	mP@100
Baseline	**5.159**	2.837	1.691
NDA	5.118	**3.116**	**2.066**

Metric	mR@20	mR@50	mR@100
Baseline	5.685	8.393	**9.869**
NDA	**6.042**	**8.548**	9.867

표 1. Visual Genome 데이터셋에서의 SGDET 결과 (mean Precision 및 mean Recall)

Metric	mP@20	mP@50	mP@100
Baseline	3.806	1.962	1.185
NDA	**3.812**	**2.038**	**1.260**

Metric	mR@20	mR@50	mR@100
Baseline	6.673	**9.378**	**10.719**
NDA	**6.675**	9.122	10.372

표 2. Open Image 데이터셋에서의 SGDET 결과 (mean Precision 및 mean Recall)

실험 결과, NDA 를 적용한 모델은 Baseline 대비 mean Precision 지표에서 전반적으로 향상된 성능을 보였다. Visual Genome 의 경우 mR@20 과 mR@50 구간에서 소폭 증가하였으며, mR@100 에서는 거의 동일한 수준을 유지하였다. 반면 mP@20 에서는 다소 감소가 관찰되었으나, mP@50 과 mP@100 에서는 뚜렷한 개선 효과가 확인되었다. Open Images 에서는 mR@20 이 거의 동일하고, mR@50 과 mR@100 에서 소폭 감소하는 양상을 보였으나, mean Precision 은 모든 cutoff 구간에서 일관되게 향상되었다.

Ⅴ. 결 론

본 연구는 Scene Graph 생성 모델에서 학습 데이터의 클래스 편향으로 인해 발생하는 환각 문제를 완화하기 위해 부정 데이터 증강(Negative Data Augmentation, NDA) 기법을 제안하였다. NDA 는 학습 과정에서 빈번히 잘못 예측되는 관계 클래스를 기반으로 negative label 을 생성하고 이를 fine-tuning 에 활용함으로써, 모델이 잘못된 관계를 반복적으로 예측하는 경향을 억제하여 일반화 능력을 확보한다.

실험 결과, NDA 는 mean Precision 지표에서 뚜렷한 개선 효과를 보였으며, 특히 top-K 평가에서 K 값(상위 예측 개수)이 커질수록 개선 폭이 두드러져 모델이 전체 예측 분포 전반에서 성능을 향상시켰음을 확인할 수 있다. mean Recall 은 데이터셋의 특성에 따라 소폭의 증감은 있었으나 전반적으로 baseline 과 유사한 성능을 유지하였다. 이러한 결과는 추가적인 데이터 수집 없이도 Scene Graph 생성 모델의 일반화 능력을 강화할 수 있음을 보여준다.

ACKNOWLEDGMENT

이 논문은 2025 년도 정부(과학기술정보통신부)의 재원으로 2025 년 혁신거점 인공지능 데이터 융합과제 사업의 지원을 받아 수행된 연구임(S2201-24-1002)

This thesis was conducted with the support of the 2025 innovation base artificial intelligence data convergence project with the funding of the 2025 government (Ministry of Science and ICT) (S2201-24-1002)

참 고 문 헌

[1] Open Images V7 Dataset: Krasin, Ivan, et al. OpenImages: A Public Dataset for Large-Scale Multi-Label and Multi-Class Image Classification. Google AI, 2017.

[2] GQA Dataset: Hudson, Drew A., and Christopher D. Manning. GQA: A New Dataset for Real-World Visual Reasoning and Compositional Question Answering. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

[3] Li, Wei, et al. "Ppdl: Predicate probability distribution based loss for unbiased scene graph generation." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022.

[4] Sinha, Abhishek, et al. "Negative data augmentation." arXiv preprint arXiv:2102.05113 (2021).

[5] Cong, Yuren, Michael Ying Yang, and Bodo Rosenhahn. "Reltr: Relation transformer for scene graph generation." IEEE Transactions on Pattern Analysis and Machine Intelligence 45.9 (2023): 11169-11183.

[6] Krishna, Ranjay, et al. "Visual genome: Connecting language and vision using crowdsourced dense image annotations." *International journal of computer vision* 123.1 (2017): 32-73.

SHAP 기반 Subpopulation 분석과 머신러닝을 활용한 OLED 증착 공정 PPA 편차 원인 인자 도출 방법론

양상식, 서동주
삼성디스플레이

sangsik.yang@samsung.com, dongjoo.seo@samsung.com

요 약

본 연구는 OLED 증착 공정에서 발생하는 PPA(Pixel Position Accuracy) 편차의 원인 인자를 체계적으로 도출하기 위해 XGBoost 와 LightGBM 기반 예측 모델을 구축하고 SHAP 을 활용한 분석 방법론을 제안하였다. 증착 공정의 FMM(Fine Metal Mask) ID, 온도, Tact, Glass, 증착기 인자 등 정형화된 테이블 데이터를 활용해 불량률 예측 모델을 학습시킨 후, SHAP Feature Importance 를 기반으로 Subpopulation 분석 및 공정 인자 그룹화를 수행하여 불량 원인을 도출하였다. 실제 양산 데이터를 활용한 결과, 상위 4% 순위 내에서 정확한 원인 인자 식별과 그룹별 평균 중요도에서 유의미한 증가를 확인함으로써 제안 방법론의 효과성을 검증하였다. 이는 엔지니어의 경험 의존적 분석에서 벗어나 데이터 기반의 신속하고 정확한 불량 원인 규명을 가능하게 한다.

Ⅰ. 서 론

본 연구는 OLED 증착 공정에서 발생하는 픽셀 위치 정확도(PPA) 편차의 주요 원인을 도출하기 위해 머신러닝 기반 분석 방법론을 제안하였다. OLED 제조 공정은 R/G/B 서브픽셀별로 정밀한 증착이 요구되며, FMM 형상, 온도, 증착 Tact Time 등 복합적인 설비 상태 변수가 불량 발생에 영향을 미친다. 기존에는 엔지니어의 경험에 의존하는 분석 방식으로 인해 분석 속도와 정확도가 개인 역량에 크게 좌우되는 한계가 존재하였으며, 이를 개선하기 위해 본 연구는 트리 기반 모델과 SHAP 을 결합한 접근법을 적용하여 데이터 기반의 체계적인 원인 분석을 수행하였다.

SHAP(SHapley Additive exPlanations)은 게임 이론의 샤플리 값(Shapley Value)을 기반으로 하여 머신러닝 모델의 각 입력 특성이 예측 결과에 기여한 개별 영향도를 수량화하는 모델 무관한 해석 기법이다. 주어진 예측 $f(x)$에 대해 각 특성 i 의 SHAP 값 ϕ_i 는 다음과 같이 정의된다.

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!\,(|N| - |S| - 1)!}{|N|!} \left(f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S) \right),$$

N 은 전체 특성 집합, S 는 특성 부분집합, f_s 는 S 에 속한 특성만을 사용한 모델의 기여도를 나타낸다. SHAP 은 샤플리 값의 공리인 효율성, 대칭성, 영향력 부재, 그리고 선형성을 만족하며, 각 특성의 기여를 더하여 기준값과의 차이로 예측값을 완전히 분해할 수 있는 장점이 있다. SHAP 은 모델 예측 결과의 특성 중요도를 수치화하여 해석 가능성을 확보하는 데 효과적이며, 제조업 및 반도체 공정에서 불량 예측 및 설비 진단에 널리 활용되어 왔다[1,2]. 특히 Causal SHAP 은 인과 관계를 반영한 보다 정확한 설명을 목표로 하나, 정확한 인과 그래프 구축의 어려움, 계산 복잡도 증가, 동적 변수 처리 한계 등의 문제로 산업 현장 적용에 제약이 있어[3], 본 연구는 이러한 한계를 고려하여 SHAP 기반 분석에 초점을 맞추며, 실제 증착 데이터를 활용해 제안 방법론의 효과성을 검증함으로써 데이터 기반 공정 최적화에 기여하고자 한다.

Ⅱ. 본론

증착 공정에서 생성되는 데이터의 수집, 전처리부터 모델학습 및 불량 원인 분석을 Figure 1 의 Flow 를 따라 진행하였다. 실시간 데이터 수집을 위해 Dataiku 시스템을 활용하여 최근 1 개월간의 증착 공정 내(In-Fab) 데이터를 원인 인자로, PPA 편차에 의한 불량률을 결과 인자로 수집하였다. 이후 XGBoost 및 LightGBM 모델을 적용해 불량률 예측 모델을 학습하고, 학습된 모델에 SHAP Tree Explainer 를 적용하여 SHAP Feature Importance 를 계산하였다. 불량률의 통계적으로 유의미한 차이를 기준으로 증착주기 및 FMM 교체주기 기반 Subpopulation 을 'REFERENCE' 그룹과 'COMPARE' 그룹으로 분류한 후, 원인 인자를 FMM ID, 온도, Tact Time, Glass, 증착기 인자로 구분하고, 각 인자별 평균 SHAP Feature Importance 를 산출하였다.

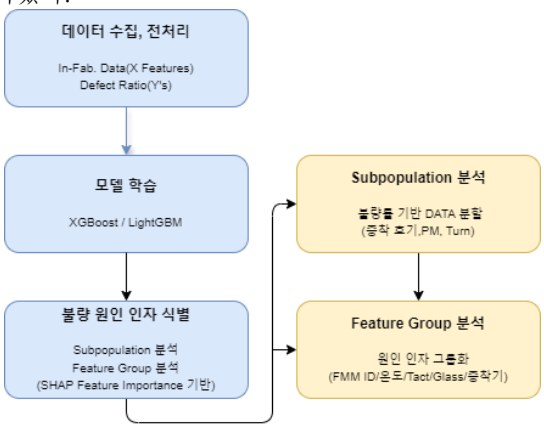


Figure 1. 불량 분석 Flow.

학습에 적용된 학습 인자는 Table 1 과 같다. 분석 방법론의 강건성을 확인하기 위하여 XGBoost 와 LightGBM 을 적용하였으며, Tree Size 와 정규화 강도를 Split 하여 모델의 Capacity 가 원인계 예측에 미치는 영향도 변화를 검증하였다. 그 외 인자는 모델 기본값을 적용하였다.

학습에 적용한 데이터는 실 양산 데이터를 적용하였으며 Subpopulation 은 특정 FMM 이 적용된 시점을 기준으로 설정하였다. Table 2 에 학습된 DATA 및 REFERENCE/COMPARE Subpopulation 및 원인계별 인자 수를 정리하였다.

XGBoost			
Parameter	reference	weak_reg.	strong_reg.
lambda (L2)	1	0.1	10
max_depth	6	8	4
LightGBM			
Parameter	reference	weak_reg.	strong_reg.
lambda_l2 (L2)	0	0	10
max_depth	6	8	4
num_leaves	31	90	10

Table 1. 모델 학습 Parameter

학습 DATA	
Number of features	학습 DATA Rows
123	3106
Subpopulation	
REFERENCE	COMPARE
171	149
인자 Group	
FMM ID	19
Tact	2
온도	35
증착기	22
모델	12
Glass	33

Table 2. 학습 DATA

Figure 2 와 3 에 각각 XGBoost/Strong Reg. 모델의 그룹별 평균, 상위 10 개 인자의 Feature Importance 를 나타내었다. 실제 불량은 COMPARE 구간의 Tact 변동 및 특정 ID 의 FMM 에 의하여 발생하였으며, Figure 2 에서 Tact 그룹의 Feature Importance 가 타 그룹 대비하여 큰 Importance 를 얻었으며, REFERENCE 대비하여 COMPARE 에서 두배 수준의 증가를 보여준다. 상위 중요도 기준 상위 10 개 인자 확인시 불량 발생을 유발한 FMM ID 인자와 Tact 인자가 1, 4 위로 나타남을 확인할 수 있었다. 이것은 해당 인자의 수준이 높은, 특정 ID FMM 이 반영되었거나 Glass to Glass 시간이 증가된 경우 불량률이 높은 현상을 반영한 결과이다.

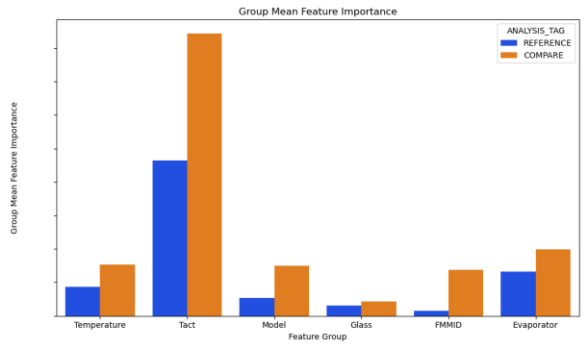


Figure 2. 그룹별 평균 Feature Importance

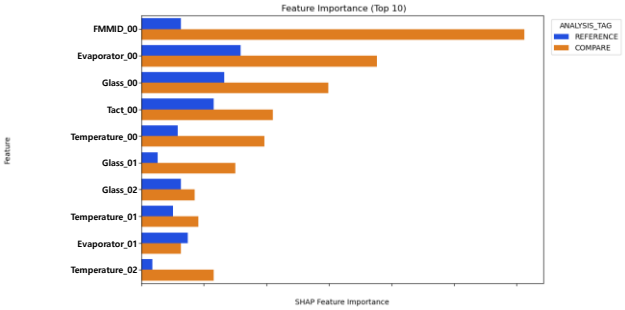


Figure 3. Top 10 Feature Importance

제안된 모델링 및 분석 방법론이 실 제품생산에서 적용될 경우 데이터의 Capacity 가 변화하는 상황에서도 안정적인 예측성능을 확보하는 것이 중요하며, 이를 검증하기 위하여 모델 Type/Capacity 에 따른 불량 인자 예측성능을 Table 3 에 나타내었다. XGBoost 의 경우 모델 Capacity 변화에 둔감한 성능을 보여주었으며, LightGBM 의 경우 XGBoost 대비하여 검출 성능이 떨어지는 사례가 발생하였다. 이는 LightGBM 이 특성간 상호작용을 학습하는 과정에서 노이즈를 과적합하여 나타났다고 추정되며, XGBoost 의 기본 parameter 가 보수적이고 안정적인 성능을 목표로 설계된 점에서 기인한다고 판단할 수 있다[4,5].

모델	조건	FMM ID 인자 순위	Tact 인자 순위
XGBoost	reference	1	5
	strong reg.	**1**	**4**
	weak reg.	1	4
LightGBM	reference	6	7
	strong reg.	1	5
	weak reg.	1	5

Table 3. 불량 원인 인자 순위

Ⅲ. 결론

본 연구는 OLED 증착 공정의 픽셀 위치 정확도(PPA) 편차 원인을 규명하기 위해 서브모집단 기반 피처 중요도 분석을 수행하였다. 트리 기반 모델(XGBoost, LightGBM)을 활용해 전체 데이터와 그룹별 특성 중요도를 비교한 결과, XGBoost 의 경우 상위 4%(Top 5 이내) 중요 변수 내에서 불량 원인을 효과적으로 식별했으며, 불량 그룹에서 해당 변수의 중요도가 유의미하게 증가함을 확인하였다. 이는 기존 경험적 분석 대비 데이터 기반의 객관적 근거를 제공하며, SHAP 을 통한 해석 가능성 확보로 제조 공정 최적화에 기여할 수 있음을 보여준다.

참 고 문 헌

[1] Lundberg et al. " A Unified Approach to Interpreting Model Predictions," Advances in Neural Information Processing Systems (NeurIPS), 30, 2017.

[2] Lee et al. “SHAP Value-Based Feature Importance Analysis for Short-Term Load Forecasting,” J. Electr. Eng. Technol. 18, pp579- 588 2023.

[3] Heskes et al. “Causal Shapley Values: Exploiting Causal Knowledge to Explain Individual Predictions of Complex Models,” NeurIPS, 33, 2020.

[4] Chen et al. “XGBoost: A Scalable Tree Boosting System,” KDD, 2016.

[5] Guolin Ke et al. “LightGBM: A Highly Efficient Gradient Boosting Decision Tree,” Advances in Neural Information Processing Systems (NeurIPS), 30, 2017.

BLE fingerprinting 을 활용한 DRSS-DNN 기반 실내 위치 추정에 관한 연구

공연정, 권예은, 진형준, 강주호, 김민희, 권재욱*, 김민석**
충남대학교, *와이파이브, **충남대학교

202550597@o.cnu.ac.kr, yeeuncnu27@o.cnu.ac.kr, wlsgudwns28@o.cnu.ac.kr, ang3495@o.cnu.ac.kr, oliviathree@o.cnu.ac.kr, ck1015@wifive.co.kr, kms4105@cnu.ac.kr**

A Study on Indoor Localization Using BLE Fingerprinting Based on DRSS-DNN

Gong Yeon Jeong, Kwon Ye Eun, Jin Hyoung Joon, Kang Ju Ho, Kim Min Hee, Gwon Jae Wook*, Kim Min Seuk**
Chungnam National Univ., *WiFive Co., Ltd., **Chungnam National Univ.

요 약

본 논문에서는 DRSS (Differential Received Signal Strength)를 딥러닝에 적용한 실내 위치 추정 기법을 제안한다. 무선 통신 표준 BLE (Bluetooth Low Energy)의 신호 세기인 RSSI 를 fingerprinting 하여 실내 위치 추정이 가능한 상황이다. 하지만 신호 불안정성과 장애물로 인한 multi-path 문제로 인해 정확한 위치 추정 성능을 얻기 힘들다. 따라서 본 논문에서는 다방향 BLE beacon 을 활용하여, 딥러닝의 전처리 과정에서 수신기의 안테나 간 RSSI 의 차이인 DRSS 를 입력으로 활용하는 기법을 제안한다. 이를 통해 DRSS 를 활용한 MLP (Multi-Layer Perceptron)의 입력 차원 확장이 BLE 기반 실내 위치 추정의 정확도 향상에 효과적이며 실용적이라는 것을 입증하였다.

Ⅰ. 서 론

실내 위치 추정은 스마트 빌딩, 물류 자동화, 자율주행 등 다양한 분야에서 필요한 기술로서 주목받고 있으며, 특히 GPS 신호가 도달하지 않는 실내 환경에서는 그 중요성이 대두되고 있다. 이를 위해 BLE (Bluetooth Low Energy)의 신호 세기인 RSSI (Received Signal Strength Indicator)를 기반으로 한 fingerprinting 방식은 구현이 간단하고 저비용이라는 장점을 가지므로 위치 추정 분야에서 활발히 연구되고 있다 [1].

해당 방식은 다수의 수신기가 송신기로부터 수신한 신호의 RSSI 값들을 특정 위치별로 fingerprinting 하고, 딥러닝을 활용해 2 차원 또는 3 차원 공간상의 좌표값이 매핑 될 수 있도록 학습한다. 이후 새로운 RSSI 값들을 입력으로 받아들였을 경우, 학습된 모델이 해당 위치의 좌표값을 예측하는 구조이다. 실제로 많은 딥러닝 기법 (Deep Neural Network, DNN; Convolutional Neural Network, CNN [2]; Recurrent Neural Network, RNN [3] 등)이 BLE-RSSI 기반 실내 위치 추정에 적용되었으나, RSSI 데이터 자체의 환경 변화에 대한 민감성과 다중경로 (multi-path) 및 비가시경로 (NLoS, Non-Line of Sight) 환경에서의 신호 왜곡 문제로 인해 위치 추정 정확도는 실내 기준 2~3 m 이상의 오차를 보이는 경우가 많다 [4].

이러한 문제를 해결하고자 우리는 다방향 BLE beacon [5]의 안테나 간 RSSI 차이인 DRSS (Differential RSS)를 입력으로 사용하는 전처리 기법을 도입한다. DRSS 는 RSSI 의 절대값 대신 수신 안테나 간 RSSI 의 상대적 차이를 사용함으로써, 환경 잡음 및 간섭에 대한 민감도를 줄여 보다 안정적인 특성 표현이 가능하다. 따라서 본 논문에서는 BLE-RSSI 에 대해 DRSS 를 추출하여 전처리를 수행한 뒤 해당 데이터를 딥러닝 모델의 입력으로 추가하여 학습함으로써, 실내 위치 추정의 정확도와 일관성을 향상시키는 방법을 제안한다.

Ⅱ. 본 론

가. DRSS 의 원리

본 논문에서 사용한 DRSS 는 하나의 다방향 BLE beacon 의 두 안테나에서 수신한 RSSI 간 차이를 기반으로 하며, 수신 전력 p_{rx} 및 전송 이득 g_{tx} 을 통해 DRSS δ_{ij}를 아래와 같이 정의한다 [6].

$$\delta_{ij} = p_{rx,j} - p_{rx,i} \approx g_{tx}(\tilde{\theta}_j) - g_{tx}(\tilde{\theta}_i) \quad \cdots \quad (1)$$

이를 제안 모델에 추가적인 입력 특성으로 활용하면, 왜곡 가능성이 존재하는 단일 RSSI 값에 대한 의존도를 낮출 수 있으며, 위치 추정의 정밀도가 향상될 수 있다.

그림 1 은 BLE 4 번 beacon 의 DRSS 분포를 시각화한 것으로, 측정 지점 변화에 따른 DRSS 의 공간적 분포 양상을 보여준다. 이는 곧 DRSS 가 위치 변화에 대해 높은 민감도를 가지고 있음과 동시에 공간적 정보를 내포하는 신호임을 나타내며, 딥러닝의 입력 특성으로 적합함을 시사한다.

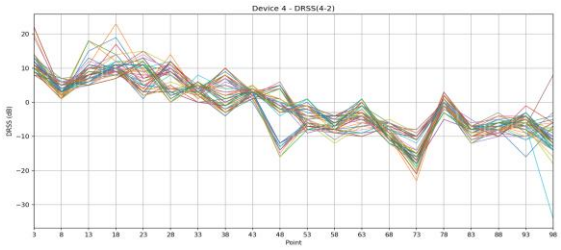


그림 1. BLE 4 번 beacon 의 DRSS 분포도

나. 데이터 구성

본 논문에서는 BLE 를 활용한 실내 위치 추정의 정확도 향상을 위해 고차원 입력 구조를 설계하고, 이를 DNN 모델에 학습시키는 구조를 제안한다. 실내 테스트의 경우 총 100 개의 포인트에서 실제 좌표와 더불어 포인트마다 4 개의 안테나가 장착된 8 대의 BLE beacon 으로부터 RSSI 데이터를 수집하였다. 이로써 RSSI 32 차원 (8 × 4)과 각 장비의 안테나 쌍 (4 – 2, 3–1)에서 계산한 DRSS 16 차원을 추가하여 총 48 차원의 입력 벡터를 구성하였다.

다. 딥러닝 구조

제안하는 딥러닝 구조는 다음과 같다. 해당 모델은 총 2 개의 은닉층으로 구성되어 있으며, 각 은닉층은 384 개의 뉴런과 선형 변환 (Linear Transformation), GELU (Gaussian Error Linear Unit) 활성화 함수, 그리고 과적합 방지 및 안정적인 특징 학습을 위한 Dropout 을 순차적으로 적용한 하나의 블록 형태로 설계하였다. 이렇게 구성된 두 개의 블록을 통해 입력 특징을 심층적으로 학습한 후, 최종 출력층에서 2 차원 (x, y) 좌표를 회귀 (Regression) 방식으로 예측한다. 모델 학습에는 AdamW 옵티마이저를 사용하였고, 손실 함수로는 SmoothL1 loss 를 적용하였다.

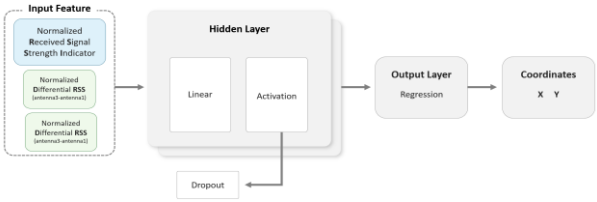


그림 2. 제안하는 딥러닝 구조

Ⅲ. 실험결과

학습된 모델은 총 14 개의 테스트 포인트에서 위치를 예측하며, 성능 평가는 예측 좌표와 실제 좌표 간 유클리드 거리 (Euclidean distance) 기반으로 수행되었다. 실내 환경에서의 실제 좌표는 cm 급 위치 추정 오차를 보이는 Ultra-WideBand (UWB) anchor 들과 tag 를 활용하였으며, 실제 좌표를 기준으로 전체 평균 오차, 중앙값 오차, 최대 오차를 산출하여 모델의 정밀도와 일관성을 확인하였다.

표 1 은 DRSS 적용 후 모든 평가 지표에서 성능이 향상되었음을 수치로 보여준다. 특히 평균 오차는 1.62 m 에서 1.43 m 로 약 11 % 개선되었고, 최대 오차 또한 4.83 m 에서 3.72 m 로 약 22 % 감소하여 안정성이 크게 향상되었다.

표 1. DRSS 적용 여부에 따른 성능 평가 결과

평가 항목	RSSI (m)	DRSS (m)
Mean Error	1.62	1.43 (- 0.19)
Median Error	1.58	1.35 (- 0.23)
Max Error	4.83	3.72 (- 1.11)

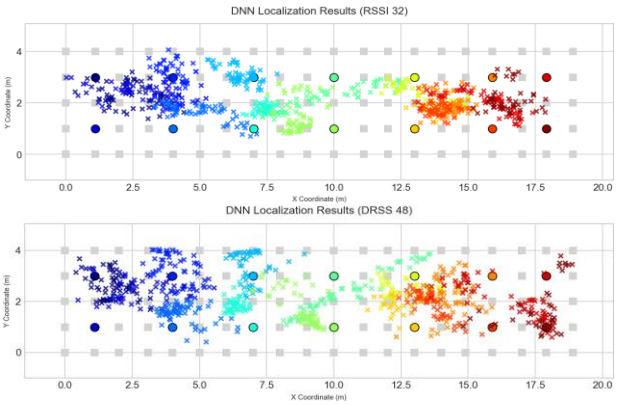


그림 3. DRSS 적용 전(상)과 적용 후(하)

그림 3 은 DRSS 적용 여부에 따른 위치 추정 결과를 비교한 것으로, DRSS 를 적용한 경우 예측된 위치가 실제 테스트 지점 주변에 더 밀집된 양상을 보인다. 이는 DRSS 가 위치 추정의 정확도를 향상시키는 데 기여함을 나타낸다.

Ⅳ. 결 론

본 논문은 안테나 간 RSSI 차를 활용한 딥러닝 기반의 실내 위치 추정 실험을 통해, DRSS 가 단일 RSSI 대비 정확도 향상에 유의미한 영향이 있음을 실증적으로 규명하였다. 본 연구는 다양한 측정 지점 및 환경에서 수집한 RSSI 와 DRSS 데이터를 통합하여 학습에 활용함으로써, 제안한 모델이 환경 변화에 대한 일반화 성능을 확보할 수 있음을 실험을 통해 입증하였다. 또한 기존 RSSI fingerprinting 기법의 상용화 한계를 극복할 수 있는 대안으로서 BLE 를 활용한 실내 위치 추정 기술의 성능 향상과 상용화 가능성 제고에 기여한다.

ACKNOWLEDGMENT

본 연구는 충남대학교 학술연구비에 의해 지원되었음.

참 고 문 헌

[1] R. Safwat, E. Shaaban, S. M. Al-Tabbakh, and K. Emara, “Fingerprint-based indoor positioning system using BLE: real deployment study,” *Bulletin of Electrical Engineering and Informatics,* pp. vol. 12, no. 1, pp. 523–532, 2023.

[2] R. S. Sinha and S.-H. Hwang, “Comparison of CNN Applications for RSSI-Based Fingerprint Indoor Localization,” *Electronics,* 2019.

[3] M. T. Hoang *et al.*, “Recurrent Neural Networks for Accurate RSSI Indoor Localization,” *IEEE Internet of Things Journal,* p. 10639–10651, 2019.

[4] L. Polak *et al.*, “Received Signal Strength Fingerprinting-Based Indoor Location Estimation Employing Machine Learning,” *Sensors,* p. 4605, 2021.

[5] 성태경, “실내위치 측정시스템 및 측정방법”. 대한민국 특허권 2616631, 18 12 2023.

[6] P. Park *et al.*, Multidirectional Differential RSS Technique for Indoor Vehicle Navigation,” *IEEE Internet of Things Journal,* p. 241–253, 2023

코바늘 객체 인식 기반 기호도 자동 생성 연구

오장훈, 이송희, 김문기, 전휘빈, 김유현, 유진희, 권민규*
성균관대학교, 건국대학교, 동국대학교, 상지대학교, 런던대학교, 안양대학교, *서울대학교

orichardev@g.skku.edu, sh713@konkuk.ac.kr, rlaansrlssp@dgu.ac.kr, 201574031@sangji.ac.kr,
yuhyeon.kim.20@alumni.ucl.ac.uk, 2020E7219@gs.anyang.ac.kr, *kmk0610@snu.ac.kr

Automatic Generation of Crochet Symbol Charts Based on Object Recognition

Oh Jang Hun, Lee Song Hee, Kim Moon Gi, Jeon Hwi Bin, Kim Yu Hyeon, Ryu Jin Hee, Kwon Min Kyu*
Sungkyunkwan Univ., Konkuk Univ., Dongguk Univ., Sangji Univ., University College London., Anyang Univ., *Seoul National Univ.

요 약

본 연구는 코바늘 직물의 스티치를 객체로 인식해 기호도(symbol chart)로 자동 변환하는 가능성을 검증하였다. 295 장의 이미지를 기반으로 7 개 클래스, 13,861 개 스티치 데이터셋을 구축하고, YOLO 와 Detectron2 모델을 비교하였다. 실험 결과 YOLOv11x 가 F1-score 0.869 로 최고 성능을 보였으며, 탐지 결과를 정렬·보정해 기호도 생성을 시범 적용하였다. 본 연구는 코바늘 기호도의 디지털 자동화를 통해 전통적 수작업 의존성을 줄이고, 패턴 공유 및 창작의 효율성을 높일 수 있는 확장 가능성을 제시한다.

Ⅰ. 서 론

코바늘 도안은 완성된 작품을 기호(symbol)로 해석하여 작성되며, 이는 작품 제작의 필수 과정이지만 높은 숙련도와 많은 시간이 요구된다. 현재 기호도의 대부분은 개인 간 거래나 소규모 공유에 의존하고 있어 접근성 및 확장성이 부족한 상황이다. 또한 코바늘 도안 제작 도구의 부재로 제작 과정은 과도한 창작자의 수작업에 의존해 비효율적이고 오류 가능성이 있다.[1]

기존 연구는 주로 대바늘 뜨개질(knitting)의 기하학적 모델링이었으며[2], 일부 코바늘(crochet) 패턴 자동화 연구[3]는 새로운 무늬 창작에 국한되어 주어진 작품을 표현하는 기호도 생성의 연구는 없다. 본 연구는 이러한 상황에서 코바늘 기법별 객체 인식을 기반으로 코바늘 기호도 자동 생성을 위한 연구를 제안한다.

Ⅱ. 본론

가. 데이터셋 구축

기존 공개 데이터셋(Crochet Stitch Detect CV Model, 72 장, 4 클래스)[4]은 해상도와 라벨 품질의 한계가 있어, 직접 Pinterest, 블로그, 상업 이미지 사이트 등에서 고해상도 이미지를 기반으로 총 295 장의 원본 데이터셋을 수집하였다. 각 이미지에는 평균 47 개 개별 코(stitch)가 포함되어 전체 13,861 개의 개별 코 데이터셋을 구축하였다. 개별 코의 종류는 기본적인 사슬뜨기, 한길긴뜨기, 짧은뜨기와 긴뜨기 및 두길긴뜨기는 앞/뒤 형태 차이를 반영하여 Figure1 과 같은 총 7 종류로 구분하였다.

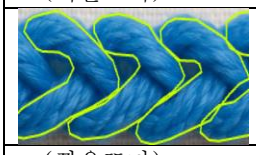
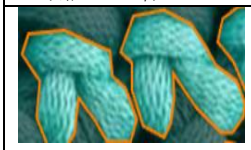
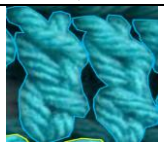
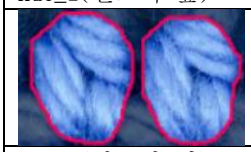
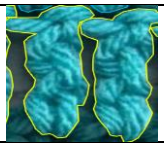
클래스	개수	클래스	개수
ch(사슬뜨기)		dc(한길긴뜨기)	
	2,725		2,333
sc(짧은뜨기)		tr_1(두길긴뜨기_앞)	
	3,612		1,342
hdc_1(긴뜨기-앞)		tr_2(두길긴뜨기_뒤)	
	1,458		1,215
hdc_2(긴뜨기_뒤)			
	1,176		

Figure 1. 코바늘 데이터셋 클래스별 객체 분포

나. 모델 선정 및 연구 프로세스 개요

본 연구는 코바늘 직물의 기호도 생성을 위해 4단계 절차를 구성하였다. 첫째, 객체 탐지 단계는 직물 이미지에서 스티치 단위를 인식하고 위치를 추출하는 과정이다. 둘째, 정렬 및 보정 단계는 탐지된 객체를 도안의 격자 구조에 맞게 배열하고, 위치 오류와 중복을 보정하는 과정이다. 셋째, 기호 매핑 단계는 정렬된 객체를 스티치 유형별 정의된 기호에 대응시키는 과정이다. 넷째, 도안 생성 단계는 변환된 기호를 체계적으로 배열하여 최종 기호도를 출력하는 과정이다.

이 중 객체 탐지 단계에서는 대표적 one-stage 계열인 YOLO와 two-stage 계열인 Detectron2를 비교 대상으로 선정하였으며, 두 모델이 코바늘 스티치 객체 탐지 및 기호도 변환 과정에 적용되는 적합성을 평가하였다.

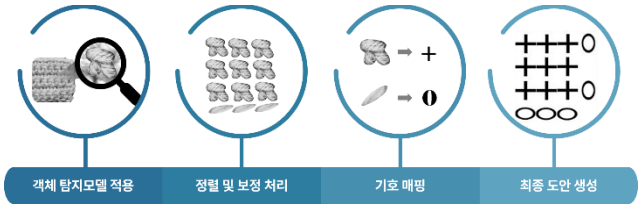


Figure 2. 연구 프로세스 개요 다이어그램

Ⅲ. 실험 및 결과

가. 평가 지표 및 선정 기준

본 연구에서는 제일 핵심적인 각 코바늘 스티치 검출과 관련된 모델평가는 F1-Score 주된 지표로 하였으며, 세밀한 성능 차이를 설명하기에 부족한 MAP50 은 보조 지표로만 사용하였다.

나. 실험 절차 및 결과

1. 모델 성능 비교

여러 객체 탐지 모델들로 자체적으로 구축한 데이터를 기반으로 실험한 결과, 가장 성능이 좋게 나온 YOLO, Detectron 모델을 본 연구의 주 모델로 선정하였다. 그 결과 Figure 3 에서와 같이 YOLOv11x 의 모델 기반으로 수집된 데이터 증강을 활용한 모델이 F1-Score 0.869 성능을 보였다. 특히 hdc_1, hdc_2, dc, sc 등 구별이 난해한 클래스에 있어서도 Precision 과 Recall 이 모두 0.8 이상의 성능이 나와 이를 코바늘 이미지에 대한 패턴 인식 최종 모델로 선정하였다.

순위	모델명	F1-score
1	YOLOv11x	0.869
2	YOLO11x-seg	0.852
3	YOLOv8	0.845
4	MaskRCNN_R50	0.703
5	FasterRCNN_R50	0.690

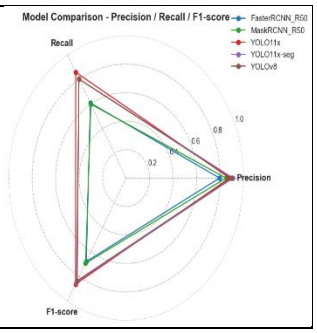


Figure 3. 성능 지표 그래프

2. 기호도 생성 시범 적용

탐지된 객체를 기호도로 변환하기 위해 먼저 스티치 좌표를 기반으로 층간 높이를 클러스터링하여 행 단위로 정렬하였다. 기호도는 규칙적 행·열 구조를 전제로 하므로, 실제 직물의 불규칙 배치를 보정할 필요가 있기 때문이다. 이후 누락된 스티치는 연속적인 패턴을 유지하기 위해 보간 처리로 채워 넣었으며 X 축 방향에서는 균등한 배열을 확보하기 위해 비거나 초과되는 개수를 보정하였다. 이렇게 보정된 스티치 행렬을 클래스별 기호(symbol)로 매핑하여 최종 기호도를 생성하였다(Figure 4).

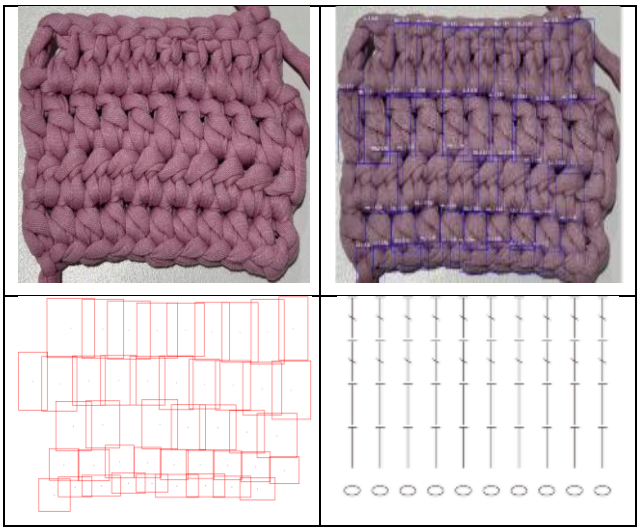


Figure 4. 기호도 변환 절차 (탐지-정렬-기호화)

Ⅳ. 결론

본 연구는 코바늘 직물의 스티치 단위를 객체로 인식하고 이를 기호도로 변환하는 자동화 가능성을 검증하였다. YOLO 와 Detectron2 계열 모델을 비교한 결과, YOLOv11x 가 가장 안정적인 성능을 보여 코바늘 직물의 세밀한 패턴 인식에 적합함을 확인하였다. 나아가 구축한 데이터셋과 탐지-정렬-기호화 과정은 단순한 성능 비교를 넘어, 코바늘 기호도의 디지털 보존과 자동 제작으로 확장될 수 있음을 시사한다. 향후에는 여러 스티치가 복합적인 직물을 기반으로 전체 도안을 자동 생성하는 시스템 구현으로 발전시킬 계획이다.

참 고 문 헌

[1] Seitz, K., Rein, P., Lincke, J., Hirschfeld, R., "Digital Crochet: Toward a Visual Language for Pattern Description", In Proc. Onward! 2022 (ACM), pp. 48-62, 2022.

[2] Gray, K., Bell, B., Sieper, D., Kobourov, S., Schreiber, F., Klein, K., Hong, S. "Wooly Graphs: A Mathematical Framework For Knitting," arXiv, (https://arxiv.org/abs/2407.00511), 2024.

[3] Martinez, M., Lipnicki, A. T. "Automating Crochet Patterns for Surfaces of Revolution," arXiv, (https://arxiv.org/abs/2302.02205), 2023.

[4] Roboflow, "Crochet Stitch Detect CV Model", https://universe.roboflow.com/akinorewek/crochet-stitch-detect(Accessed: 2025.08.16)

초심층 잔차 네트워크의 그래디언트 흐름 분석과 격자 기반 분화-병합 패러다임

박병준, 한동석
경북대학교 대학원 전자전기공학부
qudwns7171@knu.ac.kr, dshan@knu.ac.kr

Gradient Flow Analysis of Ultra-Deep Residual Networks and a Grid-Based Bifurcation-Merge Paradigm

Byeong-Jun Park, Dong Seog Han
School of Electronic and Electrical Engineering, Kyungpook National Univ.

요 약

초대규모 합성곱 신경망은 다양한 분야에서 성능을 크게 향상시켰으나, 잔차 연결 기반 구조는 깊이가 증가할수록 그래디언트 불안정성이 발생한다. 본 논문에서는 그래디언트 흐름의 관점에서 순차적 잔차 네트워크를 분석하고, 깊이에 따른 그래디언트 분산 감소와 장경로 지배 현상을 이론적으로 규명한다. 이를 극복하기 위해, 본 논문은 레이어와 블록을 일차원적으로 적층하는 대신 정사각 격자 기반 분화-병합 구조를 제안한다. 제안 방식은 입력을 다중 경로로 분화하고 병합 지점에서 평균 연산을 적용하여 균형 잡힌 그래디언트 전파를 유도한다. 실험 결과, 본 구조는 기존 순차형 네트워크 대비 더 안정적인 그래디언트 분포와 빠른 수렴 특성을 달성하며, 추가 파라미터 없이도 학습 효율과 정확도를 향상시킬 수 있음을 입증한다.

Ⅰ. 서 론

초대규모 합성곱 신경망(convolutional neural network, CNN)은 컴퓨터 비전, 자연어 처리, 멀티모달 학습 등 다양한 분야에서 성능을 크게 향상시켰다. 잔차 네트워크(ResNet [1])와 그 변형 구조는 항등 기반 지름길 연결을 통해 기울기 소실 문제를 완화하여 매우 깊은 모델의 학습을 가능하게 했으나, 수백~수천 레이어로 확장되는 초대규모 모델에서는 여전히 그래디언트 불안정성이 발생한다. 특히 순차적 잔차 구조에서는 깊이에 따라 그래디언트 분산이 누적되어 최적화 효율과 일반화 성능이 저하되는 한계가 보고되었다 [2].

이러한 문제를 완화하기 위해 Inception [3], ResNeXt [4]와 같은 다중 브랜치 구조가 제안되었으나, 이는 주로 특징 융합 효율에 초점을 두고 있어 그래디언트 전파 안정성을 직접적으로 해결하지는 못한다. 우리는 초대규모 네트워크에서 그래디언트가 분화(bifurcation)되고 병합(convergence)되는 지점이 불안정성이 핵심이라고 보고, 이를 구조적으로 재구성함으로써 내부 연산을 변경하지 않고도 그래디언트 안정성을 향상시킬 수 있다는 가설을 제시한다.

Ⅱ. 기울기 흐름 분석

딥러닝 네트워크를 방향성 비순환 그래프(directed acyclic graph, DAG)로 나타내면, 각 노드의 순전파와 역전파는 다음과 같이 정의된다.

$$\mathbf{x}_i = \mathcal{F}_i(\mathbf{z}_i), \quad \mathbf{z}_i = \sum_{j \in \mathcal{P}(i)} \mathbf{x}_j \cdot w_{j,i} \tag{1}$$

여기서 $F_i(\cdot)$는 해당 노드의 변환 연산(예: 합성곱 + 정규화 + 활성화 함수), $P(i)$는 노드 i의 부모 노드 집합, $w_{j,i}$는 연결 가중치(잔차 연결의 경우 보통 1)를 나타낸다.

역전파(backpropagation)에서, 손실 함수 L에 대한 노드 i 출력의 그래디언트는 다음과 같이 표현된다.

$$\mathbf{g}_i = \frac{\partial \mathcal{L}}{\partial \mathbf{x}_i} = \sum_{k \in \mathcal{C}(i)} \mathbf{g}_k \cdot \frac{\partial \mathbf{x}_k}{\partial \mathbf{x}_i} \tag{2}$$

여기서 $C(i)$는 노드 i의 자식 노드 집합이다.

순차형 잔차 네트워크에서는 대부분의 경우 $|P(i)| = 1$이며, 스킵 연결이 있을 때만 2가 된다. 이 경우 그래디언트 경로 수가 제한되는데, 분화 구간에서는 $|P(i)| > 1$이 되어 그래디언트 신호가 여러 경로로 분리되고, 병합 구간에서는 $|P(i)| = 2$가 되어 서로 다른 경로의 그래디언트가 합쳐진다. 병합 연산은 다음과 같이 정의된다.

$$\mathbf{x}_m = \frac{\mathbf{x}_{p_1} + \mathbf{x}_{p_2}}{2} \tag{3}$$

$$\mathbf{g}_{p_1} = \frac{1}{2}\mathbf{g}_m, \quad \mathbf{g}_{p_2} = \frac{1}{2}\mathbf{g}_m \tag{4}$$

여기서 p_1, p_2는 병합 노드 m의 두 부모 노드이다.

이 평균 연산 규칙은 두 부모 노드의 기여도를 동일하게 유지하도록 하여, 병합 지점에서의 그래디언트 분포를 균형 있게 만든다. 이론적으로, 분화-병합 사이클을 반복적으로 적용하면 깊이에 따른 그래디언트 분산의 급격한 감소를 완화하고, 순차형 구조에서 나타나는 장경로(long-path) 지배 현상을 줄일 수 있다.

깊이 d에서의 그래디언트 분산을 $\sigma_g^2(d)$라고 하면, 순차형 모델에서는 경험적·이론적 분석 모두 다음과 같은 경향을 보인다.

$$\sigma_g^2(d) \propto \alpha^d \tag{5}$$

여기서 $|\alpha| < 1$이면 기울기 소실 구간이 된다. 반면, 제안하는 분화-병합 패러다임에서는 다음과 같이 변화한다.

$$\sigma_g^2(d) \approx \beta^d + \frac{1-\beta^d}{2} \tag{6}$$

여기서 d가 커질수록 순차 구조보다 그래디언트 감쇠 속도가 느려지고 깊이에 따라 보다 균일한 그래디언트 에너지를 유지한다.

Ⅲ. 격자 기반 분화-병합 패러다임

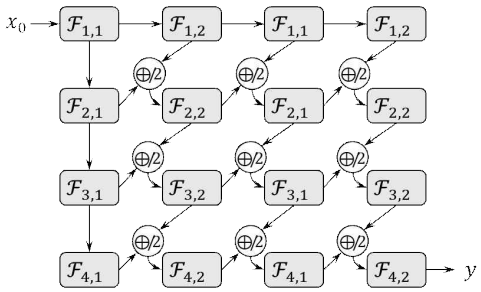


그림 1. 격자 구조의 분화-병합 패러다임

기존의 1차원 순차형 잔차 구조의 한계를 극복하기 위해, 본 논문에서는 레이어 또는 블록을 배치하는 데 있어 2차원 정사각 격자 토폴로지를 제안한다. 그림 1에 나타난 바와 같이, 네트워크는 $R \times R$ 격자로 구성되며 각 노드는 연산 블록 $F_{i,j}$에 해당한다. 데이터는 격자의 수평 및 수직 방향으로 흐르며, 이 과정에서 구조적 그래디언트 분화가 발생하고, 이후 그래디언트 병합이 이루어진다.

노드 연결 방식은 두 가지로 분류된다. 분화 노드는 하나의 입력 피처맵이 두 방향으로 복제되어 각각 다른 경로로 전달된다. 병합 노드는 두 개의 입력 피처맵이 다음과 같이 평균 연산을 거친 후 다음 블록으로 전달된다.

$$\mathbf{x}_{i,j} = \frac{\mathbf{x}_{i-1,j} + \mathbf{x}_{i,j-1}}{2} \tag{7}$$

이 평균 규칙은 두 입력의 기여도를 동일하게 유지하도록 하며, 역전파 과정에서는 다음과 같이 균형 잡힌 그래디언트 분포를 보장한다.

$$\mathbf{g}_{i-1,j} = \frac{1}{2}\mathbf{g}_{i,j}, \quad \mathbf{g}_{i,j-1} = \frac{1}{2}\mathbf{g}_{i,j} \tag{8}$$

그림 1의 격자는 하나의 구조적 모듈을 나타낸다. 이 모듈을 K번 반복하여 대규모 네트워크를 구성할 경우, 분화-병합 이벤트의 수는 격자 해상도 R의 증가에 따라 제곱 비율로 늘어난다. R이 커질수록 입력과 출력 사이의 평균 그래디언트 경로 길이가 다양해져, 특정 경로에 그래디언트가 집중되는 현상을 줄이고 다중 경로 기반의 그래디언트 안정성을 촉진한다.

Ⅳ. 실험

모든 실험은 Windows - Conda 환경에서 NVIDIA RTX 4090 GPU, CUDA 12.6, PyTorch 2.7을 기반으로 수행되었다. 재현성을 위해 랜덤 시드를 42로 고정하였다. 결과는 세 번의 독립 실행 평균으로 보고하였다.

실험에는 CIFAR-10 [5]데이터셋을 사용하였다. 모델은 SGD(Nesterov 모멘텀 0.9, weight decay 0.0004)를 사용해 300 에폭 동안 학습되었으며, 초기 학습률 0.1은 학습 스케줄의 50%와 75%에서 10배씩 감소시켰다. 손실 함수는 교차 엔트로피를 적용하였고, 배치 크기는 128로 설정하였다. 데이터 증강은 랜덤 크롭과 수평 반전을 사용하였다.

역전파 과정에서 발생하는 그래디언트 소실 현상을 정량적으로 평가하기 위해, 모델의 모든 학습 가능한 파라미터에 대해 그래이언트 크기의 분포를 분석하였다. 각 파라미터 텐서에 대해, 검증 데이터를 기반으로 손실 함수의 그래디언트를 계산하고, 이를 1차원으로 평탄화한 후 전체 모델의 그래디언트 값을 하나의 집합으로 결합하였다. 이 때 그래디언트의 절댓값이 미리 정의된 임계값 ϵ보다 작으면 해당 그래디언트를 '소실된(vanished)' 것으로 간주한다. 이에 따라, 그래디언트 소실 비율은 다음과 같이 정의된다.

$$\text{Vanished Ratio} = \frac{\#\{|\nabla_\theta \mathcal{L}| < \epsilon\}}{\#\{\nabla_\theta \mathcal{L} \text{ elements}\}} \tag{9}$$

표 1에서 확인할 수 있듯이, 제안 구조는 파라미터 수와 FLOPs에서 기존 순차형 네트워크와 동일한 수준을 유지한다. 그러나 표 1의 정확도와 표 2의 분포 비교 결과, 모델의 규모가 커질수록 제안 구조는 순차형 네트워크에 비해 분류 정확도가 높으며, 그래디언트 값이 소실 범위(ϵ) 내에 치우치지 않고 더 균형적으로 분포한다. 이는 분화-병합 지점이 그래디언트의 분산을 완화하며, 학습 안정성을 개선하는 데 기여함을 시사한다.

표 1. 제안 구조와 순차형 구조의 모델 크기, 연산량 및 정확도 비교

#layers	#Params.	#FLOPs			Acc (%)		
		VGGs	ResNets	Ours	VGGs	ResNets	Ours
26	0.37M	55.70M	56.01M	56.01M	91.34	**93.07**	92.68
98	1.54M	228.32M	228.63M	228.63M	64.56	94.03	**94.43**
386	6.20M	918.81M	919.11M	919.11M	10.00	93.55	**94.93**
1538	24.87M	3.68G	3.68G	3.68G	10.00	94.21	**95.09**

표 2. 초심층 네트워크에 대한 기울기 분포 및 소실도 비교

Model	Total gradients	Standard deviation	Vanished gradients ratio (%)			
			$\epsilon < 10^{-4}$	$\epsilon < 10^{-5}$	$\epsilon < 10^{-6}$	$\epsilon < 10^{-7}$
ResNet-1538	$\approx 2.49 \times 10^9$	2.60×10^{-3}	64.64	39.55	22.68	14.72
Ours (1538 layers)	$\approx 2.49 \times 10^9$	1.52×10^{-3}	40.10	16.45	9.31	6.51

Ⅴ. 결론

본 논문에서는 초심층 CNN의 순차형 잔차 아키텍처를 기울기 흐름 관점에서 분석하고, 다중 브랜치의 위상적 장점에서 영감을 받은 격자 기반 분화-병합 구조를 제안하였다. 제안 방식은 추가 파라미터나 연산 수정 없이 기존 연산 단위를 재배치하는 것만으로 그래디언트 분포를 균형 있게 유지하며, 초대규모 환경에서 학습 안정성과 효율을 동시에 향상시킴을 실험을 통해 입증하였다. 향후 연구에서는 분화-병합 구조의 효과를 극대화할 수 있는 새로운 설계 방식을 고안하고, 수학적 분석과 더불어 더 큰 데이터셋과 다양한 아키텍처로 확장성을 검증할 예정이다.

ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원-대학ICT연구센터(ITRC)의 지원을 받아 수행된 연구임 (IITP-2025-RS-2020-II201808).

참 고 문 헌

[1] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.

[2] C. Zhang, S. Bengio, and Y. Singer, "Are all layers created equal?" *Journal of Machine Learning Research*, vol. 23, no. 67, pp. 1-28, 2022.

[3] Szegedy, Christian, *et al.*, "Going deeper with convolutions." *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, 2025, pp. 1-9.

[4] S. Xie, R. Girchick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 5987-5995

[5] A. Krizhevsky, G. Hinton *et al.*, "Learning multiple layers of features from tiny images," 2007, accessed: 2025-08-15.

화재 진압용 드론 시스템을 위한 카메라-LiDAR 센서 융합 기반의 건물 유리창 감지 및 파손 위치 추정 기술 연구

이신우, 신수용*

국립금오공과대학교

nakgongs@kumoh.ac.kr, *wdragon@kumoh.ac.kr

Research on Camera-LiDAR Sensor Fusion-based Building Window Detection and Damage Location Estimation Technology for Fire Suppression Drone Systems

Shin Woo Lee, Soo Young Shin*

Kumoh National Institude of Technology

요 약

본 논문은 백드래프트 현상과 같은 위험을 회피하면서 고층 건물 화재 시 소화 활동을 지원하는 드론 시스템을 제안한다. 딥러닝 기반 이미지 처리와 LiDAR 센서의 융합을 통해 화재 현장의 창문 파손 상태를 실시간으로 정확하게 탐지하는 방법을 제시한다. 정밀한 카메라-LiDAR 캘리브레이션과 투영 기술을 활용하여 파손된 유리창의 3차원 위치를 정확히 파악하고, 이를 바탕으로 소화탄 발사 또는 드론의 안전한 내부 진입을 위한 최적의 지점을 제공한다. 이 시스템은 소방관의 위험을 최소화하고 화재 진압의 효율성을 향상시킬 것으로 기대된다.

Ⅰ. 서 론

그림 1. 런던 그린펠 타워 화재 현장 [1]

최근 30층 이상 고층 건축물의 급격한 증가로 고층 건물 화재의 위험성이 높아지고 있다. 통계에 따르면 2000년대 이후 화재 발생 건수가 지속적으로 증가하고 있으며, 특히 고층 건물 화재의 비중이 두드러지게 높아지고 있다. 런던 그렌펠 타워 화재와 같이 가연성 외장재로 인해 화재가 급속도로 확산되는 사례가 보고되면서 고층 건물 화재 진압의 어려움이 더욱 부각되고 있다 [1]. 기존 소방 장비의 명확한 한계로 인해 혁신적인 소화 시스템의 필요성이 대두되고 있다. 특히, 고층 건물 외벽의 유리창이 파손될 경우 공기 유입으로 인한 급격한 연소 확대, 즉 백드래프트(Backdraft) 현상이 발생하여 진압 활동에 심각한 위험을 초래할 수 있다. 이러한 문제를 해결하기 위해 본 연구는 카메라와 LiDAR를 융합한 센서 시스템을 탑재한 드론을 활용하여 화재 현장의 창문 파손 여부를 탐지하고, 효율적인 소화탄 발사를 위한 파손 위치를 추정하는 기술에 중점을 둔다. 이 기술은 실시간으로 화재 현장의 3차원 정보를 분석하여 드론의 자율적인 임무 수행을 가능하게 하며, 궁극적으로 소방관의 인명 피해를 최소화하고 화재 진압 성공률을 높이는 데 기여할 수 있다.

Ⅱ. 본론

본 논문에서는 제안하는 드론 시스템의 핵심 기술인 카메라-LiDAR 센서 융합 기술을 상세히 설명한다. 기존 연구에서는 카메라 이미지만으로 파손된 유리를 감지하려 시도했으나, 유리창의 투명한 특성 때문에 정확한 감지에 어려움을 겪었다 [2]. 본 연구는 이러한 한계를 극복하기 위해 딥러닝 기반 이미지 처리와 LiDAR 센서의 융합 방법을 제안한다. 먼저, 드론에 장착된 RGB 카메라로 건물 외벽 이미지를 획득하고, 딥러닝 모델을 사용하여 창틀로 추정되는 영역을 실시간으로 식별한다.

그림 2. 창틀 커스텀 트레인 결과

특히, 기존 객체 탐지 모델의 한계를 극복하고 유리창의 경계인 창틀을 더욱 정확하게 인식하기 위해 창틀 데이터셋을 추가하여 커스텀 트레이닝을 진행했다. 그림 2은 이러한 학습 결과의 시각적 예시이다.

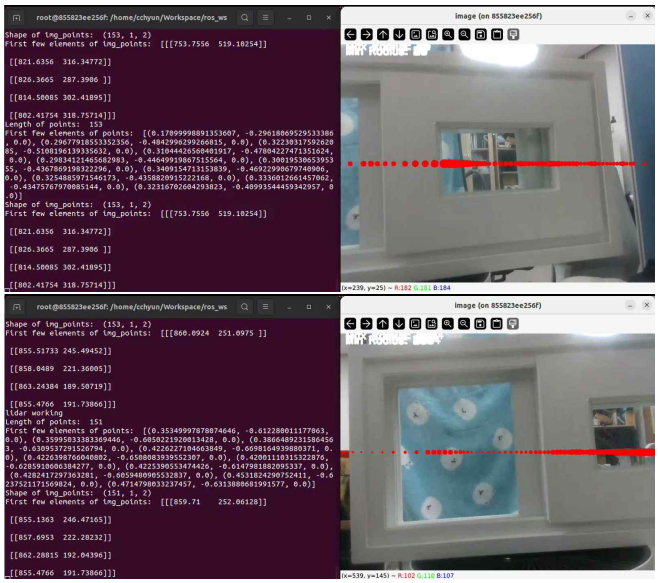


그림 3. 카메라 이미지-LiDAR 좌표 투영 결과
(상:유리가 있는 창틀, 하: 유리가 없는 창틀)

동시에, LiDAR 센서는 건물과의 정확한 거리 및 3차원 점군 데이터를 수집한다. 이 두 센서의 데이터를 효과적으로 결합하기 위해 정밀한 Calibration 과정을 거쳐 카메라 이미지의 픽셀 좌표와 LiDAR의 3차원 좌표를 일치시키는 실시간 투영 기술을 구현했다. 투영된 빨간 점은 LiDAR에서 감지한 거리 데이터를 바탕으로 거리가 가까울수록 점의 두께가 두꺼워지도록 설정했다. 그림 3에서는 2차원 이미지 상에서 감지된 유리창 영역에 대응하는 3차원 공간상의 위치를 정확히 파악하는 과정을 보여준다. 이러한 과정을 통해 화재 진압에 필요한 유리창의 파손 지점을 정확하게 계산하여 소화탄 발사 또는 드론의 건물 내부 진입을 위한 최적의 목표 지점을 제공하며, 이는 드론의 안정적인 위치 제어 및 효율적인 화재 진압 전략 수립에 필수적이다.

Ⅲ. 결론

본 논문에서는 화재 진압용 드론 시스템의 핵심 기술로 카메라와 LiDAR 센서의 융합을 통해 건물 유리창을 감지하고 파손 위치를 추정하는 기술을 제안하였다. 딥러닝 기반 이미지 처리와 LiDAR 데이터의 통합을 통해 화재 현장의 유리창 파손 상황을 식별할 수 있음을 보였다. 실험 결과, 카메라와 LiDAR 간 실시간 투영 기술이 성공적으로 구현되었으며, 이는 드론의 소화 임무 수행을 위한 중요한 토대를 마련했다. 향후 연구에서는 실제 화재 환경에서 시스템의 성능을 평가하고, 소화탄 발사 메커니즘과 결합된 통합 시스템 개발을 논의하고자 한다. 본 연구 결과는 고층 건물 화재 진압 기술의 발전에 의미 있는 이정표를 제시할 것으로 기대된다.

ACKNOWLEDGMENT

"이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원-지역지능화혁신인재양성사업의 지원을 받아 수행된 연구임 (IITP-2025-RS-2020-II201612, 50%)"

"본 연구는 과학기술정보통신부 및 정보통신기획평가원의 ICT혁신인재 4.0 사업의 연구결과로 수행되었음" (IITP-2025-RS-2022-00156394, 50%)

참 고 문 헌

[1] 한국소방안전원, "초고층건축물 + 다중밀집지역 = 대형화재의 불씨," 2017, (https://www.kfsi.or.kr/contents/webzine/201712/sub01-01.html).

Davies R. W." The Data Encryption standard in perspective,"Computer Security and the Data Encryption Standard, pp. 129-132.

[2] 이신우, 서정형, 김동현, & 신수용. (2024). 빌딩의 화재 진압을 위한 유리 감지 및 소화 드론에 관한 연구. 한국정보처리학회 학술대회논문집, 31(2), 896-897.

Asynchronous Reputation Management Framework For V2X Networks Under Intermittent Connectivity

Hope Leticia Nakayiza, Love Allen Chijioke Ahakonye †, Dong-Seong Kim, Jae Min Lee,
IT Convergence Engineering, † ICT Convergence Research Center,
Kumoh National Institute of Technology Gumi, South Korea
hopeleticia, loveahakonye, dskim, ljmpaul@kumoh.ac.kr

***Abstract*—Vehicular networks depend on continuous connectivity for reputation management, but disruptions compromise reliability. This paper proposes a decentralized V2X trust system that operates during offline periods using a hashed trust lock protocol to generate secure local trust logs and synchronize them with IPFS and blockchain upon reconnection. Experiments show an average processing latency of $1.2s$ per vehicle with reliable detection of replayed and conflicting logs prior to on-chain anchoring.**

***Index Terms*—Blockchain, IPFS, Offline communication, Trust and Reputation management, V2X**

I. Introduction

Vehicle-to-everything (V2X) networks enable communication among vehicles, infrastructure, and other road users [1]. They employ both IEEE 802.11p-based direct short-range communication (DSRC) and cellular V2X (C-V2X) technologies [2], supporting both direct exchange and broadband-assisted connectivity. These systems depend on roadside units (RSUs), base stations, cloud-edge infrastructure, and the internet for trust management and incentive mechanisms [3]. Yet, coverage gaps in rural areas, tunnels, or during disasters can disrupt connectivity [4]. Without effective fallback measures, such outages may be exploited by malicious actors, threatening system integrity [5].

This paper, therefore, introduces a system for continuous trust management that does not require constant internet or infrastructure access. Our key contributions include the Hashed Trust Lock Protocol (HTLP) for secure offline trust log generation and storage, a robust conflict detection and replay prevention mechanism, and the integration of the Interplanetary File System (IPFS) with blockchain for decentralized storage and reputation verification.

II. Methodology

Vehicles equipped with connectivity technologies such as C-V2X or Wi-Fi interact with RSUs and infrastructure to synchronize trust management data with IPFS and blockchain nodes. In the absence of internet connectivity, vehicles can communicate with each other through V2V links over DSRC. Figure 1 depicts the proposed offline reputation management architecture. The proposed HTLP ensures verifiable cryptographic commitments of trust logs during offline operations. A vehicle V_a observing the behavior of a peer vehicle V_p initiates the trust evaluation, logging the trust events locally

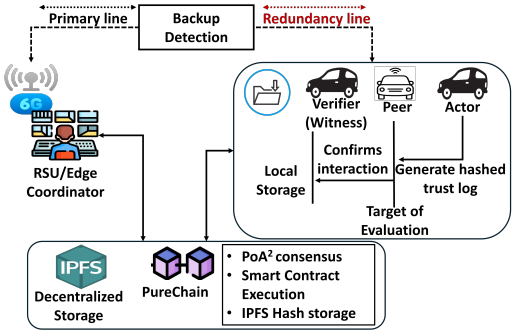


Fig. 1: HTLP-Based Offline Reputation Management Architecture

and creating a hash lock to commit to the trust assessment cryptographically. A witness vehicle, V_w, endorses this interaction by appending its signature to the log, thereby boosting its credibility. The trust log L is defined in Equation 1.

$$L = \{V_a, V_p, s_{ap}, h_i, t, \sigma_a, V_k, \sigma_k\}, \qquad h_i = H(n_i) \quad (1)$$

where s_{ap} is the trust score assigned by V_a to V_p, $H(n_i)$ is the SHA-256 hash of the secret nonce n_i and locks the log, t is the timestamp, σ_a and σ_k are the ECDSA signatures of V_a and V_w, respectively. Trust scores are updated incrementally, where $T_{ap}(t)$ at time t is given by Equation 2.

$$T_{ap}(t) = T_{ap}(t-1) + \Delta s_{ap}(t), \quad (2)$$

with $\Delta s_{ap}(t) \in [-0.3, +0.2]$, where negative values denote malicious behavior and positive values indicate cooperation.

A conflict detection algorithm ensures consistency by checking if the same vehicle logs contradictory trust scores for the same peer within a short time window Δt. Two logs, L_1 and L_2, are deemed conflicting as shown in Equation 3.

$$|s_1 - s_2| > \theta \quad \text{and} \quad |t_1 - t_2| \leq \Delta t, \quad (3)$$

where $\theta = 0.3$ and $\Delta t = 60$ seconds. To prevent replay attacks, which involve resubmitting old logs to manipulate reputation, each log is assigned a unique identifier ID_i as in Equation 4.

$$\mathrm{ID}_i = H\big(V_a \parallel V_p \parallel t \parallel H(n_i)\big), \quad (4)$$

Logs are flagged as replayed if their ID_i matches any previously observed identifiers stored in set $\mathcal{S}$.

Once connectivity is restored, vehicles upload filtered logs to IPFS, generating a content identifier (CID) that is recorded on the blockchain. The Edge Coordinator retrieves the CID from the blockchain, verifies the logs, and checks for duplicates or conflicts before updating trust scores. Replay attack protection is also enforced at the blockchain level, and after validation, the Edge Coordinator triggers reward distribution via a smart contract, advancing the reputation system.

III. Experimentation and Result Discussion

We validated the system using a simulated vehicular network of 50 vehicles, each generating 50 interactions with unique blockchain addresses and ECDSA key pairs. Vehicles recorded 1–6 offline events before uploading logs to a local IPFS node. A Solidity smart contract was deployed on a custom proof-of-authority and association blockchain [6]. Figure 2 shows part of a hashed trust log stored locally and uploaded to IPFS.


```
{
  "actor": "V1",
  "peer": "V7",
  "score": 0.2,
  "hash_lock": "b31baf04e267c3603cb88b5ae3d07128cc
  "timestamp": 1754970486,
  "signature": "c30dd249486498d2040e76962795bfc642
  "nonce": "1660c6ab7e1c4cd0af8f267c8f5cdc79",
  "verifier": "V10",
  "verifier_signature": "094e4f7a52c0e0dbe46e3ad76
},
```


Fig. 2: Hashed Trust Log Stored on IPFS

A. Computation Overhead Analysis

The computational costs, including IPFS upload latency, ECDSA signing/verification on vehicle onboard units, blockchain transaction latency, and total end-to-end latency from connectivity to on-chain confirmation, are shown in Figure 3.

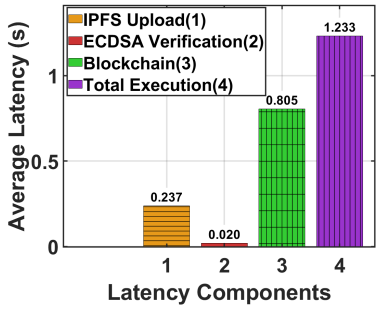


Fig. 3: Average Vehicle Communication System Latency Analysis

B. Security Analysis

We evaluated robustness against replay and conflict attacks during simulated interactions. Table I shows system totals, retained logs, and security filter removals. The smart contract prevents replay attacks by rejecting reused hash locks and duplicate IPFS CIDs, while an offline detector identifies conflicting entries. Invalid logs are filtered before IPFS storage, preventing malicious data from reaching distributed storage or blockchain, thus maintaining data integrity.

TABLE I: System-Wide Performance Summary.

	Initial logs	Filtered logs	Conflicts	Replays	Logs Removed
Total	60	44	6	10	16

IV. Conclusion

The proposed offline trust management for vehicular networks ensures continuity of trust during connectivity disruptions. HTLP, IPFS-based decentralized storage, and blockchain anchorage enable the system to securely maintain and verify trust logs without requiring continuous network access. Experimental results demonstrate its efficiency, with an average processing latency of $1.2s$ and strong security guarantees. Future work aims to develop a fully offline blockchain channel for reputation management.

Acknowledgment

This work was partly supported by Innovative Human Resource Development for Local Intellectualization program through the IITP grant funded by the Korean government(MSIT) (IITP-2025-RS-2020-II201612, 25%) and by Priority Research Center's Program through the NRF funded by the MEST(2018R1A6A1A03024003, 25%) and by the MSIT, Korea, under the ITRC support program(IITP-2025-RS-2024-00438430, 25%), and by the IITP(Institute of Information & Communications Technology Planning & Evaluation)-ICAN(ICT Challenge and Advanced Network of HRD) grant funded by the Korea government(Ministry of Science and ICT)(IITP-2025-RS-2022-00156394, 25%).

References

[1] H. L. Nakayiza, L. A. C. Ahakonye, D.-S. Kim, and J. M. Lee, "Trust-Aware Relay Selection for Message Delivery to Isolated Malicious Vehicles in V2X Networks," in *Proceedings of Symposium of The Korean Institute of Communications and Information Sciences Summer Conference (KICS SUMMER 2025)*, 06 2025.

[2] S. A. Yusuf, A. Khan, and R. Souissi, "Vehicle-to-Everything (V2X) in The Autonomous Vehicles Domain–A Technical Review of Communication, Sensor, and AI Technologies For Road User Safety," *Transportation Research Interdisciplinary Perspectives*, vol. 23, p. 100980, 2024.

[3] H. L. Nakayiza, L. A. C. Ahakonye, D.-S. Kim, and J.-M. Lee, "Blockchain-Enhanced Feature Engineered Data Falsification Detection in 6G In-Vehicle Networks," *IEEE Internet of Things Journal*, vol. 12, no. 15, pp. 30 036–30 048, 2025.

[4] R. Xing, Z. Su, T. H. Luan, Q. Xu, Y. Wang, and R. Li, "UAVs-Aided Delay-Tolerant Blockchain Secure Offline Transactions in Post-Disaster Vehicular Networks," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 11, pp. 12 030–12 043, 2022.

[5] C. Li, H. Jin, W. Wu, M. Yang, Q. Wang, and Y. Pei, "Path Loss and Auxiliary Communication Analysis of VANET in Tunnel Environments," *Symmetry*, vol. 15, no. 6, 2023.

[6] D.-S. Kim, I. S. Igboanusi, L. A. C. Ahakonye, and G. O. Anyanwu, "Proof-of-Authority-and-Association Consensus Algorithm for IoT Blockchain Networks," in *The 43rd IEEE International Conference on Consumer Electronics (ICCE 2025)*, 2025.

가우시안 혼합 모델을 활용한 방수 궤적의 확률적 추론

임건호, 이영준, 김승혁, *주백석
국립금오공과대학교
geon815@kumoh.ac.kr, joon2121008@naver.com, tmdgur3919@naver.com,
*bschu@kumoh.ac.kr

Probabilistic Inference of Water Jet Trajectories with Gaussian Mixture Model

Geonho Yim, Youngjun Lee, Seunghyeok Kim, Baeksuk Chu*
Kumoh National Institute of Technology

요 약

최근 화재 환경에서 발생하는 인적 및 물적 피해를 최소화하기 위하여 무인 소방 로봇 연구가 활발히 진행되고 있다. 소방 로봇을 활용한 효과적인 화재 진압을 위해서는 방수 궤적을 예측하고 궤적에 따른 최적 각도를 생성하는 것이 중요하다. 그러나 방수 궤적은 각도 이외에 노즐 구경, 바람, 압력 등의 물리적 요소의 영향을 크게 받아 각도 기반 단일 궤적 예측의 정확도가 제한된다. 특히 물 궤적의 경우 외란의 영향으로 동일 각도임에도 불구하고 분산된 데이터를 보이는 경향이 있다. 본 논문은 분산된 각도별 물 궤적을 GMM(Gaussian Mixture Model) 기반 확률 추론 모델을 생성하여 화재를 진압하는 방법을 제안한다. GMM은 다층 퍼셉트론 모델과 유사한 형태의 인공신경망으로 비선형 근사의 목표 값을 상수가 아닌 조건부 확률 밀도의 가우시안 분포로 설정한다. 일반 인공신경망 궤적 추론 모델은 데이터의 불안정성으로 인해 궤적 추론이 실패할 가능성이 존재한다. 확률 추론 모델은 하나의 화재 위치 입력에 화재 진압이 가능한 궤적의 분포를 제시하여 1차 궤적 추론이 실패하여도 2차 궤적 추론이 가능하다. 이는 무인 소방 로봇을 통한 자동 화재 진압에 도움이 될 것이라 기대한다.

Ⅰ. 서 론

최근 사람이 접근하기 어려운 화재 환경을 진압하기 위해 무인 소방 로봇에 대한 개발이 활발히 진행되고 있다. 무인 소방 로봇은 소방대원을 투입하기 어려운 환경에 투입되어 사전 진화 및 진화 경로 확보 등의 임무를 수행하는 것을 목적으로 개발되고 있다. 무인 소방 로봇은 원격 조종 및 자율 주행, 화재 인식 및 진화 등 다양한 동작을 수행할 수 있다. 특히 화재 진압 동작의 자동화는 소방대원이 진입하기 힘든 환경의 화재 진압을 가능하게 하여 활용도가 높다. 그러나 환경에 영향을 크게 받는 물 궤적의 특성상 화재 진압을 위한 추론 궤적이 실제와는 다른 양상을 보이는 경우가 다수 존재한다. 방수 궤적을 각도에 대해 단일 궤적으로 추론하는 경우 추론 궤적과 실제 궤적이 다른 경우에는 자동 화재 진압에 어려움을 겪게 된다.

위 문제를 해결하기 위해 확률론적 추론 모델을 활용하여 방수 궤적을 각도별 화재 진압 확률로 탐색하는 방법을 제안한다. 기존에 제시한 방수 궤적 추론 모델은 일반적으로 각도와 궤적이 일대일 대응하는 형태이다. 각도와 궤적이 일대일로 대응하는 모델은 입력 x에 따라 단일 출력 y를 추론하는 선형적 추론 모델이다. 이는 화재 진압을 위한 방수 궤적 생성 과정에서 하나의 화재 위치에 한 가지 궤적만을 추론 가능하다. 그러나 GMM(Gaussian Mixture Model)을 활용한 궤적 추론 모델의 경우 하나의 화재 위치에 다수의 궤적을 추론 가능하고, 궤적이 화재에 도달할 확률 또한 계산할 수 있다. 본 연구에서는 화재 진압을 목표로 방수 궤적을 추론하는 모델에 확률적 추론 모델을 적용하여 무인 소방 로봇을 활용한 자동 화재 진압 성능을 개선하고자 한다.

Ⅱ. 본 론

본 연구는 기존 방수 궤적 추론 모델인 일대일 대응 형태의 각도-궤적 선형 추론 모델의 단점을 개선하고자 GMM을 활용한 확률 각도-궤적 추론 모델을 제안한다. 제안하는 확률 추론 모델은 일반 인공신경망 기반 추론 모델과 다르게 한 개의 입력에 확률 기반 출력 분포를 가지는 모델이다.

1. 확률 추론 모델

본 연구는 GMM을 활용하여 입력 x와 출력 y의 전체 확률 분포 $p(y|x)$ 를 직접 모델링하는 것을 목표로 한다. GMM은 신경망을 통해 가우시안 분포의 가중치, 평균, 공분산을 예측하여 전체 확률 분포를 나타내며, 각도 θ에 따른 방수 궤적 $\boldsymbol{c}$를 예측하는 모델로 설정하였다. 표현 방법은 아래와 같다.

$$p(\boldsymbol{c}|\theta)=\sum_{k=1}^{K}\pi_k(\theta)N(y;\mu_k(\theta),\Sigma_k(\theta)) \quad (1)$$

식(1)의 K는 혼합 성분 개수, π_k는 k번째 성분의 혼합 가중치를 의미하며 μ_k와 Σ_k는 각각 가우시안 평균과 공분산을 의미한다. $N(y;\mu_k,\Sigma_k)$ 은 평균 μ, 공분산 Σ를 갖는 정규분포의 밀도함수를 뜻한다. GMM은 주어진 학습쌍 $\{(\theta_i,\boldsymbol{c}_i)\}_{i=1}^{N}$ 에 대하여 로그우도를 최대화하는 방식으로 학습을 수행한다. 이때, 로그우도의 최대화는 음의 로그우도(Negative Log Likelihood, NLL)를 최소화하는 것과 같다.

$$L=-\frac{1}{N}\sum_{i=1}^{N}\log[\sum_{k=1}^{K}\pi_k(\theta)N(c_i;\mu_k(\theta_i),\Sigma_k(\theta_i))] \quad (2)$$

식(2)의 N은 학습 샘플 수, $(\theta_i,\boldsymbol{c}_i)$ 는 각 샘플을 의미한다. $\Sigma_k\pi_k(\theta_i)N(\bullet)$ 은 각도 θ_i가 주어지는 경우 샘플 $\boldsymbol{c}_i$가 나올 확률 밀도 $p(\boldsymbol{c}_i|\theta_i)$ 를 혼합 모형

으로 계산한 것이다.

2. 데이터 수집 및 전처리

학습에 사용할 방수 궤적 데이터를 수집하기 위해 방수 실험을 진행하였다. 방수 압력은 6bar, 방수 노즐 구경은 65A, 방수 각도는 0°~35°의 조건에서 실험을 진행하였다. 방수 궤적 데이터를 활용해 GMM으로 학습한 전체 확률 분포 $p(\boldsymbol{c}|\theta)$ 는 각도 θ에서 가능한 방수 궤적 $\boldsymbol{c}$의 확률적 샘플 집합을 생성할 수 있다. 이를 활용하면 화재 위치에 따른 진압 가능 방수 궤적을 확률에 따라 다수 생성할 수 있다. 방수 궤적 $\boldsymbol{c}$는 각도별 실험 궤적을 $y=f(x)$ 형태의 르장드르 2차 기저로 커브 피팅을 진행한 후 계수 c_0, c_1, c_2 로 표현하였다. 식(3)은 이를 식으로 표현한 것이다.

$$y(x)=c_0P_0(x)+c_1P_1(x)+c_2P_2(x), \quad x\in[-1,1] \qquad (3)$$

그림 1은 데이터 전처리 과정을 이미지 형식으로 표시한 것이다. 실험을 통해 확보한 궤적 데이터를 GMM의 출력 벡터 $\boldsymbol{c}=[c_0, c_1, c_2]^T$ 로 설정하여 각도 θ와 궤적 계수 $\boldsymbol{c}$의 확률 분포 $p(\boldsymbol{c}|\theta)$ 를 학습한 모델을 출력하도록 하였다.

3. 모델 훈련 및 평가

모델 훈련을 위해서 각도별로 300개 내외의 르장드르 계수 데이터를 확보하였다. 표 1은 모델의 파라미터 조정 범위를 나타낸 것이다. 파라미터 K(가우시안 분포 개수), Cov(공분산 구조), Reg(공분산 수치 안정화 상수) 수치를 조정하며 진행하였고, 모델 최적 파라미터를 구할 수 있었다. 모델 파라미터 조정 결과 K는 9, cov는 full, reg는 5e-5에서 가장 좋은 학습 성능을 보였다. 최적 파라미터를 사용하여 모델 학습을 진행하였고, 각도에 따른 궤적 분포를 추론하는 모델을 생성하였다. 그림 2는 생성한 모델로 추론한 궤적을 이미지 위에 표시한 결과이다. 시각적으로 비교한 결과 실제 물 궤적과 유사한 분포를 보이는 것을 확인할 수 있었다.

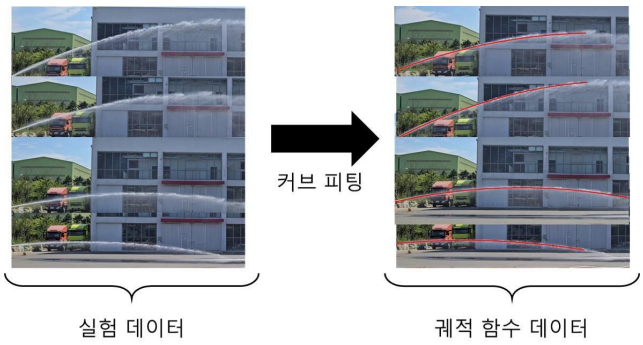


그림 1. 데이터 전처리

표 1. 모델 파라미터별 수치 범위

K	Cov	Reg
1 … 12	full, tied, diag	1e-5 … 1e-2

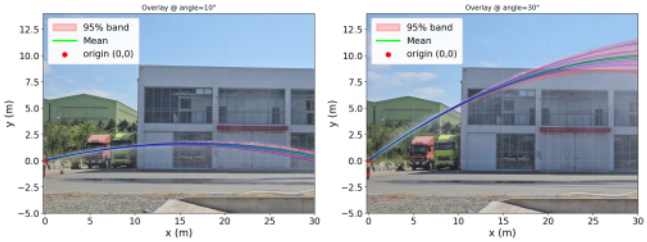


그림 2. 실제 궤적과 추론 궤적 비교

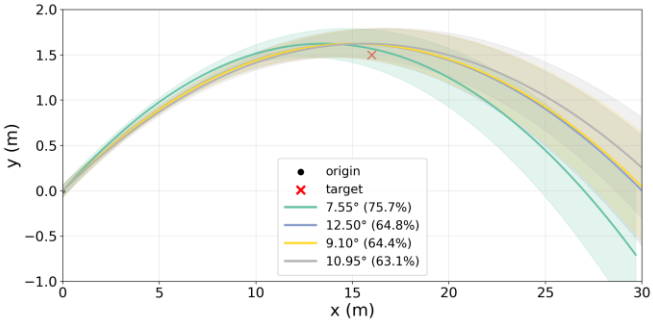


그림 3. 임의 좌표 도달 가능 궤적 추론

확률 추론 모델의 화재 진압 유용성을 확인하기 위해 임의의 좌표를 설정하고 해당 좌표를 지나는 궤적을 생성하는 각도를 찾는 시뮬레이션을 진행하였다. 생성한 추론 모델에 각도를 대입하고 출력 궤적 분포와 좌표에 도달하는 확률을 계산하였다. 그림 3은 시뮬레이션 결과를 그래프로 표시한 것이다. 좌표를 지나는 다수의 방수 궤적과 각도별 좌표 도달 확률을 확인할 수 있다.

Ⅲ. 결 론

본 연구에서는 무인 소방 로봇의 자동 화재 진압을 위해 각도-궤적 추론 모델을 확률 분포 $p(\boldsymbol{c}|\theta)$ 로 모델링하는 가우시안 혼합 기반 GMM 모델링을 제시한다. 각도에 따른 궤적의 분포를 학습함으로써 화재 위치에 따라 각도 별 화재 도달 확률을 계산할 수 있다. 확률 기반 궤적 추론을 통해 자동 화재 진압을 위한 최초 궤적 추론이 실패하여도 2차 및 n차 궤적 추론을 진행하여 차선책을 적용할 수 있다. 본 연구를 통해 무인 소방 로봇을 통한 자동 화재 진압 임무 수행 성능이 향상될 것이라 기대할 수 있다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신부 및 정보통신기획평가원의 ICT 혁신인재 4.0 사업의 연구결과로 수행되었음 (No.IITP-2025-RS-2022-00156394).

참 고 문 헌

[1] Lin, Y., et al., "Two-Stage Water Jet Landing Point Prediction Model for Intelligent Water Shooting Robot." Sensors, Volume 21, Issue 8, Article 2704. 2021.

[2] Liu, J., Cai, D., and He, X., "Gaussian Mixture Model with Local Consistency." Proceedings of the AAAI Conference on Artificial Intelligence, Volume 24, pp512-517. 2010.

[3] 이광국 외, "혼잡한 환경에 적합한 적응적인 배경모델링 방법." 멀티미디어학회논문지, Volume 11, Issue 5, pp597-609. 2008.

대규모 언어 모델에서 차별적 계층 식별 방법

곽지호, 정유철*
국립금오공과대학교
wlgh6709@kumoh.ac.kr, *jyc@kumoh.ac.kr

Discriminative layer Identification Methods in Large Language Models

Jiho Gwak, Yuchul Jung*
Kumoh National Institute of Technology

요 약

본 연구는 언어 모델의 레이어별 임베딩이 서로 다른 표현적 특성을 갖는다는 점에 착안하여, 다운스트림 과제에 최적화된 레이어를 선택하는 방법을 제안한다. 선행 연구에서는 중간 레이어가 최종 출력 레이어보다 의미적·구문적 정보를 풍부하게 담는 경향이 있음이 보고되었으나, 기존 접근은 레이어 선택 과정이 체계적으로 정의되지 않았다는 한계를 가진다. 이를 보완하기 위해 본 연구는 각 레이어의 판별력을 정량적으로 측정할 수 있는 평가 과정을 설계하고, 두 개의 텍스트 분류 데이터셋을 통해 그 효과를 검증하였다. 실험 결과, 제안된 방법은 항상 최종 레이어보다 더 높은 성능을 보이는 레이어를 사전에 선택할 수 있었으며, 이는 레이어 인지적 임베딩 선택이 모델의 표현 학습을 실질적으로 향상시킬 수 있음을 보였다.

Ⅰ. 서 론

대규모 언어 모델(LLM)의 임베딩은 다양한 자연어 처리 과제에서 널리 활용되며, 보편적으로 최종 레이어의 출력이 사용된다. 그러나 최근 연구들은 모델의 각 레이어가 상이한 표현적 특성을 가지며, 특정 과제에서는 최종 레이어보다 중간 레이어가 더 풍부한 의미적·구문적 정보를 담을 수 있음을 보여주었다[1]. 이는 레이어 선택이 다운스트림 성능에 직접적인 영향을 미칠 수 있음을 시사한다.

특히 모델 파라미터를 직접 업데이트하는 전이학습이 제한적이거나 불가능한 경우에도, 사전학습된 언어 모델의 임베딩만으로도 텍스트 분류와 같은 다운스트림 과제에서 충분히 높은 성능을 확보할 수 있다는 점에서 레이어별 특성에 대한 이해는 더욱 중요하다[2]. 그러나 임베딩의 표현력을 직접 분류 성능으로 평가하는 방식은 시간적 비용이 크고, 데이터셋마다 최적의 레이어가 달라질 수 있다는 한계를 가진다.

이에 본 연구는 레이어별 임베딩의 판별력을 정량적으로 분석하는 방법을 통해, 과제 최적화에 유리한 레이어를 사전에 식별할 수 있는 가능성을 탐구한다.

Ⅱ. 본론

가. 제안 기법

본 연구에서는 언어 모델 각 레이어에서 추출한 임베딩이 가지는 판별력을 정량적으로 평가하기 위하여 실루엣 계수(Silhouette coefficient)[3]를 사용하였다. 실루엣 계수는 동일 클래스 내에서의 응집도와 클래스 간의 분리도를 동시에 고려할 수 있는 지표로, 레이어별 표현의 질을 비교하는 데 적합하다.

특정 샘플 i에 대하여 $a(i)$는 샘플 i와 동일한 클래스에 속한 다른 샘플들과의 평균 거리(클래스 내부 거리)를 의미하며, $b(i)$는 샘플 i와 가장 가까운 다른 클래스에 속한 샘플들과의 평균 거리(최근접 클래스 간 거리)를 의미한다. 실루엣 계수 $s(i)$는 $-1 \le s(i) \le 1$ 범위를 가지며, 1에 가까울수록 샘플이 같은 클래스 내에서 잘 응집되어 있고 다른 클래스와 명확히 구분됨을 의미한다. 반대로 0값에 가까우면 경계에 위치한 샘플임을, 음수일 경우 잘못된 군집에 속했을 가능성을 나타낸다. 이때 각 개별에 대한 실루엣 계수는 다음과 같이 계산된다.

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad ...(1)$$

데이터셋 전체에 대한 평균 실루엣 계수(S)는 다음과 같이 정의된다.

$$S = \frac{1}{N}\sum_{i=1}^{N} s(i) \quad ...(2)$$

여기서 N은 전체 샘플 수를 의미한다. 이론적인 측면에서는 평균 실루엣 계수가 높을수록 분류 정확도가 향상될 가능성이 크지만, 임베딩의 형태만으로 분류 성능을 직접적으로 보장하기는 어렵다. 따라서 본 연구에서는 실루엣 계수의 절대값보다는 레이어별 계수의 형태와 상대적 변화율의 변곡 지점에 주목하여, 판별력이 높은 레이어를 식별하였다..

나. 실 험

본 연구의 실험 설정은 다음과 같다. 제안한 방법의 효과를 검증하기 위하여 두 개의 성격이 다른 텍스트 분류 데이터셋을 사용하였다. 첫 번째는 AG News[4]로, 4개의 클래스로 구성된 뉴스 기사 주제 분류 데이터셋이며, 정치, 경제, 과학·기술, 스포츠와 같은 다양한 도메인을 포함한다. 두 번째는 Emotion[5]으로, 6개의 클래스를 가진 문장 단위 감정 분류 데이터셋이며, joy, sadness, anger, fear, surprise, love 등의 감정 레이블을 포함한다. 두 데이터셋은 서로 다른 언어적 특성과 난이도를 가지므로, 제안된 방법이 다양한 조건에서 일관적으로 성능을 보일 수 있는지를 평가하는 데 적합하다.

모델은 두 가지 Decoder 기반 대규모 언어 모델을 실험에 활용하였다. 하나는 Qwen3-1.7B[6]로 총 26개의 레이어를 가진 중형 규모 모델이며, 비교적 깊은 구조를 통해 풍부한 계층적 표현을 제공한다. 다른 하나는 LLaMA-3.2-1B[7]로 총 16개의 레이어를 가지며, 경량 구조임에도 불구하고 최신 아키텍처의 강점을 활용할 수 있다. 서로 다른 크기와 레이어 수를 가진 두 모델을 병행하여 실험함으로써, 제안된 방법이 모델 규모나 구조에 관계없이 적용 가능함을 확인하기 위함이다.

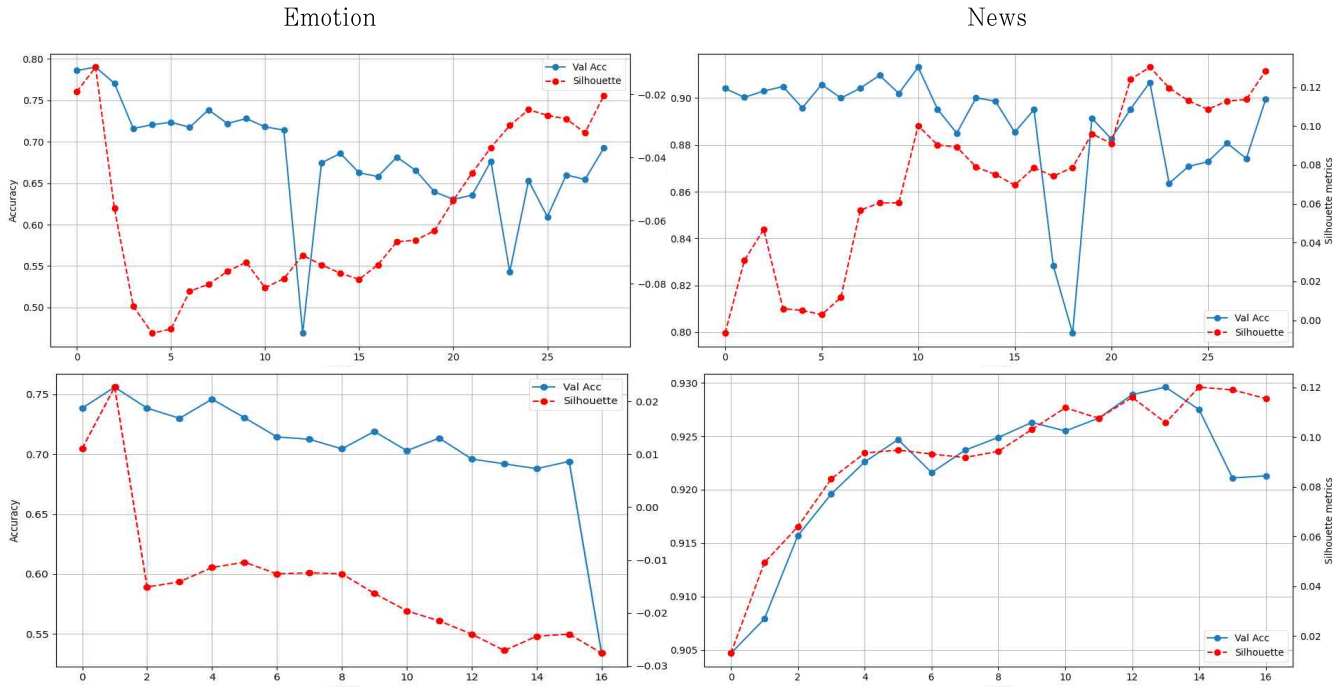


그림 1 각 모델에 대한 레이어별 정확도와 평균 실루엣 계수

실험 방법은 다음과 같다. 각 모델로부터 레이어별 임베딩을 추출한 뒤 평균 실루엣 계수를 산출하여 판별력을 비교하였다. 더 나아가 단순 Linear Layer를 활용한 분류 성능을 함께 평가하고, 실루엣 계수의 변곡점과 상대적 변화 추세를 관찰함으로써 사전에 최적의 레이어를 식별할 수 있는 가능성을 탐구하였다.

다. 결 과

그림 1은 각 모델과 데이터셋에 대해 레이어별 분류 정확도와 평균 실루엣 계수의 변화를 함께 나타낸 것이다. 상단의 두 그래프는 Qwen-3 1.7B 모델의 결과이며, 하단의 두 그래프는 LLaMA-3.2-1B 모델의 결과를 보여준다. 각 그래프는 좌측이 Emotion 데이터셋, 우측이 AG News 데이터셋을 대상으로 한 실험이다.

실험 결과에서 확인할 수 있는 주요 시사점은 두 가지이다. 첫째, 전반적으로 분류 정확도의 추세가 평균 실루엣 계수와 유사하게 나타났다. 이는 실루엣 계수가 단순한 군집 품질 지표를 넘어 실제 분류 성능과 일정 수준의 상관성을 가진다는 점을 보여준다. 둘째, 평균 실루엣 계수의 변곡점에 해당하는 구간에서 대부분의 경우 분류 정확도가 가장 높은 값을 기록하였다. 이는 단순히 최종 레이어에 의존하기보다는, 중간 레이어가 판별력 있는 표현을 담고 있음을 시사한다.

흥미로운 점은 두 모델과 두 데이터셋 모두에서 이러한 경향이 일관되게 나타났다는 것이다. 특히 감정 분류 태스크의 경우 두 모델 모두 극 초반 레이어에서 가장 높은 성능을 보였는데, 이는 짧고 직관적인 문장 단위 감정 분류 과제가 비교적 얕은 계층의 표현만으로도 효과적으로 수행될 수 있음을 보여준다. 반면 뉴스 주제 분류에서는 중간 레이어에서 뚜렷한 변곡점이 관찰되었으며, 이는 보다 복잡한 의미적 구분에 중간 수준 표현이 중요함을 나타낸다. 모델 규모(26-layer Qwen vs. 16-layer LLaMA)와 데이터셋 성격(뉴스 주제 분류 vs. 감정 분류)이 상이함에도 불구하고, 실루엣 계수의 상대적 변화 추세는 중요한 레이어를 식별하는 근거로 작동하였다. 이는 본 연구에서 제시한 분석 방법이 특정 모델이나 특정 데이터셋에 종속되지 않고, 다양한 환경에서 일반적으로 활용될 수 있음을 시사한다. 표 1은 모델, 데이터셋에 따른 평균 실루엣 계수와 정확도의 변화 추이와 변곡점을 나타낸 표이다.

	Dataset	변화 추이		변곡점(정확도)
		실루엣	정확도	
Qwen	Emotion	↘	↘	**1(0.7900)**, 24
	AGNews	→	↘	2, **10(0. 9132)**, 22
Llama	Emotion	↘	↘	**1(0.7560)**, 5
	AGNews	↗	↗	12(0.9289), 14

표 1 평균 실루엣 계수, 정확도 변화 추이와 변곡점

	Dataset	최대 정확도	및 레이어
Qwen	Emotion	0.7900	1
	AGNews	0.9132	10
Llama	Emotion	0.7560	1
	AGNews	0.9296	13

표 2 최대 정확도 및 레이어

Ⅲ. 결론

본 연구에서는 대규모 언어 모델의 레이어별 임베딩이 가지는 판별력을 정량적으로 분석하고, 이를 통해 중요한 레이어를 식별할 수 있는 가능성을 탐구하였다. 실루엣 계수를 활용하여 레이어별 응집도와 분리도를 평가하였으며, 단순 분류 성능과 비교한 결과, 실루엣 계수의 상대적 변화와 변곡점이 분류 정확도와 일정 수준의 상관성을 가지는 것을 확인하였다. 결론적으로, 본 연구는 레이어별 실루엣 계수 분석이 중요한 레이어를 사전에 판별하는 근거로 활용될 수 있음을 보여주었으며, 이는 fine-tuning이 제한적인 환경에서 임베딩 기반 학습을 위한 실질적 대안이 될 것으로 기대된다. 향후 연구에서는 보다 다양한 언어 모델과 태스크에 본 접근을 확장 적용하고, 단순 레이어 선택에 그치지 않고 Attention 메커니즘을 포함한 복합적인 평가 기법을 활용함으로써, 중요한 표현을 더욱 정교하게 식별할 수 있는 방법론을 모색할 예정이다. 이를 통해 본 연구의 접근법을 일반화된 레이어 판별 프레임워크로 발전시키는 것이 목표이다.

ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원 - 학·석사연계ICT핵심인재양성 지원을 받아 수행된 연구임 (IITP-2025-RS-2022-00156394)

참 고 문 헌

[1] Ju T., Sun W., Du W., Yuan X., Ren Z., Liu G. "How Large Language Models Encode Context Knowledge? A Layer-Wise Probing Study," arXiv preprint arXiv:2402.16061

[2] Skean O., Arefin M. R., Zhao D., Patel N., Naghiyev J., LeCun Y., Shwartz-Ziv R. "Revisiting Layer-Wise Information in Language Models," arXiv preprint arXiv:2502.02013.

[3] Rousseeuw, P.J. "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," Journal of Computational and Applied Mathematics, vol. 20, pp. 53 - 65, 1987.

[4] Zhang X., Zhao J., LeCun Y. "Character-level Convolutional Networks for Text Classification," Advances in neural information processing systems 28, pp. 649-657.

[5] Saravia E., Liu H. C. T., Huang Y. H., Wu J., Chen Y. S. "Carer: Contextualized affect representations for emotion recognition," Proceedings of the 2018 conference on empirical methods in natural language processing, pp. 3687-3697.

[6] Bai J., et al. "Qwen Technical Report," arXiv preprint arXiv:2309.16609.

[7] AI @ Meta. "The Llama 3 Herd of Models," arXiv preprint arXiv:2404.11225.

온디바이스 AI 기반 웨어러블 스토킹 방지 시스템

안제혁[1], 이재민[2], 김동성*

금오공과대학교 IT융복합공학과[1,2,*], 주식회사 엔에스랩 기술연구소*

{wpgur2018[1], ljmpaul[2], dskim*}@kumoh.ac.kr

Wearable Anti-Stalking System Based On-Device AI

Je-Hyeok Ahn[1], Jae-Min Lee[2], and Dong-Seong Kim*

Kumoh National Institute of Technology Dept. of IT Convergence Eng.[1,2,*]

NSLab Co., Ltd. Technology Research Institute*

요 약

본 논문은 웨어러블 카메라·센서로부터 수집한 데이터를 토대로 온디바이스(on-device)에서 추론을 수행하고 이를 모바일앱과 연동하여 스토킹 상황을 실시간 탐지·증거 수집하는 시스템을 제안한다. 제안 시스템은 후방 인체 탐지와 스켈레톤-실루엣 융합 기반 보행 패턴 분석으로 스토킹범 재식별, 데이터베이스의 등록 ID와 유사도가 임계값을 초과할 경우 증거 버퍼 저장 및 사용자에게 경고 및 신고를 연계한다. 실제 환경에서 사용 시, 이동식 영상정보처리기기 촬영으로 인해 초상권 등 권리가 침해될 가능성이 있다. 따라서, 원본 영상을 직접 전송하거나 활용하지 않고, 영상으로부터 추출된 비식별화 데이터인 스켈레톤과 실루엣만을 모바일 앱에 스트리밍한다. 모든 추론 과정은 이 비식별화 데이터를 기반으로 온디바이스에서 수행되므로, 프라이버시 침해 위험을 최소화하고 피해자 보호에 기여한다.

Ⅰ. 서 론

스토킹 행위는 목적과 유형이 다양하며, 행위자가 범죄 인식을 못 한 채 지속·반복하는 사례가 많다. 이에 따라 수사관의 적극적 개입으로 피해자가 증거를 신속·정확하게 수집·제출할 수 있도록 지원해 2차 피해 이전에 신속한 수사가 이루어져야 한다. 그러나 전담 인력이 제한적이어서 실시간 대응과 피해자 보호에는 구조적 한계가 있다. 가해자의 접근 사실을 입증하려면 피해자 본인이 증거를 확보해야 하는데, 이는 현실적으로 어렵다. 보완책으로 거주지 CCTV 설치, 스마트워치 지급 등이 시행되지만, CCTV는 촬영 범위·사각지대 탓에 이동 상황에 취약하고, 스마트워치는 GPS 오차와 수동 신고 의존 제약이 있다[1,2]. 기존 관련 연구로는 얼굴 인식 기술을 활용하는 경우가 많다. 하지만 실제 스토킹 상황에서 확보되는 영상은 대부분 저해상도에 흐릿하며, 가해자의 얼굴을 정면으로 선명하게 포착하는 것이 거의 불가능하여 명확한 한계가 존재한다. 이러한 이유로 사람 재식별(Person Re-ID) 기술은 얼굴이 잘 보이지 않는 상황에서도 신체나 걸음걸이 같은 단서로 동일인을 찾아내는 방향으로 발전해 왔다. 또한, 얼굴과 같은 민감 정보를 통해 학습·추론하는 접근은 프라이버시 위험과 규제 준수 부담을 수반한다[3].

본 논문은 이러한 문제를 해결하기 위해 온디바이스 AI 기반 웨어러블 스토킹 방지 시스템을 제안한다. 제안 시스템은 기존의 얼굴 인식 중심 접근이 가지는 한계를 보완하기 위해, 개인 식별성을 제거한 실루엣·스켈레톤 기반 데이터 송출과 온디바이스 추론 방식을 채택하였다. 이를 통해 저해상도·비정면 환경에서 발생하는 성능 저하 문제를 완화하는 동시에, 민감한 원본 영상 노출을 최소화하여 프라이버시 위험을 줄인다. 웨어러블 카메라의 후방 영상은 기기 내에서 비식별화되어 실시간 추론에 사용되고, 스토킹으로 판별되면 RTSP(Real Time Streaming Protocol)로 비식별 데이터만 앱에 전송되며 자동 신고·증거 수집이 연계된다. 이러한 시스템을 구축하여 피해자 보호에 기여할 수 있다.

Ⅱ. 딥러닝 기반 동일 인물 재식별 기술 분석

모델 / 기법	데이터셋	지표	성능
CentreFace	MegaFace	Rank-1 정확도	65.2%
CentreFace	SurvFace	Rank-1 정확도	29.9%
VDSR + CentreFace	SurvFace	AUC	35.1%
CentreFace (baseline)	SurvFace	AUC	37.3%

그림 1 기존 얼굴 인식 기반 동일 인물 재식별 기술 성능 비교표

사람 재식별(Person Re-ID)은 여러 카메라 뷰나 연속된 프레임 속에서 동일 인물이 사라졌다 다시 나타날 경우, 동일 인물을 식별하는 기술로, 감시·보안 분야에서 중요한 역할을 한다. 기존 Person Re-ID 연구는 주로 얼굴 인식에 기반해 발전했으며, CentreFace와 같은 모델은 정제된 대규모 고화질 얼굴 데이터셋을 학습해 얼굴 특징을 벡터로 변환하고 이를 비교하여 동일인 여부를 판별한다. 이러한 방식은 정면에 가까운 선명한 얼굴 이미지가 주어졌을 때 높은 성능을 보인다. 예를 들어, CentreFace는 MegaFace 벤치마크에서 65.2%의 Rank-1 정확도를 기록하였다. 그러나 실제 스토킹 범죄와 같은 비제어 환경에서는 성능이 유지되지 못한다. 실제 감시 환경을 반영한 SurvFace 데이터셋으로 검증한 결과, CentreFace의 Rank-1 식별률은 65.2%에서 29.9%로 급락해 54% 이상 저하되었으며, 이는 저해상도·모션 블러·비정면 포즈 등 열악한 이미지 품질 때문이었다. 이를 개선하기 위해 초해상도(Super-Resolution, SR) 기법이 도입되었으나, 대표 모델인 VDSR(Very Deep Super-Resolution) 적용 시에도 성능 지표 AUC는 37.3%에서 35.1%로 오히려 하락해 판별력 향상에 실패하였다. 즉, SR은 유용한 특징을 복원하지 못하고 오히려 노이즈를 생성해 성능을 저하시킬 수 있다[4]. 따라서 이러한 접근은 실제 감시·스토킹 환경의 저품질 영상에 적합하지 않으며, 얼굴 의존도를 줄이고 신체나 보행 패턴을 활용하는 사람 재식별(Person Re-ID) 방식이 요구된다[5]. 본 논문은 위와 같은 한계를 극복하기 위해 실루엣과 스켈레톤 기반 보행 패턴을 활용하는 Person Re-ID 접근을 제안한다.

Ⅲ. 온디바이스 AI 기반 웨어러블 스토킹 방지 시스템

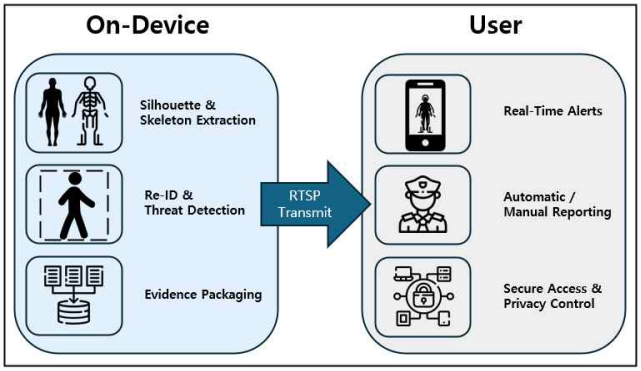


그림 2 온디바이스 AI 기반 웨어러블 스토킹 방지 시스템 구조

본 논문에서 제안하는 온디바이스 AI 기반 웨어러블 스토킹 방지 시스템 구조는 그림2와 같이 구성된다. 제안하는 시스템은 웨어러블 기기에서 영상을 비식별화하여 실루엣·스켈레톤 추출, 사람 재식별 및 위협 탐지, 증거 패키징을 수행하고, 결과를 RTSP(Real Time Streaming Protocol)를 통해 사용자 앱으로 전송한다. 사용자는 앱에서 실시간 알림을 받고, 필요 시 자동·수동 신고를 실행하며, 관리자 허가를 통해 원본 영상에 접근할 수 있다. 따라서 이러한 시스템 구조를 기반으로 스토킹 상황에서 피해자 근처를 지속적으로 따라오는 스토킹범의 외형 변화(마스크, 옷)나 저해상도에서도 강건하게 탐지하기 위해 실루엣·스켈레톤 기반 보행 패턴 분석을 통한 사람 재식별(Person Re-ID) 기법을 적용한다. 기존 얼굴인식 방식은 저해상도·비정면 환경에서 성능 저하가 크므로, 본 시스템은 웨어러블 카메라를 통해 취득한 후방 영상을 실루엣·스켈레톤 추출 및 보행 패턴 임베딩으로 변환해 개체를 구분한다. 시스템은 사전에 등록된 인물의 보행 특징과 실시간 입력 영상을 비교하며, 일정 임계치 이상의 유사도가 반복적으로 검출될 경우 스토킹으로 판단한다. 사용자가 웨어러블 기기를 착용하고, 이와 연결된 CSI 카메라를 통해 후방 영상을 실시간으로 촬영한다. 영상은 GStreamer 파이프라인으로 하드웨어 가속 디코딩되어 RGB 프레임으로 변환된 뒤 추론 스레드로 전달된다. 프레임은 전처리를 거쳐 배경 제거와 인체 분할을 통해 실루엣 마스크를 생성하고, OpenPose 기반 키포인트 검출로 17개 관절 좌표를 기반으로 스켈레톤 맵을 산출한다. 60프레임 입력을 45프레임으로 샘플링해 실루엣과 스켈레톤 시퀀스를 구성한 뒤 SkeletonGait++ 모델에 입력한다. 모델은 두 모달을 융합해 256차원 보행 임베딩 벡터를 생성한다. 마지막으로, 추출된 임베딩은 데이터베이스에 등록된 ID와 코사인 유사도를 비교하여 기준 임계값 0.8 이상을 만족할 경우 동일인으로 인식한다. 그림 3은 제안 시스템의 동작 결과를 예시적으로 나타낸 것이다. 좌측 상단 텍스트 영역에는 인식된 인물의 정보가 표시되며, 추론 결과 ID=1(안제혁)으로 인식되었고, 유사도는 89.1%, 버퍼 크기는 120/120 프레임, 스켈레톤 품질은 0.667로 기록되었다. 중앙 화면에서는 식별된 사람 영상 위에 스켈레톤 키포인트가 시각화되었으며, 우측 상단에는 해당 인체의 실루엣 마스크가 출력되었다. 이를 통해 제안 기법이 저해상도 환경에서도 안정적으로 보행 임베딩을 산출하고 동일 인물을 0.89 이상의 신뢰도로 정확히 재식별함을 확인할 수 있었다.

따라서 본 논문에서 제안한 기법은 저품질 영상 및 비정면 보행 상황에서도 동일인을 효과적으로 재식별할 수 있음을 보여준다. 특히 얼굴 기반 인식이 어려운 조건에서도 실루엣·스켈레톤 기반 보행 임베딩은 높은 강건성을 유지하며, 피해자의 후방 카메라에서 동일 인물이 마스크 착용이나 의복 변화와 같은 외형 변화를 보이더라도 보행 패턴을 통해 스토킹범을 조기에 탐지할 수 있음을 시사한다.

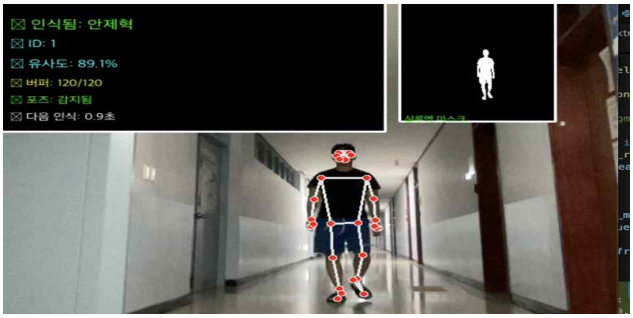


그림 3 SkeletonGait ++을 이용한 보행패턴 인식

Ⅳ. 결론

본 논문에서는 실제 환경에서 발생할 수 있는 스토킹 상황에 대응하기 위해 온디바이스 AI 기반 웨어러블 스토킹 방지 시스템을 제안하였다. 제안 기법은 실루엣·스켈레톤 기반 보행 패턴 재인식을 활용하여, 기존 얼굴 인식 방식이 가지는 저해상도·비정면 환경에서의 취약성을 효과적으로 극복하였다. 실험 결과, 동일인 재식별 과정에서 0.89 이상의 유사도를 달성함으로써 저해상도 영상 및 비정면 보행 상황에서도 높은 강건성을 입증하였다. 그러나 현재 추론 시간이 약 1.8초로 실시간성에는 제약이 존재하였다. 이에 따라 향후 연구에서는 정확도를 유지하면서도 경량화된 최적화 모델 설계 및 하드웨어 가속 기법을 적용하여 실시간 탐지가 가능한 시스템으로 발전시키는 것을 목표로 한다. 또한 이러한 시스템의 성능은 OpenGait 등 대규모 공개 데이터셋과 자체 수집 데이터를 병행하여 검증하고, Precision, Recall, ROC-AUC 등 다양한 성능 지표를 통해 다각적으로 평가할 계획이다.

ACKNOWLEDGMENT

본 논문은 한국 정부(MSIT)가 지원하는 IITP 지원금의 지역 지식화를 위한 혁신적 인적자원 개발 프로그램(IITP-2025-RS-2020-II201612, 25%)과 MEST가 지원하는 NRF의 우선연구센터 프로그램(2018R1A6A1A03024003, 25%), ITRC 지원 프로그램(IITP-2025-RS-2024-00438430, 25%) 및 한국 정부(과학기술정보통신부)의 지원을 받아 IITP(정보통신기술기획평가원)-ICAN(ICT Challenge and Advanced Network of HRD)의 지원을 받아 수행되었음(IITP-2025-RS-2022-00156394, 25%).

참 고 문 헌

[1] E.-A. Sun, “A Study on the Prevention and Blockade of Stalking Crimes under the Act on the Prevention of Stalking and Protection of Victims”, Public Land Law Review, vol. 109, pp. 335-358, Feb. 2025.

[2] H.-Y. Lee, “A study on stalking crime: Case analysis and victim protection measures”, M.S. thesis, Dept. of Police Administration, Soonchunhyang Univ., Asan, Republic of Korea, pp. 1-79, Feb. 2022.

[3] S. Bhardwaj, D.-S Kim, “Federated learning-based resource allocation for V2X communications”, IEEE Transactions on Intelligent Transportation Systems, vol. 26, pp. 382-396, Jan. 2025.

[4] Z. Cheng, X. Zhu, and S. Gong, “Face re-identification challenge: Are face recognition models good enough?”, Pattern Recognition, vol. 107, pp. 1-12, Nov. 2020.

[5] C. Fan, S. Hou, J. Liang, C. Shen, J. Ma, and D. Jin, “OpenGait: A Comprehensive Benchmark Study for Gait Recognition Towards Better Practicality”, IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1-18, Jun. 2025.

자연어 기반 자율이동을 위한 Dual-LLM 접근법

김인곤, 신수용*
국립금오공과대학교
20246104@kumoh.ac.kr, *wdragon@kumoh.ac.kr

Dual-LLM Approach to Natural Language-Guided Autonomous Navigation

In Gon Kim, Soo Young Shin*
Kumoh National Institute of Technology.

요 약

본 연구는 로봇의 자율이동 과정에서 자연어 명령을 효과적으로 처리하기 위해 Dual-LLM 기반 접근법을 제안한다. 로컬 환경에서는 Gemma-3 모델을 활용하여 사용자의 명령을 저지연으로 해석하고, 클라우드 환경에서는 GPT-4 mini 모델을 이용해 고차원 추론 및 보조적 의사결정을 수행한다. 이를 통해 기존 ROS2 내비게이션 대비 성공률과 안정성이 향상됨을 시뮬레이션을 통해 확인하였다.

Ⅰ. 서 론

최근 로봇의 자율이동(autonomous navigation)은 물류, 서비스, 재난 대응 등 다양한 응용 분야에서 핵심 기술로 자리잡고 있다. 기존의 ROS2 내비게이션 스택(Nav2)은 지도 기반 경로 계획과 비용맵 기반 주행 제어를 통해 높은 수준의 자율주행 기능을 제공한다[5]. 그러나 이러한 방식은 정확한 좌표 입력과 사전 파라미터 튜닝에 크게 의존하며, 사용자의 자연어 명령을 직접적으로 반영하기 어렵다는 한계를 가진다.

한편, 대규모 언어모델(LLM: Large Language Model)은 자연어 이해와 추론 능력을 기반으로 로봇과 사용자의 상호작용을 혁신할 수 있는 잠재력을 보여주고 있다[1][2]. 사용자가 “A 구역으로 가”와 같이 모호한 명령을 내렸을 때, LLM은 이를 구조화된 명령으로 변환하거나, 불확실한 경우 사용자에게 재질문을 생성하여 명확성을 확보할 수 있다. 하지만 단일 LLM만으로는 두 가지 문제가 존재한다. 경량 모델은 응답 속도는 빠르지만 복잡한 추론 능력이 부족하고, 대규모 모델은 추론 성능은 뛰어나지만 응답 지연(latency)과 연산 비용이 크다는 점이다[4].

본 연구에서는 이러한 한계를 극복하기 위해 Dual-LLM 기반 자율이동 접근법을 제안한다. 구체적으로, 로컬 환경에는 Gemma-3(4b-qat) 모델을 배치하여 저지연 자연어 명령 해석을 수행하고[3], 클라우드 환경에는 GPT-4mini 모델을 활용하여 고차원 추론과 보조적 의사결정을 담당하도록 설계한다. 이를 통해 사용자는 직관적인 자연어 명령을 통해 로봇을 제어할 수 있으며, 시스템은 상황에 따라 동적으로 두 모델을 적절히 활용하여 효율성과 안정성을 동시에 확보한다.

Ⅱ. 본론

제안하는 Dual-LLM 기반 자율이동 시스템은 로컬 모델(Gemma)과 클라우드 모델(GPT)이 상호 보완적으로 동작하도록 설계되었다. Gemma는 Edge 단에서 Candidate Generator 역할을 수행하며, 사용자의 자연어 명령을 빠르게 처리하여 잠재적 실행 후보를 생성한다. 여기에는 NLU(Natural Language Understanding), 맵 정보 활용, 비전 인식 결과를 종합한 행동 계획이 포함되며, ASK/ACT/WAIT와 같은 기본 정책을 통해 즉각적인 실행 가능 여부를 판정한다.

반면 GPT는 Cloud 단에서 Policy Selector로 동작하며, Gemma가 생성한 후보를 입력으로 받아 보다 정교한 의사결정을 수행한다. GPT는 MoA(Mixture of Agents), ToT(Tree of Thoughts), Reflexion과 같은 고차원 추론 전략을 적용할 수 있으며, 긴 문맥(Long-context)과 멀티 타스크(Multi-task) 추론을 통해 단일 후보 이상의 최적 정책을 선택한다. 이러한 구조는 Gemma가 경량 모델로서 빠른 반응성을 제공하고, GPT가 복잡한 상황에서 심층 추론을 담당하는 이중 계층 구조를 형성한다[2][3]. 두 모델 간의 통신은 비동기적 요청-응답 구조로 이루어지며, Gemma가 먼저 후보 실행안을 생성한 뒤 필요 시 GPT API에 전달하여 보정된 정책을 수신한다. 최종적으로 결정된 결과는 ROS2 Nav2 스택으로 전달되어 /navigate_to_pose 실행, 주행 파라미터 수정, 혹은 사용자 재질문과 같은 구체적 동작으로 이어진다[5].

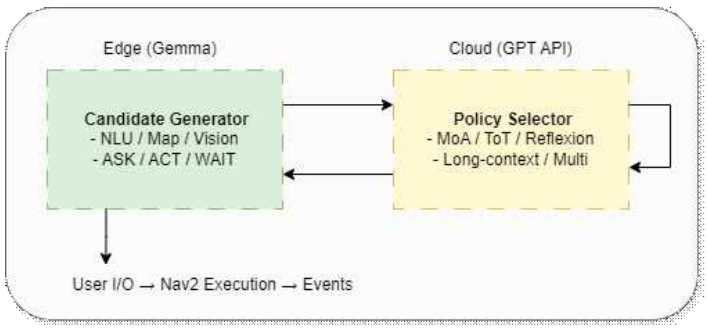


그림 1. Dual-LLM 통신 구조 (Gemma와 GPT 간 역할 분담 및 데이터 흐름)

그림 1. 은 제안하는 Dual-LLM 구조에서 Gemma와 GPT 간의 통신 및 역할 분담을 보여준다. 좌측의 Candidate Generator(Gemma)는 사용자 입력과 센서 데이터를 기반으로 실행 후보를 생성하고, 우측의 Policy Selector(GPT)는 이를 고차원적으로 평가하여 최종 정책을 결정한다. 이

러한 이중 구조는 단일 모델 대비 높은 신뢰성과 적응성을 제공하며, 특히 자연어 기반 자율이동에서 모호한 명령을 해석하거나 복잡한 주행 조건을 처리할 때 효과적이다.

Ⅲ. 실험 및 결과

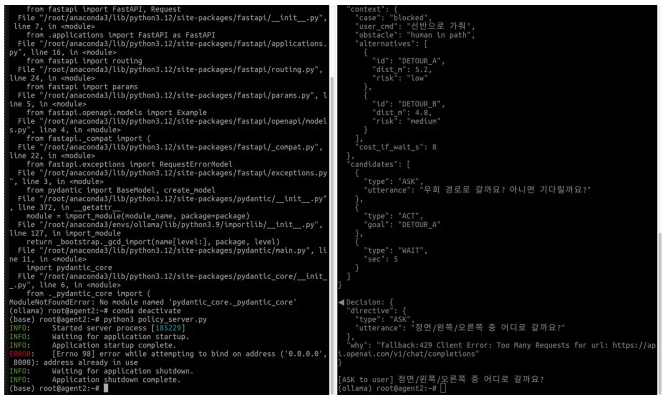


그림 2 Dual-LLM 시스템 실행 결과 로그 (좌: 서버 실행, 우: 후보안 생성 및 최종 질문)

제안한 Dual-LLM 기반 자율이동 시스템의 성능을 검증하기 위해 ROS2 Jazzy와 Gazebo Harmonic 환경에서 시뮬레이션 실험을 수행하였다. 실험 환경은 단순 창고 형태의 맵으로 구성하였으며, 사용자는 “선반으로 가줘”와 같은 자연어 명령을 입력하였다.

그림 2. 는 실제 실행 과정 중 시스템의 후보 행동 생성과 최종 의사결정 로그를 보여준다. 좌측은 서버 실행 및 FastAPI 기반 통신 과정, 우측은 “선반으로 가줘” 명령에 대해 탐지된 장애물(사람)을 인식한 뒤, **우회 경로(DET0UR_A, DETOUR_B) 및 대기(Wait)**의 후보안을 생성하고, 이들에 대해 “우회 경로로 갈까요? 아니면 기다릴까요?”라는 질문을 사용자에게 제안하는 모습을 나타낸다.

이와 같은 로그는 제안한 프레임워크가 단순히 좌표를 따라가는 것이 아니라, 자연어 기반 후보 생성(Gemma) + 정책 선택(GPT) 과정을 통해 상황에 맞는 행동을 도출함을 보여준다. 특히 Gemma는 ASK/ACT/WAIT 형태로 빠르게 후보안을 만들고, GPT는 이를 종합적으로 평가해 최종적으로 사용자에게 질문을 생성하거나 특정 경로를 선택하도록 한다.

Ⅳ. 결론

본 논문에서는 로봇의 자율이동 과정에서 자연어 명령을 처리하기 위해 Dual-LLM 기반 접근법을 제안하였다. 로컬 단에서는 Gemma-3 모델을 통해 저지연 명령 해석과 후보 행동 생성을 수행하고, 클라우드 단에서는 GPT-4mini를 통해 고차원 추론과 정책 선택을 담당하도록 설계하였다. 이를 통해 기존 ROS2 내비게이션 대비 명령 해석의 정확성과 주행 성공률을 향상시킬 수 있음을 ROS2 Jazzy + Gazebo Harmonic 환경에서의 실험을 통해 확인하였다.

제안된 구조는 단일 LLM 접근법의 한계를 극복하고, 실시간성(로컬)과 심층 추론(클라우드)을 결합한 새로운 형태의 자율이동 프레임워크로서 의의가 있다. 향후 연구에서는 멀티모달 센서 데이터를 통합하여 시각·언어 기반 의사결정을 확장하고, 실제 로봇 플랫폼에 적용하여 실증 실험을 진행할 예정이다.

ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원 - 학·석사연계ICT핵심인재양성 지원(IITP-2025-RS-2022-00156394, 50%)과 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원-지역지능화혁신인재양성사업의 지원을 받아 수행된 연구임 (IITP-2025-RS-2020-II201612, 50%)..

참 고 문 헌

[1] Vemprala, S., et al. "ChatGPT for Robotics: Design Principles and Model Abilities." arXiv preprint arXiv:2306.17582, 2023.

[2] OpenAI. "GPT-4 Technical Report." arXiv preprint arXiv:2303.08774, 2023.

[3] Google. "Gemma: Open Models for Responsible AI." arXiv preprint arXiv:2403.08295, 2024.

[4] Huang, W., et al. "Language Models as Zero-Shot Planners: Extracting Actionable Knowledge for Embodied Agents." ICML, 2022.

[5] Macenski, S., Foote, T., Gerkey, B., Lalancette, C., & Woodall, W. "Robot Operating System 2: Design, architecture, and uses in the wild." Science Robotics, 7(66), eabm6074, 2022.

경량화된 행동 인식 분류 알고리즘을 위한 시계열 데이터 변환 및 특징 추출 시스템 제안

박재한, 이만희, 신수용

초연결 기반 협력형 무인자율이동체 연구센터, IT융복합공학과

국립금오공과대학교

rimitpark@kumoh.ac.kr, fordmore@kumoh.ac.kr, wdragon@kumoh.ac.kr

Time-Series Data Transformation and Feature Extraction System for a Lightweight Action Recognition Algorithm

Jae Han Park, Lee Man Hee, Soo Young Shin

Super Connected and Collaborative Unmanned Autonomous Mobility Research Center, Department of IT Convergence Engineering

Kumoh National Institute of Technology

요 약

본 논문은 행동 인식 데이터 분류를 위한 CNN 기반 시간특성 분석 시스템을 제안한다. 제안된 시스템은 시계열 특성 분석을 위하여 시계열 데이터의 2차원 변환 알고리즘을 사용한다. 입력된 데이터는 2차원 이미지 데이터로 변환되어 CNN 입력으로 사용되며, 기존 RNN 혹은 Attention 알고리즘과 달리 고정된 연산량을 사용한다. 이런 변환 입력을 통하여 경량화된 행동 인식 데이터 분류를 시도한다.

Ⅰ. 서 론

최근 Teansformer 기반의 Attention 알고리즘을 사용한 다양한 비디오 분석 기법이 등장했다.[1][2] Transformer 기반 모델의 경우 입력 데이터를 시간 단위로 분리하여 Attention 알고리즘의 입력으로 사용하며, 모든 입력 요소 사이의 상관관계를 파악하기 때문에 매우 높은 분석 정확도를 보인다. 하지만 이런 특성은 입력 길이에 따라서 높은 연산량을 필요로 하는 문제가 있다. 비디오 데이터 기반의 행동 인식(Action Recognition) 분야 또한 많은 성능 향상이 이루었으나, 동일한 연산 효율 문제를 겪고 있다. 본 논문에서는 시계열 데이터를 GAF(Gramian Angular Field) 알고리즘을 사용하여 이미지화 하고, 경량화된 device에서도 사용 가능한 only-CNN(Convolution Neural Networks) 기반 행동 인식 데이터 분류 딥러닝 모델을 제안한다.

Ⅱ. 본론

그림 1은 본 논문의 시스템에서 사용된 데이터 변환 예시를 나타낸다. 비디오에서는 공간 특성을 포함하는 프레임 2개를 추출하고, t 축에 대하여 x변화와 y 변화를 확인할 수 있는 시간 평면 2개로 변환한다. 변환된 시간 평면은 시간의 진행에 따라 변화하는 x값 혹은 y값을 가지는 1차원 데이터로 변형 가능하며, 1차원 데이터를 기반으로 GAF 알고리즘을 사용하여 2차원 이미지로 변환한다. 기존의 시계열 변화를 분석하기 위해서는 비디오를 프레임 단위로 분리하여 CNN을 사용하여 프레임의 특성을 추출하고 그 정보들은 Attention이나 RNN(Recurrent Neural Network)계열의 모델 입력으로 사용하는 입력 길이에 따라 변화하는 구조이다. 본

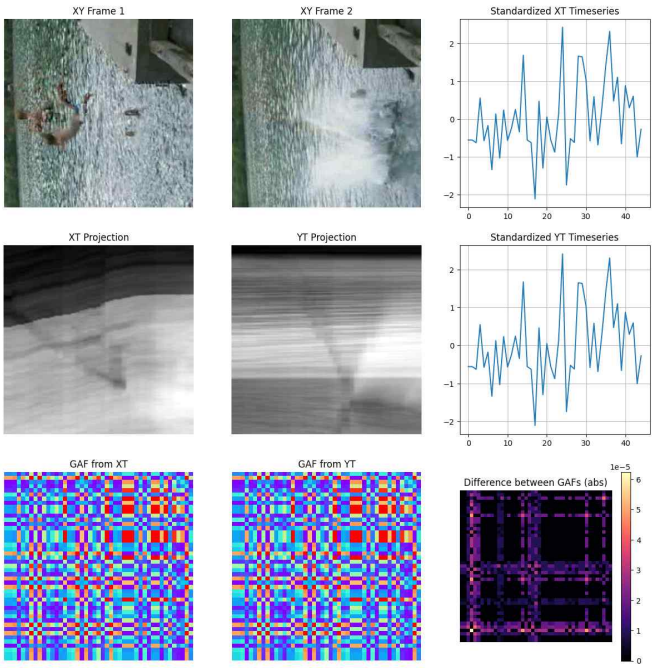


그림 1 비디오 데이터세트 HMDB51 다이빙 비디오 변환 예시

시스템에서는 비디오의 시간축에 따라 변화하는 x, y 데이터를 GAF 알고리즘의 입력으로 사용하여 2차원 이미지로 재구성한다. 이렇게 생성된 이미지는 데이터의 시간적 특성을 포함하고 있어, 공간적 특징 분석에 특화된 CNN 모델의 입력으로 사용 가능하다.[3][4][5] 또한, 시계열 데이터를

이미지로 변환하였기 때문에 기존의 다양한 이미지 증강(Augmentation) 기법을 그대로 적용할 수 있다는 장점이 있다.
이렇게 변환된 GAF 이미지(시간적 특징)는 원본 비디오 프레임(공간적 특징)에서 추출된 정보와 함께 CNN 모델의 입력으로 사용된다. CNN은 GAF 이미지의 고유한 텍스처와 패턴을 분석하여 원본 좌표 데이터의 동적 특성을 학습한다. 예를 들어, 주기적인 움직임은 GAF 이미지에서 반복적인 패턴으로 나타나며, 빠르고 급격한 움직임은 고주파(high-frequency) 텍스처로 표현된다. 모델은 이러한 시각적 단서들을 통해 움직임의 속도, 주기성, 방향 전환과 같은 복합적인 시간 정보를 포착한다. 이렇게 추출된 시간적 특징 벡터는 공간적 특징 벡터와 결합 되며, 모델은 이 통합된 정보를 바탕으로 비디오의 클래스를 최종적으로 판별한다.

Ⅲ. 결론

본 논문에서는 Attention 기반 모델의 높은 연산량 문제로 인해 행동 인식 시스템을 저사양 기기에 구현하기 어려운 점을 해결하고자 하였다. 이를 위해 비디오의 시간 정보를 담고 있는 1차원 데이터를 GAF 알고리즘을 사용하여 2차원 이미지로 재구성하고, 대표 프레임이 내포한 공간 정보와 함께 CNN 기반 알고리즘으로 처리하는 경량화된 행동 인식 모델을 제안하였다.

제안된 시스템은 비디오의 모든 프레임을 처리하지 않고도 GAF 이미지를 통해 움직임, 속도, 주기성과 같은 동적 특징을 효과적으로 표현하는 방법을 시도하였다. 이를 통해 기존 CNN 기반 이미지 분석 기술의 장점을 사용하여 전체적인 모델 경량화를 달성하며, 모바일, 임베디드 시스템 등의 경량 환경에서 비디오 분석 및 분류를 시도하였다.

하지만, 본 연구는 단일 객체나 특정 지점의 좌표를 추적하는 것을 전제로 하며, x, y의 변화량이 압축된 1차원 데이터를 사용하므로 여러 객체가 복잡하게 상호작용하는 영상에서는 특징 추출에 한계가 있을 수 있다. 현재는 다중 객체의 움직임과 x,y의 변화량을 압축하지 않은 시계열 데이터를 GAF나 다른 이미지 인코딩 기법으로 시각화하여, 더 복잡하고 섬세한 행동을 인식하는 모델로 확장하는 연구를 진행 중이다.

ACKNOWLEDGMENT

"본 연구는 과학기술정보통신부 및 정보통신기획평가원의 ICT혁신인재 4.0 사업의 연구결과로 수행되었음" (IITP-2025-RS-2022-00156394, 50%)

"본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터사업의 연구결과로 수행되었음" (IITP-2025-RS-2024-00437190, 50%)

참 고 문 헌

[1] Han, Kai, et al. "A survey on vision transformer." IEEE transactions on pattern analysis and machine intelligence 45.1 (2022): 87-110.

[2] Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems 30 (2017).

[3] He, Kaiming, et al. "Identity mappings in deep residual networks." Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14. Springer International Publishing, 2016.

[4] Huang, Gao, et al. "Densely connected convolutional networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[5] Tan, Mingxing, and Quoc Le. "Efficientnet: Rethinking model scaling for convolutional neural networks." International conference on machine learning. PMLR, 2019.

Transformer의 Feed-Forward Network 가속화 연구

이창근, 신새빈, 임완수*
성균관대학교 전자전기컴퓨터공학과
*wansu.lim@skku.edu

A Study on Acceleration of the Transformer Feed-Forward Network

Changgeun Lee, Saebin Shin, Wansu Lim
Sungkyunkwan University

요 약

Transformer는 self-attention 기반의 인코더 - 디코더 구조를 사용하여 높은 성능을 보이나, 많은 가중치로 인해 모델 크기가 크고 메모리 사용량이 높아 메모리 자원이 제한된 엣지 디바이스에서 동작이 제약된다. 특히 Feed-Forward Network 레이어는 대규모 행렬 곱 연산과 높은 가중치 비중으로 인해 연산 및 메모리 병목의 주요 원인이 된다. 본 논문은 4비트 대칭 양자화와 니블 패킹 기법을 적용하여 FFN 가중치를 압축하고, 이를 리소스가 제한된 Kria KV260 FPGA의 온칩 메모리에 모두 맵핑하는 방법을 제안한다. Vitis HLS 환경에서 합성한 결과, BRAM 사용량이 3,813개에서 242개로 감소(93.7% 절감)하여 BRAM 한도 내 온칩 맵핑을 달성하였다. 이를 통해 외부 DRAM 접근으로 인한 메모리 병목을 효과적으로 완화하였다.

Ⅰ. 서 론

Transformer는 self-attention기법을 사용하는 인코더와 디코더로 구성된 기계 번역 모델이다[1]. Transformer는 높은 성능을 지니나, 가중치 수가 많아져 모델 크기가 크다. 이는 모델 추론 속도를 저하시킬 뿐만 아니라 메모리 자원이 제한된 엣지 보드에서의 동작을 제약한다. 특히 단어 토큰 임베딩 차원이 큰 경우, 대규모 행렬 곱 연산을 수행하는 Feed-Forward Network(FFN) 레이어가 주요 병목으로 작용한다. 또한 FFN의 가중치는 Transformer 전체 가중치에서 큰 비중을 차지해 메모리 사용량을 크게 증가시킨다. FFN의 연산 병목은 FPGA를 통해 가속할 수 있으나[2], 리소스가 제한된 FPGA 보드에서는 모든 FFN 가중치를 온칩(BRAM/URAM)에 맵핑 시킬 수 없어 메모리 접근 병목(memory bottleneck)이 발생한다.

본 논문에서는 FFN 가중치에 양자화 기법과 니블 패킹을 적용하여 리소스가 제한된 Kria KV260 FPGA의 온칩 메모리에 모든 FFN 가중치를 맵핑 시킴으로써, 메모리 접근 병목을 최소화하는 Transformer 가속화를 제안한다.

Ⅱ. 제안한 가속화 알고리즘

그림1은 FPGA를 통한 FFN 가속화에 대한 전체적인 시스템을 나타낸다. 본 논문에서는 학습이 완료된 FFN 레이어의 모든 가중치들을 FPGA의 온칩 메모리에 효율적으로 맵핑하기 위해, 4비트 대칭 양자화와 니블 패킹을 적용한다. 가중치는 고정된 스케일과 오프셋으로 압축되며 니블 단위로 연속 저장된다. 또한 FFN의 연산 경로는 그대로 유지시켜 정확도 저하를 억제한다.

1. Feed-Forward Network

FFN은 Transformer의 인코더와 디코더 모두에 위치하며 self-attention과 residual connection을 거친 출력에 적용된다. 입력 차원은 $X_{FFN} \in \mathbb{R}^{S \times d_{model}}$이며, 여기서 S는 입력 문장 길이를 의미한다.

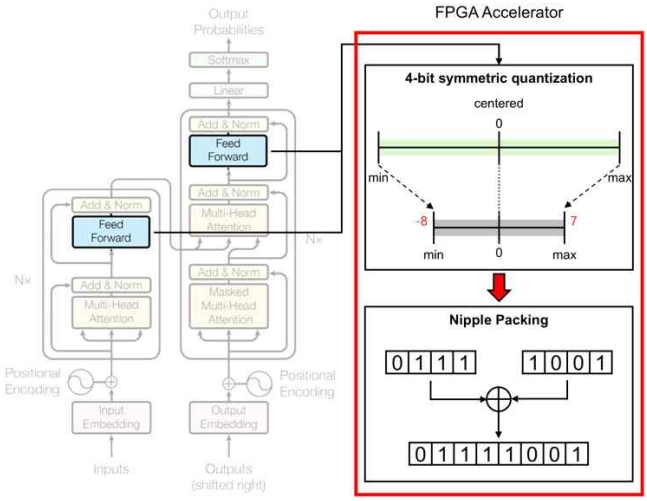


그림 1. FFN 레이어 및 FPGA 가속화를 위한 가중치 압축

FFN 레이어는 2개의 선형 변환과 $W_{hidden} \in \mathbb{R}^{d_{model} \times hidden}$으로 구성된다. 이때 필요한 연산량은 수식(1)과 같다.

$$S \times d_{model}^2 \times hidden + d_{model} \times hidden^2 \times d_{model} = d_{model}^2 \times hidden(S + hidden) \quad (1)$$

FFN은 연산량과 가중치 메모리 지배성을 이유로 병목으로 작용한다. 한 블록의 FFN은 토큰마다 두 차례의 대형 행렬 곱을 수행한다. 특히 어텐션의 주요 연산량은 문장 길이에 비례하기 때문에 엣지 추론 상황에서와 같이 문장 길이를 짧게 유지할 때는 FFN의 비중이 급격히 커져 실제 지연의 주된 원인이 된다. 또한 FFN은 어텐션의 선형 투영을 가중치를 모두 합친 크기와 비슷하거나 더 큰 비중을 차지한다. 이처럼 큰 가중치를 반복적으로 DRAM에서 스트리밍 하는 것은 대역폭과 온칩 메모리(BRAM/URAM) 용량에 영향을 받는다. BRAM 리소스가 낮은 보드에서는 외부 메모리 왕복이 전체 처리량을 제한하는 메모리 병목으로 이어진

표 1. HLS 합성 결과

	BRAM	DSP	FF	LUT
양자화 전	3813	5	18731	5102
양자화 후	242	5	14006	4975

다. 따라서 FFN의 모든 가중치를 온칩에 맵핑시켜 외부 접근을 최소화하여 지연과 처리량을 동시에 개선하고자 한다.

2. 4비트 대칭 양자화

양자화는 부동소수점을 정수 자료형으로 변환하여 모델 크기를 축소하고 추론 속도를 향상하는 기법이다[3,4]. 본 논문에서는 학습 재훈련이 필요 없는 사후 양자화(post-training) 방법을 적용한다. 또한 대칭 스케일(symmetric scale) 방식을 적용하여 가중치를 정수 형태로 변환한다. 대칭 스케일 변환은 수식(2)와 같다.

$$q = clip(round(s \times W), -8, 7), \qquad s = \frac{7}{\max|W|} \tag{2}$$

이후, 가중치 저장 효율을 위해 오프셋을 적용하여 수식(3)의 형태로 변환한다.

$$\tilde{q} = q + 8 \; IN[0, 15] \tag{3}$$

3. 니블 패킹(Nibble Packing)

니블 패킹은 1바이트를 상위 4비트와 하위 4비트로 나누어 각각 별도의 데이터를 저장하는 방식이다. 본 연구에서는 오프셋이 적용된 가중치를 1바이트의 상·하위 니블에 저장하여 메모리 사용량을 절감한다. 이후 부동소수점 복원은 수식(4)와 같다. 본 연구에서는 데이터 패킹과 부동소수점으로 복원하는 과정은 파이프라인으로 겹쳐 실행하여, 언패킹 단계가 지연의 새로운 원인이 되지 않도록 설계한다.

$$W \approx \frac{\tilde{q} - 8}{s} \tag{4}$$

Ⅲ. FPGA 구현

1. 시뮬레이션 세팅

본 시뮬레이션은 FPGA 보드 리소스 내에서 인코더와 디코더의 두 개 FFN 레이어의 모든 가중치를 온칩에 맵핑하는 것을 목표로 하였다. 사용된 FFN 가중치 차원은 512×2048이며, 실험에는 BRAM 자원이 288개로 비교적 제한적인 리소스를 가진 Kria KV260 FPGA 보드를 사용하였다. Transformer는 Pytorch로 학습하였고, FPGA 가속 시뮬레이션은 Vitis HLS 환경에서 수행하였다. Python으로 작성한 PS(Processing System)에서 FPGA로 FFN 레이어의 입력 데이터를 전송하고, FPGA의 PL(Programmable Logic) 측에서 FFN의 행렬곱 연산을 수행하였다. 이후 연산 결과는 다시 PS(Processing System) 측으로 반환되어 트랜스포머 연산의 다음 입력으로 사용되었다.

2. 합성 결과

FFN 가중치는 HLS 커널에서 니블 패킹된 int8 배열로 저장하여 용량을 최대 4배까지 절감하였다. 커널은 니블 데이터에서 상·하위 4비트를 시프트 및 마스킹 연산으로 추출하고, 사전에 정의된 스케일 값을 적용해 복원한 뒤 FFN 연산을 수행하였다. 이때 출력 값은 정확도 유지를 위해 부동소수점으로 유지하였다.

표1은 HLS에서 FFN 레이어 합성 결과를 나타낸다. 양자화 전 BRAM 사용량은 3813개로, KV260의 최대 한도(288개)를 초과하였다. 4비트 양자화 및 니블 패킹을 적용 후 BRAM 사용량이 242개로 줄어, 양자화 전 대비 93.7% 감소하였다. 이를 통해 모든 FFN 가중치를 BRAM 한도 내에 온칩에 맵핑할 수 있었다. 결과적으로 외부 DRAM 왕복으로 인한 메모리 접근 병목을 효과적으로 최소화하였다.

한편 니블 언패킹과 스케일 복원 연산이 조합 논리를 사용함에 따라 LUT 사용량은 증가하고 DSP는 변화가 없었다. 이는 스케일 값을 고정소수점으로 변환해 곱셈 연산을 DSP에 할당하거나, 부분합을 BRAM/URAM으로 이동시키는 방식으로 LUT-DSP-BRAM 간 리소스 균형을 조정할 수 있음을 보였다.

Ⅳ. 결론

본 논문에서는 Transformer의 주요 연산 병목인 Feed-Forward Network(FFN) 가속화를 위해 4비트 대칭 양자화와 니블 패킹 기법을 적용하여, 리소스가 제한된 Kria KV260 FPGA의 온칩 메모리에 FFN의 모든 가중치를 맵핑하는 방법을 제안하였다. Vitis HLS 합성 결과, 제안 기법은 BRAM 사용량을 3,813개에서 242개로 감소시켜(93.7% 절감) BRAM 리소스 내 온칩 맵핑을 가능하게 하였다. 이를 통해 외부 DRAM 접근에 따른 메모리 병목을 효과적으로 완화하였다. 또한 제안된 기법은 모델과 데이터 의존성이 낮아 다양한 Transformer 변형의 FFN에도 재학습 없이 적용할 수 있다. 향후 실제 FPGA 하드웨어 구현 및 Self-Attention 모듈까지 최적화 대상을 확장하여 Transformer 전체의 가속 성능을 극대화할 예정이며, 실제 보드 기준 엔드-투-엔드(end-to-end) 지연, 전력, 온도 프로파일을 계측하여 엣지 상황에서의 이점을 검증할 계획이다.

ACKNOWLEDGMENT

본 연구는 2023년도 산업통상자원부의 재원으로 한국에너지기술평가원(KETEP)과 2024년도 과학기술정보통신부의 재원으로 한국연구재단의 지원을 받아 수행된 연구임. (RS-2023-00266248, RS-2024-00349885)

참 고 문 헌

[1] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., and Polosukhin, I., "Attention is all you need," Advances in Neural Information Processing Systems 30, 2017.

[2] Li, B., Pandey, S., Fang, H., Lyv, Y., Li, J., and Ding, C., "FTRANS: energy-efficient acceleration of Transformers using FPGA," Proc. ACM/IEEE International Symposium on Low Power Electronics and Design (ISLPED), 2020.

[3] Choukroun, Y., Kravchik, E., Yang, F., and Kisilev, P., "Low-bit quantization of neural networks for efficient inference," IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), 2019.

[4] Wu, X., Li, C., Yazdani Aminabadi, R. Y., Yao, Z., and He, Y., "Understanding INT4 quantization for language models: latency speedup, composability, and failure cases," International Conference on Machine Learning (ICML), PMLR, 2023.

Impact of Sampling Rate on Deep Learning-based Human Activity Recognition Using Wearable Sensors

Nguyen Thi Hoai Thu, Dong Seog Han*
School of Electronic and Electrical Engineering, Kyungpook National University

thunguyen@knu.ac.kr, *dshan@knu.ac.kr

Abstract

In the wearable sensor-based human activity recognition (HAR) domain, it is crucial to select an appropriate sampling rate at which the sensor data would be captured before being forwarded to a classification model. This sampling rate needs to be selected carefully since sampling rates that are too high can require high storage capacity, wide transmission bandwidth, and high computational cost, while low sampling rates can reduce the classification accuracy. This study presents a statistical analysis of how sampling rates affect the performance of HAR systems, especially those whose classification methods are deep learning algorithms. With seven different sampling rates ranging from 100 Hz down to 20 Hz; and three advanced deep learning algorithms: multi-layer perceptron (MLP), convolutional neural network (CNN), and bidirectional long short-term memory network (BiLSTM), a series of experiments are carried out on the PAMAP2 benchmark dataset to find out the most effective sampling rate and determine which deep learning algorithm is best suited in handling low sampling rates.

I. Introduction

Wearable sensor-based human activity recognition systems (HAR) have been growing rapidly over the past decade, thanks to the advancement of microelectromechanical systems. With the advantages of protecting user privacy, wearable sensor-based HAR has dominated over the camera-based HAR approach and become a crucial part of the healthcare domain [1]. It can help users track their activities and health conditions, thereby promoting a healthier lifestyle. Some abnormal activities, such as freezing of gait from Parkinson's disease [2]. Various types of wearable sensors have been applied for the HAR systems to measure motions such as inertial measurement units (IMUs) and stretch sensors [3]. IMUs is one of the most broadly used sensors in HAR since they can capture human motion accurately. Therefore, this study will focus on the sampling rate of these sensors. The commonly used sampling rate of these sensors lies between 10 Hz and 100 Hz [1].

In a HAR system, the sensors capture the state changes at a fixed sampling rate and send this information to the processing unit for data preprocessing (filtering, windowing), feature extraction, and activity classification. Using a high sampling rate can provide more details on the motion of the user's body, thus helping to improve the accuracy of activity recognition. However, sampling sensor data at a high rate often requires consuming more energy, which is limited in wearable devices. Furthermore, to capture the data at a high frequency, high-quality sensors are often required, which can increase the production cost. On the other hand, using a low sampling rate can overcome the problems of power consumption, storage capacity, and data transmission loss. However, sensors with a low sampling rate might not be able to capture all the body movements. According to the Nyquist-Shannon sampling theorem, the sampling rate should be greater than twice the Nyquist frequency for a lossless reconstruction of a particular signal. Therefore, some minor differences between activities that have high similarity, such as walking and walking downstairs or jumping and falling, can be neglected when using a low sampling rate, since they contain high-frequency components. In a nutshell, it is a trade-off between accuracy and efficiency when deciding to use either a high sampling rate or a low sampling rate. Therefore, it is important to find out the most effective sampling rate. In this paper, we want to find out how different sampling rates can affect the deep learning-based HAR system, and which deep learning model can handle low sampling rates.

II. Methodology

All the experiments in this study are carried out on the PAMAP2 dataset [4]. The dataset contains 18 types of daily-life, household, and sport activities performed by nine subjects wearing three IMUs and a heart rate monitor. All data are sampled at a rate of 100 Hz. The acceleration and angular velocity data from the three IMUs are used in this study. The sequential data of each IMU sensor can be expressed as $\mathbf{S}_i = [\mathbf{a}_x, \mathbf{a}_y, \mathbf{a}_z, \mathbf{g}_x, \mathbf{g}_y, \mathbf{g}_z]$. The data is collected at an original sampling rate of 100 Hz. To generate the sensor data at sampling rates of 80, 60, 50, 40, 30, 20 Hz, we use an FIR anti-aliasing low-pass filter and compensate the delay of the output signal. Finally, the sensor signals are split into fixed-length windows with a length of 2 seconds and an overlap of 50%.

Table 1. Detailed Architecture of three deep learning-based HAR models.

CNN	BiLSTM	MLP
Conv(3×3, 32, 'relu', same)	BatchNorm()	Dense(units=128, 'gelu')
Conv(3×3, 32, 'relu', same)	BiLSTM(units=128, 'concat')	Dense(units=128, 'gelu')
MaxPool(2,1)	BiLSTM(units=128, 'concat')	Dense(units=128, 'gelu')
Conv(3×3, 32, 'relu', same)	BiLSTM(units=128, 'concat')	Dense(units=128, 'gelu')
Conv(3×3, 32, 'relu', same)	Dense(units=18, 'softmax')	Flatten()
MaxPool(2.2)		
Conv(3×3, 32, 'relu', same)		
Conv(3×3, 32, 'relu', same)		
GlobalAveragePool()		
Dense(units=18, 'softmax')		

Three commonly used deep learning (DL) models are considered in this study, including convolutional neural network (CNN), bi-directional long short-term memory (BiLSTM), and multi-layer perceptron (MLP). The detailed architecture of these models is shown in Table 1. The CNN is constructed from a set of 2D convolutional layers with (3×3) kernels and pooling layers. In the BiLSTM model, we perform batch normalization on the raw sensor data as a regularization mechanism. Three BiLSTM layers with concatenation merge mode are then deployed for further feature extraction. The MLP model used in this study contains four hidden layers, each has 128 hidden nodes and a GELU activation function.

The recognition accuracies of the three deep learning algorithms with seven different sampling rate values are illustrated in Fig. 1. It can be easily seen that the CNN-based system gains the best performance out of the three deep learning models in all different sampling rates. It outperforms the other two models with big gaps of 3-4%. In the CNN-based system, it is surprising that even though the sampling rate is reduced fivefold from 100 Hz to 20 Hz, the system performance only decreases 1.6% from 97% to 95.35%. Therefore, CNN with 2D kernels can directly handle raw sensor data with low sampling rates without any additional steps. In contrast, the MLP-based HAR system has better performance at low sampling rates than high sampling rates, as it obtains the highest accuracy at $F_s =$ 20, 30 Hz, and the lowest accuracy at a sampling rate of 60 Hz. Similarly, although the BiLSTM model gains the highest accuracy at $F_s =$100 Hz, there is not much difference compared to other sampling rate values. These results show that a higher sampling rate does not always provide higher accuracy, especially when the deep learning model is applied directly to the raw sensor signal. One possible reason is that long input sequences resulting from a high sampling rate can hamper the model in extracting and maintaining important features.

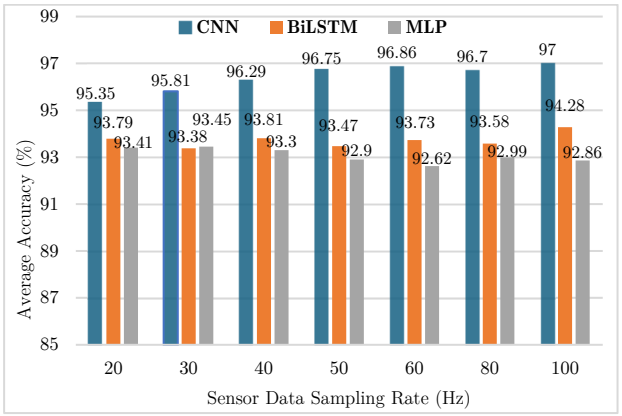


Fig 1. Performance comparison of three deep learning-based HAR systems on different sampling rates.

III. Conclusion

In this paper, experiments on different sensor sampling rates and different deep learning algorithms are carried out to find out the impact of sampling rate on DL-based HAR system performance. Experimental results from three deep learning algorithms and seven sampling rate values show that CNN is the best model for handling raw sensor data with low sampling frequencies. One interesting finding is that higher sampling rates with more detailed sensor signals do not always provide better classification results due to some limitations in feature extraction.

ACKNOWLEDGMENT

This work was supported by the IITP (Institute of Information & Communications Technology Planning & Evaluation)-ITRC (Information Technology Research Center) grant funded by the Korea government (Ministry of Science and ICT) (IITP-2025-RS-2020-II201808).

References

[1] O. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors," *IEEE Communications Surveys and Tutorials*, vol. 15, no. 3, pp. 1192–1209, 2013.

[2] N. T. H. Thu and D. S. Han, "Freezing of gait detection using discrete wavelet transform and hybrid deep learning architecture," in *2021 Twelfth International Conference on Ubiquitous and Future Networks (ICUFN)*. IEEE, 2021, pp. 448–451.

[3] N. T. H. Thu and D. S. Han, "An investigation on deep learning-based activity recognition using IMUs and stretch sensors," in *2022 International Conference on Artificial Intelligence on Information and Communication (ICAIIC)*. IEEE, 2022, pp. 377–382.

[4] A. Reiss and D. Stricker, "Creating and benchmarking a new dataset for physical activity monitoring," in *Proceedings of the 5th International Conference on Pervasive Technologies Related to Assistive Environments*, 2012, pp. 1–8.

엣지-디바이스 협업 환경에서 지연 시간 최소화를 위한 ResNet 동적 분할 전략 연구

정준호, 박진효, 윤주상*

동의대학교

jeong@junho.dev, jhpark3679@gmail.com, *jsyoun@deu.ac.kr

A Study on Dynamic Splitting Strategies for ResNet to Minimize Latency in Edge-Device Collaborative Environments

Junho Jeong, Jinhyo Park, Joosang Youn*

Dong-Eui Univ.

요 약

로컬 디바이스에서의 고성능 인공지능 모델 실행 수요가 증가함에 따라, AI 모델을 디바이스와 엣지 서버에서 분할하여 실행하는 분할 컴퓨팅 기술의 중요성이 대두되고 있다. 그러나 최적의 분할 지점은 디바이스의 연산 성능, 네트워크 대역폭 등 동적으로 변화하는 시스템 환경에 따라 민감하게 달라진다. 본 논문에서는 총 지연 시간 최소화를 목표로, 다양한 운영 환경에 따른 최적의 모델 분할 전략을 분석하기 위해 시뮬레이션 기반 연구를 수행한다. 이를 위해 총 지연 시간을 디바이스 연산 시간, 데이터 전송 시간, 엣지 연산 시간의 합으로 모델링하고, 대표적인 심층신경망인 ResNet-50을 대상으로 시스템 환경 변화에 따른 최적 분할 지점을 도출하였다. 실험 결과, 로컬 처리가 대부분의 저대역폭 구간에서 우위를 보였으며, 네트워크 대역폭이 높아질수록 연산 효율이 높은 SP2와 SP1으로 최적점이 전환되는 양상을 확인하였다.

Ⅰ. 서 론

최근 자율행동체 및 사물인터넷 기술의 발전으로, 제한된 컴퓨팅 자원을 가진 디바이스에서 복잡한 AI 모델을 실시간으로 실행해야 할 필요성이 급증하고 있다. 이러한 한계를 극복하기 위해 디바이스의 연산 부담을 엣지 서버로 오프로딩하는 기술이 제안되었으나, 원본 데이터 전체를 전송하는 방식은 높은 통신 지연과 프라이버시 문제를 유발할 수 있다. 이를 극복하기 위해 AI 모델을 분할하여 엣지 서버와 로컬 디바이스에서 나누어 처리하는 분할 컴퓨팅(Split Computing) 기술이 주목받고 있다[1].

분할 컴퓨팅의 핵심 과제는 모델의 최적 분할 지점을 선정하는 것이다. 최적의 분할 지점은 엣지 서버와 로컬 디바이스의 연산 성능이나 무선 네트워크의 대역폭 등 동적으로 변하는 요소에 크게 의존한다. 따라서 모든 상황에 적용할 수 있는 정적인 최적 분할 지점은 존재하기 어려우며, 환경 변화에 적응할 수 있는 동적 전략이 필수적이다. 본 논문에서는 이러한 동적 분할 전략의 기반을 마련하고자, 시뮬레이션을 통해 시스템 환경 변화에 따른 최적의 분할 지점을 분석한다.

Ⅱ. 관련 연구

엣지-디바이스 환경에서 DNN 모델을 효율적으로 실행하기 위한 분할 컴퓨팅 연구는 활발히 진행되어 왔다[1]. 분할 컴퓨팅은 DNN 모델을 디바이스에서 실행되는 Head 모델과 엣지 서버에서 실행되는 Tail 모델로 분할하여 연산량과 통신량을 절충하는 것을 목표로 한다. 분할 컴퓨팅의 접근 방법은 크게 모델의 구조를 변경하지 않는 방식과, 통신 효율을 높이기 위해 의도적으로 병목(bottleneck)을 삽입하는 방식으로 나눌 수 있다.

네트워크 수정 없는 분할 방식은 사전 학습된 모델의 구조와 가중치를 그대로 유지한 채 특정 계층을 분할 지점으로 선택하는 기법이다. 이 분야의 초기 연구인 Neurosurgeon[2]은 총 지연 시간과 에너지 소모를 최소화하는 최적의 분할 계층을 탐색했다. 해당 연구의 주요 결과 중 하나는 다수의 DNN 모델에서 최적 분할 지점이 모델의 가장 처음 또는 마지막 계층으로 나타났다는 점이다. 이는 완전한 로컬 컴퓨팅 또는 엣지 컴퓨팅이 중간에서 분할하는 것보다 더 효율적일 수 있으며, 중간 분할이 실질적인 이점을 갖는 조건에 대한 상세한 분석의 필요성을 제기한다.

본 연구는 네트워크 수정 없는 분할 방식을 기반으로 한다. 기존 연구들이 특정 조건에서 중간 분할의 효율성이 낮다고 보인 반면, 본 연구는 디바이스 성능, 네트워크 대역폭 등 다양한 시스템 환경 변수를 시뮬레이션하여 어떤 조건에서 중간 분할 지점이 최적의 전략이 되는지에 대한 정량적인 지표를 제시한다.

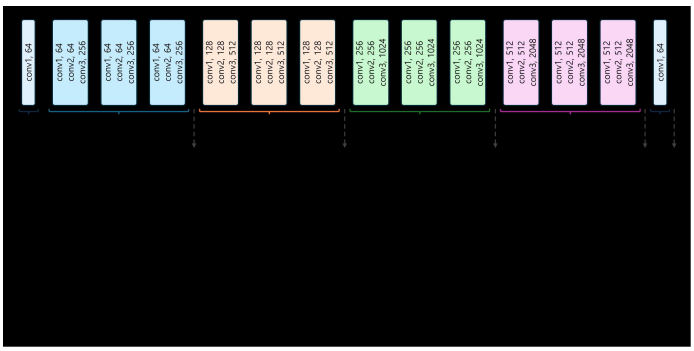


그림 1. ResNet-50 분할 후보 지점 및 후보 지점별 주요 파라미터

Ⅲ. 시뮬레이션 모델링

본 연구의 분석 대상으로는 ResNet-50[3] 모델을 선정하였다. 이는 컴퓨터 비전 분야에서 널리 사용될 뿐만 아니라, 최근 주목받는 비전-언어 모델(VLM)과 같은 더 복잡한 AI 모델의 백본으로 널리 활용되므로, 다양한 모델과 서비스에 확장하여 적용할 수 있는 대표성을 가지기 때문이다.

그림 1은 ResNet-50의 주요 블록 경계를 기준으로 선정한 분할 후보

지점과 각 지점의 핵심 파라미터를 나타낸다. 본 시뮬레이션은 표 1과 같은 조건으로 실험 환경을 구성하였다.

표 1. 시뮬레이션 환경 변수

환경 변수	값
로컬 디바이스 연산 성능(F_D)	5 - 100 GFLOPs
엣지 서버 연산 성능(F_E)	100 GFLOPs
네트워크 대역폭(B)	10 - 1,000 Mbps
분할 지점 후보 집합(K)	[SP1, SP2, SP3, SP4, LC]

분할 실행 시 발생하는 총 지연 시간(T_k)은 디바이스에서의 연산 시간(T_D^k), 데이터 전송 시간(T_{trans}^k), 엣지 서버에서의 연산 시간(T_E^k)의 합으로 정의하였다. 본 연구의 목표는 주어진 환경에서 T_k를 최소화하는 최적의 분할 지점 k^*를 찾는 것이다.

$$T_k = T_D^k + T_{trans}^k + T_E^k$$
$$= \frac{Head\ FLOPs}{F_D} + \frac{Data\ Size}{B} + \frac{Tail\ FLOPs}{F_E}$$
$$k^* = \arg\min_{k \in K} T_k$$

Ⅳ. 결과 및 분석

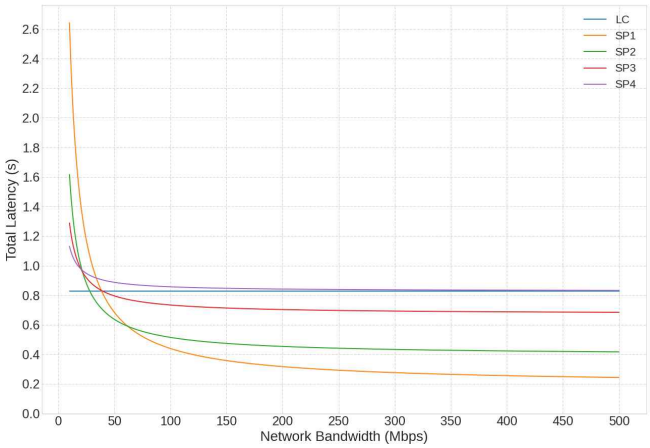


그림 2. 네트워크 대역폭에 따른 분할 지점별 총 지연 시간

본 장에서는 3장에서 정의한 시뮬레이션 결과를 제시하고 분석한다. 먼저 특정 조건에서 각 분할 지점의 성능 변화를 살펴본 뒤, 전체 환경에 대한 최적 지점 분포를 확인한다.

그림 2는 디바이스 성능을 5 GFLOPs로 고정했을 때, 네트워크 대역폭 변화에 따른 각 분할 지점(SP)의 총 지연 시간을 나타낸다. 네트워크를 사용하지 않는 로컬 컴퓨팅(LC)은 대역폭과 무관하게 일정한 지연 시간을 보여주며, 분할을 수행하는 경우 대역폭이 상승함에 따라 지연 시간이 크게 감소하는 경향을 보인다. 특히 네트워크 대역폭이 28Mbps가 넘어가는 지점에서 SP2의 성능이 LC를 넘어서며, 62Mbps가 넘어가는 구간에서는 SP1이 가장 우수한 성능을 보인다.

그림 3은 디바이스 성능과 네트워크 대역폭에 따른 최적 분할 지점을 히트맵으로 나타낸 결과이다. 네트워크 대역폭이 낮은 구간에서는 전송 지연이 시스템의 주된 병목으로 작용하므로 통신이 없는 로컬 컴퓨팅이 최적 전략으로 선택되었다. 네트워크 대역폭이 특정 임계점을 넘어서면, 최적 전략은 LC에서 SP2 및 SP1으로 전환되었다. SP3과 SP4는 모든 구간에서 최적점으로 나타나지 않은 이유는 이들이 갖는 장점이 특정 구간에서 다른 전략에 의해 항상 상쇄되기 때문이다. 네트워크가 느릴 때는 이들의 작은 전송량 이점에도 불구하고, 전송 자체가 없는 로컬 컴퓨팅에 비해 불리하며, 반면 네트워크가 충분히 빨라지는 시점에서는 전송량보다 연산량 감소가 더 중요해지기 때문에 SP1과 SP2에 비해 불리하다.

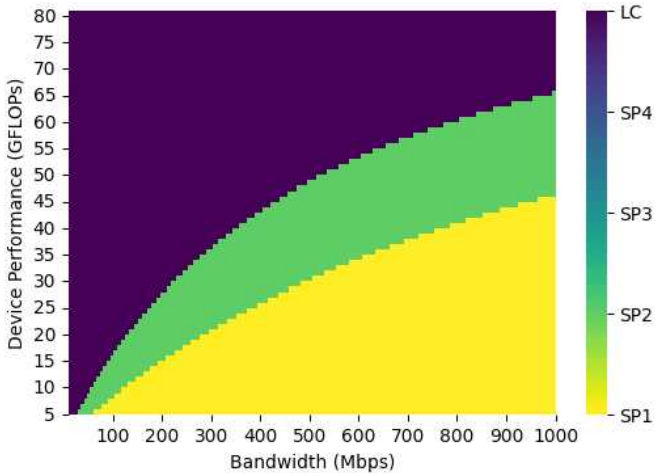


그림 3. 디바이스 성능과 네트워크 대역폭에 따른 최적 분할 지점

Ⅴ. 결 론

본 논문에서는 시뮬레이션을 통해 ResNet-50 모델의 최적 분할 지점이 디바이스 성능과 네트워크 대역폭에 따라 동적으로 변화함을 정량적으로 분석하였다. 이를 통해 실시간 분할 컴퓨팅 시스템에서 현재 시스템 상태에 맞는 최적의 분할 지점을 동적으로 선택하는 정책 수립에 필요한 정량적 근거를 제시하였다.

다만 본 연구는 ResNet-50 모델과 특정 시뮬레이션 환경에 기반하였기에, 다양한 모델 아키텍처 및 실제 네트워크 환경에서의 추가적인 검증이 필요하다. 향후 연구에서는 다양한 모델과 실제 환경에서 검증을 수행하고, 지연 시간 외에 에너지 효율성 및 프라이버시 등 다중 목표를 고려하는 심층강화학습 기반 정책 에이전트를 개발할 계획이다.

ACKNOWLEDGMENT

본 논문은 과학기술정보통신부 및 정보통신기획평가원의 복합지능자율행동체SW핵심기술개발사업(RS-2024-00346798)과 부산광역시 및 (재)부산테크노파크의 BB21plus 사업으로 지원된 연구결과임.

참 고 문 헌

[1] Matsubara, Y., Levorato, M., and Restuccia, F., "Split Computing and Early Exiting for Deep Learning Applications: Survey and Research Challenges," ACM Computing Surveys, vol. 55, no. 5, Article 90, Dec. 2022.

[2] Kang, Y., et al., "Neurosurgeon: Collaborative Intelligence Between the Cloud and Mobile Edge," in Proc. 22nd Int. Conf. on Architectural Support for Programming Languages and Operating Systems (ASPLOS '17), pp. 615-629, 2017.

[3] He, K., Zhang, X., Ren, S., and Sun, J., "Deep Residual Learning for Image Recognition," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-778, 2016.

베이지안 GMM 기반 탐지를 활용한 노이즈 견고 연합학습 기법

조예지, 김정현, 송홍엽*
세종대학교, 연세대학교*
yezy0390@sju.ac.kr, j.kim@sejong.ac.kr, hysong@yonsei.ac.kr*

Noise-robust Federated Learning with Bayesian GMM Detection

Yeji Cho, Junghyun Kim, Hong-Yeop Song*

Sejong Univ., Yonsei Univ.*

요 약

본 논문은 연합 학습의 복잡한 레이블 노이즈 문제를 해결하기 위한 종단간 레이블 수정 모델을 제안한다. 제안 모델은 2단계 프레임워크인 FedELC 모델의 로짓 조정 과정에 새로운 하이퍼파라미터를 도입하고, 클라이언트 탐지를 위해 4-성분 bayesian gaussian mixture model을 적용하였다. 이를 통해 제안 모델은 클래스 불균형 문제가 있는 로컬 데이터의 사전 분포 비율을 조절하여 편향 문제를 완화하고, 클라이언트의 구분을 더욱 세분화하여 불확실한 클라이언트에 대해 더욱 정교한 분류를 진행할 수 있다. 실험 결과, 제안 모델은 세 가지 유형의 레이블 노이즈 환경에서 모두 기존 모델보다 우수한 성능을 달성하였다.

Ⅰ. 서 론

연합 학습은 데이터 공유 없이도 분산된 클라이언트의 정보를 활용할 수 있어 개인 정보 보호가 필요한 다양한 애플리케이션에서 큰 성공을 거두었다. 그러나 연합 학습은 클라이언트별 데이터 클래스 분포가 다른 데이터 이질성 문제와 다양한 수준의 복잡한 레이블 노이즈로 인해 모델의 성능 저하가 발생한다. 특히 노이즈 비율이 높은 클라이언트는 잘못된 정보를 많이 포함하고 있어 성능 저하가 발생하기 쉽다. 따라서 노이즈가 많은 클라이언트의 부정적인 영향을 완화하기 위한 효과적인 방안이 필요하다.

본 논문에서는 복잡한 레이블 노이즈 문제를 해결하기 위한 2단계 프레임워크 모델을 제안한다. 제안 모델은 기존 FedELC 모델[1]의 로짓 조정 과정에 새로운 하이퍼파라미터 τ를 도입하였으며, 노이즈 클라이언트 탐지 방법을 2-성분 gaussian mixture model(GMM)에서 4-성분 bayesian gaussian mixture model(BGMM)으로 대체하였다. 이러한 방법을 통해 제안 모델은 로컬 클래스 분포의 반영 비율을 조절하여 클래스 불균형 데이터로 인한 편향 문제를 효과적으로 완화할 수 있으며, 네 가지로 세분화된 클라이언트 구분을 통해 더욱 정교한 노이즈 클라이언트 탐지가 가능하다. 실험 결과, 세 가지의 노이즈 환경을 고려한 CIFAR-10 데이터셋[2]에 대해서 제안 모델은 기존 모델보다 우수한 성능을 보여주었다.

Ⅱ. 본론

본 논문에서는 10개의 클래스로 구성된 CIFAR-10 데이터셋을 사용하였다. 우리는 데이터셋의 전체 평균과 표준 편차로 이미지를 정규화하였으며, 클라이언트별 클래스 분포가 다른 데이터 이질성 환경을 구현하기 위해 Dirichlet 분포를 사용하였다. 각 클라이언트에는 집중(concentration) 매개변수 γ를 갖는 Dirichlet 분포를 통해 분할된 훈련 데이터가 할당되며, γ 값이 낮을수록 Non-IID 정도가 더 높다는 것을 의미한다.

본 논문의 연합 학습 시스템은 N명의 클라이언트와 하나의 서버로 구성된다고 가정한다. 제안 모델은 두 단계로 구성되며, 1단계에서는 전체 클라이언트를 노이즈 클라이언트와 깨끗한 클라이언트로 구분하고, 2단계에서는 노이즈 클라이언트의 노이즈 레이블을 점진적으로 수정한다.

먼저, 1단계는 총 T_w번의 글로벌 라운드 동안 지속되며, FedAvg[3]를 기반으로 워밍업 모델을 훈련한다. 1단계의 각 라운드 t에 대해 선택된 각 클라이언트 k는 크로스 엔트로피(Cross-entropy, CE) 손실로 로컬 모델을 업데이트하고, FedAvg의 모델 집계 방법으로 글로벌 모델을 업데이트한다. 이때 각 클라이언트는 로컬 업데이트 과정에서 다음과 같이 클래스 분포를 고려하여 출력 예측을 보정한다.

$$\tilde{p} = p + \tau \log(\pi), \tag{1}$$

여기서 $p = f(x;\theta)$는 로컬 모델 f의 출력 예측을 의미하며, π는 로컬 데이터셋의 클래스 분포, θ는 현재 훈련된 로컬 모델 가중치를 나타낸다. 이 과정에서 제안 모델은 하이퍼파라미터 τ를 새롭게 도입함으로써 클래스 분포의 반영 비율을 조절할 수 있으며, 이는 로컬 데이터의 클래스 불균형으로 발생하는 편향 문제를 완화할 수 있다. 결과적으로 제안 모델의 로컬 업데이트를 위한 분류 손실(L_{cls})은 다음과 같이 정의할 수 있다.

$$L_{cls} = CE(p, \pi, \tau, \hat{y}), \tag{2}$$

여기서 $\hat{y}$은 노이즈가 있을 수 있는 원-핫 하드 레이블을 나타내며, τ는 실험적으로 확인한 최적값인 0.5로 설정하였다.

1단계 글로벌 학습 라운드 후, 제안 모델은 노이즈 클라이언트를 구분하기 위한 방법으로 기존의 2-성분 GMM 대신 4-성분 BGMM을 적용하였다. 이를 통해 제안 모델은 클라이언트를 4개의 그룹(깨끗한 그룹, 덜 깨끗한 그룹, 애매한 그룹, 노이즈 그룹)으로 구분하여 부분적 노이즈가 있는 클라이언트에 대해 더 유연한 대처를 할 수 있으며, 베이지안 추론으로 얻은 사후 분포를 통해 노이즈 클라이언트의 판정 신뢰도를 더 정교하게 반영할 수 있다.

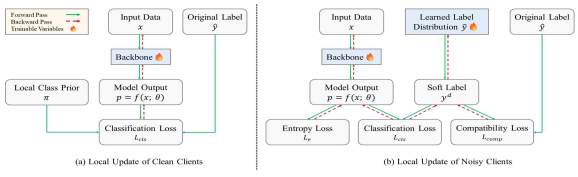


그림 1. 클린 클라이언트와 노이즈 클라이언트의 로컬 업데이트 과정.

그림 1과 같이, 2단계에서는 깨끗한 그룹에 속하는 클라이언트는 수식 (2)와 동일한 분류 손실을 사용하여 로컬 업데이트를 수행하고, 나머지 세 그룹은 로컬 레이블 수정 과정을 적용한다. 로컬 레이블 수정 과정은 모델 파라미터 학습과 레이블 수정을 동시에 최적화하는 것을 의미하며, 이를 위해 학습 가능한 소프트 레이블 변수 y^d를 도입한다. y^d는 원-핫 하드 레이블 $\hat{y}$을 기반으로 초기화되며 역전파를 통해 모델 파라미터와 함께 업데이트 된다. 구체적으로 소프트 레이블 y^d는 다음과 같이 정의할 수 있다.

$$y^d = Softmax(\tilde{y}), \quad (3)$$

여기서 $\tilde{y} = K\hat{y}$는 학습 가능한 변수이고, K는 큰 상수로 본 논문에서는 기존 모델과 동일한 값인 10을 사용한다. 이러한 과정을 통해 제안 모델은 노이즈 레이블을 점진적으로 수정하여 실제 정답 y^*에 근접한 소프트 레이블 y^d로 수렴하도록 유도할 수 있다.

노이즈 클라이언트의 로컬 최적화 목표는 다음과 같이 세 가지로 나눌 수 있다. 첫 번째, 분류 손실(L_{cls})은 하드 레이블 $\hat{y}$을 사용하는 대신, 모델 예측 p와 학습 가능 분포 $\tilde{y}$간의 분류 손실을 계산하도록 하며 다음과 같이 정의할 수 있다.

$$L_{cls} = CE(p, \tilde{y}). \quad (4)$$

두 번째, 호환성 정규화 손실(L_{comp})은 소프트 레이블 y^d와 하드 레이블 $\hat{y}$이 크게 차이나지 않도록 정규화하기 위해 사용되며 다음과 같이 정의할 수 있다.

$$L_{comp} = Compatibility(\hat{y}, y^d) = -\sum_{m=1}^{M} \hat{y}_m \log(y_m^d), \quad (5)$$

여기서 M은 CIFAR-10 데이터셋의 클래스 개수를 의미한다.

세 번째, 엔트로피 정규화 손실(L_e)은 모델이 더욱 정확하고 신뢰할 수 있는 예측을 출력하도록 유도하는 것이며 다음과 같이 정의할 수 있다.

$$L_e = Entropy(p) = -\sum_{m=1}^{M} p_m \log(p_m), \quad (6)$$

여기서 p_m은 m번째 클래스에 대한 모델의 출력 소프트맥스 확률이다.

위의 세 항을 모두 결합하여, 탐지된 노이즈 클라이언트에 대한 트리플렛 손실은 다음과 같이 정의할 수 있다.

$$L = L_{cls} + \alpha L_{comp} + \beta L_e, \quad (7)$$

여기서 α와 β는 손실 항목의 균형을 잡기 위한 하이퍼파라미터로, 기존 모델과 동일하게 각각 0.2와 0.5로 설정하였다. 이러한 트리플렛 손실에서 L_{cls}와 L_{comp}의 학습 가능한 $\tilde{y}$의 최적화 과정은 다음과 같이 정의할 수 있다.

$$\tilde{y} = \tilde{y} - \eta(\nabla \tilde{y}), \quad (8)$$

여기서 η는 $\tilde{y}$의 학습률을 의미하며, 기존 모델과 동일하게 1000으로 설정하였다. 결과적으로 실제 레이블 y^*에 가까운 수정된 레이블은 로컬 업데이트를 거친 모델의 예측 p와 업데이트된 $\tilde{y}$를 결합하여 얻어진다. 이후 제안 모델은 로컬 모델 가중치를 조절하는 Distance-Aware(DA) aggregation 방법을 적용하여 노이즈 클라이언트의 기여도를 조절하였다.

제안 모델의 성능을 평가하기 위해 클래스별 평균 정밀도(Precision)와 재현율(Recall)을 평가 지표로 사용하였으며, 세 가지 유형의 레이블 노이즈 환경을 고려한 CIFAR-10 데이터셋에서 실험을 진행하였다. 이때 실험에서 사용되는 세 가지 노이즈는 대칭 노이즈(Symmetric), 비대칭 노이즈(Asymmetric), 혼합 노이즈(Mixed)가 있으며, 혼합 노이즈는 클라이언트를 두 그룹으로 나눠 각각 대칭 노이즈와 비대칭 노이즈가 적용되도록 하였다. 이때 노이즈 비율은 0에서 0.4까지 선형적으로 증가하도록 설정하였다. 훈련에는 ResNet-18을 백본으로 사용하였으며, 혼합 정밀도 학습[4]을 적용하여 모델 학습을 가속화하였다. 학습 과정에서 총 $N = 100$명의 클라이언트를 생성하고, 매 글로벌 라운드 t마다 무작위로 선택된 10명의 클라이언트가 연합 학습에 참여한다. 전체 글로벌 라운드는 120번이며, 이 중 20번은 1단계 워밍업 라운드 T_w이고, 나머지 100번은 2단계의 글로벌 라운드로 구성된다. 각 클라이언트의 로컬 에포크 E는 5, 로컬 배치 크기는 64로 설정하였다. 모델 최적화에는 SGD 최적화기를 사용했고, 초기 학습률은 0.01, 모멘텀은 0.9, 가중치 감소는 0.0005로 설정하였다.

실험 결과, 제안 모델은 세 가지 유형의 레이블 노이즈 환경에서 모두 기존 모델보다 우수한 성능을 보여주었으며, 특히 혼합 노이즈 환경에서 가장 우수한 성능을 보였다. 이러한 결과는 표 1에서 확인할 수 있다.

Method	Dirichlet γ=1.0					
	Sym.(0.0-0.4)		Asym.(0.0-0.4)		Mixed(0.0-0.4)	
	Pre.	Rec.	Pre.	Rec.	Pre.	Rec.
기존 모델[1]	76.81	76.72	77.97	76.98	77.98	77.24
제안 모델	**78.36**	**77.83**	**79.17**	**78.69**	**79.10**	**78.40**

표 1. CIFAR-10 데이터셋에 대한 제안 모델과 기존 모델의 성능 비교.

Ⅲ. 결론

본 논문은 연합 학습에서 노이즈 클라이언트의 노이즈 레이블을 수정하기 위한 2단계 레이블 수정 기법을 제안한다. 제안 모델은 기존 로짓 조정 과정에 새로운 하이퍼파라미터 τ를 도입하여 로컬 클래스 분포의 반영 비율을 조절하고, 클라이언트 탐지를 위해 4-성분 BGMM을 적용하였다. 이를 통해 제안 모델은 클래스 불균형으로 인한 편향 문제를 효과적으로 완화하며, 기존 모델보다 세분화된 구분을 통해 노이즈 클라이언트를 보다 정교하게 탐지할 수 있다. 실험 결과, 세 가지 노이즈 환경을 고려한 CIFAR-10 데이터셋에서 제안 모델은 클래스별 평균 정밀도와 재현율 모두 기존 모델보다 우수한 성능을 달성하였다.

ACKNOWLEDGMENT

이 (성과)는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No.RS-2023-00209000).

참 고 문 헌

[1] Jiang, X., Sun, S., Li, J., Xue, J., Li, R., Wu, Z., and Liu, M. "Tackling noisy clients in federated learning with end-to-end label correction," in Proc. 33rd ACM international conference on information and knowledge management (CIKM), Boise, ID, USA, pp. 1015-1026, Oct. 21-25, 2024.

[2] Krizhevsky, A. and Hinton, G. "Learning multiple layers of features from tiny images," 2009.

[3] McMahan, B., Moore, E., Ramage, D., Hampson, S., and y Arcas, B. A. "Communication-efficient learning of deep networks from decentralized data," in Procs. 20th Artificial intelligence and statistics (PMLR), Fort Lauderdale, FL, USA, pp. 1273-1282, Apr. 20-22, 2017.

[4] Micikevicius, P., Narang, S., Alben, J., Diamos, G., Elsen, E., Garcia, D., and Wu, H. "Mixed precision training," arXiv preprint arXiv:1710.03740, 2017.

Predicting Deepfake Virality on Social Networks

Seoyoon Jeong, JeeEun Kim, S. Shyam Sundar, Jinyoung Han*
Sungkyunkwan University

jsyoon0503@g.skku.edu, wldms15@g.skku.edu, sss12@psu.edu, *jinyounghan@skku.edu

Abstract

The proliferation of deepfake content on social media threatens information integrity. The role of deepfake media's unique visual characteristics in propagation has been largely overlooked by prior research. This paper proposes a proactive predictive model that forecasts a deepfake image's total shares over a 15-day period using its visual features and early propagation data. We constructed a unique deepfake propagation dataset and analyzed how image attributes correlate with virality. Our model, which integrates GraphSAGE with an LSTM network and CNN-extracted visual features, achieves superior predictive accuracy. The results show our approach can effectively identify deepfake content with high viral potential, contributing to the early prediction and mitigation of deepfake dissemination.

Ⅰ. Introduction

With the rapid advancement of generative models, deepfake technology has become widely accessible, enabling the creation of highly realistic synthetic media [1]. Malicious actors are exploiting this technology to spread misinformation and manipulate public opinion on social media [2]. The viral nature of deepfakes on social networks is dangerous, as their influence is difficult to control once they gain traction, posing a significant threat to public trust and safety. This underscores the critical need for early intervention strategies [3].

While deepfake detection technologies have made progress, most are reactive, identifying synthetic media only after it has been widely disseminated. This approach is often too slow to counter the rapid spread of deepfakes. Although research has explored content virality, the unique characteristics of deepfake media have been largely overlooked [4]. Prior studies have mainly focused on text-based content or benign images, ignoring the fact that deepfakes are a highly visual form of stimulus. As such, they are primarily shared based on their visual content, and user attributes or textual metadata are often limited or unreliable [5].

To address these challenges, we propose a novel predictive model that forecasts the total number of shares a deepfake image post will accumulate over a 15-day period, based on the deepfake image and its initial propagation data. Our model identifies posts with high virality potential shortly after they are published, offering a proactive solution to complement existing detection efforts and serving as a crucial step toward building a safer digital environment.

Ⅱ. Methods

Ⅱ-1. Data Collection and Processing

To train and evaluate our model, we created a novel dataset to track the spread of deepfake content, as no public dataset for this purpose existed. From November 2023 to March 2025, we tracked 6,833 deepfake image posts on Tumblr, chosen for its "reblog" function, which is ideal for studying content propagation. For each post, we collected the deepfake image, its full reblog cascade, and all relevant timestamps.

To better analyze the characteristics of deepfake images that tend to go viral, we used automated tools to extract visual and demographic attributes. For example, we used the DeepFace model to infer labels for age, gender, and race. We also used Nudity-Detection to calculate a nudity score for each image, which ranges from 0 to 100, with higher values indicating greater visual exposure.

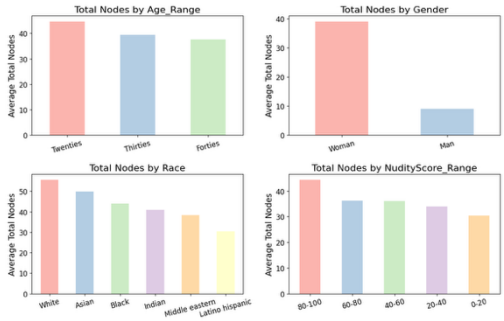


Figure 1. Influence of Demographics and Nudity on Shares

Ⅱ-2. Data Analysis

To better understand how specific visual and demographic attributes of deepfake images correlate with their virality, we measured the size of their propagation cascades by the average total number of reshares. Deepfakes of individuals categorized as White and Asian also tend to have a higher average number of reshares. We found a positive correlation between visual exposure and virality, as deepfakes with a higher nudity score had the largest average propagation cascades.

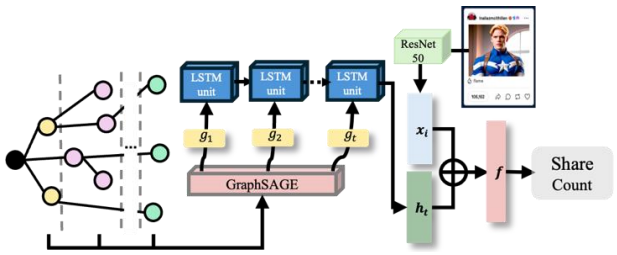


Figure 2. Overall architecture of the proposed model pipeline

This analysis confirms that the visual and demographic attributes of deepfake images are key factors in their propagation, providing crucial insights into the characteristics that drive content virality.

Ⅱ-3. Propagation Prediction Model

To predict the total number of shares, we developed a two-stage model that combines a GraphSAGE-based graph encoder with a Long Short-Term Memory (LSTM) network. This approach captures both the structural and temporal dynamics of content propagation.

In the first stage, a GraphSAGE encoder processes snapshots of the propagation graph over time. This encoder effectively captures the structural information of the sharing cascade, even with limited node-level data. The resulting sequence of graph embeddings is then fed into an LSTM network, which models the temporal evolution and learns how early propagation patterns influence the final share count.

Finally, the LSTM's final hidden state is concatenated with a visual feature vector extracted from the deepfake image using a pre-trained ResNet-50 model. This combined vector is then fed into a fully connected layer to predict the total share count, allowing our model to make a more accurate forecast by integrating both propagation and visual features.

Ⅱ-4. Result

To evaluate our model's performance, we used three metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Normalized Discounted Cumulative Gain (nDCG). While MAE and RMSE measure the accuracy of our numerical predictions for total shares, nDCG assesses the model's ability to rank content by its propagation potential, which is crucial for identifying high-risk, high-virality content for early intervention.

For benchmarking, we compared our model against two established temporal graph-based models, TGN (Temporal Graph Networks) and TGAT (Temporal Graph Attention Network), and a baseline model combining a GCN with an LSTM. These models are well-suited for capturing the evolving dynamics of social networks. To ensure a fair comparison, all models were provided with the same initial propagation graph structure constructed from the content sharing events.

Our model achieved the lowest MAE and RMSE values and the highest nDCG score, consistently outperforming the TGN and TGAT baselines in predicting the total number of shares. This demonstrates our model's superior predictive accuracy and its strong ability to identify content with high virality potential. By combining the structural and temporal dynamics of the sharing cascade with semantic features from the deepfake image, our approach proves to be a more effective way to model content propagation.

Methods	MAE	Share RMSE	nDCG
GCN + LSTM	5.339	8.431	0.653
TGN	6.207	9.005	0.602
TGAT	5.614	8.292	0.713
Ours	**4.821**	**7.901**	**0.781**

Table 1. Comparison of our method with baseline models

Ⅲ. Conclusion

In this paper, we proposed a novel predictive model to forecast the total number of shares of deepfake content using both the image's visual features and its early propagation data. We constructed a dedicated deepfake propagation dataset from Tumblr and analyzed how deepfake image features correlate with their virality. We employed a graph encoder combined with an LSTM network to capture spatiotemporal diffusion patterns, which were integrated with CNN-extracted visual features to predict the total number of shares. The results show that our model effectively identifies content with high viral potential. These findings suggest that our approach can contribute to the early prediction, thereby supporting the development of safer social networking environments.

ACKNOWLEDGMENT

This work was supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.RS-2023-00230337, Advanced and Proactive AI Platform Research and Development Against Malicious Deepfakes) and under the Global Scholars Invitation Program (RS-2024-00459638).

References

[1] Amerini, Irene, et al. "Deepfake media forensics: Status and future challenges." Journal of Imaging 11.3 (2025): 73.

[2] Xu, Zhaoxiang, et al. "Public perception towards deepfake through topic modelling and sentiment analysis of social media data." Social Network Analysis and Mining 15.1 (2025): 16.

[3] Abbas, Fakhar, and Araz Taeihagh. "Unmasking deepfakes: A systematic review of deepfake detection and generation techniques using artificial intelligence." Expert Systems with Applications 252 (2024): 124260.

[4] Wu, Cheng-Lin, et al. "MUFFLE: Multi-modal fake news influence estimator on Twitter." Applied Sciences 12.1 (2022): 453.

[5] Sundar, S. Shyam, Maria D. Molina, and Eugene Cho. "Seeing is believing: Is video modality more powerful in spreading fake news via online messaging apps?." Journal of Computer-Mediated Communication 26.6 (2021): 301-319.

논문 서지 데이터를 활용한 AI 분야의 한·미·중 연구자 이동성 분석

권태훈
한국과학기술정보연구원
kth78@kisti.re.kr

Analysis of Researcher Mobility in the AI Field across Korea, USA, and China Using Bibliometric Data

Taehoon Kwon
Korea Institute of Science and Technology Information

요 약

본 논문에서는 AI 분야의 한, 미, 중 연구자 이동성 분석을 수행하였다. 이를 위해 Openalex논문을 기반으로 1995년부터 2024년 AI 분야 논문 데이터 1,667,754건를 수집하여, 해당 저자들 1,801,975명의 소속 기관의 국가 정보의 변화를 분석하여 이동성을 분석하였다. 한국의 경우 유입 비율 4.33%, 유출 비율 4.36%, 미국의 경우 유입 비율 9.04%, 유출 비율 10.15%, 중국의 경우 유입 비율 4.55%, 유출 비율 2.60%로 분석되었다. 또한, 한국 연구자의 해외 유출에 미국이 가장 큰 비중을 차지하며, 중국과 미국은 서로에 대해 해외 유출에 있어 비중이 큰 국가로 분석되었다.

Ⅰ. 서 론

연구자 이동은 지식의 글로벌 확산과 국제 공동연구 활성화에 핵심적 역할을 하며, 나아가 혁신 생태계의 강화에도 기여한다[1][2]. 그러나 때로는 이러한 이동이 국가 간 연구력 불균형과 인력 격차로 이어져 정책적 개입이 요구되기도 한다[1][2].

최근 정부는 AI 기술을 국가 핵심 전략 산업으로 육성하기 위해 대규모 투자를 단행하고, 동시에 핵심 인재 확보 전략을 발표하였다[3]. 효과적인 인력 전략 수립을 위해서는 무엇보다도 연구자 이동성에 대한 체계적인 분석이 선행되어야 한다. 하지만 기존의 설문조사 중심 접근법은 확장성과 시의성 측면에서 한계가 있다[1][2].

이에 본 논문에서는 서지 데이터를 활용하여 AI 분야에서 한국, 중국, 일본 연구자들의 이동성을 분석하고자 한다.

Ⅱ. 본론

본 논문에서는 연구자 이동성 분석을 위해 공개 논문 데이터베이스인 OpenAlex를 활용하였다[4]. OpenAlex는 학문 분류 체계를 기반으로 한 Concept 정보를 제공한다. 이에 본 연구에서는 1985년부터 2024년까지 출판된 논문 중 AI를 Concept으로 갖는 논문을 수집하였다. Concept에 대한 확률을 의미하는 Concept Score는 0.5 이상을 기준으로 설정하였으며, 수집한 논문 가운데 Journal과 Conference Proceeding에 해당하는 Article 유형만을 분석 대상으로 사용하였다.

수집된 논문 데이터 1,667,754건으로부터 OpenAlex가 제공하는 저자 식별 ID를 활용하여, 총 1,801,975명의 저자에 대해 최초 논문 소속 국가와 최종 논문 소속 국가 정보를 추출하였다. 최초 소속 국가가 최종 소속 국가 목록에 포함되지 않는 경우 이를 유출로 정의하였으며, 해당 저자의 최종 소속 국가들을 유출국으로 간주하였다. 반대로, 최종 소속 국가가 최초 소속 국가 목록에 포함되지 않는 경우 이를 유입으로 정의하고, 해당 저자의 최초 소속 국가들을 유입국으로 간주하였다.

이러한 정의를 바탕으로 한국, 미국, 중국을 대상으로 한 상위 10개의 유출국별 유출·유입 규모를 분석하였으며, 그 결과는 각각 그림 1, 그림 2, 그림 3에 제시하였다.

한국 연구자는 해외 유출에서 미국 비중이 가장 크지만, 유출입을 함께 고려하면 미국과의 관계에서 유입이 더 우세하였다. 미국 연구자는 해외 유출에서 중국 비중이 가장 크고, 유출입을 고려해도 중국과의 관계에서 유출이 더 두드러졌다. 중국 연구자는 해외 유출에서 미국 비중이 가장 크지만, 유출입을 함께 보면 미국과의 관계에서 유입이 더 우세하였다.

또한 그림 4에서는 한국, 미국, 중국의 유출·유입 연구자 비율을 비교하였다. 앞의 유출입 정의를 따라 연구자의 최초 소속 국가는 해당 연구자의 출신 국가로 간주한다. 유출 연구자 비율은 출신 국가가 최종 소속 국가 목록에 포함되지 않는 연구자의 수를, 출신 국가 연구자 전체 수로 나누어 정의한다. 유입 연구자 비율은 해당 최종 소속 국가에는 포함되지만 출신 국가는 아닌 연구자의 수를, 해당 최종 소속 국가에 포함된 연구자 전체 수로 나누어 정의한다.

한국의 경우 유입 비율 4.33%, 유출 비율 4.36%, 미국의 경우 유입 비율 9.04%, 유출 비율 10.15%, 중국의 경우 유입 비율 4.55%, 유출 비율 2.60%로 분석되었다.

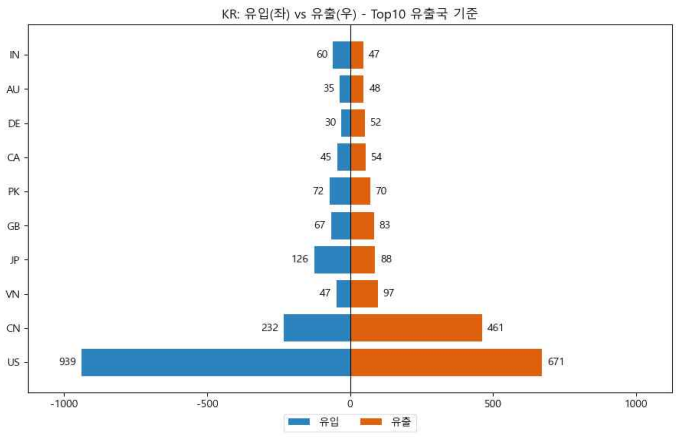


그림 1 한국의 국가별 연구자 유출입 비교 (유출 상위 10개국 기준)

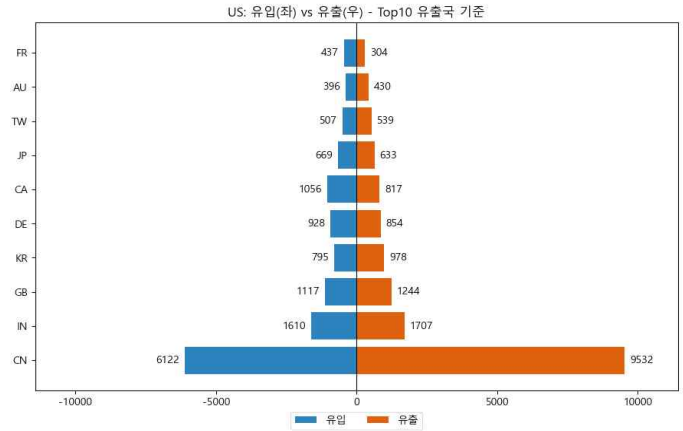


그림 2 미국의 국가별 연구자 유출입 비교 (유출 상위 10개국 기준)

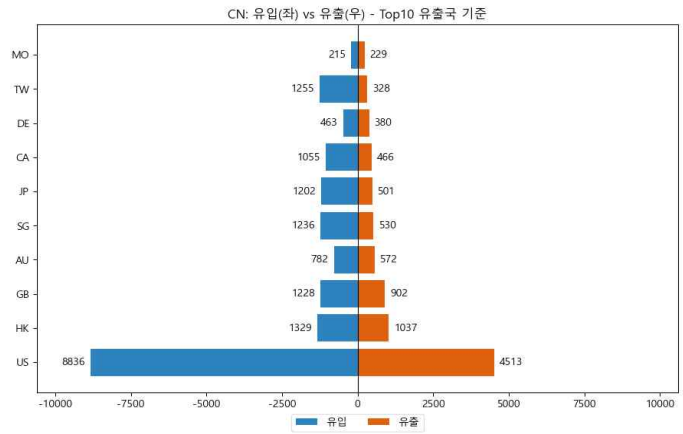


그림 3 중국의 국가별 연구자 유출입 비교 (유출 상위 10개국 기준)

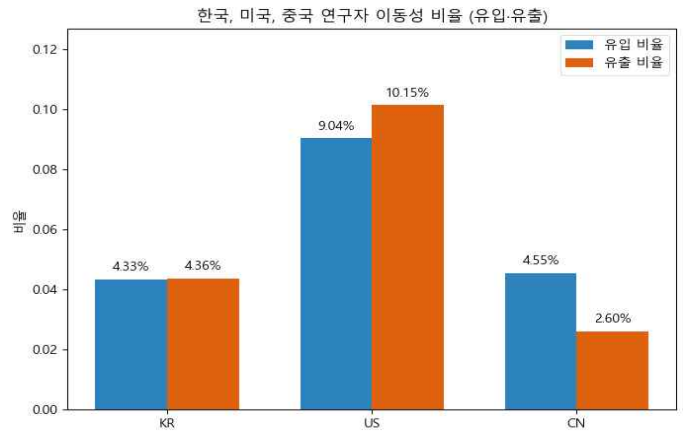


그림 4 한국, 미국, 중국의 유출입 연구자 수 비율

Ⅲ. 결론

본 연구는 OpenAlex 데이터를 활용하여 1985~2024년 AI 분야에서 한국, 중국, 일본 연구자의 이동성을 분석하였다. 유출만 기준으로 보면, 한국은 미국으로의 유출이 가장 컸고, 미국은 중국으로의 유출이 두드러졌으며, 중국은 미국으로의 유출이 가장 높았다. 유출입을 함께 고려하면, 한국은 미국과의 관계에서 유입이 우세했고, 미국은 중국과의 관계에서 유출이 더 컸으며, 중국은 미국과의 관계에서 전반적으로 유입이 우세하였다. 이러한 결과는 국가별 연구자 이동 특성이 상이함을 보여주며, 연구자 이동성 분석과 이에 기반한 맞춤형 전략이 과학기술 경쟁력 강화를 위해 중요함을 시사한다.

ACKNOWLEDGMENT

본 논문은 2025년 한국과학기술정보연구원(KISTI)의 기본사업으로 수행된 연구입니다.(K25L4M2C3).

참 고 문 헌

[1] Robinson-Garcia, N., Sugimoto, C. R., Murray, D. S., Yegros Yegros, A., Larivière, V., & Costas, R. (2019). The many faces of mobility: Using bibliometric data to measure the movement of scientists. Journal of Informetrics, 13(1), 50 - 63.

[2] Sugimoto, C. R., Robinson-Garcia, N., & Costas, R. (2016). Towards a global scientific brain: Indicators of researcher mobility using co-affiliation data. OECD Blue Sky III Forum on Science and Innovation Indicators, Ghent, September 19-21. https://arxiv.org/abs/1609.06499

[3] Ministry of Science and ICT. (2025, February 4). Public report on the implementation of key tasks in 2025 [Press release]. https://www.msit.go.kr/

[4] OpenAlex. (2025). OpenAlex database. Accessed June 2025, from https://openalex.org

실시간 공항 운영을 위한 PPO 기반 자원 배정 최적화 시스템 설계

김준규, 천민범, 안광모
주식회사 케파스
jkkim@kephas.kr, mbcheon@kephas.kr, gmahn@kephas.kr

Design of a PPO-Based Optimization System for Real-Time Airport Resource Allocation

Kim Jun Kyu, Cheon Min Beom, Ahn Kwang Mo
Kephas.

요 약

전 세계 항공 운송 수요의 지속적 증가에 따라 공항 운영에서 자원의 효율적이고 실시간적인 배정은 핵심 과제로 부상하고 있다. 기존 혼합정수선형계획법(Mixed-Integer Linear Programming, MILP)이나 메타휴리스틱 기반 기법은 환경 변화에 대한 유연성이 낮고, 실시간 운영 상황에서의 적용에 한계가 존재한다. 본 연구에서는 수학적 모델링과 정책 기반 강화학습(Proximal Policy Optimization, PPO)을 결합한 실시간 자원 배정 시스템을 제안한다. 제안 시스템은 공항 데이터 입력과 API 연동을 통해 자원 배정 요청을 수신하고, 전처리된 데이터를 기반으로 MILP 모델을 구성한 후 사전 학습된 PPO 정책을 호출하여 실시간 배정 결과를 산출한다. 인천국제공항의 실제 항공편 데이터를 활용한 실험에서 배정률 100%, 1,000편 기준 평균 실행시간 258초, 제약 위반률 0.27%를 기록하였다. 또한 동일 조건에서 수행한 기존 연구의 A3C 기반 REGAPS 모델이 평균 실행시간 185초를 보인 것과 비교해, 제안 방식은 평균 15.1초로 현저히 빠른 응답 속도를 달성하였다. 이를 통해 본 시스템이 실시간 운영 환경에서 기존 접근법 대비 우수한 성능을 제공함을 입증하였으며, 향후 다양한 운영 시나리오와 다중 자원 배정 문제로의 확장 가능성을 제시한다.

Ⅰ. 서론

전 세계 항공 운송 수요는 지속적으로 증가하고 있으며, 이에 따라 공항 자원 배정의 효율성과 실시간성은 운영 안정성을 결정짓는 핵심 요소로 부상하고 있다. 국제항공운송협회(IATA)는 2037년까지 전 세계 항공 여객 수가 약 82억 명에 이를 것으로 전망하며[1], 게이트, 주기장, 체크인 카운터, 수하물 처리 시스템 등 핵심 인프라 자원 수요 역시 급격히 증가할 것으로 예상된다. 특히 항공편 지연, 기종 변경, 긴급 착륙 등 예외 상황이 빈번한 실제 운영 환경에서는 단순 계획 수립을 넘어 실시간 유연한 자원 배정 체계가 필수적이다[2].

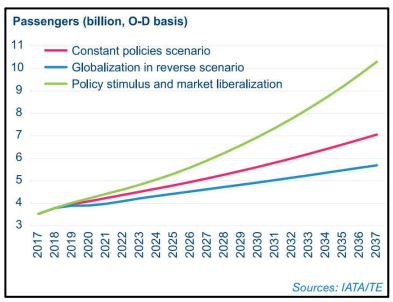


그림 1. IATA 시나리오별 항공 여객 수 증가 전망(2017-2037)

공항 자원 배정 문제는 시간 충돌, 기체 수용 조건, 자원 선호도 등 다양한 제약을 동시에 고려해야 하는 조합 최적화 문제이며, 일반적으로 NP-Hard에 속한다. 기존 연구들은 혼합정수선형계획법(Mixed-Integer Linear Programming, MILP), 유전 알고리즘(Genetic Algorithm, GA) 등 수치 기반 또는 메타휴리스틱 기법을 적용해 문제를 해결해왔다[3]. 그러나 이러한 접근은 대부분 정적 계획 기반으로 구성되어 있어 환경 변화 시 전체 계획을 반복적으로 재계산해야 하는 한계를 가진다. 특히 실시간성이 중요한 운영 환경에서는 계산 복잡도 증가, 수작업 개입, 낮은 자동화 수준 등으로 인해 실제 적용이 어렵다.

이러한 한계를 해결하기 위해서는 단순한 알고리즘 개선을 넘어 문제 정의 - 정책 실행 - 결과 반환 전 과정을 통합적으로 처리할 수 있는 구조적 설계가 필요하다. 본 연구에서는 자원 배정 문제를 수학적으로 정의하고, **정책기반강화학습(ProximalPolicyOptimization,PPO)**을 결합한 실시간 자원 배정 시스템을 설계하였다. 사용자의 요청을 입력받아 최적화 모델을 구성하고, 사전 학습된 정책을 호출하여 최적의 배정 결과를 실시간으로 도출하는 전 과정을 하나의 구조로 통합하였다. PPO는 정책 업데이트 안정성과 빠른 수렴 속도에서 강점을 가지며, 보상 구조를 통해 다양한 제약을 내재화할 수 있어 수리 모델과의 결합에도 효과적이다[4]. 제안 시스템은 입력, 최적화, 정책 호출, 결과 반환까지의 전 과정을 실시간으로 자동화함으로써, 공항 운영 환경의 실질적 문제를 해결할 수 있는 설계 구조를 제시한다.

Ⅱ. 본론

본 연구에서는 실시간성과 자동화를 동시에 충족하기 위해, 공항 자원 배정 시스템을 모듈화된 구조로 설계하였다. 제안하는 시스템은 공항 데이터 입력부터 최적화 모델 생성, 정책 기반 배정 실행, 결과 반환까지의 전 과정을 통합된 프로세스로 구성하며, 그림 2와 같이 단계별 모듈이 유기적으로 연동되어 실시간 운영 환경에 대응할 수 있도록 설계하였다.

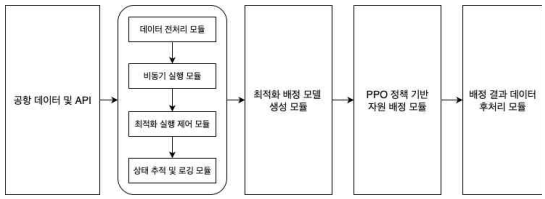


그림 2. PPO 기반 실시간 공항 자원 배정 시스템 구조도

시스템은 공항 데이터 입력 및 API를 통해 자원 배정 요청과 관련 데이터를 직접 입력받으며, 이를 기반으로 내부 처리 흐름을 시작한다. 이후 **데이터전처리모듈**이 입력된 데이터를 정제한다. 이 모듈은 입력 정보를 시계열 기반으로 재구성하고, 최적화 모델에 적용할 수 있도록 필요한 형식

과 구조로 가공한다. 다음으로 **비동기실행모듈**이 각 배정 작업을 독립적으로 병렬 실행하며, **최적화실행제어모듈**은 요청 상태를 관리하고, 필요 시 실행을 중단하거나 재요청할 수 있도록 제어한다. **상태추적및로깅모듈**은 각 단계의 진행 상황을 실시간으로 모니터링하고 기록하여 전체 흐름의 안정성과 투명성을 확보한다. 이렇게 준비된 데이터는 **최적화모델생성모듈**을 통해 수학적으로 문제화된다. 이때 자원 충돌, 수용 조건, 선호도 등 다양한 제약을 포함한 배정 모델을 생성하고, 이어서 **정책기반자원배정모듈**이 사전에 학습된 정책을 호출해 최적의 배정 결과를 도출한다. 마지막으로, **결과후처리모듈**이 배정 결과를 정제하고 응답 형태로 구성하여 공항 운영 시스템에 반환한다. 제안하는 시스템은 각 모듈이 독립적으로 구성되어 있어 유지보수가 용이하며, API 기반으로 연동되어 다양한 운영 환경에 유연하게 적용 가능하다. 특히 데이터 입력부터 결과 반환까지의 전 과정을 실시간으로 자동 처리함으로써, 기존의 정적 계획 방식보다 빠르고 유연한 대응이 가능하다. 제안하는 시스템의 핵심은 자원 배정 문제를 강화학습 환경으로 변환하고, 사전에 학습된 정책을 활용해 실시간으로 배정하는 구조다. 이를 위해 상태-행동-보상 구조의 마르코프 결정 과정(Markov Decision Process, MDP)을 기반으로 학습 환경을 정의하고, 정책 기반 강화학습 기법인 PPO로 자원 배정 정책을 학습하였다. 상태(state)는 항공편의 도착 시간, 기체 종류, 항공사 코드, 사용 가능한 자원 목록 등으로 구성되며, 행동(action)은 특정 항공편에 하나의 자원을 선택하는 방식이다. 보상(reward)은 유효한 배정 여부, 제약조건 충족, 선호 자원 배정 등을 반영해 설계하여 다양한 제약을 내재화하도록 했다. PPO는 학습 안정성과 수렴 속도에서 우수하며, 실시간 추론 환경에 적합하다. 정책 업데이트 시 불안정을 방지하기 위해 클리핑(clipping) 기법을 도입해 이전 정책과 크게 벗어나지 않는 범위에서 점진적으로 성능을 향상시킨다. 아래 수식은 새로운 정책과 이전 정책의 비율에 기반해 어드밴티지 값을 최소화하는 PPO의 핵심 손실 함수다.

$$L^{CLIP}(\theta) = E_t\left[\min\left(r_t(\theta)A_t, clip\left(r_t(\theta), 1-\epsilon, 1+\epsilon\right)A_t\right)\right] \quad (1)$$

수식 (1)에서 $r_t(\theta)$는 새로운 정책과 이전 정책의 비율, A_t는 어드밴티지 추정값, ϵ은 클리핑 계수로 설정되며, 이 수식을 통해 과도한 정책 업데이트를 방지하고 안정적인 학습을 유도한다. 제안하는 학습 구조는 반복적 시뮬레이션을 통해 다양한 상태-행동 조합을 경험하고, 자원 배정 환경의 제약 조건을 보상 함수에 자연스럽게 녹여냄으로써 현실성과 효율성을 동시에 확보한다. 학습이 완료된 정책은 실제 운영 시스템 내에서 호출되어, 입력 정보에 기반한 자원 배정을 실시간으로 수행하게 된다.

실험에는 인천국제공항의 실제 항공편 데이터를 활용하였다. 국토교통부 항공정보포털(UBIKAIS)[5]에서 2025년 3월 21일부터 25일까지 수집한 5,000편의 운항 데이터를 기반으로, 총 228개의 주기장을 포함하는 환경을 구성하였다. 자원 배정 모델은 MILP 기반으로 수립되었으며, 시간 충돌 방지, 단일 자원 배정, 수용성, 선호도 등 네 가지 제약 조건을 함께 고려하였다.

정책 학습은 PPO 알고리즘을 기반으로 총 500 에피소드에 걸쳐 수행되었으며, 학습률 0.001, 할인율 0.99, GAE λ값 0.95, 클리핑 계수 0.2 등의 하이퍼파라미터를 사용하였다. 실험은 5,000편의 운항 데이터를 1,000편 단위로 구분하여 총 5회 수행하였으며, 주요 평가지표는 배정률, 실행 시간, 제약 위반률로 설정하였다. 그 결과, 모든 실험에서 배정률 100%, 1,000편 기준 평균 실행 시간 258초, 제약 위반률 평균 0.27%를 기록하였다. 이는 제안한 정책이 제약 조건을 효과적으로 반영하면서도 실시간 대응 능력을 유지함을 보여준다.

특히 기존 연구[6]에서는 A3C 기반의 REGAPS 모델을 활용하여 124편의 항공편과 14개의 게이트로 구성된 실제 공항 데이터를 대상으로 게이트 배정 문제를 해결하였으나 평균 실행 시간이 약 185초로 측정되어 실시간 운영에 적용하기에는 한계가 있었다. 반면, 본 연구는 동일한 조건에서 PPO 정책을 적용하여 평균 실행시간을 15.1초로 단축하였으며, 기존 연구 대비 현저히 빠른 응답 속도를 달성하였다. 이를 통해 제안한 방식이 실시간 운영 환경에서 우수한 성능을 제공함을 입증하였다. 이는 단일 실행만으로 배정 결정을 수행할 수 있으며, 추론 속도와 실시간성 측면에서 강화된 자동화 구조를 제공한다는 점에서 차별성을 가진다.

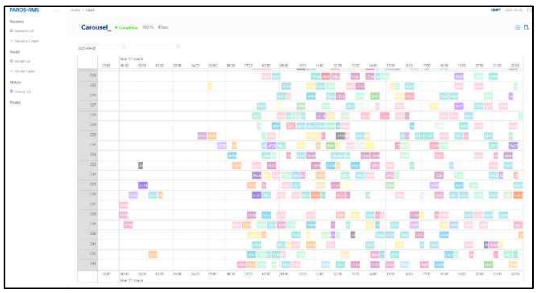


그림 3. PPO를 활용한 항공편-주기장 실시간 배정 결과

Ⅲ. 결론

본 연구는 공항 자원 배정 문제에 대해 정책 기반 강화학습(PPO)을 적용한 실시간 자동화 시스템을 설계·구현하였다. 제안 시스템은 사용자 입력부터 최적화, 정책 호출, 결과 반환까지의 전 과정을 통합 구조로 구성하여 기존 정적 계획 방식의 한계를 극복하였다. 인천국제공항 실제 데이터를 활용한 실험에서 제안 방식은 배정률 100%, 낮은 제약 위반률(0.27%)을 달성하며 실시간 대응이 가능함을 입증하였다. 또한 기존 A3C 기반 기법 대비 추론 속도와 자동화 수준에서 우수한 성능을 보였다. 본 연구는 향후 다양한 운영 시나리오와 예외 상황을 반영한 정책 고도화, 다른 강화학습 알고리즘과의 비교, 다중 자원 동시 배정 문제로의 확장 등 지능형 공항 운영 시스템 발전을 위한 기반을 제시한다.

ACKNOWLEDGMENT

이 논문은 2024년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원 SW컴퓨팅산업원천기술개발사업의 지원을 받아 수행된 연구임 (RS-2024-00398050).

참 고 문 헌

[1] International Air Transport Association (IATA), "20-Year Air Passenger Forecast" 2018, (https://www.iata.org/en/publications/store/20-year-passenger-forecast/)

[2] G. S. Daş, F. Gzara, and T. Stützle, "A Review on Airport Gate Assignment Problems: Single versus Multi Objective Approaches," *Omega*, vol. 92, p. 102146, 2020.

[3] C. H. Papadimitriou and K. Steiglitz, Combinatorial Optimization: Algorithms and Complexity, Mineola, NY: Dover Publications, 1998.

[4] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal Policy Optimization Algorithms," *arXiv preprint*, arXiv:1707.06347, 2017.

[5] UBIKAIS: Korean Air Traffic Information System, (https://ubikais.fois.go.kr:8030)

[6] H. Li, X. Wu, M. Ribeiro, B. Santos, and P. Zheng, "Deep reinforcement learning approach for real-time airport gate assignment," *Operations Research Perspectives*, vol. 14, Article 100338, 2025.

연산 스토리지 기반 전처리 오프로딩을 적용한 AI 컨테이너 오케스트레이션 아키텍처 설계에 관한 연구

길주현, 안재훈*, 김영환
한국전자기술연구원 지능형IDC 사업단
wngus0331@keti.re.kr, *corehun@keti.re.kr, yhkim@keti.re.kr

A Study on the Design of an AI Container Orchestration Architecture with Computational Storage-based Preprocessing Offloading

Juhyun Kil, Jea-Hoon An*, Younghwan Kim
Intelligent IDC Project Office, Korea Electronics Technology Institute

요 약

AI 학습 및 추론 워크로드의 데이터셋 규모 증가로 GPU 메모리 부족, 스토리지 I/O 병목, 운용 비용 증가 문제가 심화되고 있다. 특히 기존 Kubermetes 및 Kubeflow 기반 오케스트레이션은 시간에 따라 변하는 자원 사용량과 대규모 전처리의 네트워크 부하를 충분히 고려하지 못한다. 본 연구에서는 Computational Storage Device(CSD) 기반 인스토리지 전처리 오프로딩을 통합한 AI 컨테이너 오케스트레이션 아키텍처를 제안한다. 제안 구조는 자원 예측, 분석 기반 스케줄링과 CSD 연동 전처리 실행을 통해 AI 파이프라인의 효율성을 높인다. 이를 통해 자원 활용 최적화, 전처리 지연 감소, 비용 절감이 가능하며 대규모 데이터 처리화 학습 작업에서의 성능 향상을 기대할 수 있다.

Ⅰ. 서 론

최근 인공지능(AI) 학습 및 추론 파이프라인은 데이터셋 규모와 복잡성의 급격한 증가로 인해 막대한 컴퓨팅 자원과 효율적인 데이터 전처리 기술을 필요로 한다. 특히, GPU 메모리 부족과 스토리지 입출력(I/O) 병목 현상, 그리고 대규모 데이터 전송에 따른 높은 네트워크 부하는 AI 워크로드의 처리 속도와 비용 효율성에 심각한 제약으로 작용한다. 예를 들어, 시뮬레이션 및 이미지 처리와 같은 순수·응용과학 및 생명공학 분야, 대형 AI 모델 학습을 수행하는 AI 서비스 기업 등의 환경에서는 이러한 성능 저하와 스케일링 한계가 연구 및 서비스 품질 저하로 직결되고 있다. [1]

현재 널리 사용되는 Kubernetes나 Kubeflow 같은 컨테이너 오케스트레이션 플랫폼은 확장성과 자동화된 자원 관리를 지원하지만, 동적인 자원 수요 변동성과 GPU 공유 스케줄링, 대규모 전처리 과정에서의 네트워크 및 스토리지 I/O 병목 문제를 충분히 반영하지 못한다. [2] 이로 인해 데이터 전처리 단계에서 시스템 병목 현상이 발생하여, 전체 AI 파이프라인의 처리 지연과 자원 낭비가 발생하고 있다. 또한 기존 스케줄러들은 자원 사용 예측 및 정책 기반 최적화 기능이 미흡하여, AI 컨테이너의 효율적인 생명주기 관리에 한계를 보이고 있다. [3]

이러한 한계를 극복하고자 본 연구는 Computational Storage Device(CSD) 기반 인스토리지 전처리 오프로딩을 통합한 AI 컨테이너 오케스트레이션 아키텍처를 제안한다. 아키텍처는 마스터 노드의 자원분석 엔진, 자원 사용량 예측 엔진, 오케스트레이션 정책 최적화 엔진, 클러스터 오케스트레이터로 구성된다. 전처리 작업은 Kubernetes의 CRD(Custom Resource Definition)을 통해 정의되며, 워커 노드의 전처리 매니저가 이를 CSD 내부 도커 런타임 기반 전처리 모듈을 통해 실행된다. 이 방식은 전처리를 스토리지 내에서 수행해 네트워크 부하와 I/O 지연을 줄이며, 산업·연구 환경 전반에서의 비용 절감을 실현할 수 있다. [4]

Ⅱ. 본론

1. 제안하는 오케스트레이션 아키텍처

기존 시스템은 동적인 자원 수요 변동, GPU 메모리 공유 한계, 대규모 전처리 시 발생하는 네트워크·스토리지 병목으로 인한 AI 워크로드 처리 효율이 저하되는 한계가 있다. 이에 자원 분석·예측 기반 스케줄링과 CSD 기반 인스토리지 전처리 오프로딩을 통합한 AI 컨테이너 오케스트레이션 아키텍처를 정의하였다.

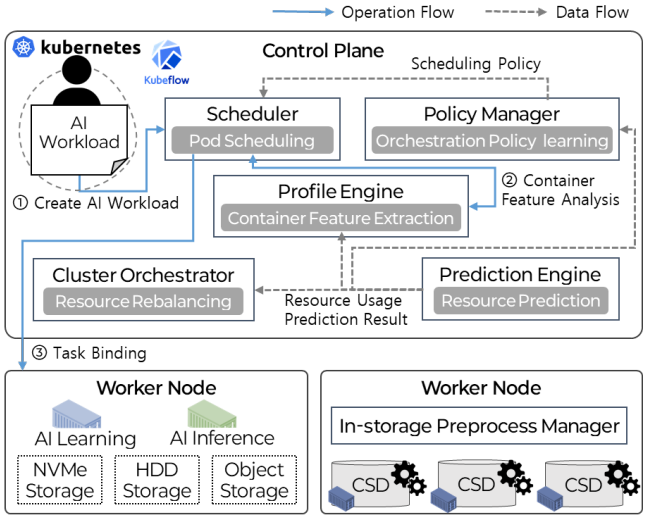


그림 1 AI 컨테이너 오케스트레이션 아키텍처

제안 아키텍처는 Kubernetes 컨테이너 오케스트레이터 위에 Kubeflow 환경이 구축된 클러스터 상에 구현되며, 클러스터는 컨트롤 플레인과 다수의 워커 노드로 구성된다. 컨트롤 플레인에는 클러스터 관리

* 교신저자

와 컨테이너 오케스트레이션을 담당하는 핵심 컴포넌트들이 위치하며, 워커노드에서는 전처리·학습·추론 등 다양한 AI 워크로드가 실행된다.

AI 컨테이너 스케줄링 과정은 다음과 같다. 사용자가 AI 작업을 수행하는 컨테이너를 정의해 요청하면 Custom Scheduler가 스케줄링이 필요한 작업을 감지한다. Custom Scheduler는 GPU 메모리 공유 스케줄링을 지원하여 자원을 효율적으로 활용하며, Profile Engine이 AI 작업 특성과 입력 데이터 특성을 분석하여 자원 요구량을 산출하고 이를 기반으로 배치 결정을 수행한다. 스케줄링 정책은 Policy Manager가 클러스터 상태에 맞추어 학습하고 최적화하며, 워커 노드에 배치된 AI 작업은 실행 중에도 시간에 따른 자원 사용량 변화가 모니터링된다.

Cluster Orchestrator는 클러스터 내 전반적인 리소스를 조정하며, 실행 중인 AI 워크로드를 포함해 노드 간 자원 활용 균형을 유지한다. 또한 Prediction Engine은 클러스터 내 모든 작업의 향후 자원 사용량을 예측하여 Cluster Orcestrator·Policy Engine·Policy Manager 등에 결과를 제공함으로써 자원 최적화를 지원한다.

2. CSD(Computational Storage Device)

CSD는 저장장치에 내장된 프로세서를 통해 데이터를 저장하는 동시에, 저장장치 내부에서 직접 데이터 처리가 가능한 차세대 스토리지 기술이다. 기존 NVMe SSD와 유사하게 서버에 장착되지만, 자체 CPU 및 운영체제가 내장되어 있어 사용자가 데이터 이동 없이 저장장치 내에서 AI 전처리, 머신러닝 인퍼런스 등의 작업을 직접 실행할 수 있다. [5] 본 논문에서는 이러한 CSD 내부에 도커 컨테이너 런타임 환경을 구비하여, 데이터 전처리 작업을 컨테이너화된 형태로 저장장치 안에서 직접 수행하도록 설계하였다. 이는 전처리 단계에서 네트워크 부하와 데이터 이동 지연을 최소화함으로써 전체 AI 워크로드의 성능을 향상하고, 자원 활용의 효율성을 극대화하는 데 기여한다.

3. 전처리 오프로딩 아키텍처

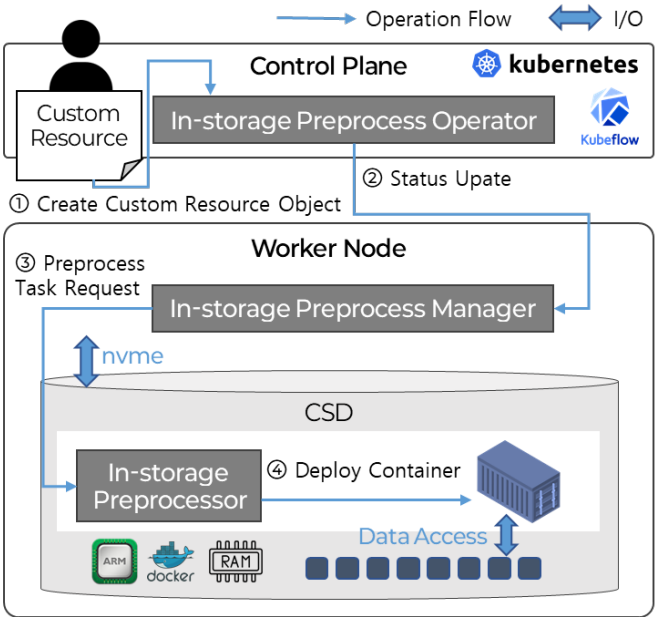


그림 2 전처리 컨테이너 오프로딩 아키텍처

AI 컨테이너의 전처리 오프로딩 작업은 Kubernetes의 CRD(Custom Resource Definition)로 정의 및 등록된다. 이 CRD는 In-stroage Preprocess Operator가 관리하며, CSD 환경에 맞춘 스케줄링 및 배치를 지원한다. CRD에는 작업 메타정보, 컨테이너 이미지, 데이터 위치, 스케줄링 힌트 등이 포함된다.

워크플로우는 먼저 사용자가 오프로딩할 전처리 작업을 CRD 오브젝트로 정의하는 것으로 시작된다. Control Plane의 In-storage Preprocess Operator는 이를 감지하고, CSD가 탑재된 워커 노드 중 작업 수행에 적합한 노드를 선정하여 작업 요청을 전달한다. 워커 노드 내 In-storage Preprocess Manager는 컨테이너 실행 정보를 수신하고 이를 CSD로 전달한다. CSD 내부의 In-storage Preprocessor는 Docker 컨테이너 런타임을 기반으로 전처리 컨테이너를 실행한다.

이 과정은 저장장치 내부 연산을 활용해 데이터 이동 및 네트워크 부하를 최소화하며, 입출력 지연을 줄여 자원 활용도와 전체 AI 워크로드 처리 성능을 최적화한다.

Ⅲ. 결론

본 논문에서는 기존 AI 컨테이너 오케스트레이션 환경에서 나타나는 동적 자원 수요 변동, GPU 메모리 공유의 한계, 대규모 전처리 단계의 네트워크 및 스토리지 병목 문제를 해결하기 위한 자원 분석·예측 기반 스케줄링과 CSD 기반 인스토리지 전처리 오프로딩을 통합한 아키텍처를 제안하였다. 제안 아키텍처는 Kubernets/Kubeflow 기반 클러스터에서 Profile Engine, Prediction Engine, Policy Manager, Custom Scheduler 등을 통해 최적의 자원 배치 결정을 수행하고, 전처리 작업을 CRD를 통해 CSD 내부에서 실행함으로써 데이터 이동을 최소화하였다.

이를 통해 네트워크 부하 및 스토리지 I/O 지연 감소, GPU 및 CPU 자원의 효율적 활용, AI 파이프라인 전반의 처리 지연 최소화, 운영 비용 절감 등의 효과를 기대할 수 있다. 또한 제안 구조는 기상 예측, 바이오 데이터 분석, 스마트 팩토리, 클라우드 AI 서비스, 자율주행 등 다양한 산업 및 연구 분야에 적용 가능하다.

향후 연구에서는 실제 시스템 구현 및 대규모 워크로드에 대한 실험을 통해 제안 아키텍처의 성능을 정량적으로 검증하고, 자원 예측 모델의 정확도 향상과 오케스트레이션 정책 최적화를 위한 강화학습 기반 기법을 추가로 적용할 계획이다.

ACKNOWLEDGMENT

이 논문은 2024년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구 결과임 (No.RS-2024-00461572, AI 연산 가속기 최적화 고효율 병렬 스토리지 SW 기술 개발)

참 고 문 헌

[1] Rodriguez, R. & Buyya, R., Machine Learning-based Orchestration of Containers: A Taxonomy and Future Directions, arXiv, 2021.

[2] Kaul, D., AI-Driven Self-Healing Container Orchestration Framework for Energy-Efficient Kubernetes Clusters, Emerging Science Research, 2024.

[3] Rothman, A. et al., Reinforcement Learning Based Kubernetes Scheduler for Efficient Resource Utilization, arXiv, 2025.

[4] Smith, J. et al., Resource-aware multi-task offloading architectures for edge AI, ScienceDirect, 2023.

[5] NGD Systems, Newport In-Situ Development Platform – Hardware Product Brief, Apr. 2019.

다중 테넌트 환경에서 GPU 커널 스케줄링 기반 적응형 자원 할당 시스템 연구

권용진, 안재훈*, 김영환
한국전자기술연구원 지능형 IDC 사업단
(rrnjs0721, corehun, yhkim)@keti.re.kr

A Study on Adaptive Resource Allocation Systems Based on GPU Kernel Scheduling in Multi-Tenant Environments

YongJin Kwon, JaeHoon An*, YoungHwan Kim
Intelligent IDC Project Office, Korea Electronics Technology Institute

요 약

현대 클라우드에서는 여러 사용자가 GPU를 공유하는 다중 테넌트 환경이 일반화되면서, 효율적인 GPU 자원 할당과 성능 보장이 중요한 과제로 떠오르고 있다. 하지만 기존 GPU 스케줄링 기법들은 주로 단일 고정 정책만을 사용하여 다양한 워크로드 특성과 실시간 사용자 요구 변화에 효과적으로 대응하지 못했다. 본 연구에서는 사용자 정의 정책에 따라 실행 중에도 동적으로 GPU 커널 스케줄링을 조정함으로써 이러한 한계를 극복하는 적응형 다중 테넌트 GPU 스케줄링 시스템을 제안한다. 제안 시스템은 JCT최소화, SLO보장, 하이브리드 최적화의 세 가지 핵심 정책 모드를 단일 프레임워크에서 지원하며, Bubble-less 커널 스케줄링과 시공간 하이브리드 자원 분할 기법을 결합하여 GPU 유휴 시간을 제거하고 자원 활용률을 극대화한다. 또한 실시간 모니터링 및 우선순위 조정을 통해 워크로드 실행 도중에도 정책 목표 달성을 위해 정책 전환과 자원 재할당을 수행한다. 이러한 동적 스케줄링 접근법은 기존 정적 방식의 한계를 극복하여, 단일 GPU 내에서 여러 테넌트의 다양한 요구사항을 유연하고 효율적으로 만족시킬 수 있음을 보인다.

Ⅰ. 서 론

GPU 컴퓨팅의 급속한 발전과 함께 클라우드 환경에서의 다중 테넌트 GPU 활용이 필수적인 요구사항으로 자리잡았다. 그러나 현재 생산 환경에서 GPU 활용률은 평균 52%에 머물러 있으며, 이는 비효율적인 자원 할당과 워크로드 간 간섭으로 인한 성능 저하가 주요 원인이다. [1]

기존 GPU 스케줄링 연구들은 주로 단일 최적화 목표에 집중되어 왔다. NVIDIA MPS, MIG[2]와 같은 하드웨어 기반 솔루션은 정적 자원 분할로 인한 유연성 부족 문제를 겪고 있으며, Bless[3], Gpulet[4]와 같은 소프트웨어 기반 접근법들도 고정된 정책 적용으로 인해 다양한 워크로드 특성 변화에 적응하지 못하는 한계를 보인다.

현재 시스템들이 직면한 핵심 문제점은 세 가지로 요약된다. 첫째, 시스템 초기화 시점에 정해진 단일 정책만 사용하여 다양한 워크로드 특성 변화에 유연하게 대응하지 못한다. 둘째, JCT 최소화, SLO 보장, 공정성 등 서로 다른 목표를 가진 사용자들의 요구사항을 동시에 만족시키기 어렵다. 셋째, 워크로드 실행 중 성능 목표나 우선순위가 변경되어도 이를 반영할 수 있는 동적 조정 메커니즘이 부족하다.

본 연구는 이러한 문제점들을 해결하기 위해 사용자 정의 정책에 따라 동적으로 커널 스케줄링을 수행할 수 있는 적응형 다중 테넌트 GPU 스케줄링 시스템을 설계한다. JCT 최소화, SLO 보장, 하이브리드 최적화라는 세 가지 핵심 정책을 단일 시스템에서 동적으로 전환할 수 있는 다중 정책 지원 아키텍처를 설계한다. Bubble-less 커널 스케줄링을 결합하여 워크로드 특성 변화와 사용자 요구사항 변경에 즉시 대응할 수 있는 실시간 정책 적응 메커니즘을 개발한다.

Ⅱ. 본론

본 논문에서 제안하는 적응형 다중 테넌트 GPU 스케줄링 시스템은 정책 관리 계층, 스케줄링 엔진 계층, 자원 할당 계층으로 구성된 계층적 구조로 설계된다.

정책 관리 계층은 사용자가 정의한 성능 목표와 제약 조건을 해석하고 적절한 스케줄링 전략을 선택한다. 세 가지 핵심 정책은 독립적인 모듈로 구현되어 런타임에 동적으로 전환될 수 있다.

스케줄링 엔진 계층은 선택된 정책에 따라 실제 커널 스케줄링을 수행한다. Bubble-less 커널 스쿼드 생성, 우선순위 기반 작업 선택, 동적 정책 적응의 세 가지 주요 기능을 제공한다.

자원 할당 계층은 물리적 GPU 자원을 논리적으로 분할하고 할당한다. SM 할당, 메모리 분할, 대역폭 관리를 통해 워크로드 간 격리를 보장하면서도 최대한의 자원 활용률을 달성한다.

1. 다중 정책 지원 메커니즘

JCT 최소화 정책은 수정된 SJF 방식을 사용하되, 예상 완료 시간과 자원 공유 가능성을 종합적으로 고려한다. 작업 간 간섭 모델링을 통해 동시 실행 가능한 작업 조합을 식별하고, 전체 시스템 처리량을 최대화한다. 실행 중 성능 모니터링을 통해 예상과 실제 진행 상황의 차이를 추적하여 스케줄링 순서를 동적으로 재조정한다.

SLO 보장 정책은 지연 시간, 처리량, 가용성이라는 세 가지 핵심 SLO 메트릭을 지원한다. 실시간 성능 모니터링을 통해 SLO 위반 위험을 감지하고, 해당 작업의 우선순위를 동적으로 조정한다. 중요도가 높은 작업에 대해서는 자원 예약 메커니즘을 적용하여 시스템 부하 증가 시에도 최소한의 서비스 수준을 유지한다.

하이브리드 최적화 정책은 Sia 스케줄러의 설계 철학을 차용하여 처리량과 공정성을 동시에 고려한 전체 Goodput 최대화를 목표로 한다.[5] 이를 위해 수식1의 다중 모델 간섭을 고려한 실효 처리량 예측 모델을 도입한다. 여기서 T_m^{eff}는 모델 m이 다른 모델들과 동시 실행될 때의 실제 처리량이며, p_{mk}는 모델 간 간섭 계수를 나타낸다. 이는 기존의 단순한 SM

분할 방식과 달리 메모리 대역폭 경합 등의 실제 간섭 효과를 정량화하여 보다 정확한 성능 예측을 가능하게 한다.

$$T_m^{eff}(s_m, C) = T_m^{solo}(s_m) \times \Pi_{k \in C, k \neq m}(1 - \rho_{mk} \times \phi(s_m, s_k))$$

[수식1] 다중 모델 간섭을 고려한 처리량 예측

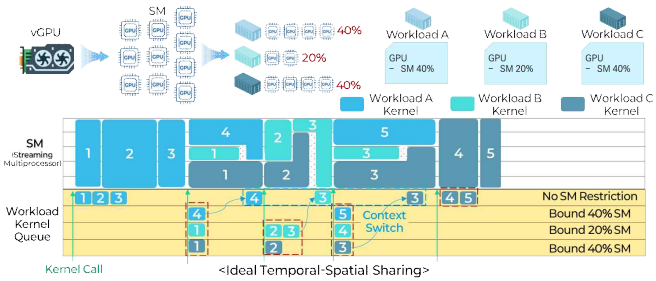
실효 처리량 모델을 기반으로 SLO 달성도를 고려한 다중 모델 최적화 목적 함수를 구성한다. 수식2의 목적 함수는 각 모델의 가중 Goodput의 합을 최대화하되, SLO 달성도 기반 품질 보정 함수 Q_m을 통해 단순 처리량이 아닌 서비스 품질을 반영한 실질적 성능 지표를 사용한다. 가중치 기반 목적함수를 통해 JCT 최소화와 SLO 보장 간의 트레이드오프를 관리한다. 이를 통해 이질적인 워크로드 환경에서도 안정적인 성능을 제공한다.

$$\max_{s_1,\ldots,s_M} \sum_{m=1}^{M} w_{m(t)} \times T_m^{eff}(s_m, C) \times Q_{m(SLO_{m(t)})}$$

[수식2] SLO 인식 다중 모델 Goodput 최적화

2. Bubble-less 커널 스쿼드 메커니즘

Bubble-less 커널 스쿼드는 GPU 자원의 유휴 시간을 최소화하기 위해 서로 다른 작업의 커널들을 효율적으로 조합한다. 스쿼드 생성 알고리즘은 커널 실행 시간 예측, 자원 요구량 분석, 간섭 영향 평가를 통해 호환성 매트릭스를 생성하고, 현재 적용된 정책의 목표에 따라 최적의 커널 조합을 선택한다.



[그림1] 커널 스쿼드 기반 GPU 커널 스케줄링

자원 할당은 SM 개수, 메모리 용량, 대역폭이라는 세 가지 주요 자원을 관리한다. SM 할당은 공간적 분할과 시간적 공유를 동적으로 조합하고, 메모리 할당은 각 작업의 사용 패턴을 분석하여 최적화한다. SLO 보장이 필요한 작업에게는 별도의 메모리 영역을 예약하여 성능 예측 가능성을 높인다.

스쿼드 실행 중 작업 완료로 인해 자원이 해제되면, 남은 작업들에게 즉시 재분배하여 bubbles를 최소화한다. 이는 그림에서 보이는 바와 같이 기존 방식에서 발생하는 자원 낭비를 효과적으로 제거한다. 이러한 메커니즘을 통해 전체 GPU 활용률을 극대화하면서도 개별 작업의 성능 요구사항을 만족시킬 수 있다.

3. 실시간 정책 적응 시스템

실시간 정책 적응을 위해 하드웨어 성능 카운터, 소프트웨어 메트릭, 사용자 피드백을 통합한 다층 모니터링 구조를 채택한다. 모니터링 데이터는 시계열 데이터베이스에 저장되어 패턴 분석과 예측에 활용되며, 기계학습 기반 이상 감지 알고리즘을 통해 성능 저하나 SLO 위반 위험을 사전에 감지한다. 수집된 모니터링 데이터를 바탕으로 수식3의 실시간 가중치 업데이트 메커니즘을 적용한다.

$$w_{m(t+1)} = w_{m(t)} \times \exp(\alpha \times (SLO_{target}^{m} - SLO_{actual}^{m(t)}))$$

[수식3] 실시간 정책 적응을 위한 가중치 업데이트

$w_{m(t+1)}$는 다음 스쿼드에서 적용할 모델 m의 우선순위 가중치이며, $SLO_{actual}^{m(t)}$는 지연 시간과 처리량을 종합한 현재 SLO 달성률을 나타낸다. 가중치 조정을 통해 SLO 위반 위험이 감지된 모델의 우선순위를 즉시 상승시켜 다음 커널 스쿼드에서 더 많은 자원을 할당받도록 한다.

Ⅲ. 결론

본 논문에서는 사용자 정의 정책에 따라 동적으로 커널 스케줄링을 수행할 수 있는 적응형 다중 테넌트 GPU 스케줄링 시스템을 설계하였다. 제안 시스템의 핵심 기여는 기존 정적 스케줄링 방식의 한계를 극복하고 다양한 사용자 요구사항을 동시에 만족시킬 수 있는 유연한 프레임워크를 제공한다는 점이다. JCT 최소화, SLO 보장, 하이브리드 최적화라는 세 가지 핵심 정책을 단일 시스템에서 지원함으로써 사용자는 자신의 요구사항에 가장 적합한 최적화 전략을 선택할 수 있다.

Bubble-less 커널 스쿼드와 정책 기반 우선순위 조정을 결합한 통합 최적화 접근법은 기존 시스템 대비 향상된 성능과 예측 가능성을 제공할 것으로 기대된다. 특히 다중 테넌트 환경에서 워크로드 간 격리를 보장하면서도 최대한의 자원 활용률을 달성할 수 있는 구조를 제시했다.

향후 연구에서는 실제 구현과 성능 평가를 통해 이론적 설계의 유효성을 검증할 예정이다. 기계학습 기반 예측 모델의 정확도 향상, 다중 GPU 클러스터 환경으로의 확장, 보안과 프라이버시 측면에서의 추가 연구가 필요하다.

ACKNOWLEDGMENT

이 논문은 2025년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구 결과임 (No.RS-2025-02220502, AI 반도체 컴퓨팅 자원분해 및 자원풀링 기술 개발)

참 고 문 헌

[1] Jeon M. et al. "Analysis of Large-Scale Multi-Tenant GPU Clusters for DNN Training Workloads," 2019 USENIX Annual Technical Conference (USENIX ATC 19), pp. 947-960, July 2019.

[2] NVIDIA Corporation, "Multi-Instance GPU User Guide," NVIDIA Documentation, 2025, (https://docs.nvidia.com/datacenter/tesla/mig-user-guide/).

[3] Zhang S. et al. "Improving GPU Sharing Performance through Adaptive Bubbleless Spatial-Temporal Sharing," Proceedings of the Twentieth European Conference on Computer Systems (EuroSys 25), pp. 573-588, March 2025.

[4] Choi S. et al. "Serving Heterogeneous Machine Learning Models on Multi-GPU Servers with Spatio-Temporal Sharing," 2022 USENIX Annual Technical Conference (USENIX ATC 22), pp. 199-216, July 2022.

[5] Subramanya S. J. et al. "Sia: Heterogeneity-aware, goodput-optimized ML-cluster scheduling," Proceedings of the 29th Symposium on Operating Systems Principles (SOSP 23), pp. 642-657, October 2023.

AI 반도체 자원 풀링을 위한 시공간 통합 관리 아키텍처에 관한 연구

안성모, 안재훈, 김영환
한국전자기술연구원 지능형 IDC 사업단
(seongmo619, corehun, yhkim)@keti.re.kr

A Study on the Spatio-Temporal Integrated Management Architecture for AI Semiconductor Resource Pooling

Seongmo Ahn, JaeHoon An, YoungHwan Kim
Intelligent IDC Project Office, Korea Electronics Technology Institute

요 약

거대언어모델(LLM)과 같은 대규모 AI 모델의 확산은 기존 데이터센터의 정적 자원 할당 방식의 비효율성을 드러내고 있다. 고가의 AI 가속기 활용률은 저조하고 운영 비용은 증가하는 문제가 발생하고 있다. 본 논문은 이러한 한계를 극복하기 위해 설계된 '시공간 통합 컴퓨팅 자원 관리 아키텍처'를 제안한다. 본 아키텍처는 워크로드 프로파일링, 시공간 자원 분해, 동적 자원 공유 비율 변경이라는 세 가지 핵심 기술을 기반으로 한다. 구체적으로, 예측과 실시간 보정을 결합한 프로파일링으로 워크로드 요구사항을 분석하고, 물리적 AI 가속기를 미세 단위의 논리 자원으로 분해하여 시공간적으로 공유하며, 실시간 변화에 맞춰 자원 할당을 동적으로 최적화한다. 클라우드 네이티브 쿠버네티스(Kubernetes) 환경 내에서 다양한 LLM 워크로드에 대한 자원 활용률과 성능을 극대화함으로써, 본 시스템은 모놀리식(monolithic) 인프라의 경직성을 탈피하는 유연하고 효율적인 대안을 제시한다.

Ⅰ. 서 론

현대의 AI 인프라는 고성능 가속기(GPU, NPU 등)가 호스트 CPU 및 메모리에 물리적으로 종속된 모놀리식 서버 설계가 지배적이다. 이러한 경직된 결합 구조는 심각한 자원 낭비를 초래한다. [1] 예를 들어, 메모리 중심적(memory-bound) 작업은 연산 코어를 유휴 상태로 남겨두고, 연산 중심적(compute-bound) 작업은 메모리 대역폭을 충분히 활용하지 못하는 자원 고립 현상이 발생한다. 특히 LLM의 학습, 추론 등 다양하고 동적인 워크로드 요구사항에 정적 할당 방식은 효과적으로 대응하기 어려워, 고가의 AI 하드웨어 도입에 따른 총소유비용(TCO) 대비 효율성이 저하되는 문제를 안고 있다. [2]

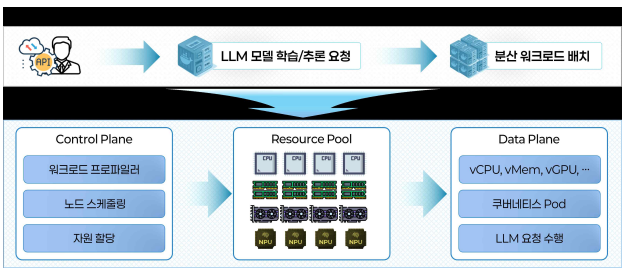
이러한 문제에 대한 해결책으로 '자원 분해' 기술이 부상하고 있다. 이는 연산, 메모리 , I/O 자원을 논리적, 물리적으로 분리하여 네트워크로 연결된 자원 풀(pool)로 구성하는 패러다임이다. 이 기술은 애플리케이션 요구에 따라 필요한 자원만 정확하게 조합하여 가상 서버를 동적으로 구성하는 인프라 구조를 가능하게 한다.

본 논문은 자원 분해 환경에서 분산된 AI 자원을 효과적으로 오케스트레이션하기 위한 소프트웨어 정의 관리 아키텍처를 상세히 기술한다. 이 아키텍처의 독창성은 자원을 공간과 시간의 두 축에서 통합적으로 관리하여 효율성을 극대화하고 다양한 워크로드의 서비스 수준 목표를 충족시키는 데 있다. 본 시스템의 핵심 기술은 (1) 워크로드 프로파일링, (2) 시공간 자원 분해, (3) 동적 자원 공유 비율 변경 기술로 구성된다. 이는 단순한 자원 풀 관리를 넘어, 워크로드 특성을 지능적으로 분석하여(워크로드 프로파일링), 자원을 미세 단위로 분해하고 시공간적으로 공유하며(시공간 자원 분해), 실시간 변화에 대응하여 자원 할당을 동적으로 최적화하는(동적 자원 공유 비율 변경) 지능형 오케스트레이션을 통해 자원 가용성을 워크로드 중심의 최적화로 전환한다.

Ⅱ. 본론

2.1 시스템 아키텍처

본 시스템은 의사결정을 담당하는 컨트롤 플레인(Control Plane)과 실제 워크로드를 실행하는 데이터 플레인(Data Plane)으로 구성된다. 관리 아키텍처는 컨트롤 플레인에 위치하며, 워크로드 분석, 스케줄링, 자원 할당의 전 과정을 책임진다. 사용자로부터 LLM 작업 요청이 들어오면, 컨트롤 플레인은 해당 요청을 프로파일링하고, 자원 풀에서 필요한 자원을 할당한 수, 데이터 플레인의 가상 AI 가속기(vGPU, vNPU 등)에서 실행되도록 작업을 분배한다. 이 구조는 관리 로직과 실행 환경을 명확히 분리하여 시스템의 모듈성과 확장성을 보장한다.



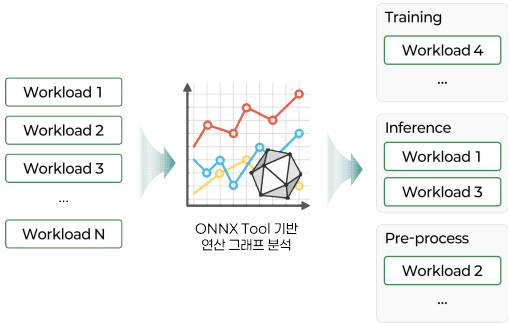
[그림 1] AI 반도체 자원 풀링 아키텍처 개요

2.2 워크로드 프로파일링

워크로드 프로파일링 기술은 연산 그래프 분석 결과를 기반으로 워크로드의 타입을 식별하고, 그 결과를 바탕으로 연산 요구 사항과 실행 패턴을 분석하여 최종적으로 워크로드의 우선순위를 결정하는 핵심 기술이다.

프로세스는 먼저 연산 그래프 분석을 통해 제출된 작업이 학습(Training), 추론(Inference), 전처리(Pre-process)인지 등 타입을 식별하

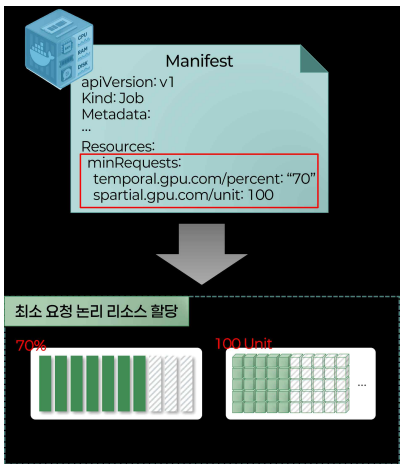
는 것에서 시작한다. 다음으로, 식별된 타입을 근거하여 작업이 연산 중심적(Compute-bound)인지, 메모리 중심적(Memory-bound)인지 등 구체적인 연산 요구 사항을 분석한다. 이어서 반복성, 지연 시간 민감도, 처리량 요구 등 워크로드의 고유한 실행 패턴을 파악한다. 최종적으로, 이러한 다각적인 분석 결과를 종합하여 워크로드의 우선순위를 결정한다. 이 과정을 통해 시스템은 각기 다른 특성을 가진 워크로드에 대해 최적의 자원을 선별적으로 할당할 수 있게 된다.



[그림 2] 연산 그래프 분석을 통한 워크로드 프로파일링

2.3 시공간 자원 분해

본 아키텍처의 근간은 물리적 하드웨어를 미세 단위의 논리 자원으로 분해하는 능력에 있다. [3][4] 예를 들어, 하나의 물리 GPU는 컴퓨팅 유닛에 대한 시분할 점유율을 나타내는 *temporal.gpu.com/percent*로 추상화된다. 이렇게 정의된 논리 자원들은 Custom Device Plugin을 통해 쿠버테니스 스케줄러에 등록된다. 이러한 접근 방식은 DevOps팀이나 오케스트레이션 시스템이 복잡한 공유 메커니즘을 알 필요 없이, 표준 쿠버네티스 메니페스트를 사용해 미세 단위의 분할된 AI 자원을 요청할 수 있게 한다.



[그림 3] 하드웨어 자원 미세 단위 분해

이렇게 분해된 자원들을 관리하는 시공간 공유 엔진은 자원 할당 메커니즘의 핵심으로, 여러 워크로드가 단일 물리 가속기를 동시에 공유할 수 있게 한다. 시간적 공유(Temporal Sharing)는 워크로드들을 시간 축에서 다중화하여 지연 시간에 민감한 추론 작업과 처리량 중심의 학습 작업을 교차 실행하여 전체 처리량을 극대화한다. 공간적 공유(Spatial Sharing)는 NPU 코어나 가속기 메모리(VRAM)와 같은 자원을 격리된 세그먼트로 분할하여 각 워크로드에 전담시켜 멀티테넌트 환경의 안정성을 확보한다.

2.4 동적 자원 공유 비율 변경 기술

본 기술의 핵심은 여러 워크로드가 단일 물리 가속기를 효율적으로 공유하도록 하는 이상적인 시공간 공유 모델에 있다. [5] 이는 시간 공유 기술과 공간 공유 기술의 유기적인 결합을 통해 자원 사용률을 극대화한다. 먼저, 워크로드 커널 그룹 실행 기반 시간 공유 기술은 개별 커널 호출마다 컨텍스트를 전환하는 대신, 여러 커널을 하나의 그룹으로 묶어 실행함으로써 잦은 컨텍스트 전환(Context Switch)으로 인한 오버헤드를 최소화한다. 이를 통해 지연 시간에 민감한 추론 작업과 처리량 중심의 학습 작업을 효과적으로 교차 실행하여 전체 시스템의 처리량을 높인다.

다음으로, 공간 자원 재구성 기반 공간 공유 기술은 물리적 자원을 공간적으로 분할한다. 예를 들어, GPU의 SM과 VRAM을 재구성 가능한 독립된 논리적 파티션으로 나누고, 각 워크로드에 특정 비율(예: Workload A에 SM 40%)을 할당한다.

이 두 기술을 통합한 시공간 공유를 통해, 서로 다른 공간적 자원 요구사항을 가진 워크로드들이 시간의 흐름에 따라 빈틈없이 교차 실행되면서 유휴 자원을 최소화하고 가속기 전체의 활용률을 극대화한다. 이러한 공간적 분할 비율과 시간적 실행 순서는 고정된 것이 아니라, 워크로드 프로파일링 결과와 실시간 모니터링 데이터를 기반으로 동적으로 변경된다. 이를 통해 하드웨어 자원의 유연한 제어가 가능하다.

Ⅲ. 결론

본 논문은 분해된 AI 가속기를 위한 새로운 시공간 통합 관리 아키텍처를 제안한다. 워크로드 프로파일러, 쿠버네티스 네이티브 자원 추상화, 그리고 동적 제어 메커니즘을 통해, 본 아키텍처는 AI 자원 활용률을 크게 향상시키고, 스케줄링 지연을 단축하며, LLM 워크로드에 대한 데이터 센터의 총소유비용을 절감할 것으로 기대된다. 이 기술은 모놀리식 AI 인프라의 한계를 극복하고 차세대 AI 데이터센터로 나아갈 경로를 제시한다. 향후 본 기술의 개발 로드맵은 단일 노드 프로토타입에서 시작하여 다중 노드의 이기종 환경으로 점진적으로 확장하는 것을 목표로 한다. 주요 과제는 다양한 제조사의 GPU, NPU, PIM 등을 포함하는 이기종 자원 풀을 통합 관리하고, CXL 기반 메모리 풀링과 연동하여 완벽하게 분해된 컴퓨팅 환경을 구축하는 것이다. 최종적으로 개발된 아키텍처를 대규모 멀티-랙 환경으로 확장하여 차세대 AI 데이터센터의 참조 모델을 제공하고자 한다.

ACKNOWLEDGMENT

이 논문은 2025년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구 결과임 (No.RS-2025-02220502, AI 반도체 컴퓨팅 자원분해 및 자원풀링 기술 개발)

참 고 문 헌

[1] M. Jeon, S. Venkataraman, A. Phanishayee, J. Qian, W. Xiao, and F. Yang, “Analysis of Large-Scale Multi-Tenant GPU Clusters for DNN Training Workloads,” in Proc. USENIX Annu. Tech. Conf. (USENIX ATC), Renton, WA, USA, Jul. 2019, pp. 947 - 960.

[2] Z. Ye, S. Chen, T. Yu, and X. Wu, “Deep Learning Workload Scheduling in GPU Datacenters: A Survey,” ACM Comput. Surv., vol. 56, no. 6, art. 146, pp. 1 - 38, Jun. 2024.

[3] F. Strati, A. Sotiropoulos, P. Pandey, M. S. Andersen, S. Z. György, and A. Katevenis, “Orion: Interference-aware, Fine-grained GPU Sharing for ML Applications,” in Proc. Eur. Conf. Comput. Syst. (EuroSys), Athens, Greece, Apr. 2024, pp. 18:1 - 18:18.

[4] P. Yu and M. Chowdhury, “Salus: Fine-Grained GPU Sharing

Primitives for Deep Learning Applications," in Proc. MLSys Conf., Austin, TX, USA, Mar. 2020, pp. 1 – 14.

[5] Xianzhe Ma et al., "Towards Efficient and Practical GPU Multitasking in the Era of LLM," arXiv preprint 2508.08448, 2025

멀티 클러스터 환경에서 AI 워크로드 GPU/NPU 자원 할당과 운용 비용 최적화

조성민, 안재훈*, 김영환
한국전자기술연구원
{another0306, corehun, yhkim}@keti.re.kr

GPU/NPU Resource Allocation and Operational Cost Optimization for AI Workloads in Multi-Cluster Environments

Sungmin Cho, Jaehoon An*, Younghwan Kim
Intelligent IDC Project Office, Korea Electronics Technology Institute

요 약

최근 초거대 인공지능 모델의 등장과 함께 클라우드 인프라 비용이 폭증하고 있다. 실제로 크라우드 AI 워크로드에 대한 지출은 2025년까지 연 평균 35%를 증가하여 360억 달러에 이를 것으로 전망된다. 이에 따른 학습, 추론, 서빙 등 다양한 AI 워크로드의 특성에 맞추어 GPU/NPU 등 가속기 자원을 효율적으로 배치함으로써 클라우드 운용비용의 최적화하려는 요구가 높아지고 있다. 본 논문에서는 쿠버네티스(Kubernetes) 기반 멀티 클러스터 환경에서 AI 워크로드의 비용 효율적인 자원 할당 전략을 제안한다. 제안하는 시스템 구조는 중앙 Control Plane을 통해 이기종 GPI/NPU 자원을 통합 관리하며, 워크로드별 스케줄링 정책을 적용하여 AI 가속기 자원 활용도를 극대화하고 유휴 자원을 최소화한다. 이를 통해 대규모 AI 모델 서비스의 비용 부담을 완화하고, 클라우드 환경에서 AI 모델을 효율적으로 운영이 가능할 것으로 기대된다.

Ⅰ. 서론

인공지능 모델의 규모와 활용이 폭발적으로 증가하면서 클라우드 인프라 비용이 기업과 연구기관에 큰 부담이 되고 있다. 특히 최신 딥러닝 모델의 학습과 추론에는 대용량 GPU 등 전용 가속기 자원이 필수적이며, 이러한 자원은 매우 고가로서 클라우드 비용의 상당 부분을 차지한다.

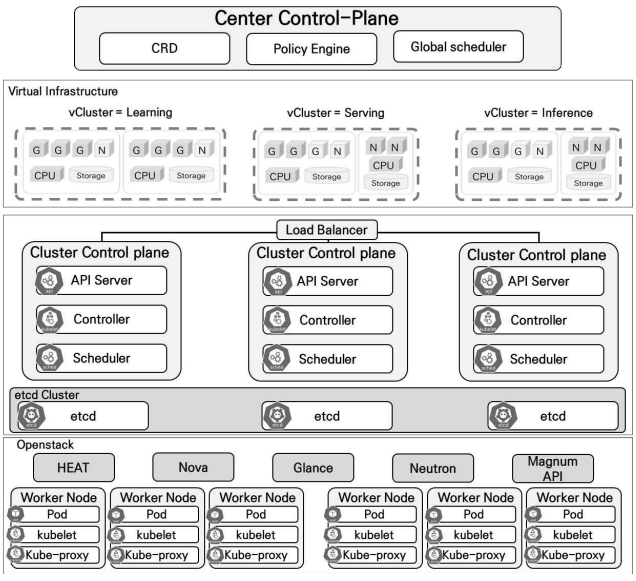
다만 대다수 조직은 가속기 집약 워크로드의 시간적·공간적 변동성을 정확히 예측하기 어렵고, 이를 보수적으로 처리하기 위해 과잉 할당(over-provisioning)과 단편화(fragmentation)를 초래한다[1]. 단일 클러스터 내부에서도 모델 병렬화 제약, 노드 간 상호연결(PCIe/NVLink) 토폴로지, 데이터 지역성 등으로 인해 사용가능한 자원과 실제로 배치 가능한 자원 사이에 괴리가 발생하며, 멀티 클러스터로 확장될수록 가용성·비용·지연시간의 상충 관계가 더 커진다[2].

본 논문에서는 이러한 문제를 해결하기 위해 멀티 클러스터 환경에서의 가상 클러스터·가상 인프라 구성과 C3-ATS(Cost-Carbon-Constraint Accelerator-aware Two-level Scheduler) 알고리즘을 제안한다. 제안 시스템은 다양한 클라우드 제공자가 이기종 가속가 자원을 통합된 풀로 관리하면서, 테넌트별 격리와 민첩성을 보장한다. C3-ATS 알고리즘은 비용·탄소·SLA 제약을 동시에 고려하는 이중 단계 스케줄링을 통해, 전역 클러스터 간 최적화와 로컬(노드) 최적화를 연계하여 운용 비용을 최소화한다[3].

Ⅱ. 본론

1. 운용 비용 최적화를 위한 멀티 클러스터 아키텍처 설계

본 연구에서 제안하는 멀티 클러스터 환경의 가상인프라는 네가지 핵심 설계 목표를 달성하기 위해 계층화된 아키텍처를 채택한다. 첫째 클러스터의 이기종 가속기를 포함한 모든 가용 자원을 통합 풀로 관리하여 워크로드 특성별 비용을 최소화한다. 둘째, 테넌트, 팀, 서비스별로 독립적인 제어면을 제공하면서도 물리 자원은 공유하여 유휴 자원을 최소화한다. 셋째, GPU, NPU 등 다양한 가속기의 특성 차이를 통합 API를 통해 인터페이스로 제공한다. 넷째, 전역 클러스터 간과 로컬(노드) 레벨에서 서로 다른 목적함수와 제약을 처리하는 이중 단계 스케줄링을 구현한다.



[그림 1] 멀티 클러스터 가상 인프라 아키텍처

그림1은 워커 노드와 컨트롤 플레인을 분리하고, API 서버는 L4/L7 로드 밸런서 뒤에서 노출한다. etcd는 전용 클러스터로 분리해 컨트롤 플레인을 보호하며, 과반수 합의로 일관성을 보장한다.

중앙 Control Plane은 멀티 클러스터 차원의 정책·카탈로그·스케줄링 결정을 담당하며, 각 지역에 배치된 고가용 컨트롤 플레인과 gRPC/REST/CRD 동기화로 연동된다. 중앙 평면은 상태 저장을 자체 스토리지(예: RDB/메시지 버스)로 관리하되, 워크로드의 상태는 각 클러스터별 외부 etcd

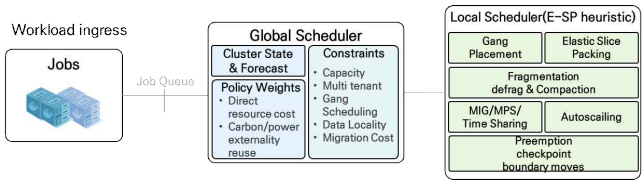
에 보관한다. 따라서 중앙 Control Plane 장애 시에도 각 클러스터는 독립적으로 안정 동작하며, 중앙 Control Plane이 복구되면 조정자(reconciler)가 최종 상태를 재수렴한다.

가상 인프라 계층은 노드와 가속기 가상화를 담당하는 핵심 구성요소다. vNode 풀(vNode Pool)은 실제 노드 풀을 가상 노드 클래스로 추상화한다. 가속기 가상화는 하드웨어 분할(MIG)과 소프트웨어 공유(MPS), 그리고 시분할 방식을 병행해 구현하며, 각 가상 가속기는 큐 기반 또는 슬라이스 기반의 단위로 제공된다.

자원 모델링 측면에서 각 가속기 프로파일은 타입, 메모리 용량(GB), 연산 성능(TOPS), 대역폭, 지원 데이터 타입, 공유 가능 여부를 포함한다. 비용 모델은 클러스터와 시간대별로 산정한 단위 자원 비용률(원화/시간)을 사용한다. 필요에 따라 클러스터별 탄소 강도와 전력 지표를 비용 항에 가중치로 포함해 환경 비용을 함께 고려한다[4].

2. C3-ATS: 비용-탄소-제약 인지형 이중 단계 가속기 스케줄러

C3-ATS(Cost-Carbon-Constraint Accelerator-aware Two-level Scheduler)는 멀티 클러스터 환경에서 비용, 탄소 배출, 그리고 서비스 수준 합의(SLA) 제약을 동시에 고려해 작업을 배치하는 이중 단계 스케줄링 알고리즘이다. 시간에 따라 도착하는 다양한 작업을 멤버 클러스터와 그 하위의 노드·가속기 자원에 적절히 배치하는 문제를 다룬다.



[그림 2] C3-ATS 스케줄러 동작방식

그림 2는 스케줄러의 전반적인 동작 방식을 나타낸다. 목표는 주어진 기간 동안의 총 운용비용을 최소화하는 것이다. 여기에는 자원 사용에 따른 직접 비용, SLA 위반으로 인한 패널티, 전력·탄소에 기인한 간접 비용이 포함되며, 반대로 모델 가중치의 재사용과 데이터 지역성으로 얻는 절감 효과는 비용에서 공제된다.

전역 스케줄링에서는 시간에 따른 가격과 수요의 변동은 자연스럽게 반영되며, 최적화는 다음 제약을 모두 만족하는 범위에서 수행된다.

(i) 노드와 가속기의 슬라이스·메모리·대역폭 등 용량 제약,

(ii) 테넌트 격리와 우선순위 기반의 공정성,

(iii) 학습 작업의 동시 배치(gang scheduling) 요구,

(iv) 모델·데이터의 지역성,

(v) 인스턴스 이동 시 발생하는 마이그레이션 비용.

가중치는 온라인 관측에 기반해 점진적으로 조정된다. 혼잡이 발생한 클러스터에는 가격을 높여 실효 단가를 증가시키고, 그 결과 신규 작업이 다른 클러스터로 분산되도록 유도한다. SLO 위반이 관측되면 지연 민감도를 강화하고, 모델·데이터 재사용이 잦을수록 지역성의 중요도를 높이는 식으로 정책이 조정된다.

로컬 스케줄링 단계에서는 Elastic Slice Packing(E-SP) 휴리스틱을 사용해 노드·가속기 단에서의 실제 배치를 최적화한다. 학습 작업은 통신 오버헤드를 줄이기 위해 가능하면 동일 노드의 연속 자원에 우선 배치하고, 배치 추론 작업은 MIG, MPS, 시분할 등 공유 메커니즘을 적극 활용해 빈 슬롯을 촘촘히 채운다[5]. 온라인 서빙 작업은 목표 지연을 보장하기 위해, 큐 길이와 지연 신호에 기반한 자동 확장으로 워밍 인스턴스를 유지한다.

자원 모델 측면에서 각 가속기는 메모리, 연산, 대역폭의 다차원 용량을 가진다. 가상화된 슬라이스 단위는 이들 차원을 일정 비율 점유하며, 작업은 각 차원에 대한 요구량을 명시한다. 서빙 워크로드는 추가로 요청률과 목표 지연을 포함한다.

E-SP는 우선순위가 높은 작업과 적합한 슬라이스를 먼저 짝지어 파편화가 최소화되도록 배치한다. 일정 주기마다 조각모음(defragmentation)과 압축(compaction)을 수행해 잘게 나뉜 슬라이스를 병합하고, 필요시 가장 영향이 작은 인스턴스를 선별적으로 이동시켜 큰 연속 슬라이스를 확보한다. 지연이 목표치를 초과할 조짐이 관측되면 버스트를 흡수하도록 슬라이스를 일시 증설하고, 부하가 안정되면 단계적으로 축소한다.

마이그레이션은 다음 원칙을 따른다. 첫째, 기대 절감액이 사전에 정한 임계치를 넘을 때만 이동한다. 둘째, 서빙 워크로드는 대체 인스턴스를 먼저 웜업해 무중단 전환을 보장한다. 셋째, 학습 워크로드는 체크포인트 경계에서만 이동을 허용해 진행 중인 학습에 미치는 영향을 최소화한다.

Ⅲ. 결론

본 논문에서는 멀티 클러스터 환경에서 이기종 가속기 자원을 효율적으로 관리하기 위한 가상 인프라 아키텍처와 C3-ATS(Cost-Carbon-Constraint Accelerator-aware Two-level Scheduler) 알고리즘을 제안하였다. 제안된 시스템은 계층화된 아키텍처를 통해 GPU/NPU 등 다양한 자원을 통합 관리하면서도 테넌트별 격리와 독립성을 보장한다. C3-ATS는 전역 단계에서 비용, 탄소 배출, SLA를 종합적으로 고려한 클러스터 선택과 로컬 단계에서 워크로드의 최적화된 자원 할당을 수행하는 이중 단계 접근법을 채택하였다. 본 연구는 AI/ML 워크로드가 급증하는 현 시점에서 비용 효율성과 지속 가능성을 동시에 달성할 수 있는 실용적인 솔루션을 제시하며, 클라우드 환경에서의 자원 관리에 새로운 방향을 제시한다. 향후 강화학습 기반 적응형 최적화와 엣지-클라우드 연속체로의 확장을 통해 더욱 발전된 시스템으로 진화할 것으로 기대된다.

ACKNOWLEDGMENT

이 논문은 2025년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구 결과임 (RS-2025-02293869, AI 반도체를 활용한 클라우드 플랫폼 구축 및 최적화 기술 개발)

참 고 문 헌

[1] Wencong Xiao, Romil Bhardwaj, et al. “Gandiva: Introspective Cluster Scheduling for Deep Learning.” OSDI 2018. USENIX Association, 2018.

[2] Juncheng Gu, Mosharaf Chowdhury, Kang G. Shin. “Tiresias: A GPU Cluster Manager for Distributed Deep Learning.” NSDI 2019. USENIX Association, 2019.

[3] Peifeng Yu, Mosharaf Chowdhury. “Salus: Fine-Grained GPU Sharing Primitives for Deep Learning Applications.” MLSys 2020. 2020.

[4] Jiaqi You, Yubo Li, et al. “Zeus: Understanding and Optimizing GPU Energy for DNN Training.” NSDI 2023. USENIX Association, 2023.

[5] Zhisheng Ye, Wei Gao, Qinghao Hu, Peng Sun, Xiaolin Wang, Yingwei Luo, Tianwei Zhang, and Yonggang Wen. “Deep Learning Workload Scheduling in GPU Datacenters: A Survey.” ACM Computing Surveys 56(6), Article 146, 2024. ACM, 2024.

Kubernetes 마이크로서비스의 리소스 효율성을 위한 예측 기반 오토스케일링 기법

정현수[1], 길준민[2*]

[1,2]제주대학교 컴퓨터공학과

[1]jhs990909@stu.jejunu.ac.kr, [2]jmgil@jejunu.ac.kr

A Prediction-Based Autoscaling Technique for Resource Efficiency in Kubernetes Microservices

Hyunsu Jeong[1], Joon-Min Gil[2*]

[1,2]Dept. of Computer Engineering, Jeju National University

요 약

현대 클라우드 네이티브 환경에서 애플리케이션의 확장성과 자원 최적화는 서비스 품질과 안정성을 좌우한다. 이를 지원하는 쿠버네티스(Kubernetes)는 HPA(Horizontal Pod Autoscaler)를 통해 부하 변화에 따라 파드 수를 자동 조정하지만, 반응형 설계 특성으로 인해 콜드 스타트(cold start)로 인한 지연과 성능 저하가 발생한다. 본 논문에서는 이러한 한계를 극복하기 위해 LSTM(Long Short-Term Memory) 기반 예측 모델을 활용하여 미래 CPU 사용량을 사전에 추정하고 이를 오토스케일링 파이프라인에 반영하는 방식을 제안한다. 예측 결과는 프로메테우스(Prometheus)와 KEDA(Kubernetes Event-Driven Autoscaler)를 통해 HPA로 전달되며, 서비스 부하가 발생하기 전에 파드를 선제적으로 확장할 수 있도록 한다. Online Boutique 마이크로서비스 애플리케이션과 Alibaba Cluster Trace 2017 데이터셋을 기반으로 수행한 실험에서 제안 기법은 기존 HPA 대비 90번째 및 95번째 백분위수에서 지연을 줄이고 요청 실패율을 감소시키며, 불필요한 자원 과잉 확장을 억제하는 효과를 보인다. 이를 통해 클라우드 네이티브 환경에서 마이크로서비스의 성능, 안정성, 자원 효율성을 동시에 확보한다.

Ⅰ. 서 론

현대 클라우드 네이티브 환경에서는 애플리케이션의 탄력성(elasticity), 확장성(scalability)과 자원 최적화를 통해 서비스 품질을 높인다. 클라우드 네이티브 환경 중 하나인 쿠버네티스(Kubernetes)는 선언적 자원 관리와 운영에 필요한 기본 기능을 제공하여 대규모 마이크로서비스를 안정적으로 운영한다. 쿠버네티스에서 파드(pod)는 컨테이너(container)를 실행하는 기본 단위로, 하나 이상의 컨테이너와 이를 실행하는 데 필요한 네트워크, 스토리지 등을 포함하는 가장 작은 컴퓨팅 단위이다. 쿠버네티스는 이러한 파드의 개수를 동적으로 조정하여 애플리케이션의 부하 변화에 대응한다. 특히, 복제본(replica)은 동일한 애플리케이션을 실행하는 다수의 파드로, 부하가 증가할 때 여러 개의 파드를 배포하여 서비스의 가용성을 유지하고 성능 저하를 방지하는 역할을 한다. 이러한 자동 확장을 담당하는 핵심이 쿠버네티스에서 HPA(Horizontal Pod Autoscaler)이며, 메트릭 서버(metric server)로부터 수집된 CPU, 메모리 사용량 등의 자원 지표를 기반으로 목표 임계치와 비교하여 파드 개수를 동적으로 조정하는 역할을 한다. 그러나 근본적으로 반응형 방식으로 동작하기 때문에, 새 파드가 생성되어 준비되는 과정에서 콜드 스타트(cold start) 문제가 필연적으로 발생하며, 이로 인해 순간적인 사용자 응답 지연이나 요청 실패율 증가가 나타난다[1]. 이러한 문제를 해결하기 위해 본 논문에서는 LSTM(Long Short-Term Memory) 기반의 예측 모델을 활용하여 미래의 CPU 사용량을 사전에 추정하고, 이를 쿠버네티스 오토스케일링 파이프라인에 반영한다. 본 논문의 제안 기법은 예측기를 통해 산출된 미래 부하 정보를 프로메테우스(Prometheus)와 KEDA(Kubernetes Event-Driven Autoscaler)를 통해 HPA로 전달하여, 부하가 발생하기 전에 파드를 선제적으로 확장할 수 있도록 한다. 이를 통해 서비스 지연을 최소화하고, 불필요한 자원 소모를 줄이며, 마이크로서비스의 안정성과 효율성을 동시에 확보한다.

Ⅱ. 제안 기법

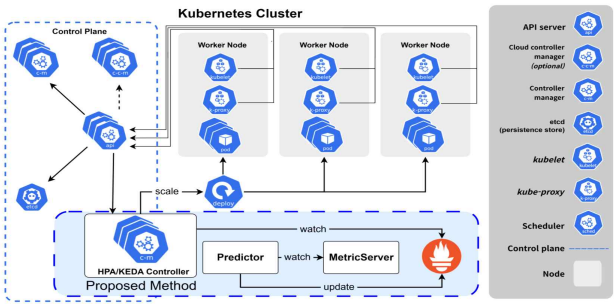


그림 1 예측 기반 오토스케일링을 위한 쿠버네티스 아키텍처

본 논문에서는 그림 1과 같이 쿠버네티스 환경에서의 예측 기반 오토스케일링 기법을 제안한다. 제안하는 기법은 메트릭 수집 단계에서 시작된다. 메트릭 서버는 각 디플로이먼트(deployment)의 CPU 사용량을 15초 주기로 수집해 쿠버네티스 API로 제공하며, 본 논문에서는 이렇게 수집된 연속된 값을 누적하여 최근 10분간의 시계열 데이터를 구성한다. 이 시계열 데이터는 LSTM 예측 모델의 입력으로 사용되며, 본 논문에서 구현한 예측기(predictor)에서 이를 기반으로 향후 15초 이후의 CPU 부하를 추정한다. 산출된 결과는 프로메테우스가 인식할 수 있는 메트릭 형식으로 변환되어 전달된다. 프로메테우스는 수집된 메트릭을 자체 시계열 데이터베이스에 저장하고 PromQL을 통해 질의할 수 있으며, 프로메테우스에서 수집한 예측된 메트릭 값들은 KEDA의 입력으로 전달된다. KEDA는 프로메테우스와 같은 외부 이벤트 소스를 활용하여 HPA를 동적으로 생성·관리할 수 있는 컴포넌트이다. 즉, KEDA는 프로메테우스에서 수집한 예측 메트릭을 기반으로 각 디플로이먼트에 대응하는 HPA를 자동으로 생성하고 해당 HPA는 실제 부하가 발생하기 전에 파드 복제본(replica) 수

표 1. k6 부하 테스트 결과: 예측 기반 오토스케일링 기법과 HPA의 비교

방식	frontend 누적 복제본 수	currencyservice 누적 복제본 수	전체 누적 복제본 수 합계	평균 응답 시간	p(90)	p(95)	요청 실패율
제안 기법	16,807	21,231	47,977	135.85ms	209.63ms	301.97ms	0.0064%
HPA	20,618	18,893	49,521	134.69ms	210.04ms	320.12ms	0.0092%

를 선제적으로 조정한다. 이 과정을 통해 마이크로서비스는 동시에 확장되며, 기존 HPA의 반응형 한계로 인해 발생하는 콜드 스타트 지연을 줄이고 불필요한 과잉 확장을 방지한다.

Ⅲ. 성능 평가

3.1 실험 환경 구성

제안한 오토스케일링 기법의 성능을 검증하기 위해서는 실제 워크로드 환경에서 부하 상황을 재현하는 과정이 필요하다. 본 논문에서는 Alibaba Cluster Trace 2017 데이터셋을 활용하여 부하 시나리오를 설계한다. 해당 데이터셋은 실제 대규모 클라우드 환경에서 수집된 워크로드를 포함하고 있다. 수집된 데이터셋은 학습과 테스트 평가 목적으로 8:2 비율로 분할하여 LSTM 기반 예측 모델의 훈련과 평가에 활용한다. 또한, 본 논문에서는 데이터셋 기반의 워크로드 패턴을 쿠버네티스 클러스터 상에서 재현하기 위해 k6[2] 부하 테스트 도구를 사용한다.

실험을 위해 사용한 마이크로서비스 애플리케이션은 Google이 공개한 Online Boutique Architecture[3]이다. 이는 전자상거래 플랫폼을 모사하도록 설계된 클라우드 네이티브 애플리케이션으로, frontend, productcatalogservice, cartservice, currencyservice, paymentservice 등 다수의 마이크로서비스로 구성된다. 이러한 구조는 서비스 간 의존성과 트래픽 전파 효과를 반영하므로, 제안 기법의 성능 검증에 적합하다.

학습 데이터셋을 기반으로 부하 패턴을 분석한 결과, 웹페이지 접속에 따른 실제 부하는 주로 frontend와 currencyservice에서 집중적으로 발생하였다. 따라서, 본 논문에서는 이 두 서비스에 대해서만 LSTM 기반 예측 모델을 적용하여 사전 확장을 수행하며, 나머지 마이크로서비스들은 기존 HPA 방식을 적용한다. 실험은 두 가지 방식으로 비교 수행한다. 첫 번째 방식은 제안한 예측 기반 오토스케일링 기법에 적용되는 frontend와 currencyservice는 LSTM 모델 예측 결과를 반영하여 선제적으로 스케일링하고 다른 서비스는 HPA에 의해 확장되는 방식이다. 두 번째 방식은 모든 서비스가 HPA에 의해 확장되는 기존 방식이다. 각 파드는 CPU 요청량(request) 100 millicore와 상한(limit) 200 millicore 조건으로 구성되었으며, 이를 기반으로 스케줄링과 자원 사용이 이루어진다. 오토스케일링의 임계 조건은 두 가지 방식 모두에서 동일하게 80%로 설정함으로써, 기존 HPA 방식과 제안하는 예측 기반 오토스케일링 방식 간 성능을 공정하게 비교할 수 있도록 한다.

3.2 성능 결과

제안 기법의 성능을 측정하기 위해 본 논문에서는 누적 복제본 수, 응답 시간(평균, 90번째 백분위수, 95번째 백분위수), 요청 실패율을 주요 지표로 설정하였다. 이때 응답 지연 시간은 k6에서 측정한 request duration으로, 이는 클라이언트 요청이 전송된 시점부터 응답이 완전히 수신될 때까지 걸린 총 소요 시간을 의미한다. 표 1은 제안한 LSTM 기반 예측 오토스케일링 방식과 기존 HPA 방식의 성능 비교 결과를 보여준다. 또한, frontend 누적 복제본 수와 currencyservice 누적 복제본 수는 k6 부하 테스트 시나리오 실행 시간 동안 15초 간격으로 측정된 복제본 수를 합산한 값이며, 시간에 따른 파드 사용량을 누적으로 반영하는 지표이다. 실험 결과, 제안한 예측 기반 오토스케일링 기법은 frontend에서 불필요하게 과잉 확장되는 현상을 줄이고 currencyservice와의 확장을 보다 균형 있게 수행한다. 이로 인해 테스트 시나리오 구간에서의 전체 누적 복제본 수 합계는 기존 HPA 대비 약 3.12% 감소한다. 또한 응답 지연 측면에서도 90번째 및 95번째 백분위수에서 예측 기반 오토스케일링 방식 기법이 개선 효과를 보이며, 요청 실패율 역시 0.0064%로 HPA의 0.0092%보다 낮게 나타나 서비스 안정성이 향상되었음을 확인할 수 있다.

Ⅳ. 결론

본 논문에서는 쿠버네티스 환경에서 HPA의 반응형 한계와 콜드 스타트 문제를 해결하기 위해, LSTM 기반 예측 모델을 활용한 오토스케일링 기법을 제안한다. 제안 기법은 CPU 사용량을 사전에 추정하고 이를 프로메테우스와 KEDA를 통해 HPA로 전달함으로써, 실제 부하 발생 이전에 파드를 선제적으로 확장할 수 있다. Online Boutique 마이크로서비스 애플리케이션을 활용한 실험 결과는 제안 기법의 효과를 뚜렷하게 보여준다. Online Boutique은 사용자가 웹 브라우저를 통해 frontend 서비스에 접속하면 frontend가 내부적으로 여러 마이크로서비스에 요청을 전달하는 구조를 가진다. 특히 상품 가격 표시 과정에서 frontend는 gRPC를 통해 currencyservice에 환율 변환 정보를 요청하고 그 결과를 받아 최종 응답을 구성한다. 이러한 구조적 특성으로 인해 외부 요청이 증가하면 부하가 우선 frontend에 집중되며, currencyservice는 frontend의 호출을 통해서만 부하가 발생하므로 확장이 상대적으로 늦게 이루어진다. 기존 HPA 방식에서는 frontend가 먼저 급격히 확장되고 currencyservice는 뒤늦게 확장되면서 병목이 발생한다. 반면, 본 논문의 제안 기법은 frontend와 currencyservice를 동시에 예측·확장함으로써 이러한 구조적 한계를 완화하고 특정 서비스에만 과도하게 부하가 몰리는 현상을 줄이며 서비스 간 확장을 균형 있게 수행한다.

ACKNOWLEDGMENT

본 과제(결과물)는 2025년도 교육부 및 제주도의 재원으로 제주RISE센터의 지원을 받아 수행된 지역혁신중심 대학지원체계(RISE)의 결과입니다(2025-RISE-17-001).

참 고 문 헌

[1] A. P. Jegannathan, R. Saha and S. K. Addya, “A Time Series Forecasting Approach to Minimize Cold Start Time in Cloud-Serverless Platform,” 2022 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom), Sofia, Bulgaria, 2022.

[2] B. Serracanta, A. Lukács, A. Rodriguez-Natal, A. Cabellos and G. Rétvári, “On the Stability of the Kubernetes Horizontal Autoscaler Control Loop,” in IEEE Access, vol. 13, pp. 7160-7166, 2025.

[3] H. Ahmad, C. Treude, M. Wagner and C. Szabo, “Smart HPA: A Resource-Efficient Horizontal Pod Auto-Scaler for Microservice Architectures,” 2024 IEEE 21st International Conference on Software Architecture (ICSA), Hyderabad, India, 2024

강화학습 기반 GNFS 분석을 활용한 RSA 보안성 평가 연구

이영도, 김수리*, 윤기순
㈜엔에스에이치씨, *성신여자대학교
{ydlee, ksyoon}@nshc.net, *suhrikim@sungshin.ac.kr

RSA Security Evaluation Using Reinforcement Learning-based GNFS Analysis

Youngdo Lee, Suhri Kim*, Kisoon Yoon
NSHC Inc. (Associate Researcher, Chief Cryptographer),
*Sungshin Women's University (Assistant Professor)

요 약

RSA 암호체계는 큰 정수의 소인수분해 문제 기반하여 안전성을 보장하며, 그 안전성은 GNFS 알고리즘의 복잡도에 의해 평가된다. GNFS는 하지수 시간 복잡도를 가지는 현재까지 가장 효율적인 정수 인수분해 알고리즘으로 알려져 있으나, 실제 구현에서는 이론적 예측과 성능 간 차이가 존재한다. 이러한 차이는 알고리즘 단계별 파라미터 선택의 복잡성, 기존 휴리스틱 기반 접근법의 한계, 그리고 하드웨어적 제약에서 비롯되는 것으로 분석된다. 본 연구에서는 강화학습을 활용하여 GNFS의 파라미터 선택 과정을 자동화하고 최적화하는 새로운 접근법을 제안한다. 강화학습은 환경과의 상호작용을 통해 보상을 기반으로 최적 정책을 학습함으로써 기존 휴리스틱 기반 방식보다 효율적인 파라미터 조정과 실행 성능 향상을 가능하게 한다. 이를 통해 GNFS의 탐색 공간을 줄이고 인수분해 속도를 개선하며, 실제 환경에서 이론적 복잡도에 근접하는 성능 향상을 기대할 수 있다.

Ⅰ. 서 론

RSA(Rivest - Shamir - Adleman) 암호체계는 1977년 제안된 이후 전자서명, 키 교환 등 다양한 보안 프로토콜에서 핵심적인 역할을 수행하며 오늘날까지 가장 널리 사용되는 공개키 암호 알고리즘이다 [1],[2]. RSA의 안전성은 큰 정수의 소인수분해 문제를 기반으로, 현재까지 금융 및 네트워크 보안 분야에서 표준으로 활용되고 있다 [3].

RSA의 보안 수준은 키 길이 비트 n에 의해 결정되며, 이를 분석하는 대표적 알고리즘은 GNFS(General Number Field Sieve)이다 [4],[5]. GNFS는 하지수(sub-exponential) 시간 복잡도를 갖는 가장 효율적인 인수분해 알고리즘으로 알려져 있으며, RSA 안전성 평가의 핵심 지표로 활용된다. 그러나 이론적으로 RSA-1024 (n : 1024비트)가 2020년 이전에 인수분해 가능할 것으로 예측되었음에도 실제로는 RSA-250(n :829비트)까지만 인수분해가 성공하였다 [6]. 이는 이론적 복잡도 모델과 실제 하드웨어 환경 간의 차이, 파라미터 최적화의 한계 등이 복합적으로 작용한 결과로 분석된다.

GNFS는 크게 5단계로 구성되며, 각 단계의 파라미터 선택은 전체 성능에 중요한 영향을 미친다 [7]. 기존 연구는 주로 경험적 방법과 휴리스틱을 통해 파라미터를 조정해왔으나, 단계 간 복잡한 상호작용을 충분히 반영하기에는 한계가 있었다. 이에 본 연구에서는 강화학습(Reinforcement Learning)을 활용하여 GNFS 파라미터를 동적으로 최적화함으로써 실제 인수분해 성능을 향상시킬 수 있는 가능성을 연구한다.

Ⅱ. 관련 연구

RSA암호체계는 제안된 이후 현재까지 가장 널리 사용되는 공개키 암호 방식 중 하나로, 큰 정수의 소인수분해 문제의 계산 복잡도에 기반해 보안성을 보장한다. RSA는 두 개의 큰 소수를 선택해 곱한 값을 n으로 사용하며, 이를 효율적으로 소인수분해하지 않는 한 비밀키를 추론할 수 없다는 가정 위에서 동작한다[3].

이론적으로 RSA의 안전성은 현재 알려진 가장 효율적인 정수 인수분해 알고리즘인 GNFS(General Number Field Sieve)의 시간 복잡도를 통해 추정된다 [4],[5]. GNFS는 하지수(sub-exponential) 시간 복잡도를 가지며, 이론적 계산 모델에서는 컴퓨팅 파워의 발전 추세를 고려할 때 RSA-1024(1024비트)가 2020년경 실질적으로 인수분해될 수 있다는 예측도 제기된 바 있다. GNFS의 복잡도는 다음과 같으며,

$$L_n[1/3,(64/9)^{1/3}] = e^{(((64/9)^{1/3}+o(1))\cdot(\log N)^{1/3}(\log\log N)^{2/3})}$$

이는 알려진 알고리즘 중 가장 효율적인 것으로 알려져있다.

그러나 위 복잡도를 기반으로 실제 실험된 결과는 이론적인 복잡도와 간극이 존재한다. 2020년 기준으로 가장 큰 n을 성공적으로 인수분해된 것은 RSA-250(n :829비트)에 불과했으며[6], 이는 1024비트의 큰 수 n을 인수분해하기 위해 요구되는 계산 자원이 여전히 확보되지 않았음을 의미한다. 이러한 차이는 이론적 모델에서 가정한 최적화된 알고리즘, 완벽한 병렬화, 이상적인 하드웨어 환경과 실제 구현에서 발생하는 메모리 대역폭, 병렬 처리 오버헤드, 프로세서 성능 한계 등의 제약 사이에서 발생한다. 따라서 RSA의 보안성을 평가할 때는 단순히 이론적 복잡도 추정치에 의존하기보다, 실험적 분석 결과와 실제 구현상의 한계를 함께 고려할 필요가 있다.

GNFS는 현재까지 알려진 가장 빠른 정수 인수분해 알고리즘으로, 크게 (1) 다항식 선택(Polynomial Selection), (2) 체(Sieving), (3) 필터링(Filtering), (4) 선형대수(Linear Algebra), (5) 제곱근(Square Root) 단계로 구성된다. 특히 다항식 선택과 체 단계에서의 파라미터 설정은 알고리즘 전체 성능에 결정적인 영향을 미치며, 현재까지는 경험적 접근과 휴리스틱 기반 파라미터 수정에 의존하는 경우가 많다.

이러한 방식의 한계는 딥러닝, 특히 강화학습 기법을 활용하여 분석하고 개선 가능성을 탐구할 수 있을 것으로 판단된다. 강화학습은 환경과의 상호작용을 통해 보상을 기반으로 최적 정책을 학습함으로써, 기존의 수작업 기반 최적화보다 더 효율적이고 안정적인 인수분해 성능 향상을 기대할 수 있다.

Ⅲ. 강화학습 기반 GNFS 분석 제안 방법

강화학습(Reinforcement Learning, RL)은 에이전트가 환경과 상호작용하며 보상 함수를 최대화하는 방향으로 정책을 학습하는 기계학습 기법으로, 명시적인 지도(supervision) 없이도 순차적 의사결정 문제를 해결할 수 있는 장점을 가진다. GNFS는 초기 단계에서의 파라미터 선택이 이후 단계의 계산량과 인수분해 성공률에 큰 영향을 미친다. 이러한 특성은 강화학습에서 파라미터 조합을 탐색·학습·보상 기반으로 최적화할 수 있는 가능성을 제공한다.

본 연구에서는 GNFS 파라미터 최적화 문제를 강화학습 프레임워크로 정의하고, 알고리즘 수행 과정에서의 다항식 품질 평가 지표인 Murphy E-값을 보상으로 설정한다. Murphy E-값는 다항식의 소인수분해 가능성을 추정하는 척도로, 일반적으로 다음과 같이 정의된다[8].

$$\frac{6}{\pi^2}\int\int_{\Omega}\rho\left(\frac{\log|F(x,y)|+a(F)}{\log B_1}\right)\rho\left(\frac{\log|G(x,y)|+a(G)}{\log B_2}\right)dxdy$$

위 식에서 $\rho(u)$는 Dickman-de Brujin 함수로, 주어진 정수 또는 다항식 값이 특정 smoothness bound를 초과하지 않을 확률을 근사하며, a는 다항식의 값들이 특정 소수로 얼마나 잘 나눠지는 지를 측정하는 함수이며, B_1, B_2 는 smoothness bound를 나타낸다. 결과적으로 Murphy E-값이 높을수록 체(Sieving) 단계에서 더 많은 유용한 관계식(relations)을 발견할 가능성이 크다.

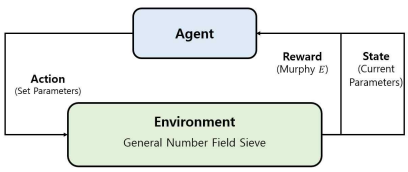


그림 1 강화학습 기반 GNFS 분석 제안 방법

본 논문에서 제안하는 방식은 [그림 1] 과 같다. 강화학습에서의 에이전트(Agent)는 환경(Environment)에서 다양한 파라미터 조합을 실험하며, 장기적으로 계산 비용을 최소화하고 높은 Murphy E-값을 유도하는 최적 전략을 학습한다. 이를 통해 기존 휴리스틱 기반 접근이 놓칠 수 있는 비직관적 최적 해를 발견하고, 동일한 계산 자원으로 더 큰 n를 인수분해할 가능성을 높일 수 있다. 강화학습 기반 접근법은 특히 다항식 선택 단계에서 후보 다항식의 Murphy E-값을 보상으로 활용해 품질을 평가하고 최적 후보를 선택하는 과정, 체 단계에서 소수 체 크기 등의 매개변수를 조정하는 과정에서 강력한 개선 효과를 기대할 수 있다. 본 논문에서 제안하는 방법은 GNFS를 활용한 RSA 분석을 보다 효율적으로 수행할 수 있는 새로운 최적화 패러다임을 제시하며, 대규모 정수 인수분해 문제의 실질적 계산 가능성 평가에 기여할 것으로 전망된다.

Ⅳ. 기대효과 및 향후 과제

본 연구에서 제안하는 강화학습 기반 GNFS 분석 방법은 기존 휴리스틱 또는 경험적 접근법으로는 발견하기 어려운 최적 파라미터 조합을 인공지능을 이용하여, 효율적으로 탐색할 수 있는 가능성을 제공한다. 이를 통해 동일한 하드웨어 자원에서 더 빠르고 효율적인 인수분해를 수행하거나, 제한된 자원으로도 기존보다 큰 n에 대한 실질적 분석 가능성을 높이는 데 기여할 수 있다. 특히 RSA 암호체계의 보안성 평가 과정에서 기존 이론적 복잡도 분석과 실제 구현 간의 간극을 줄이고, 보다 정밀한 실험 기반 보안 수준 추정치를 제공할 수 있을 것으로 기대된다.

향후 과제는 다음과 같다. 첫째, 강화학습 모델을 GNFS 환경에 적합하도록 구체화하고, 실제 인수분해 시뮬레이션 데이터를 기반으로 학습 효율을 높이는 방법을 마련해야 한다. 둘째, GNFS 단계별로 독립적 파라미터 최적화뿐 아니라 단계 간 상호 의존성을 고려한 통합 정책 학습 기법을 개발할 필요가 있다. 셋째, 제안된 강화학습 기반 최적화 기법을 대규모 분산 환경 또는 GPU 가속 기반 클러스터에서 실험하여, 실제 고성능 컴퓨팅(HPC) 환경에서의 확장성과 실효성을 검증해야 한다. 마지막으로, 강화학습 모델이 GNFS 알고리즘의 근본적인 계산 복잡도를 줄일 수 있는 잠재적인 패턴을 학습하는지, 아니면 단순히 기존 휴리스틱을 근사하는지에 대한 이론적 분석이 필요하다.

본 연구는 강화학습을 활용한 새로운 정수 인수분해 최적화 접근법을 제안함으로써 RSA 보안성 평가 및 암호 강도 측정에 있어, 보다 실질적이고 정량적인 근거를 마련할 수 있는 기반을 제공하며, 향후 고전적 인수분해 알고리즘의 효율화와 차세대 암호 안전성 분석 분야에서 중요한 연구 방향으로 확장될 수 있을 것으로 기대된다.

ACKNOWLEDGMENT

이 논문은 2025년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (RS-2024-00399401, 양자안전 보안인프라 전환 및 대양자 복합 안전성 검증기술 개발)

참 고 문 헌

[1] R. L. Rivest, A. Shamir, and L. Adleman, “A Method for Obtaining Digital Signatures and Public-Key Cryptosystems,” Communications of the ACM, vol. 21, no. 2, pp. 120 - 126, 1978.

[2] W. Stallings, “Cryptography and Network Security: Principles and Practice”, 8th ed., Pearson, 2023.

[3] A. J. Menezes, P. C. van Oorschot, and S. A. Vanstone, “Handbook of Applied Cryptography”, CRC Press, 1996.

[4] A. K. Lenstra and H. W. Lenstra Jr., “The Development of the Number Field Sieve”, Springer, 1993.

[5] R. P. Brent, “Recent Progress and Prospects for Integer Factorisation Algorithms,” Computational Methods in Number Theory, 1995.

[6] B. B. Arjen Lenstra et al., “The Factorization of RSA-250,” Mathematics of Computation, 2020. [Online]. Available: https://eprint.iacr.org/2020/1492

[7] T. Kleinjung, “Polynomial Selection for the General Number Field Sieve,” Mathematics of Computation, vol. 75, pp. 2037 - 2047, 2006.

[8] B. A. Murphy, “Polynomial Selection for the Number Field Sieve Integer Factorization Algorithm”, Ph.D. Thesis, The Australian National University, 1999.

보이스피싱 탐지 AI 강화를 위한 ATT&CK - RAG 프레임워크

설동명, 김지용
한국전자통신연구원

dmsul@etri.re.kr, kjy@etri.re.kr

ATT&CK-RAG Framework for Strengthening AI for Voice Phishing Detection

Seol Dong Myung, Kim Ji Yong
Electronics and Telecommunications Research Institute.

요 약

본 논문은 보이스피싱의 사회공학적 특성을 체계적으로 분석하기 위해 설계된 보이스피싱 ATT&CK(Adversarial Tactics, Techniques, and Common Knowledge) 프레임워크를 제안한다. 본 프레임워크는 최근 급증한 AI 음성 모방, 정부지원 빙자 대출, 가상자산 투자 사기 등 최신 사례를 기반으로, 공격 전술(Tactics)과 기술(Techniques)을 명확히 분류한다. 특히 피해자 중심의 관점에서 심리적 조작 과정을 구조화하여, 효과적인 예방·대응 전략 수립에 기여한다. 또한, 프레임워크를 활용한 RAG(Retrieval-Augmented Generation) 기반 시나리오 자동 생성 방법론을 제시함으로써, 변종 및 신종 보이스피싱 탐지 AI 학습 데이터 확장과 성능 향상을 동시에 달성한다. 제안한 접근법은 기관과 개인의 능동적 위협 대응을 지원하며, 향후 보안 교육·실시간 탐지 시스템에 활용 가능하다.

Ⅰ. 서 론

보이스피싱은 단순한 금전적 피해를 넘어 사회 전반에 심각한 신뢰 훼손을 초래하는 범죄로 발전하고 있다[1]. 최근 공격자는 발신 번호 조작, AI 음성 합성, 맞춤형 사회공학 기법 등 첨단 기술을 적극 활용하고 있다. 그러나 기존 사이버 보안 위협 모델링은 주로 네트워크·시스템 취약점에 초점을 맞춰, 인간 심리의 취약성을 악용하는 보이스피싱 특성을 충분히 반영하지 못했다.

이에 본 연구는 피해자 관점의 심리·행동 기반 위협 모델인 보이스피싱 ATT&CK 프레임워크를 설계하였다. 또한 해당 프레임워크를 RAG 기법과 결합해 변종 시나리오를 자동 생성함으로써, 실제 탐지 AI 의 학습 데이터 부족 문제를 해결하고자 한다.

Ⅱ. 보이스피싱 ATT&CK 프레임워크

보이스피싱 범죄는 발신자 번호 조작, 정교한 시나리오 개발, AI 음성 합성 등 기술적 진보와 결합하여 끊임없이 고도화되고 있다. 특히 코로나 19, 고금리 등 사회적 변화를 악용한 수법들이 급증하며, 피해가 더욱 광범위해지고 있다. 이러한 공격 방식은 IT 시스템의 취약점보다 인간의 심리적 취약점에 의존하는 사회공학적 기법을 핵심으로 한다. 따라서 단순히 기술적 방어에 집중하는 기존 모델로는 보이스피싱 공격을 효과적으로 방어하기 어렵다. 본 프레임워크는 이와 같은 배경에서 보이스피싱 공격의 전술과 기술을 표준화된 방식으로 분류하고 공유함으로써, 대중 및 기관의 위협 이해도를 높이고 피해를 최소화하는 것을 목표로 한다.

2.1 보이스 피싱 ATT&CK 프레임워크의 구성 요소

보이스피싱 ATT&CK 프레임워크는 기존 IT 중심의 ATT&CK 과 차별화되는 고유한 구성 요소를 포함한다. 이 구성 요소들은 보이스피싱의 비기술적이고 심리적인 복잡성을 포착한다.

2.1.1 전술 (Tactics)

프레임워크는 공격의 목적에 따라 6 가지 주요 전술을 정의한다. '초기 접촉 및 접근'을 시작으로, 피해자의 경계심을 허무는 '신뢰 구축', 비합리적 판단을 유도하는 '심리적 압박 및 조작', 민감 정보를 탈취하는 '정보 획득', 금전적 이득을 취하는 '금전 편취', 그리고 추적을 어렵게 만드는 '흔적 제거 및 회피'로 구성된다. 이 전술들은 기술적 시스템이 아닌 '인간'이라는 가장 취약한 고리를 목표로 한다는 보이스피싱의 본질을 반영한다.

2.1.2 기술 (Techniques)

각 전술 아래에는 특정 행동을 수행하기 위한 구체적인 방법인 기술이 정의되어 있다. 프레임워크는 2020 년 이후의 최신 사례를 반영하여 다음 기술들을 포함한다.

- 초기 접촉: 전화 발신(T1001.001), 문자/메시징 앱(T1001.002)을 통한 접근과 함께, AI 음성

모방(T1004.005)과 같은 첨단 기술을 활용한 지인 사칭(T1002.002) 등이 포함된다.

- 정보 획득: 피해자의 개인 식별 정보(T1004.001) 및 금융 정보(T1004.002)를 직접 요구하는 것 외에도, 악성 앱 설치 유도(T1004.003)나 가짜 웹사이트 유도(T1004.004)를 통해 정보를 탈취하는 기술이 상세히 명시된다.

- 금전 편취: 단순히 계좌 이체(T1005.001)를 유도하는 것을 넘어, 피해자 명의로 고액 대출을 실행하고 편취(T1005.004)하거나 비대면 계좌를 개설하여 악용(T1005.005)하는 최신 수법들이 포함되어 있다.

2.2 보이스 피싱 ATT&CK 프레임워크 설계

본 프레임워크는 다음의 핵심 철학을 기반으로 설계되었다. 첫째, 피해자 중심 관점이다. 공격자의 기술적 침투 경로보다는 피해자가 겪는 인지적, 감정적 과정에 초점을 맞춤으로써, 실제 피해 상황을 더 잘 이해할 수 있도록 돕는다. 둘째, 실제 사례 기반이다. 금융감독원 등 공신력 있는 기관의 발표 자료와 2020 년 이후의 최신 사례를 면밀히 분석하여 프레임워크의 실용성과 신뢰성을 확보했다. 셋째, 예방 및 교육 중점이다. 공격 전술과 기술을 명확히 이해함으로써 효과적인 예방 교육 프로그램을 개발하고, 개인의 보안 의식을 높이는 데 활용될 수 있도록 했다. 마지막으로 확장성 및 유연성을 고려하여 새로운 보이스피싱 수법에 유연하게 대응할 수 있도록 했다.[2]

Ⅲ. 보이스피싱 탐지 AI 를 위한 RAG 기반 시나리오 생성

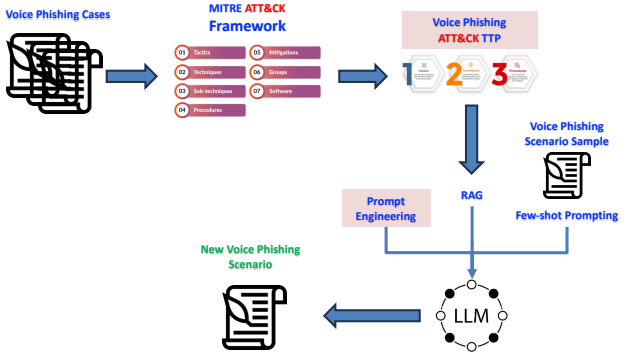
보이스피싱 탐지 AI 의 성능은 방대한 양의 고품질 학습 데이터에 크게 의존한다. 그러나 기존의 실제 사례 데이터만으로는 끊임없이 진화하는 변종 및 신종 공격 시나리오에 효과적으로 대응하기 어렵다. 이에 본 논문은 보이스피싱 ATT&CK 프레임워크를 활용하여 RAG(Retrieval-Augmented Generation) 기반으로 다양한 시나리오를 자동으로 생성하는 방법론을 제안한다.[3]

3.1 RAG 기반 시나리오 생성 방법

RAG 모델은 '검색(Retrieval)'과 '생성(Generation)'의 두 가지 핵심 단계로 구성된다.

- 검색(Retrieval): 보이스피싱 ATT&CK 프레임워크의 전술 및 기술 데이터베이스, 실제 보이스피싱 대본, 그리고 보이스피싱에 사용될 수 있는 일상 대화 패턴 등 다양한 정보 소스를 벡터화하여 지식 기반을 구축한다. 특정 시나리오를 생성하고자 할 때, 사용자는 결합하고자 하는 전술(예: T1004.005 AI 음성 모방)과 기술(예: T1005.004 대출 실행)을 프롬프트로 입력한다. 시스템은 이 프롬프트와 가장 연관성이 높은 정보(예: T1004.005 와 T1005.004 에 대한 설명, 관련 실제 대화 예시)를 지식 기반에서 검색하여 가져온다.

- 생성(Generation): 검색된 정보는 생성 모델(Generative Model), 즉 LLM(Large Language Model)의 입력으로 들어간다. LLM 은 이 정보를 바탕으로 특정 전술과 기술이 결합된, 현실적이고 자연스러운 대화 스크립트를 생성한다. 이 과정에서 다양한 화법, 말투, 상황(예: 자녀 빙자, 금융 기관 사칭 등)을 적용하여 다채로운 변종 시나리오를 만들어낼 수 있다.



[신종 보이스 피싱 시나리오 생성]

3.2 RAG 기반 시나리오의 장점

이 방법론은 다음과 같은 이점을 가진다. 첫째, 데이터 희소성 문제 해결: 실제 데이터의 한계를 넘어, 무한에 가까운 변종 시나리오를 생성하여 AI 학습 데이터의 양을 극적으로 늘릴 수 있다. 둘째, 실용성과 다양성 확보: 프레임워크의 전술과 기술을 기반으로 생성되므로 현실성이 높고, 다양한 조합을 통해 공격의 복합성을 반영한 시나리오를 만들어낼 수 있다. 셋째, 능동적 대응: 새로운 공격 트렌드가 발견되면 해당 정보를 지식 기반에 추가하여, 즉시 이를 반영한 시나리오를 생성하고 탐지 모델을 업데이트할 수 있다.

Ⅳ. 결론

본 논문에서는 보이스피싱의 사회 공학적 특성을 체계적으로 분석한 보이스피싱 ATT&CK 프레임워크를 제안하고, 나아가 이 프레임워크를 활용하여 보이스피싱 탐지 AI 의 성능을 혁신적으로 향상시킬 수 있는 RAG 기반 시나리오 생성 방법론을 제시했다. 이 프레임워크는 보이스피싱 예방 교육 뿐만 아니라, AI 기반 보안 시스템 개발의 핵심 도구로 활용될 수 있다. RAG 기반 생성 모델은 기존 데이터의 한계를 극복하고, 현실적이고 다양한 변종 시나리오를 만들어냄으로써 AI 모델이 미래의 예측 불가능한 위협에 더욱 효과적으로 대응할 수 있도록 돕는다.

ACKNOWLEDGMENT

이 논문은 2025 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.RS-2025-02215393, (2 세부) 알려지지 않은 신종 보이스피싱 탐지·예측 기술개발).

참 고 문 헌

[1] 금융감독원, “보이스피싱 피해 현황 및 예방 대책,” 2024.

[2] MITRE ATT&CK, https://attack.mitre.org/

[3] AWS, “Retrieval-Augmented Generation,” https://aws.amazon.com/ko/what-is/retrieval-augmented-generation/

RF Fingerprinting 과 DNN 을 이용한 NFC 태그 인증 시스템

이은규, 유호근, 허경무, 이재훈
고려대학교

lek7146@korea.ac.kr, hogeunyoo@korea.ac.kr, trivialdevil@korea.ac.kr, ejhoon@korea.ac.kr

Deep Learning-based RF Fingerprinting for NFC Tag Authentication

Eunkyu Lee, Hogeun Yoo, Kyongmoo Hur, Jaehoon Lee
Korea Univ.

요 약

본 논문에서는 신호 수집 시스템을 통해 ISO/IEC 14443 표준을 따르는 NFC 태그의 ATQA 신호에서 하드웨어의 특성을 반영하는 RF Fingerprint 를 추출하였다. 추출된 신호 샘플을 입력으로 하는 정규화 및 완전 연결 신경망 모델을 설계·학습하여 태그를 높은 정확도로 분류함을 실험적으로 검증하였다. 추가적인 신호 재수집 실험을 통해 시간과 환경 변화에도 안정적인 NFC 태그 분류 성능을 확인하였다.

Ⅰ. 서 론

NFC(Near Field Communication) 기술은 13.56 MHz 대역에서 작동하는 근거리 무선 통신 기술로, 다양한 분야에서 광범위하게 사용되고 있다. 그러나 NFC 기술의 광범위한 보급과 함께 다양한 보안 위험이 대두되고 있으며, 특히 카드 복제 공격과 중계 공격 등이 주요한 보안 취약점으로 지적되고 있다.

최근 연구들에서는 이러한 한계를 극복하기 위한 대안으로 RF Fingerprinting 기술이 주목받고 있다[1][2]. RF Fingerprinting 은 무선 송신기의 하드웨어 제조 과정에서 발생하는 미세한 물리적 불완전성을 이용하여 각 디바이스의 고유한 특징을 추출하는 기술이다.

본 논문에서는 RF Fingerprinting 기술에 딥러닝 모델을 적용하여, 기존 접근법의 한계를 극복하고 높은 분류 정확도를 달성하는 NFC 태그 인증 시스템을 제안한다. 정밀한 신호 수집 시스템 구축부터 ATQA(Answer To Request of Type A) 신호의 전처리 기법, 완전 연결 신경망 기반의 딥러닝 분류 모델 설계 및 최적화 과정에 이르기까지 전체적인 시스템 개발 방법론을 제시한다. 또한, 시간 경과에 따른 성능 안정성 분석을 통해 RF Fingerprinting 기술의 보안성 향상 효과를 검증한다.

Ⅱ. 본론

NFC 태그의 고유한 RF Fingerprint 는 제조 과정에서 발생하는 회로 부품의 미세한 하드웨어 변이에서 기인한다. 이러한 물리적 불완전성은 NFC 통신 과정에서 태그가 필수적으로 송신하는 ATQA 신호에 나타나며, 이를 활용하여 태그 식별이 가능하다. ATQA 신호는 ISO/IEC 14443 표준에 따라 리더의 REQA(Request Command, Type A) 신호에 대한 응답으로 전송되는 2 바이트 신호로, 모든 태그가 동기적으로 응답하는 특성을 가진다.

본 연구에서는 정밀한 RF Fingerprint 추출을 위해 SDR(Software Defined Radio) 기반의 NFC 신호 수집 시스템을 구축하였다(그림 1). GNU Radio 를 이용하여 ISO/IEC 14443 표준을 준수하는 REQA 신호를 생성하고, 이를 USRP-2954R 에 연결된 안테나를 통해 13.56 MHz 반송파로 송신한다. 송신 안테나와 동일한 모델의 수신 안테나를 사용하여 REQA 송신과 ATQA 응답을 포함한 전체 NFC 통신 과정을 모니터링한다. 중심 주파수는 13.56 MHz, 샘플링 레이트는 10 MSamples/s 로 설정하여 신호를 수집한다.

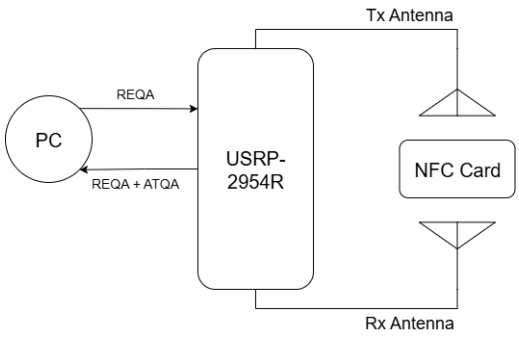


그림 1. NFC 카드 RF Fingerprinting 신호 수집 시스템 구조

수집된 원시 데이터는 REQA 와 ATQA 신호가 모두 포함되어 있으므로, RF Fingerprint 추출을 위해서는 후처리 과정이 필요하다. 수집된 연속 데이터를 REQA 신호를 기준으로 개별 세션(요청-응답 쌍)으로 분할한 뒤, ATQA 신호의 정확한 위치를 식별하기 위해 적응형 임계값 기반 검출 알고리즘을 적용한다. 각 샘플에 대해 1/2 비트 구간의 평균값을 계산하고, 이 평균값이 동적 임계값을 초과하는 지점들 중 가장 마지막에 위치한 지점을 종료 비트로 식별한다. 검출된 ATQA 종료점을 기준으로 역산하여 ATQA 신호의 시작점을 결정하고, 해당 구간의 신호를 정밀하게 추출한다. 이러한 데이터

수집 및 후처리 과정을 통해 NFC 카드 특성이 반영된 고품질 ATQA 신호를 추출할 수 있다(그림 2).

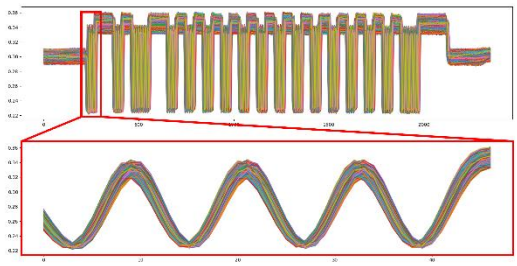


그림 2. 카드의 ATQA 신호와
딥러닝 학습에 활용되는 초기 신호 구간

RF Fingerprinting 기법을 이용한 개별 NFC 카드의 식별 가능성을 검증하기 위해 심층 신경망 기반의 분류 모델을 구축하였다. ATQA 신호에서 하드웨어 특성이 명확하게 나타나는 첫 번째 비트의 전반부 구간을 특징 추출 영역으로 선정하였다. 이 구간은 카드의 하드웨어에 기인한 고유한 신호 특성이 집중적으로 나타나는 영역으로, 해당 구간의 47 개 신호 샘플을 입력 벡터로 사용하며, Z-score 정규화를 적용하는 전처리 과정을 거친다. 이는 신호의 상대적 분포를 표준 정규분포로 변환하면서 학습 과정의 안정성을 향상시키고, 모델이 유의미한 특징을 보다 용이하게 학습하도록 돕는다.

본 연구에서는 완전연결 신경망(Fully Connected Neural Network, FCNN) 구조를 채택하였다. 네트워크는 그림 3 과 같이 구성된다.

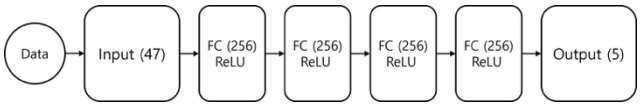


그림 3. NFC 태그 인증 시스템의 딥러닝 모델 구조

입력층과 출력층의 노드 수는 각각 47 개와 5 개이다. 은닉층은 크기가 256 인 4 개의 완전연결층으로 구성되어 있으며, 활성화 함수로 ReLU 함수를 적용하였다. 중요 특징을 보존하면서 과적합을 방지하기 위해 Dropout 기법을 적용하였다.

데이터셋은 목적에 따라 학습(train), 검증(validation), 시험(test)로 분리하였다. 학습 데이터셋을 이용해 신경망 모델 파라미터를 최적화하였으며, 전체 데이터 중 올바르게 예측한 비율을 계산하여 정확도를 측정하였다. 검증 데이터셋에서 가장 높은 정확도를 기록한 모델을 최종 모델로 선정하였고, 해당 모델에 시험 데이터셋의 정확도를 평가함으로써 모델의 성능을 측정하였다.

하이퍼파라미터는 학습률 0.001, 미니 배치 크기 128 로 설정하였고, Adam 옵티마이저를 사용하여 최적화하였다. 카드마다 약 1000 개의 데이터를 5000 Epoch 동안 반복하여 학습된 모델은 시험 데이터셋에서 98.3% 이상의 높은 정확도로 카드를 구별할 수 있음을 확인하였다(그림 4). 이는 추출된 RF Fingerprint 가 각 카드의 고유 특성을 효과적으로 반영하고 있음을 시사한다.

초기 높은 분류 성능이 데이터 수집 시점이나 환경적 요인 등에 의존할 가능성을 확인하기 위해 추가적인 검증 실험을 수행하였다. 기존에 수집한 데이터와는 상이한 시점에 수집한 데이터도 동일한 분류 성능을 갖는지 확인하기 위해, 기존 카드들에 대해 새로운 세션에서 데이터를 수집하고 기존 데이터만 학습한 동일한 모델을 통해 추론을 진행하였다(그림 5). 그 결과 99.1% 이상의 정확도로 카드를 구분할 수 있었고, 이를 통해 시간적 안정성과 학습되지 않은 데이터에 대한 모델의 강건성을 확인할 수 있었다.

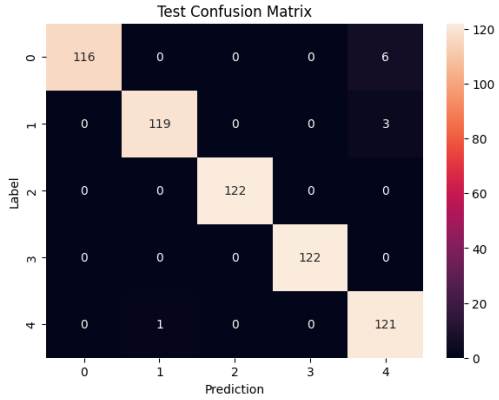


그림 4. 단일 시점 데이터에 대한 학습 및 추론 결과

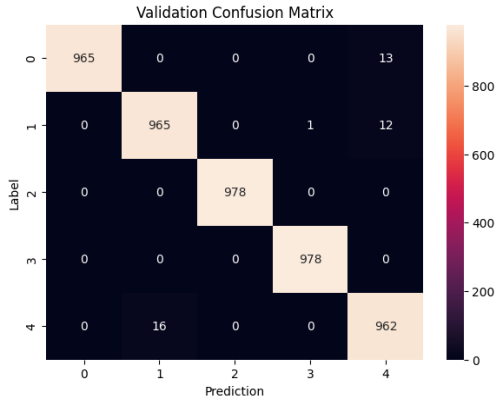


그림 5. 상이한 시점 데이터에 대한 사전 학습 모델 추론 결과

Ⅲ. 결론

본 연구에서는 신호 수집 시스템과 완전 연결 신경망을 결합하여 NFC 태그의 ATQA 신호에서 하드웨어 불완전성을 효과적으로 추출·분류함으로써 높은 카드 인증 정확도를 확인하였다. 이를 통해 RF Fingerprinting 기법의 실용 가능성을 검증하였으며, 향후 다양한 딥러닝 아키텍처와 다중 주파수 특징 공간을 도입하여 보안성과 일반화 성능을 강화할 예정이다.

참 고 문 헌

[1] W. Lee, S. Y. Baek and S. H. Kim, "Deep-Learning-Aided RF Fingerprinting for NFC Security," IEEE Communications Magazine, vol. 59, no. 5, pp. 96-101, May 2021, doi: 10.1109/MCOM.001.2000912.

[2] Y. Yang et al., "Jump Out of Resonance: A Practical NFC Tag Fingerprinting Scheme," IEEE Transactions on Mobile Computing, vol. 23, no. 9, pp. 8694-8709, Sept. 2024, doi: 10.1109/TMC.2024.3354813.

[3] W. Lee, S. Y. Baek, "Deep neural network-aided radio frequency fingerprinting for identification of near field communication tags," Expert Systems with Applications, vol. 296, Part B, 2026, 129016, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2025.129016.

노이즈에 강인한 아크탐지 알고리즘

이순우*, 강지명, 이주승, 이효동, 최성수
한국전기연구원

*rheesw@keri.re.kr

Arc detection algorithm robust to environmental noise

Soonwoo Lee, Jimyung Kang, Juseung Lee, Hyodong Lee, Sungsoo Choi
Korea Electrotechnology Research Institute (KERI)

요 약

무인화된 소규모 재생에너지 자원이 증가함에 따라 화재에 대한 효과적인 대비가 필요하다. 이를 위해 태양광 화재의 대부분을 차지하는 아크의 탐지와 차단에 대하여 IEC63027, UL1699B 와 같은 가이드라인을 규정하고 있다. 본 논문은 태양광 시스템과 같은 직류 전원(패널)의 직렬 아크의 탐지 방식을 제안한다.

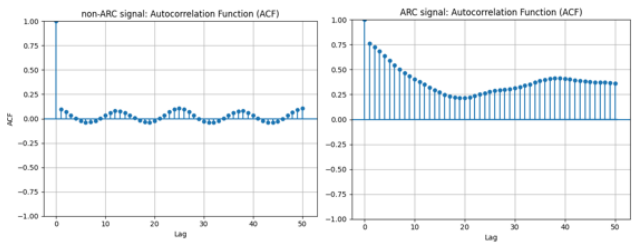
Ⅰ. 서 론

직렬 아크 발생시 전류의 증가가 미미하기 때문에 일반적으로는 전압, 전류 및 이의 주파수 등의 특정 패턴을 찾아 아크를 검출한다. 그러나 부하의 종류 및 상태의 급격한 변화, 주변의 EMI, 스위칭 노이즈 등 다양한 환경의 변화에 의해 정확한 아크의 탐지가 쉽지 않다[1].

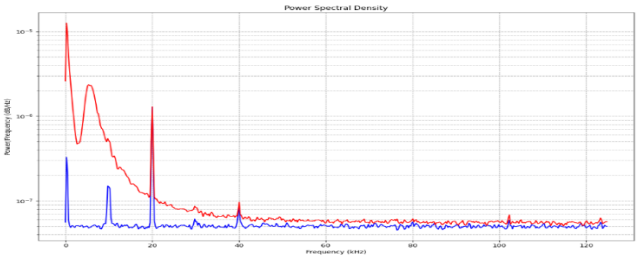
본 연구에서는 직렬 아크 전류의 50kHz 이하 대역에서 공통적으로 나타나는 pink-noise 성분[2]을 이용하여 다양한 노이즈에 보다 강인한 아크탐지 방식을 제안한다.

Ⅱ. 본 론

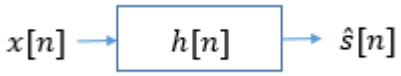
태양광 시스템에서 인버터 변환 이전의 직류 전류는 일반적으로 화이트 노이즈와 함께 인버터의 스위칭 주파수 성분이 강하게 나타난다. [그림 1]은 약 10kHz 의 스위칭 주파수를 갖는 인버터 앞단의 직류 전류에 대한 auto-correlation function (ACF)을 나타낸 것으로 정상 전류의 경우 매우 낮은 값을 갖는 반면 아크 발생시에는 correlation 이 강한 pink noise 성분에 의해 상대적으로 높은 ACF 값을 갖는다[3].



[그림 1] 정상전류(왼쪽) 및 아크전류(오른쪽) ACF



[그림 2] 정상전류(위) 및 아크전류(아래) PSD



[그림 3] MMSE 추정을 위한 LTI 필터

[그림 2]의 power spectral density (PSD) 그래프에서도 정상전류와 아크전류의 경우 대략적으로 50kHz 이하의 성분에서의 차이가 나타나는데 이러한 경향은 앞선 논문들에 의해 언급된다[3].

이와 같은 pink noise 성분을 강화하기 위해 [그림 3]과 같은 LTI 시스템에서 배경잡음 $\mathbf{w[n]}$ 이 포함된 측정신호 $\mathbf{x[n]}$ 로부터 아크신호 $\mathbf{s[n]}$ 를 MMSE (minimum mean square error) 관점에서 가장 잘 표현하는 출력신호 $\hat{s}[n]$ 를 갖는 필터 $\mathbf{h[n]}$ 즉, wiener filter 를 설계하였다[4].

식(1)과 같이 에러를 정의하면 MMSE 는 식(2)로, MSE 는 다시 식(3)과 같이 나타낼 수 있다.

$$\mathbf{e[n]} \triangleq \hat{s}[n] - s[n] \qquad \text{식 (1)}$$

$$\min_{h[\cdot]} \epsilon = E\{e^2[n]\} \qquad \text{식 (2)}$$

$$\epsilon = \mathrm{E}\{(\sum_{k=-\infty}^{\infty} h[k]x[n-k] - s[n])^2\} \qquad \text{식 (3)}$$

식(3)의 MSE 를 최소화하는 필터 $\mathbf{h[n]}$를 얻기 위해서는 식(4)에 의한 식(5)의 결과가 도출된다[4]. 이때 다음과 같은 가정이 필요하다.

아크신호 s[n] 는 평균이 0 인 WSS 확률과정이고, 배경잡음 w[n]은 아크신호 s[n]와 orthogonal 하다.

$$\frac{\partial \epsilon}{\partial h[m]} = 0 \qquad \text{식 (4)}$$

$$\mathrm{h[m]} * R_{xx}[m] = r_{sx}[m] \qquad \text{식 (5)}$$

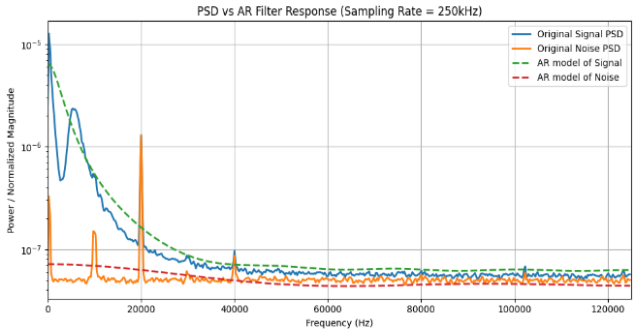
식(5)에서 $R_{xx}[m]$ 는 측정신호 x[n] 의 auto-correlation matrix 이고, $r_{sx}[m]$ 는 측정신호 x[n] 와 아크신호 s[n] 의 cross-correlation vector 이다. 식(5)를 z-Domain 으로 변환하면 식(6)과 같이 보다 간단한 형태로 나타낼 수 있다.

$$H(z) = \frac{S_{sx}(z)}{S_{xx}(z)} \qquad \text{식 (6)}$$

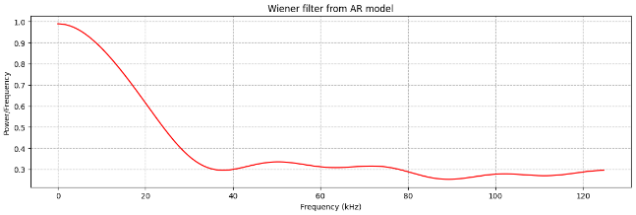
한편, 특정 부하의 스위칭 노이즈 등이 포함된 [그림 2]와 같은 신호로부터 기본적인 pink-noise 성분을 추출하기 위하여 [그림 4]와 같이 정상전류와 아크전류에 대하여 근사화된 Autoregressive(AR) 모델 $H_w(z)$, $H_s(z)$ 을 사용하여 식(7)과 같은 필터를 설계하였다. 설계된 필터 응답은 [그림 5]와 같다.

$$H(z) = \frac{|H_s(z)|^2 - k_1|H_w(z)|^2}{k_2(|H_s(z)|^2 - k_1|H_w(z)|^2) + k_3|H_w(z)|^2} \qquad \text{식 (7)}$$

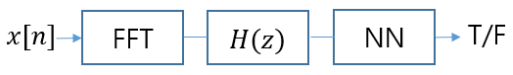
식(7)에서 $k_1 \sim k_3$ 은 튜닝 파라미터, $|H_s(z)|^2 - k_1|H_w(z)|^2$ 는 신호성분을, $k_3|H_w(z)|^2$ 는 배경잡음을 표현한다.



[그림 4] 정상전류 및 아크전류의 PSD 와 AR model



[그림 5] 설계된 필터의 주파수 응답



[그림 6] 아크 탐지 구조

최종적인 아크탐지 구조는 [그림 6]과 같으며, 우선 설계된 필터 $H(z)$ 를 이용하여 신호성분을 강조하고 이어서 신경망(NN) 알고리즘을 통해 아크 유무(T/F)를 판단한다. 신경망은 FFT 입력층과 1 단의 은닉층으로 구성된 multi-layer perceptron (MLP) 구조를 사용하였으며, 필터 설계에서 고려하지 못한 비선형 성분을 처리할 수 있다.

테스트를 위해 총 4 개의 서로 다른 스위칭 주파수(5kHz, 10kHz, 15kHz, 20kHz)를 갖는 인버터 부하를 사용하였다(각각 PV system 1~4). 만일, system (1)의 전류신호(정상 및 아크)로부터 필터 설계와 신경망 학습을 하였다면, 테스트는 나머지 system (2)~(4)의 전류 신호(정상 및 아크)로 함으로써, 학습되지 않는 새로운 환경에서의 성능을 평가하였다. 필터를 적용하여 테스트한 결과 [표 1]과 같이 모든 실험에서 필터가 없을 때보다 F1 score 가 증가하였다.

[표 1] 테스트 결과

실험 환경	평균 F1 score
w/o filter	0.92
system (1)	0.95
system (2)	0.96
system (3)	0.96
system (4)	0.96

Ⅲ. 결 론

본 연구에서는 다양한 노이즈 환경에서 직렬아크를 효과적으로 탐지하기 위한 알고리즘을 제안하였다. 아크 발생시 공통적으로 관찰되는 pink-noise 성분을 강조하기 위해 선형 필터를 설계하고, 신경망 구조를 연결함으로써 학습에 사용되지 않는 새로운 환경에서도 높은 성능을 나타내었다.

ACKNOWLEDGMENT

이 연구는 2025 년도 정부(과학기술정보통신부)의 재원으로 국가과학기술연구회의 지원을 받아 수행된 한국전기연구원 기본사업임. (No. 25A01033)

참 고 문 헌

[1] Jay Johnson, et al., “Arc-Fault Unwanted Tripping Survey with UL 1699B-Listed Products”, 2015 IEEE 42th Photovoltaic Specialist Conference, 2015

[2] Jonathan C. Kim, et al., “Isolation and Detection of Arc Fault Noise in a Real PV System Using Current Demodulation and Autocorrelation Coefficients”. IEEE Trans. Power Electronics, Vol. 39, No. 8, 2024

[3] Jay Johnson, et al., “Parametric study of PV arc-fault generation methods and analysis of conducted DC spectrum”, 2014 IEEE 40th Photovoltaic Specialist Conference, 2014

[4] Steven M. Kay, “Fundamentals of statistical signal processing: Estimation theory”, Prentice Hall PTR 1993

심층 강화학습을 활용한 비동기식 Self-Play 사이버 공방 연구

김정현 [1], 정재혁 [1], 장준원 [2], 김민석 [2*]

[1] 상명대학원 전자정보시스템공학과
[2*]상명대학교 휴먼지능로봇공학과

[1]jungkim9898@gamil.com, [1]2023D1013@smu.ac.kr, [2*]minsuk.kim@smu.ac.kr

Asynchronous Self-Play Simulation of Cyber-attack/defense via Deep Reinforcement Learning

Jung Hyun Kim [1], Jae Hyeok Jeong [1], Jun Won Jang[2], Min-Suk Kim [2*]
[1]Sangmyung University, Dept. of Electronic Information System Engineering
[2*]Sangmyung University, Dept. of Human Intelligence & Robot Engineering

요 약

최근 매년 증가하는 사이버 위협으로 인해 경제적 손실과 국가적 피해가 심각하게 확대되고 있다. 본 논문에서는 사이버 공방 대응을 위한 비동기식 시뮬레이션 환경을 설계하고, 이를 활용하여 심층 강화학습 기반의 적대적 멀티 에이전트 학습 기법을 제안한다. 제안된 환경에서 공격자는 방어자가 보호하는 노드를 점거하는 것을 목표로 하며, 방어자는 해당 노드를 방어하는 것을 목표로 한다. 또한, 공격자의 정상적인 공격 시퀀스 생성을 확인하기 위해 DQN, PPO, SAC 기반 강화학습 모델을 적용하여 성능을 검증하였으며, 방어자는 이를 대상으로 한 방어 실험을 통해 성능을 평가하였다. 실험 결과, 공격자의 공격 성공률은 SAC 44.44%, DQN 2060%, PPO 300%로 나타났으며, 그 중 SAC 가 상대적으로 가장 안정적이고 우수한 성능을 보였다.

Ⅰ. 서 론

사이버 위협은 개인, 기업, 국가를 대상으로 매년 증가하고 있다. 한국인터넷진흥원(KISA)에서 파악한 국내 사이버 공격 시도는 23 년 1,277 건이며 24 년은 1887 건으로 510 건이 증가하여 약 15%증가함을 보였다. 2025 년 올해 사이버 공격 횟수는 SKT 침해사고 여파에 따라 상반기에만 1034 건을 달성하였고, 이는 2023 년 상반기에 발생한 664 건, 24 년 상반기에 발생한 899 건에 비해 각각 55.72%, 15.02%가 증가 했음을 확인할 수 있다.[1] 또한 World Economic Forum(WEF)에서 전세계의 사기 및 사이버범죄 피해는 최근 12 개월 동안 1 조 달러 이상이 발생했으며 일부 국가의 GDP 의 3%의 손실을 얻었음을 확인하였다.[2] 이러한 사이버 위협은 경제적으로 막대한 피해를 초래할 뿐만 아니라 국가 안보를 위협하는 수단으로 활용되고 있다. 이에 대응하기 위한 사이버 보안 기술은 현대 사회에서 핵심적인 안보 수단으로 자리 잡고 있으며, 최근에는 인공지능과의 융합을 통해 그 성능과 효율성이 지속적으로 발전하고 있다. 특히, 강화학습은 복잡한 사이버 공격 방식을 학습하고, 다양하게 생성된 공격 패턴에 대한 방어 기술을 개발하는 데 유용한 연구 방법[3]이다. 특히 공격과 방어를 독립적으로 학습하는 기존 방식과 달리, 멀티 에이전트 강화학습(MARL)을 활용하면 공격자와 방어자가 상호작용 속에서 동시에 학습할 수 있어 사이버 보안 연구에 효과적으로 적용되고 있다.

본 논문에서는 Unreal 엔진 기반의 사이버 공방 시뮬레이션 환경을 구축하고, 이를 활용하여 MARL 기반의 비동기식 사이버 공방 학습 환경을 설계하였다. 기존의 사이버 공방 시뮬레이션 환경은 동기식으로 제작되어 공격과 방어가 차례대로 학습하는 방식[4]을 일반적으로 사용한다. 이러한 구조는 공격과 방어 행위에 대한 정확성을 높일 수 있는 반면, 실시간성을 확보하기 어려움이 있다. 본 논문에서는 이러한 문제점을 해결하기 위해 Unreal 엔진을 활용하여 공격과 방어의 행위 과정을 시각화 함으로써, 실시간으로 학습 과정을 관찰할 수 있는 환경을 구축하였다. 또한, 실시간으로 변화하는 네트워크 환경을 대상으로, 공격자는 특정 노드를 점령하고 방어자는 이를 저지하는 것을 목적으로 시뮬레이션을 설계하여 실험을 진행하였다.

Ⅱ. 본문

2.1 사이버 공방 기반 시뮬레이션 환경

본 논문에서는 기존의 동기식 사이버 공방 구조가 아닌 비동기식 시뮬레이션 환경을 구현하기 위해 Unreal 기반의 사이버 공방 시뮬레이션을 구축하였다. 전체 시뮬레이션 구조는 (1) 환경 상태를 관리하는 클라이언트(Unreal), (2) 환경과 에이전트 간의 통신을 담당하는 서버, (3) 환경 내 행동을 수행하는 클라이언트(Agent)로 구성된다.

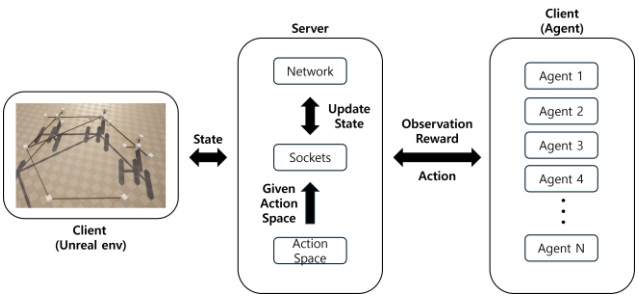


그림 1 Unreal 엔진 기반 사이버 공방 환경 구조

그림 1 은 사이버 공방 환경 시뮬레이션의 전체 구조를 보여주며, 학습 에이전트는 공격팀과 방어팀으로 구분되어 학습이 진행된다. 이때 공격팀은 공격자 단말을 통해 점진적으로 정보를 획득하여 학습을 수행하며, 방어팀은 개별 에이전트가 각 노드를 독립적으로 담당하도록 설계되었다. 공격도구와 방어도구는 단순화 작업을 최소화한 상태에서 MITRE ATT&CK 프레임워크에 기반하여 구현하였으며, 환경 조건에 따른 성공 유무를 판단하여 보상함수를 설계하고 이를 기반으로 학습을 진행하였다.[4]

2.2 Method: Decentralized Training with Decentralized Execution

MARL 학습은 크게 두 가지 접근 방식으로 구분된다. 첫째, 중앙에서 에이전트의 학습을 통합적으로 관리하는 방식인 CTDE(Centralized Training with Decentralized Execution)이며, 다른 하나는 학습과 실행 모두 에이전트가 독립적으로 수행하는 DTDE(Decentralized Training with Decentralized Execution)이다. 본 논문에서는 공격 에이전트가 점진적으로 정보를 취득하여 학습하는 POMDP(Partially observable Markov decision Process) 구조와 방어 에이전트가 자신이 관리하는 노드의 정보를 기반으로 학습하는 MDP(Markov Decision Process) 구조를 채택하기 때문에 두방식을 동시에 중앙에서 관리하는 것이 어렵다. 이에 각자의 역할에 최적화된 학습을 수행할 수 있도록 DTDE 방식을 기반으로 학습을 진행하였다.

Ⅲ. 실험

3.1 실험 환경 구성

본 논문에서는 하나의 공격 에이전트와 세 개의 방어 에이전트로 학습 환경을 구성하였으며, 최종 학습 목표는 방어 중인 세 개의 노드를 탈취하는 것으로 설정하였다. 반대로, 방어 에이전트는 각자 담당하는 노드의 점거를 방지하는 것을 목표로 학습을 수행하며, 이를 통해 공격자의 침투 시도를 저지하도록 구성하였다. 공격 순서는 실제 해킹 순서와 동일하게 미리 점거한 단말기와 연결된 노드들 스캔하고 공격하고자 하는 노드를 기준으로 취약점을 찾는 행위를 통해 정보를 취득하여 최종 Exploit 을 통해 노드를 탈취하는 방식이다. 방어는 공격의 스캔 행위가 실제 환경에서 대응하기 어렵거나 불가능하기 때문에, 공격 에이전트가 생성한 취약점 정보를 패치(Patch)하거나 공격 수행에 필요한 상태를 사전에 변경하는 방식으로 학습을 진행하였다.

3.2 MARL 기반 강화학습 모델 성능 검증

본 논문의 실험 환경에서는 학습에 앞서, 공격 에이전트가 실제 해킹 절차와 동일한 순서로 공격 과정을 생성하는지 확인하였다. 또한, 해당 에이전트가 주어진 공격 도구를 사용하여 상태공간을 확인하고 상황에 맞는 공격 학습이 진행됨을 확인하였다. 그림 2 는 강화학습 모델의 성능 검증을 위해 DQN(Deep Q-learning Network), PPO(Proximal Policy Optimization), SAC(Soft Actor-Critic)을 학습시킨 Reward-Episode 학습 성능 결과 그래프이다.

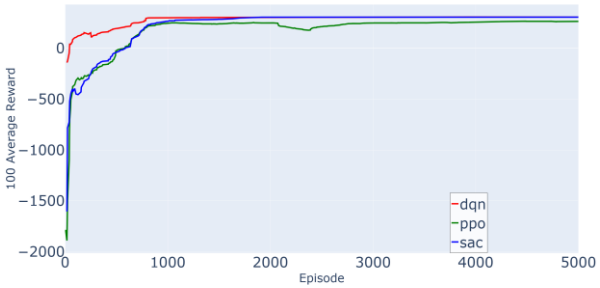


그림 2 Single Attacker Result for DQN, PPO, SAC Reward

DQN 의 공격 시퀀스는 최종 10 개, SAC 는 9 개, PPO 는 27 개의 공격 시퀀스를 생성함을 확인하였다. 공격 순서 역시 현재 설계된 시나리오에서 가정한 공격 순서와 유사하게 생성했기 때문에 현재 제작된 사이버공방 환경에서 의도한 결과를 도출했음을 확인하였다. 또한, 방어 에이전트를 도입하여 에이전트간 Self-Play 학습을 진행하였다. 그림 3 은 방어 에이전트를 적용했을 때 나타난 DQN Reward-Episode 결과 그래프이며, 결과적으로 기존 단일 공격자에 비해 도달한 보상함수의 최대 값이 감소함을 확인하였다.

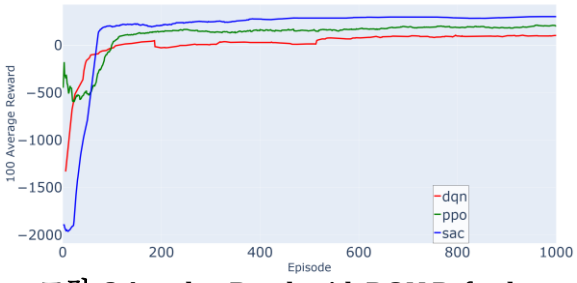


그림 3 Attacker Result with DQN Defender

최종 1000 에피소드를 기준으로 SAC 는 13 의 시퀀스를 생성했으며, PPO 는 108, DQN 은 216 의 시퀀스를 생성함을 확인하였고, 이는 방어자가 수행한 행동이 공격자의 유효한 행위 실행을 성공적으로 저지했음을 나타내고 있다.

Ⅲ. 결론

본 논문은 비동기식 사이버 공방 시뮬레이션 환경을 설계하고, 공격/방어 에이전트 간의 상호 적대적 학습을 통해 해당 환경의 신뢰성을 검증하였다. 비동기식 Self-play 학습방식은 동기식과는 다르게 통신 지연 및 작동 순서에 영향을 받는다. 따라서 단일 공격 행위만을 학습한 경우, 설계된 환경에서 선택할 수 있는 최선의 수인 9 개의 시퀀스가 아닌, 평균 15 개의 시퀀스를 생성하므로 일부 비효율적으로 작동함을 확인하였다. 또한, 공격 에이전트의 학습을 확인한 뒤 비동기식 Self-Play 환경에서 방어 에이전트를 적용하여 학습 진행하였을 때, 공격 에이전트의 공격 시퀀스 개수가 기존 대비 평균 약 646.67% 증가함을 확인할 수 있다.

ACKNOWLEDGMENT

This work was supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea Government (MSIT) (No. RS-2022-II220961)

참 고 문 헌

[1] 한국인터넷진흥원, “2025 년 상반기 사이버 위협 동향 보고서” 2025, (https://www.kisa.or.kr)

[2] R. Muggah and M. Margolis. Global Cybersecurity Outlook 2025. 2025, (https://www.weforum.org/publications/global-cybersecurity-outlook-2025/)

[3] Landolt, Christoph R., et al. "Multi-Agent Reinforcement Learning in Cybersecurity: From Fundamentals to Applications." arXiv preprint arXiv:2505.19837 (2025).

[4] MITRE, “MITRE ATT&CK,” 2024 (https://attack.mitre.org)

비학습 기반 하이브리드 기법을 이용한 LiDAR 차선 검출

강대근, 은승우, 한동석*
경북대학교 전자전기공학부
kangdg@knu.ac.kr, sweunwave@knu.ac.kr, *dshan@knu.ac.kr

Non-Learning-Based Hybrid Framework for LiDAR Lane Detection

Dae Geun Kang, Seung Woo Eun, Dong Seog Han*
Kyungpook National Univ.

요 약

최근 자율주행 및 ADAS의 확산으로 조도 변화·악천후에 취약한 카메라 기반 차선 검출의 한계가 부각되고 있다. 본 논문에서는 외부 광원에 덜 민감한 LiDAR의 반사강도와 기하 정보를 직접 활용하여 비학습 기반으로 동작하는 차선 검출 파이프라인을 제안한다. Region of Interest로 도로 영역을 선별하고 RANSAC 알고리즘으로 노면을 분리한 뒤, 진행 방향 수직축의 반사강도 히스토그램 피크로 좌·우 차선을 초기화한다. 이후 동적 프로그래밍과 평행성 제약으로 프레임 전역에서 연속적이며 잡음·이상치에 강한 차선 경로를 복원한다. 제안하는 연구 방법의 목적은 학습 데이터나 사전 훈련 없이도 다양한 실도로 조건에서 신뢰할 수 있는 차선을 검출하는 데 있다. 실도로 데이터로 검증한 결과, 제안 기법은 직선뿐 아니라 곡률의 변화가 있는 곡선 구간에서도 차선을 안정적으로 추종하며, 거짓 피크에 덜 민감하게 동작하여 형상 안정성을 유지하였다.

Ⅰ. 서 론

자율주행 시스템 및 첨단 운전자 보조 시스템(ADAS, advanced driver-assistance systems) 기술이 빠르게 발전함에 따라, 차량의 안전성과 편의성을 극대화하기 위한 주변 환경 인식 기술의 중요성이 그 어느 때보다 부각 되고 있다. 다양한 인식 기술 중에서도 차선 검출은 차량의 정확한 위치를 파악하고, 주행 경로를 계획하며, 차선 유지 보조(lane keeping assist)와 같은 핵심적인 차량 제어 기능의 필수적인 기술이다.

기존의 차선 검출 연구는 주로 카메라 기반 접근 방식에 의존해 왔다. 카메라는 색상과 질감 정보를 효율적으로 제공하며 낮은 비용으로 구현이 가능하다는 장점이 있으나, 외부 환경 변화에 매우 취약하다. 특히 야간이나 터널, 그림자, 역광과 같은 조도 변화 상황에는 차선의 명암 대비가 약화 되거나 소실되어 인식 성능이 급격히 저하되며, 눈·비·안개 등 악천후 조건에서는 노면 반사 및 렌즈 오염 등으로 인해 인식 신뢰도가 크게 낮아지는 문제가 발생한다.

이러한 한계를 극복하기 위한 대안으로 라이다(LiDAR, Light Detection and Ranging) 센서가 각광받고 있다. 라이다는 능동형 센서로, 외부 광원에 의존하지 않고 야간이나 그림자 환경에서도 안정적인 성능을 제공하며, 3차원 공간에 대한 정밀한 거리 정보를 통해 도로의 구조를 보다 정확하게 파악할 수 있다. 특히 라이다의 반사 강도 정보는 차선 도색과 배경 도로 간의 재질 차이를 기반으로 한 명확한 대비를 제공함으로써, 조도 변화에 상대적으로 강인한 차선 검출이 가능하다는 장점을 가진다[1].

본 논문에서는 이러한 라이다의 특성을 활용하여, 학습 없이도 동작 가능한 비학습 기반 차선 검출 알고리즘을 제안한다.

Ⅱ. 본론

본 논문에서 제안하는 LiDAR 기반 차선 검출 알고리즘의 성능을 검증하기 위해, Ouster OS1-128 3D LiDAR 센서를 장착한 자율주행 실험 차량을 활용했다. 실험은 경북대학교 대구캠퍼스 내 다양한 도로 환경(직선, 곡선, 분기점, 주차장 진입로 등)을 포함하는 구간을 주행하면서 데이터를 수집하는 방식으로 진행된다. 이 과정에서 ROS 2 기반의 센서 드라이버를 통해 LiDAR 데이터를 수집하였으며, 각 프레임은 .pcd 형식으로 변환하여 총 3,310장의 포인트 클라우드를 확보했다.

그림 1은 수집된 LiDAR 포인트 클라우드 데이터를 시각화한 결과를 나타낸다. 그림1에서 도로의 윤곽뿐만 아니라, 차선 도색에 해당하는 영역이 주변 아스팔트에 비해 높은 반사 강도를 가지는 것을 확인할 수 있으며, 이는 차선 검출 알고리즘의 초기 입력으로 사용된다.

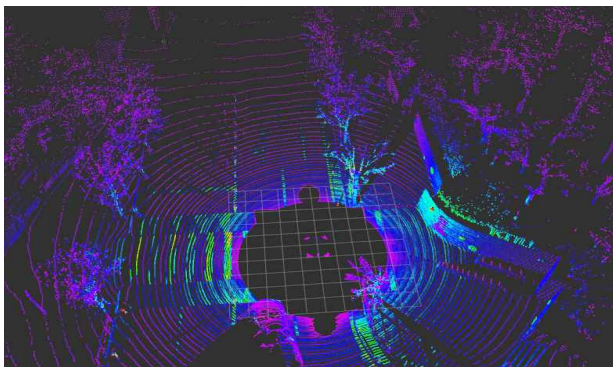
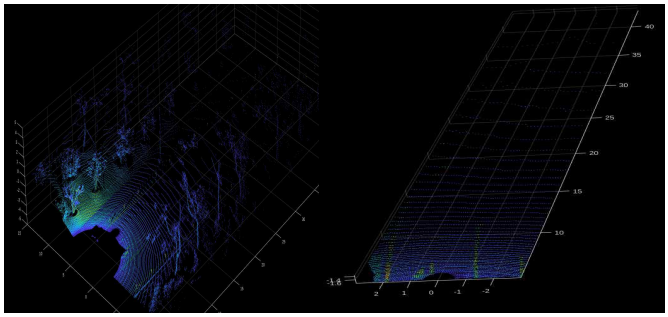


그림 1. 수집된 LiDAR 포인트 클라우드 예시

원본 PCD 데이터에는 도로 이외에도 주변 건물, 가로수, 보행자, 차량 등이 포함되어 있으므로, 전체 포인트 클라우드를 그대로 차선 검출에 사용하는 것은 비효율적이다. 에이 본 연구에서는 차량 좌표계를 기준으로 전방 0 ~ 50m, 좌우 6m, 높이 -2 ~ 2m의 범위를 ROI(Region of Interest)로 설정하였다. ROI 필터링은 차선이 존재할 가능성이 높은 도로 영역만

을 남겨 연산량을 줄이고, 이후 과정의 정확도를 높이는 역할을 한다.

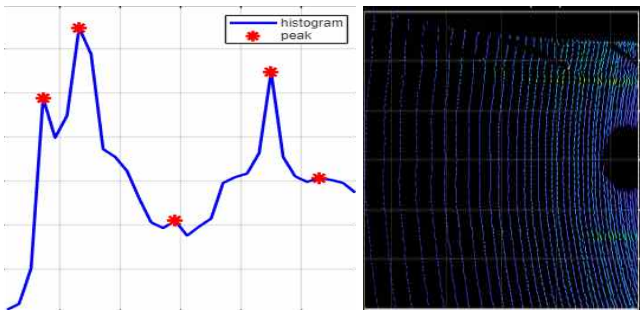
그림 2는 ROI 적용 전과 적용 후를 비교한 것으로, 차선 검출에 필요한 도로 표면 데이터만이 효과적으로 남아 있음을 보여준다.



(a) ROI 적용 전 (b) ROI 적용 후

그림 2. ROI 적용 전과 후 결과

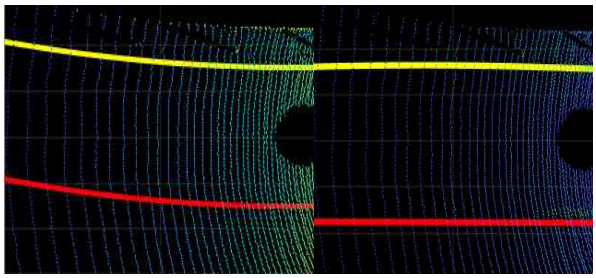
ROI가 적용된 포인트 클라우드에 대해서는 도로 평면을 추출하기 위해 RANSAC(random sample Consensus) 기반 평면 분리 기법을 적용하였다. 이는 잡음과 이상치가 포함된 환경에서도 도로 평면을 안정적으로 모델링할 수 있도록 한다. 평면 분리 후, 도로 위의 반사 강도 값 분포를 분석하여 차선 후보 영역을 찾는다. 아스팔트 대비 차선 도색은 상대적으로 높은 반사율을 보이기 때문에, 이를 이용해 차량 진행 방향과 수직한 축에 대해 히스토그램을 생성한다[2]. 히스토그램 상의 뚜렷한 피크는 좌우 좌선에 해당하며, 이를 통해 초기 차선 후보 지점을 검출한다. 그림 3-(a)는 ROI 내 반사 강도 분포와 히스토그램 피크 검출 결과를 시각화한 것으로, 그림 3-(b)의 실제 차선 위치와 높은 일치를 보인다.



(a) 히스토그램 피크 검출 결과 (b) 실제 차선 위치

그림 3. 히스토그램 피크와 실제 차선 위치

그러나 단순히 히스토그램 피크만을 이용하면 검출 결과가 불연속적이거나 잡음에 민감할 수 있다. 이를 해결하기 위해 본 연구에서는 동적 프로그래밍(Dynamic Programming, DP)을 활용하여 전체 후보 지점 중에서 전역 최적 경로를 탐색하였다. DP 기반 경로 탐색은 프레임 전체에 걸쳐 연속적이고 매끄러운 차선을 형성하며, 특히 곡선 구간 등 잡음으로 인한 이상치(outlier)를 효과적으로 제거한다. 그림 4-(a)는 히스토그램 피크 기반 차선 검출 방법과 히스토그램 피크와 DP 최적 경로를 비교한 결과로, 본 논문의 방법이 도로 차선 형상을 잘 나타내고 있음을 확인할 수 있다.



(a) 제안된 방법 (b) 기존 방법

그림 4. 차선 검출 결과

Ⅲ. 결론

본 논문에서는 학습 기반 접근법을 배제하고 LiDAR 포인트 클라우드 데이터의 기하학적·반사 강도 특성을 직접 활용하여 차선을 검출하는 알고리즘을 제안하였다. 제안된 방법은 ROI 필터링을 통해 도로 영역만을 효과적으로 분리하고, 반사 강도 분포에 기반한 히스토그램 분석으로 초기 차선 후보를 검출하였다. 이후 동적 프로그래밍 기반 최적 경로 탐색과 병렬성 제약 조건을 통해 직선 및 곡선 도로뿐만 아니라 연속적이고 안정적인 차선을 형성할 수 있음을 확인하였다.

실험 결과, Ouster OS1-128 LiDAR 센서를 이용하여 수집한 실제 도로 주행 데이터에서 본 연구의 알고리즘은 기존의 학습 기반 모델이 요구하는 대규모 학습 데이터나 복잡한 학습 절차 없이도 신뢰성 있는 차선 검출이 가능함을 입증하였다. 또한, 각 모듈은 독립적으로 설계되어 있어 시스템 확장성 측면에서도 유리하며, 향후 정밀 지도 제작 및 자율주행 시스템의 경로 계획 단계에서 직접적으로 활용될 수 있다[3].

다만, 본 논문에서는 강도값이 약하거나 마모된 차선의 경우 검출 성능이 저하되는 한계가 있다. 이러한 문제를 보완하기 위해, 향후 연구에서는 차선 도색 패턴의 기하학적 일관성을 추가로 활용하거나, 카메라와 같은 다른 센서와의 융합을 통해 강건성을 더욱 향상 시킬 예정이다.

ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 과학기술사업화진흥원의 지원을 받아 수행된 연구임('학연협력플랫폼구축 시범사업' RS-2023-00304695).

참 고 문 헌

[1] Hata, A., and Wolf, D., “Road Marking Detection Using LiDAR Reflective Intensity Data and Its Application to Vehicle Localization,” Proc. IEEE Intelligent Transportation Systems Conference (ITSC), pp. 584 - 589, 2014.

[2] Jung, J., and Bae, S.-H., “Real-Time Road Lane Detection in Urban Areas Using LiDAR Data,” Electronics, 7(11):276, 2018. DOI: 10.3390/electronics7110276.

[3] Dong, H., Gu, W., Zhang, X., Xu, J., Ai, R., Lu, H., Kannala, J., and Chen, X., “SuperFusion: Multilevel LiDAR - Camera Fusion for Long-Range HD Map Generation,” Proc. IEEE Int. Conf. Robotics and Automation (ICRA), 2024.

비지도 학습 기반의 TEM 이미지 결손 원자 탐지 알고리즘

김민성, 이상철*
한국기초과학지원연구원
kms00737@kbsi.re.kr, *sclee02@kbsi.re.kr

요 약

물성 분석에서 투과전자현미경 이미지를 통한 결손 원자 검출은 전문가의 수작업 판독에 의존하는 한계가 있다. 이러한 점을 해결하기 위해 본 연구에서는 주기적인 원자 배열로 구성된 결정 구조 내 결손 원자를 자동으로 탐지하는 비지도 학습 기반 알고리즘을 제안한다. 제안된 기법은 원본 영상에서 원자 중심점을 추출한 영상을 입력으로 사용한다. 이후 들로네 삼각분할과 필터링 과정을 통해 구조적 연결 정보를 확보하고, 원자 간 최근접 이웃 선분의 개수에 기반한 지역 벡터 패턴을 추출한다. 이렇게 얻어진 패턴들을 일반화하여 대표적인 벡터 집합으로 단순화한 뒤, 각 원자점 주변의 이웃 관계와 대조함으로써 결손 가능성이 높은 위치를 효과적으로 추정한다. 제안된 기법의 성능을 검증하기 위해 TEMImageNet 데이터셋의 다양한 격자 구조를 대상으로 실험을 수행하였다. 그 결과, 결손 비율이 증가하더라도 정밀도가 안정적으로 유지되어 결손 위치 판별의 신뢰성을 확인할 수 있었다.

Ⅰ. 서 론

결정 재료 내 원자 단위의 결함은 물성에 중대한 영향을 미치므로, 이를 탐지하는 기술은 재료 특성 분석에 필수적이다. 이 과정에서 투과전자현미경은 원자 수준의 해상도를 바탕으로 결함 구조를 직접 관측할 수 있는 핵심 도구로 활용되고 있다. 그러나 촬영 조건, 대비의 민감도와 같은 요인으로 인해 영상 내 결함 식별이 어렵다. 또한, 결함 탐지 자동화 도구가 일부 개발되었으나 샘플의 구조적 다양성과 사용자의 요구를 충분히 반영할 수 없기 때문에 실제 현장 분석은 여전히 전문가의 경험에 크게 의존하고 있다.

본 연구에서는 결손 원자를 자동으로 식별하기 위해 비지도 학습 기반의 탐지 기법을 제안한다. 제안하는 방법은 사전 학습 데이터나 결정 구조에 대한 사전 정보 없이, 원자 간의 기하학적 배치를 분석하여 반복적으로 나타나는 격자 패턴을 도출한다. 이후 각 원자 위치에서 도출된 패턴과 대조함으로써, 결손 발생 가능성이 높은 지점을 효과적으로 추출한다. 이러한 접근은 전문가의 경험과 수작업에 의존하던 기존 분석 방식을 대체할 수 있으며, 다양한 결정 구조와 복잡한 재료 시스템에도 확장 가능한 자동화 탐지 기반을 제공한다.

Ⅱ. 본론

본 연구의 전체 과정은 규칙적인 배열 구조를 갖는 원자 좌표를 기반으로, 결손 가능성이 높은 위치를 추출하는 것을 목표로 한다. 이를 위해 세 단계 과정을 거쳐 결손 원자 위치를 도출한다. 입력 이미지는 256×256 해상도에서 검정 배경 위에 각 원자 중심을 1픽셀 흰점으로 표시한 형태이며, 제안된 방법의 세부 내용은 다음 절에서 설명한다.

2.1. 들로네 삼각분할을 이용한 그래프 생성 및 지역 벡터 추정

규칙적인 원자 중심 좌표에서 결손 위치를 탐지하기 위해서는 점간의 구조적 연결 관계를 파악할 필요가 있다. 들로네 삼각분할은 각 점을 삼각형으로 연결하되 외접원 내부에 다른 점이 존재하지 않도록 함으로써, 점 간의 관계를 격자 구조로 변환하여 구조적 연결성을 나타낸다. 그러나 이 방법은 모든 점을 최소 삼각형으로 연결하기 때문에 그림 1의 (a)에서 보이는 것처럼 불필요한 장거리 연결선이나 규칙적 구조 내의 대각선 변이 포함될 수 있다. 이에 따라, 일정한 면적을 유지하는 규칙적 원자 배열을 얻기 위해 전역 평균 길이보다 짧은 변만을 남기는 방식을 적용하여 불필요한 연결을 제거한다. 이후 그림 1의 (b)와 같이 구축된 그래프를 바탕으로 각 원자에 대해 인접 원자까지의 벡터를 계산하여 거리와 방향 정보를 포함하는 지역 벡터를 추정한다. 지역 벡터는 점 간의 각도와 규칙적 배열 패턴을 반영하고 있어, 추후 결손 원자 후보 위치를 예측하는 핵심 정보로 활용된다.

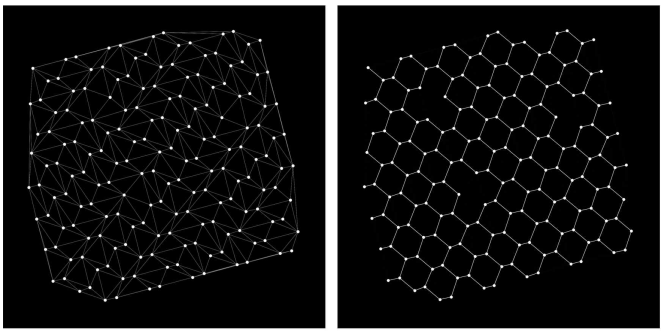


그림 1. (a) 들로네 삼각분할 결과, (b) 평균 선분 길이 기반 필터링 결과

2.2. 들로네 삼각분할 기반 필드 추정을 이용한 저밀도 후보 탐색

모든 원자 점에 대해 지역 벡터를 직접 적용해 결손 후보를 탐색하는 방식은 불필요한 연산이 많아 계산 비용과 시간이 크게 증가할 수 있다. 따라서, 결손 가능성이 높은 영역을 우선 검출해 탐색 대상을 줄이는 과정이 필요하다. 이를 위해 본 연구에서는 들로네 삼각분할 기반 필드 추정(Delaunay Triangulation Field Estimator, DTFE)을 사용한다. DTFE는 들로네 삼각분할을 기반으로 각 삼각형의 면적을 인접 정점에 분배해 국소 밀도를 추정하는 기법으로, 면적간의 밀도 변화를 통해 결손 위치 인근 지점을 효과적으로 탐지할 수 있다. 그림 2의 (a)는 DTFE 결과로, 짙은 회색의 저밀도 영역이 결손 원자 위치와 근접하며, 인접 영역의 결손 후보점은 그림2의 (b)에서 X로 표시하였다.

2.3. 결손 위치 추정

앞서 선별한 각 저밀도 후보점에 대해 인접 원자와의 벡터 방향을 분석

하여, 사전에 정의된 지역 벡터 중 가장 유사한 벡터를 선정한다. 이 과정에서 인접 원자와의 방향이 다수 매칭되고 결손 발생 가능성이 높은 벡터를 선택함으로써, 그림 2(b)에 나타난 노란색 삼각형과 같은 결손 후보 위치를 산출한다. 이때, 생성된 후보 위치 중 기존 원자점과 과도하게 가까운 경우는 오탐 가능성이 높으므로 기존 원자 간 최소 평균 거리를 기준으로 필터링을 수행한다. 이후 남은 후보 점들은 위치 중복을 제거하고 대표 좌표를 결정하기 위해 DBSCAN(Density-Based Spatial Clustering of Applications with Noise) 알고리즘을 적용한다. DBSCAN은 인접 밀도 기반의 군집화 방식으로, 노이즈를 배제하면서 인접한 후보 점들을 하나의 군집으로 묶을 수 있어 결손 위치의 안정적인 산출이 가능하다. 이 과정을 통해 최종 후보 점들의 각 군집 중심점을 결손 원자 위치로 확정한다.

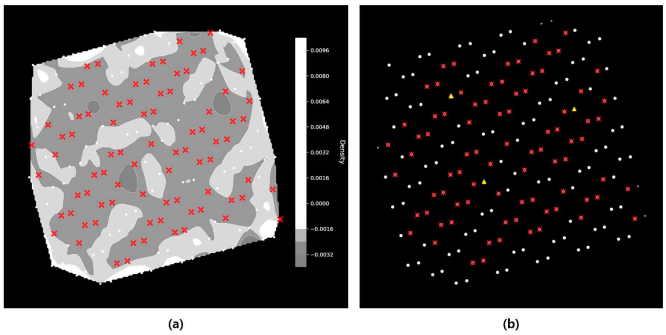


그림 2. (a) DTFE 기반 후보점 생성 결과 (b) 결손 원자 후보 위치 추정 결과 (○: 기존 원자, ×: 저밀도 후보 지점, △: 결손 원자 위치 추정 지점)

Ⅲ. 실험

제안한 결손 원자 탐지 알고리즘의 성능을 검증하기 위해 TEMimagenet [1] 데이터셋을 기반으로 실험을 수행하였다. TEMimagenet에서 규칙적인 원자 배열을 갖는 이미지를 선별하기 위해 원자 개수가 30개 이상인 6,588장을 추출하였다. 성능 평가는 검출된 좌표가 실제 결손 위치로부터 1픽셀 이내에 존재할 경우 정탐(True Positive)으로 판정하는 기준을 적용하였다. 또한, 오탐과 실제 결손 검출 비율을 정량화하기 위해 정밀도(Precision)와 재현율(Recall)을 사용하였으며, 종합적인 평가를 위해 F1-score를 함께 활용하였다.

표 1은 중앙부에 1~3개의 랜덤하게 결손을 삽입한 경우의 결과를 나타낸다. 이 조건에서 전체 합산 기준 정밀도는 0.962, 재현율은 0.952, F1-score는 0.957을 기록하여, 제안된 기법이 규칙적인 원자 배열 구조에서 결손 위치를 높은 정확도와 재현율로 안정적으로 탐지함을 보여준다.

또한, 이미지 내 원자 수 대비 결손 비율을 10%, 20%, 30%로 달리하여 성능을 비교한 결과는 표 2와 같다. 결손 비율이 증가함에 따라 재현율과 F1-score는 점진적으로 감소하는 경향을 보였으나, 정밀도는 모든 경우에서 0.92 이상으로 높게 유지되었다. 이는 제안된 알고리즘이 결손 비율이 높아져도 불필요한 오탐(False Positive)을 최소화하며, 검출한 결손 위치의 신뢰도가 높다는 것을 의미한다. 결손 비율이 높아질수록 재현율이 하락한 이유는, 결손이 많아짐에 따라 격자의 규칙적 패턴이 점점 사라져, 지역 벡터 기반 패턴 분석에서 일부 결손 위치를 탐지하지 못했기 때문이다. 이러한 결과는 규칙적인 패턴이 유지되는 영역에서는 제안된 알고리즘이 항상 높은 예측 신뢰도를 보이며 결손 위치를 정확히 식별할 수 있음을 의미한다.

그림 3은 원본(GT) 이미지와 결손 비율을 10%, 20%, 30%로 변화시켰을 때의 결과를 보여준다. 흰색 동그란 점은 기존 원자 위치를, 노란색 네모 점은 제거된 원자(결손 위치)를 나타낸다.

표 1. 랜덤한 결손 생성 시 성능 결과

vacancy	Precision	Recall	F1-score
1~3	0.962	0.952	0.957

표 2. 원자 수 대비 결손의 비율에 따른 성능 비교

vacancy ratio	Precision	Recall	F1-score
10%	0.955	0.759	0.845
20%	0.948	0.725	0.822
30%	0.922	0.672	0.777

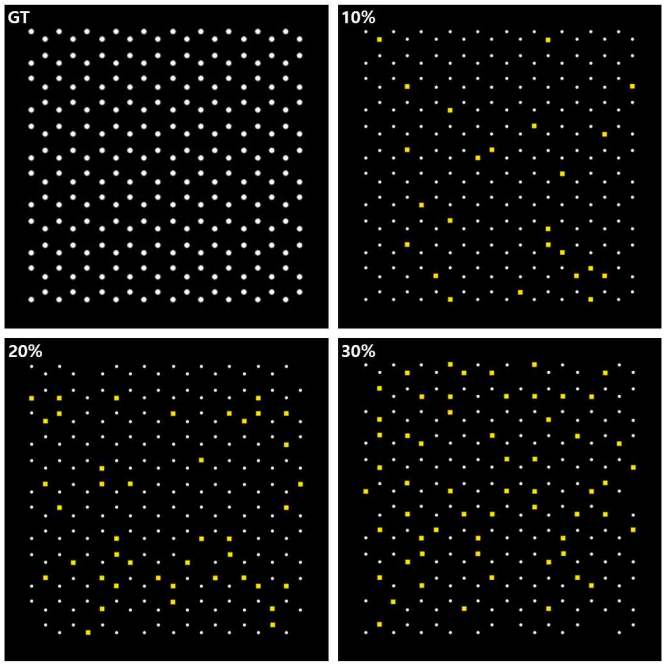


그림 3. 결손 비율에 따른 결손 원자 탐지 결과 (○: 기존 원자, □: 최종 결손 원자 위치 추정 지점)

Ⅳ. 결론

본 논문에서는 사전 학습 데이터나 결정 구조에 대한 사전 정보 없이 결손 원자를 자동 탐지할 수 있는 비지도 학습 기반 기법을 제안하였다. 제안 방법은 원자 좌표로부터 지역적인 규칙 격자 패턴을 추출하고, 국소 밀도 분석을 통해 결손 가능성이 높은 영역을 선별한 뒤, 군집화를 통해 최종 결손 원자 위치를 산출하는 방식이다. TEMimagenet 데이터셋을 활용한 실험에서 제안 기법은 높은 정밀도와 재현율을 달성하였다. 이러한 결과는 본 방법이 다양한 규칙적 원자 배열 구조에서 결손 위치를 신뢰성 있게 탐지할 수 있음을 보여준다.

ACKNOWLEDGMENT

This research was supported by the basic science research program (C524200) funded by the Korea Basic Science Institute.

참 고 문 헌

[1] J. Madsen, M. R. Clausen, T. S. Jespersen, et al., "TEMImageNet: A large-scale multi-domain dataset for electron microscopy image analysis," arXiv preprint arXiv:2003.10585, 2020.

현장 신뢰성 향상을 위한 심각도 추정 · 인과 분석 통합 기반 웨이퍼 결함 분석 및 자연어 리포트 생성

김서현, 박현희*
명지대학교 인공지능융합학과
seohy2on, hhpark* @ mju.ac.kr

Wafer Defect Analysis and NL Report Generation Based on Integrated Severity Estimation and Causal Analysis for On-Site Reliability

Seohyeon Kim, Hyunhee Park*
Dept. of Artificial Intelligence Convergence, Myongji University

요 약

반도체 제조 공정에서 웨이퍼 결함의 신속하고 정확한 분석은 수율 및 제품 신뢰성에 직결된다. 미세 결함은 제품 성능에 치명적이지만, 기존 검사 방법은 탐지 및 분류 정확도에 한계가 있다. 본 논문에서는 실제 현장에서 결함 판별이 어려운 점을 해결하기 위해, Mask2Former 기반의 픽셀 단위 세그멘테이션을 통한 결함 탐지를 수행한다. 나아가, 심각도 정량화 모델링과 통계적 패턴 분석을 통한 인과적 편향 분석을 통합하여, 근거 중심의 자연어 리포트를 생성한다. 이를 통해 해석 가능성과 신뢰성을 높이는 통합형 결함 분석 시스템을 제안한다.

Ⅰ. 소 개

반도체 제조 공장에서 웨이퍼 결함은 제조 경쟁력 확보에 중요한 요인이다. 이를 효과적으로 탐지하고 분석하는 것은 현대 반도체 산업의 경쟁력 확보에 필수적이다.

웨이브릿 변환, 공간 필터링, KNN 등 기존의 웨이퍼 결함 분석 방법들은 복잡한 배경 노이즈와 미세한 결함 간의 구별에서 탐지 성능이 낮았다[1]. 특히 서로 다른 결함 유형들이 유사한 특성을 보이는 경우 분류 성능이 매우 저하되었다. 또한, 대부분의 기존 연구는 탐지 및 분류에 치중했고, 원인과 공정의 상관관계 분석은 수동으로 이루어졌다. 이와 더불어, 딥러닝 모델의 특성으로 인해 분석 결과에 대한 신뢰성과 설명 가능성이 부족하였다.

기존 연구들의 한계점을 해결하기 위해, 본 논문에서는 웨이퍼 결함 분석 시스템을 제안한다. 이는 딥러닝 기반의 정밀 결함 탐지, 회귀 모델 기반의 심각도 추정, 그리고 통계적 패턴 분석을 통한 인과 관계 추론을 통합한 웨이퍼 결함 분석 시스템이다. 또한, 분석된 데이터를 기반으로 근거 중심의 자연어 리포트를 생성한다. 이를 통해 신뢰성을 높이며, 웨이퍼 상태를 직관적으로 이해하고 신속하게 의사결정을 내릴 수 있도록 한다.

Ⅱ. 시스템 아키텍처

본 논문에서 제안하는 웨이퍼 결함 분석 시스템은 다섯 모듈(탐지, 기하학 정보 추출, 심각도 추정, 인과 분석, 리포트 생성)로 구성된다. 각 모듈은 독립적으로 동작하면서도 상호 연동하여 결함을 종합적으로 분석한다.

데이터셋 및 전처리 본 논문에서는 WM811K 웨이퍼 맵 데이터셋을 사용한다. 전체 데이터셋 중 'Edge-Loc', 'Loc', 'Scratch' 세 가지 결함 클래스를 선별하여 실험을 진행한다. 'Scratch' 클래스의 경우 다른 두 클래스에 비해 상대적으로 적은 샘플 수를 보여 데이터 불균형 문제를 해결하기 위해 데이터 증강 기법을 적용한다.

딥러닝 기반 결함 탐지 모듈 결함 탐지의 정확도 향상을 위해 Mask2Former[2] 아키텍처를 활용한 픽셀 단위 세그멘테이션 모델을 구현한다. Mask2Former 는 범용 세그멘테이션 아키텍처로, 픽셀 단위에서 객체와 영역을 정밀하게 구분할 수 있다. 학습된 모델을 활용하여 웨이퍼 이미지의 각 픽셀을 분석하고 결함 영역을 탐지한다.

웨이퍼 이미지의 특성상 주기적 패턴과 미세 결함이 겹쳐 보이기 때문에, 딥러닝 모델만으로는 탐지 성능이 제한될 수 있다. 이를 보완하기 위해, 색상 분석을 적용한다. 넓은 영역에 걸쳐 나타나는 결함의 경우 노란색 계열의 색상 특성을 활용한 탐지 알고리즘을 구현한다. 또한, 미세한 결함에 대해서는 밝기 대비 분석, 경계선 검출 등의 기법을 적용하여 탐지 성능을 향상시킨다.

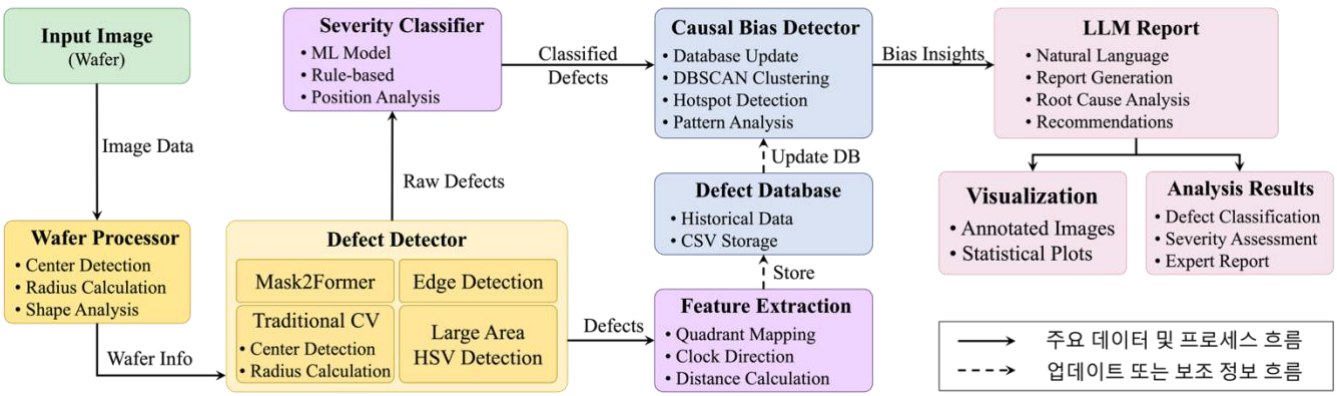


그림 1. 웨이퍼 결함 분석 시스템 아키텍처

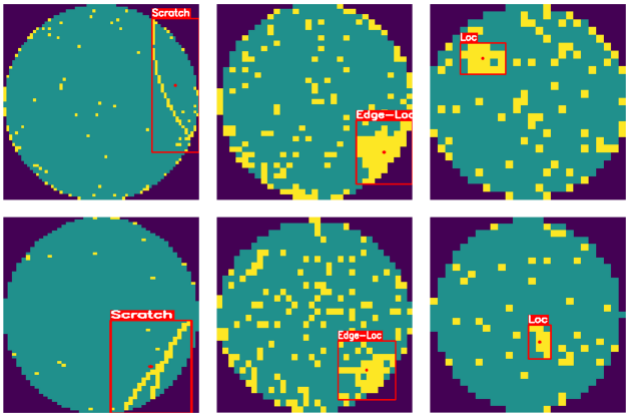


그림 2. 결함 탐지 시각화 예시

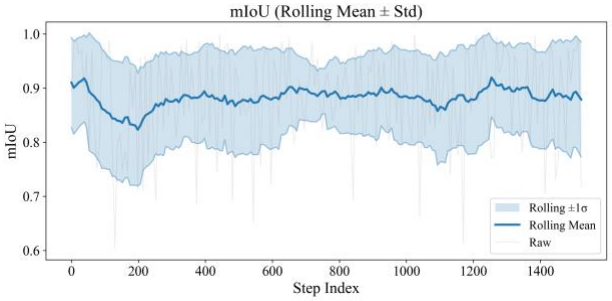


그림 3. Mask2Former 학습 과정별 mIoU 성능 변화

그림 3 은 Mask2Former 기반 세그멘테이션 모델의 학습 과정에서 mIoU(mean Intersection over Union) 지표의 변화를 보여준다. 초기 100 스텝에서 급격한 성능 향상을 보인 후, 200-300 스텝 구간에서 일시적인 성능 변동이 관찰되었다. 그러나, 400 스텝 이후부터는 안정적으로 수렴하는 양상을 나타내며, 최종적으로 0.94 의 mIoU 값을 달성하였다. 이는 웨이퍼 결함 영역과 정상 영역을 정밀하게 구분할 수 있음을 보여주며, 후속 모듈인 기하학적 특성 분석과 심각도 추정의 정확도 향상에 직접적으로 기여한다.

웨이퍼 기하학적 특성 분석 웨이퍼의 중심점과 반지름을 자동으로 추출하는 알고리즘을 포함한다. 이는 허프 변환(Hough Transform)과 원 검출 알고리즘을 기반으로 구현한다. 탐지된 결함이 웨이퍼의 중심부와 가장자리 중 어디에 위치하는지를 정확히 판단할 수 있게 한다. 이러한 위치 정보는 심각도 추정과 인과 관계 분석에서 중요한 특징으로 활용된다.

회귀 모델 기반 심각도 추정 탐지된 결함의 심각도를 추정하는 회귀 모델을 구현한다. 결함의 면적, 웨이퍼 중심에서의 거리, 가장자리 근접성 등 다양한 수치적 특징을 추출하고, 이를 바탕으로 머신러닝 회귀 모델을 학습한다. 모델 학습 과정에서 결함 특징과 실제 심각도 점수 간의 상관관계를 학습하여, 신규 결함에 대한 심각도를 자동으로 예측할 수 있다. 예측된 점수는 1(낮음), 2(중간), 3(높음)의 3 단계 등급으로 분류되어 우선순위를 쉽게 판단할 수 있도록 한다. 심각도 추정을 통해 결함의 정량적 위험도를 평가함으로써 생산 라인 의사결정의 우선순위를 직접 지원한다.

통계적 패턴 분석을 위한 인과적 편향 탐지 누적된 웨이퍼 결함 데이터를 분석하여 시스템적 편향과 공정상의 근본적 문제를 식별한다. 모든 분석 결과는 CSV 형태로 축적되며, 결함들의 정규화된 위치 좌표를 DBSCAN(Density-Based Spatial Clustering of Application with Noise) 알고리즘을 통해 분석한다. 따라서 특정 위치에 결함이 비정상적으로 집중되는 '핫스팟(Hotspot)' 영역을 자동으로 탐지하고, 해당 영역과 관련된 경고 메시지를 생성한다.

자연어 리포트 생성 각 분석 모듈에서 생성된 결함 종류, 위치, 심각도, 편향 분석 결과 등을 종합하여 근거 기반의 해석 가능한 보고서를 생성한다. 본 논문에서는 경량 LLM 인 TinyLlama[3]를 활용하여 결함의 위치(사분면, 시계 방향), 인과적 특성 등 핵심 정보를 바탕으로 신뢰성 높은 리포트를 생성한다. 그 결과, 설명 가능성(Explainability)과 실용성을 동시에 확보한다.

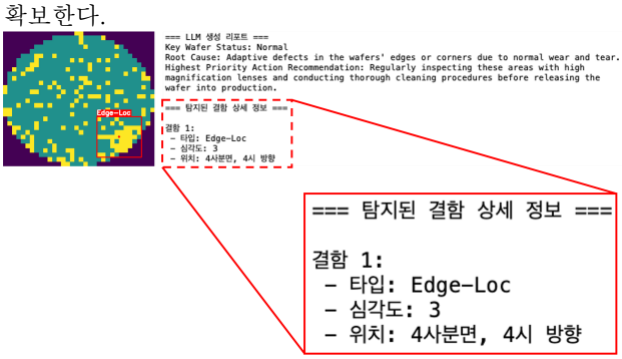


그림 4. 자연어 리포트 생성 예시

Ⅲ. 결 론

본 논문에서는 반도체 제조 현장의 신뢰성 향상을 위한 통합형 웨이퍼 결함 분석 시스템을 제안하였다. Mask2Former 기반의 세그멘테이션을 활용한 결함 탐지, 회귀 모델을 활용한 심각도 추정, DBSCAN 알고리즘을 이용한 패턴 기반 인과적 편향 분석을 통합하여 종합적인 웨이퍼 결함 탐지 분석 시스템을 구현하였다.

본 시스템은 단일 목적 탐지의 한계를 넘어, 탐지부터 심각도 추정 및 분석, 리포트 생성까지의 통합 과정을 구현한다. 이를 통해 반도체 제조 현장에서의 불량률 감소와 공정 최적화에 기여할 것으로 기대한다. 향후 연구에서는 더 다양한 결함 유형에 대한 확장과 실시간 분석 성능 최적화를 통해 실제 생산 라인에 적용 가능한 수준으로 발전시킬 예정이다.

ACKNOWLEDGMENT

이 논문은 정부(교육부)의 재원으로 한국연구재단과 중소벤처기업부의 기술개발사업 지원을 받아 수행된 연구임(No. 2022R1A2C2005705, 분산 머신러닝 기반 지능형 플라잉 기지국을 위한 AI-MAC 프로토콜, RS-2025-02217793, 제습모듈을 내장한 웨이퍼 이송장비 및 편축보정기술개발과 딥러닝을 활용한 광학문자인식 및 Chipping 검출을 포함한 하이브리드 검사장비 기술 개발)

참 고 문 헌

[1] Jianhong Ma, Tao Zhang, Cong Yang, Yangjie Cao, Lipeng Xie, Hui Tian, and Xuexiang Li. "Review of Wafer Surface Defect Detection Methods." *Electronics*, vol.12, no.8, Apr.2023, doi:10.3390/electronics12081787.

[2] Bowen Cheng, Ishan Misra, Alexander G. Schwing, Alexander Kirillov, and Rohit Girdhar. "Masked-attention Mask Transformer for Universal Image Segmentation." arXiv preprint arXiv:2112.01527 (2021).

[3] Peiyuan Zhang, Guangtao Zeng, Tianduo Wang, and Wei Lu. "TinyLlama: An Open-Source Small Language Model." arXiv preprint arXiv:2401.02385 (2024).

주파수 대역 분할 및 그룹 컨볼루션을 이용한 개선된 고속 푸리에 컨볼루션

허재호, 배주한, 주현철, 나태영
에스케이텔레콤

{jaeho.hur, juhan.bae, hc.joo, taeyoung.na}@sk.com

Improved Fast Fourier Convolution Using Frequency Band Division and Grouped Convolution

Jaeho Hur, Juhan Bae, Hyunchul Joo, Taeyoung Na
SK Telecom

요 약

본 논문에서는 기존 Fast Fourier Convolution(FFC)의 모델링 역량을 개선하기 위한 새로운 접근법을 제안한다. 기존 FFC 는 주파수 도메인에서 모든 주파수 성분에 동일한 1×1 Convolution 가중치를 적용하는 한계를 가지고 있어, 주파수 성분별로 차별화된 학습이 불가능하다. 본 논문에서는 입력 Feature 를 균등한 크기의 여러 하위 주파수 대역으로 분할하고, 각 주파수 대역마다 독립적인 가중치를 적용할 수 있는 Grouped Convolution 기반의 새로운 FFC 구조를 제안한다.

Ⅰ. 서 론

Deep Neural Network 에서 Convolutional Neural Network 의 Receptive Field 를 효과적으로 확장하여 모델링 역량을 높이기 위해 Fast Fourier Convolution (FFC)[1]이 제안되었다. FFC 는 Fast Fourier Transform 을 기반으로 주파수 도메인에서 하나의 Feature 값을 변경하여 Spatial Domain 의 입력 Feature 전체에 영향을 미칠 수 있는 성질을 이용한다. 이를 통해 보통의 Convolution 과 달리 Receptive Field 의 범위를 전역으로 넓히는 것이 가능하며, Local Receptive Field 와 Non-local Receptive Field 의 수용 및 처리를 하나의 Unit 에서 동시에 가능하도록 하는 효과적인 앙상블 구조를 제안하였다.

그러나 기존 FFC 의 핵심 구성 요소인 Fourier Unit(FU)에서는 주파수 도메인에 적용되는 1×1 Convolution 의 가중치가 모든 주파수 성분에 동일하게 적용되기 때문에, 주파수 성분마다 다른 Convolution 가중치로 학습하는 것이 불가능한 한계가 존재한다.

본 논문에서는 이러한 기존 FFC 의 제한된 모델링 역량을 개선하기 위해 주파수 대역 분할 및 Grouped Convolution[2] 기반의 새로운 FFC 구조를 제안한다.

Ⅱ. 제안하는 방법

이번 장에서는 기존 FFC 의 핵심 연산 단위인 FU 에 대해 간략히 설명하고, 제안하는 방법이 주파수 대역 분할과 Grouped Convolution 을 이용하여 기존 FU 의 한계점을 어떻게 개선하는지 설명한다.

(1) 기존 FU 구조

FU 는 먼저 Fast Fourier Transform(이하 FFT)을 통해 입력 Feature 를 주파수 도메인으로 변환하고, 주파수 도메인에서 1×1 Convolution, Batch Normalization, ReLU 를 통해 일반적인 Deep Neural Network 의 연산을 수행한다. 마지막 과정에서 Inverse FFT 를 통해 다시 Spatial Domain 으로 Feature 를 복원하여 마무리한다. 여기에서 주파수 도메인에 적용되는 1×1 Convolution 의 가중치는 모든 주파수 성분에 동일하게 적용되기 때문에, 주파수 성분 마다 다른 Convolution 가중치로 학습하는 것이 불가능한 한계가 있다. 그림 1 의 (a)는 기존 FU 의 구조를 나타낸다.

(2) 제안하는 개선된 FU 구조

제안하는 FU 는 기존 FU 와 마찬가지로 FFT 를 통해 입력 Feature 를 주파수 도메인으로 변환하는 과정부터 시작한다. 이후, Frequency Splitter, 1×1 K-Grouped Convolution, Frequency Band Merger 와 같은 새로운 구성 요소들을 거쳐 최종적으로 Inverse FFT 를 통해 Feature 를 복원한다. 이러한 구조는 연산 효율성을 극대화하면서도 각 주파수 대역별로 정교한 모델링을 가능하게 하여 전체 모델의 표현력을 향상시킨다. 그림 1 의 (b)는 제안하는 개선된 FU 구조를 나타내며, 세부 동작에 대한 설명은 다음과 같다.

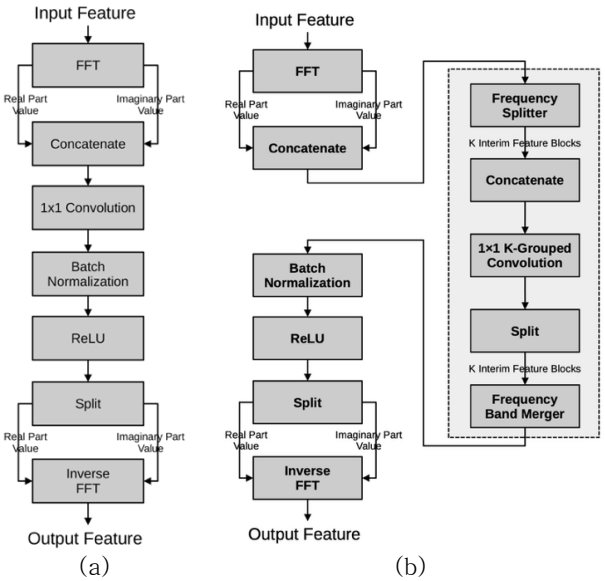


그림 1. 기존의 FU 와 제안하는 개선된 FU 구조: (a) 기존 FU 구조, (b) 주파수 분할 및 Grouped Convolution 기반의 개선된 FU 구조

Frequency Splitter: FFT 를 통해 얻어진 주파수 도메인의 Feature Tensor 를 균등한 크기의 여러 개의 하위 주파수 대역으로 분할한다. 분할의 기준은 텐서의 공간 차원(Height, Width) 상에서 N×M 크기로 정확하게 K 개로 나누는 것이다. 이 과정을 통해 K 개의 Interim Feature Block 들이 생성되며, 이는 주파수 도메인에서의 저주파 및 고주파 성분들을 물리적으로 분리하는 효과를 가져온다.

Concatenate: 분할된 K 개의 Interim Feature Block 들은 첫 번째 채널 축을 기준으로 연접(Concatenate)되어, 1×1 K-Grouped Convolution 의 입력으로 사용할 수 있는 단일한 Tensor 형태로 변환된다.

1×1 K-Grouped Convolution: 연접된 Feature Tensor 는 K 개의 그룹을 가진 1×1 Grouped Convolution 레이어에 입력된다. 이 레이어는 각 그룹(즉, 각 주파수 대역)에 대해 독립적인 1×1 Convolution 을 수행하여, 각 주파수 대역의 특성에 맞는 개별적인 가중치를 학습하게 된다. 이 단계는 제안하는 방법의 핵심으로, 기존 FU 의 단일 1×1 Convolution 이 가지는 모델링 역량의 한계를 극복하는 역할을 한다.

Split/Frequency Band Merger: 1×1 K-Grouped Convolution 의 출력은 다시 첫 번째 채널 축을 기준으로 분리(Split)되어 개별 Frequency Band 에 대응하는 Interim Feature Block 들로 복원된다. 이후, Frequency Band Merger 를 통해 분리된 주파수 대역들이 원래의 공간 차원 위치로 재배열 및 병합된다. 이 과정은 Grouped Convolution 으로 처리된 정보가 기존의 FFC 구조에 다시 원활하게 통합되도록 보장한다.

제안하는 새로운 FU 는 기존 FFC 의 FU 를 완전히 대체할 수 있으며, 전체 FFC 네트워크 구조의 큰 변경 없이도 모델의 표현력을 획기적으로 향상시킬 수 있는 이점을 가진다. Grouped Convolution 의 채택은 구현의 용이성과 병렬 연산 효율성 극대화 측면에서도 긍정적인 효과를 가져올 수 있다.

Ⅲ. 실험 결과

제안하는 방법의 효과를 객관적으로 검증하기 위해, Image Inpainting Task 를 실험 대상으로 선정하였다. Image Inpainting 은 이미지의 손상되거나 가려진 부분을 주변 맥락에 맞게 자연스럽게 복원하는 작업으로, 전역적인 이미지 구조와 질감을 이해하는 능력이 요구된다. 따라서 FFC 와 같은 전역적 정보 처리 모델의 성능을 평가하기에 매우 적합한 Task 이다.

성능 비교를 위해, 기존 FFC 기반 Image Inpainting 모델인 Large Mask Inpainting[3] 을 기준으로 설정하였다. 제안하는 모델은 이 기준 모델의 FU 를 새롭게 개선한 FU 로 교체하여 구성하였다. 모델의 성능은 FID, LPIPS, SSIM 등의 세 가지 주요 평가 지표를 사용하여 측정하였다.

표 1. Image Inpainting Task 성능 비교

평가 지표	기존 FFC 기반 Inpainting Model	제안하는 FFC 기반 Inpainting Model
FID (↓)	42.94	41.514
LPIPS (↓)	0.126	0.121
SSIM (↑)	0.871	0.873

표 1 에서 나타내는 바와 같이, 실험 결과는 FID, LPIPS, SSIM 모든 지표에서 새로운 FFC 기반 모델이 기존 모델을 능가하는 성능을 보였다. 특히, FID 가 42.94 에서 41.514 로 감소한 것은 매우 중요한 결과이다. FID 는 생성된 이미지의 전체적인 품질과 사실성을 종합적으로 평가하는 지표로, 이 수치의 개선은 제안하는 FU 구조가 기존 구조보다 더 풍부한 특징을 학습하고, 실제 이미지의 데이터 분포에 더 가까운 고품질의 이미지를 생성할 수 있게 되었음을 의미한다. 이는 본 논문이 해결하고자 했던 기존 FFC 의 제한된 모델링 역량 문제를 성공적으로 해결했음을 입증한다.

또한, LPIPS 와 SSIM 의 개선은 생성된 이미지가 원본 이미지와 지각적 및 구조적으로 더 유사하다는 것을 의미한다. 이는 복원된 영역이 주변부와 더 자연스럽게 어우러지고, 이미지의 핵심적인 구조적 정보가 손상 없이 잘 복원되었음을 나타낸다. 이러한 결과는 주파수 대역별로 특화된 학습을 통해 모델이 저주파 성분(전역적 구조)과 고주파 성분(국소적 디테일)을 모두 효과적으로 처리할 수 있게 되었기 때문이다. Grouped Convolution 을 주파수 대역 분할과 결합한 것은 단순한 연산 효율성 향상을 넘어, FFC 의 표현력을 근본적으로 향상시키는 구조적 효과를 가져왔다.

Ⅳ. 결론

본 논문은 기존 FFC 의 제한된 모델링 역량 문제를 해결하기 위한 새로운 방법을 제시하였다. 제안하는 방법은 기존 FFC 의 핵심 유닛인 FU 의 구조를 개선하기 위해 주파수 대역 분할 및 Grouped Convolution 을 도입하여, 주파수별 특성을 독립적으로 학습할 수 있는 메커니즘을 구축했다. 이를 통해 Grouped Convolution 이 가진 연산 효율성 이점을 활용하면서도, 주파수 도메인에서의 정교한 모델링을 통해 FFC 의 표현력을 성공적으로 향상시킬 수 있음을 입증하였다.

ACKNOWLEDGMENT

본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2024 년도 문화체육관광연구개발사업으로 수행되었음 (과제명: 전통예술 가무악의 융복합 공연제작 활성화를 위한 융복합 공연 기획/제작 플랫폼 기술 개발, 과제번호: RS-2024-00398536, 기여율: 100%)

참 고 문 헌

[1] Lu Chi, Borui Jiang, and Yadong Mu. Fast fourier convolution. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, Advances in Neural Information Processing Systems, volume 33, pages 4479–4488. Curran Associates, Inc., 2020.

[2] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. Advances in neural information processing systems, 25.

[3] Roman Suvorov, Elizaveta Logacheva, Anton Mashikhin, Anastasia Remizova, Arsenii Ashukha, Aleksei Silvestrov, Naejin Kong, Harshith Goka, Kiwoong Park, and Victor Lempitsky. Resolution-robust large mask inpainting with fourier convolutions. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pages 3172–3182, 2022.

다목적 강화학습 기반 최적 산업구조 탐색에 관한 연구

홍윤기 [1,2], 이상철 [1], 조한얼 [1,3], Nabil Belacel[4], 김찬수 [1,2,*]
1 한국과학기술연구원, 2 과학기술연합대학원대학교, 3 세종대학교, 4 Canada National Research Council

orange4381@kist.re.kr, *eau@ust.ac.kr

Searching for the Optimal Industrial Structure using Multi-Objective Reinforcement Learning

Yoongi Hong[1,2], Sangcheol Lee[1], Haneol Cho[1,3], Nabil Belacel[4], Chansoo Kim[1,2,*]
1 KIST (Korea Institute of Sci. and Tech.), 2 UST (Univ. of Sci. and Tech.), 3 Sejong University, 4 Canada National Research Council,

요 약

본 연구는 산업구조 조정만으로 경제성장과 온실가스(GHG, Greenhouse Gas) 감축을 동시에 달성하기 위한 최적 전략 도출을 위해, 시스템 다이내믹스(SD, System Dynamics) 시뮬레이션을 다목적 강화학습(MORL, Multi-Objective Reinforcement Learning)과 결합한 프레임을 제안한다. 에이전트의 행동은 최종연도(2050)의 산업구조(투입산출계수/산업비중)를 선택하는 것으로 정의하고, 중간연도(2021~2049)는 선형 보간으로 생성하여 현실 제약(기준년도 데이터 고정, 급격 변화 억제)과 고비용 시뮬레이션 문제를 동시에 완화한다. 벡터형 보상은 (i) GDP 누적 성장과 (ii) GHG 누적배출 감소량으로 구성되며, 파레토 관점에서 비지배 정책 집합을 학습한다.

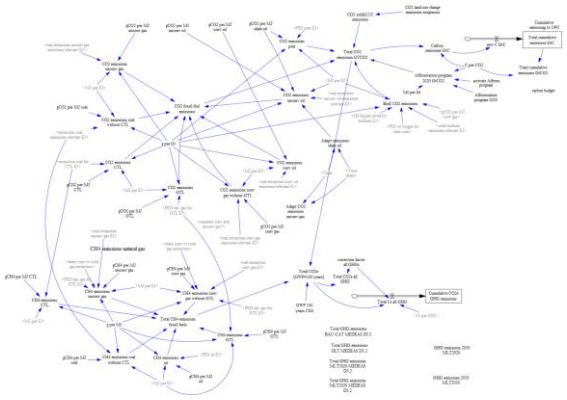
Ⅰ. 서 론

경제성장과 GHG 감축은 전통적으로 상충 관계로 인식되어 왔으나, 산업 간 연계를 고려하면 구조적 재배치만으로도 양 목표를 동시에 개선할 여지가 존재한다. 이때 산업 간 상호의존과 파급효과를 정량화하는 분석틀로 투입산출분석(IO, Input-Output Analysis)과 레온티예프 역행렬이 표준적으로 활용된다 [5,6]. SD-IO 통합 접근의 정책·성장 시나리오 응용은 최근 연구에서도 확인된다 [1].

SD 관점에서 에너지-경제-환경을 통합적으로 모의하는 모델 프레임은 MEDEAS 계열 모형을 준거로 삼았고[2], 산업구조 데이터를 연결하기 위해 환경확장 다지역 투입산출(EE- MRIO) 데이터베이스의 구축·활용 연구를 참고했다 [3,4].

본 연구의 목적은 어떤 산업구조가 두 목표의 균형을 가장 잘 이룰지를 파레토 관점에서 탐색하여 그 결과를 현재 구조와 수치로 비교하며, 누적 배출 및 임계점 도달 시기를 함께 제시하는 데 있다. 다만 현실의 제약(초기년 고정, 급격 변화 불가, 계산비용)을 고려하면 전 기간의 산업구조를 자유롭게 탐색하기 어렵다. 본 저자는 목표연도 구조만 결정하고 중간 경로는 선형 보간으로 단순화하여, 학습효율과 현실성을 동시에 확보하는 실용적 설계를 채택한다.

Ⅱ. 본론



2.1 시뮬레이션 환경 및 행동(액션)

환경은 경제-에너지-환경 모듈이 결합된 MEDEAS 모델을 사용하여 2020-2050 연도별 산업산출, 에너지사용, 온실가스 배출을 계산한다.

에이전트의 행동은 2050 년 산업구조(IO 계수/산업비중) 선택이다. 2021-2049 년 구조는 2020↔2050 선형 보간으로 자동 생성하여, ① 기준년도(실측/보정값) 고정, ② 급격 변화 억제, ③ 일관성 제약(비음수, 정규화 등)을 동시에 만족한다. IO 계수의 의미·해석, 산업 간 상호의존 관계에서의 제약과 전제는 고전·표준 참고문헌을 따른다 [5,6]. 데이터의 환경 확장은 EE-MRIO 구축·활용 문헌을 참고하여 구성했다 [3,4].

2.2 보상(다목적)과 파레토 학습
보상은 벡터형(R_1, R_2)으로 구성한다.

- R_1 : 2020-2050 GDP 누적 성장
- R_2 : 2020-2050 GHG 누적배출 감소량

가중합 스칼라화의 임의성/비볼록 프론트 미포착 문제를 피하기 위해, 파레토 지배·하이퍼볼륨 등 대표적 다목적 기준으로 비지배 정책 집합을 유지·갱신한다. 스칼라 단일 해가 아닌 파레토 최적 정책 집합의 근사를 목표로 탐색한다.

2.3 최적화 기준
탐색된 최적 산업구조 정책들은 파레토 프론트 형태로 제시된다. 그러나 실제 정책 결정 과정에서는 모든 파레토 해를 고려하기보다는, 사용자가 설정한 제약조건(예: 특정 GDP 성장률 이상, 특정 배출량 이하)에 따라 선택 범위를 좁힌다. 즉, 먼저 전체 최적 정책 집합을 도출한 뒤, 이러한 제한조건에 따라 선택된 범위의 파레토 바운더리만을 다시 추출하고, 그 안에서 정책 간 성과를 다음과 같은 지표로 비교한다.

- 구조 차이: IO/산업비중의 L_1/L_2 거리, 변화 상위 부문 Top-k, 부문별 배출집약도 변화
- 거시 성과: 2020-2050 누적 GDP, 2020-2050 GHG 누적 배출 감소량
- 임계점 시기: 선택된 파레토 해 각각에 대해 기후변화 임계온도 기준의 최초 도달 연도

이러한 방식으로 제약조건에 따라 좁혀진 파레토 바운더리를 비교함으로써, 사용자가 원하는 정책 우선순위에 부합하는 후보군 내에서 정책 간 상대적 장단점을 직관적으로 확인할 수 있다.
결과는 선택된 파레토 프론트의 부분집합 그래프, 누적배출-예산 곡선, 구조 차이 요약표로 시각화하여 정책 선택에 참고자료로 제공한다.

Ⅲ. 결론

본 논문에서는 SD 와 MORL 을 결합하여 2050 년 최적 산업구조를 파레토 관점에서 탐색하고 현재 구조와 정량 비교하는 방법을 제안하였다. 시뮬레이션을 기반으로 누적 배출량과 기후 임계점 도달 시기까지 함께 제시함으로써, "어떤 산업구조가 언제까지 어떤 성과를 보장하는가"를 정책 논의에 활용 가능한 형태로 정리했다.
물론 정책 실현 가능성과 경로의 내생적 최적화는 심화할 연구주제이다. 이에 대한 개선으로, 연도별 산업구조 스텝을 직접 최적화하는 경로 기반 MORL 을 수행하고, 연간 변화율 상한·고용/투자 제약을 내재화하여 보다 현실적인 지표를 탐색한다.

ACKNOWLEDGMENT

This research was funded by the grant Nos. 2023-00262155; 2024-00339583; 2024-00460980; and 2025-02304717 (IITP) funded by the Korea government (the Ministry of Science and ICT).

참 고 문 헌

[1] Perez- Montiel, J., Nieto, J., Carpintero, Ó., & Lobejón, L. F., "Is a Green Growth Strategy the Road to Energy Transition? An Energy- Constrained Supermultiplier Approach," SSRN Working Paper, 2024.

[2] Capellán- Pérez, I., de Blas, I., Nieto, J., de Castro, C., Miguel, L. J., Carpintero, Ó., et al., "MEDEAS: a new modeling framework integrating global biophysical and socioeconomic constraints," Energy & Environmental Science, 13, 986-1017, 2020.

[3] Stadler, K., Wood, R., Bulavskaya, T., Södersten, C.- J., Simas, M., Schmidt, S., et al., "EXIOBASE 3: Developing a Time Series of Detailed Environmentally Extended Multi- Regional Input-Output Tables," Journal of Industrial Ecology, 2018.

[4] Timmer, M. P., Dietzenbacher, E., Los, B., Stehrer, R., & de Vries, G. J., "An Illustrated User Guide to the World Input-Output Database: The Case of Global Automotive Production," Review of International Economics, 23(3), 575-605, 2015.

[5] Leontief, W., Input-Output Economics, 2nd ed., Oxford University Press, 1986.

[6] Miller, R. E., & Blair, P. D., Input-Output Analysis: Foundations and Extensions, 2nd ed., Cambridge University Press, 2009.

Trend-Seasonal 분해를 적용한 Transformer 기반 전기차 배터리 SoH 예측

천유림, 윤유정, 변예린, 임완수*
성균관대학교 전자전기컴퓨터공학과
*wansu.lim@skku.edu

State of Health Prediction for EV Batteries Using a Trend-Seasonal Decomposed Transformer

Yurim Cheon, Youjeong Yoon, Yerin Byeon, Wansu Lim
Sungkyunkwan University

요 약

본 논문은 전기차 배터리의 상태지수(State of Health, SoH) 예측 정확도를 향상시키기 위해 시계열 분해 기반 Transformer 모델을 제안하였다. 제안 모델은 Autoformer의 이동평균 기반 시계열 분해 기법을 적용하여 배터리 열화 신호를 추세(Trend) 성분과 주기적(Seasonal) 성분으로 분리한 후, 각 성분을 Transformer 인코더를 통해 독립적으로 학습하고, 이를 재조합하여 최종 SoH를 예측하도록 설계하였다. 모델 학습에는 미국 메릴랜드 대학교 CALCE 센터의 CS2 계열 배터리 실험 데이터를 활용하였으며, 내부저항과 CCCT(Constant Current Charging Time) 측정 데이터를 입력 피처로 사용하였다. 실험 결과, 제안 모델은 배터리 열화 특성을 효과적으로 반영하여 기존 방법 대비 RMSE와 MAE 지표에서 우수한 성능을 나타내었다.

Ⅰ. 서 론

전기차 배터리의 SoH 예측은 배터리의 안정성과 수명 관리를 위해 중요한 기술이다[1]. 기존에는 RNN, LSTM 등 순환 신경망 기반의 모델이 주로 사용되었으며, 최근에는 transformer 계열의 모델이 시계열 데이터의 장기 의존성을 효과적으로 학습할 수 있어 활용이 확대되고 있다. 특히 Autoformer는 시계열 데이터를 추세 성분과 주기적 성분으로 분해한 후 각각을 독립적으로 학습함으로써 예측 정확도를 향상시키는 구조로 주목받고 있다[2-3].

본 연구는 시계열 분해 기반 예측 구조를 전기차 배터리 SoH 문제에 적용하고, 추세 성분과 주기적 성분의 특성을 분리 학습하는 transformer 모델을 제안한다. 시계열 데이터를 입력으로 받아 추세 성분과 주기적 성분을 분리하고 각각을 transformer로 학습한 후 재조합하여 최종 SoH를 예측한다. 제안하는 모델은 배터리 열화 특성에 적합한 학습 구조를 갖추고 있으며, 실험을 통해 예측 성능의 우수성을 확인한다.

Ⅱ. 본론

2.1 데이터셋

본 연구에서는 전기차 배터리 SoH 예측 모델의 성능을 검증하기 위해 미국 메릴랜드 대학교의 CALCE(Computer-Aided Life Cycle Engineering) 센터에서 공개한 배터리 열화 실험 데이터를 활용하였다. 사용된 데이터셋은 CS2 실험군 중 CS2_35, CS2_36, CS2_37, CS2_38의 총 네 개 시리즈로 구성되며, 각각 동일한 초기 상태에서 반복적인 충·방전 조건 하에서 측정된 시계열 데이터를 포함한다. 이 데이터셋은 다양한 열화 조건에서 배터리의 전압, 전류, 온도, 내부저항 등의 특성을 측정한 결과를 제공하며, 실환경에 가까운 배터리 열화 양상을 반영하고 있어 SoH 예측 모델의 학습 및 검증에 적합하다.

특히 본 연구에서는 데이터셋 내 CCCT 측정값을 기반으로 하였으며, 입력 피처로는 배터리의 내부저항과 CCCT 구간에서 수집된 주요 변수를 활용하였다. SoH는 초기 용량 대비 현재 용량의 비로 정의하고, 각 사이클마다 계산된 SoH를 회귀 모델의 학습 타깃으로 설정하였다. 또한 다양한 환경 조건에서 전압, 전류, 온도, 내부저항 등을 종합적으로 반영하여 기초 데이터를 구축함으로써 모델의 일반화 성능을 강화하였다.

2.2 전처리

CALCE 데이터는 실험 환경에서 얻어진 raw 데이터이기 때문에, 시계열 분석 모델에 직접 적용하기에는 시간축 간격이 불균일하고 일부 데이터가 누락되어 있는 문제가 존재한다. 이를 해결하기 위해 본 연구에서는 큐빅 스플라인 보간법(Cubic Spline Interpolation)을 활용하여 시간축 기준으로 데이터를 균등하게 보간하였다. 큐빅 스플라인 보간법은 각 데이터 구간을 3차 다항식으로 근사하여 전체 시계열을 부드럽게 이어주는 방식으로, 급격한 값의 변화를 방지하면서 연속성을 유지할 수 있는 장점을 가진다.

보간된 시계열 데이터는 고정된 입력 시퀀스 길이를 갖도록 윈도우 슬라이싱을 적용하였으며, 학습 데이터와 테스트 데이터는 시계열 순서를 유지한 채 비율로 분할하였다. 데이터 전처리 과정에서는 단순한 노이즈 제거에 그치지 않고, 시계열의 장기적 경향과 단기 변동성을 보존하기 위한 특수 필터링 기법을 적용하였다. 이 과정은 후속 Transformer 모델이 학습 시 불필요한 잡음을 줄이고, 실제 배터리의 열화 패턴을 더욱 정밀하게 학습할 수 있도록 돕는다. 나아가, 기존의 평균화 방식이 갖는 한계를 극복하여 세밀한 SoH 변화 추이를 반영하는 데 더욱 유리하도록 설계하였다.

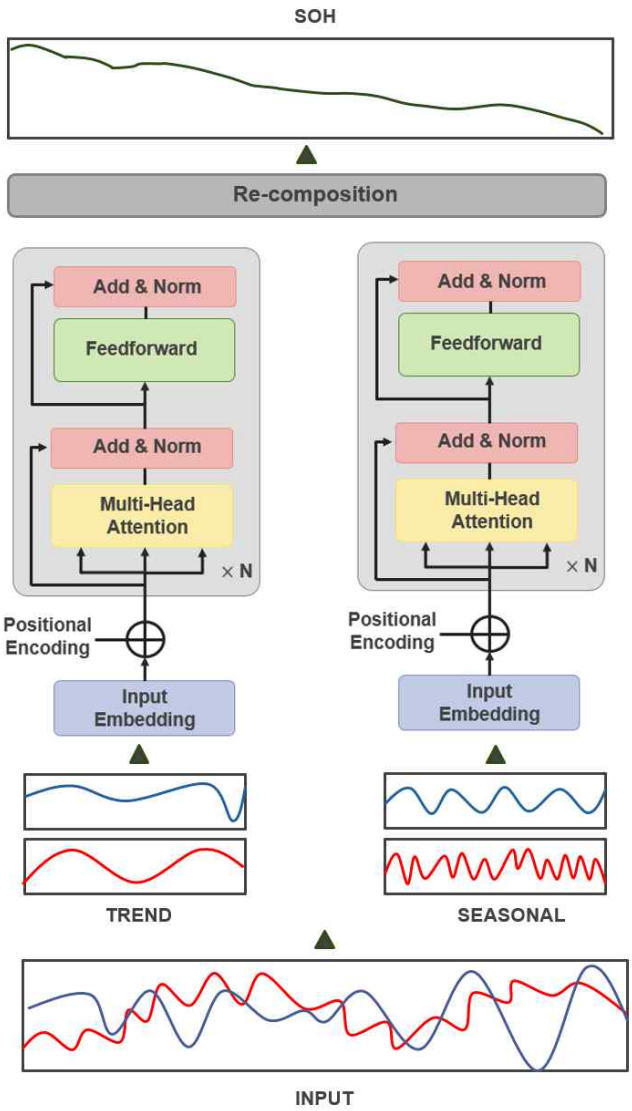


그림 1. 제안한 모델 구조

2.3 모델 구조

제안하는 모델은 Transformer 기반의 시계열 예측 프레임워크이다. 입력 시계열 데이터를 추세 성분과 주기적 성분으로 분해한 후, 각 성분을 Transformer 인코더를 통해 독립적으로 학습하고 결과를 통합하여 최종 SoH를 예측한다.

우선, 입력 시계열 X에 대해 Autoformer에서 제안된 이동평균(Moving Average) 기반 시계열 분해 기법을 적용하였다. 이를 통해 전체 신호는 장기적인 추세 성분과 단기적 주기 성분으로 구분되며, 수식적으로는 다음과 같이 표현된다.

$$X_t = AvgPool(Padding(X)) \quad (1)$$

$$X_s = X - X_t \quad (2)$$

이러한 분해 과정을 통해 모델은 배터리 열화의 추세 성분과 주기적 성분을 분리하여 학습할 수 있다. 분리된 각 성분은 Transformer 인코더 내의 멀티헤드 Self-Attention 모듈을 통해 시점 간의 장기 의존성을 학습하며, 이를 통해 비선형적 열화 패턴과 반복적 충·방전 변동성을 동시에 반영할 수 있다.

특히 추세 성분 학습은 배터리의 장기 열화 경향을 안정적으로 포착하는 데 기여하고, 주기 성분 학습은 충·방전 과정에서 나타나는 단기 변동성을 효과적으로 반영한다. 또한 Positional Encoding은 시계열 특성을 보존하여 시간 축 정보를 유지하고, Self-Attention은 두 성분 간 상관관계를 정밀하게 학습하도록 돕는다. 이러한 과정은 기존 RNN/LSTM 기반 모델이 가지는 장기 의존성 학습의 한계를 극복하고, Autoformer 기반 구조의 장점을 그대로 살려 잡음 억제와 예측 안정성 향상에 기여한다.

모델 학습에는 평균제곱오차(Mean Squared Error, MSE)를 손실 함수로 사용하였으며, Adam optimizer를 통해 파라미터를 최적화하였다. 또한 학습의 안정성을 확보하기 위해 학습률 조정(Learning Rate Scheduler)과 Early Stopping 기법을 병행하여 과적합을 방지하였다. 모델의 성능 평가는 RMSE(Root Mean Squared Error)와 MAE(Mean Absolute Error) 지표를 활용하여 정량적으로 분석하였다.

Ⅲ. 결론

본 논문은 전기차 배터리 SoH 예측의 정확도 향상을 위해 시계열 분해 기반 transformer 모델을 제안하였다. 제안 모델은 Autoformer의 이동평균 기반 시계열 분해 기법을 적용하여 배터리 열화 신호를 장기 경향과 주기적 변동 성분으로 분리하고, 각 성분을 transformer 인코더를 통해 독립적으로 학습한 후 재조합하여 최종 SoH를 예측하였다. 미국 메릴랜드 대학교 CALCE 센터의 CS2 계열 데이터를 활용한 실험 결과, 제안 모델은 RMSE와 MAE 지표에서 기존 방법 대비 더욱 향상된 성능을 보였으며, 이는 배터리 열화 특성을 효과적으로 반영한 학습 구조의 타당성을 충분히 입증한다. 향후 연구에서는 다양한 배터리 화학 조성 및 운용 환경에 대한 확장 실험과, 경량화 및 실시간 추론을 위한 모델 최적화를 지속적으로 진행할 예정이다.

ACKNOWLEDGMENT

▷ 본 연구는 2023년도 산업통상자원부의 재원으로 한국에너지기술평가원(KETEP)의 지원을 받아 수행한 연구 과제입니다. (RS-2023-00266248)

▷ 이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(RS-2024-00349885)

참 고 문 헌

[1] M. Kannan, K. Sundareswaran, P. S. R. Nayak, and S. P. Simon, “A Combined DNN-NBEATS Architecture for State of Charge Estimation of Lithium-Ion Batteries in Electric Vehicles,” IEEE Transactions on Vehicular Technology, vol. 72, no. 6, pp. 7328 - 7338, Jun. 2023.

[2] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention is all you need,” in Advances in Neural Information Processing Systems, vol. 30, 2017.

[3] H. Wu, J. Xu, J. Wang, and M. Long, “Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting,” Advances in Neural Information Processing Systems, vol. 34, pp. 22419 - 22430, 2021.

[4] S. Wang, H. Wu, X. Shi, T. Hu, H. Luo, L. Ma, J. Y. Zhang, and J. Zhou, “TimeMixer: Decomposable Multiscale Mixing for Time Series Forecasting,” in International Conference on Learning Representations (ICLR), 2024.

전자동 생성형 이미지 증강 시스템에 대한 연구

정성윤, 정남훈, 조규태
LIG 넥스원

seongyun.jeong@lignex1.com, namhoon.jung@lignex1.com, kyutae.cho2@lignex1.com

A Study on the Fully Automated Generative Image Augmentation System

Seongyun Jeong, Namhoon Jung, Kyutae Cho
LIG Nex1

요 약

최근 생성형 모델의 발전으로 이미지 생성 기술이 빠르게 혁신되고 있으며, 특히 생성기반의 이미지 데이터 증강 기법에 대한 연구가 활발히 이루어지고 있다. 이러한 기법은 컴퓨터 비전 과제에서 분류 모델의 일반화 성능을 향상시키고 클래스 불균형 문제를 완화하는 데 기여하고 있다. 그러나 정밀하고 자연스러운 데이터 증강을 위해서는 목적에 맞는 정교한 프롬프트와 시각적 가이드와 같은 요소를 수동으로 준비해야 하며, 이는 대규모 증강에 적용하기에 시간과 작업량 측면에서 비효율적이다. 본 논문에서는 이러한 한계를 해결하기 위해 전자동 생성형 이미지 증강 시스템을 제안한다. 제안하는 시스템은 사용자가 수행했던 프롬프트 작성 및 시각적 특징 추출 과정을 언어 모델과 비전 모델들을 활용해 자동화함으로써, 이미지 입력만으로도 사용자의 개입 없이 다양한 시나리오가 반영된 고품질 이미지를 생성할 수 있도록 한다. 또한, 본 시스템을 활용해 생성한 데이터를 이미지 분류 모델 학습에 적용한 결과, 분류 성능이 향상되는 것을 실험을 통해 확인하였다.

Ⅰ. 서 론

최근 Stable Diffusion [1]과 같은 생성형 인공지능(Generative AI)의 등장으로 이미지 생성 분야에서 큰 기술적 혁신이 이루어지고 있다. 이러한 고성능 생성 모델들은 다양한 컴퓨터 비전 과제에 활발히 활용되고 있으며, 특히 데이터 증강(Data Augmentation) 분야에서 주목받고 있다. 생성형 AI 는 클래스 불균형 문제를 완화하고, 수집이 어려운 희귀 데이터를 보완하는 데 효과적으로 사용된다.

최신 이미지 생성 모델들은 입력 방식에 따라 크게 두 가지로 나뉜다. 텍스트를 기반으로 이미지를 생성하는 Text-to-Image 방식과, 이미지를 기반으로 새로운 이미지를 생성하는 Image-to-Image 방식이다. 텍스트 프롬프트는 주로 배경, 객체, 시나리오 설정 등 전반적인 컨셉을 제시하는 데 사용되며, 이미지 가이드는 객체의 형태, 위치, 구도 등 시각적 특성을 제어하는 데 활용된다. 이 두 방식은 생성 목적에 따라 혼용되며, 사용자가 의도한 이미지에 보다 가깝게 생성될 수 있도록 돕는다.

그러나 사용자가 원하는 고품질 이미지를 얻기 위해서는 정교하게 작성된 프롬프트와 명확한 시각 정보가 요구된다. 예를 들어, 프롬프트에는 배경과 객체 간의 상호작용을 포함한 구체적인 설명이 요구되며, 배경의 완전한 변환을 위해 필요한 객체와 배경의 분리에는 배경 제거 작업이 선행되어야 한다. 이러한 작업이 수작업으로 이루어지다 보니, 실제 데이터 증강 파이프라인 활용은 시간과 비용 측면에서 제약이 따른다.

본 논문에서는 위의 수작업 과정을 자동화하기 위한 전자동 이미지 증강 시스템을 제안한다. 제안하는 시스템은 사용자가 증강하고자 하는 이미지를 입력으로 받아, 클래스 이름을 자동으로 추출하고 주요 객체를 분리한 후, 객체로부터 Edge, Depth 와 같은 시각 단서들을 생성한다. 이어서 언어 모델이 자동으로 이미지 생성을 위한 프롬프트를 작성하고, 최종적으로 다양한 경우의 수에서 새로운 이미지를 생성한다. 사용자는 단순히 이미지를 입력하기만 하면, 시스템이 객체의 시각적 특징을 추출하고 자동으로 시나리오를 설계해 반영된 이미지를 생성하는 구조이다.

제안하는 시스템은 대규모 데이터셋 증강에 적용 가능하며, 복잡한 수작업 없이도 다양한 조건의 데이터를 효과적으로 생성할 수 있다. 본 논문에서는 시스템의 전체 구조를 상세히 설명하고, 생성된 데이터를 이미지 분류 모델 학습에 활용했을 때의 성능 향상 효과를 실험을 통해 입증한다.

Ⅱ. 본론

1. 구조

제안하는 전자동 생성형 이미지 증강 시스템은 먼저 입력 이미지를 Vision-Language Model(VLM)을 통해 분석하여 주요 키워드를 추출하고 (본 연구에서는 WD14 Tagger 사용), 이를 클래스 이름으로 활용한다. 추출된 클래스 이름은 GroundingDINO [2] 모델의 입력으로 사용되어 객체의 위치가 Annotated Image 형태로 출력되며, 이는 다시 SAM [3] 모델에 전달되어 객체의 Mask 이미지를 생성한다. 이 과정을 통해 입력 이미지로부터 객체를 분리한 후, Rotation, Flipping과 동의 이미지 변환 기법을 적용하여 다양한 시점(View Point)의 이미지를 생성한다. 다음으로 생성한 이미지를 Canny Edge 알고리즘과 Depth Anything [4] 모델을 통해 중요 시각적 특징들을 추출한다. 이는 이미지

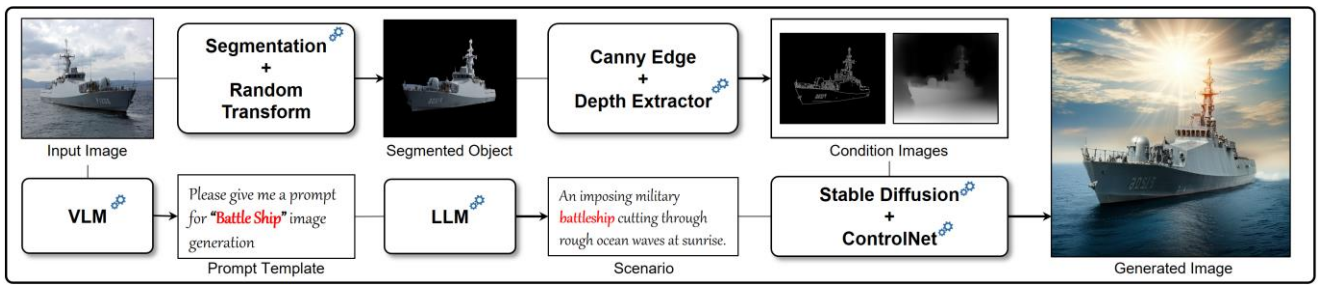


그림 1 전자동 생성형 이미지 증강 시스템의 구조 및 동작 예시

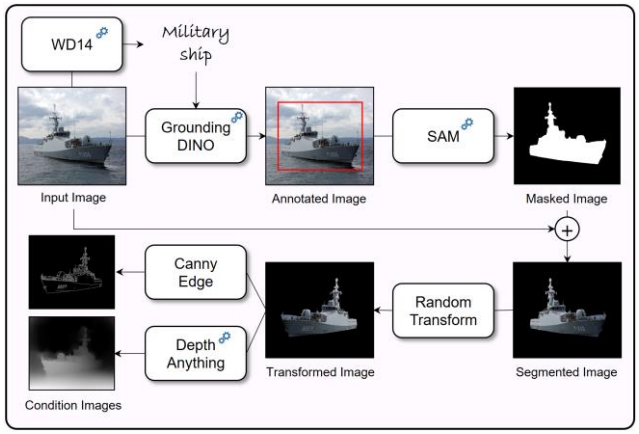


그림 2 시각적 특징 추출 과정

생성시 시각적 가이드로 활용된다.

추출한 키워드는 사전에 정의된 프롬프트 템플릿의 빈칸을 채우는 데 사용되며, 완성된 템플릿은 Large-Language Model(LLM)의 입력으로 활용되어 이미지 생성을 위한 최종 프롬프트를 생성한다. 시스템의 전체 동작 과정은 그림 1에 제시된 파이프라인을 통해 확인할 수 있다. 그림 2를 통해 시각적 특징을 추출하는 자세한 과정을 확인할 수 있다. 그림 3은 해당 작업을 여러 번 수행하였을 때 생성된 데이터의 예시를 보여준다.

그림 3 제안하는 시스템을 활용해 생성한 이미지 예시

2. 실험

제안한 시스템의 이미지 분류 과제에서의 유효성을 검증하기 위해 ResNet-18 모델을 Vehicle-512 데이터셋을 증강해가며 Random initialized된 모델을 50epoch, 0.001 Learning Rate, 512 Batch size로 설정하고, Adam Optimizer를 활용해 학습 시키고 이미지 분류 성능을 확인했다.

학습에 사용한 Vehicle-512 데이터셋은 16개의 클래스 별 200장의 이미지로 구성되어 있고 모든 실험은 NVIDA A100 GPU를 통해 수행되었다.

표 1에서 분류 모델 학습시 데이터의 증강 비율이 증가할수록 정확도가 증가하는 모습을 확인할 수 있다. 이 결과를 통해 제안하는 시스템을 통해 증강한 데이터가 분류 모델의 일반화 성능을 개선할 수 있음을 보였다.

표 1 ResNet-18 이미지 분류 성능 비교표

Synthesis Ratio	Trainset Size	Accuracy	Δ
0%	1x	69.06%	-
25%	1.25x	71.88%	2.82
50%	1.5x	75.00%	5.94
100%	2.0x	78.44%	9.38

Ⅲ. 결론

본 논문에서는 생성형 인공지능을 활용하여 데이터 증강을 자동화하는 이미지 증강 시스템을 제안하였다. 이 시스템은 입력 이미지로부터 클래스 이름과 주요 객체를 자동으로 추출하고, 객체의 시각적 특징을 분석한 뒤, 언어 모델을 통해 프롬프트를 자동으로 생성함으로써 최종적으로 사용자가 원하는 이미지와 유사한 증강 데이터를 생성한다. 이를 통해 기존 방식에서 수작업으로 이루어졌던 프롬프트 작성과 객체 분리 작업의 부담을 획기적으로 줄였다. 제안한 시스템을 이미지 분류 모델 학습에 적용한 결과, 증강된 데이터셋이 모델의 분류 성능을 유의미하게 향상시킴을 확인하였다. 이는 본 시스템이 단순한 자동화 도구를 넘어, 실제 학습 성능 향상에 기여할 수 있는 실용적인 데이터 증강 시스템임을 보여준다.

참 고 문 헌

[1] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer. High-resolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 1068410695, 2022.

[2] S. Liu, Z. Zeng, T. Ren, F. Li, H. Zhang, J. Yang, Q. Jiang, C. Li, J. Yang, H. Su, et al. Grounding dino: Marrying dino with grounded pre-training for open-set object detection. In European Conference on Computer Vision, pages 38– 55. Springer, 2024.

[3] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, et al. Segment anything. In Proceedings of the IEEE/CVF international conference on computer vision, pages 4015– 4026, 2023.

[4] L. Yang, B. Kang, Z. Huang, Z. Zhao, X. Xu, J. Feng, and H. Zhao. Depth anything v2. Advances in Neural Information Processing Systems, 2024.

Synthetic-driven Text-to-Image Generation with Preference-based Reinforcement Learning

Bumkyu Choi, Namhoon Jung and Kyutae Cho

LIGNEX1 CO., Ltd
{bumkyu.choi, namhoon.jung, kyutae.cho2}@ligenex1.com

Abstract

Most existing subject-driven text-to-image generation methods predominantly utilize real-world images as their reference data. However, in fields where obtaining real data is difficult or impossible such as military applications or speculative futuristic concepts, this approach is quite limited. In this work, we address this challenge by using virtually generated reference data, such as computationally synthesized designs or conceptual prototypes, as the starting point for subject-driven image generation. By leveraging these artificial references with preference-based reinforcement learning, we can flexibly produce new, scenario-specific visuals that align with the intended applications and requirements. Our experiments suggest that this approach could enables the creation of highly relevant datasets for tasks such as object detection, and recognition highlighting the potential of virtually seeded data pipelines for practical deployment.

Ⅰ. INTRODUCTION

Recent advances in subject-driven text-to-image generation, such as DreamBooth [1] and Subject-Driven Text-to-Image (SuTI) [2, 3], have enabled users to generate highly customized images from textual prompts and reference images. However, these methods are often grounded in the availability of real-world reference data, which presents a major limitation in domains where data acquisition is constrained or impractical. For instance, sensitive areas such as the military or speculative fields involving futuristic or conceptual designs often lack access to sufficient real images.

To address this gap, we propose a new approach that utilizes virtually generated reference data, produced via generative models and conceptual design, as a starting point for subject-driven image generation. Leveraging synthetic references and fine-tuning with preference-based reinforcement learning, our approach supports the creation of diverse and scenario-specific datasets even in the absence of real-world imagery.

This virtually seeded generation pipeline extends the applicability of personalized generation techniques to previously inaccessible domains. While we do not directly evaluate downstream tasks, the semantic quality and consistency of the generated images suggest their potential utility in applications such as object detection or recognition. Extensive experimentation highlight the practicality and adaptability of the proposed method for dataset creation in challenging or novel environments.

Ⅱ. RELATED WORKS

Ⅱ-1. Subject-Driven Text-to-Image Generation

Subject-driven text-to-image generation aims to produce images that accurately reflect a specific subject based on textual prompts and reference images. Previous works like DreamBooth [1] and SuTI [2, 3] enable personalized image generation by fine-tuning diffusion models with a small number of real images to preserve subject identity in diverse contexts. However, a common limitation among these approaches is their reliance on real-world reference data, which restricts their applicability to domains where acquiring such data is challenging or unfeasible.

Building upon these methods, we focus on extending their applicability to scenarios lacking real-world reference images by leveraging synthetic virtual data and preference-based reinforcement learning. We employ a two-stage pipeline, where the first stage involves generating reference images, and the second stage utilizes these generated images as training references.

Ⅲ. METHODS

Ⅲ-1. Reference Images Generation

To generate virtual reference images, we utilized recent advanced generative models including Stable Diffusion (XL, 3.5), and the GPT-Image Generation model. The objective was not merely to produce high-quality individual images, but rather to ensure

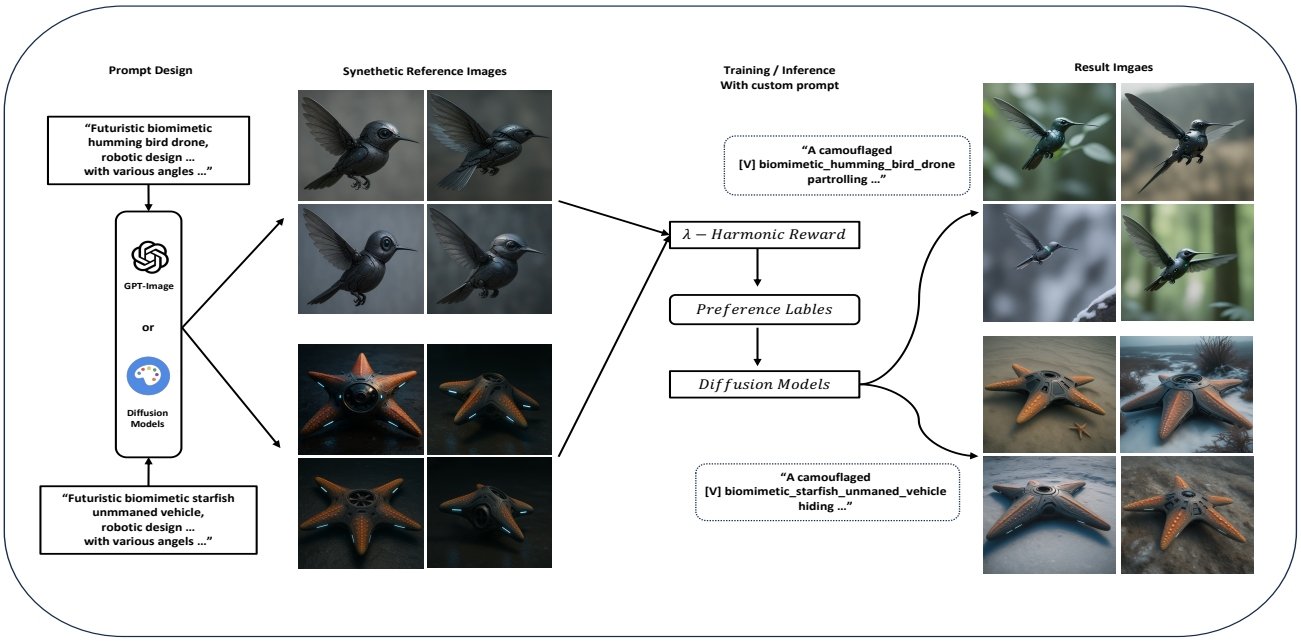


Figure 1. The proposed framework generates subject-driven images using synthetic virtual references and preference-based reinforcement learning. Illustrated examples include biomimetic hummingbird drones and starfish-shaped unmanned vehicle.

consistency across multiple viewpoints, creating cohesive multi-angle visualizations of each conceptual design. As practical examples, we applied this generation strategy to futuristic biomimetic designs such as 'biomimetic hummingbird drones' and 'biomimetic starfish unmanned vehicle'.

Ⅲ-2. Fine-tuning with RPO

In the training phase, we applied Reward Preference Optimization (RPO) [3] to fine-tune the model using a limited set of virtual reference images. RPO leverages a reward function that promotes alignment between image and text representations, allowing effective fine-tuning without relying on human-labeled feedback. This approach remains robust even when synthetic references are used in place of real-world image datasets.

As illustrated in **Figure 1.** we can use these artificially generated images, which do not exist in the real world, to create various scenarios. Moreover, our two-stage pipeline eliminates the need for meticulously crafted prompts; instead, specialized image generation can be guided using simple, unique tokens.

Quantitative evaluation yielded the following similarity scores (synthetic reference / generated result): Biomimetic hummingbird — **DINO**: 0.888 / 0.699, **CLIP-I**: 0.961 / 0.833, Biomimetic starfish — **DINO**: 0.796 / 0.682, **CLIP-I**: 0.922 / 0.867

IV. CONCLUSION

This paper introduces a practical and adaptable framework for text-to-image generation targeting domains where real-world reference images are difficult or impossible to obtain. By synthesizing semantically coherent reference images from text, our proposed pipeline enables subject-driven generation in fields where the target concepts are hypothetical, conceptual, or beyond current real-world observation. Future work may explore the integration of these high-quality synthetic datasets into downstream pipelines, such as object detection, and recognition to further assess their practical utility.

REFRENCES

[1] N. Ruiz, Y. Li, V. Jampani, Y. Pritch, M. Rubinstein, K. Aberman, "Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation," presented at the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 22500-22510 (2023).

[2] W. Chen, H. Hu, Y. Li, N. Ruiz, X. Jia, M.-W. Chang, W. W. Cohen, "Subject-driven text-to-image generation via apprenticeship learning," Advances in Neural Information Processing Systems, vol. 36, pp. 30286-30305 (2023).

[3] Y. Miao, W. Loh, S. Kothawade, P. Poupart, A. Rashwan, Y. Li, "Subject-driven text-to-image generation via preference-based reinforcement learning," Advances in Neural Information Processing Systems, vol. 37, pp. 123563-123591 (2024).

도메인 일반화를 위한 프롬프트 앙상블 기반 Zero-shot 화재 탐지 기법

정성균, 최흥준, 이민호, 이원일, 이수빈, 유연승, 김태형*
Seonggyun.Jeong@lotte.net, heungjunchoi@lotte.net, minho_lee@lotte.net, wonil.lee@lotte.net, leesubin@lotte.net, yys4000@lotte.net, thkim07@lotte.net*

롯데이노베이트㈜

Prompt Ensemble-based Zero-shot Fire Detection for Domain Generalization

Jeong Seonggyun, Choi Heungjun, Minho Lee, Wonil Lee, Lee Subin, Yu Yeon Seung, Kim Taehyung*
Seonggyun.Jeong@lotte.net, heungjunchoi@lotte.net, minho_lee@lotte.net, wonil.lee@lotte.net, leesubin@lotte.net, yys4000@lotte.net, thkim07@lotte.net*

LOTTE INNOVATE Co.,Ltd.

요 약

최근 증가하는 화재 사고를 방지하기 위해 다양한 딥러닝 모델이 제안됐으나 기존 지도 학습 방식의 모델의 경우 일반화 부족으로 환경이 바뀌었을 때 성능이 감소할 수 있는 문제가 존재한다. 이를 해결하기 위해 사전 학습된 텍스트-이미지 모델인 CLIP 과 프롬프트 앙상블(prompt ensemble) 방식을 결합하여 zero-shot 화재 탐지 구조를 본 논문에서 제안한다. 해당 방식의 경우 기존 방식에 비해 비교적 작은 연산량과 더불어 효과적인 탐지 성능을 보이고, 실시간 화재 감지 시스템에 적극적으로 활용될 수 있다.

Ⅰ. 서 론

최근 국내 발생 화재 건수는 전년 대비 약 10% 증가한 3,444 건 [1]이며 화재 사건은 사회적·경제적 손실을 초래하기에 해결이 시급한 문제이다.

해당 문제를 해결하기위해서 규칙 기반의 실시간 화재 경보 시스템이 개발되었지만, 이 시스템의 경우 규칙의 일반화가 다소 어려워 특정 환경에서는 정확도가 감소하는 모습을 보인다. 따라서 다양한 환경에서도 신뢰성 있는 결과를 출력하는 시스템이 필요하며 이에 따라 딥러닝 기반의 영상 화재 경보 시스템이 최근 활발히 개발 이 진행되고 있다. 주로 실제 화재 이미지를 수집 후 정답을 만들어 지도학습 방식으로 이를 구현한다. 그러나 이러한 방식은 수집된 데이터가 특정 지역 혹은 국가에 국한되기에 해당 특성과 다른 도메인의 데이터가 입력될 경우에 여전히 오 예측이 발생할 가능성이 존재한다.

따라서, 도메인의 일반화를 위해서 모든 데이터를 수집하는 것이 아니라, 기존의 사전 학습된 모델을 이용해 재학습 없이 일관성 있는 예측을 하는 zero-shot 화재 탐지 모델 [2], [3] 이 필요로 해 최근 논문에서 제안됐다. 해당 구조는 CLIP [4]이라는 구조를 이용해서 zero-shot 화재 탐지를 구현했는데, CLIP 은 텍스트-이미지 기반의 모델로서 대용량 사전 데이터로부터 이미지와 텍스트 간의 유사도를 출력하도록 사전 학습됐기에, 다양한 도메인의 화재 이미지를 탐지할 수 있다. 구체적으로 화재 이미지가 들어왔을 때 화재 관련 텍스트와 비화재 텍스트의 유사도를 비교해서 판별하는 것이며, 해당 방식을 통해 다양한 데이터 셋에서 효과적인 탐지 성능을 보였다.

하지만 여러 개의 화재 및 비화재 프롬프트 각각에 대한 유사도를 계산하므로 계산량이 증가하며 통합된 프롬프트를 사용하면 성능이 더 향상될 여지가 존재한다.

따라서 본 연구에서는 통합된 프롬프트를 사용해 zero-shot 기반의 화재 탐지 구조를 제안한다. 구체적으로 프롬프트 앙상블 방식을 통해 프롬프트를 통합했으며, 여러 프롬프트의 특성 벡터를 종합해 평균 화재 벡터와 비화재 벡터를 구해 이에 대한 유사도로 화재를 탐지한다. 본 방식은 여러 번의 유사도 계산을 하지 않으므로 비교적 적은 연산량을 가지면서도 향상된 탐지 성능을 보인다.

실험에 사용된 모델은 별도의 재학습 없이, CLIP ViT-B/16 [5]의 이미지 인코더와 텍스트 인코더의 사전 학습된 가중치를 그대로 활용했으며, 텍스트 프롬프트는 **표 1** 과 같이 기존 연구 [2], [3]에서 제안된 문장을 사용했다.

표 1 화재, 비화재 프롬프트

화재 프롬프트
There is no person. There is a moving person. There is a person doesn't move.
비화재 프롬프트
There is smoke rising. There is an arsonist. There is something that shines brightly. There is a campfire. There is a flame and fire soaring.

Ⅱ. 본론

본 논문에서 제안하는 화재 탐지 모델의 전체 구조는 **그림 1** 과 같으며 우선, 추론 대상 이미지를 이미지 인코더에 입력하여 이미지 임베딩 벡터 v_1^I를 얻는다.

이어서, 사전에 정의된 화재 관련 텍스트 프롬프트와 비화재 관련 텍스트 프롬프트 각각 텍스트 인코더에 입력하여, 화재 프롬프트에 대한 임베딩 벡터 집합 $[v_1^n, v_2^n, ..., v_n^n]$과 비화재 프롬프트에 대한 임베딩벡터 집합 $[v_1^f, v_2^f, ..., v_n^f]$를 생성한다. 기존 방식 [2], [3]에선 이미지 임베딩 벡터와 화재, 비화재 프롬프트 임베딩 집합의 원소에 대해 모두 유사도를 계산해서 가장 높은 유사도를 보이는 3 개의 텍스트 프롬프트 모두 화재 프롬프트에 속해야 화재로 판별한다.

하지만 본 연구에서 각 텍스트 임베딩 벡터 집합의 전체적인 의미를 통합적으로 반영하기 위해, **그림 1** 과 같이 유사도 모두 계산하는 것이 아니라 각각의 벡터 집합에 대해 평균 연산을 수행하여 화재 평균 임베딩 벡터 v^{-n}, 비화재 평균 임베딩 벡터 v^{-f}를 계산한다. 그다음, 이미지 임베딩 벡터 v_1^I와 평균 텍스트 임베딩 벡터들 간의 유사도를 계산하기 위해 매트릭스 곱을 수행한다. 즉, $v_1^I \cdot v^{-n}$ 는 화재 점수, $v_1^I \cdot v^{-f}$는 비화재 점수가 된다. 각 점수는 0 과 1 사이의 값으로 표현되며 두 점수를 비교해, 화재 점수가 더 높은 경우에 화재로 분류한다.

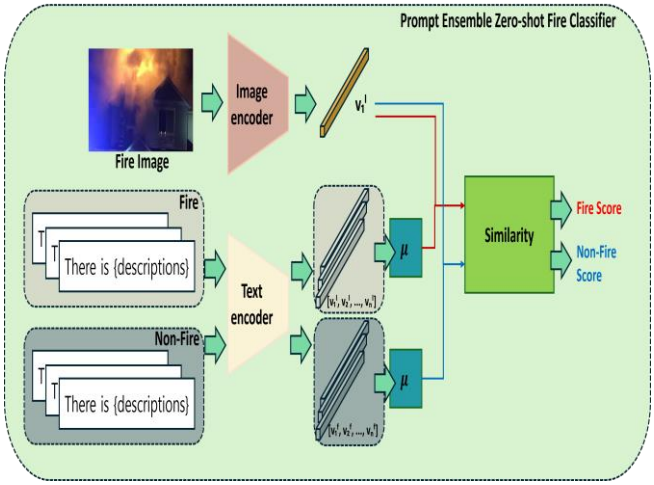


그림 1 제안하는 화재 탐지 파이프라인

Ⅲ. 실험

제안하는 구조의 화재 탐지 성능을 검증하기 위해, FireNet [6] 데이터셋을 활용하였다. FireNet 은 다양한 해상도의 화재 이미지로 구성되어 있으며, 훈련용 412 장, 검증용 90 장의 이미지를 포함하고 있다. 모든 이미지는 이미지 인코더에 입력하기 위해 224×224 크기로 리사이즈 한 후 사용하였다.

표 2 은 FireNet 데이터셋에서의 실험 결과를 정리한 것이며 Zero-shot 분류 성능 비교를 위해, 기존 텍스트 프롬프트 기반 방식 추론이 가능한 YOLOE [7]을 비교군으로 추가했으며, 구체적으로 YOLOE-11-L 가중치를 사용하였다.

또한, CLIP 모델의 경우 패치 크기에 따라 성능 차이가 발생하므로, 본 모델 (CLIP ViT-Base/16) 외에도 CLIP ViT-Base/32 를 추가로 비교하였다. 사용된 그래픽 카드의 경우 NVIDIA H100 80GB 이며 추론 배치는 1 로 고정해서 FPS 를 측정하였다. Accuracy, F1 score 는 1 에 가까울수록 모델의 예측 결과가 정답에 유사함을 의미하며, FPS 는 초당 추론한 이미지 개수로 값이 클수록 추론속도가 빠름을 의미한다. 제안하는 프롬프트 앙상블 기반 구조가 정확도 0.9976, F1 score 0.998, FPS 96.24 로 가장 우수한 성능과 빠른 추론속도를 기록하였다.

표 2 FireNet zero-shot 성능 비교

모델	Accuracy	F1 Score	FPS (Img/s)
YOLOE-11-L	0.313	0.476	22.73
CLIP (ViT-B/32)	0.895	0.944	85.47
CLIP (ViT-B/16)	0.946	0.972	64.1
+Prompt Ensemble	**0.997**	**0.998**	**95.24**

Ⅳ. 결론

본 연구를 통해서 기존의 텍스트-이미지 모델을 활용한 화재 탐지 모델의 탐지 성능을 향상시킬 뿐만 아니라 추론 속도 또한 효율적으로 높일 수 있었다.

제안한 방법은 텍스트 프롬프트를 종합하는 방식이기에 다양한 구조에 손쉽게 결합될 수 있으며, 기존의 화재 방지 CCTV 시스템에 적극적으로 도입해 화재 사건 방지에 기여할 수 있다.

현재 모델을 단순 화재와 방화 사건을 구별하지는 못하므로 추후엔 텍스트 프롬프트와 구조의 고도화를 통해서 화재만 판단하는 것이 아닌 원인 분석까지 가능한 파이프라인 개발로 확장할 예정이다.

참 고 문 헌

[1] 소방청, 화재통계, 2025, (https://www.nfds.go.kr/stat/general.do).

[2] Radford Alec, "Learning transferable visual models from natural language supervision." International conference on machine learning.

[3] Jeon, Hobeom, "Zero-shot Fire And Arson Detection Using Textual Descriptions." *2022 13th International Conference on Information and Communication Technology Convergence (ICTC)*. IEEE, 2022.

[4] Jeon Hobeom, "PASS-CCTV: Proactive Anomaly surveillance system for CCTV footage analysis in adverse environmental conditions.", Expert Systems with Applications 254, 2024.

[5] Dosovitskiy Alexey, "An image is worth 16x16 words: Transformers for image recognition at scale." *arXiv preprint arXiv:2010.11929*, 2020.

[6] Jadon Arpit, "FireNet: a specialized lightweight fire & smoke detection model for real-time IoT applications.", arXiv preprint arXiv:1905.11922, 2019.

[7] Wang Ao, "Yoloe: Real-time seeing anything.", *arXiv preprint arXiv:2503.07465*, 2025.

한국어 기반 Text-to-Video 프롬프트 엔지니어링 시스템 연구 : CogVideoX를 중점으로

임영기[1], 최지희[1], 엄태선[1], 이우성[1], 구영현*

㈜피씨엔[1], *세종대학교

yglim@pcninc.co.kr, jhchoi@pcninc.co.kr, tseom@pcninc.co.kr, wooslee@pcninc.co.kr, *yhgu@sejong.ac.kr

A Study on the Text-to-Video Prompt Engineering System Based on Korean : Focusing on CogVideoX

Younggi Lim, Jihui Choi, Taeseon Eom, Wooseong Lee, Yeonghyeon Gu*

PCN R&S Headquarters., *Sejong Univ.

요 약

대다수의 Text-to-Video(T2V) AI 모델은 영어 사용을 전제로 연구되고 있기 때문에, T2V 모델 비전문가 혹은 국내 사용자들은 텍스트 프롬프트 작성에 많은 어려움을 겪는다. 이를 해결하기 위해 본 논문은 한국어 기반 Text-to-Video 프롬프트 엔지니어링 시스템(Video Prompt Engineering based on Korean, VPE-Ko)을 제안한다. VPE-Ko는 한국어 대화를 통해 사용자로부터 영상 생성에 필요한 핵심 정보를 수집하여 T2V 모델에 최적화된 프롬프트를 생성하는 시스템이다. 본 연구는 CogVideoX를 중점으로 진행되었다. VPE-Ko는 T2V 프롬프트의 5가지 핵심 요소를 도출해 이를 활용하여 VPE-Ko 시스템 프롬프트 및 멀티-턴 챗봇 시스템을 구축한다. 본 연구는 Qwen2 언어 모델을 활용해 수행됐으며, 한국어 멀티-턴 데이터셋을 생성 및 사용한다. 실험 결과 VPE-Ko는 G-Eval 4.9, 프롬프트 만족도 점수 4.5, 영상 만족도 점수 4.6을 기록하며 우수한 성능을 보였다. 더욱 자세한 내용 GitHub에서 확인할 수 있다 : https://github.com/wndaasa/VPE-Ko/

Ⅰ. 서 론

최근 Text-to-Video(T2V) 생성형 AI 모델의 발전이 활발히 이루어지며 텍스트 프롬프트 엔지니어링의 중요성이 강조되고 있다. 하지만 대부분의 T2V 모델은 기본적으로 영어 사용을 전제하고 있으며, 프롬프트 가이드 또한 영어 작성을 권장하고 있다. 이는 한국어가 익숙한 국내 사용자들, 특히 T2V 모델 비전문가들에게 큰 어려움으로 작동한다.

프롬프트 엔지니어링을 위한 연구는 다방면으로 진행되어 왔다. Sarkar 외[1]는 소규모 LLM으로도 효율적인 프롬프트 작성이 가능함을 보여준다. Ein-Dor 외[2]는 LLM을 활용해 단계별 논의 수행을 통해 효과적인 프롬프트를 제작하는 프레임워크를 제안한다. Zhang 외[3]는 효과적인 대화형 AI 개발을 위해서는 전반적인 기술 발전이 필요함을 강조한다. 손민준 외[4]는 데이터셋에 따라 적합한 프롬프트 엔지니어링 기법이 달라질 수 있음을 분석했다. 이처럼 프롬프트 엔지니어링을 위한 AI 시스템 연구는 많으나, 한국어 혹은 T2V 모델에 특화된 프롬프트 엔지니어링 시스템 연구는 많이 부족한 상황이다.

이에 본 논문은 한국어 기반 T2V 프롬프트 엔지니어링 시스템(T2V Prompt Engineering system based on Korean, VPE-Ko)을 제안한다. 본 논문의 제안점은 두 가지이다. 첫 번째는 T2V 프롬프트의 핵심 요소를 도출해 설계한 시스템 프롬프트이다. 두 번째는 프롬프트 엔지니어링을 위한 한국어 멀티-턴 챗봇 구현이다. VPE-Ko의 유효성은 사용자 작성 프롬프트와 VPE-Ko 기반 프롬프트를 한 쌍으로 묶어 G-Eval과 사용자 설문조사의 2가지 방법으로 측정했다.

Ⅱ. 방법론

본 연구는 한국어 기반 T2V 프롬프트 엔지니어링 시스템인 VPE-Ko를 제안한다. VPE-Ko는 T2V 모델을 처음 사용하는 한국어 사용자가 전문적인 프롬프트 지식 없이도 고품질 프롬프트를 작성하도록 지원한다.

본 논문은 오픈소스 T2V 모델 중 프롬프트 이해도가 높으며, 1360×768 고해상도 영상 생성이 가능한 CogVideoX[5]를 중심으로 연구됐다. 우선 공식 문서를 분석해 VPE-Ko 시스템을 위한 고품질 영상 생성 핵심 요소를 도출했다. VPE-Ko 핵심 요소는 표 1에서 확인할 수 있다.

표 1 VPE-Ko 핵심 요소

핵심 요소	예시
주요 객체	건물, 사람, 동물 등
카메라 구도	전경을 담는 구도, 특정 객체에 집중한 구도 등
카메라 종류	드론, 1인칭 카메라 등
영상의 속도	느린 속도, 빠른 속도 등
조명	자연광, 네온 조명 등

VPE-Ko 학습에 사용된 데이터셋

VPE-Ko 핵심 요소에 중점을 둔 한국어 대화 데이터셋을 자체 생성하여 연구를 진행했다. 데이터셋은 두 개의 챗봇에 각각 Assistant와 User 역할을 부여한 뒤 자연스럽게 대화시켜 구축했다. Assistant는 VPE-Ko 시스템 역할로 핵심 요소를 수집하고, User는 VPE-Ko 사용자로서 핵심 요소 정보를 제공한다. 이렇게 생성된 5,000개의 대화 내용과 이를 기반으로 한 24,769건의 한국어 대화쌍 데이터셋은 Qwen2[6]의 학습에 사용됐다. 데이터셋 구축을 위한 챗봇은 ChatGPT-API를 사용했다.

VPE-Ko 시스템 프롬프트 설계

VPE-Ko 시스템 프롬프트는 멀티-턴 챗봇의 지시문으로 사용되며, 설계 방향성은 아래와 같다.

1. VPE-Ko 핵심 요소의 체계적인 수집을 위한 순차적 질문 전략 설계
2. VPE-Ko 핵심 요소 외에도 키워드 가중치, 부정 프롬프트 등 일반적인 프롬프트 가이드 포함
3. T2V 프롬프트 스타일 분석을 위한 Few-Shot 예시 프롬프트 제시

* 교신저자

4. 영어 프롬프트, 한국어 번역 프롬프트, 부정 프롬프트, 개선 사항 제안의 네 가지 표준 출력 형식 제안

VPE-Ko 멀티-턴 챗봇 시스템

본 논문이 제안하는 VPE-Ko 멀티-턴 챗봇 시스템은 사용자와의 연속적인 한국어 대화를 기반으로 고품질 영어 프롬프트를 생성한다. 멀티-턴 챗봇 구현을 위해 이전 대화 내용 전체를 다음 입력값으로 사용했다.

VPE-Ko는 사용자가 입력한 주제에 알맞은 핵심 요소를 한국어 대화를 통해 수집한다. VPE-Ko가 핵심 요소를 충분히 수집했다고 판단하면 영어 프롬프트를 포함한 네 가지 출력을 생성한다. 이때 사용자는 일상적인 구어체 대화를 통해 프롬프트 피드백을 전달하고, VPE-Ko는 이를 반영하여 더욱 개선된 프롬프트를 생성한다. VPE-Ko 멀티-턴 챗봇 시스템 구조도는 그림 1에서 확인할 수 있다.

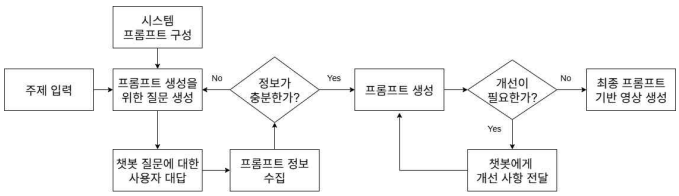


그림 1 VPE-Ko 멀티-턴 챗봇 시스템 구조도

Ⅲ. 실험

VPE-Ko 시스템의 유효성을 검증하기 위해 본 연구에서는 A-B 테스트를 진행했다. 실험 참가자는 T2V 모델 사용 경험이 없는 한국인으로 구성되었다. 우선 실험 참가자가 직관적으로 작성한 프롬프트 A와 VPE-Ko 시스템을 활용해 생성한 프롬프트 B를 준비한다. 작성된 프롬프트 A와 B는 CogVideoX 모델을 통해 각각 영상 A', B'를 생성한다.

본격적인 실험에 앞서 프롬프트 A, B와 영상 A', B'의 특징은 다음과 같다. 아무런 사전 지식 없이 작성한 프롬프트 A는 매우 짧고 정보가 단편적이다. 영상 A' 또한 CogVideoX가 이해하지 못하는 한국어 문장을 기반으로 하기 때문에 프롬프트 A와 무관한 영상이 생성됐다. 반면 프롬프트 B는 VPE-Ko의 5가지 핵심 요소가 전부 포함되어 사용자의 의도가 상세하게 반영된 영어 문장이다. 덕분에 영상 B' 또한 사용자의 생각과 의도를 그대로 반영하고 있다. 프롬프트 A, B와 생성된 영상 A', B'의 대표 예시는 표 3에서 확인할 수 있다.

실험 참가자는 프롬프트 A, B와 영상 A', B'에 대해 자신의 의도가 정확하게 반영됐는지 5점 만족도 조사를 진행한다. 또한 프롬프트 A, B는 생성형 AI 평가에 특화된 G-Eval[7] 지표를 사용하여 객관적인 유효성을 획득한다. VPE-Ko 유효성 실험은 10명의 참가자로 진행되었다.

실험 결과, G-Eval 점수는 A 3.6, B 4.9로 VPE-Ko 기반 프롬프트가 더욱 우수한 품질임을 보였다. 또한 사용자 만족도 점수에서도 A 3.6, B 4.5로 VPE-Ko가 사용자의 의도를 보다 정확하게 반영했음을 확인했다. 결정적으로 CogVideoX는 한국어 처리가 불가능해 한국어 프롬프트로 생성된 영상 A'의 만족도 점수는 2.1로 아주 낮은 반면, VPE-Ko의 영어 프롬프트를 입력하여 생성한 영상 B'의 만족도는 4.6점으로 매우 높았다. 이러한 결과는 한국어를 사용하는 T2V 비전문가에게 VPE-Ko가 매우 유용했음을 의미한다. 실험 결과는 표 2에서 확인할 수 있다.

표 2 VPE-Ko 실험 결과

종류	G-Eval	사용자 만족도	
		프롬프트	영상
A	3.6	3.6	2.1
B	**4.9**	**4.5**	**4.6**

Ⅳ. 결론

본 논문은 한국어 기반 T2V 프롬프트 엔지니어링 시스템 VPE-Ko를 제안한다. 해당 시스템은 LLM 기반 시스템 프롬프트와 멀티-턴 챗봇 구조를 활용하여 효과적인 프롬프트 엔지니어링 환경을 제공한다. 이를 위해 CogVideoX 프롬프트 핵심 요소 도출하고 2개의 LLM으로 생성한 24,769건의 한국어 대화쌍 데이터셋으로 Qwen2을 학습시켰다. VPE-Ko 실험은 사용자 만족도와 G-Eval을 활용한 A-B 테스트로 진행됐다. VPE-Ko 기반 프롬프트가 G-Eval 4.9점, 사용자 만족도 4.5점을 기록했으며, 특히 생성된 영상 만족도 4.6점을 달성하며 매우 유효한 연구임을 증명했다.

따라서 한국어 대화 기반 고품질 영어 프롬프트를 생성하는 VPE-Ko 시스템은 한국어에 익숙한 T2V 비전문가에게 큰 도움이 될 수 있으며, 더 나아가 다양한 영상 생성 AI 모델에도 적용될 수 있을 것으로 기대된다.

acknowledgment

본 논문은 2025년도 정부(과학기술정보통신부)의 재원으로 정보통신산업진흥원의 지원을 받아 수행되었으며(고성능 컴퓨팅 지원(GPU, NPU)), 연구 수행 과정에서 귀중한 조언과 도움을 주신 이혜원 前 차장님께 깊이 감사드립니다.

참 고 문 헌

[1] Sarkar, R., et al., (2025). Conversational user-ai intervention: A study on prompt rewriting for improved llm response generation. arXiv preprint arXiv:2503.16789.

[2] Ein-Dor, et al., (2024). Conversational prompt engineering. arXiv preprint arXiv:2408.04560.

[3] Zhang, C., Dai, et al., (2025). A survey on multi-turn interaction capabilities of large language models. arXiv preprint arXiv:2501.09959.

[4] 손민준, & 이성진. (2025). 대규모 언어 모델에서 프롬프트 엔지니어링 기법에 관한 연구. 한국통신학회논문지, 50(1), 9-21.

[5] Yang, Z., et al., (2024). Cogvideox: Text-to-video diffusion models with an expert transformer. arXiv preprint arXiv:2408.06072.

[6] Team, Q. (2024). Qwen2 technical report. arXiv preprint arXiv:2407.10671.

[7] Liu, Y., et al., (2023). G-eval: NLG evaluation using gpt-4 with better human alignment. arXiv preprint arXiv:2303.16634.

표 3 VPE-Ko 대표 예시

A-B	텍스트 프롬프트 (Input)	영상 (Output)
A (A')	뜨겁게 타오르는 화산	
B (B')	A scorching volcano erupts in a breathtaking display of fiery majesty, its molten lava (flowing down:2.0) the mountain's sides like a river of liquid fire. From above, the camera captures the scene in sweeping aerial shots, emphasizing the scale and intensity of the eruption. The natural light bathes the landscape in an ethereal glow, highlighting the (crimson flames:1.8) that dance and flicker against the night sky.	

다중모달 신체·행동 지표를 이용한 머신러닝 기반 경도인지장애 단계 예측에 관한 연구

김여진, 박인규*, 이은주*

강동성심병원, *한림대학교

yjhelena@kdh.or.kr, *qkrdlsrb3946@naver.com, *ejlee@hallym.ac.kr

A Study on the Machine Learning-Based Estimation of the Mild Cognitive Impairment Stage

Yeo Jin Kim, Ingyu Park*, Unjoo Lee*

Kangdong Sacred Heart Hospital., *Hallym University

요 약

세 가지 비영상 기반 신체·행동 지표, 즉 보행, 체성분, 수면 특징 데이터를 활용하여 인지기능 저하 정도를 예측하고, 나아가 뇌 구조적 변화를 반영할 수 있는 디지털 바이오마커로 활용될 수 있는지를 규명하고자 하였다. 보행, 체성분, 수면 특징 데이터들의 조합으로 구성된 단일모달 및 다중모달 데이터셋을 활용하여 서포트 벡터 머신, 랜덤 포레스트, 다층 퍼셉트론, 합성곱 신경망 등 다양한 모델에서 분류 성능을 평가하였다. 결과적으로 신체 및 행동 데이터만으로 학습된 머신러닝 모델은 조기 단계와 후기 단계 MCI를 구분하는 데 있어 최대 94%의 AUC의 성능을 보여 주었으며, MRI 기반 모델과 비교해도 손색이 없는 높은 분류 정확도를 보였다.

Ⅰ. 서 론

인지기능저하 전 단계인 경도인지장애(MCI)는 일상생활 기능은 유지되지만 인지적 손상이 뚜렷하게 나타나는 상태로, 치매로 진행하기 전 조기 발견·개입이 가능한 중요한 시기이다. 기존 표준 신경심리검사는 정확도가 높지만 1.5 내지 2시간이 소요되어 실시간·연속적 모니터링에는 한계가 있다. 최근에는 웨어러블과 모바일 기기를 활용한 디지털 헬스기술이 발달하면서, 보행, 체성분, 수면과 같은 일상에서 쉽게 측정 가능한 신체·행동 지표를 디지털 바이오마커로 활용하려는 연구가 주목받고 있다. 이러한 지표는 비용이 높고 접근성이 낮은 영상 기반 마커에 비해 실용적이며, 뇌 구조 변화와 인지저하를 반영할 수 있어 인지기능 예측과 예방적 관리에 유용한 대안으로 기대된다. 본 연구에서는 인지기능과의 연관성이 있는 것으로 보고된 바 있는 [1-3] 세 가지 비영상 기반 신체·행동 지표, 즉 보행, 체성분, 수면 특징 데이터를 활용하여 인지기능 저하 정도를 예측하고, 나아가 뇌 구조적 변화를 반영할 수 있는 디지털 바이오마커로 활용될 수 있는지를 규명하고자 하였다.

Ⅱ. 본론

본 논문에서는 MCI로 진단된 환자 80명을 모집하여 Mini-Mental State Examination(MMSE) 점수에 따라 초기 단계 그룹과 후기 단계 그룹으로 분류하였다. 참여자들은 임상 평가(Consortium to Establish a Registry for Alzheimer's Disease Assessment Packet 한국판), GAITRite®를 이용한 보행 분석, 이중에너지 X선 흡수계측법(DEXA)을 통한 체성분 평가, 다중수면검사(PSG)에 기반한 수면 평가를 받았다. 또한 구조적 뇌 영상을 얻기 위해 뇌 MRI도 시행하였다. 단일모달 및 다중모달 데이터셋을 활용하여 서포트 벡터 머신, 랜덤 포레스트, 다층 퍼셉트론, 합성곱 신경망 등 다양한 모델에서 분류 성능을 평가하였다.

Ⅲ. 결론

신체 및 행동 데이터만으로 학습된 머신러닝 모델은 조기 단계와 후기 단계 MCI를 구분하는 데 있어 최대 94%의 AUC의 성능을 보여 주었으며, MRI 기반 모델과 비교해도 손색이 없는 높은 분류 정확도를 보였다. 신체·행동 지표와 MRI 특징을 결합했을 때 예측 성능은 소폭 향상되는 수준이었다. 이 중에서도 보행 속도, 제지방량, 수면 효율이 인지기능을 예측하는 핵심 변수로 나타났다. 결과적으로 다중모달 디지털 바이오마커, 즉 다양한 신체 및 행동 신호를 결합한 지표는 MCI의 중증도를 효과적으로 추정할 수 있으며, 실제 생활 환경에서 인지기능 저하의 조기 발견 및 모니터링을 위한 확장 가능하고 비용 효율적인 접근법이 될 수 있음을 제안한다.

ACKNOWLEDGMENT

이 연구는 한국 정부가 지원하는 한국연구재단(NRF) 연구비 (NRF-2022R1A2C1011286; RS2022NR069669)와 기초과학연구사업 (2020R1F1A1048281; RS202300240976)의 지원을 받았다.. 또한 보건복지부가 주관하는 국민건강증진 R&D 사업(HS21C0000)의 지원을 받았다. 또한 본 과제(결과물)는 교육부 및 한국연구재단의 재원으로 지원을 받아 수행된 한림대학교 글고콜대학30 사업 (GLOCAL-202504370001)의 연구결과이다.

참 고 문 헌

[1] Tolea, M. I. *et al.* Gait, balance, and physical performance as markers of early Alzheimer's disease and related dementia risk. *J Alzheimers Dis*, 13872877241313144 (2025) (https://doi.org:10.1177/13872877241313144)

[2] Chen, L. *et al.* Association between cognitive function and body

composition in older adults: data from NHANES (1999–2002). *Front Aging Neurosci* **16**, 1372583 (2024). (https://doi.org:10.3389/fnagi.2024.1372583)

[3] Zimmerman, M. E. *et al.* The effects of insufficient sleep and adequate sleep on cognitive function in healthy adults. *Sleep Health* **10**, 229–236 (2024). (https://doi.org:10.1016/j.sleh.2023.11.01.)

시각 정보를 이용한 3DGS 기반 안전 경로 계획

함형찬, 안희진*
한국과학기술원
hyeongchan.ham@kaist.ac.kr, heejin.ahn@kaist.ac.kr

Safe Path Planning using Visual Information on 3D Gaussian Splatting Maps

Hyeongchan Ham, Heejin Ahn*
Korea Advanced Institute of Science and Technology (KAIST)

요 약

로봇이 임무를 수행하기 위해 위험에 노출되지 않는 안전한 경로를 계획하는 것은 중요한 문제이다. 주어진 환경 표현 지도가 높은 표현력을 가질수록 보다 많은 정보를 이용하여 안전한 경로 계획이 가능하다. 하지만 기존의 방법들은 물리적인 충돌만을 회피하며 물리적인 부피를 가지지 않는 의미론적인 안전은 달성하지 못하는 한계점이 존재한다. 예를 들어 사족보행 로봇은 도로 배수구 틈에 다리가 빠져 위험에 빠질 수 있으며, 이는 물리적 충돌 회피 만으로는 방지할 수 없는 문제이다. 본 연구에서는 3DGS 로 표현된 환경 지도에서 시각 정보를 활용하여 부피를 가지지 않는 의미론적 위험을 회피할 수 있는 경로 계획 알고리즘을 제안한다. 이를 검증하기 위해 물리적 위험과 의미론적 위험이 복합적으로 존재하는 시뮬레이션 환경에서 실험을 진행하였으며, 제안하는 방법은 부피를 가지지 않는 의미론적 위험을 회피할 수 있다는 것을 확인하였다.

Ⅰ. 서 론

오늘날 로봇의 활용도는 높아지고 있으며 물리적 환경과 상호작용하며 다양한 분야에서 활용되고 있다. 특히 navigation 로봇의 경우 무인 택시, 배달 로봇, 탐사 로봇 등 많은 산업 분야에서 수요가 증가하고 있다. 이러한 navigation 로봇이 동작하기 위해서는 주위 환경을 인식하고, 목적을 달성하기 위한 경로를 계획한 뒤, 경로를 따라 이동할 수 있는 제어 입력을 생성하여 동작하게 된다. 이때 로봇은 목적을 달성할 수 있는 최적의 경로를 계획함과 동시에 물리적 충돌, 손상 등의 위험을 회피할 수 있어야 하며, 경로가 부정확하거나 위험에 노출되는 경우 임무의 실패로 이어지게 된다.

경로 계획은 환경 표현 지도 (environment representation map)와 로봇의 상태, 목표에 대한 상태를 입력으로 받는다. 이때 환경 표현 지도는 point cloud, voxel 등의 방법을 통해 표현될 수 있다. 이러한 환경 표현은 쉽게 생성이 가능하다는 장점이 있지만, 낮은 표현력으로 인해 정보의 손실이 있을 수 있다는 단점이 존재한다. 특히 안전을 고려하기 위한 경우에는 낮은 표현력 문제로 인해 경로가 위험에 노출될 수 있다.

최근 환경을 표현하는 방법으로 radiance field 방법이 주목을 받고 있다. 이미지를 입력으로 받아 새로운 view point 를 만들어 낼 수 있는 방법 (novel view synthesis)으로 풍부한 표현력을 가질 수 있다는 장점이 있다. 이러한 방법들 중 기하적인 정보를 explicit 하게 가지는 3DGS[1]는 경로 계획을 위한 환경 표현으로 직접적으로 활용될 수 있다. Splat-Nav[2]는 3DGS 로 표현된 환경에서 Gaussian ellipsoid 와의 충돌을 회피하는 collision measure 를 제안하였고, 기하적인 공간을 가지는 장애물과의 충돌을 회피할 수 있는 경로 계획 알고리즘을 제안하였다.

하지만 기존의 방법은 환경과의 물리적인 충돌만 회피할 수 있다는 한계점이 존재한다. [그림 1] 왼쪽의 경로 시각화에서 부피를 가지는 장애물(보라색 기둥, 초록색 상자)은 회피가 가능한 반면, 부피가 존재하지 않는 바닥면의 장애물(보라색 원)은 회피하지 못하는 것을 볼 수 있다. 현실에서 안전한 경로를 계획하기 위해서는 배수구, 진흙과 같이 접근해서는 안되는 영역을 회피해야 하지만 이는 의미론적인 (semantic) 위험으로 간주되어 기하적인 표현으로는 일반적인 바닥면과 구분하기가 어렵다. 따라서 의미론적인 위험을 회피할 수 있는 안전한 경로 계획 알고리즘이 필요하다.

본 연구에서는 의미론적인 위험을 회피할 수 있는 시각적 정보 기반의 safe path planning 알고리즘을 제안한다. Radiance field 방식의 환경 표현이 novel view synthesis 를 통해 좀 더 풍부한 정보를 표현할 수 있다는 점을 활용하여, 기하적인 정보 뿐만 아니라 시각적 정보를 활용하여 안전 조건을 만족시키는 경로 계획을 진행한다. 이를 통해 부피를 가지는 장애물을 회피하는 것 뿐만 아니라 부피를 가지지 않지만 위험에 노출될 수 있는 장애물 또한 회피가 가능하게 된다. 이를 검증하기 위해 물리적인 부피가 있는 장애물과 물리적인 부피가 없는 위험한 장애물이 복합적으로 존재하는 환경에서 경로 계획의 성능을 비교하였다.

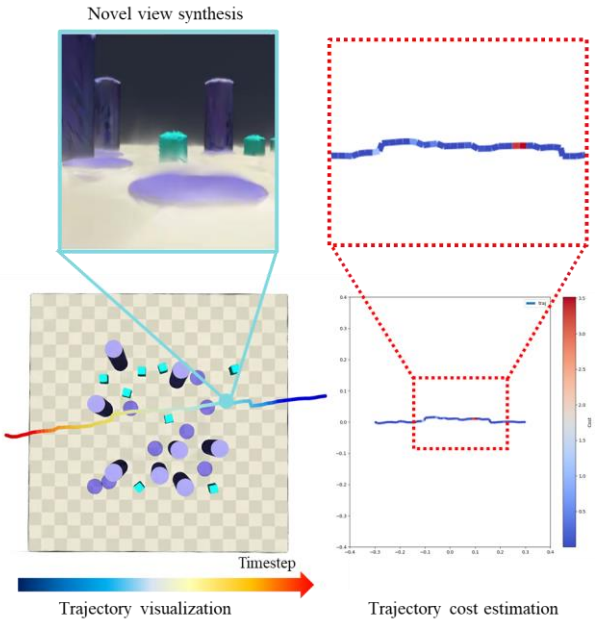


그림 1 기존[2] 경로계획 결과와 위험 비용 시각화

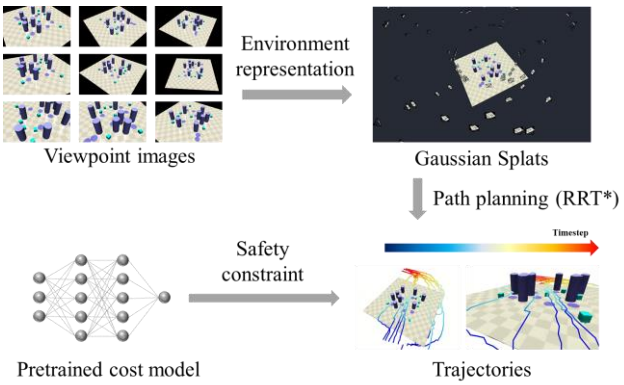


그림 2 제안된 안전 경로계획 알고리즘

Ⅱ. 본론

본 연구에서 제안하는 안전한 경로 계획 과정은 [그림 2]와 같다. 먼저 주위 환경을 관측하며 이미지 데이터를 수집한다. 충분한 이미지가 수집된 후 3DGS 를 이용하여 Gaussian Splatting map 을 생성한다 [1]. 이와 같이 생성된 환경 표현에서는 수집된 이미지에 존재하지 않는 새로운 viewpoint 에서의 이미지를 생성할 수 있게 된다. 이 과정과 별개로 환경과 상호작용하며 학습된 위험 비용 추정 함수인 cost model 을 확보한다. 이 cost model 은 로봇마다 다르게 학습이 될 수 있으며, safe reinforcement learning 을 통해 학습할 수 있다 [3]. 관측 이미지가 cost model 의 입력으로 들어오게 되면 현재 상태에 대한 위험 비용을 추정하여 출력하게 된다. 이를 통해 [그림 1] 우측 시각화와 같이 경로 계획 상에서 현재 경로 위의 상태가 안전한지 여부를 시각적 정보를 통해 확인할 수 있다. 앞서 확보된 환경 표현인 Gaussian Splatting map 과 cost model 을 활용하여 안전한 경로를 생성한다. 이때 경로 생성 알고리즘은 RRT*를 사용한다 [4]. RRT*는 샘플링 기반의 경로 생성 알고리즘이며 상태 공간에서 임의의 상태를 샘플링한 뒤 환경과의 물리적 충돌 여부를 확인 후 이상이 없을 경우 최단 경로가 되는 상태로 추가한다. 이때 물리적 충돌뿐만 아니라 시각적 정보를 통한 안전 또한 확보하기 위해 본 연구에서는 cost model 의 출력에 대한 안전 충족 여부 조건을 추가하였다. 샘플링 된 상태가 물리적 충돌이 발생하지 않더라도 cost model 로 추정한 위험 비용이 기준보다 높으면 해당 상태는 위험한 상태로 간주되어 경로로 포함되지 않고 삭제된다. 이러한 과정을 반복하여 안전 조건을 충족하는 상태들로 이루어진 안전한 경로를 생성할 수 있다.

Ⅲ. 실험 및 결과

제안한 방법을 검증하기 위해 Safety Gymnasium [5] 환경에서 실험을 설계하였다. 경로 계획의 목표는 출발지로부터 안전을 위반하지 않으며 목적지까지 최단거리로 도달하는 경로를 찾아내는 것이다. 환경에는 물리적 부피가 있는 pillar 와, 물리적 부피가 존재하지 않고 지면의 영역으로 표시된 hazard 를 안전 위반을 유발하는 장애물로 설정하였다. 비교하는 방법은 Gaussian Splatting map 에서 기하적인 정보를 이용하여 충돌 회피를 검증하는 기존의 알고리즘을 사용하였으며 [2], 시각적 정보를 이용하여 위험 비용을 추정하기 위한 cost model 은 안전을 고려한 강화학습 모델에서 학습된 모델을 사용하였다 [3].

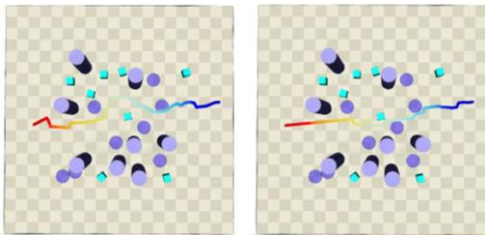


그림 3 안전한 경로 생성 시각화

실험 결과 제안한 방법은 시각적 정보를 통해 안전을 확인하여 부피가 존재하지 않는 hazard 또한 안전하게 회피하였다. [그림 3]에서의 경로 생성 결과와 같이 제안 알고리즘은 기존의 물리적 장애물(pillar)을 회피하는 것뿐만 아니라 기존의 방법으로는 회피할 수 없었던 지면의 hazard 를 회피하는 안전한 경로를 생성하였다. 이를 통해 경로 계획 단계에서 물리적 충돌만을 회피하는 것뿐만 아니라 시각적 정보를 함께 활용하여 의미론적인 안전에 대한 조건을 만족할 수 있게 되었다.

Ⅳ. 결론

본 연구는 로봇의 안전한 경로 계획을 위해 Gaussian Splatting map 으로 표현된 환경에서 시각적 정보를 활용한 방법을 제안하였다. 물리적 충돌만을 회피할 수 있는 기존의 방법과 달리, 본 연구는 상태의 시각적 정보와 cost model 을 활용하여 위험 비용을 계산함으로써 부피를 가지지 않는 의미론적인 위험도 회피가 가능하다. 이를 검증하기 위해 물리적 부피를 가진 장애물과 부피를 가지지 않는 장애물이 혼재된 환경에서 실험하였으며, 위험에 노출되지 않고 안전한 경로를 생성할 수 있는 것을 입증하였다. 향후 연구에서는 경로를 안전하게 추종할 수 있도록 로봇의 동역학을 함께 반영한 경로 계획 연구가 진행될 것이다.

ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원-대학 ICT 연구센터(ITRC)의 지원(IITP-2025-RS-2023-00259991, 50%)과 현대차 정몽구 재단 장학생으로서 지원(50%)을 받아 수행된 연구임

참 고 문 헌

[1] Kerbl, B., Kopanas, G., Leimkühler, T., & Drettakis, G., “3D Gaussian splatting for real-time radiance field rendering,” ACM Trans. Graph., 42(4), 2023, 139-1.

[2] Chen, T., Shorinwa, O., Bruno, J., Swann, A., Yu, J., Zeng, W., ... & Schwager, M, “Splat-nav: Safe real-time robot navigation in gaussian splatting maps,” IEEE Trans. Robot., 2024.

[3] Huang, W., Ji, J., Xia, C., Zhang, B., & Yang, Y, “SafeDreamer: Safe Reinforcement Learning with World Models,” In Proc. Int. Conf. Learn. Represent., 2024.

[4] Karaman, S., & Frazzoli, E, “Sampling-based algorithms for optimal motion planning,” Int. J. Robot. Res., 30(7), 2011, 846-894.

[5] Ji, J., Zhang, B., Zhou, J., Pan, X., Huang, W., Sun, R., ... & Yang, Y, “Safety gymnasium: A unified safe reinforcement learning benchmark,” In proc. Adv. Neural Inf. Process. Syst., 36, 2023, 18964-18993.

DQN 기반 적응형 전송 기법을 통한 NR-V2X 혼합 트래픽의 환경의 성능 개선

강태구, 황경호*
국립한밭대학교 컴퓨터공학과
bigninth@hanbat.ac.kr, gabriel@hanbat.ac.kr*

Performance Improvement in NR-V2X mixed traffic situations through DQN-based adaptive transmissions

Tae-Gu Kang, Gyung-Ho Hwang*

요 약

NR-V2X(5G New Radio V2X) 통신은 자율주행 및 도로 안전 기술의 핵심이지만, 차량 밀도 증가로 인한 채널 혼잡은 통신 신뢰도를 크게 저하시킨다. 주기적인 CAM(Cooperative Awareness Message)과 비주기적인 DENM(Decentralized Environmental Notification Message)이 혼재하는 환경에서는 PRR(Packet Reception Ratio)의 급격한 감소가 발생한다. 본 논문은 이러한 문제를 해결하고자 심층 강화학습(Deep Reinforcement Learning) 기반의 적응형 전송 파라미터 제어 기법을 제안한다. NR-V2X의 물리 계층 및 채널 혼잡 상황을 모델링한 환경에서 DQN(Deep Q-Network) 에이전트를 훈련시켰다. 제안하는 DQN 기반 정책은 고정된 파라미터를 사용하는 기존 방식 대비 향상된 PRR 성능을 나타냈다. 특히, CAM과 DENM 트래픽이 혼재하여 채널 혼잡이 극심한 환경에서 더욱 높은 성능 개선 효율을 보였으며, 이는 제안하는 DQN 에이전트가 통신 환경 변화에 효과적임을 확인할 수 있었다.

Ⅰ. 서 론

3GPP Release-16에서 표준화된 5G 기반 NR-V2X 통신은 기존 LTE-V2X 대비 높은 신뢰성과 낮은 지연을 목표로 하며, Release-17/18을 통해 지속적으로 개선되고 있는 차세대 자율주행 및 지능형 교통 시스템의 핵심 기술이다[1]. NR-V2X Mode 2에서는 차량 간 주기적인 상태 인지 정보 교환을 위한 CAM과 사고 등 비주기적인 위험 상황 발생 시 즉각적인 정보 공유를 위한 DENM이 기본 통신 트래픽으로 활용된다[2].그러나 이러한 메시지들은 도심 및 고속도로와 같은 고밀도 환경에서 혼합 전송될 경우, 자원 충돌과 지연 증가로 인해 안전 메시지의 신뢰도가 저하되는 문제가 발생한다[3].

본 논문에서는 동적으로 변화하는 V2X 통신 환경에 효과적으로 대응하기 위해 DQN(Deep Q-Network) 기반의 적응형 전송 제어 기법을 제안한다. 제안하는 기법은 DQN 에이전트가 실시간 채널 상태 및 트래픽 부하를 관찰하여, 통신 신뢰도를 높일 수 있는 최적의 전송 파라미터(전송 전력, 변조 및 코딩 방식 등)를 동적으로 선택하는 정책을 학습한다. 고정된 방식과 학습된 DQN 에이전트의 성능을 비교하여 PRR(Packet Reception Ratio)의 개선 효과를 정량적으로 분석하고, 이를 통해 제안하는 기법의 성능 향상을 확인한다.

Ⅱ. 시스템 모델

본 연구에서 제안하는 알고리즘의 성능을 평가하기 위해 5G New Radio (NR) 표준을 기반으로, 차량 밀도가 높은 실제 도로 환경의 특성을 반영하여 모델링 하였다[4]. 시스템은 10MHz 대역폭을 사용하며, 이는 50개의 PRB(Physical Resource Block)에 해당한다. 차량수(N_{UE})는 저밀도(50대), 중밀도(100대), 고밀도(200대)의 세 가지로 하여 각 차량은 주기적으로 CAM 메시지를 전송하고 필요에 따라 DENM 메시지를 전송한다. 수신기에서의 신호 대 잡음비는 전송 전력, 경로 손실, 잡음 전력, 그리고 간섭 영향을 종합적으로 고려하여 계산된다. 잡음 전력은 -104dBm으로 설정되며, 송수신기의 잡음 지수는 9dB, 안테나 이득은 송신기와 수신기 각각 3dB로 적용된다. 특히 간섭의 영향은 CBR(Channel Busy Ratio)에 비례하는 페널티 항목으로 모델링되며, 최대 15dB까지의 성능 저하가 발생할 수 있다. CAM 메시지는 300바이트, DENM 메시지는 800바이트의 크기를 가지며, 각각의 전송에 필요한 PRB 수는 사용되는 MCS에 따라 달라진다. 100ms의 관찰 윈도우 동안 발생하는 모든 메시지 전송 요구를 합산하여 전체 자원 수요를 계산하고, 이를 가용 자원과 비교하여 CBR을 도출한다.

CBR은 V2X 통신 환경에서 채널의 혼잡도를 나타내는 핵심 지표이다. 이는 특정 시간 동안 전체 가용 무선 자원 중 차량들의 메시지 전송에 활용된 자원의 비율을 의미하며, 0과 1 사이의 값으로 표현된다. CBR 값이 1에 근접할수록 채널이 포화 상태에 도달했음을 나타낸다. 이러한 CBR은 V2X 표준에 의거하여 다음 식 (1)과 같이 정의된다.

$$CBR = \frac{(N_{UE} \cdot R_{CAM}) + (N_{DENM} \cdot R_{DENM})}{R_{Total}} \quad (1)$$

식 (1)은 모든 차량(N_{UE})이 전송하는 주기적인 CAM(R_{CAM}) 메시지와 전체 차량 중 1~10%의 무작위 차량(N_{DENM})이 추가로 전송하는 비주기적인 DENM(R_{DENM}) 메시지를 합산하여, 이 총량이 전체 통신 채널 용량(R_{Total})에 얼마나 많은 부하를 주는지를 나타낸다. 이를 통해 특정 시간 동안 요구되는 총 무선 자원량을 전체 가용 자원량과 비교하여 채널의 부하, 즉 CBR을 산출한다. 이렇게 계산된 CBR 값은 시스템의 채널 모델에서 수신 신호 대 잡음비(SNR)를 직접적으로 감소시키는 간섭 페널티(Interference Penalty)로 작용하며, 채널 혼잡이 심화될수록 통신 품질이 저하되는 현실적인 네트워크 동작을 시뮬레이션에 반영하였다.

III. DQN 기반 적응형 전송 기법

기존 NR-V2X 통신의 정적 규칙 기반 방식은 실시간 채널 가변성 대응에 한계가 있다. 이에 본 논문에서는 변화하는 환경에 지능적으로 적응하는 심층 Q-네트워크(DQN) 기반 강화 학습 접근법을 제안한다. DQN 시스템에서 상태는 0.1m부터 300m까지의 연속 통신 거리로, 행동은 표 1에 정의된 전송 전력-변조 코딩 방식(MCS) 조합으로 구성된다. 보상 함수는 통신 성능 최적화를 위해 PRR로 설정된다. 수신 SNR은 거리 기반 경로 손실과 CBR에 비례하는 간섭 페널티를 고려하여 계산하며, 실제 페이딩 채널의 영향을 반영하고자 채널의 통계적 특성을 기반으로 한 성능 예측 모델을 통해 유효 SNR을 계산한다.

표 1. DQN 에이전트의 행동 공간(Action Space) 정의

행동	전송 전력(dBm)	MCS	변조 방식	전략적 목표
α_1	23	7	16QAM	균형
α_2	23	10	64QAM	효율성
α_3	29	5	QPSK	신뢰성

제안하는 DQN 에이전트는 현재 통신 거리 상태를 입력으로 받아 전송 전력-MCS 조합을 선택하고, 산출된 PRR을 보상으로 활용하여 최적 정책을 학습한다. 경험된 데이터는 손실 함수 (2)를 최소화하는 방향으로 Q-네트워크를 업데이트하고, 이를 통해 에이전트는 장기적으로 통신 성능을 향상시키는 행동을 습득한다.

$$L(\theta) = \mathrm{E}\left[(y_i - Q(s, a; \theta))^2\right] \quad (2)$$

여기서 Q-값은 다음과 같이 정의 된다.

$$y_i = r + \gamma \max_{a'} Q(s', a'; \theta^-) \quad (3)$$

즉, 목표 Q-값(y_i)은 즉각적인 보상(r)과, 할인율(γ)이 적용된 다음 상태(s')의 최대 기대 가치의 합으로 정의된다. 이때 신경망의 파라미터(θ)는 에이전트가 학습을 통해 축적한 지식으로, 현재 상태에서 각 행동의 Q-값을 추정하는데 사용된다. 또한, 학습 과정에서 에이전트는 탐색(exploration)과 활용(exploitation)의 균형을 위해 ε-greedy 정책을 적용한다. 에이전트는 확률 ε으로 무작위 행동을 선택하여 아직 경험하지 않은 상태-행동 조합을 탐색하고, 확률 $(1-\varepsilon)$으로 현재 학습된 Q-값을 기반으로 최적 행동을 선택하여 기존 지식을 효과적으로 활용한다. 이러한 과정을 반복함으로써 제안하는 DQN 기반 전송 기법은 CBR 기반 간섭 페널티와 거리 변화를 동시에 고려하여, 기존의 고정 규칙 기반 접근법에 비해 높은 적응성과 안정성을 확보할 수 있다.

IV. 시뮬레이션 및 결과

본 연구는 제안하는 DQN 기반 적응형 전송 정책의 성능을 검증하고, 기존 고정 파라미터 기반 전송 방식과 비교 분석하였다. 시뮬레이션은 도심 환경을 가정하고 다양한 차량 밀도 조건에서 수행되었으며, 주요 파라미터는 표 2와 같다. 성능 비교를 위해 통신 환경의 변화와 관계없이 항상 전송 전력 23dBm, MCS 7의 고정된 파라미터 조합으로 동작하는 방식을 고정 방식으로 설정하였다. 제안한 DQN 정책은 모든 통신 거리 구간에서 기존의 고정 방식보다 높은 PRR 성능을 보였다. CAM과 DENM의 혼합 트래픽 상황에서 제안한 DQN 정책은 차량 밀도에 따른 CBR 변동과 거리 변화를 효과적으로 감지하고 대응하여 기존 방식에 비해 안정적이고 향상된 통신 성능을 확인할 수 있었다.

표 2. 시뮬레이션 파라미터

파라미터	값
채널 대역폭	10 MHz
기준 경로 손실	47.86 dB
경로 손실 지수	2.75
잡음 전력	-104 dBm
CAM, DENM 크기	300 Byte, 800Byte
차량수	50, 100, 200 대
안테나 이득(G_tx, G_rx)	3 dB, 3 dB
전송 전력	23, 29 dBm
잡음 지수	9 dB
MCS	5, 7, 10

그림 1은 세 가지 차량밀도(50, 100, 200대)조건에서 적응형 DQN 학습 정책과 고정 정책(Baseline) 간의 통신 거리에 따른 PRR 성능을 비교한 결과이다.

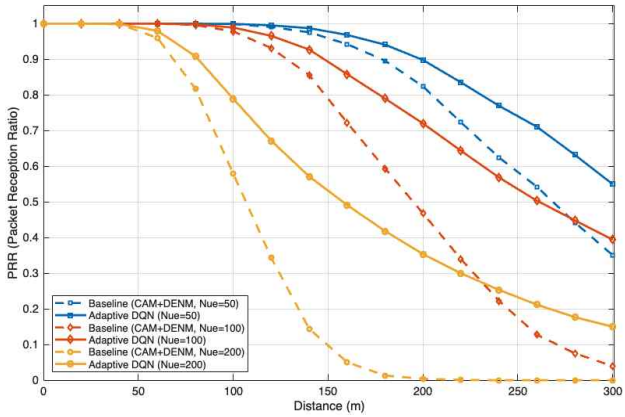


그림 1. 거리에 따른 PRR 성능 비교

제안한 DQN 기반 시스템은 저중밀도 환경(50-100대)에서 baseline 대비 5-15%의 PRR 성능 개선되었다. 고밀도 환경(200대)에서는 CBR 포화로 인해 성능 개선이 제한적이었으나, 최소한 baseline 수준의 성능을 유지하는 것을 확인하였다. 이런 결과는 DQN 에이전트가 채널 변화에 맞춰 전송 파라미터를 적응적으로 선택하여 NR-V2X 통신 신뢰성과 안정성을 향상시켰음을 보여준다.

V. 결론

본 논문은 NR-V2X 통신 환경의 채널 혼잡 문제를 해결하고자 심층 강화학습 기반의 지능형 전송 정책을 제안하고 그 성능을 검증하였다. 제안하는 DQN 에이전트는 통신 환경의 핵심 변수인 거리를 기반으로 PRR을 최대화하는 전송 파라미터를 최적화하여, 기존 방식 대비 통신 신뢰성을 크게 향상시켰다. 향후 연구로는 에이전트에 CBR와 같은 실시간 환경 정보를 추가하고 멀티 에이전트 강화학습으로 확장하여 성능을 높이는 방법을 고려할 수 있다.

참 고 문 헌

[1] 3GPP TR 38.885, "Study on NR Vehicle-to-Everything (V2X)," Rel-16, v16.0.0, Jul. 2018.

[2] ETSI, "Specifications of Decentralized Environmental Notification Basic Service (DENM)," ETSI EN 302 637-3 V1.3.1, Apr. 2019.

[3] S. S. Lee and H. H. Choi, "Design of C-V2X CAM/DENM Separate Resource Pool," J. Korea Inst. Inf. Commun. Eng., vol. 25, no. 6, pp. 821-828, Jun. 2021.

[4] W. Anwar, N. Franchi, and G. Fettweis, "Physical Layer Evaluation of V2X Communications Technologies," in Proc. IEEE 90th VTC-Fall, Sep. 2019.

자율주행 환경에서 상황 인지 기반 적응형 시맨틱 통신 기법

오성현, 임영훈, 김정곤*
한국공학대학교 전자공학부
osh119@tukorea.ac.kr, dladjwlsrh@tukorea.ac.kr, jgkim@tukorea.ac.kr

Adaptive Semantic Communication Based on Context Awareness in Autonomous Driving Environments

Sung Hyun Oh, Young Hoon Lim, Jeong Gon Kim*
Dept. of Electronic Engineering, Tech University of Korea

요 약

자율주행 차량은 다양한 센서와 V2X(Vehicle to Everything) 통신을 통해 주변 환경을 인지하고 주행 의사결정을 수행한다. 그러나 센서가 생성하는 데이터의 양은 매우 방대하며 이를 그대로 전송하는 기존 방식은 제한된 대역폭과 초저지연 요구 조건을 동시에 충족하기 어렵다. 이러한 문제를 해결하기 위한 대안으로 시맨틱 통신이 주목받고 있다. 시맨틱 통신은 단순한 비트 단위 복원이 아닌 의사결정에 필요한 의미 정보를 선별하여 전송하는 기법이다. 하지만 자율주행 환경에서는 동일한 의미라도 상황에 따라 중요도가 달라지므로 상황을 고려하지 않은 단순 시맨틱 통신만으로는 한계가 존재한다. 이에 본 논문에서는 상황 인지 기반 적응형 시맨틱 통신 기법을 제안한다. 제안 기법은 상황 인지 모듈을 통해 현재 주행 맥락을 파악하고 그에 따라 핵심 의미 정보를 선별하며 네트워크 상태에 적응적으로 전송 전략을 조정한다. 이를 통해 긴급 상황에서는 안전과 직결되는 정보를 우선적으로 공유하고 평상시에는 효율적인 데이터 교환이 가능하다.

Ⅰ. 서 론

최근 자율주행 기술의 발전과 함께 V2X(Vehicle-to-Everything) 통신의 중요성이 크게 증가하고 있다[1]. 자율주행 차량은 카메라, LiDAR(Light Detection and Ranging), RADAR (RAdio Detection And Ranging) 등 다양한 센서를 통해 주변 환경을 인지하고 이러한 데이터를 다른 차량 및 인프라와 공유하여 안전한 주행을 수행한다. 그러나 자율주행 환경에서 방대한 데이터가 발생하며 이를 그대로 무선 채널을 통해 전송하는 것은 제한된 대역폭과 초저지연 요구 조건을 동시에 만족시키기 어렵다. 특히 네트워크 혼잡이나 긴급 상황 발생 시 핵심 정보가 제때 공유되지 못한다면 심각한 안전 문제로 이어질 수 있다.

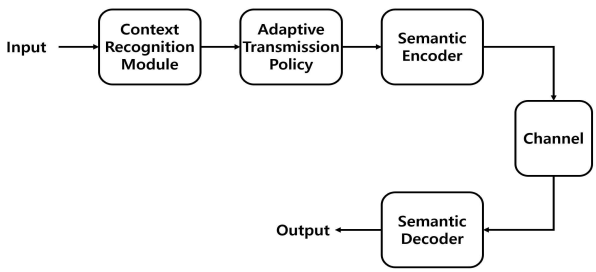
이러한 한계를 해결하기 위해 최근에는 시맨틱 통신이 주목받고 있다. 시맨틱 통신은 단순히 원본 데이터를 비트 단위로 정확히 복원하는 것이 아니라 수신 측 의사결정에 필요한 의미 정보만을 추출하여 전송하는 새로운 통신 패러다임이다[2]. 그러나 기존의 시맨틱 통신 기법은 모든 상황에서 동일한 방식으로 의미를 전송하기 때문에 상황별로 달라지는 중요도를 충분히 반영하지 못한다는 한계가 있다. 예를 들어 고속도로 주행 시에는 차량 속도 및 차간거리 정보가 핵심적이지만 도심 교차로에서는 보행자 위치와 신호등 상태가 더욱 중요하다.

따라서 본 논문에서는 상황 인지 기반 적응형 시맨틱 통신 기법을 제안한다. 제안 기법은 상황 인지 모듈을 통해 현재 주행 맥락을 파악하고 이를 바탕으로 핵심 의미 정보를 선별하여 네트워크 상태에 맞게 적응적으로 전송한다. 이를 통해 긴급 상황에서는 안전과 직결되는 정보를 우선적으로 공유하고 평상시에는 효율적인 데이터 교환이 가능하다.

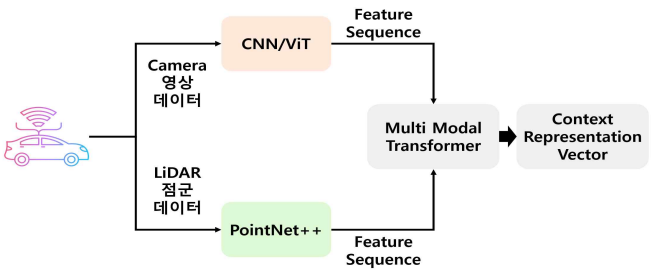
본 논문의 구성은 다음과 같다. 2절에서는 제안하는 시스템 모델을 설명하고 상황 인지 기반 적응형 시맨틱 통신 구조를 구체적으로 제시한다. 끝으로 3절에서는 본 연구의 결론과 향후 연구 방향을 논의한다.

Ⅱ. 본론

본 절에서는 제안하는 상황 인지 기반 적응형 시맨틱 통신 기법을 설명한다. 제안하는 시스템의 전체 구조는 [그림 1]과 같으며, 상황 인지 모듈, 적응형 전송 정책 모듈, 그리고 시맨틱 인코더/디코더로 구성된다.



〔그림 1〕 제안하는 시스템 구성도



〔그림 2〕 상황 인지 모듈 구조

* : 교신저자

2.1 상황 인지 모듈 (Context Awareness Module)

상황 인지 모듈은 주행 환경의 의미론적 맥락을 추출하는 핵심 부분으로, 카메라 영상과 LiDAR 점군 데이터를 입력으로 하는 이중 모달 융합 구조를 가진다. 상황 인지 모듈의 구조는 [그림 2]와 같다. 본 연구에서는 공개 자율주행 데이터셋인 nuScenes-mini를 활용했으며, 입력 데이터로는 CAM_FRONT 이미지와 LIDAR_TOP 점군 데이터를 사용했다.

1) 영상 특징 추출 (Vision Branch)

영상 특징 추출을 위해 사전 학습된 ResNet-50과 ViT(Vision Transformer)[3] 모델을 결합하여 사용했다. 먼저, ResNet-50은 영상의 지역적 특징을 추출하고, ViT는 추출된 특징을 입력받아 Self-Attention을 통해 전역적 패턴이 반영된 시퀀스를 식 1과 같이 도출한다.

$$F_{vision} \in R^{N_v \times d} \tag{1}$$

2) LiDAR 특징 추출 (LiDAR Branch)

3차원 공간 정보를 처리하기 위해 PointNet++[4] 아키텍처를 적용했다. 두 개의 Set Abstraction(SA)으로 구현하여, FPS+반경 기반 그룹화(ball query)로 점군의 지역 - 전역 기하 특징을 계층적으로 추출하였다. 이를 통해 시각 정보만으로는 파악하기 어려운 정밀한 거리, 위치, 형상 정보를 담은 특징 벡터가 식 3과 같이 생성된다.

$$F_{lidar} \in R^{N_l \times d} \tag{2}$$

3) 멀티모달 융합 (Multi-Modal Fusion)

두 Branch에서 추출된 영상 특징과 LiDAR 특징은 멀티모달 Transformer를 통해 최종적으로 융합된다. 멀티모달 Transformer는 Cross-Attention 메커니즘을 통해 두 정보의 상호 보완적 관계를 학습하며, 주행 환경의 복합적인 의미를 담은 상황 표현 벡터 z를 생성한다. 생성된 상황 표현 벡터는 식 3과 같이 나타낼 수 있다.

$$Z = \mathrm{Transformer}(F_{vision}, F_{lidar}) \tag{3}$$

2.2 적응형 전송 정책 모듈 (Adaptive Transmission Policy Module)

본 연구에서 적용한 적응형 전송 정책 모듈은 경량 DNN(Deep Neural Network)을 기반으로, 상황 표현 벡터와 채널 SNR(Signal-to-Noise Ratio)을 입력받아 데이터의 전송 우선순위와 압축 수준을 결정한다. 여기서, DNN 모델 학습에는 전송 우선순위 분류를 위한 Cross-Entropy 손실과 압축 수준 예측을 위한 MSE(Mean Squared Error) 손실을 가중합하여 사용하였다.

본 연구에서 제안한 전송 정책은 주행 상황에 따라 차별화된다. 도심 및 교차로에서는 보행자를 최우선 순위로, 압축 수준은 0.7-0.8을 목표로 한다. 반면 고속도로에서는 차량을 우선순위로 하여 0.5-0.6 사이의 압축 수준을 목표로 학습하였다. 또한, 최종 압축 수준은 실시간 채널 품질에 따라 동적으로 조절되도록 설계하였다.

2.3 시뮬레이션 및 결과

제안 방식의 성능 평가를 위해 nuScenes-mini의 CAM_FRONT와 LIDAR_TOP 데이터를 사용하였다. 이때 학습 파라미터는 epoch 500, batch size 32, learning rate 0.0001로 고정하였다. 채널은 SNR 기준으로 GOOD(25 - 30 dB)과 POOR(15 - 20 dB) 두 단계로 설정했으며, 결과는 [그림 3]에서 확인할 수 있다. [그림 3]에서 볼 수 있듯이, 고속도로 환경과 교차로 환경에서 목표하는 압축률이 도출됨을 확인할 수 있다. 고속도로

(a) (b)

[그림 3] 주행 상황에 따른 압축 결과. (a) 고속도로, (b) 교차로

는 다른 차량과의 상대 거리 정보가 주된 특징이므로, 장면 내 의미적 요소가 비교적 단순하다. 따라서 시맨틱 레벨에서 중요하지 않은 배경 정보나 반복적인 패턴은 과감히 제거할 수 있어, 높은 압축률(0.6)로도 충분히 의미 보존이 가능하다. 반면, 교차로는 다양한 방향에서 접근하는 차량, 보행자, 신호등, 표지판 등 위험 요소와 상황 변화가 많은 복잡한 상황이므로, 지나친 압축은 안전-critical 정보를 손실할 수 있다. 따라서 교차로에서는 낮은 압축률(0.8)로 설정하여 더 많은 정보를 전송함으로써, 상황 인지 성능을 보장하도록 하였다. 결과적으로 제안 방식은 환경의 특성을 고려하여 압축률을 동적으로 조정함으로써, 통신 자원 효율과 안전성 사이의 균형을 달성한다.

Ⅲ. 결론

본 논문에서는 자율주행 환경에서 효율적인 정보 전달을 위해 상황 인지 기반 적응형 시맨틱 통신 기법을 제안하였다. 제안한 방식은 멀티모달 Transformer 기반의 상황 인지 모듈을 통해 카메라와 LiDAR 데이터를 융합하여 고차원 상황 표현을 생성하고 전송 정책 모듈은 DNN을 활용하여 상황 및 채널 조건에 따른 전송 우선순위와 압축 수준을 결정한다. 이를 통해 단순 데이터 전송이 아닌 의미 기반 정보 교환을 가능하게 하여 제한된 자원을 효율적으로 활용하면서도 신뢰성 있는 통신을 지원할 수 있음을 보였다. 향후 연구에서는 실험 기반 검증과 정책 최적화를 통해 제안 기법의 성능을 고도화할 예정이다.

ACKNOWLEDGMENT

이 논문은 2025년도 정부(산업통상자원부)의 재원으로 한국산업기술진흥원의 지원을 받아 수행된 연구임 (RS-2024-00415938, 2024년 산업혁신인재성장지원사업)

참 고 문 헌

[1] Khezri, R., Steen, D., and Tuan, L. A. "Vehicle to Everything (V2X) – A Survey on Standards and Operational Strategies," Proc. EEEIC / I&CPS Europe, pp.1-6, June 2022.

[2] Nguyen, L. X., Raha, A. D., Aung, P. S., Niyato, D., Han, Z., and Hong, C. S. "A Contemporary Survey on Semantic Communications: Theory of Mind, Generative AI, and Deep Joint Source-Channel Coding," arXiv preprint arXiv:2502.16468 [cs.CV], pp. 1-30, February 2025.

[3] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., and Houlsby, N. "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," arXiv preprint arXiv:2010.11929, pp.1-35, October 2020.

[4] Qi, C. R., Yi, L., Su, H., and Guibas, L. J. "PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space," arXiv preprint arXiv:1706.02413, pp.1-18, June 2017.

모달리티 불균형 완화를 위한 개별 헤드 기반 예측 융합

윤영우, 최원준, 이동규
경북대학교
dbsduddn@knu.ac.kr, sangju@knu.ac.kr, dglee@knu.ac..kr

Per-Modality Heads with Logit Fusion for Mitigating Modality Imbalance

Young-Woo Youn, Wonjun Choi, Dong-Gyu Lee
Kyungpook National University.

요 약

본 논문은 멀티모달 인간 행동 인식(MHAR)에서 발생하는 모달리티 간 최적화 불균형 문제를 해결하기 위한 새로운 학습 전략을 제안한다. 제안 방법은 학습 초기에는 대조 학습(contrastive learning)을 통해 모달리티 간 표현 차이를 줄이고, 이후 단계에서는 개별 손실과 융합 손실을 병행하는 지도 학습을 수행한다. 이를 위해 각 모달리티별 독립적인 classifier head를 두고 logit 수준에서 융합을 수행함으로써 균형 있는 학습을 유도하였다. 실험 결과, 제안한 방법은 기존 융합 방식 대비 안정적이고 일반화 성능이 높은 모델을 확보할 수 있음을 보여준다.

Ⅰ. 서 론

멀티모달 인간 행동 인식 (Multimodal Human Action Recognition, MHAR)은 인간의 다양한 활동을 인식하기 위해 RGB, skeleton, IMU 등 이질적인 센서 데이터를 결합하는 대표적인 응용 분야이다. 여러 모달리티를 동시에 활용함으로써 단일 모달리티 기반 인식보다 풍부한 표현력을 얻을 수 있다는 장점이 있다[1]. 그러나 멀티모달 학습에서는 모달리티 간 표현력과 학습 난이도의 차이로 인해 최적화 불균형이 발생할 수 있다. 이 경우 모델은 상대적으로 빠르게 최적화되는 모달리티에 의존하게 되며, 다른 정보는 충분히 활용되지 못한다. 이러한 현상은 멀티모달 감정 분야에서 이미 보고된 바 있으며, 특정 모달리티에 대한 과도한 의존이 전체 성능 저하로 이어질 수 있음이 지적되었다[2,3,4]. 또한, 기존 MHAR 연구의 융합 접근 방식은 크게 두 가지로 구분된다. (1) feature-level concatenation 방식, (2) decision-level logit aggregation 방식이다. 그러나 두 방법 모두 모달리티별 최적화 상태를 고려하지 못해 특정 모달리티에 치중되는 문제가 존재한다[5,6,7,8].

본 연구에서도 해당 문제가 인간 행동 인식 분야에서도 발생함을 확인하였다. 특히 MMAct[1] 데이터셋의 RGB와 IMU 모달리티를 실험적으로 분석한 결과, IMU 모달리티가 상대적으로 빠르게 최적화되어 융합 과정에서 불균형이 발생하는 경향이 관찰되었다. 이는 다른 도메인에서 제기된 문제와 유사한 현상으로, MHAR에서도 모달리티 불균형이 중요한 도전 과제임을 보여준다.

위 문제를 해결하기 위해 모달리티별 개별 손실과 통합 손실을 병행하는 logit aggregation 기반 학습 방법을 제안한다. 학습 초기에 대조 학습 기반 커리큘럼 단계를 도입하여 모달리티 간 표현 간극을 줄이고, 후반에는 지도 학습으로 전환하여 개별 손실과 통합 손실을 함께 최적화하는 전략을 적용하였다. 이를 통해 두 모달리티가 균형적으로 학습되며, 최종적으로 안정적이고 일반화 성능이 높은 모델을 확보할 수 있음을 보인다. 더 나아가, 대표적인 벤치마크 데이터셋을 활용한 비교 실험에서 제안하는 방법이 기존 방법들 대비 우수한 성능을 달성하였으며, 이는 멀티모달 학습에서 모달리티 불균형 문제를 효과적으로 완화할 수 있음을 시사한다.

Ⅱ. 본론

본 논문에서는 멀티모달 인간 행동 인식에서 발생하는 모달리티 간 최적화 불균형 문제를 해결하기 위하여 커리큘럼 기반 학습 프레임워크를 제안한다. 제안 방법은 크게 두 단계로 구성된다. 초기 단계에서는 대조 학습을 통해 모달리티 간 표현 간극을 줄이고, 후기 단계에서는 cross-entropy 기반 지도 학습을 통해 개별 모달리티와 융합 출력을 동시에 최적화한다. 전체 프레임워크는 그림 1과 같이 구성된다.

제안한 프레임워크의 입력은 두 가지 모달리티로 구성되어 있다. RGB 데이터는 ResNet-50을 backbone으로 사용하여 특징을 추출하며, 최종 fully-connected layer를 제거한 후 projection head를 통해 embedding 공간으로 투영한다. IMU 데이터는 시계열 신호의 장기 의존성을 포착하기 위해 temporal convolutional network 인코더를 사용한다. 모든 인코더의 출력은 동일한 차원의 embedding으로 변환되어 모달리티 간 비교가 가능하다.

학습 초기에는 각 모달리티의 표현 차이가 크기 때문에 직접적인 지도 학습을 수행하면 특정 모달리티로 최적화가 집중될 수 있다. 이를 방지하기 위해 대조학습을 먼저 적용한다. 동일한 클래스에 속하는 모달리티 표현은 embedding 공간에서 서로 가깝게, 서로 다른 클래스에 속하는 표현은 멀어지도록 학습한다. 대조 학습 손실은 다음과 같이 정리된다

$$\mathcal{L}_{con} = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{\exp\left(\mathrm{sim}(z_i, z_i^{+})/\tau\right)}{\sum_{j=1}^{N}\exp\left(\mathrm{sim}(z_i, z_j)/\tau\right)}, \quad (1)$$

본 단계는 비지도 학습 형태로 진행되지만 클래스 라벨 정보를 토대로 positive/negative 쌍을 구성함으로써 실제 task와 밀접하게 연관된 특징 정렬을 유도한다. 이를 통해 서로 다른 모달리티가 공통된 표현 공간에서 일정 수준의 정렬을 확보할 수 있다.

대조 학습을 마친 후에는 cross-entropy 기반 지도 학습 단계로 전환된다. 이 단계에서는 각 모달리티별 classifier head와 융합 classifier를 동시에 학습한다. 모달리티별 head는 RGB, IMU 각각의 logit 출력을 산출하

며, 여기에 개별 cross-entropy 손실을 부여한다.

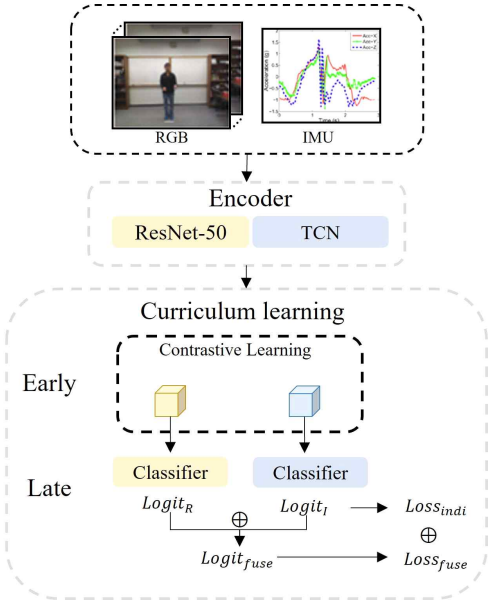


그림 1 제안하는 프레임워크의 전체 구조

동시에 두 모달리티의 logit을 평균하여 융합 결과를 얻고, 이에 대해서도 추가적인 손실을 계산한다. 최종 손실은 다음과 같이 정의된다.

$$\mathcal{L}_{total} = \sum_{m \in \{RGB, IMU\}} \mathcal{L}_{CE}^{(m)} + \mathcal{L}_{CE}^{fuse}, \quad (2)$$

최종 손실은 모달리티별 독립적인 성능을 보존하면서도 융합 결과의 정확도를 동시에 높일 수 있도록 설계되어 있다.

실험은 MMAct[1] 벤치마크 데이터셋을 사용하여 수행되었으며, 평가 척도로는 F1-score를 사용하였고 두 개의 cross-validation인 subject와 session에 대하여 평가하였다.

Method	F1-score	
	Cross-subject	Cross-session
SAKDN'21 [9]	77.23	82.77
MuMu'22 [5]	76.28	87.50
MAWKDN'23 [6]	78.77	85.26
SMTDKD'24 [7]	79.48	86.16
MSMFT'25 [8]	81.07	93.06
Ours	**84.88**	**93.40**

표 1 벤치마크 데이터셋에 대한 실험 결과

표 1의 실험결과를 통해 제안하는 방법이 기존 최신 방법들보다 우수한 성능을 보임을 확인 할 수 있다. 이는 멀티모달 학습에서 단순 융합 방식에 의존하기 보다는, 각 모달리티의 정보를 충분히 반영하여 통합 모델이 균형이 있게 학습되도록 유도하는 것이 중요함을 시사한다.

Ⅲ. 결론

본 논문에서는 멀티모달 인간 행동 인식에서 발생하는 모달리티 간 최적화 불균형 문제를 해결하기 위해 커리큘럼 기반 학습 전략을 제안하였다. 초기 단계에서는 클래스 정보를 활용한 대조 학습을 통해 모달리티 간 표현 간극을 줄였으며, 후반부에는 cross-entropy 기반 지도 학습으로 전환하여 개별 손실과 통합 손실을 동시에 최적화하였다. 이를 통해 서로 다른 모달리티가 균형적으로 학습될 수 있었으며, 최종적으로 안정적이고 일반화 성능이 높은 모델을 확보할 수 있었다.

MMAct 데이터셋을 대상으로 한 실험 결과, 제안 방법은 기존 융합 방식 대비 우수한 성능을 보였으며, 이는 멀티모달 학습에서 융합 전략뿐만 아니라 개별 모달리티의 학습을 병행적으로 고려하는 것이 중요함을 보여준다. 특히 모달리티 간 불균형 문제가 실제 데이터셋에서도 관찰되었음을 확인하였고, 제안하는 학습 전략이 이를 효과적으로 완화할 수 있음을 입증하였다.

향후 연구에서는 depth, skeleton과 같은 추가 모달리티를 포함하여 더 복잡한 환경에서의 확장 가능성을 검증하고, 다양한 커리큘럼 설계 방식을 탐구할 예정이다. 또한 본 방법을 대규모 데이터셋과 실시간 추론 환경에 적용하여 효율성과 범용성을 동시에 향상시키는 방향으로 발전시킬 계획이다.

ACKNOWLEDGMENT

이 성과는 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원-대학ICT연구센터의 지원(IITP-2025-RS-2020-II201808, 50%)과 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. RS-2025-02214941, 50%)

참 고 문 헌

[1] Kong, Quan, et al. "Mmact: A large-scale dataset for cross modal human action understanding." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019.

[2] Yang, Yang, et al. "Facilitating multimodal classification via dynamically learning modality gap." Advances in Neural Information Processing Systems 37 (2024): 62108-62122.

[3] Qian, Chengxuan, et al. "Dyncim: Dynamic curriculum for imbalanced multimodal learning." arXiv preprint arXiv:2503.06456 (2025).

[4] Wei, Yake, and Di Hu. "MMPareto: boosting multimodal learning with innocent unimodal assistance." Proceedings of the 41st International Conference on Machine Learning. 2024.

[5] Islam, Md Mofijul, and Tariq Iqbal. "Mumu: Cooperative multitask learning-based guided multimodal fusion." Proceedings of the AAAI conference on artificial intelligence. Vol. 36. No. 1. 2022.

[6] Quan, Zhenzhen, et al. "MAWKDN: A multimodal fusion wavelet knowledge distillation approach based on cross-view attention for action recognition." IEEE Transactions on Circuits and Systems for Video Technology 33.10 (2023): 5734-5749.

[7] Quan, Zhenzhen, et al. "SMTDKD: A semantic-aware multimodal transformer fusion decoupled knowledge distillation method for action recognition." IEEE Sensors Journal 24.2 (2023): 2289-2304.

[8] Zhou, Xianfa, et al. "MSMFT: Multi-Stream Multimodal Factorised Transformer for Human Activity Recognition." IEEE Sensors Journal (2025).

[9] Liu, Yang, et al. "Semantics-aware adaptive knowledge distillation for sensor-to-vision action recognition." IEEE Transactions on Image Processing 30 (2021): 5573-5588.

모빌리티 환경을 위한 AI 모델 분할 기반 지능형 컨테이너 오케스트레이션

박기철, 안재훈, 김영환
한국전자기술연구원
{kcpark, corehun, yhkim93}@keti.re.kr

Intelligent Container Orchestration Based on AI Model Partitioning for Mobility Environments

Kicheol Park, Jaehoon An, Younghwan Kim
Korea Electronics Technology Institute

요 약

본 논문은 모빌리티 서비스는 자율주행, 스마트 교통, 도심 항공 모빌리티(UAM) 등 다양한 형태로 발전하고 있으며, 실시간 인지·판단을 위한 AI 모델의 성능과 효율성이 핵심 경쟁 요소가 되고 있다. 그러나 대규모 AI 모델을 차량이나 모빌리티 디바이스에서 전부 실행하는 것은 연산 부담과 지연 문제를 야기한다. 본 논문에서는 AI 모델을 Backbone과 Neck+Head로 분할하여, 모빌리티 디바이스와 엣지 서버에 적절히 배치함으로써 처리 지연을 최소화하는 지능형 컨테이너 오케스트레이션 기법을 제안한다. 제안 기법은 모빌리티 환경에서 분할된 워크로드와 모빌리티 디바이스의 특성을 반영하여 최적의 실행 경로를 결정함으로써, 미션 수행에 지장을 주지 않으면서 안정적이고 효율적인 처리를 가능하게 한다.

Ⅰ. 서 론

모빌리티 환경에서는 자율주행 차량, 도심 항공 모빌리티(UAM)와 같이 다양한 형태의 이동체가 센서 데이터를 실시간 분석하여 주변 상황을 인지하고, 경로를 계획하며, 주행을 제어해야 한다. 이러한 과정에서 AI 모델은 높은 연산량과 짧은 응답시간이 동시에 요구되며, 특히 안전과 직결되는 작업에서는 성능 저하 없이 즉시 처리하는 것이 필수적이다[1].

그러나 고성능 AI 모델을 모빌리티 디바이스에서 전부 실행하는 경우, 연산 자원 부족과 전력 제약, 네트워크 통신 지연 등의 문제로 안정적인 성능을 유지하기 어렵다. 이를 해결하기 위해 SDI(Software-Defined Infrastructure)와 엣지 컴퓨팅을 결합하여, 컴퓨팅·네트워킹·스토리지 자원을 유연하게 관리하고, 단말과 인근 엣지 서버 간에 연산을 적절히 분배하여 처리 속도와 효율성을 향상시킬 수 있다[2].

특히 AI 모델을 계층별로 분할하여, Backbone 는 모빌리티 디바이스에서 수행하고, 후속 통합·추론 단계는 엣지 서버에서 실행하는 구조는 자원 부담을 분산하고 지연을 최소화하는 효과적인 방법이다[4].

본 논문에서는 SDI 기반 환경에서 AI 모델 분할 실행을 지원하는 지능형 컨테이너 오케스트레이션 기법을 제안한다. 제안 기법은 분할된 워크로드와 모빌리티 디바이스의 특성을 반영하여 최적의 실행 경로를 결정함으로써, 미션 수행에 지장을 주지 않으면서 안정적이고 효율적인 처리를 가능하게 한다.

Ⅱ. 본론

1. 모빌리티 환경에서의 AI 추론 특성 및 제약 요인

모빌리티 환경에서 제공되는 자율주행, 지능형 교통관리, 실시간 객체 인식 등의 AI 서비스는 초저지연성과 고신뢰성을 동시에 요구한다[2]. 차량·드론·로봇 등 이동형 디바이스는 카메라, 라이다(LiDAR), 레이더 등 복수 센서로부터 초당 수십~수백 프레임의 데이터를 수집하며, 이를 실시간 분석하기 위해 GPU/NPU와 같은 연산 자원이 필요하다[3].

본 논문에서는 대표적인 소형 모빌리티 플랫폼인 터틀봇3에 라즈베리파이5를 온보드 컴퓨팅 장치로 탑재한 환경을 모빌리티 디바이스로 정의하고 실험을 진행하였다. 이와 같은 소형 플랫폼은 연산·전력·발열·폼팩터 제약을 동시에 받으며, 특히 모델 학습까지 수행하기에는 연산 자원이 절대적으로 부족하다. 추론 역시 단말 CPU만으로는 높은 프레임레이트를 유지하기 어렵고, 이동에 따른 네트워크 품질 변동성으로 클라우드 단독 처리 시 일관된 지연 보장이 불가능하다[1].

이러한 제약을 완화하기 위해, AI 모델을 분할하여 온보드 디바이스(라즈베리파이5)와 외장 NPU(Hailo-8)에 연산을 분산(offload)하는 방식을 적용하였다. 컨볼루션 등 연산집약적인 처리를 NPU로 이전함으로써 추론 지연을 줄이고, 단말 CPU의 전력·열 부담을 경감하는 것을 목표로 한다.

2. SDI 기반 지능형 오케스트레이션 아키텍처

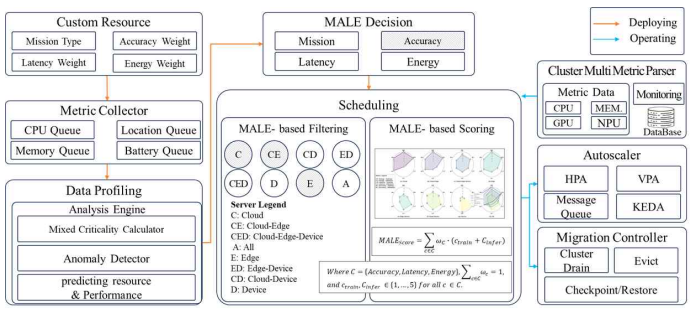


그림 1 모빌리티 환경을 위한 AI 모델 분할 기반 지능형 컨테이너 오케스트레이션 아키텍처

그림 1은 Mission, Accuracy, Latency, Energy(MALE) 지표를 기반으로 한 지능형 오케스트레이션 아키텍처를 보여준다. MALE은 미션 목표 달성을 위한 정확도, 지연 시간, 에너지 효율을 종합적으로 고려하여 최적의 자원 할당 및 컨테이너 배포를 수행한다. 이를 위해 시스템은 Kubernetes를 기반으로 컨테이너 오케스트레이션 환경을 구성하며, 다음과 같은 기능을 구현한다.

Data Profiling: 모델의 복잡도, 실행 환경의 이기종 자원 특성, 워크로드의 혼합 중요도(Mixed-Criticality)를 분석한다.
Metric Collector: CPU, 메모리, 배터리, 위치 등의 상태 데이터를 실시간으로 수집한다.
MALE Decision: 수집된 지표를 바탕으로 정확도, 지연 시간, 에너지 효율 간의 균형을 고려하여 스케줄링 기준을 결정한다.
Scheduler: 필터링과 점수화를 통해 컨테이너를 실행할 최적의 디바이스(Cloud, Edge, Mobility Device 등)를 선택한다.
Autoscaler & Migration Controller): 워크로드 변화에 따라 컨테이너 수를 자동 조정하고, 필요 시 체크포인트를 기반으로 컨테이너를 다른 노드로 이동시킨다.

3 모델 분할 기반 엣지-모빌리티 협력 구조

본 연구에서는 YOLOv5 계열의 객체 인식 모델을 대상으로, Backbone 단계와 Neck+Head 단계를 분리하는 계층 분할 방식을 채택하였다. Backbone 단계는 입력 영상으로부터 저차원에서 고차원에 이르는 특징(feature)을 추출하는 CNN 계층으로 구성되며, 연산량은 크지만 데이터 크기 축소 효과를 제공한다. 반면, Neck+Head 단계는 다중 스케일 feature fusion과 객체 검출을 수행하며, 높은 연산량과 메모리 접근이 요구된다.

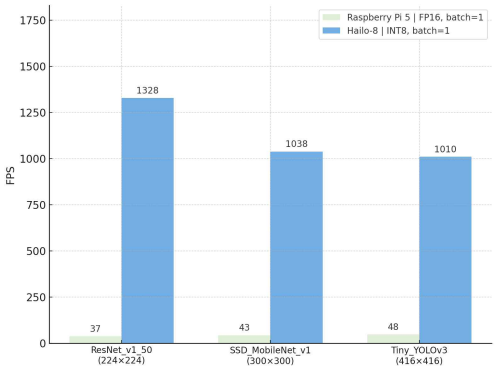


그림 2 라즈베리파이 5와 Hailo-8 가속기의 AI 모델별 FPS 성능 비교

그림 2는 라즈베리파이 5와 Hailo-8 가속기에서 다양한 AI 모델의 FPS 성능을 비교한 결과를 나타낸다. 실험 결과, 라즈베리파이 5 단독 환경에서는 ResNet, SSD-MobileNet, Tiny YOLOv3 등 상대적으로 경량화된 모델만 실시간 처리 성능을 확보할 수 있었으며, 이는 하드웨어 스펙상 복잡도가 높은 모델 운용에 제약이 있음을 시사한다.
이러한 한계를 극복하기 위해, 본 연구에서는 YOLOv5를 Backbone과 Neck+Head로 분할하고, 각 모듈을 컨테이너 단위로 패키징 하여 Kubernetes 환경에서 배포하는 방식을 제안한다. 그림 3은 이러한 방식을 통해 실행 위치가 변경되더라도 동일한 이미지를 재사용할 수 있으며, 환경 의존성을 최소화하면서 고성능 모델의 분산 처리가 가능함을 보여준다.

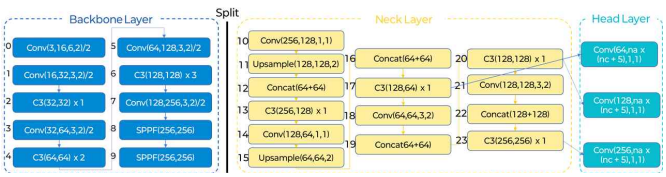


그림 3 : YOLOv5 모델 분할 테스트베드 Backbone - Neck - Head 계층 분할 지점

Ⅲ. 결론

본 논문에서는 AI 모델을 계층 단위로 분할하여 모빌리티 디바이스와 엣지 서버에 분산 배치하는 지능형 컨테이너 오케스트레이션 기법을 제안하였다. 초기 단계 연산은 모빌리티 디바이스에서, 후속 단계 연산은 엣지 서버에서 수행함으로써 연산 부하를 분산하고 지연을 최소화하였다. 제안 구조는 네트워크 상태와 시스템 자원 변동에 따라 실행 경로를 동적으로 조정할 수 있도록 설계되었으며, 실험 결과 모빌리티 디바이스 단독 처리 대비 지연 시간 단축과 처리 효율 향상이 기대된다. 향후에는 다양한 AI 모델과 서비스 환경으로의 확장 적용 및 실제 모빌리티 디바이스-엣지 테스트베드 기반 성능 검증을 수행할 예정이다.

ACKNOWLEDGMENT

이 논문은 2024년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구 결과임 (No.RS-2024-00406245, 미래 모빌리티를 위한 소프트웨어 정의형 인프라스트럭처 기술 개발)

참 고 문 헌

[1] W. Rafique, L. Qi, I. Yaqoob, et al., “Complementing IoT Services Through Software Defined Networking and Edge Computing: A Comprehensive Survey,” IEEE Communications Surveys & Tutorials, vol. 22, no. 3, pp. 1761 - 1804, 2020.

[2] K. C. Park, J. H. An, Y. H. Kim, H. Kim, “Scalable Orchestration Design for Mixed-Critical Task Scheduling in SDI-based Mobility Systems,” in Proc. ACM Symposium on Applied Computing (SAC), pp. 1407 - 1411, 2025.

[3] F. Lumpp, F. Fummi, H. D. Patel, N. Bombieri, “Enabling Kubernetes Orchestration of Mixed-Criticality Software for Autonomous Mobile Robots,” IEEE Transactions on Robotics, vol. 40, no. 2, pp. 540 - 553, 2024.

[4] J. C. Lee, Y. Kim, S. T. Moon, J. H. Ko, “A Splittable DNN-Based Object Detector for Edge-Cloud Collaborative Real-Time Video Inference,” in Proc. IEEE AVSS, pp. 1 - 8, 2021.

[5] P. Amanatidis, D. Karampatzakis, G. Iosifidis, T. Lagkas, A. Nikitas, “Cooperative Task Execution for Object Detection in Edge Computing: An IoT Application,” Applied Sciences, vol. 13, no. 8, Art. 4982, 2023.

지진파 신호 노이즈 제거를 위한 8개 딥러닝 모델 성능 비교 연구

황선준† , 박세희*, 고강민**, 백지윤***

† 연세대학교, *연세대학교, **연세대학교, ***연세대학교

† sunjun7559012@yonsei.ac.kr, *serra0508@yonsei.ac.kr,

**kkm4629@yonsei.ac.kr, ***bjy0609@yonsei.ac.kr

PerformanceComparisonof8DeepLearningModelsforSeismicSignalDenoising

Sunjun Hwang† , Sehee Park*, Gangmin Koh**, Jiyun Baik***

† Yonsei Univ, *Yonsei Univ, **Yonsei Univ. ***Yonsei Univ.

요 약

본 논문은 지진 예측 시스템의 신뢰성 향상을 위해 8개 딥러닝 모델의 지진파 노이즈 제거 성능을 체계적으로 비교 분석하였다. BiLSTM, Denoising Autoencoder, FFTformer, Informer, PatchTST, TCN, UNet1D, WaveNet 8개 모델을 동일한 조건에서 평가하여 각 모델의 특성과 성능을 정량적으로 분석하였다. 실험 결과 PatchTST가 MSE 0.0004, 정확도 99.98%로 가장 우수한 성능을 보였으며, UNet1D와 TCN이 그 뒤를 이었다. 특히 PatchTST는 오탐률 0.025%, 미탐률 0%를 달성하여 지진 예측 시스템에 가장 적합한 모델임을 입증하였다. 본 연구는 지진파 신호 처리 분야에서 Transformer 기반 모델의 우수성을 보여주며, 실용적인 지진 예측 시스템 구축에 기여할 것으로 기대된다.

Ⅰ. 서 론

지진은 막대한 인명 피해와 경제적 손실을 초래하는 자연재해로, 미국 지질조사국(USGS)에 따르면 매년 약 130만 건이 발생하며, 이 중 약 100건이 심각한 피해를 유발한다[1]. 2023년 터키-시리아 지진에서 5만명 이상이 사망하였고, 2025년 7월 캄차카 M7.4 지진은 신뢰할 수 있는 예측 시스템의 필요성을 다시 한번 보여주었다[2]. 국내에서도 2016년 경주지진(규모5.8)과 2017년 포항 지진(규모5.4) 이후 지진에 대한 관심이 높아지면서, 한반도 지질 특성에 맞는 예측 시스템 개발의 필요성이 대두되고 있다.

기존 지진 예측 시스템은 오탐(false positive)과 미탐(false negative)의 딜레마에 직면해 있다. 오탐은 사회적 혼란과 경제적 비용을 초래하며, 미탐은 치명적인 인명 피해를 야기한다. 이러한 문제는 기존 딥러닝 모델이 예측 불확실성을 적절히 정량화하지 못하는데서 기인하며, 최근 AI 안정성 연구에서는 예측과 함께 확신도를 제공하여 신뢰성을 향상시키는 방안이 연구되고 있다[3].

최근 딥러닝 기법의 발전으로 CNN 기반의 PhaseNet, Transformer 기반의 EQTransformer 등이 제안되어 상당한 성능 향상을 보였다. 하지만 기존 연구들은 대부분 단일 모델의 성능 향상에 집중되어 있으며, 다양한 아키텍처의 체계적 비교 연구는 상대적으로 부족한 상황이다. 따라서 본 연구는 지진파 노이즈 제거를 위한 최적의 딥러닝 모델을 찾고자 대표적인 딥러닝 모델들을 체계적으로 비교 분석한다.

본 연구에서는 Transformer 계열(FFTformer, Informer, PatchTST), 순환 신경망(BiLSTM), 오토인코더(Denoising Autoencoder), 합성곱 신경망(TCN, UNet1D, WaveNet) 등 다양한 아키텍처를 동일한 조건에서 평가하여 지진파 신호 처리에 가장 적합한 접근법을 제시한다.

Ⅱ. 본론

2.1 데이터셋 및 전처리

본 연구에서는 MSEED 형식의 지진파 데이터를 사용하여 체계적인 전처리 과정을 수행하였다. 데이터셋은 총 40,000개의 지진파 샘플로 구성되었으며, 이 중 30%가 실제 지진 신호, 70%가 배경 노이즈로 구성되었다. 각 샘플은 100Hz 샘플링 레이트로 10초간 기록된 1,000개의 데이터 포인트로 구성되었다.

전처리 과정은 다음과 같이 수행되었다. 먼저 선형 트렌드 제거(demean)를 통해 신호의 DC 성분을 제거하였다. 이후 코사인 테이퍼(20%)를 적용하여 신호 양단의 급격한 변화를 완화하였으며, 지진파의 주요 주파수 대역인 0.1-10.0Hz 범위에서 밴드패스 필터를 적용하였다. 마지막으로 각 샘플을 [-1, 1] 범위로 정규화하고 1,000 샘플 길이의 고정 윈도우로 분할하였다. 이러한 전처리 과정은 한국 지역 지진파의 특성을 고려하여 최적화되었으며, 기상청 및 한국지질자원연구원의 관측 데이터 형식에 맞춰 설계되었다.

2.2 딥러닝 모델 구성

본 연구에서는 8개의 대표적인 딥러닝 모델을 선정하여 성능을 비교 분석하였다. Transformer 계열로는 주파수 도메인 처리에 특화된 FFTformer, Sparse attention mechanism을 활용하는 Informer, 그리고 패치 기반 시계열 변환을 수행하는 PatchTST를 포함하였다. 순환 신경망 계열에서는 양방향 정보를 활용할 수 있는 BiLSTM을 선택하였으며, 오토인코더 계열에서는 전통적인 노이즈 제거 접근법인 Denoising Autoencoder를 포함하였다. 합성곱 신경망 계열에서는 Temporal Convolutional Networks(TCN), 1차원 U-Net 아키텍처(UNet1D), 그리고 Dilated convolution 기반의 WaveNet을 선정하였다.

모델	MSE	MAE	정확도	F1-score	오탐률	미탐률
PatchTST	**0.0004**	**0.0165**	**99.98%**	**0.0**	**0.025%**	**0.0%**
UNet1D	0.0328	0.0958	96.53%	0.941	1.19%	2.29%
TCN	0.0104	0.0712	96.46%	0.940	1.19%	2.34%
WaveNet	0.0144	0.0792	95.32%	0.920	1.53%	3.16%
BiLSTM	0.0016	0.0249	97.88%	0.0	0.0%	2.13%
Informer	0.0046	0.0591	97.88%	0.0	0.0%	2.13%
DenoisingAE	0.7426	0.4561	69.78%	0.0	0.0%	30.23%
FFTformer	0.7426	0.4560	69.78%	0.0	0.0%	30.23%

표1. 8개 딥러닝 모델 성능 비교 결과

2.3 실험 설정

모든 모델에 대해 동일한 실험 조건을 적용하여 공정한 비교를 수행하였다. 옵티마이저로는 Adam을 사용하였으며 학습률은 1e-3으로 설정하였다. 배치 크기는 16, 에포크는 20으로 제한하여 과적합을 방지하였다. 손실 함수로는 MSE Loss를 사용하였으며, 데이터셋은 훈련 70%(28,000 샘플), 검증 15%(6,000 샘플), 테스트 15%(6,000 샘플)로 분할하여 사용하였다. 실험 환경은 NVIDIA RTX 4090 GPU와 PyTorch 1.13 프레임워크, Python 3.8 환경에서 수행되었다.

2.4 평가 지표

모델 성능 평가를 위해 다층적 접근법을 사용하였다. 회귀 성능 지표로는 MSE(Mean Squared Error)와 MAE(Mean Absolute Error)를 사용하여 복원된 신호가 원본 신호와 얼마나 유사한지를 정량적으로 측정하였다. 동시에 예측 시스템의 실용적 성능을 평가하기 위해 분류 성능 지표로 정확도(Accuracy), F1-score, 오탐률(False Positive Rate), 미탐률(False Negative Rate)을 계산하였다. 분류 성능 평가를 위해서는 복원된 신호의 최대 진폭이 0.5를 초과하는 경우를 지진으로 분류하는 임계값 기반 접근법을 적용하였다.

2.5 실험 결과

실험 결과 PatchTST가 MSE 0.0004, MAE 0.0165, 정확도 99.98%로 가장 우수한 성능을 보였다. 특히 미탐률 0%를 달성하여 인명 보호라는 최우선 목표를 완벽히 달성하였으며, 오탐률 또한 0.025%로 사회적 비용을 최소화할 수 있음을 보여주었다. UNet1D와 TCN이 각각 두 번째, 세 번째로 우수한 성능을 보였으며, 두 모델 모두 96% 이상의 높은 정확도와 0.94 이상의 F1-score를 달성하였다.

반면 DenoisingAE와 FFTformer는 예상과 달리 매우 저조한 성능을 보였다. 두 모델 모두 MSE 0.74 이상, 정확도 69.78%로 실용적 사용이 어려운 수준이었으며, F1-score가 0.0으로 나타나 지진 신호를 전혀 탐지하지 못하는 것으로 분석되었다. 이러한 결과는 DenoisingAE의 경우 단순한 인코더-디코더 구조로 지진파의 복잡한 시간적 의존성을 모델링하지 못하였고, FFTformer의 경우 주파수 도메인 변환 과정에서 중요한 시간 정보가 손실되었기 때문으로 추정된다.

아키텍처별 특성을 분석한 결과, Transformer 계열 중에서는 PatchTST의 패치 기반 접근법이 지진파 신호의 국소적 특성과 전역적 패턴을 동시에 효과적으로 모델링할 수 있음을 확인하였다. 합성곱 신경망 계열에서는 UNet1D, TCN, WaveNet 모두 실용적으로 사용 가능한 성능을 보였으며, 특히 UNet1D의 skip connection이 원본 신호 복원에 효과적임을 확인하였다. 순환 신경망인 BiLSTM은 중간 수준의 회귀 성능을 보였으나, F1-score 0.0으로 지진 탐지에는 한계가 있는 것으로 나타났다.

Ⅲ. 결론

본 연구에서는 지진 예측 시스템의 신뢰성 향상을 위해 8개 딥러닝 모델의 지진파 노이즈 제거 성능을 체계적으로 비교 분석하였다. 실험 결과 PatchTST가 MSE 0.0004, 정확도 99.98%, 오탐률 0.025%, 미탐률 0%를 달성하며 가장 우수한 성능을 보였다. 이는 실제 지진 예측 시스템에 즉시 적용 가능한 수준의 예측력임을 입증한다.

UNet1D와 TCN이 각각 두 번째, 세 번째로 우수한 성능을 보였으며, 합성곱 기반 아키텍처의 효과를 확인하였다. 반면 DenoisingAE와 FFTformer는 예상과 달리 매우 저조한 성능을 보여 추가적인 구조 개선이 필요함을 시사하였다.

본 연구의 주요 기여점은 다음과 같다. 첫째, 8개 딥러닝 모델에 대한 체계적이고 공정한 비교 분석을 통해 지진파 신호 처리에 최적인 모델을 명확히 제시하였다. 둘째, 동일한 실험 조건과 평가 지표를 사용하여 객관적이고 재현 가능한 결과를 도출하였다. 셋째, 회귀 성능과 분류 성능을 모두 고려한 다층적 평가를 통해 실용적 관점에서의 모델 성능을 종합적으로 분석하였다.

향후 연구 방향으로는 DenoisingAE와 FFTformer의 구조적 한계 분석 및 개선, PatchTST와 UNet1D를 결합한 앙상블 접근법 연구, 모델 경량화를 통한 실시간 지진 예측 시스템 구축, Monte Carlo Dropout을 활용한 예측 신뢰도 향상 연구 등이 있다.

본 연구는 신뢰할 수 있는 AI 기반 지진 예측 시스템 구축에 중요한 기반을 제공하며, 특히 PatchTST 모델의 우수한 성능은 실제 운영 환경에서의 적용 가능성을 높였다. 궁극적으로 이러한 연구 성과는 지진 재해로부터 인명과 재산을 보호하는데 실질적으로 기여할 것으로 기대된다.

참 고 문 헌

[1] United States Geological Survey (USGS). “Earthquake Statistics.”

[2] United States Geological Survey (USGS).“M7.4 Kamchatka Peninsula, Russia” *USGS Earthquake catalog*, July 2025.

[3] Gal, Y., & Ghahramani, Z. (2016). “Dropout as a Baysian Approximation: Representing Model Uncertainty in Deep Learning.” In *Proceedings of the 33rd International Conference on Machine Leargning*(ICML), pp.1050-1059.

감성-VMD-기술지표 융합 cAR-GAN 합성데이터와 주가 로그수익률 예측 성능 검증

이지훈, 김재윤*
순천향대학교

20201468@sch.ac.kr, *kimym38@sch.ac.kr

Evaluating Stock Log-Return Forecasting Performance with cAR-GAN Synthetic Data Fusing Sentiment, VMD, and Technical Indicators

Lee Jihun, Kim Jaeyun*
Soonchunhyang Univ.

요 약

본 연구는 뉴스 감성과 VMD-PSO 로 추정한 주파수 모드를 조건으로 하는 조건부 AR-GAN(cAR-GAN)으로 주가 로그수익률을 합성하고, SMA(5·20)·RSI·20 일 롤링 변동성·시차 로그수익을 결합해 입력을 확장하였다. 합성물의 분포·의존 구조는 MMD, Jensen-Shannon divergence, ACF/PACF 로 점검하였다. LSTM, GRU, Transformer, TCN, CNN1D 모델을 KOSPI·KOSDAQ·NASDAQ·S&P500 네 지수에 대해 학습·평가하고, 실제(Real) 대비 실제+가상(Real+Synthetic) 학습을 동일 테스트 구간에서 비교한 결과, RMSE/MAE 기준으로 실제+가상이 전반적으로 더 낮은 오차를 보였다. 또한 변동성 분위, 감성 삼분위, IMF 대역별로 층화해 평가한 결과, 고변동성·부정 감성·고주파 구간에서 개선 폭이 특히 컸다. 이는 감성·주파수·기술지표의 구간화 결합이 실제+가상 학습의 예측력을 일관되게 강화함을 시사한다.

Ⅰ. 서 론

금융 빅데이터의 확산에도 불구하고 비정상성·구조적 변화로 인해 기존 딥러닝·머신 러닝의 일반화 성능은 여전히 제약을 받는다. 특히 시가·고가·저가·종가 중심 입력만으로는 시장의 복합 동학을 충분히 포착하기 어렵다는 지적이 지속되어 왔다[1].

이를 보완하기 위해 신호를 대역별 IMF 로 분해해 노이즈와 비선형성을 완화하는 변분 모드 분해(VMD)가 제안되었고, 금융 예측에서는 PSO 와 결합한 VMD 파라미터 최적화 사례가 축적되었다[1].

한편 뉴스·소셜미디어 감성은 단기 수익률 및 거래활동을 예측한다는 근거가 다수 보고되었다[2].

생성 모델 측면에서는 조건부 GAN 이 조건부 분포 모사를 위한 표준 틀을 제공하고, 시계열 특화 TimeGAN 계열이 시간 의존성 보존을 목표로 발전해 왔다[3].

본 연구는 이러한 흐름을 통합하여, 주파수 모드(VMD), 뉴스 감성, 기술적 지표(SMA(5·20)/EMA, RSI, 20 일 롤링 변동성, 시차 로그수익)를 조건으로 하는 조건부 AR-GAN(cAR-GAN)을 구성하고, 이를 통해 주가 로그수익률 합성 데이터를 생성한다.

생성물의 분포 및 시계열 의존 구조는 커널 기반 이표본 거리(MMD)와 ACF/PACF 로 정합성을 검증한다[3]. 마지막으로 LSTM, GRU, Transformer, TCN, 1D-CNN 에 대해 실제와 실제+가상 학습을 비교하여 외부표본 예측 오차(RMSE/MAE)를 평가한다.

Ⅱ. 본론

본 연구는 (i) 가격·거래 자료에 뉴스 감성, VMD-PSO 기반 주파수 모드, 기술적 지표(SMA/EMA, RSI, 롤링 변동성 등)를 결합해 특성 공간을 구성하고, (ii) 이를 조건으로 하는 조건부 AR-GAN(cAR-GAN) 으로 로그수익률 합성 데이터를 생성한다. (iii) 생성물의 통계·시계열 정합성은 MMD, JS, ACF/PACF 로 검증한다. (iv) 2014-2022 학습 / 2023-이후 테스트의 고정 분할에서 LSTM, GRU, 1D-CNN, TCN, Transformer 의 외부표본 성능을 Real vs. Real+Synthetic 으로 비교한다. (v) 더 나아가 모델 간 비교 분석을 수행하여 평균 순위, 유의성(예: Diebold-Mariano, Wilcoxon), 효과크기(예: Cliff's) 및 방향 정확도(DA) 등을 보고하고, (vi) 변동성 분위·감성 삼분위·주파수 대역 구간화와 층화 증강으로 개선 효과의 견고성을 점검한다.

Ⅱ.A 데이터셋 및 전처리

국내·해외 지수(KOSPI, KOSDAQ, S&P500, NASDAQ)와 선택 종목의 일별 OHLCV 를 사용하며, 표적은 종가 로그수익률 $r_t = \ln(P_t / p_{t-1})$, 예측 타깃은 y_{t+1} 이다. 데이터는 2014-2022 년을 학습·검증, 2023 년 이후를 테스트로 분할한다. 데이터는 FinanceDataReader(국내)·yfinance(해외)에서 수집 후 비거래일 제거, 이상치 보정, 정규화(Open, High, Low, Close, Volume)를 수행하였다.

뉴스 데이터는 크롤링 후 광고·중복 제거, 문장별 감성을 평균해 일별 감성·기사 건수로 집계하였다.

입력 X_t는 기술적 지표(SMA, EMA, RSI, 20 일 변동성, 시차 로그수익), 거래량 및 변화율, VMD-PSO 기반 IMF 모드를 포함한다. 모든 피처는 거래일 기준으로 병합하고, 학습 구간으로 스케일링을 fit 하여 검증·테스트에 동일 적용하였다. 예측 타깃 y_{t+1} 은 t 시점 입력에 대한 1-step ahead 로그수익률이다.

Ⅱ.B 주파수 분해 및 특징 구성

각 자산의 로그수익률 x_t을 변분 모드 분해(VMD)로 K개의 본질모드(IMF)$u_k(t)$로 분해한다. 모드 수 K (3-8)와 평활화 계수 α(500-5000)는 입자군집화(PSO)로 공동 탐색하며, 목적함수는 재구성 오차 MSE $(x_t \sum_{k=1}^{K} u_k, (t;\alpha))$ 최소화이다. 수치 불안정 또는

비정상 해에는 큰 패널티를 부여하고 최대 3 회 재시도로 안정화한다. 분해는 스케일링 전에 원시 로그수익률에서 수행하고, 이후 표준화/정규화는 학습 구간(2014-2022)에만 적합해 검증·테스트에 동일 적용한다. 일반적으로 앞선 모드는 중·고주파(단기 변동), 뒤쪽 모드는 저주파(추세) 성분을 반영한다.

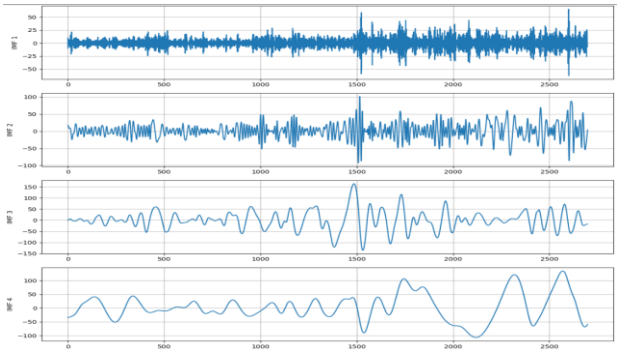


그림.1 로그수익률 원신호와 VMD 재구성(IMF) 비교

Ⅱ.C 합성 모델(cAR-GAN)과 품질 검증

본 연구의 합성기는 조건부 AR-GAN(cAR-GAN)이다. 시점 t 의 조건 c_t 에는 SMA(5·20), EMA(5·20), RSI(14), 20 일 롤링 변동성, 시차 수익률(1·5·10), 거래량·증감, VMD-PSO 로 얻은 IMF 성분, 일별 뉴스 감성(평균·건수)을 포함한다. 생성기는 잡음 z, c_t, 최근 수익률(AR skip)을 입력 받아 $\tilde{r}_{t+1}$을 생성하고, 판별기는 (r_{t+1}, c_t)의 진위를 판정한다. 조건은 임베딩 후 결합하며, 인과성 유지를 위해 슬라이딩 윈도우로 학습한다. 손실은 표준 cGAN 목적에 안정화 정규화를 더해 모드 붕괴를 억제한다.

품질 검증은 (i) 분포 정합: MMD(RBF)·Jensen-Shannon divergence, (ii) 시계열 의존: ACF/PACF(수익률·제곱수익률), (iii) 조건 일관성: 변동성 분위·감성 삼분위·IMF 대역별로 층화 후 동일 지표 재평가의 세 축으로 수행한다. 모든 평가는 학습기(2014-2022) 내 홀드아웃 블록에서도 반복해 과적합 가능성을 점검한다.

Ⅱ.D 예측 실험 및 결과

KOSPI·KOSDAQ·NASDAQ·S&P500 의 일별 로그수익률에 대해 실제 학습과 실제+가상 학습을 동일 테스트 구간에서 비교했다.

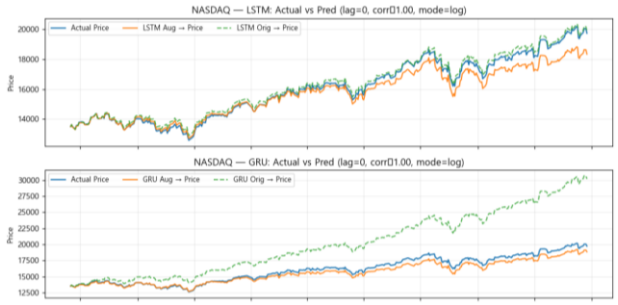


그림.2 실제 가격과 Real/Real+ Synthetic 예측 비교

모델은 LSTM, GRU, Transformer, TCN, CNN1D 의 다섯 가지를 사용하고, RMSE/MAE 및 R^2로 평가하였다. 합성데이터를 포함한 학습이 전반적으로 더 낮은 오류를 보여 20 개 조합 중 11 개에서 RMSE 가 개선되었고, 특히 GRU 는 네 시장 모두에서 오차가 유의하게 감소했으며, NASDAQ 에서는 RMSE 가 0.00821 에서 0.00423 으로, S&P500 에서는 0.00659 에서 0.00464 로 낮아지는 등 평균 약 30%의 개선을 보였다. LSTM 은 KOSPI , KOSDAQ, NASDAQ 에서 개선되었고, NASDAQ-LSTM 에서도 $DMp \approx 1.17 \times 10^{-3}, Wilcoxonp \approx 2.38 \times 10^{-3}$ 으로 통계적으로 유의미했다. 급변 구간에서도 합성 학습 모델의 설명력은 대체로 높게 유지되어 희소/스트레스 구간 R^2 가 약 0.87-0.99 범위에서 관측되었고, 정상 구간에서도 유사한 수준을 보였다.

Table 1. 실제 대비 증강 성능 비교

Model	Real		Real+ Synthetic	
	RMSE	R^2	RMSE	R^2
LSTM	0.011	0.977		
GRU	0.007	0.972		
TF	0.014	0.936		
TCN	0.011	0.977		
CNN	0.011	0.965		
평균	0.011	0.965		

Ⅲ. 결론

뉴스 감성, VMD-PSO 기반 주파수 모드, 기술적 지표를 조건으로 한 cAR-GAN 증강은 일별 주가 로그수익률 예측에서 전반적으로 외부표본 오차를 낮추는 경향을 보였다. 개선 효과는 시장과 모델 전반에서 관찰되었고, 특히 고변동성·부정 감성·고주파 구간에서 더 뚜렷했다. 합성 시계열은 분포·의존구조 측면(MMD, JS-divergence, ACF/PACF)에서 양호한 정합성을 보여, 예측 개선이 단순 과적합의 결과가 아님을 시사한다.

첫째, 감성-주파수-기술지표를 함께 조건화한 다중 조건부 증강이 가격·거래량 중심 입력의 정보 공백을 메워 다양한 예측기에 일관된 일반화 이익을 제공한다. 둘째, 변동성·감성·주파수에 따른 구간화(층화) 증강이 데이터 희소·편향 구간의 견고성을 높인다.

한계로는 일별 자료와 고정 분할에 따른 시간가변성 평가의 제약, 감성 소스·모델 편향 및 VMD 파라미터 민감도, 그리고 예측 개선을 거래비용을 포함한 실현 성과로 연결하는 후속 검증의 부재가 있다. 향후 롤링/확장 검증, 대안 생성기(확산·flow) 비교, 비용·리스크 제약을 반영한 백테스트로 확장할 필요가 있다.

ACKNOWLEDGMENT

본 연구는 2025 년 과학기술정보통신부 및 정보통신기획평가원의 SW 중심대학사업의 연구 결과로 수행되었음(2021-0-01399).

참 고 문 헌

[1] Salim Lahmiri, “Intraday stock price forecasting based on variational mode decomposition,” Journal of Computational Science, 12 (2016), 23-27.

[2] Rahul Verma, Priti Verma, “Economic News, Social Media Sentiments, and Stock Returns: Which Is a Bigger Driver?,” Journal of Risk and Financial Management, 18(1) (2025), 16.

[3] Arthur Gretton, Karsten M. Borgwardt, Malte J. Rasch, Bernhard Schölkopf, Alexander Smola, “A Kernel Two-Sample Test,” Journal of Machine Learning Research, 13 (2012), 723-773.

마스크 조건 TimeGAN 기반 비정규 센싱 데이터의 정규 시계열 복원

김현정, 이연서, 김파란하늘, 현장훈*

국립한밭대학교

20231052@edu.hanbat.ac.kr, *jhhyeon@hanbat.ac.kr

Masked TimeGAN for Regular Time-Series Reconstruction from Irregular Sensing Data

Hyeonjeong Kim, Yeonseo Lee, Paranhaneul Kim, Janghun Hyeon*

Hanbat National University

요 약

이상 탐지 시스템의 일반화를 위해서는 일관된 센싱 주기로 정렬된 입력이 요구된다. 그러나 실제 현장 데이터는 제조사나 설정의 차이와 통신 오류로 인해 배터리마다 주기가 다르며 센싱 주기가 길게 지속되는 경우, 기존 보간 기법만으로는 신호의 고정 주기적 특징을 안정적으로 복원하는 데 한계가 있다. 본 논문은 이런 한계를 보완하기 위해 생성형 모델을 마스크 기반 생성형 모델로 확장하여 관측값은 보존하고 결측만 복원함으로써 불규칙한 센싱 데이터를 단일 고정 격자로 일관되게 변환하고 이상 탐지 시스템의 이식성을 고려한 고정 주기 재구성 방법을 제안한다. 실험 결과, 제안한 모델은 유효한 조건 내에서 MAE 최대 19.3%, MSE 최대 20.5% 감소를 보여 선형 보간 대비 시점 또는 구조적 패턴을 더 안정적으로 복원함을 확인하였다.

Ⅰ. 서론

ESS(에너지 저장 시스템) 배터리에서 화재와 같은 이상 탐지 시에 일관된 센싱 주기의 데이터를 이용하면 이상 탐지 성능의 안정성을 기대할 수 있다[1]. 센싱 주기가 서로 다를 때, 센싱 주기가 길수록 정보량이 줄어들어 짧은 이상 신호에 대한 탐지가 어려워지기 때문에 이상 탐지 성능의 불안정성이 커진다. 그러나 실제 환경에서는 운영 조건과 장비 설정 차이로 인해 ESS마다 센싱 주기가 달라 일관된 센싱 주기 데이터를 확보하기 어렵다. 따라서 혼재 센싱 주기 데이터를 공통의 고정 주기 격자로 재구성하는 전처리가 필요하다.

고정 주기 격자 재구성은 선형 보간(Linear)[2]이나 3차 회선 보간(Cubic)[2]과 같은 수학적 기법으로 가능하다. 하지만 이러한 방법들은 [그림 1-ⓐ]같이 전체적인 데이터 패턴을 고려하지 않고 인접한 데이터에만 의존하기 때문에 샘플링 주기가 길어질수록 세밀한 변동을 정밀하게 복원하는 데 한계가 존재한다.

본 연구는 관측된 센싱 주기보다 짧은 센싱 주기를 갖는 고정 주기 시계열을 생성하도록 관측값을 기반으로 비관측 값만 보간하는 생성형 모델을 제안한다. 이를 통해 전체적인 패턴을 고려하면서 혼재 센싱 주기 데이터를 고정 주기의 격자로 일관되게 재구성하는 방법을 제안한다.

Ⅱ. 본론

본 연구는 [그림 1-ⓑ]와 같이 입력으로 사용하기 위해 다른 주기로 리샘플링된 데이터와 해당 센싱 주기 시점의 위치를 나타내는 마스크를 입력으로 사용한다. 또한 기존 시계열 생성형 모델의 손실을 수정하여 비관측 구간만을 복원하도록 한다. 이를 기반으로 학습된 모델은 테스트 데이터에 적용되어 임의의 혼재 주기 데이터에 대해 고정 주기 데이터로 재구성하는 성능을 평가한다.

1. 데이터 전처리

데이터는 배터리 rack을 기준으로 여러 module로 나누어지며, 각 module은 다시 여러 개의 cell로 구성된 계층적 구조를 가진다. 이 구조에서 설정된 주기를 가지도록 개별 cell 단위마다의 전압값을 저장한다.

사용한 데이터는 센서 오류 및 방전에 의해 의미 없는 0값이 많아 스케일링 과정에서 왜곡되지 않도록 하기 위해 Min-Max 스케일링을 적용한다. 스케일링 과정에서 데이터 누수를 방지하기 위해 학습 데이터에서의 Min, Max 값을 이용한다. 이후 [그림 1-ⓑ]의 리샘플링 과정을 수행한다. 리샘플링 과정은 혼재 주기 환경을 모사하기 위해 고정된 주기로 센싱된 데이터를 다양한 센싱 주기로 리샘플링하여 센싱 주기 별로 리샘플링 데이터를 생성한다. 리샘플링 시, 해당 주기 시점의 값은 관측 영역으로 판단하여 그대로 보존하고 그 외 시점은 0으로 채운다. 또한 별도의 채널을 만들어 리샘플링 데이터에서 실제 값이 존재하는 시점은 1로, 0으로 채워진 시점은 0으로 표시하는 마스크 벡터를 생성한다.

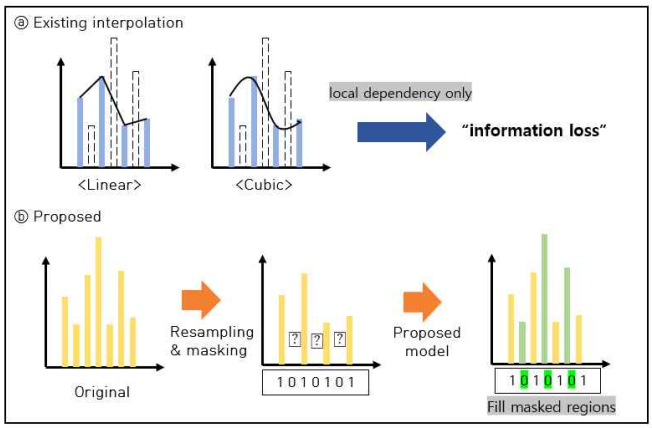


그림 1. 기존 보간의 한계 및 제안 방법

2. 마스크 조건부 생성형 모델

본 논문은 잠재 공간에서 시간적 동역학을 보존하며 실제와 유사한 시퀀스를 생성하는 시계열 생성 모델 TimeGAN[3]을 기반으로 비관측 구간 보간을 통해 고정 주기로 재구성하는 네트워크를 제안한다. 기존 TimeGAN은 전체 시퀀스 생성을 목적으로 설계되어 관측 값을 보존하고 결측 구간만 선택적으로 생성할 수 없다. 따라서 TimeGAN의 입력을 확장하고 손실을 재설계함으로써 결측 구간을 보간하여 고정 주기 센싱 데이터를 생성하도록 한다.

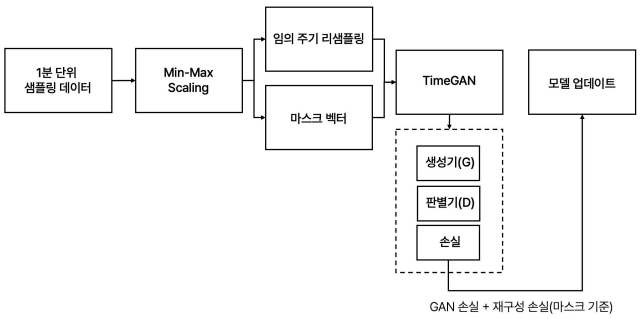


그림 2. 마스크 기반 확장 모델 학습 구조

2.1 모델 구조

제안된 모델은 혼재 주기 데이터와 해당 위치의 관측 여부를 나타내는 마스크 벡터를 입력으로 비관측 구간을 모델이 직접 복원하도록 한다. 생성기는 리샘플링된 데이터와 마스크 벡터를 입력으로 받아 관측 구간에서는 원래 값을 유지하고 비관측 구간에서는 새로운 값을 생성하여 고정 주기로 재구성한다. 판별기는 이렇게 복원된 시계열과 실제 원본 시계열을 구분한다. 학습 과정에서는 입력 데이터를 무빙 윈도우 방식으로 일정 길이의 시퀀스로 분할하여 배치 단위로 학습시키며 이를 통해 장기적 패턴과 국소적 패턴을 동시에 학습한다[그림 2].

2.2 손실 함수 설계

본 연구는 기존 TimeGAN 손실 구조를 기반으로 손실 함수를 수정 및 확장한다. 기존 TimeGAN은 재구성 손실(Reconstruction Loss), 감독 손실(Supervised Loss), 적대적 손실(Adversarial Loss), 임베딩 손실(Embedding Loss)로 구성된다. 본 연구에서는 재구성 손실을 비관측 구간에 초점을 맞추어 학습하도록 하는 인페인팅 손실(Inpainting Loss)로 수정한다. 또한 관측 구간에서 입력값과 복원된 값이 동일하게 유지되도록하는 아이덴티티 보존 손실(Identity Preservation Loss)을 추가한다.

2.2.1 인페인팅 손실

$$L_{inp} = E[\,\| (1-M)\odot(X-\hat{X}) \|\,] \quad (1)$$

기존 재구성 손실은 임베더-리커버리 네트워크의 출력 $\hat{X}$과 원본 X의 차이를 윈도우 전체 시계열 구간에서 계산한다. 그러나 본 연구는 비관측 영역 손실에 대해 최적화해야하기 때문에 해당 주기 시점의 위치는 1, 그 외 시점을 0으로 하는 마스크 벡터 M에 따라서 마스크 벡터 M이 비관측(0)일 때만 손실에 반영되도록 하는 인페인팅 손실(1)로 수정한다.

2.2.2 아이덴티티 보존 손실

$$L_{id} = E[\,\| M\odot(X-\hat{X}) \|\,] \quad (2)$$

아이덴티티 보존 손실(2)은 관측 값을 훼손하지 않도록 강제하는 역할을 한다. 마스크 벡터 M을 통해 관측된 위치(1)에서만 출력 $\hat{X}$과 원본 X가 같아지도록 오차를 계산한다.

3. 실험 설계

실험은 제안한 모델과 전통적인 보간 기법인 선형보간(Linear), 3차 회선 보간(Cubic)을 비교한다. 비교 과정에서 3.3V 이하의 값을 방전 값 또는 센서 오류로 인한 값으로 판단하여 3.3V 이하의 값은 평가에서 제외한다. 또한 정답이 NaN인 구간은 단순 결측이 아닌 측정 종료로 해석하여 이 구간에서 모델이 3.3V 이상을 예측하면 패널티를 부여하여 평균 절대 오차(Mean Absolute Error)와 평균 제곱 오차(Mean Squared Error)로 성능을 평가한다.

4. 결과

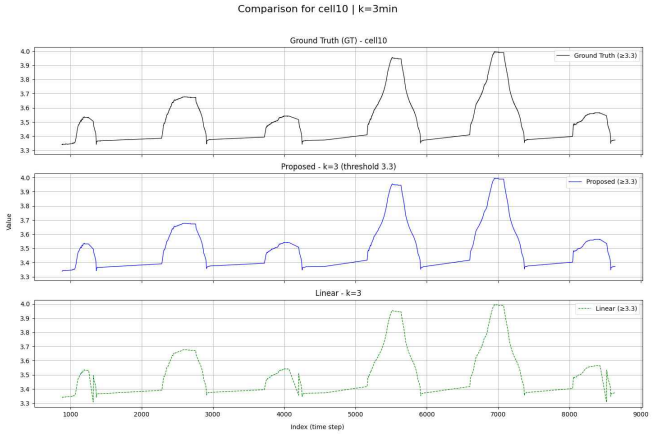


그림 3. 제안 방법 및 선형 보간 결과 예시 (k=3)

[그림 3]은 선형 보간과 제안 모델의 재구성 예를 비교한 것이다. 선형 보간은 고정 주기화 자체는 가능하지만 왜곡이 나타나는 반면, 제안 모델은 상승, 하강 시점 지속 등 구조적 패턴을 더 안정적으로 복원함을 볼 수 있다.

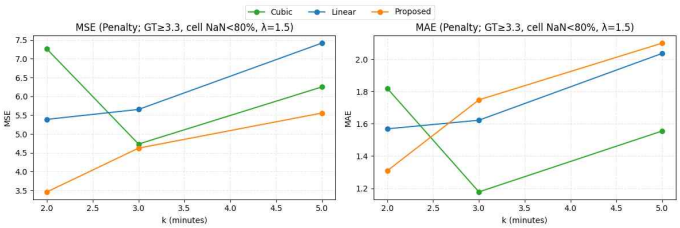


그림 4. 기존 보간기법과 제안 모델 보간의 성능 비교

[그림 4]를 볼 때, 제안된 모델은 전통적인 보간 기법보다 짧은 샘플링 간격에서 최저 오차를 보이며 긴 간격에서도 경쟁력을 유지하는 균형적인 특성을 보인다. 하지만 평균 절대 오차 관점에서는 장주기에서 선형, 3차 회선 보간이 일부 구간에서 더 낮은 값을 보이는 한계가 존재한다.

Ⅲ. 결론

본 연구는 관측값을 그대로 보존하고 결측만 복원하도록 학습하는 마스크 조건 구조를 통해 서로 다른 샘플링 주기를 일관된 주기로 정규화하는 방법을 제안한다. 제안한 모델은 유효한 조건에서 실용 범위(k<5분)에서는 평균 절대 오차 최대 19.3%, 평균 제곱 오차 최대 20.5% 감소를 확인하여 샘플링 주기가 다양한 환경에서도 일정한 주기로 보정이 가능함을 보인다. 한편 조건이 없는 환경에서는 단순 보간(Linear, Cubic)이 점추정 정확도에서 더 유리한 경우가 존재해 제안 모델의 우월성을 모든 상황으로 일반화하긴 어렵다. 그럼에도 제안한 방식은 보간법의 구조적 한계를 딥러닝으로 보완하는 대안으로서 다양한 주기 환경에서 신뢰 가능한 시계열 복원을 제공한다.

참 고 문 헌

[1] Liu W., Zhang Z., Zhao Z., and Zhang W., “A Framework for Anomaly Cell Detection in Energy Storage Systems Based on Daily Operating Voltage and Capacity Increment Curves,” Batteries, vol. 11, no. 8, p. 316, 2025

[2] Burden R. L. and Faires J. D., Numerical Analysis, 9th ed., Brooks/Cole, 2010.

[3] Yoon J., Jarrett D., and van der Schaar M., “Time-series Generative Adversarial Networks,” Advances in Neural Information Processing Systems (NeurIPS), vol. 32, 2019,

연속학습의 자율주행 엣지 스트리밍 적응성 분석

이지은, 홍서희, 김지수, 김윤희
숙명여자대학교

{mariewldms, seohee0320, smwujisue, yulan}@sookmyung.ac.kr

An Adaptability Analysis of Autonomous Edge Streaming using Continual Learning

Jieun Lee, Seohee Hong, Jisue Kim, Yoonhee Kim
Sookmyung Women's University

요 약

자율주행 차량의 주행 중 실시간 스트리밍 데이터 처리 성능 향상을 위해 엣지 내에서 딥러닝 학습 연구가 진행되고 있다. 그러나, 딥러닝 학습 중 실시간으로 모델 파라미터가 조정되면서 파괴적 망각이 발생하고 실시간 적응 능력이 떨어지므로 연속학습(Continual Learning)을 도입할 필요가 있다. 본 논문에서는 연속학습을 온라인 학습 알고리즘에 도입할 학습 모델을 제안한다. 실험을 통해 스트리밍 데이터에 연속학습을 적용하면 학습 중 GPU 사용량이 크게 증가하지 않으면서 과거 기억을 보존하고 동시에 새로운 데이터에 대한 적응 능력이 향상되었음을 보였다.

Ⅰ. 서론

온디바이스 AI 의 중요성이 부각됨에 따라, 자율주행 엣지 내 실시간 스트리밍 데이터 처리 과정에서 딥러닝 학습이 시도되고 있다. 그러나 새로운 데이터 학습 시 기존 모델 가중치가 조정되면서 과거 지식이 급격히 손실되는 파괴적 망각(Catastrophic forgetting) 문제가 발생한다. 이러한 상황에서 연속학습(Continual Learning)은 망각을 방지하면서 실시간 데이터에 적응할 수 있는 방법[1]으로 주목받고 있다. 따라서 자원이 한정된 자율주행 엣지에서 연속학습은 실시간 적응 능력 향상을 위한 해결책이 될 수 있다.

본 논문에서는 자율주행 엣지의 스트리밍 데이터 처리 과정에서 연속학습을 적용할 수 있는 온라인 학습 모델을 제시하고, 학습 및 평가 결과를 기반으로 연속학습 알고리즘의 성능을 평가한다.

Ⅱ. FM 기반 온라인 연속학습 모델

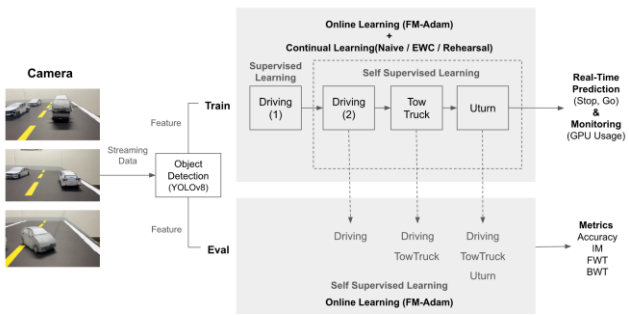


그림 1. 자율주행 엣지의 FM 기반 온라인 학습 모델

2.1 온라인 연속학습 모델

본 논문에서는 그림 1 과 같이 자율주행 엣지에 연속학습을 도입하여 주행 여부를 추론하는 온라인 연속학습 모델을 제안한다. YOLO v8 을 통한 객체 탐지 결과를 바탕으로 바운딩박스 좌표, 추정 거리, 중심점 좌표 등을 포함한 특징 벡터를 추출하고, 이를 FM(Factorization Machine)[3] 기반의 분류 모델에 입력하여 주행 여부를 결정하는 이진 분류 모델을 학습한다. 지도학습을 통한 초기 모델 구축 후 자기 지도학습을 통해 라벨이 없는 실시간 데이터에 대해 지속적인 학습을 수행한다 [2]. 자기 지도학습 단계에서는 원본 입력과 노이즈가 추가된 입력 간의 임베딩 일관성을 유지하는 일관성 손실을 활용하여 라벨이 없는 상황에서도 효과적인 학습 신호를 생성한다. 모델은 새로운 프레임이 유입될 때마다 즉시 학습을 수행할 수 있으며, Adam optimizer 를 활용한 효율적인 파라미터 업데이트를 통해 엣지의 제약사항을 고려한 실용적인 온라인 연속학습 모델을 구현한다. 각 태스크가 마무리될 때마다 자기 지도학습 기반 평가 모델 예측값과 임의 라벨(ground truth)을 비교하여 연속학습 성능 지표를 통해 성능을 종합적으로 평가한다.

2.2 연속학습

본 논문에서는 대표적인 연속학습 알고리즘[1]인 Regularization-Based Approach 방법론 기반 EWC 와 Replay-Based Approach 방법론 기반 Rehearsal 을 활용한다. EWC 는 가중치 변화를 최소화하는 방식으로, 중요한 가중치일수록 페널티가 증가해 값 이전 기억을 보존한다. Rehearsal 은 학습 과정에서 이전에 학습한 데이터 샘플들을 메모리에 저장해두고, 새로운 태스크를 학습할 때 이전 데이터를 함께 사용하여 모델이 이전 지식을 망각하지 않도록 한다.

Ⅲ. 실험 및 결과

3.1 스트리밍 데이터

본 논문에서는 실제 자율주행 상황에서의 도로와 주행 차량을 모방한 실험환경을 설계하고 RC 카 시점 전면 영상을 촬영했다. 성능 측정을 위해 기존 자율주행 알고리즘이 예측하기 어려운 돌발 상황을 설계했다. 앞 자동차 주행 상황(일반 주행), 견인차에 실린 자동차가 낙하하는 상황(견인차 낙하), 맞은편 도로에서 주행하던 자동차가 유턴하는 상황(유턴)을 가정했다. 지도학습 모델에서 일반 주행 데이터가, 자기 지도학습 모델에서 일반 주행, 견인차 낙하, 유턴 데이터가 프레임 단위로 차례대로 학습된다.

3.2 주행 여부 라벨링

327 개의 프레임에 대해 0 을 '멈춤', 1 을 '주행'으로 하는 라벨을 붙이고, 지도학습을 수행하였다. 이후 854 개의 프레임은 라벨 없이 자기 지도학습을 수행하였다.

3.3 성능 지표

평가에 활용할 성능 지표로는 IM(Intransience measure), FWT(Forward transfer), BWT(Backward transfer)를 사용하였다. 통상적인 지표 [1]보다 간소화된 버전으로 사용하였다.

IM 은 식 (1)과 같이 최종적으로 모델이 모든 평가 실험에 대해 얼마나 평균적으로 잘 수행했는지를 의미한다.

$$\mathrm{IM} = \frac{1}{T}\sum_{i=1}^{T} A_i \tag{1}$$

BWT 는 식 (2)와 같이 마지막 평가 실험이 다른 과거 평가 실험들보다 상대적으로 잘 수행되는지를 나타낸다.

$$\mathrm{BWT} = \frac{1}{T-1}\sum_{i=1}^{T-1}(A_T - A_i) \tag{2}$$

FWT 는 식 (3)와 같이 baseline_acc=0.5 대비 새로운 평가 실험 후 성능이 얼마나 향상되었는지를 보여준다.

$$\mathrm{FWT} = \frac{1}{T-1}\sum_{i=2}^{T}(A_i - \mathrm{baseline_acc}) \tag{3}$$

T는 전체 평가 실험 수, A_i는 i 번째 평가 실험의 정확도이다.

3.4 실험결과

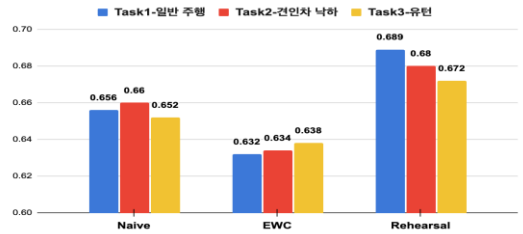


그림 2. 각 태스크 학습 종료 시점의 일반 주행 데이터 정확도 변화

학습 모델이 동작하는 자율주행 엣지로 NVIDIA Jetson AGX Orin 32GB 를 사용한다. 각 태스크별 학습이 마무리될 때마다 첫 번째 태스크에서 학습했던 일반 주행 데이터 정확도를 측정해 과거 데이터 보존 능력인 안정성을 평가했다. 그림 2 와 같이 Rehearsal 기법이 연속학습 미적용 모델인 Naive 대비 5%, 3%, 3% 정확도가 높았다. EWC 기법은 다른 두 기법에 비해 정확도는 떨어지지만, 성능 변화폭이 0.6%로 가장 작아 안정성 측면에서 뛰어나다.

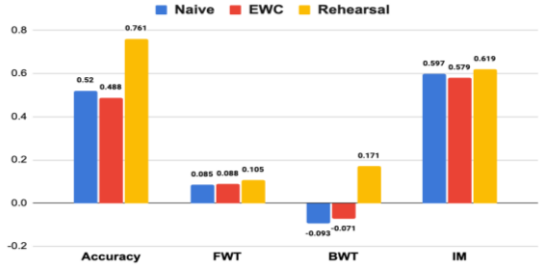


그림 3. 모든 태스크 학습 후 유턴 데이터 성능 평가 결과 비교

세 번째 유턴 태스크 학습 종료 후 유턴 데이터를 이용해 모델을 평가해 최종 성능 지표를 계산했다. 그림 3 에서 Rehearsal 기법이 Accuracy 0.761, FWT 0.105, BWT 0.171, IM 0.61 로 모든 성능 지표에서 우위를 보였다. 특히 안정성을 나타내는 BWT 에서 유일하게 양수 값을 기록하면서 적응성을 나타내는 FWT 가 가장 커 파괴적 망각 문제를 해결하면서 적응 능력이 뛰어남을 보인다.

	Naïve	EWC	Rehearsal
GPU Usage(%)	41.5	41.83	41.40

표 1. 온라인 학습 중 GPU 사용량 평균(%)

온라인 학습 중 엣지의 GPU 사용량 평균값을 측정한 결과, 표 1 과 같이 GPU 사용량이 연속학습 알고리즘을 사용했음에도 41%로 크게 차이 나지 않았다.

Ⅳ. 결론

본 논문에서는 자율주행 엣지에 실시간으로 유입되는 스트리밍 데이터를 기반으로 주행 여부를 판단하는 과정에서 온라인 학습 알고리즘에 연속학습을 적용하는 모델을 제안하였다. 또한, 엣지의 모델 학습 과정에서 GPU 사용량을 측정하고, 모델 평가 과정에서 성능 지표를 기반으로 연속학습 안정성과 적응성을 측정했다. 분석 결과, 연속학습을 적용하면 GPU 사용량이 크게 증가하지 않으면서, 과거 기억을 보존하고 동시에 새로운 데이터에 대한 적응 능력이 개선되었다. 향후 연구로는 자율주행 환경에 맞는 대용량 스트리밍 데이터에 적합한 연속학습 모델을 개발할 예정이다.

ACKNOWLEDGMENT

This work is supported by the United States Office of Naval Research Global (N62909-25-1-2082)

참 고 문 헌

[1] Liyuan Wang, Xingxing Zhang, Hang Su and Jun Zhu, "A Comprehensive Survey of Continual Learning: Theory, Method and Application", in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 46, no.8, pp. 5362-5383, Aug. 2024

[2] Botos Csaba1, Wenxuan Zhang, Matthias Müller, Ser-Nam Lim, Mohamed Elhoseiny, Philip H.S. Torr1, and Adel Bibi1, "Label Delay in Online Continual Learning, 2024(https://arxiv.org/pdf/2312.00923).

[3] A. -P. Ta, "Factorization machines with follow-the-regularized-leader for CTR prediction in display advertising," 2015 IEEE International Conference on Big Data (Big Data), Santa Clara, CA, USA, 2015, pp. 2889-2891, doi: 10.1109/BigData.2015.7364112.

ChromaTag: Delta E 색차 제어 기반의 투명 마커 포즈 추정

박유진, 김은경, 박천음*
국립한밭대학교
20191906@edu.hanbat.ac.kr, ekim@hanbat.ac.kr, parkce@hanbat.ac.kr

ChromaTag: Delta E Chromatic Difference Control Based on Transparent Marker Pose Estimation

Yu Jin Park, Eunkyung Kim, Cheoneum Park*
Hanbat National University

요 약

본 연구는 흑백 ArUco 마커를 인간의 시각적 방해를 최소화하는 투명 ChromaTag로 제작하고, 이를 이용해 6-DOF 포즈를 추정하는 프레임워크를 제안한다. 제안 방법은 인간 시각 인지 모델인 ΔE 색차를 기반으로 투명 마커를 합성하고, YOLO-Pose 모델을 통해 마커의 4개 코너점을 직접 검출하여 포즈를 추정한다. 모델 학습을 위해, 기존 A4 문서 YOLO 데이터셋 기반으로 총 12,710개의 학습데이터를 구축하며, ΔE 난이도에 따른 가중 샘플링을 적용한다. 실험 결과, 제안 방법은 포즈 추정 태스크에서 mAP@[.5:.95] 기준 0.99 성능을 달성하였다. 동일한 실험 조건에서 기존 ArUco 마커의 OKS mAP@[.5:.95]가 0.984를 기록한 것과 비교할 때, ChromaTag는 낮은 ΔE 조건에서도 포즈 추정 정확도와 키포인트 검출 안정성을 유지하여 투명 마커 시스템의 실용적 가능성을 입증하였다.

Ⅰ. 서 론

증강현실(AR)과 로보틱스 분야에서 객체의 정확한 6-DOF 포즈 추정은 널리 사용하는 기술이다. 기존의 ArUco나 AprilTag 같은 가시적 마커는 높은 인식률을 제공하지만, 뚜렷한 흑백 패턴으로 인해 시각적 방해를 유발하고 한계가 있다[1, 2]. 이러한 문제를 해결하기 위해 투명 마커에 대한 연구가 활발히 진행되고 있다. InfraredTags [3]는 적외선 형광 물질을 활용한 방법을 제시했으나, 특수한 IR 카메라와 필터링 장비가 필요하다. iMarkers [4]는 자외선 기반 방법을 제안했지만 특수 장비에 의존성이 존재한다. 이러한 접근법들은 추가 장비로 인한 비용과 확장성에 대한 어려움이 있다. 이에 본 연구는 별도의 하드웨어 없이 표준 RGB 카메라 환경에서 인간의 시각 인지 모델을 역이용하는 ChromaTag 프레임워크([6]과 다름)를 제안한다.

Ⅱ. 본론

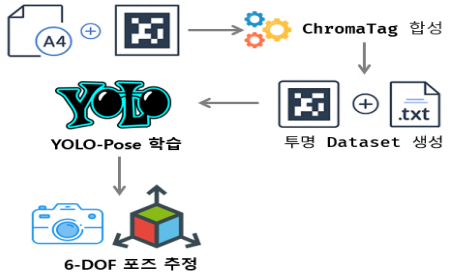


Figure 1. ChromaTag 프레임워크 파이프라인 개요

본 논문에서 제안하는 ChromaTag 프레임워크는 [Fig.1]과 같다. 이는 인간의 시각 시스템이 특정 색차(ΔE) 변화에 둔감하다는 점을 이용, 사람이 인지하기 어려운 색상 신호를 마커에 인코딩하고 YOLO-Pose 모델을 통해 검출 및 포즈를 추정하는 방식이다.

본 프레임워크는 인간의 시각 인지 모델과 유사한 CIE 2000 ΔE 색차 공식을 활용한다. 이 공식은 밝기(L)와 두 색상 채널(a, b)을 분리하는 CIELAB 색 공간을 기반으로 한다. 인간의 눈은 밝기(L)의 변화에는 매우 민감하지만, 색상(a, b)의 미세한 변화에는 상대적으로 둔감하다[5].

이를 기반으로, 제안 방법은 마커가 합성될 영역의 L 값을 원본 배경과 유사하도록 유지하면서, 마커의 패턴 정보를 인간이 잘 인지하지 못하는 a와 b 채널에 인코딩한다. 최종 픽셀 색상은 사용자가 설정한 목표 target ΔE (예: 2.0)와 A+C 필드(색차 분배 지도)를 결합하는 것으로 결정된다. A와 C 필드는 다음과 같다.

- A-Field (저주파 위장): 마커의 전반적인 영역을 담당한다. 가우시안 차분(Difference of Gaussians, DoG) 필터를 사용하여 마커의 날카로운 경계선(고주파)을 제거하고, 부드러운 저주파 구조만 남긴다. 이를 통해 마커의 존재감을 최소화하는 위장 효과를 극대화한다.
- C-Field (코너 강화): 모델이 코너를 찾도록 보조 신호를 코너에 추가하는 것으로, 특정 방향성을 가진 미세한 줄무늬 패턴인 가보르(Gabor) 필터를 네 개의 코너에 추가한다. 사람이 인지하기 어렵지만, 모델은 학습을 통해 코너의 위치 예측이 가능하게 된다.

각 픽셀에 적용될 실제 목표 색차 local ΔE는 target ΔE × A+C Field (x,y)로 계산되며, 이 값을 만족하도록 a*, b* 채널을 미세 조정하여 최종 색상을 결정한다.

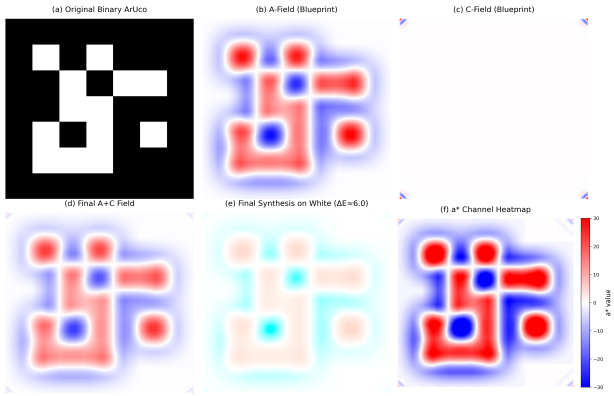


Figure 2. ChromaTag A+C 변조 필드 합성 (a) ArUco 마커. (b) A-Field. (c) C-Field. (d) A+C 필드. (e) 합성 (ΔE≈6.0), (f) 결과물의 a* 채널 히트맵 (시각화를 위해 신호 세기는 실제보다 강조됨)

[Fig.2]는 ChromaTag의 A+C 변조 과정을 시각화한다: (a)의 원본 마커로부터 (b) DoG 기반 A-Field와 (c) 가보르 기반 C-Field를 결합하여 (d) 최종 변조 필드를 생성하며, 이를 통해 (e) 합성된 결과는 색상 채널에 (f) 명확한 신호를 인코딩하면서도 시각적으로 인지하기 어렵다.

생성된 변조 필드(d)를 깨끗한 배경에 적용한 최종 합성 결과물(e)은 원본의 명확한 형태가 사라지고 배경과 자연스럽게 융화된 패턴을 보인다. 마지막으로 a* 채널 히트맵(f)은 해당 패턴이 시각적으로는 인지하기 어려우나, 색상 채널에는 데이터로서 명확히 인코딩되었음을 나타낸다. 공정한 성능 비교를 위해 동일한 학습 파이프라인에서 ArUco와 직접 비교 실험을 수행하였다.

A4 문서 YOLO 데이터셋 기반으로 ΔE 난이도별 가중 샘플링을 적용하여 총 12,710개 합성 데이터를 구축하였다. 최적화된 DoG/Gabor 파라미터로 A+C 변조를 수행하였다. 대용량 데이터셋의 실시간 ΔE 계산을 위해 colormath 라이브러리의 CPU 기반 연산을 GPU 병렬 처리로 최적화하여 처리 속도 향상을 달성하였다.

모든 데이터는 마커 당 4개의 키포인트를 갖는 YOLO Pose 형식으로 레이블링 된다. Yolo11m-pose 모델을 100 epoch 학습하여 테스트 데이터셋으로 검증한 결과는 [Table 1],[Table 2]와 같다. 성능 검증을 위해 동일한 YOLO11m-pose 파이프라인에서 ArUco와 직접 비교하였고, ChromaTag는 OKS mAP@[.5:.95] 0.995 vs 0.984로 우수한 성능을 보였다. 특히 ChromaTag는 모든 ΔE 구간에서 안정적인 정밀도/재현율을 유지하여 투명 마커의 강건성을 확인하였다.

방법	ΔE 범위	Pose OKS mAP@[5:.95]	Precision	Recall
Ours	ΔE 1~2 (Visible)	0.995	~1.0	~1.0
Ours	ΔE 2~3 (Subtle)	0.992	~1.0	~1.0
Ours	ΔE 3~4 (Invisible)	0.985	~1.0	~1.0
ArUco Marker		0.984	~1.0	0.973

Table 1. ChromaTag vs ArUco 성능 비교(YOLO11m-pose, 동일 조건)

[Fig.3]은 제안 모델의 결과를 시각적으로 보여준다. ΔE가 낮아져 육안 식별이 어려운 조건(좌측 열)에서도 2D 코너 검출(2행)이 안정적이며, 이를 기반으로 정확한 6-DOF 포즈 추정(3행)이 가능함을 확인할 수 있다.

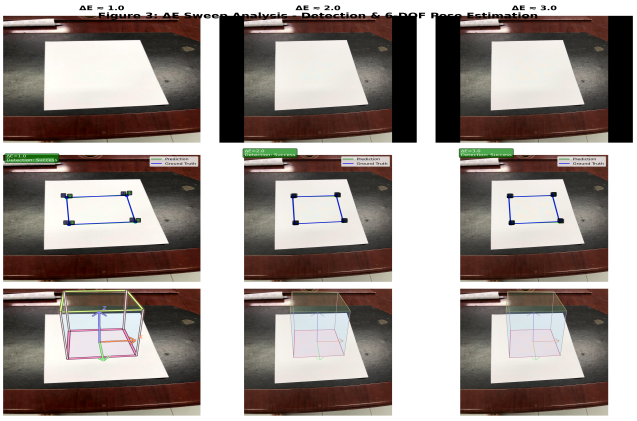


Figure 3. ΔE 난이도에 따른 테스트 데이터셋 기반 ChromaTag의 시각적 은닉성 및 단계별 포즈 추정 결과

Ⅲ. 결론

본 연구는 ΔE 색차 제어와 A+C 변조를 결합하여 표준 RGB 카메라만으로 동작하는 투명 마커 시스템 ChromaTag를 제안하고, 그 가능성을 보였다. A4 합성 데이터셋 기반 실험에서 달성한 약 0.99의 높은 포즈 mAP 성능은, 특수 장비 없이도 포즈 추정이 가능함을 보인다. 이에 따라, 본 논문에서 제안한 ChromaTag는 사용자 경험을 해치지 않는 비침해 AR을 구현할 수 있으며, 물리적 객체에 정밀한 3D 디지털 정보를 증강시키 것으로 상호작용 시스템 발전에 기여할 것이라 사료된다.

참 고 문 헌

[1] Garrido-Jurado, S. et al., “Automatic generation and detection of highly reliable fiducial markers under occlusion,” Pattern Recognition, 47 (6), 2280-2292, 2014. DOI: 10.1016/j.patcog.2014.01.005.

[2] E. Olson, "AprilTag: A robust and flexible visual fiducial system," 2011 IEEE International Conference on Robotics and Automation, Shanghai, China, 2011, pp. 3400-3407, doi: 10.1109/ICRA.2011.5979561.

[3] Mustafa Doga Dogan, Ahmad Taka, Michael Lu, Yunyi Zhu, Akshat Kumar, Aakar Gupta, Stefanie Mueller, CHI 2022, DOI: 10.1145/3491102.3501951.

[4] Ali Tourani et al., “Unveiling the Potential of iMarkers: Invisible Fiducial Markers for Advanced Robotics,” arXiv:2501.15505 (2025).

[5] Mullen KT, Plant GT. Colour and luminance vision in human optic neuritis. Brain. 1986 Feb;109 (Pt 1):1-13. doi: 10.1093/brain/109.1.1. PMID: 3942852.

[6] DeGol, J., Bretl, T., & Hoiem, D. (2017). Chromatag: A colored marker and fast detection algorithm. In Proceedings of the IEEE international conference on computer vision (pp. 1472-1481).

이미지 텍스처 특징에 기반한 센싱 데이터 이상탐지 연구

최용훈, 김민서, 김민석*
상명대학교 휴먼지능로봇공학과
choiyonghoon1103@gmail.com, *minsuk.kim@smu.ac.kr

Image Texture Feature-Driven Anomaly Detection for Sensing Data

Yong-hoon Choi, Min-seo Kim, Min-suk Kim*
Dept. of Human Intelligence & Robot Engineering. Sangmyung University

요 약

고위험 작업 공간 및 시설에서는 센서 데이터 기반의 상황 모니터링이 유지보수와 안전 확보를 위해 필수적이다. 특히 발전소와 같이 위험도가 높은 시설에서는 이상 징후를 신속하게 탐지하고 대응하는 것이 중요하다. 그러나 인력에 의존한 수동 모니터링은 탐지 지연과 실패 가능성을 내포하며, 이는 심각한 사고로 이어질 수 있다. 따라서 본 연구에서는 GADF(Gramian Angular Difference Field) 변환을 통해 센서 데이터를 이미지 형태로 변환하여 기계학습 모델의 입력으로 사용하며, 정상 패턴 학습을 위해 생성형 모델 DDPM(Denoising Diffusion Probabilistic Model)에 기반한 DDAD(Denoising Diffusion Anomaly Detection)기법을 적용한다. 또한, 변환된 GADF 이미지를 텍스처 형태로 간주하고 이를 효과적으로 처리하기 위해 CAPTN(Chebyshev Attention Depth Permutation Texture Network) 모델을 도입한다. 제안된 시스템은 환경 변화에 대한 추가적인 적응 과정 없이 일관된 이상 탐지 성능을 제공하는 것을 목표로 하며 다양한 환경에서의 센서데이터로 구성된 KAERI(한국원자력연구원)에서 수집한 데이터세트를 통해 이를 검증한다.

Ⅰ. 서 론

산업 현장을 비롯하여 고위험 작업 공간 또는 시설의 경우, 센서 데이터에 기반한 상황 모니터링 기술은 시설 유지보수와 작업자 안전을 위해 필수적으로 도입되어야 한다. 특히 발전소와 같이 위험도가 높은 시설에서는 이상 징후를 신속하게 파악하고 분석하여 적절한 조치를 취하는 것이 매우 중요하다. 그러나 인력에 의존하여 수동으로 수행되는 상황 모니터링 방식은 다양한 요인으로 인해 이상 탐지가 지연되거나 실패할 가능성이 존재하기 때문에 시스템의 일관성과 신뢰성을 보장하지 못한다. 이러한 한계점은 고위험 시설에서 치명적인 결과로 이어질 수 있으므로, 센서 데이터 기반의 자동화된 이상 탐지 연구가 요구된다.

본 논문에서는 이상 탐지 기법에 대한 검증을 위해 KAERI 에서 수집한 데이터를 기반으로 이상 탐지를 수행한다. 해당 데이터는 다양한 환경에서 발생한 이상 징후를 포함하고 있기 때문에 제안하고자 하는 이상 탐지 기법의 일관성과 신뢰성을 검증하는데 적합하다. 이를 위해 여러 환경에서 일관적인 이상 탐지를 구현하고자 센서 데이터를 GADF(Gramian Angular Difference Field) 방식으로 이미지화하여 기계학습 모델의 입력으로 사용한다. 따라서 GADF 이미지 세트에서의 정상 특징을 포착하기 위하여 본 논문에서는 생성형 모델인 DDPM(Denoising Diffusion Probabilistic Model)[1]에 기반한 이상 탐지를 제시한 DDAD(Denoising Diffusion Anomaly Detection)[2] 기법을 사용하여 다양한 형태의 변환 데이터로부터 정상 패턴을 학습한다. 또한, GADF 이미지로 변환된 센서 데이터가 텍스처적 특성을 가진다는 가정하에, 이를 효과적으로 처리하기 위해 CAPTN(Chebyshev Attention Depth Permutation Texture Network)[3] 모델을 도입하였다. 이를 통해 본 논문에서는 텍스처 모델의 도입을 통해 환경 변화에 대한 별도의 과정 없이 일관된 성능을 보이는 이상 탐지 시스템을 목표로 한다.

Ⅱ. GADF 이미지 기반 이상 탐지

2.1. 데이터세트 분석 및 처리

본 논문에서는 A, B, C 로 정의된 환경에서 수집된 센서 데이터로 구성된 데이터세트를 활용하여 이상 탐지를 수행한다. 해당 데이터셋은 20kHz 에서 100kHz 범위의 주파수 영역에서 측정된 초음파 신호로 이루어져 있으며, 미세한 이상 징후를 탐지하기 위하여 4ms 구간에서 0.25kHz 간격으로 푸리에 변환을 수행하였다. 이를 통해 각 주파수 대역에서의 대표 값이 추출되며, 최종적으로 길이 320 의 데이터 형태를 가지게 된다.

본 연구에서는 수집된 해당 데이터를 GADF 변환을 통해 이미지 형태로 변환하여 입력으로 사용하였고 변환된 데이터는 3x3 크기의 RGB 이미지 형태가 된다. 또한, 다양한 이미지 데이터로부터 공통적인 정상 패턴을 효과적으로 추출하기 위해 텍스처 기반 모델을 도입하였다. 이를 위해 학습 과정에서는 클래스 구분을 두지 않고 전체 데이터를 통합하여 학습을 수행하였으며, 그 결과를 기반으로 이상 탐지 성능을 평가하였다.

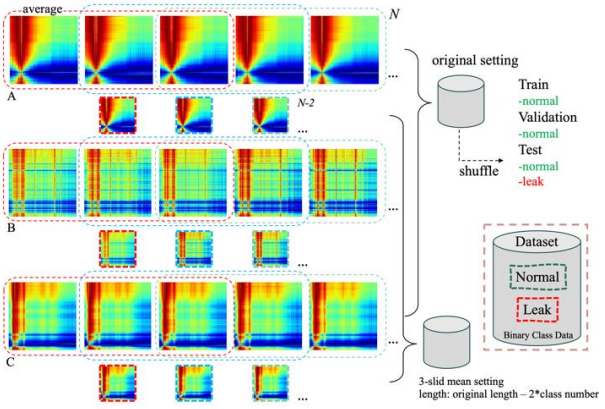


Figure 1. KAERI data formats and preprocessing workflow

또한, 상이한 데이터 간의 노이즈를 감소시키고 동일 클래스 내 데이터의 분산을 줄이기 위해, 각 이미지와 그 앞뒤 한 장을 포함하여 평균화하는 전처리 과정을 적용하였으며, 이를 통해 해당 기법의 효과를 검증하였다. 실험에 사용한 데이터세트의 구조와 처리 과정은 그림 1 을 통하여 확인 가능하다.

2.2. Proposed Method: DDAD (Denoising Diffusion Anomaly Detection)

DDAD 는 Denoising Diffusion 모델을 활용한 이상 탐지 제안 방법이며 DDPM 의 확률론적 노이즈화, 복원 구조를 기반으로 정상 데이터의 분포를 학습하고 이상 점수를 산출하는 기법이다. 기존 DDPM 은 전방 확산 과정에서 데이터에 점진적인 가우시안 노이즈를 주입하여 완전한 노이즈로 변환한 뒤, 역방향 복원 과정을 통해 데이터를 학습하는 모델이다. DDAD 는 DDPM 구조를 변형하여, 역방향 복원 시 입력 데이터 자체를 참조하는 조건부 디노이징 과정을 통해 학습된 정상 패턴은 유지되고 이상 영역은 제거되어 생성하도록 유도하게 된다. 따라서 입력 이미지와 복원 이미지의 차이를 기반으로 이상 점수를 계산한다. 또한 사전 학습된 네트워크를 중간 계층 출력을 이용한 특징 공간 거리 이상 점수를 통합하여 정밀한 이상 탐지를 수행한다. 하지만 기존의 ImageNet 와 같은 오브젝트 이미지셋을 학습한 사전학습 모델의 경우, 전체적으로 반복되는 패턴이 존재하는 형태의 텍스처 이미지인 실험 데이터에는 적합하지 않다. 따라서 본 연구에서는 DDAD 에 CAPTN 을 결합하여 텍스처 정보를 효과적으로 보존함으로써 다양한 환경에서의 정상 패턴을 효과적으로 포착하는 이상 탐지를 수행한다.

2.3. Proposed Method: CAPTN (Chebyshev Attention Depth Permutation Texture Network)

기존의 오브젝트 이미지의 경우 전역 형태를 기반으로 해당 객체를 인식하는 정보로 사용할 수 있으며, 합성곱 네트워크들을 이용하여 높은 성능을 달성하였다. 하지만 KAERI 데이터와 같은 형태의 텍스처 이미지의 경우 전역적인 형태가 아닌 특정 패턴이 가변성을 가지며 반복되는 통계적인 특성을 가지고 있음으로 이를 처리할 수 있는 모델이 요구된다. CAPTN 은 텍스처 특성이 중요한 데이터에 특화된 신경망 구조로서, 텍스처 내 복잡한 공간적 및 채널 간 관계를 효과적으로 고려하도록 설계되었다. 해당 모델의 경우, 체비쇼프 다항식 기반의 스펙트럼 그래프 컨볼루션을 활용하여 입력 이미지의 텍스처 패턴을 그래프 신호로 해석하고 국소 및 전역적 의존성을 동시에 반영한다. 또한, 순열 모듈을 통하여 특징 채널과 공간 차원을 재배열하여 텍스처내 잠재적 상관관계를 높이고 모델의 표현력을 극대화한다. 따라서 본 논문에서는 CAPTN 를 이용하여 DDAD 로 재구성된 이미지의 텍스처 특징을 추출하여 특징 공간에서의 이상 탐지 성능을 보강한다.

Ⅲ. 실험

본 논문에서는 두가지 형태의 데이터세트를 학습하고 이에 대한 성능 검증을 진행하였다. CAPTN 의 경우 사전 학습된 convnext nano 모델에 기반하여 체비쇼프 다항식 모듈을 학습하고, 학습된 CAPTN 을 이용하여 데이터의 특징을 추출하여 해당 특징 공간에서의 거리를 기반으로 이상 탐지를 수행한다. 이에 대한 실험 결과는 표 1 과 같다.

Table 1. Dataset and Model Configurations Summary (AUROC)

class	A	B	C	Binary
DDAD	70.4%	89.4%	79.6%	74.3%
DDAD + 3slid	72.4%	97.6%	76.1%	71.5%
DDAD + CATPN	70.0%	88.5%	74.5%	73.8%
DDAD + CAPTN + 3lid	79.0%	97.1%	71.5%	80.1%

실험 결과, 기존 WideResNet 기반 DDAD 모델은 평균화된 데이터 세팅에서 일부 클래스의 성능이 향상되었으나, 전체 클래스에 대한 이상 탐지 성능은 오히려 저하되어 전반으로 낮은 성능을 보였다. 이는 해당 모델이 전체 데이터의 공통적인 특징을 효과적으로 포착하는 데 한계가 있음을 시사한다. 반면, CAPTN 기반 DDAD 모델의 경우 평균화 전처리 데이터를 대상으로 일부 클래스의 성능이 감소하였음에도 불구하고 전체 데이터를 대상으로 한 이상 탐지 성능이 향상되었으며 이는 가장 높은 성능을 기록하였다. 이를 통해 CAPTN 모델과 본 논문에서 제안한 데이터 처리 방식의 유효성을 검증 가능하였다.

Ⅲ. 결론

본 논문에서는 고위험 작업 환경에서의 신뢰성 있는 모니터링 기술을 목표로, GADF 이미지 형식의 센서 데이터에 기반한 이상 탐지 기법을 제안하였다. 정상 패턴 학습을 위해 DDPM 기반의 DDAD 구조를 적용하여 이상 탐지를 수행하였으며 텍스처 특징을 효과적으로 처리하기 위해 CAPTN 모델을 도입하여 성능을 검증하였다. 또한 텍스처 특성을 강화하기 위해 평균화 전략을 적용한 결과, 높은 AUROC 성능을 달성할 수 있었으며, 이를 통해 CAPTN 모델이 데이터의 공통된 텍스처 특징을 효과적으로 포착함을 확인하였다. 결과적으로, 본 연구는 이미지 형태로 변환된 센서 데이터를 활용한 이상 탐지 방법론을 제시하였으며, 기존 모델을 기반으로 효율적인 이상 탐지 성능 향상에 기여할 수 있을 것으로 기대된다.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government. (MIST) (No. RS-2022-00144000)

참 고 문 헌

[1] Ho, J., Jain, A., Abbeel, P. "Denoising diffusion probabilistic models," Advances in Neural Information Processing Systems (NeurIPS), vol. 33, pp. 6840-6851, 2020.

[2] Mousakhan, A., Brox, T., Tayyub, J. "Anomaly detection with conditioned denoising diffusion models," DAGM German Conference on Pattern Recognition, Springer Nature Switzerland, 2024.

[3] Evani, R., Rajan, D., Mao, S. "Chebyshev Attention Depth Permutation Texture Network with Latent Texture Attribute Loss," Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR), 2025.

인공지능 기반 외국인 노동자 근로 계약서 번역 및 법률 정보 제공 앱

정민성, 이수현, 육종범, 옥주용, 강윤서, *황경호
국립한밭대학교 컴퓨터공학과
jungminsung@edu.hanbat.ac.kr, SuHyeonLee@edu.hanbat.ac.kr, yukjb5154@edu.hanbat.ac.kr,
ockjuyong@edu.hanbat.ac.kr, YunseoKang@edu.hanbat.ac.kr, *gabriel@hanbat.ac.kr

AI-based Employment Contract Translation and Legal Information App for Foreign Workers

Min-Sung Jung, Su-Hyeon Lee, Jong-Beom Yuk, Ju-Yong Ok,
Yun-Seo Kang, *Gyung-Ho Hwang
Dept. Computer Engineering, Hanbat National University

요 약

본 논문은 외국인 노동자의 권리 보호와 고용주-노동자 간 분쟁 예방을 목표로, 인공지능 기술을 접목한 근로계약서 번역 및 법률 정보 제공 플랫폼 “Legal Guide”의 설계와 구현을 다루었다. 외국인 노동자 증가와 이에 따른 분쟁 증가 문제를 해결하기 위하여, 본 플랫폼은 Optical Character Recognition (OCR) 기반 계약서 텍스트 인식, 다국어 번역, Generative Pre-trained Transformer (GPT) 기반 법률 조항 분석, 생활 언어 변환, 국가 법령정보센터 공공데이터 애플리케이션 프로그래밍 인터페이스 (API) 연계 기능을 제공하였다. 사용자는 근로계약서를 업로드 또는 촬영하여 실시간으로 번역과 법률 분석을 받을 수 있다. 본 연구는 인공지능 기술이 법률 서비스에 접목되어 외국인 노동자의 권리 보장과 사회적 신뢰도 향상에 기여할 수 있음을 보여주었다.

Ⅰ. 서 론

최근 통계에 따르면, 장기 체류 외국인 노동자는 10 년 전 대비 약 6% 증가하였으며, 이와 관련된 노동 분쟁 역시 증가 추세를 보이고 있었다[1]. 분쟁의 주요 원인은 언어 장벽, 법률 정보 접근성 부족, 그리고 표준화되지 않은 계약서 작성 관행이었다. 특히 한국어 실력이 미흡한 외국인 노동자는 불리한 조건의 계약을 체결하거나, 계약서 자체가 작성되지 않아 임금 체불이나 불법체류로 이어지는 사례가 발생하였다[2].

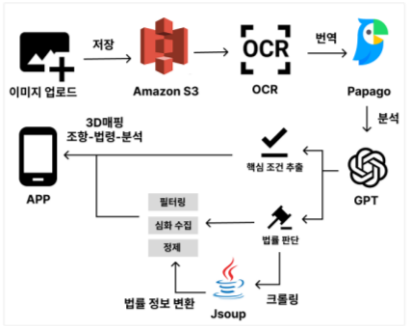
기존 법률 서비스나 AI 기반 법률 플랫폼은 대체로 일반 대중을 대상으로 하거나, 법률 검색 및 판례 예측에 치우쳐 있어, 언어 취약계층을 위한 실무형 계약 분석이나 다국어 번역 기능은 부족한 실정이다. 예를 들어, 미국의 DoNotPay 는 자동 청구소송에 강점을 보이나, 번역·계약서 분석 등에는 한계가 있으며, 한국 법률구조공단의 AI 상담도 제한적인 한국어 기반 상담에 머무르고 있다[3]. 기존 연구들 또한 GPT 와 LLM 의 법률 자문 가능성, 사용자 신뢰도 등을 분석했으나, 실질적인 다국어 기반 계약 분석 시스템 구축에는 미치지 못하였다[4].

이에 본 연구는 외국인 노동자의 언어 장벽과 법률 정보 격차 문제를 실질적으로 해결하기 위해, 인공지능 기술을 활용하여 계약서 번역과 법률 분석을 실시간으로 제공하는 실무형 플랫폼“Legal Guide”를 개발하였다. 본 플랫폼은 고용주와 노동자가 동일한 이해를 바탕으로 계약을 체결하도록 돕고, 법률 정보 접근성을 향상시키는 것을 목표로 하였다.

Ⅱ. 본론

본 플랫폼은 OCR 기반 텍스트 인식, 자동 번역 및 번역 이미지 생성, GPT 기반 법률 요소 분석, 생활 언어 변환, 그리고 법령 정보 연계로 구성된 순차적인 절차로 구성되었다.

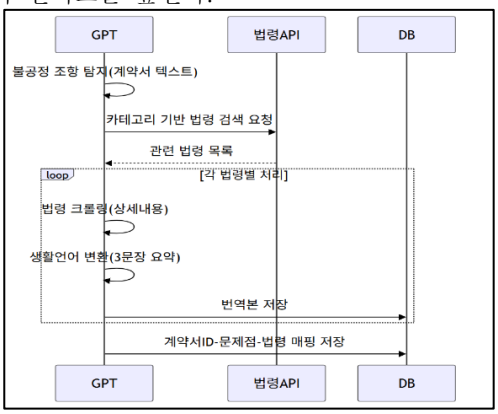
사용자는 모바일 애플리케이션을 통해 근로계약서를 업로드하거나 카메라로 촬영할 수 있으며, 업로드된 이미지 파일은 Amazon Simple Storage Service (Amazon S3) 외부 스토리지에 저장되고, 백엔드에서 해당 이미지 경로와 메타데이터를 관리하였다. 저장된 계약서 이미지는 먼저 OCR 모듈을 거쳐 계약서 내의 문자를 정확하게 추출하였다. 이후 사용자 프로파일 기반의 언어 설정에 따라 자동으로 번역 대상 언어가 선택되며, Papago API 를 활용한 문맥 기반 번역이 수행된다. 번역 결과는 단순 텍스트로 저장되는 것이 아니라, Papago 이미지 합성 기술을 통해 원본 계약서 이미지 위에 번역 내용을 시각적으로 합성하여 직관적인 이해를. 이러한 전체 동작 과정을 도식화한 시스템 설계도는 [그림 1]과 같다.



[그림 1] 시스템 설계도

번역이 완료된 계약서 텍스트는 AI 기반 불공정 조항 탐지 및 법령 분석 단계로 넘어가며, 이 과정에서 OpenAI GPT-4o API 를 통해 임금, 근로시간, 휴가, 해고 사유 등 핵심 근로 조건이 추출되고, 관련

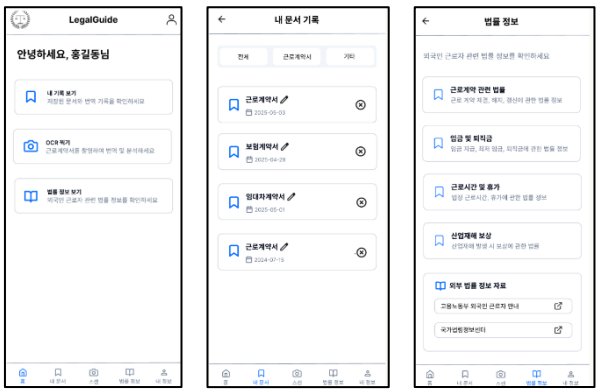
노동법과 비교하여 불리한 조건의 존재 여부가 자동으로 판단된다. 탐지된 조항은 비전문가도 이해할 수 있도록 GPT 를 활용해 생활 언어로 변환되며, 이후 국가법령정보센터 API 와 Jsoup 기반 크롤링을 통해 수집된 법령 정보와 연계된다. 법령 정보는 (1) 필터링을 통한 기본 정보 수집, (2) 심화 수집을 통한 법령 전문 확보, (3) 정제·변환 단계를 거친 생활 언어 변환의 세 과정으로 처리되며, 이 과정에서 GPT 의 요약 기능을 활용하여 긴 법령 조항도 쉽게 이해 가능한 형태로 제공된다. 이러한 연계를 통해 사용자는 계약서 내용과 관련 법령을 동시에 열람할 수 있으며, 계약서 ID 를 중심으로 조항-법령-GPT 분석 결과 간의 3D 매핑 구조가 구현되어 분석 근거 데이터 또한 함께 저장된다. 이는 단순히 결과를 보여주는 수준을 넘어, 법적 판단의 “증거 기반”을 함께 제시함으로써 사용자 신뢰도를 높인다.



[그림 2] 법령 분석 및 생활 언어 변환 절차도

[그림 2]는 [그림 1]의 시스템 개요에서 특히 불공정 조항 탐지 이후의 세부 절차를 나타낸다. GPT 가 계약서 내 잠재적으로 불리한 조항을 탐지하면, 법령 API 를 통해 관련 법령 목록을 요청·수신하고, 각 법령은 크롤링을 거쳐 상세 조항이 확보된다. 확보된 법령은 GPT 기반 요약 과정을 거쳐 생활 언어로 변환되며, 최종적으로 번역본과 함께 DB 에 저장된다. 이때 계약서 ID 와 문제 조항, 관련 법령이 함께 매핑되어 3D 구조로 관리됨으로써, 분석 결과와 그 근거가 동시에 제공된다. 정확도와 안정성 확보를 위해 JSON 고정 출력 포맷을 적용하여 파싱 오류를 최소화하고, 법률 용어 사전 기반 프롬프트 설계로 GPT 응답의 일관성을 강화하였다. 오류 발생 시 최대 3 회까지 재시도하는 로직을 추가하여 시스템 안정성을 높였으며, 사용자 프로파일 기반 언어 매핑과 Fallback 메커니즘을 통해 언어 인식 실패 상황에서도 번역의 신뢰도를 보장하였다. 이러한 다층적 보완 전략은 기존 AI 기반 번역 시스템이 지적받아온 “번역 품질의 불안정성” 문제를 해결하는 데 중요한 기여를 한다.

다음으로 본 플랫폼“Legal Guide”의 메인 홈 화면은 사용자 진입점으로, 개인화된 인사 메시지와 함께 "내 기록 보기", "OCR 찍기", "법률 정보 보기"의 세 가지 핵심 메뉴를 제공하였다. "내 기록 보기"에서는 이전에 업로드한 계약서를 날짜별로 조회 및 재분석할 수 있고, 불필요한 문서는 삭제 가능하였다. "OCR 찍기"는 촬영 또는 업로드를 통한 즉시 분석을 지원하였으며, "법률 정보 보기"는 관련 자료로 바로 이동하였다.



[그림 3] Legal Guide 메인 및 법률 정보 페이지

또한 사용자는 메인 홈 화면에서 "OCR 찍기" 메뉴를 선택하여 근로계약서 분석을 시작할 수 있었다. 첫 단계인 "근로계약서 스캔" 화면에서 문서를 카메라로 촬영하거나 파일을 업로드하면, 시스템은 즉시 OCR 을 통해 텍스트를 추출하고 개인정보를 자동 마스킹 처리하였다. 이어서 "번역 결과" 화면으로 전환되었으며, 사용자는 원문, 번역문, 전체 보기 중 원하는 탭을 선택하여 내용을 확인할 수 있었다. 번역 과정은 Naver Papago API 를 기반으로 하였으며, 사용자는 하단의 " 법률 분석 보기" 버튼을 눌러 GPT 기반의 심화 법률 분석 결과 페이지로 이동할 수 있었다.

[그림 4] OCR 및 번역 화면

Ⅲ. 결론

본 연구에서 개발한 "Legal Guide" 앱은 외국인 노동자의 계약 이해도와 법률 정보 접근성을 향상시켜, 고용주-노동자 간 불필요한 분쟁을 예방하는 데 기여하였다. 생성형 인공지능 기술을 활용해 계약서 내용을 실시간 번역·분석하고, 법률 정보를 생활 언어로 변환하여 제공함으로써 외국인 노동자의 권리 보호와 사회적 신뢰도 향상에 실질적인 도움을 주었다. 향후 본 플랫폼은 지원 언어를 확대하고, 변호사 상담 연계 기능과 다양한 계약서 유형으로의 확장을 통해 법률 서비스 범위와 활용성을 더욱 넓혀 나갈 계획이다.

참 고 문 헌

[1] 법무부, “출입국·외국인정책 통계연보”, 2024.
DOI: https://www.immigration.go.kr/
[2] “일했는데 빚더미 앉는 이주노동자…월급 뜯겼는데 되레 소송도,” 한겨레, 2024 년 4 월 30 일.
DOI: https://www.hani.co.kr/
[3] Joshua Browder, “DoNotPay: The World's First Robot Lawyer,”
DOI: https://donotpay.com/
[4] H. Li, Y. Chen, Q. Ai, Y. Wu, R. Zhang, Y. Liu, “LexEval: A Comprehensive Chinese Legal Benchmark for Evaluating Large Language Models”,
arXiv preprint arXiv:2409.20288, 2024.
DOI: https://arxiv.org/abs/2409.20288

온 프레미스(On-Premise) 환경에서의 MCP(Model Context Protocol)를 활용한 범용 Text-to-SQL 에이전트 설계 및 구현에 관한 연구

김지섭[1], 김동우[1], 이재희[2]

[1]동서울대학교 컴퓨터 소프트웨어과 학부생

[2]동서울대학교 컴퓨터 소프트웨어과 교수

wltjq6702@gmail.com, gimdongu19@gmail.com, lih7314@du.ac.kr

A Study on the Design and Implementation of an On-Premise General-Purpose Text-to-SQL Agent based on the MCP(Model Context Protocol)

Ji-Seob Kim[1], Dong-Woo Kim[1], Jae-Hee Lee[2]

[1] [2]Dept. of Computer Software, Dong-Seoul University

요 약

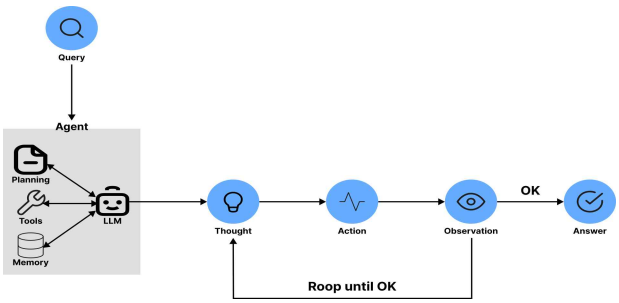
본 연구는 온 프레미스(On-Premise) 환경에서 MCP(Model Context Protocol)를 활용한 범용 Text-to-SQL 에이전트의 설계 및 구현에 관한 것이다. 본 연구에서 제안한 시스템은 직접 구현한 Host-Client-Server 구조로 동작하며 Streamlit 기반의 MCP Host를 통해 LangGraph 기반 LLM-based 에이전트가 MCP Server에 구현된 9개의 표준화된 도구(Tool)를 사용하여 데이터베이스와 상호작용하는 구조이다. 총 4종류의 데이터베이스와 연동실험을 통해 자연어 질의를 성공적으로 처리하는 실험결과를 통해 본 연구의 범용성과 실용성을 검증하였으며, 이를 통해 온 프레미스 환경에서도 안전하고 확장 가능한 범용 Text-to-SQL 에이전트의 구현 가능성을 제시하였다.

Ⅰ. 서 론

비전문가에게 SQL은 데이터 접근의 큰 장벽이다. 이를 해결하기 위한 기존 Text-to-SQL 기술은 특정 데이터베이스에 종속되어 확장성에 한계가 있다. 최근 LLM(Large Language Model)을 활용한 방식은 성능이 뛰어나지만, 고비용 상용 API 의존과 민감 데이터의 외부 전송에 따른 보안 문제를 안고 있다[1]. 본 연구에서는 이러한 한계점들을 극복하기 위해 온 프레미스 환경으로 비용 및 보안 문제를 해결하고 MCP(Model Context Protocol)로 확장성 문제를 해결한 범용 Text-to-SQL 에이전트(Agent)를 제안하였다. 본 연구에서 제안하는 시스템은 Streamlit 기반의 MCP Host, LangGraph 프레임워크를 활용한 LLM-based 에이전트 기반의 MCP Client, 그리고 직접 개발한 9개의 도구(Tool)를 갖춘 MCP Server로 구성된다. 에이전트가 자체적으로 연결된 데이터베이스 종류를 파악하여 자연어로 전달된 질의를 각 데이터베이스 종류에 맞는 SQL언어로 변환하여 각 데이터베이스에 전달하여 안전하게 상호작용하도록 설계 및 구현하였다.

Ⅱ. LLM-based 에이전트

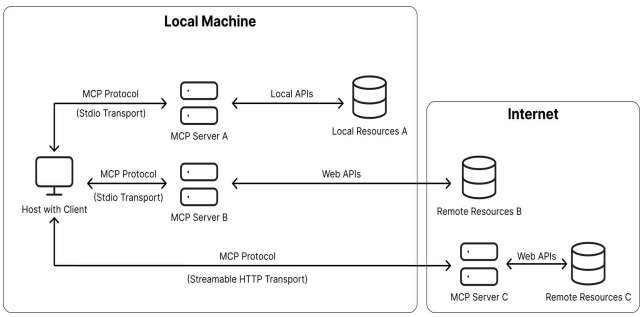
본 연구에서 제안하는 LLM-based 에이전트는 LangGraph의 React(Reasoning and Acting) 프레임워크를 기반으로 추론(Chain of Thought)과 행동(외부 도구 호출)을 교차 생성하도록 유도하였다[2]. 에이전트는 자연어 질의에 대해 생각(Thought)-행동(Action)-관찰(Observation) 단계를 반복하면서 질의를 해결한다. LLM이 추론 순위를 지정하여 행동 계획을 업데이트 하고, 행동을 통해 API 호출이나 데이터베이스 조회와 같은 외부 정보에 접근함으로써 추가 정보를 얻을 수 있다. 그림 1은 본 연구에서 제안한 LLM-based 에이전트의 구조이다.



[그림 1] LLM(Large Language Model)-based 에이전트의 구조

Ⅲ. MCP (Model Context Protocol)

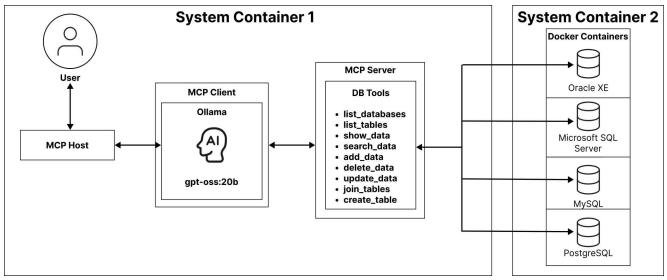
MCP는 2024년 앤트로픽(Anthropic)사에서 제안한 Host-Client-Server 구조로 동작하면서 LLM에 외부 도구와 데이터를 표준화된 방식으로 제공하는 개방형 프로토콜이다[3]. 이는 기존 통합 방식의 확장성 문제를 해결하고 개방적인 에이전트 구축을 지원한다. 그림 2는 본 연구에서 활용하는 MCP의 구조이다.



[그림 2] MCP(Model Context Protocol) 구조도

Ⅳ. 제안한 시스템 구조도

제안한 시스템은 다음과 같이 직접 구현한 Host-Client-Server 구조로 동작한다. 사용자의 요청은 Streamlit 기반의 MCP Host를 통해 OpenAI사에서 공개한 gpt-oss:20b 모델을 활용한 LLM-based 에이전트 기반의 MCP Client에게 전달되며 MCP Client는 MCP Server의 9개의 도구를 호출하여 연결된 4개의 데이터베이스와 상호작용 후 최종 결과를 반환하는 구조이다. 본 연구에서는 도커(Docker)를 이용하여 에이전트 부분과 데이터베이스 부분을 2개의 컨테이너(Container)로 각각 분리하여 구현하였다. 그림 3은 본 연구에서 제안하는 시스템의 전체 구조이다.



[그림 3] 제안한 시스템 구조도

Ⅴ. 실험 환경 및 내용

본 연구의 실험 환경 및 내용은 표 1과 같다.

<표 1> 실험 환경 및 내용

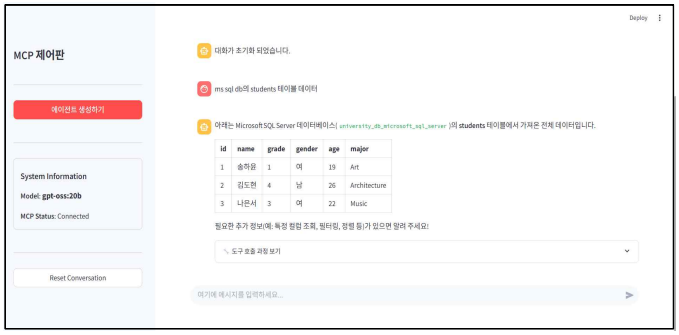
구분	항목	내용
H/W	GPU	NVIDIA V100 32GB
S/W	OS	Debian 12.0 Bookworm
	Language	Python 3.12
	LLM	gpt-oss:20b
	MCP Host	streamlit==1.44.1
	MCP Client	langchain-ollama==0.3.6
		langgraph==0.3.21
	MCP Server	mcp[cli]==1.6.0
		langchain-mcp-adapters ==0.0.7
	Database	Oracle XE(21c), Microsoft SQL Server(2022), MySQL(8.0), PostgreSQL(15)
	Container	Docker 28.3.2

Ⅵ. 제안한 시스템의 실험 결과

제안하는 Text-to-SQL 에이전트의 동작성능을 검증하기 위해 Oracle, Microsoft SQL Server, MySQL, PostgreSQL 등 4개의 데이터베이스 환경에서 에이전트의 Text-to-SQL 동작을 테스트했으며, 그림 4는 Oracle XE DB에 대한 테이블 목록 질의 결과이다.

[그림 4] Oracle XE DB에 테이블 목록 질의 결과

그림 5는 Microsoft SQL Server DB에 대한 테이블 데이터 질의 결과이다.

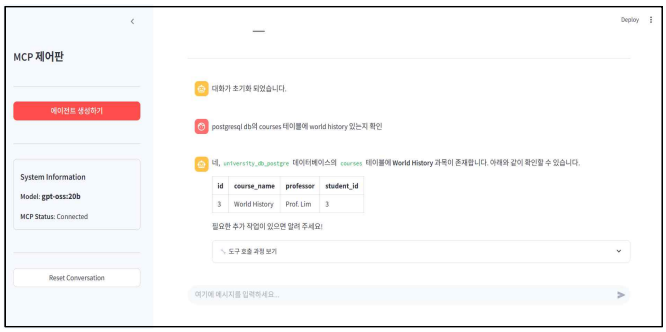


[그림 5] Microsoft SQL Server DB에 테이블 데이터 질의 결과

그림 6은 MySQL DB에 대한 데이터 추가 질의 결과이다.

[그림 6] MySQL DB에 데이터 추가 질의 결과

그림 7은 PostgreSQL DB에 대한 데이터 검색 질의 결과이다.



[그림 7] PostgreSQL DB에 데이터 검색 질의 결과

Ⅶ. 결론

본 논문은 온 프레미스 환경에서 MCP를 활용해 비용, 보안, 확장성 문제들을 해결하기 위해 범용 Text-to-SQL 에이전트를 설계 및 구현하였다. 본 연구를 통해 온 프레미스 환경에서 충분히 실용적인 Text-to-SQL 에이전트의 가능성을 확인했다. 다만 현재의 Text-to-SQL 에이전트 구조는 복잡한 작업을 병렬적으로 처리하는 능력이 부족하다. 향후 A2A(Agent-to-Agent) 아키텍처 기반의 멀티 에이전트 구현에 대한 연구를 진행하고자 한다.

참 고 문 헌

[1] D. Gao, et al., "Text-to-SQL Empowered by Large Language Models: A Benchmark Evaluation," arXiv preprint arXiv:2308.15363, 2023.

[2] S. Yao, et al., "ReAct: Synergizing Reasoning and Acting in Language Models," arXiv preprint arXiv:2210.03629, 2022.

[3] Xinyi Hou et al., "Model Context Protocol (MCP): Landscape, Security Threats, and Future Research Directions", arXiv:2503.23278, 2024.

RAG-LLM 을 활용한 고위험 환경 시계열 이상 감지: 프롬프트 최적화와 의사 이상 데이터 생성

최하진, 박지원, 김민석*
상명대학교 휴먼지능로봇공학과

pid9534@naver.com, dev_parkjiwon@naver.com, *minsuk.kim@smu.ac.kr

Time-Series Anomaly Detection in High-Risk Domains based on RAG-LLM: Prompt Optimization and Pseudo-Anomaly Data Generation

Ha Jin Choi, Ji Won Park, Min-Suk Kim*
Dept. of Human Intelligence & Robot Engineering, Sangmyung University

요 약

고위험 플랜트 시설에서의 시계열 데이터 기반 이상 감지는 데이터 수집의 제약, 센서의 다양성, 상이한 데이터 분포 환경, 그리고 시계열 전용 파운데이션 모델의 부재 등으로 해결하기 어려운 도전 과제로 남아 있다. 본 논문은 고위험 환경에서의 시계열 이상 감지 한계를 극복하기 위해, Large Language Model(LLM) 기반 시계열 데이터 이상 감지 구조에 Retrieval-Augmented Generation(RAG) 구조를 결합하여 환경 적응성을 강화하는 새로운 접근 방식을 제안한다. 또한, 시계열 이상 상태에 대한 프롬프트 전략 및 해당 프롬프트와 LLM 을 이용하여 의사 이상 데이터를 생성하고, 생성한 이상 데이터와 입력 데이터와의 유사도 비교를 통해 LLM 기반 이상 추론 과정의 불투명한 결과 분석 문제를 보완하고 이상감지 성능을 높인다. 실험 결과, 제안한 방식은 다양한 환경에서 일관되게 높은 정확도를 나타내고 있으며, 데이터가 제한적인 환경에서도 효과적인 적용 가능성을 보여주고 있다.

Ⅰ. 서 론

고위험 산업 환경의 시계열 데이터 이상 감지는 데이터 수집의 어려움과 다수의 센서별 데이터 분포 차이로 추가 모델이 요구되지만 시간과 비용이 많이 필요하다. 더불어 시계열 전용 파운데이션 모델의 부재로 인해 도메인 간 전이 학습을 적용하기도 매우 어렵다. 이러한 한계를 극복하기 위해, 최근에는 사전 학습된 Large Language Model(LLM)을 시계열 분석에 적용하려는 연구가 활발히 진행되고 있지만, 이상 감지 분야에서는 여전히 시계열 데이터의 도메인별 특수성을 반영하지 못해 도메인 적응력이 제한되고 있다.

본 논문에서는 이러한 문제를 해결하기 위해 프롬프트 설계에 기반한 LLM 시계열 이상 감지 구조인 AAD-LLM[1]을 바탕으로 Retrieval-Augmented Generation(RAG)을 결합한 프레임워크를 제안하며, 이를 통해 적용 도메인에 대한 적응성을 향상시키고자 한다. 더하여 이상을 잘 감지할 수 있도록 보조기법을 더한 프롬프트를 설계했다. 또한, LLM 을 이용해 의사 이상 데이터를 생성하고, 이를 입력 데이터와 비교하여 유사도를 산출함으로써, LLM 의 추론 과정에서 발생하는 결과 해석의 모호성을 완화하고 이상 감지 성능을 향상시킨다.

Ⅱ. 시스템 구조

Ⅱ.1. RAG 기반 도메인 적응

RAG 는 retriever 를 통해 외부 지식을 검색하여 LLM 의 응답 생성 과정에 반영함으로써, 출력의 품질과 신뢰성을 향상시키는 기법이다. 일반적으로 Retriever 은 벡터 데이터베이스나 문서 집합에서 유사도 기반으로 관련 정보를 검색하고, 이를 LLM 의 입력 프롬프트에 포함시켜 응답을 생성한다. 본 연구에서는 적용 도메인에 특화된 데이터를 csv 파일 형태로 벡터 데이터베이스에 저장하고, 입력 데이터가 새롭게 들어오면 이를 참조하도록 하여 환경 적응력을 향상시키는데 목적이 있다.

Ⅱ.2. 프롬프트 설계

INSTRUCTION
You are an anomaly detection and synthetic data generation expert for frequency-domain data.
Abnormality types examples:
- Point anomaly: a few isolated frequency points deviate sharply from the normal range.
- Collective anomaly: a contiguous frequency range shows consistent deviation.
- Level shift: a sustained increase/decrease across a wide band.
- Trend change: slope across the spectrum differs significantly from normal.

CONTEXT
FFT-based spectra from 20kHz to 100kHz at 0.25kHz intervals (320points).
The RetrievalQA chain will supply reference normal data
The reference data has the form of (k, 320), consisting of k data samples each with 320 points
< reference data from RAG DB >

INPUT_DATA
< Input data >

TASKS
1) First, evaluate overall pattern similarity:
- Compare INPUT_DATA against the reference normals in CONTEXT.
- Determine whether INPUT_DATA is normal or anomalous by considering all factors such as the mean, standard deviation, maximum, and minimum of the k normal samples referenced in CONTEXT.
- Also consider the proportion of outside points across the whole spectrum. If about 5% or more of all points consistently lie outside the normal band, that should also be considered anomalous.
2) Consider both the numerical data and the visual information, in comparison with the reference patterns provided in CONTEXT.
3) Decide NORMAL vs ANOMALOUS and provide the reasons.
- try identify whether there are anomalies in the time series.
- if anomalies are identified, try to get it's index according to it's position in the time series list.
- explain why those points should be considered as anomalies.
- Return a value between 0 and 1 (0 = normal, 1 = anomaly, anomaly threshold=0.5).
4) Generate a synthetic anomaly data based from the reference samples in CONTEXT:
- Choose an anomaly type randomly.
- Apply a ±20%~±40% modification in a contiguous range.
- Preserve overall structure outside the affected ranges.
- Ensure that modified values do not exceed twice the maximum or minimum values of the original series.

그림 1. 프롬프트 템플릿 구성

위 그림과 같이 프롬프트는 INSTRUCTION, CONTEXT, INPUT_DATA, TASKS 네 부분으로 구성되며 규칙 및 이상 상태에 대한 정의가 포함된다. 또한, 이상 탐지 성능 향상을 위해 프롬프트에 기존의 세 가지 보조 기법[2]을 적용하였다. 첫째, multi-modal instruction 을 활용하여 LLM 이 시계열 데이터를 단순한 수치정보 뿐만 아니라 시각적 패턴으로도 해석할 수 있도록 하였으며, 이를 통해

이상 패턴 인식의 정확도를 향상시켰다. 둘째, In-context learning 을 통해 RAG retriever 에서 참조한 정상 데이터 예시를 포함시켜 데이터가 제한된 환경에서도 높은 적응성을 확보할 수 있게 하였다. 셋째, Chain-of-Thought(CoT) 프롬프트 튜닝을 적용하여 모델이 이상 판단 과정과 근거를 단계적으로 서술하게 함으로써 LLM 기반 이상 탐지 정확도를 향상시켰다.

Ⅱ.3. 시스템 구조

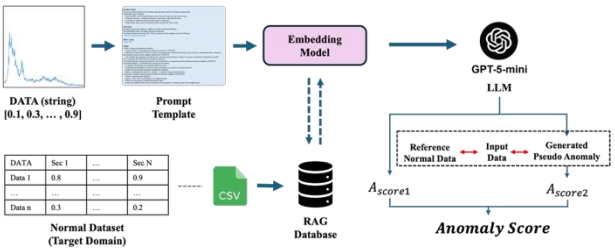


그림 2. RAG-LLM 기반 이상감지 작업흐름도

그림 2 는 본 논문에서 제안하는 RAG-LLM 기반 이상 감지의 전체 작업흐름도를 나타내고 있다. 시계열 입력 데이터 x_{in}이 시스템에 제공되면, 해당 데이터는 전처리 과정을 거친 후 앞서 언급한 바와 같이 프롬프트 템플릿에 입력된다. 프롬프트는 임베딩 모델을 통해 벡터화되고 이 벡터는 retriever 를 통해 도메인 관련 데이터셋에서 유사도가 높은 정상 데이터들을 검색하는데 사용된다. 또한, 검색된 관련 정상 데이터와 원본 프롬프트는 LLM 에 입력되고 두 가지 방식으로 이상 감지 과정을 거친다. 첫 번째 방식은 입력 데이터와 RAG 를 통해 검색된 유사 데이터 간의 비교 및 분석을 LLM 이 자체 판단을 하여 A_{score_1} 이 산출된다. 두 번째 방식에서는 LLM 이 프롬프트를 기반으로 n개의 의사 이상 데이터(pseudo anomaly data)를 생성한다. 이에 대한 평균 점인 P_a 와, RAG retriever 를 통해 뽑은 n개의 정상 데이터에 대한 평균 점인 P_n을 수식(1)과 같이 구한다. 이후 수식(2)와 같이 입력 데이터와 각 평균점과의 코사인 유사도 $\cos(\,)$ 에 대한 Softmax 기반의 두번째 이상점수인 A_{score_2}를 계산한다.

$$P = \frac{1}{n}\sum_{k=1}^{n} x_k \tag{1}$$

$$A_{score_2} = \frac{\exp(\cos(x_{in}, P_a))}{\exp(\cos(x_{in}, P_a)) + \exp(\cos(x_{in}, P_n))} \tag{2}$$

최종 이상 점수는 수식(3)과 같이 α 와 β 를 통해 기여도를 조절하여 두 방식의 이상 점수를 가중 합산하여 결정된다.

$$Anomaly\ Score = \alpha \cdot A_{score_1} + \beta \cdot A_{score_2} \tag{3}$$

Ⅱ.4. 데이터 분석 및 처리

본 연구에서는 실험을 위해 발전소의 이차계통에서 배관의 누출 상황을 가정한 실험 장비에서 수집한 시계열 데이터를 활용하였다. 초음파 신호로 수집된 데이터는 고속 푸리에 변환을 통해 전처리 하였으며, 20~100 kHz 범위를 0.25 Hz 간격으로 측정한 평균 스펙트럼 값 320 길이로 구성된다. 서로 다른 환경에서 수집한 A, B, C 클래스가 있으며 구성은 표 1 과 같다.

	A	B	C
Normal	250	250	250
Anomaly	200	200	200

표 1. 데이터셋 구조

Ⅲ. 실험

본 실험에서 LLM 은 OpenAI 의 GPT-5-mini 를 사용하였으며, RAG 모듈에서 참조할 전용 CSV 파일에는 각 환경별로 정상 데이터를 무작위로 50 개씩 저장했다. 나머지 데이터는 실험의 입력데이터로 사용하여 제안한 구조의 성능을 평가했다. 또한, 제안한 RAG-LLM 기반 이상 감지 기법(RAG-AD)을 기존 LSTM 기반 이상 감지 모델과 비교하여 실험을 진행하였다. 실험 결과, 모든 환경에서 RAG-AD-2, RAG-AD-1, LSTM 순으로 성능이 우수하게 나타났다. 특히, RAG-AD-2 는 의사 이상 데이터를 사용한 이중 평균 점수 기반 이상 감지 방식인 $Anomaly\ Score$ 를 적용함으로써, A_{score_1} 만을 사용한 RAG-AD-1 대비 모든 환경에서 성능이 향상되었다. 이는 의사 이상 데이터 생성을 통해 정상 데이터와의 차이가 더욱 명확히 하여, LLM 의 이상 판단 정확도를 향상시킴을 보여주고 있다.

	A	B	C
LSTM Autoencoder	0.5325	0.4528	0.5401
RAG-AD -1	0.8041	0.6869	0.7693
RAG-AD-2	**0.8292**	**0.7692**	**0.8182**

표 2. F1-score 측정 결과

Ⅳ. 결론

본 논문에서는 데이터 제약이 존재하는 고위험 환경에서의 이상 감지를 위해 RAG-AD 구조를 제안하였으며, 프롬프트 전략과 의사 이상 데이터 생성을 통해 LLM 의 도메인 적응성과 이상 감지 성능을 향상시켰다. 또한, 고위험군 배관 누출을 재현한 테스트베드에서 수집한 데이터기반의 제안한 모델 성능은 기존 LSTM 모델 대비 높은 성능을 달성하였다. 특히, 의사 이상 데이터 기반 이중 평균 점수 방식은 모든 환경에서 성능 향상을 보였으며, 이는 데이터가 제한된 고위험 환경에서의 실질적인 활용 가능성을 시사하고 있다.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government. (MIST) (No. RS-2022-00144000)

참 고 문 헌

[1] Russell-Gilbert, A., Sommers, A., Thompson, A., Cummins, L., Mittal, S., Rahimi, S., ... & Church, J. (2024, December). Aad-llm: Adaptive anomaly detection using large language models. In *2024 IEEE International Conference on Big Data (BigData)* (pp. 4194-4203). IEEE.

[2] Dong, M., Huang, H., & Cao, L. (2024). Can LLMs Serve As Time Series Anomaly Detectors?, *arXiv preprint arXiv:2408.03475*.

음성 지원 대화형 AI 서비스를 활용한 정서 케어 지원

[1]유영호, [1]김세현, [1]박도현, [2]홍수지, [1]이병정
[1]서울시립대학교 컴퓨터과학부, [2]서울시립대학교 인공지능학과
netface12347@gmail.com, rlatpgus5432580@naver.com, qkrehgus02@uos.ac.kr, sujihong93@gmail.com, bjlee@uos.ac.kr

Towards Supporting Emotion Care by Utilizing Conversational AI Service with Voice

[1]Youngho Yoo, [1]Sehyeon Kim, [1]Dohyeon Park, [2]Suji Hong, [1]Byungjeong Lee
[1]Dept. of Computer Science and Engineering, University of Seoul, [2]Dept. of Artificial Intelligence, University of Seoul

요 약

본 논문은 다중 에이전트 아키텍처를 기반으로 사람의 정서 케어 지원을 돕는 대화형 AI 에어전트 서비스를 소개한다. 본 서비스는 사람의 감정을 반영하기 위하여 LLM 을 미세조정하고, 음성 데이터의 임베딩을 통하여 감정을 분석함으로써 사용자의 감정 맥락을 인지한다. 또한, 사용자의 맞춤형 대화를 지원하기 위해서 클라우드 기억 관리자를 활용한다. 이를 통해 기존의 사실 기반 질의응답에서 벗어나 사용자에게 지인과 대화하는 듯한 보다 친밀한 경험을 제공한다.

Ⅰ. 서 론

최근 대화형 AI 는 빠른 기술적 발전을 이루었으나, 주로 정보의 사실적 정확성을 높이는 데 초점을 맞추어 왔다. 이로 인해 사용자의 감정적 맥락을 인지하거나 과거 대화 이력을 바탕으로 개인화된 상호작용을 제공하는 데에는 명백한 한계가 존재한다. 이러한 한계는 특히 섬세한 접근이 요구되는 정서 케어 지원 분야에서 AI 의 실용성을 제한하는 요인으로 지적된다. 본 논문은 기존 대화형 AI 의 한계를 극복하기 위해, 자율적으로 목표를 수행하는 복수개의 에이전트들이 협력하는 새로운 아키텍처를 제안한다. 본 서비스는 개인화된 정서 지원, 장기 기억, 그리고 실시간 감정 인식의 세 가지 핵심 분야를 기반으로 한다. 서비스 동작 원리는 먼저, 사용자의 음성 입력에서 음향 특징을 추출하여 감정을 분석함과 동시에, 음성 입력을 텍스트로 변환한다. 다음으로 이렇게 변환된 텍스트와 감정 정보는 외부 지식을 활용해 미세조정된 LLM 에 전달된다. LLM 은 클라우드 기억 관리자에 저장된 과거 대화, 토픽, 감정 변화 데이터를 참조하여 개인에게 맞춤화된 일관성 있는 답변을 생성한다. 최종적으로 생성된 텍스트는 음성으로 변환되어 사용자에게 제공된다.

Ⅱ. 관련 연구

최근 AI 연구는 사용자의 정신 건강을 지원하는 개인화된 시스템으로 확장되고 있다. 다중 에이전트 아키텍처를 통해 맞춤형 정신 건강 지원을 제공하는 연구[1]는 AI 의 정서적 지원 역할에 대한 가능성을 제시했다. 이러한 고차원적 지원의 전제 조건은 사용자와의 신뢰이며, 이는 AI 가 과거 대화를 일관성 있게 기억하는 능력에 의존한다. 이를 위해, 과거 대화를 성찰하고 요약하여 깊이 있는 관계 형성을 목표로 하는 '성찰적 기억 관리' 메커니즘이 제안되었다[2]. 나아가, 효과적인 정서적 교감은 사용자의 현재 감정 상태를 정확히 파악하는 능력에 크게 좌우된다. 기존 텍스트 분석의 한계를 넘어, 속삭임과 같은 미세한 음성 신호에서 감정을 직접 인식하는 연구[3]는 음성 데이터가 가진 풍부한 감정 정보의 유용성을 보였다. 이러한 연구들은 정서 지원 프레임워크[1], 기억 메커니즘[2], 음성 감정 인식[3] 등 각 요소 기술에서 중요한 진전을 이루었다. 그러나 이들을 실시간 음성 기반 시스템으로 유기적으로 통합하려는 시도는 부족했다. 따라서 본 연구에서는 이러한 기술적 공백을 참고하여, 세 가지 핵심 개념을 적용하여 대화형 AI 기반 정서 케어 지원 서비스를 제안한다.

Ⅲ. 음성 지원 대화형 AI 서비스

3.1 유즈케이스도

그림 1 은 본 음성 지원 대화형 AI 서비스 유즈케이스도이다. Converse 는 핵심적인 유즈케이스로 사용자가 음성, 텍스트를 통하여 시스템과 데이터를 주고받는다. RequestContents 는 사용자가 자신의 감정 상태에 맞는 케어 컨텐츠를 요청한다. SearchConversation 은 사용자 요청에 맞는 과거 대화 이력을 적절한 형태로 제공해준다.

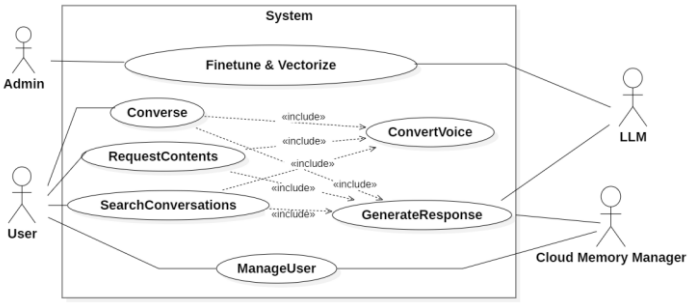


그림 1. 유즈케이스도

GenerateResponse 는 앞선 3 개의 유즈케이스에서 넘어온 입력을 미세조정된 LLM 에 전달하여 답변을 생성한다. ManageUser 는 회원정보에 관한 유즈케이스이며 사용자 별로 맞춤형 대화를 지원하기 위해 사용자 정보를 기억한다. ConvertVoice 는 음성과 텍스트를 서로 변환하고, 음성의 감정 벡터를 추출하는 기능이다. Finetune & Verctorize 은 더 정확한 감정 대화와 컨텐츠 추천을 위해 LLM 을 연관된 전문적인 데이터로 미세조정한다. 또한, 전문적인 데이터를 벡터화 한 후 저장하여 사용자 입력이 들어오면 의미 검색을 통해 정보를 추출하여 LLM 의 입력으로 사용한다.

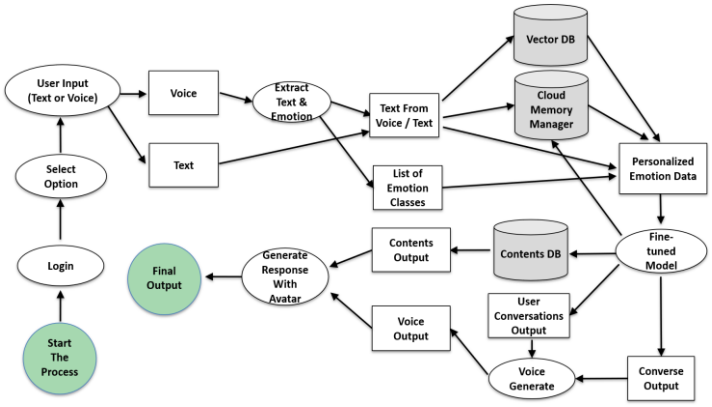


그림 2. 전체 작업흐름

그림 2 는 본 서비스의 전체적인 흐름을 보여주는 작업 흐름도이다. 사용자는 서비스를 사용하기 위해 회원가입하고 로그인한다. 그리고 Converse, RequestContents, SearchConversation 3 가지 옵션 중 하나를 선택하고 음성이나 텍스트로 원하는 입력을 넣는다. 음성 입력이라면 음성으로부터 벡터를 추출해 감정을 판단함과 동시에 음성을 텍스트로 변환한다. 텍스트 입력이라면 별도의 처리 과정 없이 진행된다. 이후 텍스트를 기반으로 외부 상담 데이터와 같은 전문적인 데이터가 들어있는 벡터 DB 에서 의미 검색을 통해 관련 정보를 추출한다. 동

시에 사용자에 대한 기억을 담당하는 클라우드 기억 관리자에서 현재 입력과 관련있는 사용자 정보를 찾는다. 이 모든 정보가 통합되어 미세조정된 LLM 에 입력으로 사용되고 선택한 옵션에 적합한 정보가 출력된다. LLM 의 응답은 사용자 입력과 함께 클라우드 기억 관리자에 저장되어 다음 응답 생성에 활용된다. Converse 와 SearchConversation 의 경우 LLM 의 답변을 그대로 음성으로 변환해 움직이는 캐릭터(아바타)와 함께 출력한다. RequestContents 의 경우 LLM 의 응답을 바탕으로 Contents 가 보관된 DB 에서 컨텐츠를 찾아서 움직이는 캐릭터(아바타)와 함께 출력한다. 본 연구의 작업흐름은 참고문헌[1]의 구조를 기반으로 하되, 음성 중심의 실시간 상호작용에 최적화하기 위해 전체 과정을 간소화하고 개인화된 경험을 위한 클라우드 기억 관리자를 추가하여 재구성했다.

3.2 시스템 아키텍처

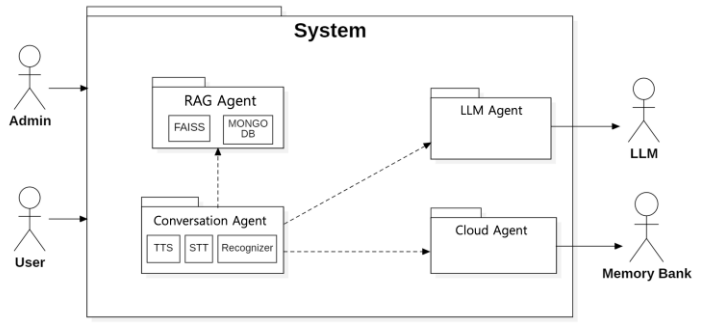
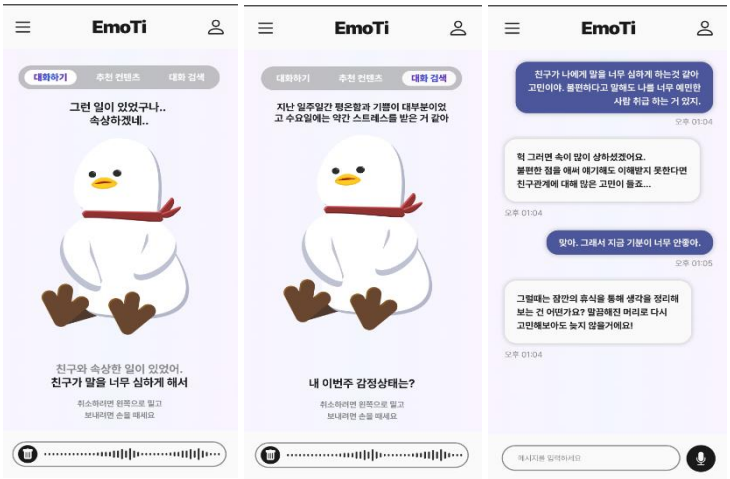


그림 3. 아키텍처도

본 연구에서 제안하는 시스템의 전체 아키텍처는 그림 3 과 같다. 본 시스템은 대화 에이전트(Conversation Agent)를 중심으로 다른 모든 컴포넌트를 유기적으로 호출하고 관리하는 구조이다. 시스템의 모든 상호작용은 대화 에이전트를 통해 시작된다. 먼저 STT(Speech-to-Text)로 사용자의 음성 입력을 인식하고, Recognizer 를 통해 그 안에 담긴 감정 상태를 파악한다. 이렇게 파악된 사용자의 의도와 감정을 기반으로, 대화 에이전트를 답변을 더욱 풍부하게 만들기 위해 필요한 정보를 수집한다. 이를 위해 RAG Agent 를 호출하여 FAISS[i] 및 MongoDB[ii]에서 관련 외부 사실을 검색하고, Cloud Agent 를 호출하여 클라우드 기억관리자, 즉 메모리 뱅크(Memory Bank)[iii]에 저장된 대화 이력이나 과거 감정 상태 같은 장기 기억 데이터를 추출한다. 이렇게 종합된 정보를 바탕으로 LLM Agent 를 호출하여 LLM 을 위한 최적의 프롬프트를 생성한다. 마지막으로, LLM 으로부터 받은 답변을 TTS(Text-to-Speech)를 통해 자연스러운 음성으로 변환하여 사용자에게 최종적으로 전달한다. 한편, 관리자(Admin)는 별도로 시스템의 전반적인 성능 향상을 위한 Finetune 및 Vectorize 작업을 수행한다.

Ⅳ. 대화 서비스 UI



(a) 대화하기 (b) 대화 검색 (c) 채팅 로그

그림 4. 서비스 UI

그림 4 는 본 연구에서 제안하는 음성 지원 AI 대화 서비스 'EmoTi'의 사용자 인터페이스를 보여준다. 본 서비스는 텍스트와 음성 입력을 모두 처리할 수 있는 모델을 기반으로 설계되었으며, 사용자는 자신의 선호나 상황에 따라 입력 방식을 자유롭게 선택하여 시스템과 상호작용한다. 시스템의 주요 기능은 UI 의 상단 탭을 통해 접근 가능하다. '대화하기'는 사용자의 일상 및 감정에 대한 전반적인 대화를 수행하는 핵심 기능으로, 시스템은 사용자의 발화에 공감하고 정서적 지지를 제공하는 상호작용을 목표로 한다. 예를 들어, 그림 4 의 좌측 화면(a)에서는 사용자가 친구와의 갈등으로 인한 속상함을 토로하자, 시스템이 공감적 반응을 보이는 것을 확인할 수 있다. '추천 컨텐츠' 기능은 대화의 맥락과 사용자의 감정 상태를 분석하여 비언어적 콘텐츠를 제공하는 역할을 수행한다. 이는 캐릭터의 제스처가 동반된 노래, 시 낭송 등의 형태로 제공되어 사용자에게 다각적인 위로와 즐거움을 전달한다. '대화 검색'은 축적된 대화 이력을 기반으로 사용자가 특정 정보를 검색하거나 자신의 감정 변화를 회고할 수 있도록 지원한다. 그림 4 의 중앙 화면(b)에 나타난 바와 같이, 사용자가 "내 이번주 감정상태는?"과 같은 질의를 하면, 시스템은 과거 대화 내용을 요약 및 분석하여 유의미한 인사이트를 제공한다. 이러한 핵심 기능들과 더불어, 본 서비스는 UI 측면에서 몰입감 높은 '캐릭터 모드'와 텍스트 기반의 '채팅 모드' 를 모두 지원한다. 그림 4 의 좌측 및 중앙 화면처럼 캐릭터와 음성 또는 텍스트로 상호작용하면, 해당 대화 내용은 우측 화면(c)의 채팅 로그에 실시간으로 기록된다. 이와 같은 유연한 UI 구조는 캐릭터와의 감성적 교감을 깊게 하면서도 대화의 연속성과 정보제공을 보장함으로써 사용자 경험을 극대화하는 효과를 가진다.

Ⅴ. 결론

본 논문에서는 사용자에 대한 심층적 기억을 참고하여 섬세한 감정적 응답으로 정서 케어 지원하는 서비스를 소개하였다. 기존의 단순 사실 기반 응답이 아닌 목소리의 비언어적 특징을 고려한 감정적 소통이 가능하며, 현재의 질문에만 의존한 답장에서 벗어나 사용자와의 대화를 체계적으로 저장하고 참조하여 사용자에게 맞춤형으로 대화를 제공한다. 또한, 각 기술을 자율적인 에이전트로 재정의하고 이들의 협력을 통해 정서 케어라는 복합적인 문제를 해결하는 새로운 아키텍처를 제안했다는 점에서 의의를 가진다. 그러나 표정과 행동 등의 또 다른 비언어적 특징을 고려하지 못한 한계가 있다. 향후에는 실시간 영상 분석을 통한 표정 인식과, 이를 기반으로 한 감정 벡터 표현 기법을 결합하여 보다 현실적이고 몰입감 있는 감정 교류형 AI 서비스로 확장할 것이다. 이러한 발전은 사람과 AI 의 상호작용을 한층 더 자연스럽고 의미 있는 방향으로 진행하게 할 것이다.

참 고 문 헌

[1] Pattar V., Patil T., Dhuri B., Karel A., Deole A., and Kulkarni S., "*Mindcare: A Multi-Agent AI Architecture for Personalized and Responsible Mental Health Support*," International Journal of Research and Scientific Innovation (IJRSI), vol. 12, no. 4, Apr. 2025.

[2] Tan Z., et al., "*In Prospect and Retrospect: Reflective Memory Management for Long-term Personalized Dialogue Agents*," in Proc. of Annual Meeting of the Association for Computational Linguistics (ACL), Aug. 2025.

[3] Qu X., Sun Z., Feng S., Chen C., and Tian T., "*Breaking the Silence: Whisper-Driven Emotion Recognition in AI Mental Support Models*," in Proc. of IEEE Conference on Artificial Intelligence (CAI), 2024.

[i] https://github.com/facebookresearch/faiss
[ii] https://www.mongodb
[iii] https://gemini.google.com/app/08823167abb96296

LLM 기반 합성 범죄수사 데이터 생성 시스템의 설계

김희두*, 우병관, 문병훈**, 김지온
*고려대학교, 한림대학교, **수사연수원
heedou123@korea.ac.kr, byeongkwan.woo@gmail.com, mbh010@police.go.kr, jion972@hallym.ac.kr

Design of an LLM-Based Synthetic Criminal Investigation Data Generation System

Hee Dou Kim*, Byeong Kwan Woo, Byung Hoon Moon**, Jion Kim
*Korea Univ., Hallym Univ., **Korean Police Investigation Academy

요 약

본 논문은 실제 수사 데이터를 활용하기 어려운 법적·윤리적 제약을 극복하기 위해, 멀티에이전트 기반 합성 데이터 생성 파이프라인과 전문가 참여형 시드데이터 설계를 제안한다. 제안된 방법은 사건 시나리오와 수사 기록물을 실제와 유사하게 합성하여 LLM 학습용 데이터로 활용 가능하며, 정합성과 품질을 검증할 수 있는 체계를 포함한다. 이를 통해 안전하고 신뢰성 있는 LLM 기반 수사지원시스템 개발의 기초를 마련하고자 한다.

Ⅰ. 서 론

최근 대규모 언어모델(LLM, Large Language Model)의 활용은 경찰 실무 영역에서도 점차 확산되고 있다[1]. 범죄 수사는 사건을 둘러싼 법률적 해석과 쟁점에 대한 추론, 피의자 신문, 공소 유지에 필요한 보고서 작성, 방대한 사건 기록 분석 등 높은 수준의 전문성을 요구하는 영역이다. 그러나 SNS나 온라인 커뮤니티를 기반으로 조직적·대규모로 발생하는 사이버 범죄 피해가 급격히 증가는 현상이 지속되면서, 수사관들은 수사의 전문성을 유지하면서도 복잡하고 방대한 사건을 신속하고 효율적으로 처리하는 데 큰 어려움을 겪고 있다. 이러한 상황 속에서 LLM은 데이터 분석을 도와주거나 법률적 쟁점을 신속히 분석해주고, 수사 보고서 초안을 자동으로 작성해주는 등 업무 효율을 높이는 핵심 도구로 주목받고 있으며, 이에 따라 범죄 수사 자동화 시스템의 필요성이 높아지고 있다[2].

그러나 범죄 수사 지원을 위한 LLM 기반 시스템 개발의 큰 난제는 수사 데이터 확보의 어려움이다. 실제 사건 기록물은 영장, 진술서, 수사보고서 작성 시 가장 유사한 예제가 될 수 있는 가치 있는 원천 데이터이지만, 개인정보보호법, KICS 운영상의 법·보안적 제한, 그리고 AI 활용 시의 감시·편향 등 윤리적 문제로 인해 직접 활용은 사실상 불가능하다[3].

이러한 배경 속에서 LLM 기반 합성 데이터(synthetic data)는 데이터 부족을 보완하면서 안전한 수사지원시스템 개발의 중요한 대안으로 주목받고 있다. 합성 데이터는 실제 데이터를 완전히 대체할 수는 없고, 생성 과정에서 편향 증폭이나 성능 저하, 새로운 윤리적 위험을 초래할 수 있는 문제가 제기되기도 하지만, 최근 발전한 LLM의 정교한 문장 생성 능력에 체계적 검증 도구와 필터링 절차를 결합한다면 특히 데이터가 희소한 경찰 AI 영역에서 학습 모델 보강 수단으로 큰 효과를 기대할 수 있다[4].

이에 본 연구는 멀티에이전트 기반 합성 수사데이터 설계 프레임워크를 제안한다. 이를 통해 국내의 수사 지원 시스템 개발 환경에서 활용 가능한, 안전하고 검증된 데이터 생성 시스템의 구축 방안을 제시하고자 한다.

그림 1과 같이 제안하는 합성 수사데이터 생성 기술은 사건 데이터를 크게 두 가지 층위로 나눈다. 첫째, 실제 수사에서 다루는 죄종별 시나리오를 실제 발생 사건과 유사하게 재현한 이벤트 정보를 생성한다. 둘째, 이러한 이벤트 전개에 따라 실제 수사 과정에서 작성되는 수사문서를 생성하여, 마치 실제 사건의 수사 기록물처럼 여겨지는 데이터를 만들어낸다. 이를 통해 LLM의 수사 추론이나 자동 수사보고서 작성을 위한 학습용·예제 데이터로 활용할 수 있다. 본 연구에서 제안하는 데이터 생성 기술을 적용하면, 법적 제약으로 인해 직접 활용할 수 없었던 실제 수사 기록물을 대신해 실제에 가까운 합성 수사데이터를 활용할 수 있으며, 이를 기반으로 LLM 기반 수사지원 시스템 개발에 기여할 수 있을 것으로 기대한다.

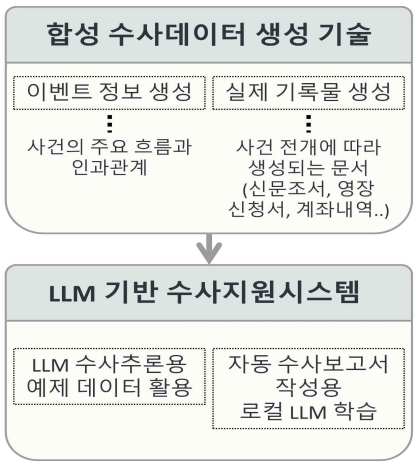


그림 1. 합성 수사데이터 생성 기술의 기대효과

Ⅱ. 기술의 개요 : 합성 수사데이터의 타당성 확보를 위한 전제조건

합성 수사데이터를 LLM 수사지원시스템 개발에 활용하기에 앞서 판단해야 할 가장 중요한 고려 요소는 생성된 데이터가 실제 세계의 사건 데이터와 얼마나 유사한가, 그리고 데이터 내 정합성이 얼마나 타당하게 검증되었는가에 있다. 이러한 요건을 충족하여 품질 높은 합성 데이터셋을 확보한다면, 실제 수사 데이터를 사용하지 않더라도 안전하게 사건 기록물을 활용한 LLM 학습을 시도해볼 수 있다. 본 연구는 합성 수사데이터의 타당성 검증이 유효하도록 하기 위해, 다음과 같은 구조를 필수적으로 따를 것을 전제 조건으로 설정한다. 첫째, 데이터 생성 단계를 세분화한 멀티에이전트 기반 파이프라인을 적용하여, 각 에이전트가 독립적으로 작동하

면서도 순차적으로 정합성을 유지하도록 설계한다. 이는 실제 사건 데이터의 내용이 아주 복잡하기 때문에 단순화된 LLM 프롬프트 설계만으로는 복잡하고 방대한 사건 데이터 내의 논리적 흐름을 유지하고, 이와 유사하게 변형시키는 작업이 어려울 것이라는 점에 착안하였다. 이와 같은 독립된 역할을 담당하는 멀티에이전트 기반의 설계를 통해 단일 문서 단위가 아닌, 사건 시나리오별로 기록이 묶여 제공되는 체계를 구현할 수 있을 것이다. 둘째, 수사 전문가가 직접 참여하여 합성 데이터 생성을 위한 품질 높은 시드 데이터를 제작한다. 이를 통해 합성 데이터 생성에 필요한 필수적인 LLM 프롬프트용 예제를 만들고, 실제 세계의 사건 데이터와 유사한 합성을 유도하는 전략을 취할 수 있다. 마지막으로, 합성된 생성 데이터의 타당성을 검증하는 전문가 기반 절차를 도입한다.

Ⅲ. 멀티에이전트 기반 합성 파이프라인

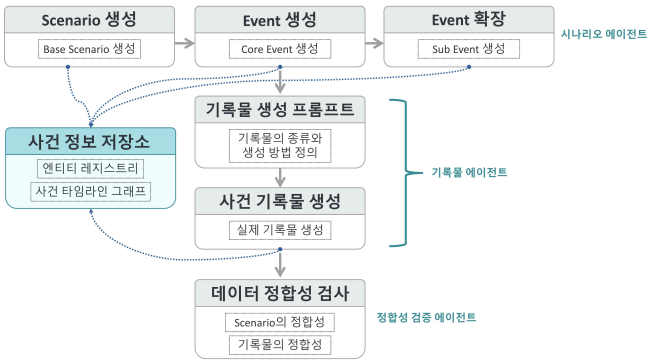


그림 2 합성 수사데이터 생성 파이프라인 구조

그림 2는 합성 데이터 파이프라인의 구조를 보여준다. 제안하는 파이프라인은 LLM의 생성 기술을 결합한 멀티에이전트 기반 시스템으로 설계되었으며, 각 에이전트가 독립적으로 작동하면서도 순차적으로 협력하는 생성 - 검증-패키징의 구조를 따른다. 또한 각 에이전트는 GPT5, Gemini 2.5와 같은 최신 상용 언어모델의 API만을 활용하되, 각 단계에서의 결과물의 완성도를 스스로 판단하여 다음 단계의 에이전트에 결과를 전달할 수 있는 기능을 지니고 있다. 먼저, 시나리오 에이전트에서는 사건의 이벤트 정보를 정교하게 생성하여 사건의 시나리오를 완성한다. 이를 위해 이벤트 정보의 속성을 Base Scenario, Core Event, Sub Event라는 하위 범주로 구분한다. 각 이벤트는 다음의 의미를 지니고 있다.

· Base Scenario : 전체 사건의 뼈대를 이루는 기본 정보
· Core Event : 10개 이내로 압축된 핵심 사건
· Sub Event : Core Event를 기준으로 세분화된 사건

이처럼 이벤트를 구분함으로써 범인·피해자·관련자 등의 사건 발생과 관련된 일련의 행위와 결과들을 독립적으로 생성할 수 있다. 또한 생성된 이벤트 정보는 공유된 사건 정보 저장소에 저장되며, 이후 정합성 검증 에이전트를 통해 논리적 오류, 예를 들어 동일 범인이 같은 시간대에 서로 다른 행위를 한 경우 등을 점검한 뒤, 재생성이나 수정 과정을 거칠 수 있다. 다음으로, 기록물 에이전트는 완성된 이벤트를 토대로 실제 기록물을 작성하여 최종 패키징을 하여 사용자에게 제공하며, 아래 두 단계로 나뉜다.

· 기록물 생성 프롬프트 작성 단계: 어떤 기록물을 생성해야 하는지 규정
· 기록물 작성 단계: 해당 프롬프트에 따라 실제 기록물을 생성

이는 기록물의 종류, 예를 들어 압수영장, 수사보고서, 신문조서 등에 따라 서술 방식과 필수 기재 항목이 모두 다르기 때문에, 기록물별로 별도의 LLM 지침이 필요하기 때문이다. 기록물 생성 과정 역시 사건 정보 저장소의 엔티티와 사건 타임라인을 기준으로 오류 검증 절차를 거친다. 최종적으로 완성된 기록물들은 시간적 순서와 정합성이 유지된 상태로 하나의 사건에 대한 수십 건의 기록물 묶음이 되며, 이를 JSON이나 마크다운(markdown) 형식 등으로 패키징하여 산출물로 저장할 수 있다.

Ⅳ. 전문가 참여형 시드데이터 제작 및 품질 평가

다음으로, 합성 수사데이터의 품질을 높이기 위해 본 연구는 전문가가 주도하여 제작한 시드 데이터셋을 도입한다. 여기서 시드데이터란 실제 사건 기록물을 바탕으로 수사 전문가가 작성한 시나리오 데이터를 의미한다. 시드데이터는 그림 2에서 제시한 합성 데이터 파이프라인과 유사한 구조를 따른다. 즉, 사건 기록물을 선정한 뒤 사건의 기본 시나리오와 이벤트 타임라인을 재작성하고, 해당 이벤트에 따라 분류된 기록물에서 핵심 내용을 추출·정리한다. 이렇게 재작성된 데이터는 합성 데이터와 동일한 형식을 갖춘 정답 데이터셋이 된다. 이때 시드데이터의 제작은 다음과 같은 의의를 지닌다. 첫째, 시드데이터는 실제 사건 기록물에서 LLM 학습에 필요한 핵심 정보만 추출한 예제 데이터로서, 합성 데이터 생성에 필수적인 참고 자료가 된다. 둘째, 시드데이터는 합성 과정에서 LLM이 강력하게 의존할 수 있는 구체적 예시를 제공한다. 셋째, 모든 세분화된 생성 단계에서 시드데이터를 기반으로 LLM의 생성을 유도함으로써, 실제 사건 데이터와 구조적으로 유사한 합성 데이터를 생성할 수 있으며, 이는 곧 few-shot 프롬프트 활용을 가능하게 한다.

마지막으로, 전문가는 합성 데이터의 품질을 평가하여 이를 LLM 기반 수사지원시스템의 학습 재료로 활용할 수 있는지 여부를 판단하는 절차에 참여한다. 이 과정에서 완성된 시드데이터는 정답셋(Golden Answer) 또는 참고용 데이터(Reference Answer)로 사용될 수 있으며, 합성 데이터와 비교하여 종합적인 품질 점수를 산출할 수 있다. 또한 LLM-as-a-Judge 기법을 활용하여 활용도, 가독성, 논리적 일관성, 사건의 다양성의 측면을 지표로 사용하여 인간 평가와의 상관성을 측정하고, 나아가 평가 과정의 자동화를 시도할 수 있다.

Ⅴ. 결론

본 연구는 실제 수사 데이터를 직접 활용하기 어려운 제약을 극복하기 위해, 멀티에이전트 기반 합성 파이프라인과 전문가 참여형 시드데이터 설계를 제안하였다. 이를 통해 실제성과 정합성이 확보된 합성 수사데이터를 제공함으로써, 향후 안전하고 신뢰할 수 있는 LLM 기반 수사지원시스템 개발에 기여할 수 있을 것으로 기대된다. 향후 연구에서는 실제 프로토타입 구현과 다양한 범죄 유형에의 적용, 성능 평가 실험을 통해 제안 시스템의 실효성을 검증할 계획이다

Acknowledgment

이 논문은 2025년도 경찰청의 재원으로 과학치안진흥센터 사이버범죄 수사단서 통합분석 및 추론시스템 개발 사업사업의 지원을 받아 수행된 연구임(No. RS-2025-02218280)

참 고 문 헌

[1] Adams, I. T. (2024). Large language models and artificial intelligence for police report writing. CrimRxiv.

[2] Kim, H., Kim, D., Lee, J., Yoon, C., Choi, D., Gim, M., & Kang, J. (2024, October). LAPIS: language model-augmented police investigation system. In Proceedings of the 33rd ACM International Conference on Information and Knowledge Management (pp.

4637-4644).

[3] 김희두, & 김대호. (2025). 대규모 언어모델의 치안 분야 도입을 위한 기술 및 정책전략 수립 연구. 치안정책연구, 203-242.

[4] Schmidhuber, M., & Kruschwitz, U. (2024). Llm-based synthetic datasets: Applications and limitations in toxicity detection. LREC-COLING 2024, 37.

수사 데이터 중복 식별을 위한 알고리즘 기반 처리 프레임워크 연구

문병훈
성균관대학교 과학수사학과
sindori1@naver.com

An Algorithmic Framework for Identifying Duplicates in Investigative Data

Moon Byoung Hun
Sungkyunkwan Univ

요 약

수사기관에서 수집한 통신·계좌 데이터를 병합하는 과정에서 발생하는 중복 문제를 분석하고, 이를 해결하기 위한 알고리즘 기반 처리 프레임워크를 제안한다. 통신 데이터에는 착발신 번호와 시작일시를 활용한 복합키 알고리즘을 적용하고, 계좌 데이터에는 무방향 계좌쌍 기반 시간·금액 매칭 절차를 적용하여 중복을 정밀하게 식별한다. 이러한 프레임워크는 표준화, 대표 레코드 선정, 시각화, 로그 기록 전 과정을 체계화함으로써 분석의 정확성과 데이터 신뢰성을 동시에 확보한다.

Ⅰ. 서론

수사 현장에서 통신·계좌 데이터 수집·분석 과정은 범죄사실 규명과 증거 확보를 위해 필수적이다[1]. 은행과 통신사로부터 개별적으로 수신한 원본 데이터를 표준화하여 통합하는 절차는 수사 데이터 활용의 출발점이 된다[2]. 병합 과정에서 동일 이벤트, 동일 거래가 중복되는 현상이 빈번히 발생하며 분석 결과의 타당성을 저해할 수 있다. 통신 데이터는 두 개 이상의 전화번호에 대해 통신사실확인자료를 요청하면, 요청한 번호 간에 발생한 송수신 기록이 각각의 요청 건에 포함되어 중복된다. 계좌 데이터에서도 두 개 이상 계좌번호를 대상으로 금융계좌추적용 압수수색영장을 집행할 경우, 동일한 자금 이체가 송금 계좌와 수취 계좌의 거래 내역에 각각 기록되어 병합 시 이중으로 남는다. 이러한 중복은 시각화 과정에서 거래나 통화 횟수를 실제보다 과대하게 표현하여 네트워크 구조와 빈도 기반 지표를 왜곡한다. 특히 관계망 분석이나 시계열 패턴 분석과 같이 정량 지표를 활용하는 분석에서는 중복으로 인한 수치 왜곡이 수사 판단의 오류로 이어질 수 있다. 이 연구는 통신과 계좌 데이터 표준화 및 병합 과정에서 발생하는 중복을 체계적으로 식별하는 방안을 제안한다. 제안 방법은 시간, 금액, 착·발신 번호, 계좌번호 등 핵심 필드를 활용한 복합키 설계와 규칙 기반 필터링 절차를 중심으로 구성되며, 이를 통해 분석의 정확성과 시각화 결과의 신뢰성을 제고하는 데 목적이 있다.

Ⅱ. 국제적 수사 데이터 품질 문제와 사례

데이터 중복과 품질 저하 문제는 국내에 국한되지 않는다. 해외에서도 수사 데이터의 변환·통합 과정에서 유사한 오류가 발생하여 심각한 법적 영향을 초래한 사례가 보고된 바 있다. 덴마크 통신 스캔들은 2010년부터 2019년까지 덴마크 경찰 텔레센터(Telecentre)가 형사사건 수사에 활용되는 과거 통화기록(Historical Call Data Records, HCDR)을 처리하는 과정에서 다수 오류가 발생한 사건이다[3]. HCDR은 기지국 위치, 수·발신 번호, 통화 시간 등 통신 메타데이터를 포함하며, 수사에서 알리바이 검증, 범행 시간대 추정, 공범 관계 식별 등 핵심 증거로 사용된다. 이 사건에서는 데이터 변환 과정에서 행(row) 누락, 형식 해석 오류, 데이터 불일치 등 중대한 문제가 발견되었다. 덴마크 정부는 사건 심각성을 인식하고 10,000건 이상 형사사건을 재검토하였다. 오류의 근본 원인은 여러 통신사로부터 제공받은 다양한 원시 데이터 형식을 단일 형식으로 변환하는 과정에 있었다. 2011년부터 2019년 사이, 외부 조사에서 확인된 원시 데이터 형식만 100종 이상이었다. 변환 시스템은 이러한 변화를 완벽히 처리하지 못했고, 일부 행이 누락 되거나 데이터 의미가 변형되는 오류가 발생하였다. 텔레센터와 경찰청 간 데이터 품질 검증 절차도 미흡하였다. 변환된 데이터와 원시 데이터 행 수를 비교하는 단순 검증만 수행되었으며 내용 정확성이나 완전성은 확인되지 않았다. 오류의 지속에는 편향(bias)과 과신(overconfidence)이 중요한 역할을 하였다. 텔레센터 직원들은 변환 시스템이 정확하게 작동한다고 믿었고, 경찰청과 검찰도 이를 신뢰하였다. 보고서와 변환 데이터가 형식상 일치하는 것이 확인되면 데이터 품질이 보장되었다고 판단하였으며, 이는 각 단계에서 신뢰를 강화하는 편향 눈덩이 효과(bias snowball)로 작용하였다. 이러한 신뢰 구조는 오류를 조기에 발견하지 못하게 만들었고, 일부 사건에서는 잘못된 기소와 재판으로 이어질 위험을 높였다.

Ⅲ. 데이터 중복 발생 구조와 영향 분석

1. 통신데이터

통신 데이터는 수사기관이 통신사실확인자료를 요청하면 각 통신사에서 엑셀 파일 형태로 제공한다. 제공되는 항목은 발신번호, 착신(역발신)번호, 통화시작일시, 통화종료일시, 통신사명(업체명), 발신위치(기지국), 서비스 구분 등이다. 대상 전화번호의 통신사에서는 해당 번호의 발신과 역발신 내역, 다른 통신사에서는 역발신 내역만 제공한다. 요청 대상 전화번호가 두 개(A, B) 이상인 경우 A의 통신사에서 제공받은 A→B 발신 내역과, B의 통신사 외에서 제공받은 B←A 역발신 내역이 중복된다. A 또는 B의 통신 데이터를 개별 분석하면 문제가 되지 않지만, 두 전화번호 간 관계를 분석하기 위해 표준화 후 통합하면 중복이 불가피하다. 이런 중복값은 통신 3사(SKT, KT, LGU+)와 같은 경우 서비스 구분을 제외하고 대부분 항목이 일치하나, 별정통신사는 서비스 구분과 함께 통화종료일시

값이 다른 경우도 발생한다.

<표1> 통신 데이터 중복값 예시

구분	발신번호	착신번호	시작일시	종료일시	업체명	발신위치	서비스 구분	비고
통신3사	01053710000	01056690000	2017-10-24 17:00:05	2017-10-24 17:01:46	KT 이동	경기도 성남시 분당구 판교동 291-0	VOLTE-망외음성(VOLTE간)	
통신3사	01053710000	01056690000	2017-10-24 17:00:05	2017-10-24 17:01:46	KT 이동	경기도 성남시 분당구 판교동 291-0	VOLTE음성	
통신3사	01053710000	01056690000	2017-10-26 16:04:31	2017-10-26 16:06:59	KT 이동	서울특별시 구로구 구로동 1125-6	VOLTE-망외음성(VOLTE간)	
통신3사	01053710000	01056690000	2017-10-26 16:04:31	2017-10-26 16:06:59	KT 이동	서울특별시 구로구 구로동 1125-6	VOLTE음성	
별정통신사	01083190000	01075530000	2022-08-17 13:45:13	2022-08-17 13:45:21	LGU+	(04323) 서울시 용산구 동자동 12-0	VOLTE 음성	
별정통신사	01083190000	01075530000	2022-08-17 13:45:13	2022-08-17 13:45:22	스마텔	(04323) 서울시 용산구 동자동 12-0	volte 음성	
별정통신사	01058490000	01072370000	2022-08-17 16:13:18	2022-08-17 16:13:53	LGU+	(61957) 광주시 서구 치평동 1171-5	VOLTE 음성	
별정통신사	01058490000	01072370000	2022-08-17 16:13:18	2022-08-17 16:13:54	큰사람	(61957) 광주시 서구 치평동 1171-5	volte 음성	

이러한 중복은 데이터 생성 단계에서부터 구조적으로 내포되어 있다. 통신사와 별정통신사 간 망 연동 과정에서 각 사업자가 자체 과금·관리 목적의 로그를 생성하고, 수사기관에는 사업자별 로그를 개별 제공하기 때문이다. 여기에 전화번호 단위로 통신사실확인자료를 요청하는 절차가 더해지면서, 요청한 번호 간에 발생한 송수신 기록이 각 요청 건에 중복된다. 수사기관에서 사용하는 표준화 프로그램(i1, A1)도 데이터 중복의 원인이 된다. 이 프로그램은 각 통신사에서 제공받은 데이터를 표준화한 뒤 하나의 파일로 병합하는 방식이므로, 동일 이벤트가 중복될 가능성이 존재한다. 데이터의 중복은 네트워크 분석에서 동일 노드 간 연결(링크)을 증가시키고, 요청 대상 간 통화 빈도를 실제보다 과대(2배) 계산하게 만든다. 시계열 분석에서는 동일 시점에 중복이벤트가 집중되는 것처럼 표현되어 특정 시기 활동량이 왜곡된다. 그 결과 범죄 연루 가능성 평가나 공범 관계 추정과 같은 분석에서 부정확한 판단이 이루어질 가능성이 존재한다.

2. 계좌거래데이터

계좌 거래 데이터는 수사기관이 금융계좌추적용 압수수색영장을 집행하여 금융기관으로부터 제공받는다. 제공되는 항목에는 입출구분, 계좌주, 금융기관명, 계좌번호, 거래일자, 거래시간, 거래구분, 취급점, 입금·출금 적요, 거래금액, 거래 후 잔액, 상대은행, 상대예금주, 상대계좌번호, 단말번호, IP 주소, MAC 주소 등이 포함된다. 요청 대상 계좌번호가 두 개(A, B) 이상인 경우, A의 은행에서 제공받은 A→B 출금 내역과 B의 은행에서 제공받은 B←A 입금 내역이 동일 거래임에도 불구하고 각각 독립적으로 기록된다. 개별 계좌를 분석할 때는 문제가 없지만, 두 계좌 간 자금 흐름을 파악하기 위해 표준화 후 병합하면 동일 거래가 이중으로 남게 된다. 통신 데이터의 경우 종료일시와 서비스 구분을 제외하면 대부분 항목이 일치하지만, 계좌 거래 데이터는 계좌번호와 거래일자를 제외하고 대부분 항목이 일치하지 않는다. 동일 거래라도 수수료 부과 여부에 따라 거래금액이 일치하지 않을 수 있다. 추가로 송금 후 입금이 이루어져야 함에도 각 금융기관의 로그 생성 시점 차이로 인해 출금 시각이 입금 시각보다 늦게 기록되는 경우도 발생할 수 있다. 예를 들어, <표2>에서 2011년 5월 6일의 동일 거래는 출금 시각이 11:44:06이고 입금 시각이 11:44:00으로, 시간상 역전된 형태로 나타나기도 한다.

<표2> 계좌 데이터 중복값 예시

구분	계좌주	금융기관	계좌번호	거래일자	거래시간	거래구분	취급점	입금적요	입금액	출금적요	출금액	상대계좌번호
출금	박OO	우체국	7004300000000	2011-04-25	08:17:00	전자금융타행이체	0310431		-	박OO	3,000,000	043400000000
입금	박OO	대구은행	043400000000	2011-04-25	08:17:17	폰뱅킹	청평우체국	체신박OO	3,000,000		-	7004300000000
출금	박OO	대구은행	043400000000	2011-05-06	11:44:06	인터넷	고객감동센터		-	체신박OO	2,000,600	7004300000000
입금	박OO	우체국	7004300000000	2011-05-06	11:44:00	전자금융타발타행이체	0310431	박OO	2,000,000		-	043400000000
출금	박OO	우체국	7004300000000	2011-08-22	11:56:00	전자금융타행이체	0724399		-	박OO	1,000,000	043400000000
입금	박OO	대구은행	043400000000	2011-08-22	11:56:46	폰뱅킹	청평우체국	체신박OO	1,000,000		-	7004300000000
출금	박OO	우체국	7004300000000	2011-09-26	09:55:00	전자금융타행이체	0724399		-	박OO	470,000	043410390343
입금	박OO	대구은행	043400000000	2011-09-26	09:55:09	폰뱅킹	청평우체국	체신박OO	470,000		-	7004300000000

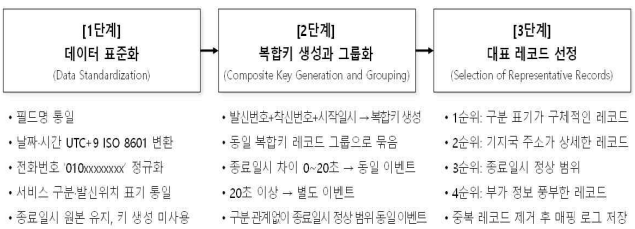
이러한 중복은 거래 데이터의 구조적 특성에서 비롯된다. 금융기관은 자금 이체 발생 시 송금자와 수취자 각각의 계좌 원장에 독립적으로 기록을 남기며, 수사기관은 각 계좌의 거래 내역을 별도로 수집한다. 특히 여러 금융기관에서 데이터를 제공받는 경우, 기관별 파일 포맷과 필드 구성, 날짜·시간 표기 방식이 달라 단순 비교만으로 동일 거래 여부를 판별하기 어렵다. 이러한 중복은 금전 흐름 네트워크에서 링크 수를 증가시키고, 총 거래금액과 거래 횟수를 실제보다 과대 산출한다. 그 결과 자금 추적의 정확성이 저하되고 수사 자원이 불필요하게 소모되며 분석 전반의 신뢰성이 떨어진다.

Ⅳ. 데이터 중복 식별을 위한 프레임워크

1. 통신데이터 중복 식별 알고리즘 설계

통신 데이터 중복 문제를 해결하기 위해 발신번호·착신번호·시작일시를 핵심 필드로 한 복합키를 중심으로 중복을 식별하되, 종료일시는 복합키에 포함하지 않고 검증 단계에서만 활용하는 알고리즘을 설계하였다. 1단계는 데이터 표준화이다. 서로 다른 사업자에서 제공받은 데이터의 필드명을 통일하고, 날짜·시간은 UTC+9 기준 ISO 8601(YYY-MM-DD hh:mm:ss) 형식으로 변환한다. 발신번호와 착신번호는 모든 구분기호를 제거하고 숫자로만 구성된 '010xxxxxxxx' 형태로 정규화한다. 서비스 구분과 발신위치는 대소문자, 공백, 약어를 일관되게 변환하고, 종료일시는 원본값을 유지하되 식별 키 생성에는 사용하지 않았다. 2단계는 복합키 생성과 그룹화이다. 발신번호, 착신번호, 시작일시 세 필드를 결합하여 복합키를 생성하고, 동일 복합키를 가진 레코드를 하나의 그룹으로 묶었다. 그룹 내에서는 종료일시를 비교하여 0~20초 범위의 차이는 동일 이벤트로 간주하고, 20초 이상의 차이는 별도 이벤트로 재분류한다. 이러한 기준은 통화 종료 기록이 사업자별로 최대 수십 초 차이가 날 수 있지만, 20초 이상 차이는 데이터 매핑 오류 가능성이 높다는 실무 분석 결과에 따른 것이다. 서비스 구분 값이 서로 다르더라도 종료일시가 정상 범위에 있으면 동일 이벤트로 처리하였다. 3단계는 대표 레코드 선정이다. 동일 이벤트 그룹 내에서 대표 레코드를 선정하기 위해 우선순위 규칙을 적용하였다. 첫째, 서비스 구분 표기가 구체적인 레코드를 우선 보존하였다. 둘째, 기지국 주소가 상세하게 기록된 레코드를 보존하였다. 셋째, 종료일시가 정상 범위에 있는 레코드를 우선하였다. 넷째, 부가 필드(비고란)의 정보가 풍부한 레코드를 보존 대상으로 선정하였다. 나머지 레코드는 중복으로 분류하고 중복 레코드와 대표 레코드 간 매핑 정보를 별도의 로그 파일로 저장하여 후속 검증과 복원에 활용할 수 있도록 하였다. 이 알고리즘은 네트워크 분석에서 동일 노드 간 불필요한 링크를 줄이고, 시계열 분석에서 특정 시점의 통화량이 과대 표시되는 현상을 완화한다. 특정 인물의 통신량이 실제보다 부풀려지는 오류를 방지하여 범죄 연루 가능성 평가와 공범 관계 추정의 정확성을 높인다.

<그림1> 통신 데이터 중복 식별 알고리즘

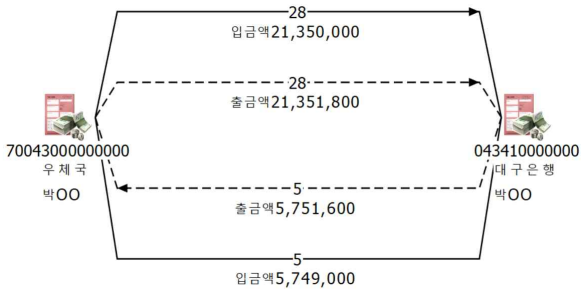


2. 계좌데이터 중복식별 알고리즘 설계

계좌 거래 데이터의 중복은 다계좌 병합 분석 과정에서 거래 건수와 금액 합계를 왜곡시켜 분석 정확성을 저하시킨다. 이를 해소하기 위해 무방향 계좌쌍을 기반으로 한 중복 식별 알고리즘을 설계하였다. 1단계는 데이터 표준화이다. 여러 금융기관에서 제공받은 데이터의 필드명을 통일하고, 거래일시는 UTC+9 기준 ISO 8601(YYYY-MM-DD hh:mm:ss) 형식

으로 변환한다. 계좌번호는 모든 구분기호를 제거하고 숫자만 남도록 정규화하며, 입출구분은 '입금'과 '출금'으로 통일한다. 거래구분·적요·취급점 표기는 대소문자, 공백, 약어를 일관되게 변환하여 비교 가능성을 높인다. 2단계는 무방향 계좌쌍 생성과 그룹화이다. 출금계좌번호와 입금계좌번호를 비교하여 작은 값을 앞에, 큰 값을 뒤에 두는 방식으로 정렬함으로써 방향성을 제거한다. 동일 계좌쌍을 하나의 그룹으로 묶고, 거래일시와 거래금액은 그룹 키에서 제외하되, 이후 비교 단계에서 보조 조건으로 활용한다. 3단계는 그룹 내 순차 비교와 후보군 생성이다. 동일 계좌쌍 그룹 내 거래를 시간순으로 정렬한 후 인접 거래를 비교한다. 거래일시 차이가 ±60초 이내이고 거래금액 차이가 허용오차(예: ±3,000원, 수수료) 범위 안에 있으면 후보군으로 지정한다. 이 과정에서 상대계좌와 상대은행의 교차 일치 여부를 우선 확인하며, 거래금액, 거래구분·채널 표준화 일치 여부, 적요·취급점의 유사도를 함께 평가한다. 4단계는 수수료 포함 여부 판별이다. 거래구분에서 이체 행위를 식별한 뒤, '전자금융타행이체', '폰뱅킹' 등 타행송금에 해당하는 경우 수수료 부과 가능성이 높다고 판단한다. 이어서 후보군의 출금액과 입금액 차이를 계산하여, 400원, 500원, 1,000원 등 사전에 정의된(은행 거래구분별 수수료 데이터 확보) 소액 패턴과 일치하면 수수료로 판정한다. 5단계는 매칭 결정과 대표 레코드 선정이다. 후보군 내에서는 출금 집합과 입금 집합 간 최소 비용 매칭을 수행하여 1:1 대응을 확정한다. 다대일·일대다 매칭의 경우 금액 부분합 검증을 통해 일치 여부를 확인한다. 대표 레코드는 출금에서 입금 방향으로 정규화하며, 순금액은 입금액, 수수료는 출금액과 입금액의 차이로 산출한다. 거래구분·채널은 구체적인 표기를 우선 채택하고, 적요·취급점은 정보량이 많은 항목을 보존한다. 6단계는 시각화와 중복 링크 표기이다. 네트워크 분석 시 실거래는 실선 링크로, 중복 거래는 점선 링크로 표시하여 분석자가 직관적으로 구분할 수 있도록 한다. 점선 링크에는 해당 거래의 출금액과 입금액을 함께 표기하고, 실선과 점선을 한 쌍으로 배치해 금액 차이와 중복 여부를 한눈에 확인할 수 있게 한다. 이 방식은 중복 거래를 단순히 제거하는 대신 시각적으로 보존함으로써 재현성과 투명성을 확보하며, 법정 제출 자료나 보고서에서 데이터 처리 과정을 설명하는 근거로 활용할 수 있다.

<그림2> 계좌 데이터 시각화 및 중복 링크 표기



7단계는 로그 기록 및 예외 처리이다. 대표 레코드와 원본 레코드 간 매핑 정보를 별도의 로그 파일에 저장하여 검증과 복원에 활용할 수 있도록 한다. 거래일시·거래금액이 허용 범위를 초과하거나 교차 일치가 확인되지 않는 건은 예외 큐로 이관하여, 보조 메타데이터 확보 또는 수동 검증 후 재처리한다. 이러한 로그는 추후 재분석이나 패턴 학습(머신러닝 기반 사기 탐지)에도 활용 가능하다. 이 알고리즘은 동일 거래의 중복 반영을 방지하여 다계좌 병합 분석 시 네트워크 분석 왜곡을 최소화하고, 거래량 및 중심성 지표의 과대 계상을 완화한다. 수수료로 인한 송금액 부풀림 오류를 방지하고, 점선 표기를 통해 중복 거래 정보를 투명하게 전달함으로써 범죄 연루 가능성 평가와 공범 구조 추정의 정확성을 높인다.

<그림3> 계좌 데이터 중복 식별 알고리즘

V. 결론

이 연구는 수사 과정에서 수집되는 통신 데이터와 계좌 거래 데이터의 병합·분석 시 발생하는 구조적 중복 문제를 실증적으로 분석하고, 이를 식별하기 위한 알고리즘 기반 처리 프레임워크를 제안하였다. 중복 데이터는 네트워크 분석과 시계열 분석 등 정량 분석에서 연결 구조와 빈도 지표를 왜곡하여 수사 판단의 신뢰성을 저하시킨다. 통신 데이터에서는 발신·착신 번호와 시작일시를 핵심 식별 요소로 한 복합키 설계와 종료일시 허용 범위 검증을 통해, 계좌 거래 데이터에서는 무방향 계좌쌍 생성과 금액·시간 차이 기준을 활용한 후보군 매칭 절차를 통해 중복을 효과적으로 식별·제거할 수 있음을 확인하였다. 제안한 프레임워크는 표준화, 복합키 생성, 그룹화, 대표 레코드 선정, 시각화 및 로그 기록 일련 과정을 구조화하여, 데이터 처리의 재현성과 투명성을 동시에 확보하였다. 특히 중복 거래를 단순히 삭제하는 것이 아니라, 시각적 표기와 로그 기록을 병행함으로써 분석 과정의 검증 가능성을 높였다. 이는 법정 제출 증거나 수사 보고서 작성 시 데이터 신뢰성을 뒷받침하는 근거로 기능할 수 있다. 제안된 방법은 수사기관이 보유한 표준화 프로그램의 기능 한계를 보완하고, 대규모 다원적 데이터 환경에서 분석 정확성을 향상시키는 실무적 대안을 제시한다. 정책적으로는 표준화 프로그램(i1, A1)의 기능 개선과 법령·내규에 데이터 표준화 및 중복 식별 가이드라인을 반영하는 것이 필요하다. 이를 통해 기관 간 데이터 호환성을 높이고, 수사자료의 신뢰성을 제도적으로 보장할 수 있다. 향후 연구에서는 다양한 사업자·기관 간 데이터 연동 환경에서 알고리즘의 성능을 자동화·고도화하는 방안을 검토해야 한다. 특히 머신러닝 기반의 이상 탐지 모델과 결합하여, 중복 제거와 동시에 의심 거래나 통화 패턴을 실시간으로 식별하는 지능형 수사 지원 체계로 확장할 수 있을 것이다.

참 고 문 헌

[1] A. T. T. Wong, "Opportunities and challenges of big data analytics in crime investigation," International Annals of Criminology, pp. 1–15, 2025.

[2] M. S. Gal and D. L. Rubinfeld, "Data standardization," New York University Law Review, vol. 94, p. 737, 2019.

[3] L. W. Lentz and N. Sunde, "The use of historical call data records as evidence in the criminal justice system—lessons learned from the Danish telecom scandal," Digital Evidence and Electronic Signature Law Review, vol. 18, p. 1, 2021.

COVID-19 감염경로 분석을 위한 역학조사서 전처리 방법론에 관한 연구

김시원, 김지온*

한림대학교

siwonkim@hallym.ac.kr, *jion972@hallym.ac.kr

A Study on Preprocessing Methodology for Epidemiological Investigation Data to Analyze COVID-19 Transmission Routes

Siwon Kim, Jion Kim*

Hallym Univ.

요 약

본 논문은 반정형 형태로 기록된 역학조사서를 전처리하여 확진자-접촉자 연결망을 구축하고, 이를 네트워크 분석에 적용한 과정을 제시한다. 역학조사서는 일정한 필드 구조를 갖추고 있으나, 실제 자료는 작성자와 시기마다 기록 방식이 달라 데이터의 불일치 및 결측이 빈번하게 발생한다. 이를 해결하기 위해 개인 식별자 부여, 역할 전환 처리, 필드 표준화 등의 전처리 과정을 수행하였다. 전처리 결과를 기반으로 관계망을 시각화하고, 중심성 지표를 산출하여 주요 전파자와 매개자를 규명할 수 있었다. 이러한 반정형 데이터 전처리 방법론은 전염병 기반의 역학 연구와 더불어 관계정보를 다루는 다양한 연구 분야에도 적용될 수 있다.

Ⅰ. 서 론

사회 연결망 분석(Social Network Analysis) 원리는 개인 간 상호작용을 구조적으로 이해하고, 사건이나 현상의 확산 경로를 규명하는 데 중요한 방법론이다 [1]. 특히, 감염병 역학 분야에서는 확진자와 접촉자 간의 관계를 네트워크로 전환함으로써 감염경로, 전파자의 특성, 슈퍼 전파자 존재 여부, 전파 경로의 특성을 밝히는 데 기여해왔다.

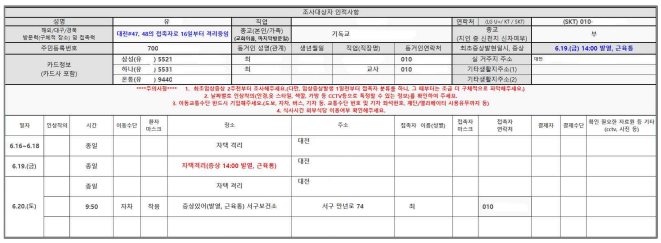
그러나 이러한 감염경로 추적을 위한 분석의 성패는 원본 자료의 품질과 정형화 수준에 크게 의존한다. 대부분의 역학조사서는 일정한 필드 구조를 갖춘 반정형(semi-structured) 데이터로 수집되지만, 실제 현장에서는 작성자, 시기, 기관마다 기록 방식이 상이하여 데이터의 일관성이 부족한 경우가 많다 [1]. 역학조사서는 성명, 성별, 연령 등의 정형 필드와 증상 발현일, 감염 경위와 같은 자유 기입식 서술형 텍스트가 혼재된 형태로, 동일한 항목이 문서마다 다르게 표현되거나 일부 필드가 공란으로 남겨지는 사례가 빈번하다. 반정형 데이터의 이러한 비표준적인 특성은 특히 의료 및 역학 분야에서 데이터 전처리 과정에 소모되는 시간을 늘리고 복잡하게 만들며 연결망 분석의 정확성을 저해하는 주요 요인이 된다 [2]. 따라서 원본 데이터를 정제하고 관계정보의 추출이 가능한 형태로 재구성하기 위한 전처리 과정이 필수적이다. 그러나 기존 연구들이 대부분 정형화된 자료에 기반을 두고 있어, 실제 현장의 반정형 데이터를 분석 가능한 형태로 전환하는 과정에 대한 논의는 상대적으로 미흡한 실정이다 [3].

이에 본 연구는 COVID-19 역학조사서를 사례로 반정형 형태의 역학조사서를 관계분석이 가능한 형태로 전처리하여 연결망 분석 가능한 데이터셋으로 정제하는 과정을 제시하고자 한다. 이러한 전처리 방법은 전처리 역학 분야를 넘어, 사회 조사, 범죄 수사 기록 등 비정형 또는 반정형 데이터가 빈번히 발생하는 다양한 분야에 확장 적용될 수 있는 잠재력을 탐색하고자 한다.

Ⅱ. 연구 방법

본 연구는 역학조사서 원본 데이터를 연결망 분석에 활용하기 위해 원본 자료→ 1차 전처리 → 2차 전처리의 세 단계 과정을 거쳤다.

1) 원본 데이터 (심층 역학조사서)



조사대상자 인적사항

성명	성	직업			연락처	(SKT) 010
확진자구분	대전#47, 48의 접촉자로 16일부터 격리중임	종교(본인/가족)	기독교		증상	부
발문확구분자 있으면 및 접촉력		(무직자는 학기업분일)			(1일 이상 시점지 상자 여부)	
주민등록번호	700	증거인 성명(관계)	생년월일	직업(직장명)	증거인연락처	최초증상발현일시, 증상
카드정보 (카드사 포함)	성성(무) 5521	희			010	실 거주지 주소
	하나(무) 5531	희		교사	010	기타생활지주소(1)
	롯데(무) 9440					기타생활지주소(2)

일자	진술확보	시간	이동수단	동선 마스크	장소	주소	접촉자 이름(성별)	접촉자 마스크	접촉자 연락처	클래스	클래스수업	확인 필요한 자료명 등 기타 (cctv, 사진 등)
6.16~6.18		종일			자택 격리	대전						
6.19.(금)		종일			자택격리(증상 14:00 발열, 근육통)	대전						
6.20.(토)		종일			자택 격리	대전						
		9:50	자차	착용	증상없이(발열, 근육통) 서구보건소	서구 만년로 74	희		010			

〈그림 1〉 비식별화 완료된 심층 역학조사서 원본 데이터

원본 데이터는 정형 필드와 함께 자유 서술형 텍스트가 혼재된 반정형 형태를 가지고 있으며, 이러한 특성으로 인해 규칙 기반 자동 파싱만으로는 일관된 값을 안정적으로 추출하기에 어려움이 존재한다. 주요 원인으로는 ① 형식의 비일관성 ② 날짜·장소·시간 등의 값 표현의 자유도와 다수의 이형(異形) 존재 ③ 동일 인물의 중복 기재나 확진자·접촉자의 역할 반복 등 사람의 의미적 판독 필요 ④ 정확도를 위한 분석 규칙 적용 등이 있다. 따라서 1차 단계에서는 사람이 직접 항목을 판독하고 핵심 필드를 정리하는 수작업 전처리가 불가피했으며, 이후 2차 단계에서 코드화된 규칙을 적용해 관계정보를 구조화 및 검증하는 이중 처리 전략을 채택했다.

2) 1차 전처리

구분	컬럼명	설명
확진자 정보	성명	확진자 이름
	전화번호	확진자 연락처
	생년	출생년도
	성별	남/여
	주소	거주지 주소
	증상 발현일	최초 증상 발현 시점
	확진일	확진 판정일
동선 및 접촉자 정보	일자	해당 동선 발생일
	시간	방문/노출 시간
	장소	방문 장소명
	업종	장소의 업종 분류
	주소	방문 장소 주소
	확진자 마스크 착용 여부	O/X
	접촉자 이름	-
	접촉자 전화번호	-
	접촉자 성별	-
	접촉자 생년	-
	접촉자 주소	-
	접촉자 마스크 착용 여부	-

〈표 1〉 1차 전처리 후 데이터 컬럼 구성

원본 역학조사서에서 분석과 무관한 법령, 작성 지침, 중복 서술을 제거한 뒤, 분석에 필요한 항목만을 추출하여 총 19개 컬럼을 가진 표 형태로 정리하였다. 성명, 성별, 생년, 주소, 증상 발현일, 확진일 등 기본 필드를 선별하고, 문서마다 달리 기재된 날짜와 장소 표현은 단순화된 규칙에 따라 통일된 형식으로 정리하였다. 예를 들어, '6월 24일 16시경'과 같은 서술식 기록은 '2020-06-24 16:00'과 같이 표준화하였다. 이를 통해 불규칙한 반정형 자료를 연결망 분석이 가능한 구조로 가공하였다.

3) 2차 전처리

컬럼명	설명	예시
person_id	인물 고유 식별자	P000001
person_name	인물 이름(익명 처리)	이OO
role	확진자/접촉자 구분	confirmed
sex	성별	F
birth_year	출생년도	1997
onset_dt	증상 발현일	2020-02-17
diagnosis_dt	확진일	2020-02-21
district_code	거주 지역 코드	25
conversion_flag	접촉자→확진자 전환 여부	0

〈표 2〉 최종 산출 데이터 구조 persons 시트 예시

1차 전처리로 정리된 데이터를 기반으로, 코드화된 규칙을 적용하여 관계정보를 추출하기 위한 데이터 형태를 구조화하였다. 동일 인물이 문서 내에서 확진자와 접촉자로 반복 등장하는 경우에는 하나의 엔티티로 통합하였고, 접촉 관계는 확진자 중심으로 일관되게 연결되도록 변환하였다. 또한, 데이터 누락과 중복 문제를 최소화하기 위해 표현이 다른 값은 동일한 표준 형식으로 변환하며 이를 통해 연결망 구축 과정에서 발생할 수 있는 불필요한 노드 분산이나 잘못된 연결을 방지하였다.

최종적으로 산출된 데이터 시트는 persons, locations, edges의 세 개의 시트로 나뉘며 확진자와 접촉자를 구분하는 person_id, role 컬럼을 중심으로 접촉이 발생한 날짜(date), 장소(locations), 접촉자 식별자(contact_id) 및 변환 과정에서 부여된 변수값(conversion_flag, mask_flag) 등으로 구성된다. 이를 통해 각 행은 '확진자-접촉자 쌍'을 의미하며, 접촉 시점과 장소 정보가 결합되면서 네트워크의 엣지(edge) 정보로 변환될 수 있도록 구조화했다.

Ⅲ. 분석 및 결과

대전지역의 심층 역학조사서 200건을 데이터셋으로 사용하여 확진자-접촉자 네트워크 구축 및 분석을 진행하였다. 전처리를 거쳐 정제된 데이터셋은 총 200명의 확진자와 1,851명의 접촉자로 구성되었으며, 이들 간의 관계는 2,231건의 엣지(edge)로 표현된다.

1. 네트워크 기본 통계

확진자-접촉자 네트워크는 노드 수, 엣지 수, 평균 차수, 네트워크 밀도, 약한 연결 집합 수, 최대 연결 집합 크기, 노드 간 최대·최소 거리 등을 통해 구조적 특성을 파악하였다.

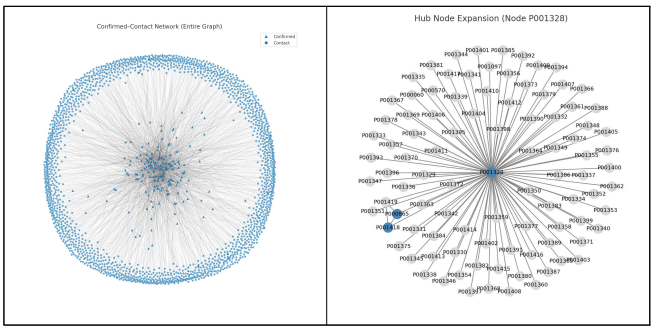
지표	값
노드 수	2,051
엣지 수	2,231
평균 차수(average degree)	2.12
네트워크 밀도	0.001
약한 연결 집합 수	44
최대 연결 집합 크기	1,698
노드 간 최대 거리	16
노드 간 최소 거리	1
노드 간 평균 거리	7.3

〈표 3〉 확진자-접촉자 네트워크의 기본 통계

평균 차수가 약 2라는 것은 대부분의 개인이 1~2명 수준의 제한적인 접촉 관계만 맺고 있다는 것을 의미한다. 그러나 일부 노드는 다수와 연결되어 있어, 전파가 특정 집단이나 개인에 의해 집중적으로 이루어졌음을 보여준다. 네트워크 밀도가 0.001로 낮게 나타난 것은 모든 사람이 서로 연결된 촘촘한 구조가 아니라, 허브 노드를 중심으로 방사형으로 뻗어 나가면서 주변 이웃 노드 간의 연결은 거의 없는 느슨한 네트워크라는 것을 의미한다. 이러한 구조적 특성은 수많은 개별 노드들이 산발적으로 분리되어 있으면서도, 동시에 슈퍼 전파자와 같은 소수 핵심 노드가 전체 확산을 좌우할 수 있는 위치에 있다는 것을 의미한다. 실제로 44개의 연결 집합은 지역 내 다수의 소규모 전파 집단이 존재함을 반영하며, 그중 최대 연결 집합이 전체의 약 82.8%를 포함하는 대규모 구조로 나타난 것은 특정 사건이나 집단이 지역 내에서 지배적인 영향력을 행사했음을 보여준다.

더불어 네트워크의 최소 거리가 1, 최대거리가 16으로 나타나는 것은 한쪽 끝의 개인에서 다른 쪽 끝의 개인까지 최장 16단계를 거쳐 연결될 수 있다는 것을 의미한다. 이는 전파 대부분이 짧은 거리에서 발생하지만, 동시에 일부 경로는 장기간 연쇄적으로 이어져 감염병이 확산될 수 있다는 구조적 가능성을 내포한다.

2. 확진자-접촉자 네트워크 구조



〈그림 1〉 확진자-접촉자 네트워크 전체 구조(왼)와 주요 허브 노드 확대(오)

앞서 확인된 바와 같이, 본 연구를 통해 도출한 대전지역 COVID-19 감염경로 네트워크는 낮은 밀도와 중심부에 다수의 소규모 연결 집합을 특징으로 한다. 이러한 구조적 특성은 시각화를 통해 명확하게 드러난다. 네트워크 주변부에 중심부와 연결된 수많은 개별 노드들이 분포되어 있으면서, 중심부에 소규모 허브 노드들의 하위집단이 감염네트워크의 코어를 구성하고 있다. 소규모 집단들은 주로 가족이나 지인 모임과 같은 밀접 생활 단위, 직장·학교·종교시설 등 특정 공간을 공유하는 단위, 혹은 특정 시점에 발생한 사건 단위로 형성되었다.

그림 1의 왼쪽 그래프는 전체 확진자-접촉자 네트워크를 나타낸다. 네트워크는 중심부에 연결 중심성이 높은 몇 명의 확진자를 중심으로 뭉쳐 있으며, 주변부에는 중심부와 연결된 다수의 개별 노드들이 분산되어 존재하는 형태를 보인다. 이는 대다수 개인이 제한적인 접촉 관계만 맺고 있었으나, 특정 확진자를 통해 네트워크가 전반적 연결이 일어났다는 것을 보여준다.

그림 1의 오른쪽 그래프는 대표적인 허브 노드인 P001328의 연결구조를 확대한 것이다. 해당 확진자는 95명의 접촉자와 직접 연결되어 있어 네트워크 내에서 스타형 구조를 형성하며, 가장 높은 연결 중심성을 보였다. 이는 감염병 전파에서 소수의 슈퍼 전파자 혹은 특정 사건이 전체 연결망 구조를 결정하는 데 큰 영향을 끼친다는 것을 의미한다.

Ⅳ. 결론 및 시사점

본 논문에서는 대전지역의 심층 역학조사서 200건을 기반으로 확진자-접촉자 네트워크를 구축하고, 이를 통해 감염병 전파 구조를 분석하였다.

첫째로, 반정형 형태의 역학조사서에 포함된 관계정보를 시각화 하여 감염경로 연결망을 분석하기 위해 2차에 걸친 전처리 방법론을 제시하였다.

둘째로 관계형 데이터로 전처리한 역학조사 데이터셋을 네트워크 형태로 시각화하여 분석한 결과, 대전지역 COVID-19 확진자 연결망은 주변부에 서로 연결되어 있지 않은 다수의 개별 노드들로 분산되어 있으면서도, 특정 사건이나 슈퍼 전파자에 의해 중심부에 소규모 집단이 형성되는 구조를 보였다. 이는 감염병 전파가 소수 핵심 노드에 의해 급격히 확산될 수 있음을 시사한다.

본 연구는 역학조사서와 같은 비정형·반정형 데이터를 네트워크 데이터로 전환할 수 있음을 입증하였으며, 이를 통해 전파 구조의 시각화, 주요 확산 경로 식별 등 감염경로 기반의 다양한 분석 가능성을 제시하였다. 나아가 제안된 전처리 방법론은 감염병 역학조사뿐만 아니라 반정형 형태의 보고서를 활용하는 다양한 분야에 적용될 수 있을 것으로 기대한다. 다만 향후 후속연구를 통해 전처리와 연결망 데이터 구축 과정을 LLM 등 AI 기술을 활용하여 최대한 자동화할 수 있어야만 이와 같은 분석 방법론을 효과적으로 실무에 적용할 수 있을 것이다.

ACKNOWLEDGMENT

이 논문은 2025년도 정부(경찰청)의 재원으로 과학치안진흥센터의 지원을 받아 수행된 연구임 (과제번호: RS-2025-02218280, 사이버범죄 수사단서 통합분석 및 추론 시스템 개발 사업).

참 고 문 헌

[1] Wasserman, Stanley, and Katherine Faust. 1994. Social Network Analysis: Methods and Applications. Cambridge University Press.

[2] 이승희, 이지원, 김연주, 김상희 and 조수현. "역학조사관의 감염병 대응 경험: 질병관리청 소속 역학조사관을 대상으로 한 포커스 그룹 인터뷰" 보건행정학회지 34, no.4 (2024) : 440-449.doi: 10.4332/KJHPA.2024.34.4.440

[3] Amiruddin, N. A., Wahid, N. A. A., & Yasin, M. M. 2024. "Pre-processing Approach for Semi-Structured Medical Data." In Proceedings of the International Conference on Innovation & Entrepreneurship in Computing, Engineering & Science Education (InvENT 2024). Atlantis Press. https://www.atlantis-press.com/proceedings/invent-24/126005585.

[4] Jo, W., Chang, D., You, M. et al. A social network analysis of the spread of COVID-19 in South Korea and policy implications. Sci Rep 11, 8581 (2021). https://doi.org/10.1038/s41598-021-87837-0

AI 기반 Web Agent 를 활용한 불법 데이터 자동 탐지 및 수집 연구

이수정, 이요셉*, 조용성**, 송경우
연세대학교, *숭실대학교, **한국전자통신연구원

drlisa@yonsei.ac.kr, *dytpq0916@soongsil.ac.kr, **yscho73@etri.re.kr,
kyungwoo.song@yonsei.ac.kr

Web Agent Approach to Automated Monitoring and Collection of Illegal Data

Lee Soojeong, Lee Joseph*, Cho Yongseong**, Song Kyungwoo
Yonsei Univ., *Soongsil Univ., **Electronics and Telecommunication Research Institute (ETRI)

요 약

사이버 범죄는 불법 웹사이트와 다크웹을 중심으로 빠르게 확산되고 있으나, 기존의 데이터 수집 기법은 변화하는 환경에 대응하기 어렵다. 본 연구는 이러한 한계를 극복하기 위해 AI 기반 Web agent 를 활용한 자동 탐지 및 데이터 수집 방법을 제안한다. Web agent 는 웹 환경을 인식하고 계획하며 실행하는 과정을 통해 자율적으로 탐색과 수집을 수행하며, 텍스트, 이미지, 영상 등 멀티모달 데이터를 실시간으로 처리할 수 있다. 제안된 체계는 불법 데이터 탐지, 수집, 불법 유형별 분류의 단계로 구성되어 사이버 범죄 단서 확보를 자동화한다. 이를 통해 수사 기관은 불법 행위의 패턴을 효과적으로 파악하고, 정책적 대응과 집행 전략을 수립하는 데 활용할 수 있다.

Ⅰ. 서 론

사이버 범죄의 종류가 날로 다양해지고, 범죄 발생 건수 또한 급격히 증가함에 따라 사이버 범죄에 대한 단서 수집을 사람이 수동적으로 수행하기에 많은 인력을 필요로 하고 있다. 특히 불법 웹사이트와 다크웹 환경은 수사망을 회피하기 위해 구조와 운영 방식을 지속적으로 변화시키고 있어, 기존의 수작업 기반 단서 수집이나 단순 크롤링 방식으로는 효율적인 대응이 어렵다. 이러한 문제를 해결하기 위해 불법 데이터를 자동으로 점검하고, 구조적으로 수집 할 수 있는 체계가 필요하다.

이에 본 연구에서는 변화하는 환경에 동적으로 대응할 수 있는 AI agent 기반 방법론을 제안한다. Web agent[1]는 다크웹을 포함한 웹 환경에서 자율적으로 탐색, 수집, 분석을 수행하는 소프트웨어 에이전트로, 텍스트, 이미지, 영상 등 멀티모달 데이터를 다루면서도 크롤링과 같은 기존 방식보다 빠른 속도로 데이터를 수집하여 실시간으로 사용 가능하다는 장점을 가진다[2]. 이러한 특성을 활용하면 불법 데이터를 효과적으로 탐지하고 수집할 수 있으며, 더 나아가 수집된 데이터를 정리하고 구조화함으로써 사이버 범죄 수사를 위한 단서 제공과 분석 지원이 가능하다. 본 연구는 Web agent 를 기반으로 한 자동 탐지 및 데이터 수집 방법을 활용하며, 수사 기관이 효율적으로 불법 행위의 패턴을 파악하고 대응 전략을 마련할 수 있도록 돕는다.

Ⅱ. 본론

가. Web agent 정의 및 특성

Web agent 는 웹 환경에서 사용자의 명령을 독립적으로 수행하는 지능형 소프트웨어 에이전트로 정의된다. 기존의 단순한 매크로나 스크립트 기반 자동화와 달리, Web agent 는 복잡하고 동적인 웹 페이지 구조를 이해하고, 주어진 목표를 달성하기 위한 합리적인 행동을 스스로 선택하고 실행할 수 있다는 점에서 차별화된다[3]. 예를 들어, 사용자가 “회의 일정을 잡아줘”라고 지시하면, Web agent 는 이메일 애플리케이션을 열고, 수신자 주소를 탐색하며, 메일을 작성하고 전송하는 일련의 과정을 자동으로 처리할 수 있다. 이러한 과정은 기존 챗봇처럼 지속적인 사용자 개입에 의존하지 않고 독립적으로 진행된다는 점에서 의의가 크다.

Web agent 의 주요 특성은 크게 세 가지로 요약할 수 있다. 첫째, 사용자 편의성이다. 반복적이고 시간이 많이 소요되는 온라인 예약, 정보 검색과 같은 작업을 자동화하여 사용자의 부담을 경감한다. 둘째, 복잡한 업무 수행 능력이다. 단순히 단일 명령만 처리하는 것이 아니라, 다단계 상호작용과 장기적 목표 달성이 가능하다는 점에서 기존 매크로 기반 자동화 도구와 구별된다. 셋째, 확장성과 일반화 가능성이다. 특정 웹사이트나 도메인에 종속되지 않고 다양한 웹 환경에서 일반화된 행동을 수행할 수 있어, 보다 광범위한 응용 가능성을 지닌다. 이러한 특성은 Web agent 가 지능형 개인 비서, 자동 데이터 수집, 온라인 협업 지원 등 다양한 분야에서 활용될 수 있는 근거가 되며, 궁극적으로 사용자의 웹 활용 경험을 향상시킨다.

나. Web agent 방법론

Web agent 는 웹 환경에서 명령을 수행하기 위해 “인식 – 계획 및 추론 – 실행”의 세 단계를 따른다[3]. 이 절에서는 각 단계의 방법론적 특징을 기술한다.

먼저, 인식 단계에서 Web agent 는 웹 환경을 이해하기 위해 정보를 수집한다. 가장 흔히 활용되는 방식은 HTML 기반의 DOM(Document Object Model) 분석과 웹페이지의 시각적 스크린샷이다. DOM 은 웹 요소의 구조적 관계를 명확히 제공하지만, 사용자가 실제로 경험하는 시각적 맥락은 충분히 반영하지 못한다. 반대로 스크린샷은 인터페이스의 실제 모습을 반영하나, 세부 구조 정보를 담기에는 한계가 있다. 최근 연구들은 이 두 가지 방식을 융합한 멀티모달 인식 기법을 활용하여 보다 정교한 웹 환경 표현을 구축하고 있으며, 이는 복잡한 사용자 상호작용을 처리하는 데 효과적이다[4].

다음으로, 계획 및 추론 단계에서는 사용자의 목표를 달성하기 위해 필요한 행동을 합리적으로 결정한다. 이 과정에는 작업을 단계적으로 분해하는 작업 분해(task decomposition), 다음 동작을 예측하는 행동 예측(action prediction), 그리고 장기적 맥락을 반영하기 위한 메모리 기반 추론(memory-based reasoning)이 포함된다. 예컨대, 특정 상품을 구매하는 경우, Web agent 는 ‘검색창 클릭 → 키워드 입력 → 결과 확인 → 상품 선택 → 결제 진행’과 같은 일련의 행위를 계획한다.

마지막으로, 실행 단계에서는 계획된 행동을 실제 웹 환경에서 수행한다. 이 과정에서 agent 는 DOM 요소를 탐색하여 클릭, 스크롤 등의 상호작용을 재현하거나, 특정 API 와의 연동을 통해 효율적으로 작업을 처리한다. Browser-use 프레임워크[5]는 이러한 실행을 지원하기 위해 Python 기반 API 호출 및 헤드리스 브라우저 제어 기능을 제공하며, 이는 웹 자동화의 대표적인 구현 방식으로 자리 잡아가고 있다.

[Figure 1] Web Agent 를 활용한 불법 데이터 탐색 및 수집 과정

다. Web agent 기반 불법 데이터 탐지 및 수집

앞 절에서 Web Agent 의 구조와 방법론을 고찰하였다. 본 절에서는 이를 토대로 불법 데이터 대응 문제에 적용 가능한 방안을 제시하며, 해당 과정을 데이터 탐지, 수집, 분류의 세 단계로 구분하였다.

첫째, 데이터 탐지 단계에서는 단순 키워드 검색을 넘어, Web Agent 가 프롬프트 기반 BFS 탐색 기법을 활용하여 다크웹을 포함한 웹 환경을 체계적으로 계획하여 추론한다. 예를 들어 ‘스포츠 베팅’과 ‘최신 영화’ 등 불법 콘텐츠 관련 키워드를 기반으로, BFS 를 사용하여 한 페이지 내 하위 링크를 최대 depth 2 까지 구조적으로 탐색하고, 추가 콘텐츠를 위해 다음 페이지 이동도 자동으로 처리한다. 이를 통해 단일 URL 에 국한되지 않고, 웹 전반의 잠재적 불법 데이터 확산 경로를 식별할 수 있다는 점에서 기존 단순 크롤링 기법과 차별성을 가진다.

둘째, 데이터 수집 단계에서는 텍스트, 이미지, 영상 등 다양한 멀티모달 데이터를 확보한다. 본 연구에서는 프롬프트에 결과 예시 형식을 직접 제공하여, URL 단위로 수집된 데이터를 마크다운 형식으로 구조화하였다. 나아가 다운로드 자동화 모듈을 추가하여 원본 파일(이미지, 영상 등)을 직접 저장할 수 있도록 구현하였다. 이를 통해 단순한 텍스트 크롤링을 넘어, 실제 수사기관에서 활용 가능한 원본 증거 데이터 확보가 가능해진다.

셋째, 데이터 분류 단계에서는 수집된 데이터를 불법 행위의 유형(불법 링크 공유, 불법 영상물 유통, 불법 도박 거래 등)과 불법 여부에 따라 정밀하게 분류한다. 단순 식별을 넘어, 정책적 대응 및 집행 전략 수립에 필요한 실질적 근거 데이터를 제공한다는 점에서, 본 연구는 기존 연구 대비 실무적 활용 가능성을 높인다.

Ⅲ. 결론

본 연구는 기존의 수작업 기반 데이터 수집과 단순 크롤링 기법이 가지는 한계를 극복하기 위해, AI 기반 Web agent 를 활용한 자동 탐지 및 데이터 수집 체계를 제안하였다. 제안된 방법은 웹 환경을 인식, 계획, 실행하는 과정을 통해 멀티모달 데이터를 실시간으로 확보하고, 이를 불법 데이터 탐지와 분류에 적용할 수 있음을 보였다. 이러한 체계는 변화하는 불법 웹사이트 및 다크웹 환경에서 안정적으로 동작할 수 있어, 수사기관이 효율적으로 단서를 확보하고 분석할 수 있는 기반을 마련한다는 점에서 의의가 크다.

ACKNOWLEDGEMENT

이 논문은 2025 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행되었으며 (No. RS-2023-00224740, 디지털성범죄 피해 예방을 위한 불법촬영물 이미지 유포 차단 및 추적 기술 개발), 또한 2025 년도 정부(경찰청)*의 재원으로 과학치안진흥센터 사이버범죄 수사단서 통합분석 및 추론시스템 개발 사업의 지원을 받아 수행된 연구임(RS-2025-02218280).

참 고 문 헌

[1] Gur, Izzeddin, et al. "A real-world webagent with planning, long context understanding, and program synthesis." arXiv preprint arXiv:2307.12856 (2023).

[2] Huang, Wenhao, et al. "AutoScraper: A progressive understanding web agent for web scraper generation." arXiv preprint arXiv:2404.12753 (2024).

[3] Ning, Liangbo, et al. "A survey of webagents: Towards next-generation ai agents for web automation with large foundation models." Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2. 2025.

[4] Deng, Xiang, et al. "Mind2web: Towards a generalist agent for the web." Advances in Neural Information Processing Systems 36 (2023): 28091-28114.

[5] Browser-use Documentation, 2025, (https://docs.browser-use.com/introduction)

통신수사에서의 통화내역 분석과 SNA 기반 대포폰 탐지 알고리즘 연구

우병관, 김지온
한림대학교

byeongkwan.woo@hallym.ac.kr, jion972@hallym.ac.kr

A Study on Call Detail Record Analysis and a Social Network Analysis-based Burner Phone Detection Algorithm for Crime Investigation

Woo Byeong Kwan, Kim Jion
Hallym Univ.

요 약

본 연구는 용의자가 대포폰으로 수사망을 피하는 문제에 대응하고자 사회연결망분석(SNA) 기반의 탐지 알고리즘을 제안한다. 용의자의 전화번호는 바뀌어도 관계망과 통화 패턴은 유지된다는 점에 착안하여, 1단계로 연결 중심성 분석을 통해 용의자의 핵심 관계망을 식별한다. 2단계에서는 구조적 등위성 분석으로 기존 용의자의 역할을 대체하는 대포폰 후보 번호를 탐색한다. 마지막으로 시계열 분석을 통해 통화 시간대 등 동적 활동 패턴의 유사성을 교차 검증하여 특정한다. 실제 사건 기반의 사례 연구를 통해 본 방법론이 수사 현장에서 적용 가능한 실용적 프레임워크임을 입증하였다.

Ⅰ. 서 론

범죄 수사에서 용의자의 통신내역은 추적을 위한 핵심 단서 및 범죄를 입증하는 결정적인 증거로 활용된다. 그렇기 때문에 용의자들은 법집행기관의 추적을 피하기 위해 대포폰을 새로 개통하여 범죄를 실행하고 있으며, 이는 법집행기관의 수사에 큰 장애로 작용한다.

한편, 용의자는 대포폰을 개통하여 통화의 '수단'은 바꿀 수 있어도, 일상 생활이나 범행 실행과정에서 통화 상대방과 맺는 '관계' 또는 고유의 통화 '패턴'까지 완전히 바꾸기는 어렵다. 즉, 용의자의 전화번호만 변경될 뿐, 용의자가 맺고 있는 사회적 관계 네트워크는 그대로 유지되거나 유사한 형태로 계승된다는 것이다.

본 연구는 이러한 전제에 착안하여, 통신 네트워크의 구조적 속성과 사용자의 행동 패턴에 집중하는 새로운 대포폰 추적 알고리즘을 제안하였다. 알고리즘은 아래의 세 가지 핵심 질문에 답하는 과정으로 설계되었다.

- 용의자의 핵심 관계망은 누구로 구성되어 있는가?
- 이 관계망 내에서 용의자의 역할을 대체하는 번호는 무엇인가?
- 그 번호가 용의자의 행동 습관까지 모방하는가?

이러한 다층적 접근을 통해, 본 연구는 실제 대포폰을 높은 정확도로 특정하는 실용적이고 정교한 방법론을 제공하고자 한다.

Ⅱ. 관련 개념 및 연구

1. 통신수사와 통신사실확인자료

'통신수사'는 범죄 수사에서 범인이 사용하는 다양한 통신수단을 조사하여 필요한 증거를 확보하는 방법을 의미[1]하며, 실무적으로 통신수사는 이러한 증거 자료를 수집·분석하여 범인을 특정하고 검거하는 수사 기법으로 정의할 수 있다.

통신수사를 통해 확보할 수 있는 증거로는 통신자료(가입자 정보), 통신사실확인자료(통화내역, 발신 기지국 위치정보 등), 감청자료 등이 있는데, 본 연구는 통화내역에 SNA 알고리즘을 적용하여 대포폰을 탐지하는 알고리즘을 제안한다는 점에서 여러 유형의 증거 자료 중 통신사실확인자료를 대상자료로 삼고 있다.

2. 사회연결망분석

사회연결망분석(이하 SNA)이란 사람들의 사회적 행위를 그들이 맺은 관계로 구성된 연결망의 특성으로 설명하려는 시도[2], 사회 구성원 간의 관계에 분석의 초점을 맞추어 이들 관계의 패턴으로부터 의미있는 시사점을 도출하는 방법론[3]을 말한다. 즉, 분석 대상 고유의 '속성'이 아니라 분석 대상이 맺고 있는 '관계'에 주목하여, 전화통화 내역, SNS 전송, 송금 내역 등 주요 범죄정보를 분석하는 것이다.

경찰은 그간 SNA를 활용하여 범인의 검거·특정, 공범 식별, 범죄혐의 증명이 가능함을 확인[4]하고, SNA 중 중심성(Centrality), 등위성(Equivalence), 고유벡터(Eigenvector) 등 알고리즘을 조합하여 범죄 데이터를 분석할 수 있음을 보이는 등[5][6] 연구를 진행하였으며, 학계에서도 사이버금융범죄 수사를 위해 2-mode 범죄 데이터를 1-mode로 변환 후 SNA를 적용하는 연구가 있었다[7].

Ⅲ. SNA 기반 대포폰 탐지 알고리즘 제안

대포폰 탐지 알고리즘은 아래와 같이 4단계로 구성하였다.

1. (1단계) 네트워크 모델링

경찰이 압수수색영장을 통해 확보한 통화내역 데이터 중 '전화번호'를 노드(Node)로, 두 전화번호 간 통화 기록을 엣지(Edge)로 정의하고, 발신전화에서 착신전화로 방향성을 갖도록 모델링한다. 두 전화번호 간 통화 횟수를 가중치로 설정하되, 시간정보를 반영한 타임라인 분석을 수행하는 경우에는 엣지에 시간정보를 입력하고, 별도의 가중치를 부여하지는 않는다.

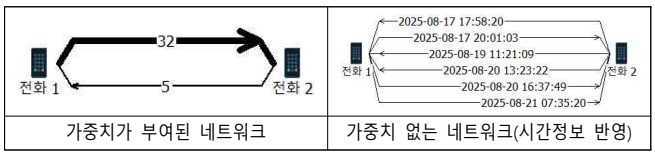


그림 1 네트워크 모델링 유형

2. (2단계) 핵심 네트워크 식별(연결 중심성 분석)

용의자는 통신 수단을 변경하더라도 기존에 맺고 있던 핵심적인 사회 관계망은 유지하는 경향이 있을 수 있으므로, 특정 노드가 네트워크 내 다른 노드들과 얼마나 많은 직접적 연결을 갖는지를 측정하

는 연결 중심성(Degree Centrality) 분석을 통해 용의자의 핵심 관계망을 식별한다.

이를 통해 먼저 용의자의 기존 번호와 가장 많이 통화한 집단, 즉 주요 의심 다수통화자를 추출하고 수사 맥락 상 중요한 의미를 갖는 대상자를 선별하여 분석의 기준인 '핵심 관계망'으로 삼고, 이들 '핵심 관계망' 번호와 많이 통화한 다수통화자를 추출하여 구조적 등위성 및 시계열 분석을 위한 대상을 특정한다.

3. (3단계) 용의자 등위 노드 탐색(구조적 등위성 분석)

용의자가 기존에 사용하던 번호를 해지하면, 해당 노드가 차지하던 네트워크 내의 구조적 위치는 공백으로 남게 되는데, 새로 개통한 대포폰은 이 공백을 대체하며 위의 핵심 네트워크 및 과거 용의자와 유사한 연결 형태를 형성한다. 이 때 특정 노드들이 네트워크 내 다른 노드들과 맺는 관계 형태의 유사성을 측정하는 기법인 구조적 등위성(Structural Equivalence)을 적용하여 유의미한 노드, 즉 네트워크 안에서 용의자의 위치를 대체하는 노드를 탐색한다. 용의자의 기존 번호가 핵심 관계망과 형성했던 연결 형태와 가장 유사한 형태를 나타내는 번호가 용의자의 대포폰일 가능성이 높기 때문이다.

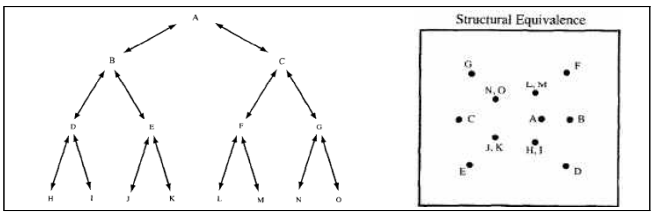


그림 2 구조적 등위성(Structural Equivalence) [8]

4. (4단계) 시계열 분석을 통한 검증

구조적 등위성 분석만으로는 통화 대상자가 우연히 겹친 제3자를 용의자로 오인할 가능성이 존재한다. 이러한 오류를 최소화하고, 억울한 수사대상자를 만들지 않기 위해서 통화의 패턴까지 유사한지 검증하는 절차가 필요하다. 타임라인 분석을 통해 용의자가 기존에 사용하던 휴대전화의 통화 기록과, 위에서 특정한 대포폰 번호 후보군의 통화 기록을 시계열적으로 비교하여, 주 통화 시간대, 평균 통화 시간, 통화 주기 등 동적 활동 패턴의 유사도를 분석한다. 이처럼 구조적 역할의 등위성뿐만 아니라 행동 패턴의 유사성까지 검증함으로써, 실제 대포폰을 특정할 확률을 높이고 분석 결과의 신뢰성을 확보한다.

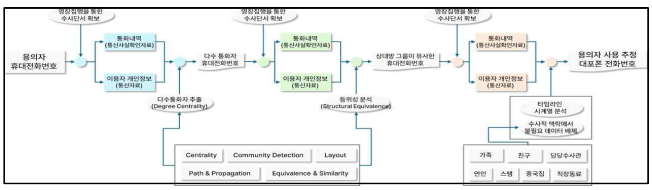


그림 3 대포폰 탐지를 위한 알고리즘

Ⅳ. 사례 연구를 통한 방법론 적용 및 타당성 검증

본 연구를 통해 제안한 알고리즘이 수사 과정에 적용할 수 있는지 방법론 타당성을 검증하기 위해 정량적 데이터셋을 통한 성능 평가 대신 실제 사건 기반의 사례 연구를 적용하였다.

수사관은 살인 혐의로 지명수배가 된 용의자의 대포폰을 특정하기 위하여 용의자가 기존에 사용하던 휴대전화번호의 통화내역에 연결중심성 분석을 적용, 부인, 내연녀, 친구 등 통화량이 많은 3명을 도출하였다. 다음으로, 부인, 내연녀, 친구 휴대전화번호 통화내역에 연결중심성 분석을 적용하여 통화량이 많은 대상을 추출하고 등위성 분석을 통해 용의자 기존 휴대전화번호와 네트워크 안에서 구조적으로 같은 위치에 있는 전화번호 3개를 특정하였다. 그 중 하나는 담당수사관의 번호였고, 다른 하나는 범행 이전에 개통된 전화번호로 범죄와 관련성이 낮아 대포폰 후보군에서 제외되었으며, 새롭게 발견된 나머지 하나를 대포폰으로 추정하게 되었다. 이후 시계열 분석을 통해 내연녀가 담당수사관과 전화를 주고받은 직후 대포폰 추정 번호로 전화를 거는 패턴이 확인되었으며, 이와 같은 교차 검증을 바탕으로 수배자의 대포폰을 특정하게 되었다.

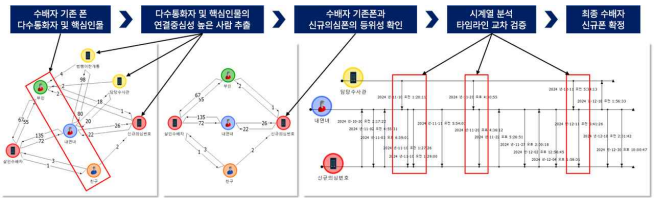


그림 4 실제 수사사례 적용 결과

Ⅴ. 결론

본 연구를 통해 통신수사 과정에서 사회 연결망 분석(SNA)을 적용하여 조직적으로 사용되는 대포폰을 효과적으로 탐지하는 알고리즘을 제안하였다. 위의 사례 분석을 통해, 제안하는 4단계 알고리즘이 복잡한 통신 데이터 속에서 논리적 흐름에 따라 대포폰 후보를 효과적으로 압축하고 특정할 수 있는 체계적인 분석 프레임워크를 제공함을 보였다. 통화내역을 네트워크로 모델링하고 커뮤니티 탐지 및 중심성 분석을 통해 대포폰 사용 집단의 구조적 특징을 식별함으로써, 기존 분석 기법의 한계를 보완하고 수사 효율성을 높일 수 있는 가능성을 제시했다는 점에서 실무적 의의가 있다.

다만, 본 연구는 두 가지 한계가 있다. 첫째로, 사회적 관계를 반드시 정량적으로만 분석할 수는 없다는 점이다. 통화 횟수가 적더라도 가족, 친인척, 업무상 관계 등 수사적 맥락에서 중요한 전화번호가 있는데, 가중치가 낮거나 연결중심성 지표가 낮아 핵심 네트워크에서 제외되는 경우가 발생할 수 있다. 둘째로, 사례 연구를 통해 제안 모델의 실효성을 확인하였지만, 여러 사건, 대규모 데이터에 대한 일반화 가능성이나 탐지 정확도(Precision), 재현율(Recall)과 같은 정량적 지표를 제시하지는 못했다.

향후 수사적 맥락을 반영한 가상의 데이터를 생성하여 정확도를 검증하고, 시간의 흐름에 따른 네트워크 변화를 동적으로 분석하는 동적 네트워크 분석(Dynamic Network Analysis) 기법을 도입하여 더욱 고도화된 탐지 모델을 개발할 계획이다.

ACKNOWLEDGMENT

이 논문은 2025년도 경찰청의 재원으로 과학치안진흥센터 사이버범죄 수사단서 통합분석 및 추론시스템 개발 사업사업의 지원을 받아 수행된 연구임(No. RS-2025-02218280)

참 고 문 헌

[1] 박노섭 · 이동희 · 이윤 · 장윤식, "범죄수사학", 경찰대학 출판부, 2013

[2] Yong Hak Kim and Young Jin Kim, "Social Network Analysis", PAKYOUNGSA, 2024

[3] Kee Young Kwahk, "Social Network Analysis", CRBooks, 2017

[4] 김지온. "사회연결망 분석원리의 범죄 수사상 활용방안에 관한 연구", 디지털포렌식연구, 13(2), pp. 89-109.

[5] Byeongkwan Woo and Ji On Kim and Ro-seop Park, "A Study on the Intelligent Using Methodology of Crime Scene Investigation Data", Journal of Data Forensics Research, 1(1), pp. 63-87

[6] 김지온 and 우병관. "사이버 범죄 탐지 및 분석기술에 대한 탐색적 연구", 한국치안행정논집, 19(3), pp. 57-76.

[7] 김현철 and 윤지원. "소셜네트워크 분석을 통한 사이버금융범죄 수사 활용사례(2-모드 중심으로)" 디지털포렌식연구, 14(4), pp. 449-465.

[8] Burt, R. S. "Detecting role equivalence". Social networks, 12(1), pp. 83-97.

사이버범죄 수사 지원을 위한 온프레미스 LLM 기반 비정형 텍스트 핵심 정보 추출

윤병휘[1], 박현호[2,3], 김지온[1,*]

[1]한림대학교, [2]한국전자통신연구원, [3]과학기술연합대학원대학교

pathofborealis04@gmail.com, hyunhopark@etri.re.kr, jion972@hallym.ac.kr

On-Premise LLM-Based Key Information Extraction from Unstructured Text for Cybercrime Investigation Support

Byung Hui Yoon[1], Hyunho Park[2,3], Ji-On Kim[1,*]

[1]Hallym University, [2]Electronics and Telecommunications Research Institute, [3]University of Science and Technology

요 약

사이버범죄 수사를 위해서는 수사 과정에서 확보한 비정형 텍스트(예: 카카오톡 대화)로부터 피해자 이름, 피해 금액 등 핵심 정보를 정확하고 신속하게 추출할 필요가 있다. LLM(Large Language Model)은 비정형 텍스트에서 핵심 정보를 정확하고 신속하게 추출할 수 있으나, 피해자 이름과 피해 금액 같은 개인정보는 유출 위험이 존재하므로 클라우드 기반 LLM(예: ChatGPT)을 활용하는 데에는 한계가 있다. 이에 본 연구는 외부 클라우드 연결 없이 활용 가능한 온프레미스(On-Premise) LLM인 EXAONE-3.5-7.8B-Instruct를 기반으로 한 핵심 정보 추출 방법을 제안한다. 또한, 본 연구에서 핵심 정보 추출 정확도 향상시킬 수 있는 프롬프트 기법을 적용하고, LLM 결과 후처리를 위한 정규표현식을 설계하였다. 본 연구에서 제안한 방법을 사이버범죄 관련 카카오톡 대화 모의 텍스트에 적용한 결과, 약 84%의 F1-score를 달성하였다. 본 연구의 핵심 정보 추출 기법은 개인정보 유출 우려 없이 사이버범죄 수사의 신속한 진행에 기여할 수 있다.

Ⅰ. 서 론

사이버범죄 수사에서 범죄의 상황을 파악하는 핵심 정보의 추출은 필수적이다. 경찰청에서 공개한 자료에 따르면 국내 사이버범죄는 2022년 217,807건에서 2023년 약 241,842건으로 증가하였고, 2023년 사이버범죄 241,842건에서 사이버사기가 전체 230,355건 중 155,715건(64.39%)을 차지하였고[1], 사이버사기의 피해액은 2022년 1조 8천억 원에 달했다[2]. 사이버사기와 같은 사이버범죄에서는 범죄자와 피해자의 관계와 범행 자금 흐름을 빠르게 식별할 수 있는 피해자의 이름과 피해 금액과 같은 핵심 정보를 신속하게 파악하는 것이 중요하다. 그러나, 수사 과정에서 확보하는 대부분의 단서는 범죄자와 피해자 사이의 카카오톡이나 문자 대화와 같은 단서 데이터들은 대부분 자연어 형태의 비정형 텍스트이기에 정형 데이터처럼 규칙 기반의 분석을 통한 핵심 정보 추출은 어렵다. 또한, 피해자가 100명 이상인 경우 수사관이 수작업으로 비정형 텍스트에서 핵심 정보를 일일이 추출하기에 상당한 시간이 소요되어 수사 진행에 부담이 될 것이다.

LLM(Large Language Model)을 활용하여 정확하고 신속하게 비정형 텍스트의 핵심 정보를 추출하는 방안에 관한 연구가 진행되었다. 연구 [3]에서 ChatGPT-3.5를 이용하여 병리보고서에서의 종양의 크기, 특징, 림프절의 침범여부 등의 정보를 추출하는 방안을 제안하였다. 그러나, ChatGPT-3.5와 같은 클라우드 기반 LLM을 통한 정보 추출은 범죄 수사와 같이 개인정보 유출에 민감한 분야에서는 적용에 한계가 있다.

본 연구는 온프레미스(On-Premise) LLM을 활용한 수사 관련 비정형 텍스트로부터 핵심 정보를 추출하는 방안을 제안한다. 온프레미스 환경은 모델과 데이터가 기관 내부 인프라에서만 운용되므로 외부 네트워크 연결이 차단되어 있어, 개인정보 및 수사 기밀 정보가 외부로 유출될 가능성을 최소화할 수 있다는 점에서 클라우드 기반 LLM보다 보안성이 우수하다. 온프레미스 LLM으로는 한국어에 대한 이해와 추론 성능이 우수한 EXAONE-3.5-7.8B-Instruct 모델을 활용하였다[4]. 핵심 정보 추출 정확도 향상을 위해 LLM의 성능을 개선시킬 수 있는 프롬프트 작성 원칙[5]에 따라 프롬프트를 설계하였다. 또한, EXAONE-3.5-7.8B-Instruct의 출력에서 핵심정보를 정제할 수 있는 후처리 방안도 설계하여 개발하였다. 본 연구에서 제안하는 온프레미스 LLM 기반 비정형 텍스트 핵심 정보 추출 기법은 사이버범죄 관련 모의 텍스트 데이터에서 약 84%의 F1-score를 기록하며 준수한 성능을 보였다. 이러한 결과는 피해자가 많은 범죄에서 핵심 정보 추출 시간을 단축 시켜 수사 진행을 지원할 수 있음을 시사한다.

Ⅱ. 온프레미스 LLM 기반 비정형 택스트의 핵심 정보 추출

온프레미스 LLM 기반 비정형 텍스트 핵심 정보 추출 방법을 설명한다. 그림 1은 온프레미스 LLM 기반 비정형 텍스트로부터 사이버 범죄의 핵심 정보 추출하는 절차를 보여준다. 사이버범죄 핵심 정보를 추출하는 절차는 크게 단계 1 “핵심 정보 추출”과 단계 2 “핵심 정보 정제”로 구분된다. 단계 1에서는 핵심 정보 추출을 위해 설계된 프롬프트를 온프레미스 LLM에 입력하여 핵심 정보를 추론하고, 그 결과를 포함한 JSON 파일을 생성한다. 이후 단계 2에서는 정규 표현식 기반 후처리를 적용하여 불필요한 정보를 제거함으로써 핵심 정보 추출 성능을 개선할 수 있다.

본 연구에서는 비정형 텍스트로부터 핵심 정보를 추출하기 위해 온프레미스 LLM 모델로 EXAONE-3.5-7.8B-Instruct 모델[4]을 활용하였다. EXAONE-3.5-7.8B-Instruct 모델이 한국어에 특화된 대규모 언어 모델

* : 교신저자

로서, 복잡한 표현과 다양한 형태로 나타나는 한국어 텍스트를 정확하게 해석하고 의미 단위를 효과적으로 구분할 수 있다. 그래서, 비정형 텍스트에서 핵심 정보 추출 과정에서 높은 신뢰성을 제공할 수 있다.

또한, 본 연구에서는 사이버사기 중 작업 대출 사기의 시나리오를 바탕으로 사이버범죄 전문가와의 자문을 통해 만들어진 모의 카카오톡 대화 100건의 텍스트 데이터를 사용하였다. 모의 카카오톡 대화 텍스트의 시나리오는 범죄자가 SNS로 신용 불량자와 같이 대출이 어려운 대출 희망자를 모집하여 대출 희망자들의 서류를 위조하여 제 2 금융권으로 대출을 받은 뒤, 대출 금액을 갈취한 이력을 다루고 있다. 모의 카카오톡 대화 텍스트에서 "대출일시", "대출자", "대출금액" 등을 중심으로 주요 핵심 정보를 추출한다면, 작업 대출 사기의 현황을 신속하게 파악할 수 있다.

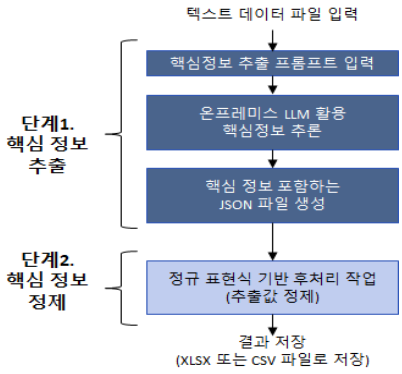


그림 1. 온프레미스 LLM 기반 사이버범죄 핵심 정보 추출 절차

단계 1에서 비정형 텍스트를 주면 온프레미스 LLM에 핵심 정보 추출을 위한 프롬프트가 입력되고, 온프레미스 LLM 모델은 프롬프트를 따라 핵심 정보를 포함하는 JSON 파일을 생성한다. 프롬프트는 연구 [5]에서 제시한 LLM 답변 품질 향상을 위한 지시문 작성 원칙을 참고하여 설계하였다. 프롬프트에서 EXAONE-3.5-7.8B-Instruct 모델의 역할을 기업용 비정형 한국어 텍스트 전용 추출 엔진으로 설정하고, 단계적 추론을 수행하도록 지시하였다. 또한 프롬프트에는 작업 대출 사기 시나리오와 관련된 "대출일시", "대출자", "대출금액" 등을 포함한 10개 필드로 구성된 핵심 정보 추출 스키마를 제시하고, 이를 단일 JSON 객체로 변환하도록 하였다. 이때 모델이 특정 필드의 핵심 정보를 식별하지 못하거나 추출된 값이 모호하다고 판단될 경우, 해당 값을 null로 표기하도록 명시하였다. 프롬프트를 따라 EXAONE-3.5-7.8B-Instruct 모델은 핵심 정보를 추론하여 JSON 파일 형식으로 출력한다.

단계 2에서는 그림 2와 같이 정규 표현식을 활용하여 추론 결과에서 핵심 정보를 정제하였다. 정보 추출 과정에서 모델이 생성한 값 중 일부는 준비된 정답과 일치했으나, 동시에 불필요한 값(garbage value)이 포함되거나 날짜 표기 방식의 불일치, 공백 미제거 등의 문제가 확인되었다. 이를 해결하기 위해 정규 표현식을 적용하여 핵심 정보의 정규화 과정을 수행하였다. 그림 2는 "대출일시"를 "10월 14일"과 같이 월과 일을 기준으로 표준화하는 정규 표현식의 예시를 보여준다.

```
_DATE_SLASH_RE = re.compile(r'(?<!\d)(\d{1,2})\s*[./]\s*(\d{1,2})(?!\d)')
_DATE_MONTHDAY_RE = re.compile(r'(?<!\d)(\d{1,2})\s*월\s*(\d{1,2})\s*일?')
_PAREN_RE = re.compile(r'\([^)]*\)')
_MULTI_SPACE_RE = re.compile(r'\s+')

def _clean_dateish(s: str) -> str:
    """대출일시: '잔금일', '까지' 제거, 10/14 → 10월14일, 공백 정리."""
```

그림 2. 정규표현식을 통한 "대출일시" 정보의 정규화

EXAONE-3.5-7.8B-Instruct 모델을 기반으로 한 작업 대출 사기 카카오톡 대화 텍스트의 핵심 정보 추출 성능은 그림 3에 제시하였다. 데이터 추출 프로그램의 weighted average F1-score는 84.33%로 산출되었다. 특히 "대출일시", "대출자", "대출금액"과 같은 주요 정보에서는 각각 87.43%, 85.71%, 93.00%의 F1-score를 기록하여 우수한 성능을 보였다. 이러한 결과는 온프레미스 LLM 기반 핵심 정보 추출이 수사관에게 범죄 현황을 신속히 파악하고 본격적인 수사 준비를 수행하는 데 실질적인 도움을 줄 수 있음을 시사한다.

```
── Per-field Precision / Recall / F1 ──────────────────────────
          대출일시  P: 94.12%  R: 81.63%  F1: 87.43%  (TP:80, FP:5, FN:18, N:100)
          사용기간  P: 96.94%  R: 97.94%  F1: 97.44%  (TP:95, FP:3, FN:2, N:100)
           대출자  P: 85.71%  R: 85.71%  F1: 85.71%  (TP:84, FP:14, FN:14, N:100)
      담보물건지 주소  P: 85.00%  R: 85.00%  F1: 85.00%  (TP:85, FP:15, FN:15, N:100)
          대출금액  P: 93.00%  R: 93.00%  F1: 93.00%  (TP:93, FP:7, FN:7, N:100)
       대출금액(숫자)  P: 87.88%  R: 87.00%  F1: 87.44%  (TP:87, FP:12, FN:13, N:100)
        대출수수료  P: 73.00%  R: 73.00%  F1: 73.00%  (TP:73, FP:27, FN:27, N:100)
          투자자  P: 70.00%  R: 72.92%  F1: 71.43%  (TP:70, FP:30, FN:26, N:100)
            금액  P: 82.00%  R: 83.67%  F1: 82.83%  (TP:82, FP:18, FN:16, N:100)
         금액(숫자)  P: 80.00%  R: 80.81%  F1: 80.40%  (TP:80, FP:20, FN:19, N:100)

Weighted F1:  84.33%  (total samples: 1000)
```

그림 3. 작업 대출 사기 대화 텍스트의 핵심 정보 추출 성능

Ⅲ. 결론

본 논문에서는 온프레미스 LLM을 기반으로 한 비정형 텍스트 핵심 정보 추출 방법을 제안하였다. 이를 위해 EXAONE-3.5-7.8B-Instruct 모델을 적용하고, 성능 향상을 위한 프롬프트 설계 원칙을 반영하였으며, 정규 표현식을 통해 추출된 정보를 정제하였다. 작업 대출 사기 대화 텍스트를 대상으로 실험한 결과, 약 84%의 F1-score를 기록하며 준수한 성능을 확인할 수 있었다. 제안한 방법은 사이버범죄 수사 초기 단계에서 범죄 상황을 신속히 파악하고 수사 준비를 지원하는 데 기여할 수 있다. 향후 연구에서는 2025년 8월에 공개된 gpt-oss를 포함한 다양한 온프레미스 LLM 모델을 비교·분석하여 최적의 모델을 탐색하고, 더 많은 사이버범죄 텍스트에 적용함으로써 핵심 정보 추출 기법을 고도화할 예정이다.

ACKNOWLEDGMENT

이 논문은 25년도 정부(경찰청)*의 재원으로 과학치안진흥센터 사이버범죄 수사단서 통합분석 및 추론시스템 개발 사업의 지원을 받아 수행된 연구임(No. RS-2025-02218280)

참 고 문 헌

[1] 김영명, "경찰청 사이버수사 인력 증가, 사이버범죄 건수·검거율은 감소," 보안뉴스, 2024. [Online]. Available: https://www.boannews.com/media/view.asp?idx=133306

[2] 김유리, "작년 사이버사기 피해 1조8천억원…4년새 8배↑," 한국세정신문, 2024. [Online]. Available: https://www.taxtimes.co.kr/news/article.html?no=26621

[3] J. Huang, et al., "A critical assessment of using ChatGPT for extracting structured data from clinical notes," NPJ Digital Medicine, vol. 7, no. 1, p. 106, 2024.

[4] LG AI Research, "EXAONE 3.5-7.8B-Instruct," Hugging Face, 2024. [Online]. Available: https://huggingface.co/LGAI-EXAONE/EXAONE-3.5-7.8B-Instruct

[5] S. M. Bsharat, A. Myrzakhan, and Z. Shen, "Principled instructions are all you need for questioning LLaMA-1/2, GPT-3.5/4," arXiv preprint, arXiv:2312.16171, Jan. 2024.

IEEE 802.15 차세대 광 무선 통신(IG NG-OWC) 관심 그룹 응용 신청

후이 응우옌, 영민 장*

국민대학교

ngochuy@kookmin.ac.kr, *yjang@kookmin.ac.kr

Call for Applications for IEEE 802.15 Interest Group on Next-Generation Optical Wireless Communication (IG NG-OWC)

Huy Nguyen, Yeong Min Jang*
Kookmin University.

요 약

The exponential growth of connected devices and the demand for ultra-high-speed, low-latency wireless communication are driving the exploration of technologies beyond traditional radio frequency (RF) systems. Optical Wireless Communication (OWC), which leverages the visible light, infrared, and ultraviolet spectrum, offers a promising solution for 6G and beyond. This paper outlines the IEEE 802.15 Interest Group for Next-Generation Optical Wireless Communication (IG NG OWC) call for applications, focusing on Optical Camera Communication (OCC), Free-Space Optics (FSO), and advanced PHY/MAC protocols. The goal is to establish a roadmap for next-generation OWC systems to achieve scalability, mobility, and secure communications across diverse applications.

Keywords—Optical Wireless Communication, OCC, FSO, IEEE 802.15.7a, 6G, PHY/MAC.

Ⅰ. 서 론

The sixth generation (6G) of wireless communication will require extremely high data rates, very low latency, and massive device connectivity. Current RF spectrum resources, especially below 40 GHz, are congested and insufficient to meet the growing demands of IoT and smart city ecosystems [1]. Optical Wireless Communication (OWC) utilizes unlicensed optical spectrum, including visible light (VL), infrared (IR), and ultraviolet (UV) bands, offering terabit-level bandwidth, immunity to electromagnetic interference, and enhanced security [2].

Recent standardization efforts, such as IEEE 802.15.7a-2024, have introduced advanced modulation and coding schemes, including OFDM, MIMO, and NOMA, to overcome OCC's limitations in communication distance and mobility support [3]. Furthermore, IEEE 802.15 IG NG OWC was formed to promote collaboration and research on next-generation OWC technologies, targeting high-speed, high-mobility applications such as autonomous vehicles, drones, and smart factories [4].

This paper summarizes the focus areas of the NG OWC initiative and calls for innovative proposals to advance OWC technologies for 6G and beyond.

Ⅱ. 본론

Next-generation OWC aims to address the limitations of current systems by enhancing performance at both the physical (PHY) and medium access control (MAC) layers. Optical Camera Communication (OCC) is a key focus area, enabling low-cost communications using existing cameras and LEDs. Enhancements are needed to improve data rates, reliability, and mobility. Techniques such as OFDM, MIMO, and NOMA have been proposed to mitigate limitations like low data rates caused by camera frame rate constraints and vulnerability in non-line-of-sight (NLoS) environments [2],[3]. Additionally, deep learning has arisen as a powerful tool for data decoding, significantly improving OCC performance under high-mobility and harsh conditions [5].

Free-Space Optics (FSO) is another vital area, offering ultra-high-speed, long-range links suitable for drone networks, satellite communication, and high-speed trains. Challenges such as atmospheric interference and precise beam alignment must be addressed through adaptive protocols and hybrid integration with RF systems [2],[4]. Enhancements to IEEE 802.15.13-2023 will further enable ultra-reliable low-latency communications and secure industrial

networks [3]. Positioning and tracking systems, such as OCC-based indoor positioning and Time-of-Flight (ToF)-based methods, are critical for autonomous systems and IoT applications, providing centimeter-level accuracy for smart factories and mobility platforms [2].

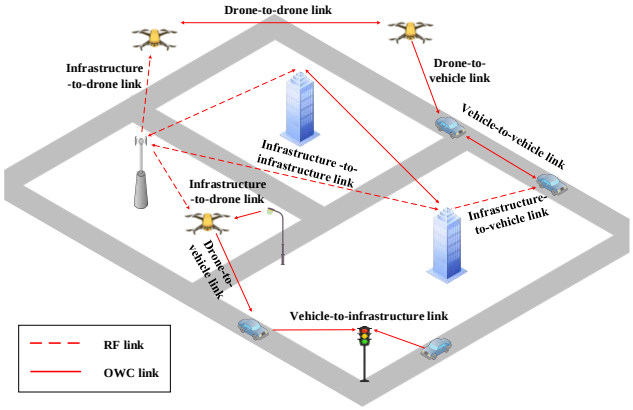


Figure 1. OWC services utilizing 6G and beyond technologies

OWC services-based UAV are a growing with applications in transportation, the military, and civilian life. Figure 1 shows that UAVs can also use for V2V, V2I, and I2I systems for 6G/beyond 6G technologies, Besides that, Table 1 show the recent advancements in NG OWC technologies for the NG-OWC technologies.

Table 1. Recent Advancements in NG OWC Technologies

Technology	Purpose / Benefit	Key Features	Reference
RS-OFDM (Rolling Shutter OFDM)	Overcomes OCC data rate limitations due to camera frame rate	Multicarrier modulation for OCC, improved data throughput	[5]
Hybrid RF/OWC Systems	Ensures reliability in dynamic and NLoS environments	Combines RF and OWC strengths for robust coverage	[2]
Deep Learning Decoders	Enhances performance in high-mobility and noisy environments	AI-based signal detection, lower BER, improved mobility support	[5]
Advanced IEEE Standards	Standardizes NG OWC for global adoption	IEEE 802.15.7a-2024, IEEE 802.15.13-2023 updates	[3],[4]

Ⅲ. 결론

NG OWC represents a paradigm shift in wireless communication, addressing spectrum scarcity and performance bottlenecks inherent to RF-based systems. By advancing OCC, FSO, and integrated sensing technologies, IEEE 802.15 IG NG OWC aims to deliver scalable, secure, and high-speed connectivity for 6G and beyond. Collaborative contributions from researchers and industry will be pivotal to shaping the future of optical wireless networks.

ACKNOWLEDGMENT

This work was supported by Korea Research Institute for defense Technology Planning and advancement (KRIT) grant funded by the Korea government (DAPA (Defense Acquisition Program Administration)) (KRIT-CT-23-041, 2025).

참 고 문 헌

[1] X. Chen et al., “Frequency selective surface toward 6G communication systems: A contemporary survey,” *IEEE Commun. Surveys Tuts.*, vol. 26, no. 3, pp. 1635-1675, 2024.

[2] H. Nguyen et al., "Survey of Next-Generation OWC Technologies for 6G and Beyond," *ICT Express*, vol. 11, pp. 576-589, 2025.

[3] *IEEE Standard for Local and Metropolitan Area Networks—Part 15.7: Short-Range Optical Wireless Communications Amendment 1: Higher Rate, Longer Range Optical Camera Communication (OCC)*, IEEE Standard 802.15.7a-2024, Feb. 2025.

[4] H. Nguyen et al., "Enabling Technologies and Challenges for IEEE 802.15.7 OWC," *IEEE Commun. Mag.*, vol. 62, no. 3, pp. 90-95, Mar. 2024.

[5] H. Nguyen et al., "Rolling OFDM for Image Sensor-Based Optical Wireless Communication," *IEEE Photonics Journal*, 2019.

AI 기반 실시간 UAV Motor Error Detection 시스템

윤희창, 김상현, 유승호*
국립부경대학교 컴퓨터공학과
{yhc122833, shkim1615}@pukyong.ac.kr, shyoo@pknu.ac.kr

AI-based Real-time UAV Motor Error Detection System

HuichangYun, Sanghyun Kim, Seungho Yoo*
Department of Computer Engineering, Pukyong National Univ.

요 약

무인항공기(UAV)는 물류, 감시, 재난 대응 등 다양한 분야에서 활용도가 증가하고 있다. UAV는 안정적인 비행, 신속한 장애 대응이 필수적이다. 특히나 UAV의 핵심 부품인 모터는 비행 안전성과 직결되며, 모터의 고장은 UAV의 비정상적인 비행을 야기하며 이는 심각한 사고로 이어질 수 있다. 이러한 사고를 막기 위해 정상 비행 여부를 판단하는 Failure Detection 기법들이 존재하지만, 기존의 대부분의 기법들은 Rule-Based 기법에 기반하고 있기 때문에 모터 오류 감지 같은 복합적인 기체 이상 진단에 어려움이 있다. 이러한 문제를 해결하기 위해 본 연구에서는 AI를 적용한 Motor Error Detection 시스템을 제안한다. Gazebo 시뮬레이션을 사용하여 모터 고장 시나리오를 구현해 데이터를 확보하였으며, 해당 데이터로 GRU 기반의 Motor Error Detection Model을 학습시키고 필터를 적용하여 신뢰성 높은 모터 오류 감지 시스템을 구현하였다. 평가 결과 제안된 시스템은 Motor Error 현상이 일어난 후 평균 0.87초 이내에 오류를 감지할 수 있었으며, 이러한 경량화된 시스템은 엣지 단에서의 실시간 적용 가능성을 보여준다.

Ⅰ. 서 론

무인항공기(UAV)는 물류, 감시, 재난 대응[1, 2] 등 다양한 분야에서 활용되고 있으며, 중요도가 계속해서 증가하고 있다. UAV 운용에 있어서 안정적인 비행과 신속한 장애 대응이 필수적이다. 특히 UAV의 핵심 부품인 모터는 비행 안전성과 직결되며, 모터 고장은 UAV의 비정상적인 비행을 일으키고 심각한 사고로 이어질 수 있다. 하지만 기존의 Rule-Based 고장 여부 판단 기법[1]은 특정 조건에서 효과적일 수 있으나, 모터 오류 감지와 같은 복합적인 기체 이상 진단에 어려움이 있다.

이러한 문제를 해결하기 위해, 본 연구에서는 무인기 비행 데이터를 활용한 AI 기반의 모터 고장을 감지하는 시스템을 제안한다. 실시간으로 정확한 모터 이상 감지를 하는 것이 제안하는 시스템의 목표이며, 이를 통해 본 연구에서는 다음의 내용들을 제안하고 확인하였다.

- 현실에서 얻기 어려운 비행 중 모터 고장 데이터를 시뮬레이션으로 구현하여 모델 학습데이터로 활용함.
- 시뮬레이션에서 생성된 데이터를 바탕으로 GRU 모델을 학습시켜 Motor Error Detection Model을 구현함.
- GRU 모델 단독으로 오류 예측을 수행하였을 때 발생하는 예측 노이즈를 해결하기 위한 후처리 필터를 적용하여 시스템의 신뢰성을 향상시킴.

Ⅱ. 데이터 생성 및 수집

본 논문에서 제안하는 시스템을 구현하고 검증하기 위하여 UAV 비행 중 모터의 고장이 일어나는 상황에 대한 데이터가 필요하였다. 하지만 현실에서 실제 UAV를 통해 이러한 데이터를 확보하는 것은 어려움이 있기에 시뮬레이션을 활용하여 해당 데이터를 수집하였다. 시뮬레이션은 Gazebo와 PX4 SITL을 사용하였고, 시뮬레이션 상에서 UAV 비행 중 모터 에러를 의도적으로 발생시켜 고장 시나리오를 구현하였다.

UAV Flight Controller인 PX4에서는 다양한 비행 관련 데이터[3]가 존재하는데, 그중 드론 기체 정보 (roll, pitch, yaw 등), 비행 목표 정보 (target_x, target_y, target_z 등), 및 모터 정보를 사용하였다. MAVLink를 통해 0.2초 간격으로 업데이트되는 UAV 실시간 비행 데이터를 수집하였다. 무인기의 이동과 모터 오류 발생에 대해 무작위적으로 이루어지도록 시뮬레이션을 구성하고, 해당 환경에서 약 15만 개의 데이터를 확보하였으며, 이를 오류 검출 모델의 학습데이터로 사용하였다.

Ⅲ. 시스템 구성

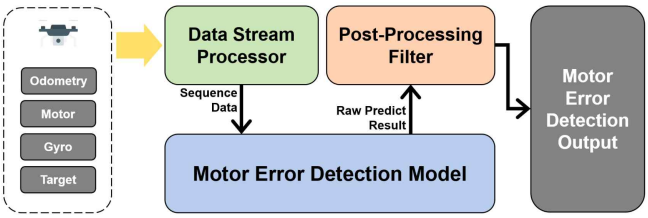


그림 1. UAV Motor Error Detection System Design

본 논문에서 제안하는 Motor Error Detection 시스템의 시스템 구성도를 그림 1에 나타내었다. 먼저, UAV의 실시간 비행 데이터를 Data Stream Processor가 받아 오고, 실시간으로 업데이트되는 비행 데이터를 정해진 윈도우의 크기만큼 이전 비행 데이터와 묶어 하나의 시퀀스 데이터를 생성한다. 이 시퀀스 데이터는 비행 데이터가 업데이트될 때마다 생성되며, 생성될 때마다 Motor Error Detection Model에 넘겨진다. Motor Error Detection Model은 받은 시퀀스 데이터를 기반으로 Motor Error 확률을 계산하여 Raw Predict Result를 생성하고 이를 후처리 필터에 전달한다. 필터는 Raw Predict Result에서 예측 노이즈를 제거하고, 최종적으로 Motor Error가 발생했는지를 결정한다.

3.1 GRU(Gated Recurrent Unit) 기반 Motor Error Detection Model

본 연구에서는 GRU 기반의 분류 모델을 사용하였다. GRU는 LSTM과 유사하게, 이전 시점의 정보를 고려하여 시계열 데이터의 시간적 패턴을 학습할 수 있으며 LSTM에 비해 구조가 단순하고 파라미터 수가 적다는 장점[4]이 있다. 대신에 LSTM에 비해 긴 시퀀스의 학습에서 약한 모습을

보이지만, 본 연구에서는 긴 시퀀스 분석은 필요로 하지 않고 오히려 실시간성이 필요로 하였기에 학습 모델로서 GRU를 선택하였다.

모델 학습에 활용된 입력 feature는 섹션 2에서 서술한 드론 기체 정보, 비행 목표 정보, 모터 정보를 사용하였고, 슬라이딩 윈도우 방식으로 데이터를 시퀀스로 변환하였다. 윈도우 크기는 30으로 설정하였으며, 각 시퀀스 데이터들은 시계열 상으로 가장 최근에 업데이트된 데이터가 정상 혹은 고장인지에 대해 라벨링하여 학습에 활용하였다.

3.2 후처리 필터

GRU 모델을 단독으로 사용하여 모터 오류 예측을 하면, 예측값이 순간적으로 튀는 예측 노이즈가 발생한다. 예를 들어 UAV가 정상 비행 중임에도 바람이 불거나 급격한 방향 전환 등의 이유로 기체가 순간적으로 흔들리는 것을 모터 에러로 착각하는 현상이 발생하는데, 이러한 오탐(False Positive)을 줄이기 위해 후처리 필터(post-processing filter)를 적용하였다. 후처리 필터 기법 중 Majority Voting, EMA(Exponential Moving Average)[4], Debounce, Kalman Filter[5]를 각각 적용하여 결과를 비교해보았고, 그 결과를 표 1에 나타내었다.

Ⅳ. 연구 결과

본 연구에서 사용한 GRU모델 및 후처리 필터 기법은 슬라이딩 윈도우를 기반으로 데이터가 업데이트된다. 따라서 모델이 모터 오류를 탐지하기 위해서는 데이터가 윈도우에 누적될 필요가 있으며, 이는 고장 상황 후 기체 데이터가 누적되는 과정(step)이 필요하다는 것을 말한다. 표 1의 Missed Detection은 테스트 중 모델이 모터 오류를 놓친 수를 나타내는데, GRU만을 썼을 때와 필터를 적용시켰을 때 모두 0인 것을 확인할 수 있다. 이는 실제 오류 탐지 자체는 모두 성공적으로 수행되었으며, 성능 차이로는 윈도우 누적 지연 및 노이즈만이 존재함을 의미한다.

따라서 성능은 다음 두 가지 항목으로 평가된다. 첫째, 실제 오류가 발생한 시점 t_e와 에러가 탐지된 탐지 시점 t_d 사이의 탐지 지연($D = t_d - t_e$, 단위: step)의 평균인 Average Delay. 두 번째, 필터 성능을 평가하기 위한 오탐률(False Positive Rate, FPR = FP / (FP + TN))이다. 여기서 FP는 정상 상태를 오류로 잘못 판단한 횟수, TN은 정상 상태를 올바르게 판단한 횟수이다. 테스트 데이터는 섹션 2에서 생성하였던 데이터와 동일한 방식으로 생성되었고, 약 3천 개의 데이터를 생성해 테스트하였다.

4.1 GRU 및 후처리 필터 적용 결과

각 필터는 Average Delay가 5 step(0.2초 업데이트 기준 1초)을 넘기지 않는 기준으로 가장 정확도가 높았던 파라미터를 적용시켜 테스트하였고, 그 결과는 표 1과 같다. 필터를 사용하면 GRU모델 단독으로 사용했을 경우보다 오탐을 확실히 줄여주는 모습을 보여준다. 그중 Majority Voting 방식이 필터를 적용한 결과 중 가장 Average Delay가 낮고, FPR 또한 가장 성능이 좋은 Debounce와 비슷하게 좋은 성능을 보여주었기에 본 연구에서는 GRU + Majority Voting을 사용하였다.

표 1. GRU 및 후처리 필터 적용 결과

	Missed Detection	Average Delay(step)	FPR
GRU Only	0	2.0	0.064
Majority Voting	0	4.36	0.006
EMA	0	4.43	0.018
Debounce	0	4.5	0.004
Kalman Filter	0	4.8	0.017

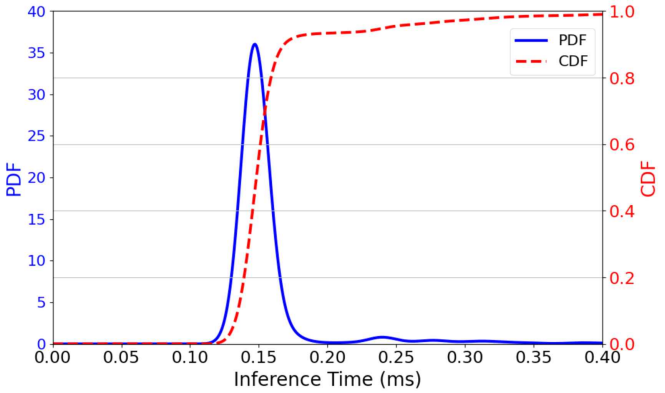


그림 2. 오류 검출 소요 시간에 대한 PDF, CDF

4.2 Inference Time

그림 2는 시퀀스 데이터가 GRU + Majority Voting을 거쳐 Motor Error Detection Output을 생성할 때까지, 하나의 step에 대한 Inference Time을 나타낸다. 90%의 경우 추론이 0.2ms 이내로 끝나며, 0.35ms 이내에 거의 모든 추론이 끝나는 것을 확인할 수 있다. 따라서 GRU + Majority Voting의 Average Delay가 4.36 step이므로 평균적으로 Motor Error 현상이 일어난 후 대부분은 0.87초 이내에 감지되는 것을 의미하며, 이를 통해 제안된 시스템이 매우 경량화되어 있음을 확인할 수 있다.

Ⅴ. 결론

본 논문에서는 UAV 비행 데이터를 사용하여 실시간으로 UAV Motor Error Detection을 수행하는 시스템을 제안한다. 시뮬레이션 환경에서 UAV의 모터 고장 시나리오를 생성하여 약 15만 개의 학습데이터를 확보하였으며, 이를 기반으로 GRU 모델을 학습시켰다. 또한 필터 적용을 통해 예측 노이즈 문제를 줄여 시스템의 신뢰성을 높였고 Average Delay, Inference Time 평가 결과, Motor Error 발생 후 대부분의 경우 0.87초 이내에 오류를 감지할 수 있음을 확인하였다.

향후 연구에서는 제안된 경량화된 시스템을 실제 UAV 기체에 탑재하여 HITL 실험을 수행함으로써 엣지 단에서 실시간 탐지 성능을 검증할 계획이다. 또한 GRU 이외의 모델과 다양한 필터 기법의 조합을 탐색하거나 UAV 비행 데이터 업데이트 주기를 빠르게 하여 더욱 정확하고 신속한 시스템을 설계할 수 있을 것으로 기대된다.

ACKNOWLEDGMENT

이 논문은 2025년 정부(방위사업청)의 재원으로 국방기술진흥연구소의 지원을 받아 수행된 연구임 (KRIT-CT-23-041)

참 고 문 헌

[1] H. Shraim, A. Awada and R. Youness, "A survey on quadrotors: Configurations, modeling and identification, control, collision avoidance, fault diagnosis and tolerant control," in IEEE Aerospace and Electronic Systems Magazine, vol. 33, no. 7, pp. 14-33, 2018.

[2] 차용준 외, "국방 분야 인공지능 기술의 적용 사례 연구," 한국통신학회 인공지능 학술대회 논문집, pp. 258 - 260, Sept. 2024.

[3] "uORB Message Reference | PX4 Guide (main)." Accessed: Aug. 25, 2025. [Online]. Available: https://docs.px4.io/main/ko/msg_docs/

[4] S. Mahjoub, L. Chrifi-Alaoui, B. Marhic and L. Delahoche, "Predicting Energy Consumption Using LSTM, Multi-Layer GRU and Drop-GRU Neural Networks," Sensors, vol. 22, no. 11, p. 4062, 2022.

[5] J. D. Borrero and J. Mariscal, "Predicting Time SeriesUsing an Automatic New Algorithm of the Kalman Filter," Mathematics, vol. 10, no. 16, p. 2915, Jan. 2022.

연합학습 기반 UAV 이상탐지 및 VC 인증·ChaCha20 기반 보안 세션 아키텍처

박성수, 김기형*
아주대학교, *아주대학교

parky@ajou.ac.kr, *kkim86@ajou.ac.kr

Federated Learning-Based UAV Anomaly Detection and VC Authentication· ChaCha20 Based Secure Session Architecture

SeongSu Park, Ki-Hyung Kim*
Ajou Univ.

요 약

본 논문은 2020 년 CMU 에서 공개한 ALFA Dataset 을 활용하여 무인항공기(UAV) 운영 환경에서의 연합학습 (Federated Learning, FL) 이상탐지(Anomaly Detection)용 VC 인증과 ChaCha20 을 활용한 보안 세션 아키텍처를 제안한다. 제안 아키텍처는 각 UAV 가 로컬에서 이상탐지 모델을 학습하고, 중앙 서버는 FedAvg 를 통해 전역 모델을 생성함으로써 데이터 프라이버시를 보호하고 통신 부담을 줄인다. 또한 탐지 결과를 신뢰도 점수(Trust Score)로 변환하여 Verifiable Credential(VC) 형식으로 발급·검증하며, 이를 Noise_XX 키합의와 ChaCha20-Poly1305 세션 암호화에 결합함으로써 UAV 미션 수행 시 강력한 인증·암호화 보안을 제공한다. 실험적으로 ALFA Dataset 의 센서·엔진 이상 로그를 기반으로 한 데모 파이프라인을 구현하였으며, VC 검증·세션 보안·재키잉(re-keying) 정책까지 포함한 결과를 제시한다. 본 연구는 UAV 연합학습 기반 보안 프레임워크 설계에 기여할 수 있다.

Ⅰ. 서론

최근 UAV(무인항공기)의 활용은 군사·민간 영역에서 빠르게 확산되고 있다. 그러나 분산된 다수 UAV 의 운용은 센서 고장, 엔진 이상, 네트워크 침입 등 다양한 위협에 노출된다. 전통적인 중앙집중식 이상탐지는 데이터 전송 비용과 보안 문제를 유발한다. 따라서 분산 데이터 처리와 보안성 강화를 동시에 만족하는 아키텍처가 필요하다.

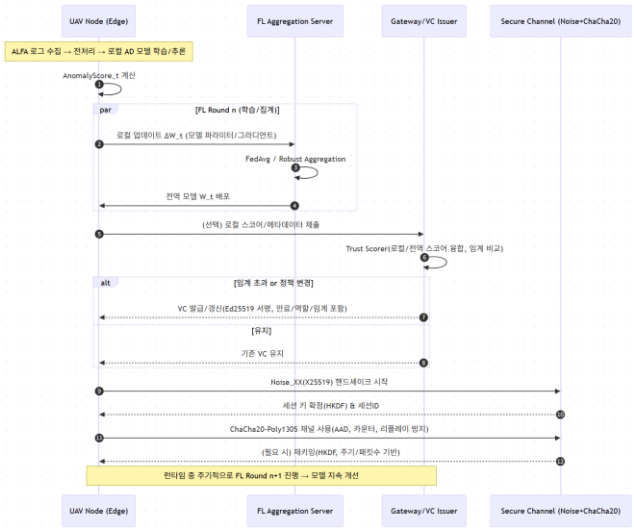
본 논문은 UAV 운용 데이터[1]를 활용하여 연합학습 기반 이상탐지를 수행하고, 탐지 결과를 VC 인증 체계[2, 3]와 ChaCha20[4] 기반 보안 세션에 결합하는 방안을 제안한다. 이를 통해 (1) 프라이버시 보존 학습, (2) 권한·정책 기반 인증, (3) 저전력 환경에 적합한 경량 암호 세션을 동시에 달성한다.

Ⅱ. 제안 아키텍처

본 논문에서 제안하는 아키텍처는 UAV 운용 환경에서 발생할 수 있는 고장 및 이상 상황을 효율적으로 탐지하고, 그 결과를 결과를 **Verifiable Credential(VC) 기반 권한 인증**[2, 3]과 **ChaCha20 보안 세션**에 결합하는 아키텍처를 제안한다. 이를 위해 연합학습(Federated Learning, FL)을 기반으로 한 이상탐지, Verifiable Credential(VC) 인증, 그리고 ChaCha20 세션 보안을 결합하였다.

우선 데이터 계층에서는 UAV 단말이 ALFA Dataset 과 같은 실제 운용 로그를 로컬에서 수집한다[1]. 각 단말은 센서·엔진·조종면 데이터를 윈도 기반으로 전처리하고, median 및 MAD(Median Absolute Deviation)를 활용한 로컬 이상탐지 모델을 학습한다[5]. 이러한 접근은 데이터가 중앙 서버로 전송되지 않고 각 단말 내부에 머무르도록 하여 프라이버시와 통신 효율성을 동시에 확보한다.

모델 계층에서는 중앙 서버가 각 UAV 에서 전송된 모델 파라미터를 FedAvg 방식으로 집계하여 전역 모델을 생성한다. 필요 시 Median-Krum Aggregation 과 같은 Byzantine-robust 기법[6]을 적용하여 악의적 업데이트나 노이즈에 강건하게 동작할 수 있다. 생성된 전역 모델은 다시 각 UAV 에 배포되어 탐지 성능을 향상시킨다.



<그림 1:운용 시퀀스(FL+ 런타임 인증/세션)>

정책 계층에서는 각 UAV 가 산출한 이상탐지 점수를 기반으로 신뢰도(trust score)를 계산하고, 이를 임계값과 비교하여 허용(allow), 2 단계 인증(2FA), 거부(deny)의 정책을 결정한다. 이 결과는 Verifiable Credential(VC) 형태로 발급된다. VC 는 Ed25519 서명을 통해 위·변조를 방지하며, 기체 DID, 미션 식별자, 신뢰도

점수, 정책 및 만료 시간 등을 포함한다. VC 는 네트워크 단절 상황에서도 자율적으로 검증 가능하도록 설계되며, revocation epoch 를 통해 최신 상태 동기화가 가능하다.

보안 계층에서는 VC 와 Noise_XX 키합의를 결합하여 세션 보안을 수립한다[7]. 키합의 시 HKDF 의 info 필드에 VC Digest, 미션 ID, UAV DID 를 포함시켜 세션 키가 권한·미션·디바이스와 강하게 결속되도록 한다. 이후 ChaCha20-Poly1305 AEAD (Authenticated Encryption with Associated Data) 암호화를 통해 UAV 와 서버 간의 통신을 보호하며, Nonce 는 세션 ID 와 카운터로 구성된다[4, 7]. AAD (Additional Authenticated Data)에는 VC Digest 와 미션, 프레임 헤더를 포함시켜 맥락 무결성을 보장한다[2]. 또한 리플레이 방지와 주기적 재키잉 정책을 통해 장기 세션에서도 안전성을 유지한다.

Ⅲ. 실험 및 데모

실험은 ALFA Dataset 과 합성 데이터를 사용하여 제안한 아키텍처의 동작을 검증하였다. 세 개의 UAV 노드를 가정하여 각 노드가 로컬 robust anomaly detector 를 학습하고, FedAvg 를 통해 전역 모델을 산출하였다. 데이터 전처리 과정에서는 센서 로그를 윈도 단위로 분할하여 이상 상황이 반영된 부분과 정상 구간을 비교하였다.

이상탐지 결과를 통해 전역 모델은 정상 데이터에 비해 엔진 고장 및 조종면 이상 데이터에서 유의미하게 높은 anomaly score 를 산출하였으며, 이를 바탕으로 trust score 가 산출되었다. 예를 들어 정상 로그에서는 높은 신뢰도가 유지되었으나, 이상 로그에서는 임계값 미만으로 떨어져 deny 정책이 실행되도록 했다.

발급된 VC 에는 UAV 의 DID, 미션 ID, 신뢰도, 정책, 만료 시간, revocation epoch 등이 포함되었으며, 서버는 이를 Ed25519 기반으로 검증하였다. 실험에서는 정상 로그의 경우 “allow” 정책이 적용되어 VC 가 발급되었고, 이상 로그의 경우 “2FA” 또는 “deny”가 적용됨을 확인하였다.

보안 세션 검증 단계에서는 Noise_XX 기반 키합의로 생성된 세션 키가 VC Digest, 미션 ID, UAV DID 에 바인딩되었다. 이후 ChaCha20-Poly1305 암호화 채널을 통해 UAV 명령 메시지를 보호하였으며, Replay 공격 시도는 카운터 검증을 통해 차단되었다. 또한 패킷 전송 횟수와 시간 기준으로 재키잉이 수행되어 장기간 세션에서도 안정적인 보안을 유지할 수 있었다.

또한, VC Digest 와 미션 ID, UAV DID 가 Noise_XX 키합의 및 ChaCha20-Poly1305 세션 암호화에 바인딩되어, anomaly detection 결과가 보안 세션까지 연계됨을 검증하였다. 데모에서는 (1) VC 발급 및 검증, (2) Noise 핸드셰이크 성공 및 세션키 확정, (3) ChaCha20 암호화 명령 전송 및 리플레이 공격 차단, (4) 패킷·시간 기준 재키잉이 성공적으로 수행되는 것도 모두 확인되었다.

이러한 데모를 통해 본 연구의 아키텍처는 (1) 분산 데이터 기반 이상탐지, (2) 정책·신뢰도 기반 VC 인증, (3) 경량 암호 기반 세션 보안을 효과적으로 결합할 수 있음을 검증하였다. 이는 UAV 운용 환경에서 프라이버시, 신뢰성, 보안성을 동시에 확보할 수 있는 실질적인 프레임워크임을 보여준다.

Ⅳ. 결론 및 향후 연구

본 연구는 ALFA Dataset 을 기반으로 UAV 연합학습 이상탐지와 VC 인증, ChaCha20 보안 세션을 통합한 아키텍처를 제시하였다. 이는 UAV 운용의 **데이터 프라이버시·인증 보장·세션 보안**을 동시에 만족시키는 프레임워크로 활용될 수 있다. 이 연구를 기반으로 향후 (1) GRU/TCN/AE 기반 시계열 모델 확장, (2) Byzantine-robust FL 알고리즘 확장, (3) VC 폐기 메커니즘 (비트맵·Accumulator) 구현, (4) 수시로 Offline 되는 네트워크 환경에서의 안정적인 운영에 대해 연구할 것이다.

ACKNOWLEDGMENT

이 논문은 2025 년 정부(방위사업청)의 재원으로 국방기술진흥연구소의 지원을 받아 수행되었고(KRIT-CT-23-041), 과학기술정보통신부 및 정보통신기획평가원의 대학 ICT 연구센터지원사업의 연구결과로 수행되었고(IITP-2025-RS-2021-II211835), 2025 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행됨 (2021-0-00590, RS-2021-II210590, 대규모 노드에서 블록단위의 효율적인 거래 확정을 위한 최종성 보장 기술개발)

참 고 문 헌

1 Keipour, A., Mousaei, M., and Scherer, S.: ‘ALFA: A dataset for UAV fault and anomaly detection’, The International Journal of Robotics Research, 2021, 40, (2-3), pp. 515-520

2 Chadwick, D.W., and Burnett, D.C.: ‘Verifiable credentials’, Self-Sovereign Identity: Decentralized Digital Identity and Verifiable Credentials, 2021, pp. 126

3 Park, S., and Kim, K.-H.: ‘Leveraging Digital Twins for ON/OFF-LINE Blockchain Networks: A Resilient Framework for Drone and IoT Communication’, in Editor (Ed.)^(Eds.): ‘Book Leveraging Digital Twins for ON/OFF-LINE Blockchain Networks: A Resilient Framework for Drone and IoT Communication’ (IEEE, 2025, edn.), pp. 426-430

4 Bernstein, D.J.: ‘ChaCha, a variant of Salsa20’, in Editor (Ed.)^(Eds.): ‘Book ChaCha, a variant of Salsa20’ (Lausanne, Switzerland, 2008, edn.), pp. 3-5

5 Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H.B., Patel, S., Ramage, D., Segal, A., and Seth, K.: ‘Practical Secure Aggregation for Privacy-Preserving Machine Learning’. Proc. Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, Dallas, Texas, USA2017 pp. Pages

6 Colosimo, F., and De Rango, F.: ‘Median-krum: A joint distance-statistical based byzantine-robust algorithm in federated learning’, in Editor (Ed.)^(Eds.): ‘Book Median-krum: A joint distance-statistical based byzantine-robust algorithm in federated learning’ (2023, edn.), pp. 61-68

7 Perrin, T.: ‘The noise protocol framework’ (2018. 2018)

Non-IID 데이터 기반 연합학습에서 채널별 대칭 양자화에 따른 통신 효율성 및 정확도 분석

이동욱, 이웅희*
한성대학교
{1971455, whlee}@hansung.ac.kr

Communication Efficiency and Accuracy Analysis of Federated Learning on Non-IID Data with Channel-Wise Symmetric Quantization

Dongwook Lee, Woonghee Lee*
Hansung University

요 약

본 논문은 non-IID 데이터 기반 연합학습 환경에서 클라이언트들이 학습 후 모델을 업로드 시 원본 모델과 채널별 양자화 모델 간에 발생하는 통신량 차이와 글로벌 정확도 변화 양상을 실험적으로 분석한다. 이를 통해 채널별 양자화가 non-IID 데이터 기반 연합학습 환경에서 전송 비트를 크게 줄이면서도 원본 모델 대비 동일 수준 이상의 정확도가 도출됨을 확인한다. 이러한 분석을 바탕으로, 본 논문은 통신 제약 환경에서의 연합학습을 위한 모델 전송 정책 수립에 필요한 실증적 정보를 제공한다.

Ⅰ. 서 론

최근 연합학습은 개인정보를 중앙에 수집하지 않고도 모델을 학습할 수 있다는 장점 때문에 엣지·모바일 환경에서 널리 활용되고 있다. 그러나 라운드마다 클라이언트가 서버로 모델을 전송해야 하므로, 통신 자원이 제한된 환경에서는 전송 비용이 학습 성능을 좌우할 수 있다. 본 논문에서는 경량 CNN 모델인 SqueezeNet[1]을 사용하고 연합학습 방식으로 널리 쓰이는 FedAVG[2]를 적용한 환경에서 모델 업로드 시 원본 모델 대비 채널별 양자화[3] 적용에 따른 통신량 절감과 전역 정확도 변화를 실험적으로 분석한다.

Ⅱ. 본론

본 논문에서는 채널별 대칭 양자화 기법을 사용하였다. 본 기법은 각 채널의 가중치 분포를 0을 중심으로 한 균일 정수 격자에 사상하는 방식이다. 채널마다 하나의 스케일 s_c를 두어 실수 가중치 $w_{c,i}$를 정수 범위 $[q_{\min}, q_{\max}]$의 최근접 격자점으로 표시한 뒤, s_c로 복원을 수행한다. 구체적 정의는 식(1)-(3)에 제시한다.

식 (1)에서 채널 c의 스케일 s_c는 해당 채널의 실수 가중치 $w_{c,i}$ 중 최대 절댓값을 양자화 범위의 최댓값 $q_{\max}$로 나누어 정의한다.

$$s_c = \frac{\max_i |w_{c,i}|}{q_{\max}} \quad \text{.....(1)}$$

식 (1)에서 도출된 채널별 스케일 s_c를 바탕으로 식 (2)에서는 채널별 가중치 $w_{c,i}$를 채널 c의 스케일 s_c로 나누어 반올림 후 $[q_{\min}, q_{\max}]$로 클리핑하여 정수 $q_{c,i}$로 변환하며, 이를 양자화 가중치로 사용한다.

$$q_{c,i} = \text{clip}(\text{round}(\frac{w_{c,i}}{s_c}), q_{\min}, q_{\max}) \quad \text{.....(2)}$$

식 (2)에서 도출된 양자화 가중치를 바탕으로 식 (3)은 채널 스케일 s_c를 곱해 양자화 가중치를 실수 가중치로 복원한다.

$$\hat{w} = s_c \cdot q_{c,i} \quad \text{.....(3)}$$

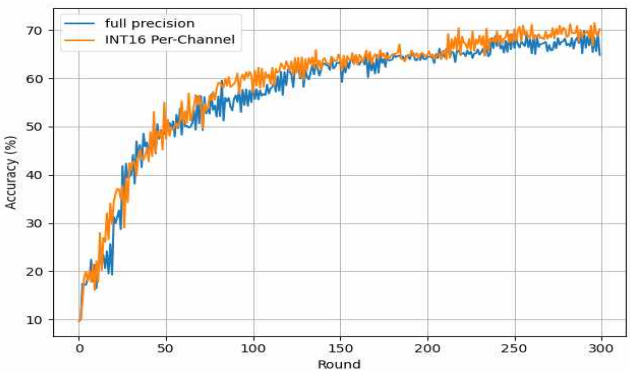


그림 1. 업로드 시 원본 모델과 양자화 모델간의 정확도 비교

그림 1은 CIFAR-10 데이터셋에 대해 Drichlet 분포(α=0.5)로 non-IID 환경을 구성하고, 클라이언트 기기 10개, 로컬 학습 5회, 글로벌 라운드 300, INT16을 양자화 비트로 설정하여 수행한 실험 결과를 나타낸다. 실험 결과, 원본 모델 대비 모든 라운드 구간에서 유사하거나 소폭 상승한 정확도를 보였고, 특히 초기 라운드에서 수렴 속도가 더 빠른 경향을 보였다. 이는 양자화가 정규화 효과를 제공하여 각 클라이언트의 로컬 데이터셋에 대한 과적합을 억제하면서 전역 일반화 성능을 개선할 수 있음을 시사한다.

Model	Traffic (MB)
SqueezeNet(FP32)	28.41
SqueezeNet(INT16)	14.66

표 1. 모델 업로드 시 통신량 크기 비교

표 1은 매 라운드마다 업로드 되는 통신량을 비교한 결과이다. 원본 모델은 28.41MB, INT16 양자화 모델은 14.66MB로 약 48.4%의 전송량 감소가 확인되었으며, 이는 통신 제약적인 환경에서도 효율적인 연합학습 운용이 가능함을 시사한다.

Ⅲ. 결론

본 논문에서는 non-IID 데이터 기반 연합학습 환경에서 모델 업로드 시 원본 모델 대비 채널별 INT16 양자화의 효과를 실험적으로 분석하였으며, 실험 결과 약 48%의 전송량 감소와 소폭 향상된 전역 정확도 및 초기 수렴 가속을 보여주었다. 이는 통신 제약 환경에서도 효율적인 연합학습 운용과 모델 전송 정책 수립이 가능함을 시사한다. 후속 연구에서는 클라이언트별 통신 상태를 인지하여 클라이언트별 모델 업로드 시 양자화를 동적으로 조정하는 적응형 양자화 연합학습 기법을 고안할 계획이다.

ACKNOWLEDGMENT

이 논문은 2025년 정부(방위사업청)의 재원으로 국방기술진흥연구소의 지원을 받아 수행된 연구임(KRIT-CT-23-041).

참 고 문 헌

[1] Iandola, Forrest N. “SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and<0.5MB model size.” arXiv preprint arXiv:1602.07360 (2016).

[2] McMahan, Brendan, et al. “Communication-efficient learning of depp networks from decentralized data.” Artificial intelligence and statistics. PMLR, 2017.

[3] Raghuraman Krishnamoorthi. Quantizing deep convolutional networks for efficient inference: A whitepaper. arXiv preprint arXiv:1806.08342, 2018.

K-공간 MRI에서 어텐션 메커니즘을 이용한 알츠하이머병 검출의 개념적 접근법

안헬리나, 김상철
국민대학교
mimili1208@kookmin.ac.kr, sckim7@kookmin.ac.kr

Application of Attention Mechanisms for Alzheimer's Disease Detection in k-space MRI: Conceptual Approach

Belausava Anhelina, Kim Sang-Chul
Kookmin Univ.

요 약

현재 치매, 특히 알츠하이머병의 초기 징후를 확인하거나 이미 진행 중인 상태를 모니터링하기 위해 신경학 및 신경외과 분야에서 가장 효과적인 방법으로 MRI 검사를 활용하고 있다. 일반적으로 MRI 결과는 시각적 이미지로 해석할 수 있는데, 딥러닝의 발전으로 치매 뇌와 정상 뇌의 자동 분류 및 예측이 가능해졌다[1]. 현재 주요 동향은 임상의가 주의를 기울일 만한 뇌 영역에 집중하여 attention 메커니즘을 정교화하는 소위 "의사의 시각"에 의해 주도되고 있다. 그러나 재구성된 기존 MRI 이미지의 본질적 기반이 무엇인지 분석하면, 최근 연구 관심사의 또 다른 주제인 주파수 표현, 즉 k-space의 작업에 도달하게 된다. 본 연구에서는 주파수 영역 데이터에 대한 attention 메커니즘을 통합하고 비교하는 방향을 검토하며, 보다 정확한 치매 진단에 대한 가능한 기여를 고찰할 것이다.

Ⅰ. 서 론

K-space는 이미지의 주파수 성분을 인코딩하는 점들로 이루어진 2D 또는 3D 행렬 표현이다. 이러한 주파수는 중심에서 주변으로 이동하면서 저주파와 고주파로 나뉘며, 각각은 전체 구조(global structures) 또는 세부 구조(fine details)를 나타낸다[2]. 따라서 이러한 데이터는 뇌의 해부학적 구조를 직접적으로 저장하지 않으며, 특정 영역에 국한되지 않고 각 요소가 전체 뇌에 대한 정보에 기여한다는 점을 이해해야 한다. 이는 attention 메커니즘을 k-space에 적용할 때의 어려움 중 하나를 나타내며, 이미지에서 잘 학습된 attention은 성능 향상을 위해 종종 ROI(Region of Interest)를 활용한다[3].

그럼에도 불구하고 k-space 데이터에 대한 연구 관심이 급속히 증가하고 있는데, 이러한 현상의 배경에는 몇 가지 핵심적인 요인들이 존재한다. 첫째, 재구성된 MRI 이미지와 비교했을 때, k-space는 역푸리에 변환(Inverse Fourier Transform, IFT) 과정에서 발생할 수 있는 손실 없이 최대한의 정보를 포함한다. 둘째, Sparse Reconstruction of Brain MRI와 같은 연구들은 k-space를 활용함으로써 불완전한 데이터로도 모델을 보다 효율적으로 학습할 가능성을 보여주며, 이는 계산 비용을 줄이는 데에도 기여한다[4][5].

이러한 k-space의 고유한 특성들을 효과적으로 활용하기 위해서는 적절한 분석 방법론이 필요하며, 이 지점에서 attention 메커니즘의 도입이 주목받고 있다[6]. Attention 메커니즘의 핵심 특징은 각 개별 특징의 중요도(weight)를 계산할 수 있다는 것이다[7]. 특히 뇌 MRI 평가에서, 특정 뇌 영역에 주목하면서 결과를 올바르게 해석하는 것이 중요한 의학적 실무와 밀접하게 관련된다. 그러나 앞서 언급했듯이 attention 메커니즘은 이미지, 오디오 등에서 잘 학습되어 있으며[8], 이제 그 효과가 주파수 영역에서도 얼마나 유효한지를 알아보고자 한다.

Ⅱ. 본 론

본 논문에서는 SE(Squeeze-and-Excitation Networks)와 CBAM(Convolutional Block Attention Module)와 같은 attention 메커니즘의 적용을 고려한다. 이들의 선택은 단순한 channel attention과 channel attention과 spatial attention의 조합을 비교한다는 관점에서도 흥미롭다. 채널에만 적용되는 attention은 공간적 의존성을 반영하지 못할 수 있으며, k-space에서는 중심을 기준으로 한 주파수 분포로 이를 고려할 수 있다. 그러나 이것은 여전히 일반적인 해부학적 표현과 일치하지 않으므로 spatial attention의 중요성과 필요성은 실험적으로 확인할 필요가 있다. 특히 k-space 도메인에서 spatial attention이 의미하는 '공간'은 주파수 좌표계 상의 위치를 나타내므로, 기존 영상 기반 모델에서 검증된 spatial attention의 효과가 동일하게 적용될 수 있는지에 대한 체계적 검증이 요구된다.

MRI 자체와 관련하여 잘 알려진 바와 같이, 치매의 가장 뚜렷한 지표는 대뇌 피질(cerebral cortex)의 위축이다[9]. 데이터의 주파수 표현은 시각적 표현을 제공하지 않으므로, 대뇌 피질이 gray matter로서 white matter와 다르다는 사실을 바탕으로, 그 위축에 대한 결론은 대비와 경계의 변화를 기반으로 도출된다. 앞서 언급했듯이 이러한 경계는 높은 주파수(high frequencies)로 표현된다. 이를 통해, 피질의 위축은 k-space에서 high frequencies의 기여와 low frequencies와의 균형 변화를 초래할 것이라고 가정할 수 있다.

이러한 저주파와 고주파 분포의 차이는 attention 메커니즘이 특징 가중치를 결정하는 주요 마커로 작용할 것으로 예상된다. 또한 여기서 중요한 feature의 weight가 실제 생리학적 의미와 어떻게 대응되는지 평가하는 것이 여전히 필수적이다.

Ⅲ. 결 론

본 연구에서 제안된 방향에 따라 연구의 주요한 어려움과 한계를 다음과 같이 정리할 수 있다.

먼저 주파수 해석 측면에서 k-space 데이터를 어떻게 해석할 것인가의 근본적인 문제가 제기된다. k-space 데이터의 본질적 특성상 이를 오디오 신호 처리 및 시간-주파수(time-frequency) 분석 기법에 가까운 파동적 표현으로 접근할 것인지, 아니면 컴퓨터 비전 및 영상 처리 기법에 가까운 점적(point-like) 표현으로 다룰 것인지에 따라 적용 가능한 어텐션 메커니즘의 설계 방향이 근본적으로 달라진다. 이러한 해석의 차이는 단순히 기술적 선택의 문제를 넘어서 모델의 학습 효율성과 최종 성능에 직접적인 영향을 미치게 된다. 또한 계산 복잡성 관점에서 k-space 데이터를 직접 활용한 학습과 처리 과정은 기존의 재구성된 MRI 영상을 사용하는 방식 대비 현저히 많은 계산 자원과 메모리를 요구할 가능성이 높다[10]. 특히 고해상도 MRI 데이터의 경우 k-space에서의 복소수 연산과 다차원 푸리에 변환 과정에서 발생하는 계산 부하는 실시간 진단 환경에서의 실용성을 제한할 수 있다.

어텐션 메커니즘 적용에 있어서는 공간적 의존성 해석의 복잡성이 중요한 도전 과제로 대두된다. channel attention 메커니즘의 경우 각 채널을 독립적인 주파수 성분으로 해석할 수 있다는 가정 하에 설계되지만, k-space에서의 spatial attention이 다루는 "공간" 개념은 일반적인 해부학적 공간 좌표계와는 본질적으로 다른 주파수 도메인 상의 위치를 의미한다. 이는 기존 영상 기반 딥러닝 모델에서 확립된 공간적 상관관계 해석과 근본적인 차이를 보이므로, k-space 도메인에 특화된 어텐션 메커니즘의 이론적 근거 수립과 체계적 검증이 선행되어야 한다[11]. 더불어 특징 추출의 제한성 측면에서 본 연구는 알츠하이머병의 주요 영상의학적 지표인 피질 두께 감소(cortical thinning) 현상을 중심으로 접근했으나, 이러한 단일 지표 중심의 분석은 해마 위축(hippocampal atrophy), 뇌실 확장(ventricular enlargement), 백질 병변(white matter lesions) 등 알츠하이머병에 수반되는 다양한 병리학적 뇌 구조 변화를 종합적으로 반영하지 못할 수 있다는 한계를 내포한다. 이는 궁극적으로 모델의 진단 민감도와 특이도에 영향을 미쳐 임상적 유용성을 제한할 가능성이 있다.

그럼에도 불구하고 k-space 도메인에서의 어텐션 메커니즘 적용은 기존 영상 기반 접근법의 한계를 극복할 수 있는 매우 유망한 연구 방향으로 평가된다. 특히 MRI 신호 획득 과정에서 발생할 수 있는 아티팩트나 노이즈에 대한 강건성 확보, 그리고 영상 재구성 과정에서 손실될 수 있는 미세한 주파수 정보의 보존 측면에서 상당한 이점을 제공할 것으로 기대된다. 따라서 이러한 접근 방식은 향후 정밀 의료 환경에서의 조기 진단 시스템 구축과 임상 현장에서의 실용화를 위한 중요한 기술적 기반이 될 것으로 전망된다.

참 고 문 헌

[1] Jin, Z., Gong, J., Deng, M., Zheng, P., Li, G. "Deep Learning-Based Diagnosis Algorithm for Alzheimer's Disease," J. Imaging 2024, 10, 333. https://doi.org/10.3390/jimaging10120333

[2] Passigan, P., Ramkumar, V. "Analyzing the Effect of k-Space Features in MRI Classification Models," arXiv:2409.13589v1 [eess.IV] 20 Sep 2024.

[3] Hossain, M.S., Shahriar, G.M., Syeed, M.M.M. et al. "Region of interest (ROI) selection using vision transformer for automatic analysis using whole slide images," Sci Rep 13, 11314 (2023). https://doi.org/10.1038/s41598-023-38109-6

[4] Du, T., Zhang, Y., Shi, X., Chen, S. "Multiple Slice k-space Deep Learning for Magnetic Resonance Imaging Reconstruction," Annu Int Conf IEEE Eng Med Biol Soc. 2020 Jul;2020:1564-1567. https://doi.org/10.1109/embc44109.2020.9175642

[5] Du, T., Zhang, H., Li, Y., Pickup, S., Rosen, M., Zhou, R., Song, H. K., & Fan, Y. "Adaptive convolutional neural networks for accelerating magnetic resonance imaging via k-space data interpolation," Medical image analysis 2021, 72, 102098. https://doi.org/10.1016/j.media.2021.102098

[6] Gösche, Erik, "Attention-based networks for brain segmentation in k-space," 2023.

[7] Hu, J., Shen, L. and Sun, G. "Squeeze-and-Excitation Networks," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, pp. 7132-7141. https://doi.org/10.48550/arXiv.1709.01507

[8] Xie, Y., Yang, B., Guan, Q., Zhang, J., Wu, Q., Xia, Y. "Attention Mechanisms in Medical Image Segmentation: A Survey," 2023. https://doi.org/10.48550/arXiv.2305.17937

[9] Nicastro, N., Malpetti, M., Mak, E., et al. "Gray matter changes related to microglial activation in Alzheimer's disease," Neurobiol Aging. 2020;94:236-242. doi:10.1016/j.neurobiolaging.2020.06.010

[10] Han, Y., Sunwoo, L. and Ye, J. C. " k-Space Deep Learning for Accelerated MRI," in IEEE Transactions on Medical Imaging, vol. 39, no. 2, pp. 377-386, Feb. 2020, https://doi.org/10.1109/tmi.2019.2927101

[11] Gösche, E., Eghbali, R., Knoll, F., & Rauschecker, A.M. (2024). Domain Influence in MRI Medical Image Segmentation: Spatial Versus k-Space Inputs. In Xuanang Xu, Zhiming Cui, Islem Rekik, Xi Ouyang, Kaicong Sun (Eds.), Machine Learning in Medical Imaging (pp. 310-319). Marrakesh, MA: Cham: Springer. https://doi.org/10.1007/978-3-031-73284-3_31

VAE-GMM 기반 생체모방 통신용 휘슬 시퀀스 생성 방법

박근호, 안종민, 김완진, 김인수, 이동훈
국방과학연구소
ghpark57935@add.re.kr, leedhun@add.re.kr

VAE-GMM-based Whistle Sequence Generation Method for Biomimetic Underwater Communication

Park Geun-Ho, Ahn Jongmin, Kim Wanjin, Kim In-Soo, Lee Donghun
Agency for Defense Development

요 약

본 논문은 생체모방 수중통신용 인공 휘슬음의 시퀀스 생성 방안을 제시한다. 본 논문에서 제안하는 생체모방용 휘슬의 시퀀스 생성 방법은 크게 심층 생성 신경망 기반 휘슬의 시간-주파수 생성, 가우시안 컴포넌트 별 주파수 할당, 휘슬 간 간격 모델링의 세 단계로 구성된다. 휘슬 형태는 변이형 오토인코더와 혼합 가우시안 밀도를 학습하여 생성하도록 구현했다. 중심 주파수는 각 가우시안 컴포넌트를 생성하는 데에 기여한 휘슬의 주파수 중에서 단순임의추출 (Simple Random Sampling)하여 결정한다. 마지막으로 휘슬 간 간격은 휘슬 추출 레이블로부터 휘슬 간 간격 분포를 일반화된 파레토 분포 (Generalized Pareto Distribution)로 모델링한 후 샘플링하였다.

Ⅰ. 서 론

돌고래와 고래는 휘슬음과 광대역 펄스 형태의 클릭음을 활용하여, 의사소통 및 먹이탐색 등을 수행하는 것으로 알려져 있다. 특히 고래류는 고정된 주파수 변조 방법을 적용한 시그니쳐 휘슬 신호로 자신을 알리거나, 상황에 적합한 휘슬음을 발생시켜 의사소통에 활용하기도 한다.

생체모방 수중통신은 고래류가 발생시키는 휘슬음을 주로 모방한다[1]. 감청자는 휘슬음 모방 통신 신호를 통신 신호로 인식하기 어려우므로, 아군의 은밀 통신을 가능하게 한다.

고래의 휘슬음의 시간-주파수 형태는 매우 다양하므로, 결정론적이지 않고 확률 분포로 나타내는 것이 적절하다. 그러므로 결정론적 심볼을 사용하는 기존 디지털 통신과는 다른 접근이 필요하다.

이러한 접근 방법의 하나로, 생체모방 통신 신호의 송신기는 휘슬의 확률 분포에서 표본을 얻는 방법으로 송신 신호를 결정할 수 있다. 이 방법의 이점은 감청자가 해당 신호를 수집하더라도, 아군의 신호 송신 체계를 파악하기 위해 많은 자료를 수집하고 분석하도록 강요한다는 점이다.

휘슬음을 확률적으로 생성하는 방법은 심층 생성 신경망을 적용하는 것이다. 본 논문에서는 [2]에서 제안한 Variational Autoencoder (VAE)와 Gaussian Mixture Model (GMM)을 학습하여 휘슬 형태를 샘플링하도록 하였다.

본 논문의 목표는 단일 휘슬 형태를 모방하는 것을 넘어서 휘슬음을 연속적으로 생성한 결과도 데이터 분포에 근접하도록 설계하는 것이다. 이를 위해서는 휘슬의 중심 주파수와 휘슬음 간격까지 고려해야 한다.

하지만 단일 휘슬 형태만 생성하는 VAE-GMM은 중심 주파수와 연속적 휘슬 간격을 고려하지 않으므로, 주파수와 휘슬 간격을 생체 신호와 유사하게 샘플링할 수 있는 별도의 방안이 필요하다. 따라서 본 논문에서는 연속 휘슬 시퀀스 생성 문제에서 휘슬의 주파수는 단순임의추출 (Simple Random Sampling)하도록 설정했고, 휘슬음 간격의 확률 분포는 간격 데이터를 활용해 일반화된 파레토 분포 (Generalized Pareto Distribution)로 모델링하였다. 관련 세부내용은 Ⅱ장부터 Ⅳ장에서 다룬다.

Ⅱ. VAE-GMM 기반 휘슬음 생성 방법

VAE-GMM 기반 휘슬 생성 방법[2]는 휘슬음의 형태를 이미지 형태로 표현한 입력 데이터에 대해 잠재 변수를 추정하는 VAE와 잠재 변수의 주변부 분포를 다수의 가우시안 분포로 모델링하는 GMM을 적용한 것이다. 휘슬을 이미지로 표현했을 때 대부분 0에 근접하는 픽셀 강도를 가지므로, 128×128 크기의 이미지를 64×1 잠재 벡터로 압축할 수 있었다. 잠재 변수의 분포, 즉 생성 모델에서의 사전 분포는 10,000개의 가우시안 분포로 모델링 하였다. 가우시안 분포의 매개변수는 Expectation-Maximization (EM) 알고리즘으로부터 추정하였다. 또한 VAE 학습 시, Wassertein Generative Adversarial Network-Gradient Penelty (WGAN-GP)를 적용하여 데이터와 유사한 형태가 되도록 신경망을 학습했다. 자세한 내용은 [2]를 참고할 수 있다.

Ⅲ. 휘슬 간격 및 주파수 샘플링 방법

휘슬 간 간격에 대한 데이터는 미국 해양대기청에서 공개하는 Detection, Classification, Localization and Density Estimation (DCLDE) Oahu [3]의 흑범고래 데이터에서 추출하였다. 먼저 DCLDE 2022 흑범고래 데이터의 레이블 작업을 통해 32,178개 휘슬의 시간-주파수 데이터를 확보한 후, 이 휘슬의 시작 시간 정보로부터 31,991개의 간격 샘플을 얻었다. 휘슬음 간격의 히스토그램은 그림 1에 나타내었다.

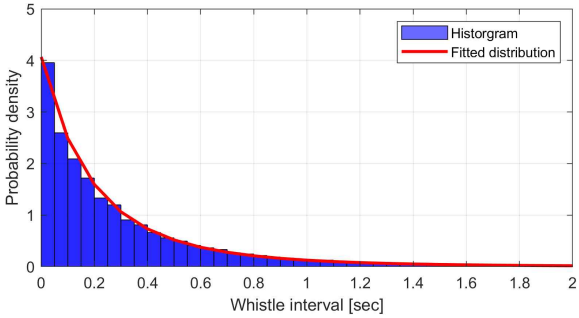


그림 1. 휘슬 간격 데이터 및 확률 분포 추정 결과.

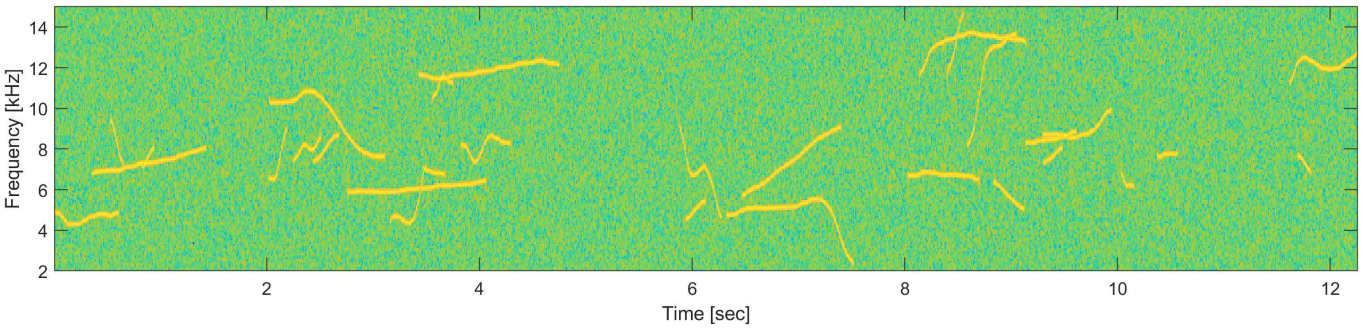


그림 2. VAE-GMM 기반 휘슬 시퀀스 생성 결과 1

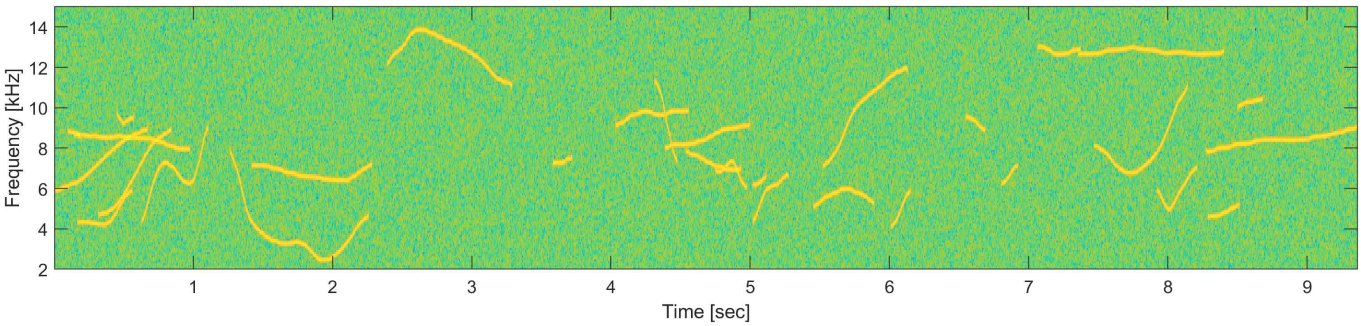


그림 3. VAE-GMM 기반 휘슬 시퀀스 생성 결과 2

본 논문은 그림 1의 휘슬 간격 데이터로부터 적색 실선의 확률 분포를 추정하였다. 확률 분포 모델은 일반화 파레토 분포 (Generalized Pareto Distribution)를 적용하였고, 이 분포의 shape, scale 및 threshold 파라미터 k, σ, θ는 각각 0.2753, 0.2458, 0과 같다.

휘슬의 중심 주파수는 GMM의 각 가우시안 컴포넌트에 주파수 샘플을 할당한 후, 단순임의추출 하도록 설정하였다. 예를 들어, 5,113번째 가우시안 분포를 생성하는 데에 기여한 5개의 휘슬 데이터가 있고, 이 데이터의 중심 주파수가 각각 6.9, 7.0, 13.8, 14.0, 4.9 kHz라면, 5,113번째 가우시안 분포로부터 휘슬을 생성할 때에 중심 주파수는 이 5개 중 하나가 임의 추출되도록 하였다.

이와 같이 주파수 생성을 위해 단순임의추출을 적용한 이유는 각 가우시안 컴포넌트에 할당된 데이터의 수가 적은 경우가 매우 많다는 점, 그리고 주파수 데이터가 가우시안 분포에 적합하지 않다는 점 때문이다. 실제 64×1 크기의 잠재 변수에 주파수 데이터 차원을 추가하여 65×1 크기의 벡터를 GMM으로 학습하도록 했을 때, 중심 주파수에 대한 가우시안 분포의 분산이 매우 커지는 한계가 발생했다. 따라서 중심주파수에 대해 단순임의추출 방법이 가장 효과적이라고 판단하였다.

Ⅳ. 휘슬 시퀀스 생성 결과

그림 2와 3은 다음 세 단계를 거쳐 30개의 휘슬을 연속 생성한 결과를 나타낸 것으로 다음 세 단계를 거친다.

(1) VAE-GMM 기반 휘슬 형태 샘플링
(2) 단순임의추출 기반 주파수 샘플링
(3) 일반화된 파레토 분포에서 휘슬 간격 샘플링

그림 2와 3에서 신호대잡음비는 -5 dB, 샘플링 주파수는 192 kHz, 스펙트로그램 FFT 크기는 2,048, 중첩 샘플 수는 1,024로 설정하였고, Hamming 윈도우를 적용하였다.

Ⅴ. 결 론

본 논문에서는 VAE-GMM 기반 휘슬 형태 샘플링, 단순임의추출 기반 주파수 샘플링, 일반화된 파레토 분포 기반 휘슬 간격 샘플링의 세 단계를 통해 휘슬 시퀀스를 생성한 결과를 제시하였다. 연속적인 휘슬 생성 결과, 데이터와 유사한 형태를 갖는 것을 확인할 수 있었다. 향후에는 휘슬 합성 결과와 원 데이터 사이의 확률 분포 차이를 수치적으로 분석하여 도출할 계획이다.

ACKNOWLEDGMENT

이 논문은 2025년 정부의 재원으로 국방과학연구소에서 수행된 연구 결과임.

참 고 문 헌

[1] Seungwhan Seol, *et. al.*, “Research trends of biomimetic covert underwater acoustic communication,” *The Journal of the Acoust. Soc. of Korea*, vol. 41, no. 2, pp. 227-234, 2022.

[2] G.-H. Park, J. Ahn, W. Kim, I. Kim, and D.-H. Lee, “A Generative Model for Cetacean Whistles Using Variational Autoencoder and Mixture of Gaussians,” *IEEE Access* (to be submitted).

[3] NOAA National Centers for Environmental Information (2017). *Passive Acoustic Data Collection*. Asheville, NC: NOAA National Centers for Environmental Information. doi: 10.25921/PF0H-SQ72.

국방 객체 데이터 부족 문제 해결을 위한 합성데이터 생성

김병현*, 정남훈, 조규태
LIG넥스원
byunghyun.kim@lignex1.com, namhoon.jung@lignex1.com, kyutae.cho2@lignex1.com

Synthetic Data Generation for Military Objects under Data Scarcity

Byunghyun Kim*, Namhoon Jung, Kyutae Cho
LIG Nex1

요 약

인공지능(AI) 모델 성능은 학습 데이터의 양과 질에 크게 의존한다. 하지만 국방 분야는 보안상의 제약뿐만 아니라 실제 데이터를 수집하는 것 자체가 어려운 근본적인 한계를 가지고 있다. 본 논문에서는 국방 분야에서 데이터 확보가 어려운 객체를 활용한 AI 모델 성능 향상을 위해 Stable Diffusion 모델을 기반으로 LoRA(Low-Rank Adaptation) 학습을 통한 합성데이터 생성 방법을 제안하였다. 제안된 방법은 소량의 실제 이미지 데이터와 프롬프트를 입력으로, 특정 국방 객체에 대한 다양하고 사실적인 이미지를 생성한다. 이러한 합성데이터 생성 방법은 국방 분야에서 데이터 부족 문제를 완화하여 국방 AI 연구 발전에 활용될 것으로 기대된다.

1. 서 론

인공지능(AI) 기술이 발전함에 따라 이미지 분류와 객체 탐지 분야에서도 뛰어난 성과들이 보고되고 있다. 그러나 국방 분야에서는 이미지 분류와 객체 탐지 등 AI를 활용한 연구가 상대적으로 어려움을 겪고 있다. 이는 국방 관련 데이터가 보안상의 제약뿐만 아니라, 실제 데이터를 수집하는 것 자체가 어려운 특수한 환경에 놓여있기 때문이다. 따라서 국방 AI 연구에 필요한 고품질 데이터 확보는 중요한 문제이며, 이를 해결하기 위해 다양한 데이터 생성 연구가 활발히 진행되고 있다. 기존 연구에서는 GAN(Generative Adversarial Network)[1]을 활용하여 합성데이터를 생성하거나 Rotation, Flip, Crop 등의 증강 기법을 통해 데이터를 확보하는 연구가 있었다. 하지만 GAN은 세밀한 부분을 구현하기 어렵고, 데이터 증강 기법은 다양성을 확보하는 데 한계가 있다. 따라서 보다 정교하고 현실감 있는 합성데이터 생성 방법이 요구된다. 합성데이터는 수집 비용이 적고 다양한 조건을 반영할 수 있어, 제한된 환경에서도 AI 모델의 유용한 학습 자원이 될 수 있다. 특히 국방처럼 데이터 접근이 제한적인 분야에서 도메인 특화된 데이터 생성을 통해 충분한 학습 효과를 기대할 수 있다. 최근 주목받고 있는 Text-to-image 모델인 Stable Diffusion[2]은 기존 기법들에 비해 뛰어난 이미지 생성 품질과 유연성을 보여주고 있다. 특히 노이즈 제거 과정을 통해 고해상도 이미지를 생성하며, 프롬프트 조건을 활용하여 구체적인 이미지 생성도 가능하다. 또한 LoRA(Low-Rank Adaptation)[3]와 같은 경량화 학습 기법을 적용하여, 적은 수의 데이터만으로도 특정 객체에 특화된 모델을 효율적으로 학습시킬 수 있다.

본 논문에서는 Stable Diffusion 모델을 기반으로 LoRA 학습을 통해 국방 분야에서 데이터 확보가 어려운 객체 이미지를 생성하는 방법을 제안한다. 소량의 실제 이미지 데이터와 프롬프트를 기반으로 다양한 조건에서 현실감 있는 이미지를 생성하고, 정성적 평가를 통해 검증하였다. 이를 통해 국방 AI 개발에 필요한 데이터 부족 문제를 완화하고, 향후 국방 AI 연구 및 실무 적용에 기여하고자 한다.

2. 국방 객체 합성데이터 생성

본 논문에서는 국방 객체 학습 데이터를 생성하기 위해 Stable Diffusion 기반의 합성데이터 생성 방법을 제안한다. 합성데이터 생성 과정은 태그 생성, 프롬프트 구성, 이미지 생성의 세 단계로 이루어지며, 전체 구조는 그림 1과 같다.

2.1. 태그 생성 (Tag Generation)

태그 생성은 LoRA 학습을 위한 핵심 전처리 과정으로, 실제 이미지에 포함된 객체의 의미와 시각적 특징을 텍스트 형태로 명시화하는 단계다. 태그 생성을 위해 소량의 실제 국방 객체 이미지를 수집한 후, WD14 Tagger[4] 모델을 활용하여 각 이미지의 시각 정보를 반영한 태그를 생성한다. 생성된 태그는 LoRA 학습에 사용되며, 이후 추론 단계에서 입력되는 프롬프트에 해당 태그가 등장했을 때 의미 있는 이미지로 재구성될 수 있도록 하는 역할을 한다.

2.2. 프롬프트 구성 (Prompt Construction)

프롬프트는 Text-to-image 모델의 입력으로 사용되는 문장 형태의 조건 정보로서, 어떤 객체를 어떤 장면에서 어떤 스타일로 생성할지를 결정하는 중요한 입력이다. 특정 객체를 LoRA를 통해 학습시킨 후, 해당 객체가 프롬프트에 명시적으로 등장해야만 생성 이미지 내에서 해당 객체가 등장한다. 따라서 프롬프트에는 정확한 객체 또는 태그명이 반드시 포함되어야 한다. 프롬프트는 “A photo of a single <category_name>”와 같이 Simple prompt[5]로 구성하였다.

2.3. 이미지 생성 (Image Generation)

Stable Diffusion 모델을 개선한 SDXL[6]을 활용하여, 주어진 프롬프트로부터 현실감 있는 이미지를 생성한다. 특히, LoRA 학습은 소량의 학습 데이터만으로도 원하는 객체를 생성할 수 있다. 이 단계는 프롬프트에 따라 다양한 배경, 시점, 조건을 가진 객체 이미지를 자동 생성하여 AI 모델 학습에 활용가능한 대규모 데이터셋을 확보하는 데 목적이 있다.

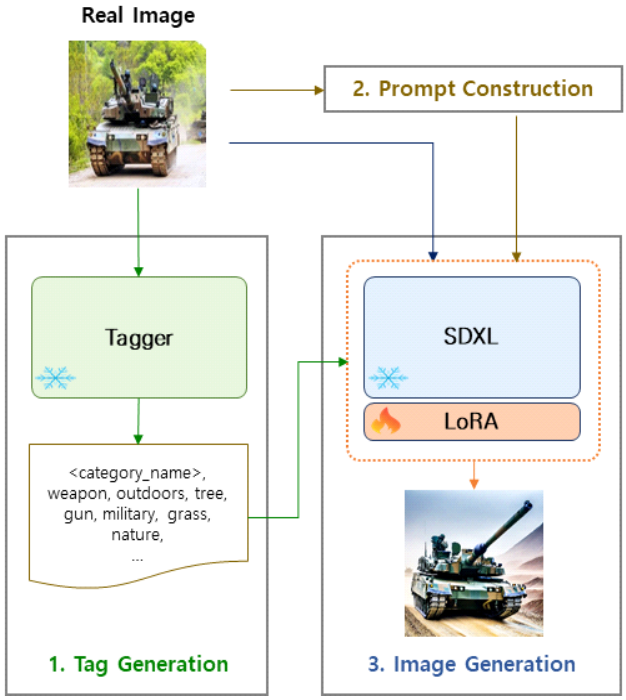


그림 1 국방 객체 합성데이터 생성 방법

그림 3 LoRA 적용 유무에 따른 시각적 비교

3. 실험 결과

실험에서 생성할 국방 객체는 K2 전차로 선정하였고, 데이터가 부족한 상황을 가정하여 인터넷에서 수집 가능한 K2 전차 이미지 10장을 LoRA 학습 데이터로 사용하였다. WD14 Tagger, SDXL 모델은 가중치를 고정했으며, 학습 시 LoRA 가중치만 업데이트하였다. 프롬프트는 Simple prompt를 사용하였고, <category_name>은 “tank”로 설정하였다. 그림 2는 합성데이터 생성의 다양성을 위해 프롬프트에 배경이나 객체의 이동 방향을 포함하여 K2 전차 이미지를 생성한 결과이다. 또한 ControlNet[7]을 적용하여 객체의 위치 및 자세를 제어하고, 학습된 LoRA를 적용한 경우와 그렇지 않은 경우로 나누어 비교하였다. ControlNet 입력은 Visual Prior 중 Depth map을 사용했으며, LoRA를 적용하지 않고 생성한 이미지로부터 추출하였다. 그림 3과 같이, 동일한 Depth map을 사용하여 전차를 생성하더라도 학습된 LoRA를 적용한 경우, 일반적인 전차가 아닌 K2 전차의 특징이 반영된 것을 확인할 수 있다.

그림 2 K2 전차 합성데이터 생성 결과

4. 결 론

본 논문에서는 국방 AI 모델 개발에 제약 중 하나인 학습 데이터 부족 문제를 해결하기 위해 Stable Diffusion 기반 모델과 LoRA 학습을 통한 합성데이터 생성 방법을 제안하였다. 제안된 방법은 태그 생성, 프롬프트 구성, 이미지 생성의 세 단계로 구성되며, 소량의 실제 이미지와 프롬프트를 기반으로 특정 국방 객체에 대한 다양하고 사실적인 이미지를 생성한다. 이러한 합성데이터 생성은 기존에 확보가 어려웠던 국방 객체에 대한 학습용 데이터를 대체하거나 보완하는 데 활용될 수 있으며, 결과적으로 이미지 분류 또는 객체 탐지 등 국방 AI 모델의 성능 향상과 연구 개발 비용 절감에 기여할 수 있을 것이다.

참 고 문 헌

[1] Goodfellow, Ian J., et al. "Generative adversarial nets." Advances in neural information processing systems 27 (2014).

[2] Rombach, Robin, et al. "High-resolution image synthesis with latent diffusion models." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022.

[3] Hu, Edward J., et al. "Lora: Low-rank adaptation of large language models." ICLR 1.2 (2022): 3.

[4] SmilingWolf, WD 1.4 Tagger, Available: https://github.com/toriato/stable-diffusion-webui-wd14-tagger

[5] Zhao, Hanqing, et al. "X-paste: Revisiting scalable copy-paste for instance segmentation using clip and stablediffusion." International Conference on Machine Learning. PMLR, 2023.

[6] Podell, Dustin, et al. "Sdxl: Improving latent diffusion models for high-resolution image synthesis." arXiv preprint arXiv:2307.01952 (2023).

[7] Zhang, Lvmin, Anyi Rao, and Maneesh Agrawala. "Adding conditional control to text-to-image diffusion models." Proceedings of the IEEE/CVF international conference on computer vision. 2023.

무선 영상 전송 위한 이중 대역 Deep JSCC 적용 기법

이승현, 오성현, 김정곤*
한국공학대학교 전자공학부
seng0424@tukorea.ac.kr, osh119@tukorea.ac.kr, jgkim@tukorea.ac.kr*

Dual Band Deep JSCC for Wireless Transmission

Seung Hyun Lee, Sung Hyun Oh, Jeong Gon Kim*
Dept. of Electronic Engineering, Tech University of Korea

요 약

본 논문은 무선 채널 환경에서 자율주행 영상 전송을 위한 Dual-Band JSCC(Joint Source Channel Coding)을 제안한다. 입력 영상은 5×5 Gaussian Blur로 얻은 저주파 성분과 Residual 고주파 성분으로 분리되어 각각 독립적인 Encoder-Channel Decoder 경로로 전송된다. 각 경로의 Latent는 전력 정규화 후 동일 SNR(Signal to Noise Ratio)의 AWGN(Additive White Gaussian Noise) 채널을 통과한다. 수신단에서는 pre-Sigmoid 복원본을 1×1 Convolution-Sigmoid 게이트로 가중합해 최종 출력을 얻는다. 성능 검증은 Deep JSCC, UEP(Unequal Error Protection) 기반 Latent Separation 및 Layer Separation과 공정 비교로 수행되었으며, 동일 전송률과 전력 조건을 적용하였다. 학습 시에는 epoch·batch마다 랜덤 SNR을 주입해 실제 무선 환경을 모사했다. Cityscapes(256×256) 데이터셋 실험 결과, 제안 기법은 특히 낮은 SNR에서 PSNR(Peak Signal to Noise Ratio) 및 SSIM(Structual Similarity Index Map) 기준 일관된 성능 향상을 보였으며, 열악한 무선 환경에서 자율주행 시스템의 상황 인지 신뢰도를 높이는 실용적 대안이 될 수 있음을 시사한다.

Ⅰ. 서 론

자율주행 자동차는 V2X(Vehicle to Everything) 통신을 통해 주변 차량이나 인프라와 실시간으로 영상 정보를 교환해야 한다. 도심의 빌딩 숲이나 터널 등에서는 통신 신호가 약해지는 낮은 SNR(Signal to Noise Ratio) 환경이 빈번하게 발생하며, 이는 안전과 직결되는 심각한 문제이다. 기존 영상 코덱은 낮은 SNR 환경에서 Cliff Effect로 인해 영상이 깨지는 상황이 발생해 자율주행 시스템의 반응에 지연을 발생시킬 수 있다. 이러한 배경에서 Deep JSCC(Joint Source Channel Coding)가 제안되어 Source Coding과 Channel Coding을 단일 신경망을 통합함으로써, 제한된 대역·지연 조건에서 강건한 화질을 입증해왔다. 초기 연구들은 AWGN(Additive White Gaussian Noise) 채널에서 전통 분리 설계 대비 경쟁력 있는 PSNR(Peak Signal to Noise Ratio)·SSIM(Structural Similarity Index Map)을 보이며, 저지연 시나리오에서 실용성을 보여주었다[1]. 이후 후속 연구들로 다양한 채널 조건과 물리 계층 제약을 모델 내에 통합했다. Deep JSCC-f는 피드백을 활용해 동일 조건에서 화질 향상 및 평균 지연 감소를 달성했다[2]. 채널 추정값 없이도 다양한 SNR에 자체 적응하도록 하는 channel-blind JSCC도 제안되어 실사용 시의 SNR 불확실성 문제를 완화하였다[3]. 또한 ViT(Vision Transformer) 등 최신 시각 Backbone을 활용한 적응형 JSCC는 MIMO(Multiple Input Multiple Output) · 가변 SNR 환경에서의 표현력과 성능을 확장하고 있다. 최근 JSCC 동향은 전송 강건성, 지연, 계산 복잡도 사이의 균형을 중심으로 빠르게 진화하고 있음을 시사한다[4].

한편, 영상 신호는 에너지의 상당 부분이 저주파 대역에 집중되고, 고주파는 세부 시각 품질을 좌우하나 채널 잡음에 취약하다는 이질성을 가진다. 전통 코딩 방식에서는 이러한 특성을 이용해 대역·계층별 자원을 달리 주는 UEP(Unequal Error Protection)를 설계해왔다.

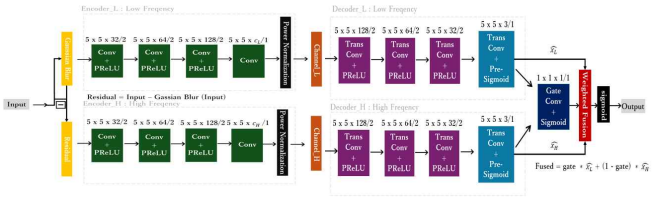
본 논문은 이러한 흐름에 착안하여, 영상 신호를 저주파와 고주파 대역으로 분리하고 이를 두 개의 독립된 신경망 경로로 전송하는 Dual-Band JSCC를 제안한다. 핵심은 고정 5 × 5 Gaussian Blur - Residual로 대역을 분해하고, 각 경로를 Encoder - AWGN - Decoder로 독립 처리하되 분기별 전력 정규화를 통해 전력을 맞추고 동일 SNR로 전송하며, 수신단에서 1×1 게이트로 두 복원본을 픽셀·공간별 가중 융합하도록 설계한 것이다. 본 연구의 핵심 기여는, Deep JSCC, Latent Separation, Layer Separation과의 엄격하고 공정한 비교를 통해 제안 방식의 실질적 우수성을 입증한 데에 있다. 이를 통해 채널 변동에 대한 높은 강건성과 구조적 효율성을 명확히 보일 것으로 시사한다.

본 논문은 다음과 같이 구성된다. 2절에선 동일 예산, 전력, SNR 프로토콜을 포함한 실험 설계 및 PSNR·SSIM 기반 정량 및 정성적 결과, 3절에선 결론 및 향후 과제를 제시한다.

Ⅱ. 본론

본 절에서는 제안하는 Dual-Band JSCC의 설계, 학습 설정, 재현성 확보 방식을 순서대로 기술한다. 전체 파이프라인은 [그림 1]과 같이 입력 영상의 정규화, 저주파 · 고주파 대역 분해, 분기별 Encoder-Channel-Decoder, 두 분기 출력의 gate 융합, 최종 Sigmoid, 역 정규화로 이어진다. 저주파는 고정 5 x 5 Gaussian Blur를 통해 추출하고, 고주파는 원본에서 저주파를 뺀 Residual로 정의하여 역할을 명확히 분리한다. 두 분기는 총 4개의 Convolution 블록을 통해 3번의 Down-sampling을 거쳐 Latent 표현을 만든다. 이후 각 Latent는 샘플별 평균 전력이 0.5가 되도록 정규화되어, 총 전송 전력 1.0을 두 경로가 공평하게 나누어 사용한다. 채널은 AWGN 하에서 두 분기 모두 동일 SNR로 통과한다. Decoder에선 분기별로 총 4개의 Transpose Convolution 블록을 통해 Up-sampling 하며, 마지막 Decoder의 Transpose Convolution 블록에서의 pre-Sigmoid logit

* : 교신저자

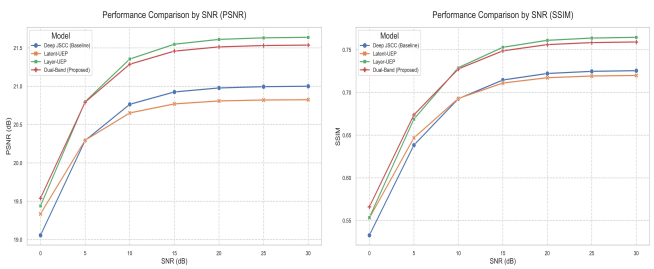


[그림 1] Dual Band JSCC 전체 파이프라인

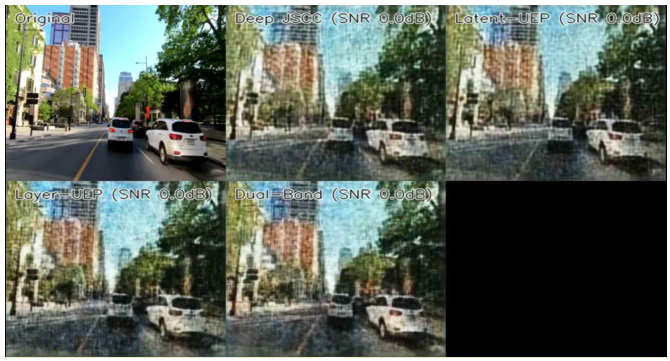
을 1 x 1 Convolution-Sigmoid로 산출한 gate로 픽셀·공간별 가중합을 진행하고, 단일 Sigmoid를 통해 최종 복원 영상을 얻는다.

제안 방식의 우수성을 검증하기 위해, 표준 Deep JSCC, 기존 UEP 기법인 Latent Separation과 Layer Separation, 그리고 주파수 대역을 독립된 경로로 분리하여 차등적으로 보호하는 새로운 UEP 방식인 제안 모델을 포함하여 총 4개의 모델을 구성했다. 공정한 비교를 위해 모든 모델은 동일한 학습 파라미터를 공유한다. 총 전송 심볼 비율은 0.2, 샘플별 평균 전력은 1.0, 학습 epoch는 30, 배치 크기는 64, 학습률은 $2 \cdot 10^{-4}$의 동일한 하이퍼파라미터를 공유하며, 손실 함수는 MSE(Mean Square Error)를 공통으로 사용하여 공정성을 확보했다. 추가적으로 연산 효율을 위해 AMP(Automatic Mixed Precision)를 활성화하여 학습을 진행했다. 모든 학습과 평가는 Cityscapes 데이터셋을 256 x 256의 고정된 해상도로 리사이즈하여 수행되었다. 이는 실험의 효율성과 재현성을 확보하고 제안 방식의 구조적 우월성이 해상도와 무관하게 나타남을 보이기 위함이다. 학습 중에는 0-30dB 사이의 SNR을 매배치마다 무작위로 샘플링하여 적용하였으며, 최종 성능 평가는 0, 5, 10, 15, 20, 25, 30dB의 고정된 SNR 환경에서 PSNR·SSIM을 측정하여 진행한다.

[그림 2]의 정량적 성능 그래프를 보면, 제안 모델이 낮은 SNR 구간에서 가장 강건함을 보인다. 이는 [그림 3]의 시각 비교에서도 확인되는데, Layer Separation 방식은 두 분기의 pre-Sigmoid를 단순 합산을 진행하기 때문에 고주파 분기의 잡음이 저주파 복원에 그대로 섞이는 오류 결합이 발생한다. 특히 0 dB 와 같은 잡음이 큰 채널 상황에선 고주파 logit이 큰 변동을 만들어 구조 윤곽이 무너지는 경향을 보인다. 반면 제안 모델은 두 분기를 독립 decoder로 복원한 뒤 1 x 1 gate-sigmoid로 픽셀·공간별 가중치를 학습하여, 낮은 SNR에서는 gate가 자연스럽게 저주파 쪽으로 무게를 두도록 유도된다. 그 결과 구조 정보 손실을 억제하고 PSNR·SSIM이 우세해진다. 한편 높은 SNR에서는 Layer Separation 방식이 미세 질감 복원에서 근소한 이점을 보이기도 한다. 잡음이 적을 때는 가중치 공유 decoder가 두 대역을 통합적으로 재구성하며, UEP 이득이 남아있어 세부 텍스처가 풍부한 반면, 제안 모델은 동일 SNR 설계이기 때문에 고주파 분기의 최대 이득이 상대적으로 제한될 수 있다. 이처럼 낮은 SNR 강건성은 제안 방식의 장점, 높은 SNR 미세 이득은 Layer Separation 방식의 장점으로 정리되며, 이는 코드 구현의 gate 가중합과 단순 합산, 동일 SNR과 차등 SNR의 구조적 차이에서 기인한다.



[그림 2] SNR에 따른 모델별 성능 비교 (좌) PSNR (우) SSIM



[그림 3] SNR에 따른 모델별 시각적 품질 비교

Ⅲ. 결론

본 논문은 자율주행의 안전성 향상을 위해, 영상의 주파수 대역을 분리하여 독립적으로 전송하는 Dual-Band JSCC 기법을 제안했다. 실험 결과, 제안 모델은 채널 오류가 심각한 낮은 SNR 환경에서 오류 전파를 효과적으로 차단하여, 다른 UEP 기법들보다 안정적으로 이미지의 핵심 정보를 보존했다. 높은 SNR 환경에서는 Layer Separation 방식이 미세 질감에서 근소한 우위를 보였으나, 최악의 상황에서의 신뢰도가 시스템 전체의 안전을 좌우하는 자율주행 분야에서는 제안 모델의 강건함이 더 큰 실용적 가치를 지닌다. 다만, 본 연구는 초기 파일럿 규모로 수행되어, 학습 예산 및 하이퍼파라미터 탐색이 제한적이었고, 일부 SNR 구간에선 뚜렷한 경향을 일반화하기 어렵다는 한계가 있다.

향후에는 더 긴 학습과 다중 seed 반복을 수행하고, 핵심 하이퍼파라미터의 체계적 sweep으로 일반화 가능한 경향을 검증할 예정이다. 또한, 주파수 대역 분해기인 Gaussian Blur를 학습형 모듈로 전환하여 SNR에 따라 가중이 적응적으로 변화하도록 하여, 낮은 SNR의 강건성과 높은 SNR의 세부 복원력을 동시에 높일 계획이다. 현재는 AWGN 모의 채널 평가를 실측 채널로 검증 범위를 확대하고, 지연·연산량 제약을 고려한 실시간 성능 평가도 병행할 예정이다. 향후 연구 방향으로는 보행자나 교통 표지판 등 의미론적 중요 영역에 전송 자원을 우선 배분하는 시맨틱 통신 기반 모델로 연구를 확장할 예정이다.

ACKNOWLEDGMENT

이 논문은 2025년도 정부(산업통상자원부)의 재원으로 한국산업기술진흥원의 지원을 받아 수행된 연구임 (RS-2024-00415938, 2024년 산업혁신인재성장지원사업)

참 고 문 헌

[1] Bourtsoulatze, E., Kurka, D. B., Gündüz, D., “Deep Joint Source-Channel Coding for Wireless Image Transmission,” IEEE Transactions on Cognitive Communications and Networking, vol. 5, no. 3, pp. 567 - 579, Sept. 2019.

[2] Kurka, D. B., Gündüz, D., “DeepJSCC-f: Deep Joint Source-Channel Coding of Images with Feedback,” IEEE Journal on Selected Areas in Information Theory, vol. 1, no. 1, pp. 178 - 193, May 2020.

[3] Yuan, H., Xiong, Y., Zhang, X., “Channel-Blind Joint Source-Channel Coding for Wireless Image Transmission,” Sensors, vol. 24, no. 12, 4005, 2024.

[4] Song, M., Kim, S., Jeong, S., “Deep Joint Source-Channel Coding for Wireless Image Transmission: A Review,” Electronics, vol. 12, no. 22, 4637, 2023.

Understanding the Role of CutMix in ETF During Neural Collapse

Nazmus Saqib, Joon Min-Gil*
Dept. of Computer Engineering, Jeju National University, South Korea.

Abstract

CutMix enhances deep neural network performance by generating convex combinations of training samples and labels. Despite its success, its effect on the data manifold during training remains underexplored. We investigate this through Neural Collapse, where last-layer features converge to a simplex equiangular tight frame (ETF). Using a controlled setup, we compare CutMix with MixUp in promoting ETF formation. Our results show that CutMix forms ETF structures more gradually than MixUp. However, it offers improved generalization and localization performance. We analyze geometric configurations to uncover mechanisms behind this behavior. This study sheds light on how CutMix influences representation learning dynamics.

Ⅰ. Introduction

Consider a classification problem characterized by an input and an output space, which are expressed as $\chi \epsilon \mathbb{R}^D$ and $\gamma := \{0,1\}^C$, respectively. Given a training set of samples, $\{(x_i, y_i)\}_{i=1}^N$, with $x_i \epsilon \chi$ denoting the i^{th} input data point and $y_i \epsilon \gamma$ representing the corresponding label, the goal is to train a model $f_\theta : \chi \rightarrow \gamma$ by finding parameters θ that minimizes the cross-entropy loss $CE(f_\theta(x_i), y_i)$ incurred by the model prediction $f_\theta(x_i)$ relative to the true target y_i, averaged over the training set, $\frac{1}{N}\sum_{i=1}^N CE(f_\theta(x_i), y_i)$.

Papyan et al. [1] observed that during loss function optimization, neural networks exhibit a phenomenon known as *Neural Collapse*, where the last-layer activations and classifier weights converge to the geometric structure of a simplex equiangular tight frame (ETF). This configuration reflects the network's inherent tendency to arrange class representations such that they are aligned with their corresponding classifiers, possess equal norms, and are equally spaced in angle—thereby achieving optimal class separation in the feature space. Understanding Neural Collapse is challenging due to the complex architecture and intrinsic non-linearity of neural networks.

A. CutMix

CutMix [2] is a popular data augmentation strategy, which generates new training examples through convex combinations of existing data points and corresponding labels.

$$x_{ii'}^{\lambda} = M \odot x_i + (1 - M) \odot x_{i'}, y_{ii'}^{\lambda} = \lambda y_i + (1 - \lambda) y_{i'} \quad (1)$$

where $M \epsilon \{0,1\}^{W \times H}$ denotes a binary mask indicating the dropout and fill which happened by exchanging information between two images. 1 is a binary mask filled with ones, and $\odot$ is the elementwise multiplication. Like Mixup [3], another popular data augmentation, CutMix also uses λ which is a symmetric Beta $\beta(\alpha, \alpha)$ distribution, where α is set to 1. The loss associated with CutMix can be mathematically represented as:

$$E_{\lambda \sim D_\lambda} \frac{1}{N^2} \sum_{i=1}^N \sum_{i'=1}^N CE(f_\theta(x_{ii'}^{\lambda}), y_{ii'}^{\lambda}) \quad (2)$$

B. Problem Statement and Contribution

Despite the widespread use and demonstrated efficacy of the CutMix data augmentation in enhancing generalization of deep neural networks, the underlying mechanism demands Neural Collapse for the following primary question:

"Does CutMix induce either similar or different geometrical configurations from Neural Collapse? Whatever the pattern it induces, how rapid or gradual the process to form the configurations is compared to other data augmentation strategies, and what is the theoretical reason behind this?"

This study aims to uncover the potential geometric configurations in the last layer activations resulting from CutMix.

Ⅱ. Result Summary

The results of our primary empirical investigation are presented in Fig. 1. This figure presents a comparative study between CutMix and Mixup to illustrate the delayed formulation of the ETF in CutMix. Our study incorporates the WideResNet-40-10 architecture on the dataset CIFAR10 using Adam as the optimizer.

Ⅲ. Experiments

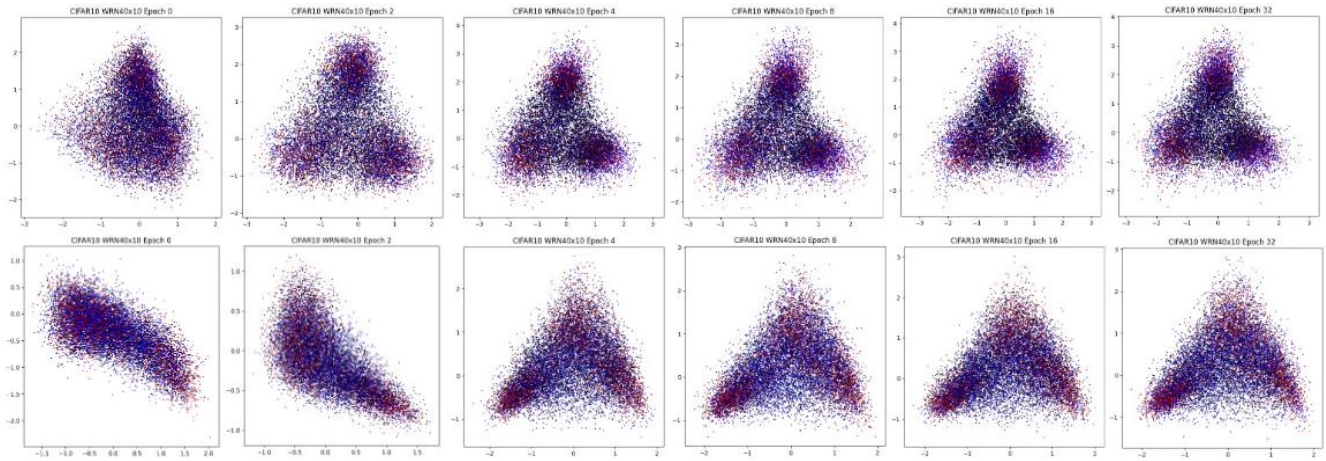


Fig 1. Evolution of three class-conditional feature activations under **Mixup** (top) and **CutMix** (bottom) training. Columns show epochs 0, 2, 4, 8, 16, and 32 (left → right). Mixup rapidly organizes the penultimate-layer features into a simplex ETF, whereas CutMix exhibits a slower, more gradual emergence of the ETF geometry.

For the above dataset and network pair, we visualize the last-layer activations for a subset of the training dataset consisting of three randomly selected classes. After obtaining the last-layer activations, they undergo a two-step projection: first onto the classifier for the subset of three classes, then onto a two-dimensional representation of a three-dimensional simplex ETF.

IV. Theoretical Justification:

CutMix tends to reduce regional dropouts, which aims to improve utilization for better generalization and localization performance. To do so, CutMix pastes inactive patches, so large portions of the image remain purely from one class while others come from another class. This creates piecewise-linear effects in feature space, where subregions belong to different classes. Thus, CutMix enjoys the property that there is no uninformative pixel during training to make it more efficient by retaining the advantages of regional dropout to attend to non-discriminative parts of objects. Our investigation reveals that while maintaining better generalization in training, CutMix fails to provide an early illustration of ETF. The theoretical reason behind this is CutMix's creation of piecewise-linear effects in feature space, while subregions belong to different classes. As a result, early in training, class features are more scattered and less aligned to an ETF structure. On the other hand, MixUp blends every pixel between two images, producing a linear combination of entire feature maps. This introduces strong linearity in feature space from the start, enforcing class-mean vectors to move quickly toward the symmetric, equidistant arrangement of ETF. We assume that, at both ends, the label mixing ratio λ directly influences the effect. MixUp's earlier formation of ETF directly influenced by λ which directly matches the pixel-mixing ratio, ensuring a consistent and immediate label-feature relationship. On the other hand, CutMix's λ depends on the patch area, and feature activations are influenced by contextual cues (background, object parts), which slows down the uniform convergence of class features toward the ETF vertices. The reduction of intra-class variance for CutMix is more gradual than MixUp, as CutMix preserves spatial structures, maintains higher-class variance longer, which slows down the ETF process. MixUp faces fast variance reduction by smoothing decision boundaries fast.

A. From data manifold perspective

As a large portion of local features are unchanged in the original image rather than the changed patches, CutMix creates piecewise manifold transitions instead of piecewise manifold transitions where local subregions belong to different class manifolds.

Ⅲ. Conclusion

We provide an investigative study about how gradual CutMix reforms the ETF in Neural Collapse. Though CutMix reforms it gradually, CutMix presents better generalization and localization performance compared to other data augmentation strategies, which present an earlier formation of ETF. In future works, we will carry out additional experiments for different datasets and networks, and theoretically characterize the optimal last-layer features.

ACKNOWLEDGMENT

This research was supported by the Regional Innovation System & Education(RISE) program through the Jeju RISE center, funded by the Ministry of Education(MOE) and the Jeju Special Self-Governing Province, Republic of Korea(2025-RISE-17-001).

References

[1] Papyan, Vardan, X. Y. Han, and David L. Donoho. "Prevalence of neural collapse during the terminal phase of deep learning training." *Proc. of the National Academy of Sciences*, vol. 117, no. 40, pp. 24652-24663, 2020.

[2] Zhang, Hongyi, et al. "mixup: Beyond empirical risk minimization." *arXiv preprint arXiv:1710.09412* (2017).

[3] Yun, Sangdoo, et al. "Cutmix: Regularization strategy to train strong classifiers with localizable features." *Proc. of the IEEE/CVF International Conference on Computer Vision*, pp. 6022-6031, 2019.

I/Q-FFT 이중 표현을 활용한 ParallelCNN 기반 광대역 스펙트럼 센싱 모델 성능 분석

권순영, 김형남*

*부산대학교

ysk1680@pusan.ac.kr, *hnkim@pusan.ac.kr

Performance Analysis of ParallelCNN-Based Wideband Spectrum Sensing Model Using I/Q-FFT Dual Representation

Soon-Young Kwon, Hyoung-Nam Kim*

*Pusan National University

요 약

본 논문은 광대역 스펙트럼 센싱의 저지연·고정확도 요구를 충족하기 위해, I/Q 데이터와 FFT 주파수 데이터를 이중 표현(dual-representation)으로 동시에 활용하는 ParallelCNN 기반 모델을 제안한다. 제안한 모델에서, I/Q 데이터 특징과 FFT 데이터 특징을 결합(concatenate)하여 두 특징의 상호보완성을 극대화하였다. 제안한 모델을 기존 모델(DeepSense, ParallelCNN)과 비교한 결과, 모든 SNR 구간에서 F1-score가 개선되었고, 특히 낮은 SNR에서 격차가 확대되는 경향을 보였다. 또한, 낮은 복잡도를 유지하면서도 낮은 SNR 구간에서 견고한 탐지 성능을 보여, 실시간 스펙트럼 센싱 적용에 적합함을 확인하였다.

Ⅰ. 서 론

라디오 주파수(radio frequency, RF) 스펙트럼은 무선통신의 핵심 자원이며, 5G·사물인터넷(internet of things, IoT) 등 신규 서비스의 확산과 더불어 희소성이 빠르게 심화되고 있다. 이러한 환경에서 인지 무선(cognitive radio, CR)은 주파수·시간 영역의 비활성 구간을 감지해 2차 사용자(secondary user, SU)가 기회적으로 접속하는 메커니즘을 제공하며, 그 출발점이자 핵심 구성요소가 스펙트럼 센싱이다. 스펙트럼 센싱은 1차 사용자(primary user, PU)의 도움 없이 채널의 점유 상태를 신뢰성 있게 판별해야 하고, 특히 광대역 환경에서는 높은 정확도와 짧은 지연(latency)을 동시에 만족해야 한다[1].

광대역 스펙트럼 센싱에는 전통적으로 두 계열의 접근이 존재한다. 나이퀴스트 기반 방식은 광대역을 여러 서브밴드로 분할해 순차적으로 혹은 동시에 탐지하지만, 지연이 증가하거나 구현 복잡도가 증가한다는 단점이 있다. 서브나이퀴스트/압축 센싱 기반 방식은 관측 축소로 하드웨어 부담을 줄이는 대신, 비선형 재구성 단계가 추가되어 지연·복잡도·성능 열화가 발생할 수 있다. 이러한 제약은 실환경 적용에서 실시간성·경량성·정확도 간 균형을 어렵게 만든다[2].

최근 딥러닝 기반 스펙트럼 센싱이 전통 알고리즘 대비 견고한 성능을 보여 주목받고 있으나, 많은 연구가 협대역 문제 설정에 머물거나 모델·전처리 복잡도가 커 실시간 처리에 불리하다는 한계가 있었다. 이를 보완하기 위해 제안된 DeepSense 모델은 원시 I/Q 데이터를 입력으로 받아 다중 서브밴드를 한 번에 판별함으로써 지연을 낮추었으나, 낮은 SNR 구간에서 정확도 저하와 샘플 길이 증가에 따른 복잡도 상승 문제가 존재한다[3]. 이에 비해 ParallelCNN 모델은 입력 I/Q를 절반으로 분할해 두 개의 CNN 분기로 병렬 처리하고, 분기 출력을 더해 정확도·지연·모델 크기의 균형을 개선한 바 있다[4].

본 논문은 ParallelCNN 모델의 경량·저지연 장점을 유지하면서, 광대역·저 SNR 환경에서의 탐지 성능을 추가로 향상시키는 것을 목표로 하였다. 구체적으로, 시간영역 I/Q 데이터와 주파수영역 FFT(fast Fourier transform) 데이터를 두 분기로 병렬 처리하되, 분기 간 정보를 더하는(add) 것이 아니라 결합(concatenate)함으로써, 각 표현이 담고 있는 상이한 특성이 소실되지 않고 최종 분류기에 전달되도록 설계하였다.

Ⅱ. 본 론

그림 1은 제안하는 기법의 모델 구조를 나타낸 그림이다. 모델의 경량화를 위해, 길이가 N 인 I/Q 데이터 $x \in \mathrm{R}^{2\times N}$를 시간 축으로 절반씩 분할하여($x_1 \in \mathrm{R}^{2\times N/2}$, $x_2 \in \mathrm{R}^{2\times N/2}$) 시간 축 분기의 두 입력으로 사용한다. 그리고 I/Q 데이터에 FFT(fast Fourier transform)를 수행하여 주파수 축 분기의 입력으로 사용한다($X \in \mathrm{R}^{2\times N}$). 각 입력은 각 CNN(convolutional neural network) 계층을 통과하며, $\mathrm{C}_i (i=1,...,8)$는 1D CNN을 의미한다. $\mathrm{M}_i (i=1,...,6)$는 1D 맥스 풀링(max pooling) 계층을 의미하며, $\mathrm{F}_1, \mathrm{F}_2$는 완전 연결(fully connected) 계층을 의미한다.

기존의 ParallelCNN 모델은 두 분기의 출력을 더해

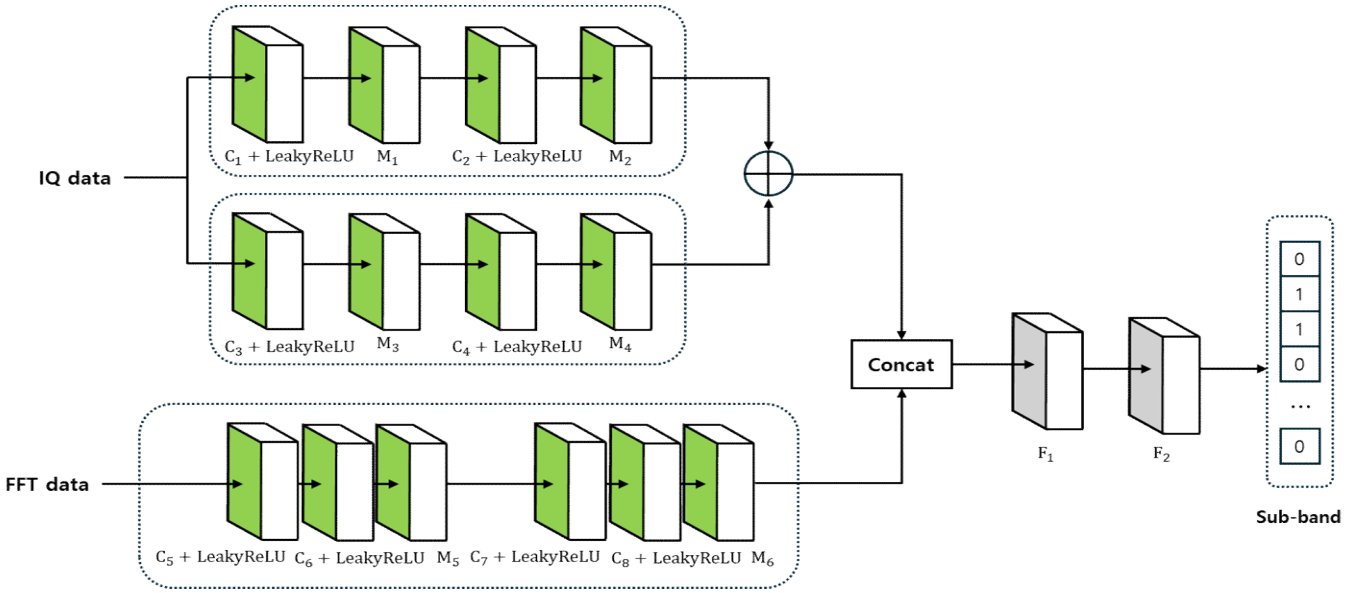


그림 2. 제안하는 신경망 모델.

(add) 하나의 표현으로 압축한다. 반면, 제안하는 모델은 시간 축 데이터의 특징과 주파수 축 데이터의 특징을 결합(concatenate)하여 완전 연결 계층이 각 특징의 상호작용을 학습하도록 유도한다.

그림 2는 기존의 신경망 모델(DeepSense, ParallelCNN)과 제안하는 신경망 모델의 F1-score 성능 비교 결과이고, 표 1에는 모델별 복잡도를 나타내었다. ParallelCNN 모델은 DeepSense 모델을 경량화하였기 때문에, F1-score 성능이 약 1~2% 열화된 것을 확인할 수 있다. 반면, 제안한 모델은 ParallelCNN 모델을 개선하여 I/Q 데이터와 FFT 데이터를 모두 활용했기 때문에, ParallelCNN 모델과 복잡도는 유사하지만 모든 SNR 구간에서 F1-score 성능이 약 2~5% 우수한 것을 확인할 수 있다.

표 1. 신경망 모델별 학습 파라미터 개수.

	Proposed	DeepSense	ParallelCNN
# parameters	20480	39472	20848

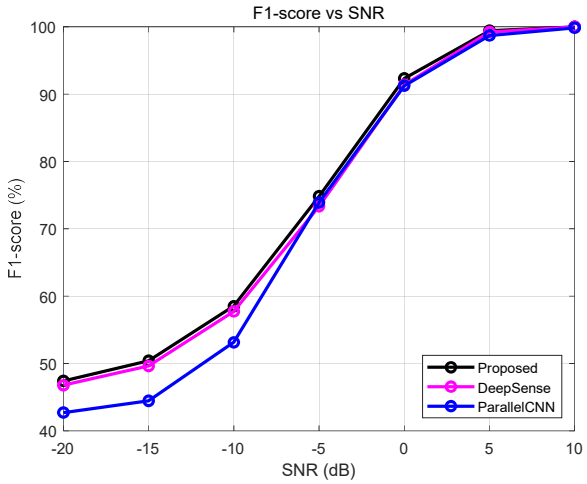


그림 1. SNR에 따른 신경망 모델별 F1-score 성능.

Ⅲ. 결 론

본 논문은 ParallelCNN 기반 구조에 I/Q - FFT 이중 표현을 도입하고 concat 결합을 적용하여, 낮은 SNR에서의 탐지 성능을 향상시키면서도 비교적 낮은 모델 복잡도를 유지하는 방법을 제시하였다. 모의실험 결과, 제안 모델은 DeepSense 및 ParallelCNN 대비 전 SNR 구간에서 F1-score 성능이 향상되었으며, 파라미터 수는 DeepSense보다는 작고, ParallelCNN과 유사한 수준으로 나타나 실시간 구현을 위한 합리적 모델임을 확인하였다.

ACKNOWLEDGMENT

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(RS-2025-00557790)

참 고 문 헌

[1] W. Zhang, Y. Wang, X. Chen, Z. Cai and Z. Tian, "Spectrum Transformer: An Attention-Based Wideband Spectrum Detector," *in IEEE Transactions on Wireless Communications*, vol. 23, no. 9, pp. 12343-12353, Sept. 2024.

[2] H. Sun, A. Nallanathan, C. -X. Wang and Y. Chen, "Wideband spectrum sensing for cognitive radio networks: a survey," *in IEEE Wireless Communications*, vol. 20, no. 2, pp. 74-81, April 2013.

[3] D. Uvaydov, S. D'Oro, F. Restuccia and T. Melodia, "DeepSense: Fast Wideband Spectrum Sensing Through Real-Time In-the-Loop Deep Learning," *IEEE INFOCOM 2021 - IEEE Conference on Computer Communications*, Vancouver, BC, Canada, 2021, pp. 1-10.

[4] R. Mei and Z. Wang, "Deep Learning-Based Wideband Spectrum Sensing: A Low Computational Complexity Approach," *in IEEE Communications Letters*, vol. 27, no. 10, pp. 2633-2637, Oct. 2023.

사이버범죄 수사지원 특화 대규모언어모델 개발을 위한 벤치마크 구축 및 보상체계 설계

계효선, 이미라, 김지형*
한국전자기술연구원
hskye@keti.re.kr, mlzr@keti.re.kr, *jkim8@keti.re.kr

Designing a Benchmark and Reward Framework for Developing Specialized Large Language Models for Cybercrime Investigation

Hyoseon Kye, Mira Lee, Jeehyeong Kim*
Korea Electronics Technology Institute

요 약

본 논문은 사이버범죄 수사 추론에 특화된 대규모언어모델(Large Language Model, LLM) 학습을 위한 보상 시뮬레이션 프레임워크를 제안한다. LLM 강화 학습 기법인 Group Relative Policy Optimization(GRPO)과 같은 방법을 활용하기 위해서는, 특수 목적 모델 튜닝에 적합한 보상함수 설계가 필수적이다. 이를 위해 본 연구에서는 보상함수의 효과를 실험적으로 검증할 수 있는 시뮬레이션 프레임워크를 설계·구현하고, 사이버범죄 수사 추론 관련 벤치마크를 구축하였다. 본 프레임워크를 통해 다양한 LLM이 해당 벤치마크와 보상체계 하에서 어떠한 성능과 경향성을 보이는지를 분석하였다.

Ⅰ. 서 론

최근 주식 리딩방, 인터넷 물품 사기, 몸캠피싱 등으로 대표되는 사이버범죄는 빠르게 고도화·조직화되고 있으며, 이에 따라 수사기관의 대응 난이도도 지속적으로 상승하고 있다. 이러한 환경에서는 범용 언어모델만으로는 수사 현장에서 요구되는 맥락 이해와 추론을 충분히 수행하기 어려우므로, 사이버범죄 수사 추론에 특화된 대규모 언어모델의 개발 필요성이 커지고 있다 [1]. 최신 API 기반 초거대 모델의 활용은 개발 난도를 낮추고 경우에 따라 파인튜닝을 생략할 수 있다는 장점이 있으나, 실제 수사 환경에서는 보안상의 이유로 외부 API 사용이 제한되는 경우가 많고, 제한된 GPU 자원에서 구축형 모델을 운용해야 하는 제약이 상존한다. 특히 이 분야는 선행 연구가 많지 않아, 제한된 자원으로도 실무에 필요한 수준의 성능을 달성할 수 있는지 불확실성이 크다. 이러한 문제를 해결하기 위해서는 기존 파운데이션 모델을 기반으로 도메인 지식을 반영한 강화 학습을 수행할 필요가 있으며, 그 핵심은 적절한 보상함수의 설계와 이를 실험적으로 검증할 수 있는 시뮬레이션 프레임워크의 마련에 있다. 본 연구는 사이버범죄 수사 추론 특화 Large Language Model(LLM) 학습을 위해 보상 시뮬레이션 프레임워크와 벤치마크를 함께 제안하고, 제한된 환경에서도 특화 모델 개발의 방향성과 실현 가능성을 체계적으로 검증할 수 있는 기반을 제공하고자 한다.

Ⅱ. 본론

도메인 특화 LLM을 구현하기 위해 단순 지도학습만을 적용하는 것은 한계가 분명하다. 사이버범죄 수사 추론은 복잡한 의사결정과 맥락 통합이 요구되는 영역으로, 정답 예시를 반복 학습하는 방식만으로는 실제 수사관의 추론 과정을 충실히 재현하기 어렵다. 이 한계를 보완하기 위해 인적 피드백을 이용한 강화 학습이 활발히 논의되어 왔으나, 전통적인 Proximal Policy Optimization(PPO) 기반 접근은 대규모 자원과 많은 피드백 데이터를 필요로 한다는 점에서 구축형 환경에는 부담이 크다. 이에 본 연구는 여러 응답 후보를 그룹 단위로 비교해 상대적 우위를 학습하는 방식의 Group Relative Policy Optimization(GRPO)[2]를 채택하였다. GRPO는 정답이 단일하지 않고 다양한 경로의 합리적 추론이 가능한 과제를 다루기에 적합하며, 절대 점수 대신 응답들 간의 상대적 선호를 학습함으로써 효율적인 미세조정이 가능하다는 장점이 있다. 또한, 보상함수 시뮬레이션의 효과적 구현을 위해 실제 수사 시나리오를 반영한 벤치마크를 구축하였다. 이 벤치마크는 총 여덟 개의 주요 문항으로 구성되며, 각 문항에는 다시 상황파악, 주요정보파악, 다음 수사 프로세스 추천의 세 가지 항목에 대해 각기 일반 응답과 전문가 응답을 제시하여 구성된다. 문항은 사기 수법의 식별과 패턴화, 계좌 명의자와 실제 사용자의 분리 판단, 피해자 진술과 기술 단서·연계 사건의 통합 정리, 가용 증거에 근거한 다음 절차의 우선순위 도출, 통화 및 거래내역 기반의 조직 구조 추론, IP 패턴과 현금 인출 및 현지 결제 내역을 활용한 위치 추정, 그리고 메신저·클라우드 접속 기록을 이용한 공범 식별 등으로 구성된다. 이러한 구성은 실제 수사관이 수행하는 상황 파악, 핵심 정보 정리, 그리고 실행 가능한 대응 제안이라는 세 가지 과업 유형을 폭넓게 포괄한다. 이러한 추론을 수행할 수 있도록 8개의 질의에 대해 별도의 컨텍스트를 포함한다. 컨텍스트는 10문장 수준의 텍스트로써, 본 추론에 활용할 수 있는 요약된 수사 정보를 의미한다. 추후 실제적인 사이버범죄 수사추론 시스템 개발을 위해서는 다양한 곳에서 데이터나 분석 결과를 조회하여 양질의 컨텍스트를 만드는 것이 성능을 향상시키는 주요 요소중 하나가 될 수 있으나, 현재 단계에서는 컨텍스트를 활용하여 모델이 얼마나 추론을 잘 할 수 있는지를 평가하는 것이 중요하므로, 이러한 컨텍스트는 충분히 양질의 컨텍스트가 사전에 생성되었다고 가정한다.

보상함수는 의미적 정합성과 내용 충실성, 그리고 언어적 완결성을 동시에 반영하도록 설계되었다. 먼저 모델 응답이 전문가 답변과 의미적으로

얼마나 근접한지를 평가하여, 단순 표현의 유사성을 넘어 추론의 방향성과 결론의 타당성을 측정한다. 다음으로 사건 분석에 필수적인 키워드의 반영 정도를 점검하여, 모델이 핵심 개념과 단서를 충분히 포괄하는지를 확인한다. 키워드는 데이터셋의 모범응답을 기반으로 주요단어를 뽑아 선정되어있다. 응답에서 해당 키워드를 얼마나 많이 포함하는지를 계산하여 점수화하는 방식이다. 마지막으로 응답의 길이와 구성, 한국어 사용의 적정성, 자연스러운 종결 등 문장 수준의 품질을 평가하여, 보고서 형식으로 바로 활용 가능한 응답을 유도한다. 사이버범죄 수사의 특수성을 고려해 최소 길이와 최적 범위를 상향 조정하고 한국어 비중 기준을 강화하는 등 도메인 맞춤형 기준을 적용하였다.

표1에서는 각 모델의 벤치마크 결과를 나타내고 있다. 전문가 응답을 정답의 기준으로 설정하고, gpt-4o, gpt-4o-mini, gpt-oss-20b[3]의 성능을 동일 벤치마크에서 비교하였다. gpt-4o와 gpt-4o-mini는 API 기반 LLM의 가장 대표적인 서비스로써 성능의 기준을 삼기 위해 선정하였다. gpt-oss-20b 는 OpenAI에서 최근에 출시한 오픈소스 Smaller Large Language Model (sLLM)이다. 특히, gpt-oss-20b는 4.5bit 양자화 튜닝으로 인해 24GB 이하의 게이밍 GPU에서도 구동이 가능한 것이 장점으로, 향후 사이버범죄 수사추론 특화 모델을 개발할 시에 적극적으로 고려될 수 있는 수준의 모델이다.

모델들의 성능을 평가하기 위해서는 벤치마크를 기반으로 보상 시뮬레이션을 구동하였을 때의 결과가 정답 데이터의 보상 시뮬레이션 결과가 얼마나 차이가 나는지를 비교해야 한다. 각 LLM 모델이 추론에 사용한 프롬프트는 상황파악, 주요정보, 다음 절차 추천으로 구분하여 구성하였으며, 핵심정보를 요약하고 이해하고 정리하여 연관성을 파악하라는 내용이 포함되어 있다. 또한 벤치마크에 포함된 컨텍스트를 프롬프트에 포함하여 LLM 모델이 활용할 수 있도록 하였다. 종합 성능에서 gpt-4o가 가장 정답에 근접했으나, 유사 수준에는 못 미치는 결과가 나왔다. 정답의 평균점수가 0.969점 인데 반해 gpt-4o의 평균점수는 0.807이었다. gpt-4o-mini는 전반적으로 gpt-4o와 큰 차이가 나지 않았다. 대부분의 모델 성능 벤치마크에서 gpt-4o이 gpt-4o-mini 보다 성능이 좋을 수 밖에 없는데도 해당 벤치마크에서 큰 차이가 나지 않는 것은 본 벤치마크가 특별히 어렵고 복잡한 역할을 LLM에게 기대하기보다 도메인 특화된 요구사항을 기대하는 것으로 해석할 수 있다. 그럼에도 불구하고, gpt-oss-20b는 다른 두 모델에 비해 현저한 성능저하가 관찰 되었다. 평균점수는 0.697 수준이었으며, 오차의 표준편차도 0.089로 다른 모델에 비해 상대적으로 일관성도 떨어졌다. gpt-4o-mini 정도의 성능만 되어도 크게 무리가 없는 수준의 복잡도를 지닌 태스크에 대해서도, 모델 성능의 한계로 인해 gpt-oss:20b는 현저한 성능차이를 보이고 있다.

표2 는 각 모델들이 세 가지 채점 항목에서 각 어떤 점수를 보이고 있는지를 나타내고 있다. 키워드 탐색에서 세 모델 다 정답 대비 적절한 성능을 보이지 못하는 것을 알 수 있다. 이는 사이버범죄 수사 도메인 특유의 단어와 표현방법을 모델들이 잘 쓰지 않기 때문에 나오는 결과로 보인다. 정답과의 유사도 점수는 gpt-4o와 gpt-4o-mini가 비슷했고, 문장 구조에 대해서는 gpt-4o가 gpt-4o-mini보다 다소 나은 결과를 보였다. 문장을 만드는 기본 능력에 있어서 gpt-4o가 다소 나은 성능을 보였다고 해석할 수 있다. gpt-4o-oss-20b의 경우, 유사도나 문장 구조에 비해 오히려 키워드 항목에서는 다른 두 모델에 비해 상대적으로 적은 차이를 보였는데, 이는 학습되지 않은 도메인 특유의 표현을 하는 것에 있어서는 다른 대규모 모델들과 큰 차이가 없었음을 의미한다. gpt-oss-20b는 단순 사실 확인형 과제에서는 준수한 결과를 보였지만, 분산된 단서를 결합해 결론을 도출하는 고난도 종합 추론 과제에서는 성능 저하가 두드러졌다.

<표1. 전문가 응답 및 각 LLM 모델들의 벤치마크 평균점수>

지표 \ 모델	전문가 응답	gpt-4o	gpt-4o-mini	gpt-oss:20b
평균점수	0.969	0.807	0.800	0.697
오차	-	0.162	0.169	0.272
오차 표준편차	-	0.043	0.047	0.089

<표2. 전문가 응답 및 각 LLM 모델들의 벤치마크 항목 별 점수>

항목 \ 모델	전문가 응답	gpt-4o	gpt-4o-mini	gpt-oss:20b
유사도	1.00	0.862	0.852	0.750
키워드	1.00	0.633	0.653	0.613
문장 구조	0.90	0.899	0.868	0.702

Ⅲ. 결론

본 연구는 사이버범죄 수사 추론에 특화된 LLM 개발을 목표로, 보상 시뮬레이션 프레임워크와 실제 수사 시나리오를 반영한 벤치마크를 제안하고 그 타당성을 실험적으로 확인하였다. 분석 결과는 상용 모델이 문장 수준의 품질에서는 높은 완성도를 보이지만, 정답에 상응하는 논리적 추론과 핵심 정보의 포괄성에서 여전히 격차가 존재함을 보여준다. 이에 따라 향후 연구에서는 벤치마크의 문항 범위와 난이도를 확대하고 항목별 평가 기준을 정교화하여, 다양한 유형의 수사 추론을 안정적으로 평가할 수 있는 체계를 구축할 것이다. 동시에 제안한 시뮬레이션을 학습 루프의 중심에 두고 도메인 피드백을 체계적으로 반영함으로써, 제한된 자원 환경에서도 일관되게 동작하는 수사 특화 모델을 단계적으로 구현할 계획이다. 아울러 실무 적용성을 높이기 위해, 수사에 필요한 정보를 능동적으로 수집·정제하는 도구 활용과 절차적 플래닝 역량을 모델 수준과 시스템 수준에서 통합하는 별도의 연구를 병행할 필요가 있다. 이를 통해 모델의 추론 능력뿐 아니라 실제 업무 흐름에서의 효율성과 신뢰성을 함께 제고할 수 있을 것으로 기대한다.

ACKNOWLEDGMENT

이 논문은 2025년도 행정안전부의 재원으로 과학치안진흥센터 사이버범죄 수사단서 통합분석 및 추론시스템 개발 사업의 지원을 받아 수행된 연구임(No. RS-2025-02218280)

참 고 문 헌

[1] J. Kim, et al., “A Conversational AI Agent System Based on Large Language Models and Retrieval-Augmented Generation to Support Investigation,” Journal of Data Forensics Research, vol. 1, no. 1, pp. 47-62, 2024.

[2] Z. Shao, et al., “DeepSeekMath: Pushing the Limits of Mathematical Reasoning in Open Language Models,” arXiv preprint arXiv:2402.03300, 2024.

[3] OpenAI, et al., “gpt-oss-120b & gpt-oss-20b Model Card,” arXiv preprint arXiv:2508.10925, 2025.

사이버범죄 수사 추론 에이전트를 위한 MCP 기반 도구 연동 플랫폼 개발

김민수, 김성섭, 권영민, 이승우*
한국전자기술연구원
{mskim92, sskim, ymkwon, seungwoo.lee*}@keti.re.kr

MCP-based Tool Integration Platform for Reasoning Agents in Cybercrime Investigation

Minsu Kim, Seongseop Kim, Youngmin Kwon, Seungwoo Lee
Korea Electronics and Technology Institute (KETI)

요 약

사이버범죄 수법의 다양화 및 고도화로 인해 수사관의 범죄수사 과정을 보조하기 위한 인공지능 기반 수사 추론 에이전트의 수요가 증가하고 있다. 기존의 수사 지원 시스템은 단일 도구 중심의 분석에 한정되어 있기 때문에 사건 장소의 변동이 크고 인물 관계가 복잡한 사이버범죄 수사 과정에서 다양한 분석·추론 도구들을 통합적으로 활용하기 어려운 한계가 있다. 본 논문에서는 사이버범죄 수사 추론 에이전트를 위한 MCP (Model Context Protocol) 기반 도구 연동 플랫폼을 제안한다. 제안하는 플랫폼은 도구 API를 통합 및 관리하고, 수사 추론 에이전트가 현재 상황에서 필요한 도구를 선택하여 실행하도록 지원하며, 수사 추론 에이전트와 도구 간의 확장 가능한 연동 아키텍처를 제공한다.

Ⅰ. 서 론

산업 및 사회 전반의 디지털 전환 가속화로 인해 사이버범죄 발생 건수가 지속해서 증가하고 있다. 사이버범죄는 비대면 특성과 투자 사기, 물품 거래 사기 등 다양한 범죄 유형이 존재하기 때문에 수사 난이도가 높다 [1]. 이에 따라 수사관의 용의자 특정 및 추적을 보조하기 위한 인공지능 기반 수사 추론이 주목받고 있다 [2]. 수사 추론은 대형 언어모델 등 추론 능력을 부여할 수 있는 모델을 에이전트로 활용하며, 디지털 증거 및 단서를 분석하여 사실관계를 요약하고 범죄 전모를 재구성함으로써 수사 방향을 결정하는 기술을 의미한다.

사이버범죄는 사이버 공간에서 행해지는 범죄로, 웹사이트, SNS, 메신저 앱 등 다양한 온라인 플랫폼을 매개로 한다 [3]. 이와 같이 복잡한 디지털 환경에서 발생하는 특성 때문에 사이버범죄에 대한 수사 추론은 다수의 분석·추론 도구를 통합하여 활용할 필요가 있다. 그러나, 개별 분석·추론 도구는 서로 다른 API 구조와 데이터 형식을 사용하기 때문에 수사 추론 에이전트의 도구 통합에 어려움이 있다. 따라서 도구 통합 및 연계를 위한 표준 프로토콜의 도입이 필요하며, 최근 MCP (Model Context Protocol)가 해결 방안으로 제시되고 있다. 에이전트와 외부 도구가 상호작용할 수 있도록 지원하는 표준 프로토콜인 MCP는 JSON-RPC 메시징 시스템을 사용하여 일관된 인터페이스를 제공한다. 수사 추론 에이전트는 MCP를 통해 네트워크 분석, 인물 관계 분석 등 다양한 분석·추론 도구를 표준화된 방식으로 연계할 수 있다.

본 논문은 사이버범죄 수사 추론 에이전트를 위한 MCP 기반 도구 연동 플랫폼을 제안한다. 제안하는 플랫폼은 다수의 도구에 대한 동적 연결 및 기능 등록을 지원한다. 기능 등록이 완료된 도구의 API 정보는 플랫폼이 관리하며, 수사 추론 에이전트에게 호출 정보가 제공된다. 수사 추론 에이전트는 직접 질의 혹은 핵심 키워드를 플랫폼으로 전달하여 현재 상황에 적절한 도구를 선택 및 실행한다.

Ⅱ. 관련 연구

기존의 에이전트 도구 연동 기술은 정적인 연결 방식을 사용하기 때문에 에이전트가 학습된 시점 이후의 최신 정보를 반영하는 것과 외부 시스템과의 연결이 어려운 문제가 있다. 따라서 에이전트가 필요한 도구를 사용하도록 하고, 원활한 작업 수행을 지원할 수 있는 표준 연동 프로토콜의 필요성이 커지게 되었다.

Anthropic에서 제안한 공개 표준인 MCP는 JSON-RPC 메시징을 통해 인공지능 어플리케이션과 외부 도구 간의 상호작용을 표준화한다. MCP는 호스트, 서버, 그리고 클라이언트의 3계층 아키텍처를 채택하여 확장성과 유연성을 제공한다 [4]. 호스트는 사용자 작업 환경을 의미하며, 서버는 도구의 기능을 정의하고 관련 메타 정보를 저장한다. 클라이언트는 서버와 인공지능 모델 혹은 에이전트 사이에서 도구 호출과 정보 전달을 중계한다. MCP는 시스템 독립적이며, 다양한 기능과 인터페이스를 가진 도구를 표준화된 방식으로 통합하여 활용할 수 있는 장점이 있다.

본 논문에서 제안하는 MCP 기반 도구 연동 플랫폼은 기존 기술과 다음과 같은 차별점이 있다.

(1) 사이버범죄 수사 추론 에이전트에서 분석·추론 도구의 통합 및 활용을 위한 동적 연결 및 기능 등록을 지원
(2) 사이버범죄 수사 추론 에이전트가 복합적 사건에 대응할 수 있도록 표준화된 도구 호출 및 연동 메커니즘을 제공

Ⅲ. MCP 기반 도구 연동 플랫폼

제안하는 플랫폼의 구조는 그림 1과 같다. 이 플랫폼은 다중 MCP 서버 지원, MCP 서버 관리, 도구 연결 및 기능 등록, 그리고 MCP 프록시로 구성된다. 연동 도구는 런타임 중에 동적으로 연결 및 해제할 수 있다. MCP 프록시는 API 파라미터, 도구 기능 목록, MCP 서버 연결 주소 등 필수적인 연동 정보를 관리한다.

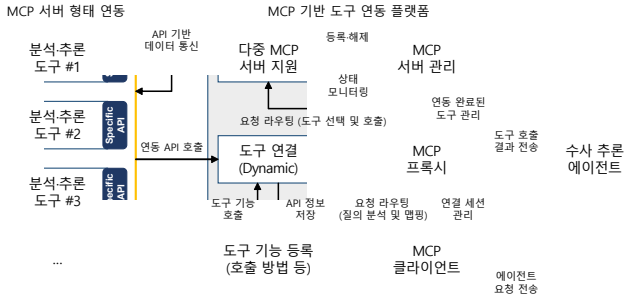


그림 1. MCP 기반 도구 연동 플랫폼 구조

개별 분석·추론 도구는 도구 연결 및 기능 등록 API를 사용하여 MCP 서버 형태로 플랫폼에 연동할 수 있다. 해당 API의 명칭 및 사용 방법은 표 1과 같다.

API 명칭	사용 방법 예시

도구 동적 연결	[POST]
	{base_url}/tool_dynamic_integration
	[필요 파라미터 목록]
	{
	"name":"human_network_analysis",
	"url":"http://123.456.789:8000",
	"timeout":30
	}

도구 기능 등록 (API 명세 정보)	[POST]
	{base_url}/tool_api_registration/{tool_name}/endpoints
	[필요 파라미터 목록]
	{
	"endpoints":[
	{
	"method":"GET",
	"path":"/users",
	"description":"Get information of all users"
	},
	…
	]
	}

도구 기능 호출	[POST]
	{base_url}/tool_api_registration/{tool_name}/call
	[필요 파라미터 목록]
	{
	"method":"GET",
	"path":"/users",
	"params":{}
	}

표 1. 도구 연결 및 기능 등록 API

MCP 프록시는 연결 및 기능 등록이 완료된 도구를 관리하며, 수사 추론 에이전트의 요청 정보를 수신한다. 에이전트로부터 라우팅되는 요청 정보는 직접 질의 혹은 핵심 키워드 형태로 정의된다. 예를 들어, "네트워크 분석"이라는 키워드가 전달되었을 경우 MCP 프록시는 등록 완료된 도구 목록을 조회하고, 키워드 매칭 혹은 기능 태그 분석을 통해 해당 기능을 보유한 도구를 탐색한다. 탐색의 결과로 도구 선택이 완료되면, 도구의 API 정보를 활용하여 필요한 기능을 호출한 다음 결과를 수신하여 에이전트로 반환한다. MCP 기반 도구 연동 플랫폼은 현재 연결 중인 도구의 상태를 주기적으로 모니터링한다. 또한, 도구 개발자를 위해 연동 규격 및 API 명세를 상세 사용 가이드 형태로 제공한다.

그림 2. MCP 기반 도구 연동 플랫폼 테스트(도구 연결 후 기능 등록)

제안하는 플랫폼을 검증하기 위해 플랫폼 세부 기능을 구현하고 테스트 환경을 구축하였다. 그림 2는 도구 연결 후 기능 동록에 대한 테스트 결과를 보여준다. 플랫폼 테스트에는 임의 구현된 네트워크 분석 도구를 활용하였다. MCP 기반 도구 연동 플랫폼은 동적 연결, 기능 등록 및 호출에 관한 절차를 API 형태로 관리함으로써 사용자가 적시적소에 수사 추론 에이전트에 필요한 도구를 연동할 수 있다. 도구 연결에 소요되는 시간은 1초 이내로 사용자 체감이 적은 수준이었으며, 이미 연결이 완료된 도구의 기능 호출은 플랫폼에서 큰 부하 없이 처리하는 것을 확인하였다.

Ⅳ. 결론 및 향후 연구

본 논문에서는 사이버범죄 수사 추론 에이전트를 위한 MCP 기반 도구 연동 플랫폼을 제안하였다. 제안된 플랫폼은 다양한 분석·추론 도구와의 표준화된 연동 인터페이스를 제공함으로써 수사 추론 에이전트가 상황에 맞는 도구를 선택 및 실행할 수 있도록 지원한다. 이를 통해 사이버범죄 사건에 대한 통합적인 분석과 수사 지원에 기여할 수 있다.

향후 연구로는 사이버범죄 수사 추론 시나리오를 구성하여 MCP 기반 도구 연동 플랫폼을 테스트하고 세부 기능을 개선하는 것이 있다. 또한, 다수의 도구에 대한 연결 및 기능 등록의 응답시간 및 처리량 측정 실험을 수행하여 플랫폼 운용 성능을 최적화할 계획이다.

ACKNOWLEDGMENT

이 논문은 2025년도 행정안전부의 재원으로 과학치안진흥센터 사이버범죄 수사단서 통합분석 및 추론시스템 개발 사업의 지원을 받아 수행된 연구임(No. RS-2025-02218280)

참 고 문 헌

[1] 이연주, 이서연, 김기범, 이성진, "사이버범죄 발생과 공간의 연관성 분석 - 공간회귀분석을 활용하여 서울특별시를 중심으로", 디지털 포렌식 연구, vol. 17. no. 4, pp. 128-145, Dec. 2023.

[2] H. Kim, D. Kim, J. Lee, C. Yoon, D. Choi, M. Gim, and J. Kang, "LAPIS: Language Model-Augmented Police Investigation System", Proc. of the ACM International Conference on Information and Knowledge Management, pp. 4637-4644, Oct. 2024.

[3] 이소현, 강일웅, 정윤혁, 김희웅, "사이버범죄 유형별 특징 분석 연구", Information Systems Review, vol. 21, no. 3, pp. 1-26, Jan. 2019.

[4] S. Liu, H. Miao, and P-T. Bremer, "ParaView-MCP: An Autonomous Visualizatiion Agent with Direct Tool Use", arXiv preprint arXiv:2505.07064, pp. 1-6, May 2025.

이종 네트워크 연계 기반 HGT 모델을 활용한 범죄 사건 내 피의자 후보군 판별 기술 연구

한현규, 백명선
세종대학교

hhg02258565@gmail.com, msbaek@sejong.ac.kr

A Study on Criminal Suspects Identification Technology using a Heterogeneous Network-based HGT Model

Hyungyu Han, Myung-Sun Baek
Sejong University

요 약

초동 수사 단계에서 신속하고 효과적으로 범죄에 대응하기 위해 수사관에게 잠재적 용의자 후보군 정보를 제공하는 것은 매우 중요하다. 기존 연구는 텍스트 분류나 개별 네트워크 분석에 초점을 맞추었으나 인물과 사건 간의 복합적인 관계를 통합적으로 모델링하는 데 한계가 있었다. 본 논문에서는 인물-사건 관계를 단일 이종 그래프(Heterogeneous Graph)로 모델링하고 Heterogeneous Graph Transformer(HGT)를 이용해 사건별 참여자 집합 내에서 피의자 여부를 분류하는 인공지능 기술을 설계한다. 제안 기술은 텍스트로 기술된 인물 및 사건 정보를 임베딩하고 이들 간의 유사도 및 소속 관계를 그래프 엣지로 구성한다. 본 기술은 초동 수사 시 제한된 자원을 효율적으로 배분하고 용의자 탐문 범위를 좁히는 데 주요한 참고 자료로 활용될 수 있다.

Ⅰ. 서 론

최근 스마트 치안 기술의 일환으로 인공지능 및 빅데이터를 활용하여 범죄에 대응하는 연구가 활발히 진행되고 있다 [1]. 효과적인 초동 수사를 위해서는 신규 사건 접수 시 해당 사건의 특성을 분석하고 관련성이 높은 인물을 신속하게 파악하는 것이 중요하다. 기존에는 텍스트 분류를 통해 범죄 유형을 예측하거나 준지도학습 기반의 네트워크를 통해 용의자를 추론하는 연구가 제안되었다[2]. 그러나 이러한 접근법들은 인물 간의 관계와 사건 간의 관계를 분리된 네트워크로 다루거나 단순 텍스트 유사도에 의존하여 다각적인 관계 정보를 통합하는 데 어려움이 있었다.

본 논문에서는 사건과 인물정보를 포함하는 누적된 가상의 사건 데이터 내 인물과 사건 정보를 하나의 이종 그래프에 통합 모델링하는 방법론을 제안한다. 본 논문에서는 사법시스템에서 가장 일반적으로 활용되는 가상의 형사사법정보시스템 (KICS: Korea Information System of Criminal Justice Services) 데이터를 활용하였다.

제안 방법은 다음과 같은 핵심 기술로 구성된다. 첫째, 최신 언어 모델을 사용해 텍스트 정보를 벡터로 임베딩한다. 둘째, 인물-인물(p-p) 및 사건-사건(c-c) 간에는 상호 k-최근접 이웃(mutual k-NN) 알고리즘으로 강건한 유사도 엣지를 생성하고 인물-사건(p-c) 간에는 원본 데이터의 소속 관계를 엣지로 연결한다. 셋째, 이종 그래프 구조 학습에 특화된 Heterogeneous Graph Transformer(HGT) [3] 모델을 적용하여 특정 사건과 관련된 인물이 '피의자' 관계로 연결될 확률을 예측한다.

상기 설계된 기술을 활용하면 신규 사건 발생 시 참여 인물들의 역할을 예측하여 수사 우선순위를 정하는 데 참고할 수 있다.

Ⅱ. 범죄유형 추론 기술 설계

입력 데이터는 KICS 양식의 사건 개요와 인물 관련 레코드로 구성되며 단일 엑셀 파일 내 KICS 입력, 피의자, 피해자, 참고인 시트를 사용한다. 인물 식별은 주민번호로 사건 식별은 최초접수번호로 정의한다. 각 시트는 스키마를 유지한 JSON 배열로 변환되어 후속 처리 단계의 일관성을 보장한다. 인물 1 인과 사건 1 건당 하나의 임베딩을 생성하며 임베딩 모델은 Qwen Embedding 4B 를 사용한다[4]. 임베딩 모델을 통해 생성된 2560 차원 벡터는 저장 후 모델 학습 시 노드 타입별 선형 투영층을 거쳐 128 차원으로 투영되며 임베딩 모델 파라미터는 동결하여 GNN 과 분류기만 학습한다. 이러한 전처리와 임베딩 정책은 대규모 배치 처리에서 메모리 사용을 예측 가능하게 만들고 운영 단계의 재현성을 높인다. 본 시스템의 파이프라인은 전처리 → 임베딩 →그래프 구성 → HGT 인코딩 → 엣지 분류기의 순서로 진행되며 그림 1 은 이 과정을 요약한다.

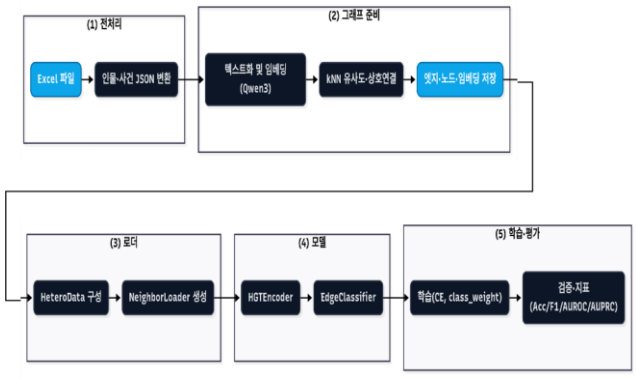


그림 1. 제안된 기술의 전체 절차를 포함하는 블록다이어그램

Ⅲ. 그래프 기반 피의자 후보군 예측 방법

그래프 $G = (V, E)$ 는 Person 과 Case 두 노드 타입으로 구성되며 인물–인물(p-p)과 사건–사건(c-c) 유사도 엣지, 인물–사건(p-c) 소속 엣지 세 관계를 포함한다. p-p 와 c-c 는 각 타입의 임베딩 코사인 유사도에 기반한 k-최근접 이웃을 계산한 뒤 서로의 이웃 목록에 동시에 포함될 때만 엣지를 부여하는 상호 k-NN 전략을 사용한다. 이때 $kpp = kcc = 15$, 임계값 $\tau = 0.4$ 를 적용하고 엣지 가중치는 $\frac{1+cos}{2}$ 로 정규화하여 스케일 안정성을 확보한다.

p-c 엣지는 인물의 사건 소속 정보를 그대로 연결하며 역할 속성은 [피의자, 피해자, 참고인] 원-핫으로 인코딩한다. 예측 대상 사건의 입력에서는 피의자·참고인 속성을 0 으로 마스킹하고 피해자 정보만 유지하여 모델이 타깃 사건에서 알려진 정답을 직접 보지 못하게 설계한다. 이러한 마스킹은 훈련·검증·추론의 전 단계에 동일하게 적용되며 피해자 존재는 사건 맥락 신호로 작용한다. HGT 는 타입과 관계에 따라 서로 다른 투영과 어텐션을 적용하는 이종 그래프용 트랜스포머로 본 연구에서는 두 개의 HGT 레이어(hidden=128, heads=4)를 사용하여 노드 임베딩을 갱신한다. 입력 단계에서 타입별 선형 투영(2560→128) 후 Swish 활성화와 RMSNorm 으로 정규화하여 안정적으로 메시지 패싱을 시작한다. 갱신된 인물 임베딩 h_p와 사건 임베딩 h_c는 Swish-GLU 두 블록과 Dropout 을 통과한 뒤 2-클래스 로짓을 출력한다. 이 구성은 임베딩의 분포 차이와 역할 정보의 누수 문제를 동시에 고려하여 설계되었으며, 사건·인물 규모가 수만 단위를 넘어서는 환경에서도 NeighborLoader 기반의 타입 인식 샘플링으로 미니배치 학습이 가능하다.

최종 확률은 소프트맥스로 계산하며 사건별로 임계 0.5 초과 인물을 후보로 제시한다. 그림 2 는 인물–사건 이층 그래프의 개념도를 나타낸다.

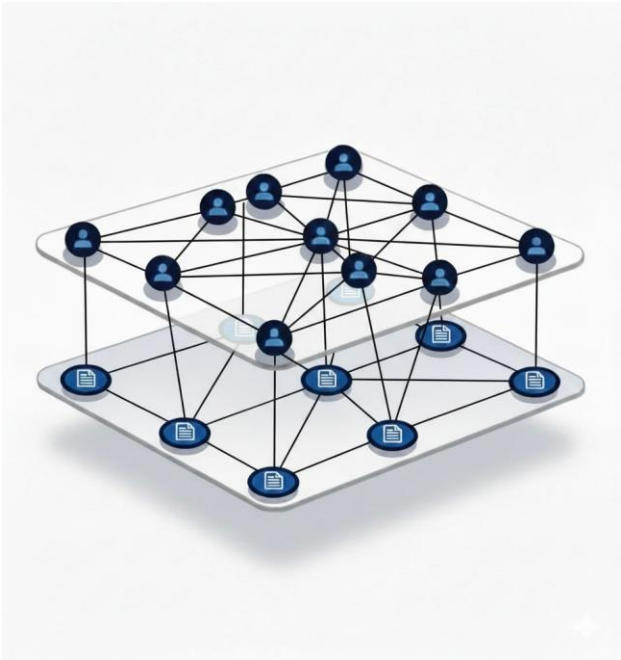


그림 2. 인물–사건 이종 그래프 개념도

Ⅳ. 개발 기술의 성능 검증 방법

제안된 모델의 학습 목적은 특정 사건 c 에 이미 참여가 확인된 인물 p 의 역할이 '피의자'인지 여부를 분류하는 것이다. 따라서 출력 확률에 대해 가중 크로스엔트로피 손실을 최소화하도록 학습된다. 클래스 가중치는 라벨 분포의 역비율로 설정하여 불균형을 완화하며 손실 집계는 타깃 사건의 p-c 엣지 중 피의자·참고인만 포함하고 피해자 엣지는 자동 제외한다. 데이터 분할의 경우, 사건 노드를 8:1:1 분할하며 전체 그래프를 사용하되 손실·평가는 해당 분할 사건 집합에만 한정한다. 최적화는 AdamW 와 Dropout 을 사용한다. 추론 단계에서는 신규 사건 임베딩을 생성하여 그래프에 연결한 후 사건 참여자 집합 내 피의자 여부 분류 및 확률과 근거 이웃(유사 사건 링크, 관련 인물)을 함께 제시해 수사관이 맥락을 확인할 수 있게 한다. 주요 지표는 F1-score, Accuracy, Precision, Recall 을 중심으로 AUROC, AUPRC 를 함께 기록한다. 본 논문은 피의자 후보군 판별 기술 구현 방법론에 대한 것으로 추후 수치를 포함한 기술 검증 연구가 추가적으로 진행될 예정이다.

Ⅴ. 결 론

본 논문은 인물–사건 이종 그래프와 HGT 를 결합해 사건별 피의자 후보 확률을 산출하는 방법을 제시하였다. 임베딩 유사도 기반 상호 k-NN 으로 유사 구조를 부여하고 역할 마스킹으로 정보 누수를 차단하며 HGT-기반 메시지 패싱과 GLU 분류기로 엣지 수준 확률을 출력하는 파이프라인을 정립했다. 향후에는 실제 데이터 분포에 대한 적응과 임계 재보정 절차를 포함한 운영형 학습 파이프라인을 확장하고 성능을 검증할 계획이다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 정보통신방송혁신인재양성 (메타버스융합대학원) 사업 연구 결과로 수행되었음(IITP-2024-RS-2023-00254529).

참 고 문 헌

[1] 장광호, “스마트치안”, 2020. 06.

[2] 백명선 외, “효과적인 초동 수사를 위한 지도/준지도 학습 기반 범죄 사건 정보 예측 및 용의자 후보군 특정,” 한국인공지능학술대회, 2020.

[3] Z. Hu, Y. Dong, K. Wang, and Y. Sun, “Heterogeneous Graph Transformer,” in *Proceedings of The Web Conference 2020 (WWW '20)*, 2020.

[4] Y. Zhang et al., “Qwen3 Embedding: Advancing Text Embedding and Reranking Through Foundation Models,” arXiv, 2025.

비지도 학습 및 다계층 네트워크를 활용한 인물 네트워크 계층 구조 추론 방법 연구

구예리[1], 김지온[2,*], 우병관[2]

[1]㈜동심우, [2]한림대학교

yerigu@dongsimwoo.com, {jion972, byeongkwan.woo}@hallym.ac.kr

Inferring Hierarchical Structures in Human Networks: An Unsupervised Learning Framework with Multi-Layer Networks

Yeri Gu[1], Jion Kim[2*], Byeongkwan Woo[2]

[1]UniSoul Friends Co., Ltd., [2]Hallym University

요 약

사이버 범죄 조직은 용의자 특정과 공범 관계 규명을 어렵게 하는 점조직 형태를 특징으로 한다. 기존의 유사도 기반 분석은 수사 단서 간의 복잡한 관계 패턴을 학습하는 데 한계가 있으며, 이는 본질적인 정보 손실을 야기한다. 이러한 문제에 대응하기 위해, 본 논문은 새로운 비지도 학습 기반 인물 계층 추론 프레임워크를 제안한다. 제안하는 프레임워크는 인물, 사건, 수사단서(계좌, IP 등)를 독립된 노드로 표현하는 다계층 이종 그래프를 구축하고, 그래프 신경망(GNN)을 이용해 관계 패턴을 직접 학습한다. GNN으로 학습된 임베딩을 활용하여 잠재적 공범 관계 예측, 은닉된 조직 구조 및 역할 분류, 그리고 네트워크 중심성과 모델의 어텐션 스코어를 통합한 핵심 인물 위험도 평가를 수행한다. 본 연구는 수사 데이터에 내재된 복합적인 구조를 효과적으로 규명하여, 수사관에게 객관적이고 정량적인 분석 근거를 제공함으로써 수사의 효율성을 제고한다는 점에서 중요한 의의를 가진다.

Ⅰ. 서 론

사이버 공간의 익명성을 악용한 지능형 범죄가 급증함에 따라, 다수의 차명 계정과 대포폰을 사용하는 점조직 형태의 범죄 조직을 규명하는 것이 수사의 핵심 과제로 부상하였다 [1]. 이러한 범죄는 명시적인 연결고리가 부족하고 관계가 복잡하게 얽혀 있어, 확보된 증거를 순차적으로 따라가는 전통적인 수사 방식으로는 전체 조직의 실체를 파악하는 데 많은 시간과 노력이 소요된다. 이로 인해 수사가 장기화되거나 미제에 빠지기 쉬우며, 이는 범죄 조직의 활동을 근절하기 어렵게 만드는 핵심적인 원인이 되고 있다.

이러한 한계를 극복하기 위해 데이터 기반의 분석 방법론이 요구되지만, 단순히 인물 간 속성(예: 동일 IP, 연락처)의 유사도를 측정하는 방식은 다양한 관계가 가진 고유한 의미와 패턴을 단일 점수로 압축하여 정보 손실을 야기하는 문제를 가진다. 따라서 복잡하게 얽혀있는 용의자 및 사건 데이터로부터 숨겨진 연결고리를 체계적으로 찾아내고, 범죄 조직의 구조를 규명하며, 핵심 인물의 위험도를 정량적으로 평가할 수 있는 새로운 분석 방법론의 필요성이 대두되고 있다.

이에 본 논문은 별도의 정답 데이터 없이 관계의 고유한 패턴을 직접 학습할 수 있는 이종 그래프 신경망(Heterogeneous GNN) 기반의 인물 계층 추론 프레임워크를 제안한다. 제안 방법론은 수사단서를 독립된 노드로 모델링하여 정보 손실을 최소화하고, GNN을 통해 데이터의 내재적 구조와 관계를 심층적으로 분석하여 ① 불상의 용의자 특정, ② 잠재적 범죄 조직 및 역할 식별, ③ 핵심 인물 및 조직의 위험도 산출을 수행한다. 이를 통해 수사 초기 단계에서 객관적이고 정량적인 분석 결과를 제공하여 수사의 효율성을 극대화하는 것을 목표로 한다.

Ⅱ. 관련 연구

범죄 네트워크 분석에 대한 초기 연구들은 주로 사회 연결망 분석(Social Network Analysis, SNA) 기법을 사용하여 특정 범죄 조직의 구조를 분석하고 핵심 인물을 식별하는 데 초점을 맞추었다 [2]. SNA는 인물 간의 관계를 시각화하고 중심성 분석 등을 통해 네트워크 내에서 영향력 있는 행위자를 식별하는 데 유용성을 보였으나, 관계의 유형을 구분하지 못하고 비정형 데이터를 처리하기 어려운 한계를 가졌다.

기계학습 기술이 발전함에 따라 더욱 고도화된 방법론이 제안되었으며, 특히 그래프 기반 준지도학습(Semi-supervised Learning, SSL)은 범죄 네트워크 분석에 효과적으로 활용되었다. Kim 등[3]은 계층적으로 구조화된 네트워크에서 SSL을 적용하여 레이블 전파(Label Propagation) 방식을 통해 소수의 레이블 된 데이터를 단서로 유사한 특성을 가지는 다른 용의자를 예측하는 프레임워크를 제시하였다. 나아가, Jhee 등[4]은 대규모 범죄 네트워크에서 그래프 기반 SSL의 느린 추론 속도 문제를 해결하기 위해, 잠재 네트워크(latent network)를 도입하여 유사 용의자를 빠르게 예측하는 프레임워크를 제안했다.

이러한 SSL 기반 연구들은 분석을 위해 일부 정답 데이터(Label)를 필요로 하며, 이를 통해 유사 용의자를 추천하는 예측 문제에 초점을 맞추었다. 반면, 최근에는 정답 데이터가 없는 상황에서 그래프의 구조적 정보를 최대한 활용하는 비지도 학습(Unsupervised Learning) 및 자기지도 학습(Self-supervised Learning) 기반의 GNN 연구가 활발히 진행되고 있다 [5, 6]. 이러한 접근법은 데이터 자체의 내재적 구조를 분석하여 네트워크의 특성을 식별하는 데 중점을 둔다.

본 연구에서 제안하는 비지도 학습 기반 접근법은 이러한 최신 연구 흐름과 맥을 같이한다. 수사단서를 독립된 노드로 취급하는 이종 그래프 모델과 GNN을 통해 관계의 의미적, 구조적 패턴을 직접 학습함으로써, 단순 유사도 계산으로 인한 정보 손실 문제를 극복하고자 한다. 이는 특정 용의자를 추천하는 것을 넘어, 데이터 내에 숨겨진 공범 관계 예측, 잠재적 범죄조직 식별, 각 인물의 위험도 산출 등 네트워크 전체에 대한 거시적이고 복합적인 통찰을 제공한다는 점에서 선행 연구와 차별성을 가진다.

Ⅲ. GNN 기반 인물 계층 추론 방법론

본 논문에서 제안하는 인물 계층 추론 방법론은 그림 1과 같이 3단계로 구성된다. 첫째, 이종 그래프 구축 단계에서는 다양한 출처의 수사 데이터를 통합하여 그래프로 모델링한다. 둘째, GNN 기반 관계 학습 및 조직 분석 단계에서는 GNN을 통해 인물 간의 잠재적 관계와 조직 구조를 학습하고 분석한다. 셋째, 위험도 평가 단계에서는 식별된 인물과 조직의 위험도를 정량적으로 평가하여 수사 우선순위를 도출한다.

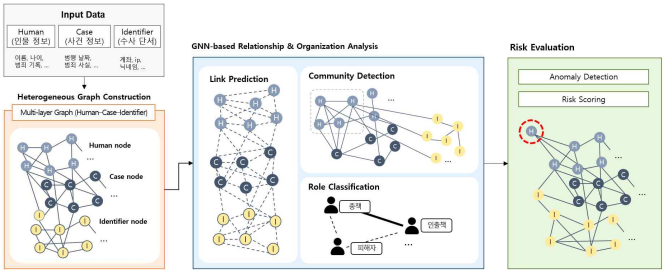


그림 1 인물 계층 추론 모듈 개념도

1. 이종 그래프 구축 (Heterogeneous Graph Construction)

분석의 첫 단계는 텍스트 형태의 범죄사실, 인물 정보, 계좌 거래 내역, 통신 기록 등 정형·비정형 데이터를 통합하여 그래프 $G = (V, E)$ 형태로 변환하는 것이다. 본 연구에서는 기존의 인물-사건 2계층 구조에서 더 나아가, 정보 손실을 최소화하고 관계의 구조적 근거를 보존하기 위해 수사단서(Identifier)를 독립된 노드 유형으로 추가한 3계층 이종 그래프를 구축한다.

- 인물 노드 $(v_h \in V_H)$: 용의자, 피해자, 참고인 등 사건에 관련된 모든 인물. 각 노드는 이름, 연락처, 계좌번호, IP 주소 등 인물 고유의 속성 정보를 가짐
- 사건 노드 $(v_c \in V_C)$: 개별 범죄 사건. 각 노드는 사건 번호, 범죄 사실 등의 속성 정보를 가짐
- 수사단서 노드$(v_i \in V_I)$: 인물과 사건을 연결하는 매개체. 계좌번호, IP 주소, 전화번호, 암호화폐 지갑 주소 등이 포함되며, 각 단서의 유형(type)과 값(value)을 속성으로 가짐
- 엣지 $(e \in E)$: 인물과 사건, 인물과 수사단서, 사건과 수사단서 간의 관계를 정의한다. 예를 들어, (Human)-(Identifier) 엣지는 특정 인물이 특정 계좌를 '소유' 또는 '사용'했음을 의미한다.

표 1 Human-Identifier-Case 이종 그래프 스키마

구분	노드/엣지 유형	설명
노드	Human(인물)	용의자, 피해자 등 사건의 모든 주체
	Case(사건)	개별 범죄사건
	Identifier(수사단서)	인물/사건 연결하는 수사단서(계좌, IP, 전화번호 등)
엣지	(Human)-(Case)	인물의 사건 참여 관계(역할: 피의자/피해자)
	(Human)-(Identifier)	인물과 단서의 관계
	(Case)-(Identifier)	사건과 단서의 연루 관계

2. GNN 기반 관계 학습 및 조직 분석

본 프레임워크의 핵심은 구축된 이종 그래프를 비지도 학습 방식으로 GNN에 학습시키는 것이다. 이는 '어떤 인물이 범죄 조직의 일원인지'와 같은 사전 정보나 정답(Label) 없이, 데이터에 내재된 관계 패턴과 구조만으로 네트워크를 이해하는 접근법이다. GNN은 각 노드 주변의 이웃 노드와 엣지 정보를 반복적으로 집계하여, 각 노드를 저차원 벡터 공간에 임베딩(Embedding)한다. 이 과정에서 '어떤 관계(금융 거래, IP 공유 등)가 인물 간의 긴밀도를 나타내는지'를 데이터로부터 자동으로 학습하여, 수동 가중치 설정의 주관성을 배제하고 객관적인 분석을 가능하게 한다.

이렇게 학습된 임베딩 벡터는 각 노드의 다차원적인 특성을 함축하고 있어, 이를 활용해 다음과 같은 심층 분석을 수행한다.

- 잠재 관계 예측 (Link Prediction): 임베딩 벡터 간의 유사도를 계산하여 명시적으로 드러나지 않은 잠재적 공범 관계를 예측한다.
- 조직 식별 (Community Detection): 임베딩된 인물 노드들을 대상으로 Louvain Method와 같은 군집 분석 알고리즘을 적용하여 밀집도가 높은 잠재적 범죄 조직을 식별한다.
- 역할 분류 (Role Classification): GNN 모델 위에 분류기(Classifier)를 추가하여, 학습된 노드 임베딩을 바탕으로 각 인물의 역할(예: 총책, 인출책 등)을 분류하는 모델을 개발할 수 있다.

3. 위험도 평가 (Risk Assessment)

식별된 인물과 조직 중 수사력을 집중해야 할 대상을 선정하기 위해 위험도를 다각적으로 평가한다. 이는 네트워크의 구조적 특성, GNN이 학습한 잠재적 영향력, 그리고 실제 범죄의 심각성을 종합적으로 고려하여 이루어진다. 본 연구에서는 인물 v의 위험도 $R_{(v)}$를 식(1)과 같이 정의한다.

$$R_{(v)} = w_1 C_{D,B}(v) + w_2 C_{imp}(v) + w_3 E_S(v) \qquad (1)$$

여기서 네트워크 구조 기반 지표인 $C_{D,B}(v)$는 연결 중심성(Degree Centrality)과 매개 중심성(Betweenness Centrality) 등 전통적인 네트워크 중심성 지표를 정규화한 점수이다. 연결 중심성은 한 인물이 얼마나 많은 다른 인물과 직접적인 관계를 맺고 있는지를 측정하며, 이 값이 높은 인물은 조직 내에서 허브(Hub) 역할을 수행할 가능성을 시사한다. 매개 중심성은 네트워크 내 정보나 자금의 흐름에서 한 인물이 얼마나 중요한 길목을 차지하는지를 측정하며, 이 값이 높은 인물은 조직 내 중간 관리자 또는 정보의 통제자 역할을 할 가능성이 높다. $C_{imp}(v)$는 GNN 모델이 학습 과정에서 산출하는 노드 중요도(Node Importance) 또는 어텐션 스코어(Attention Score)를 의미한다. 이는 단순히 연결 관계의 수나 위치뿐만 아니라, 주변 노드와의 복합적인 관계 패턴을 통해 학습된 해당 인물의 잠재적 영향력을 나타낸다. 예를 들어, 직접적인 연결은 적더라도 중요한 인물들과 다수 연결된 경우, GNN은 해당 노드의 중요도를 높게 평가할 수 있어 전통적 지표가 포착하지 못하는 영향력을 측정할 수 있다. $E_S(v)$는 해당 인물이 연루된 사건들의 평균 심각도를 나타내는 도메인 지식을 정량화하여 반영한 점수이다.

최종적으로, 이 세 가지 지표를 가중치 w_1, w_2, w_3을 통해 결합함으로써, 한 인물에 대한 다면적이고 종합적인 위험도 점수를 산출한다. 이 점수는 수사 우선순위를 결정하고 한정된 수사 자원을 효율적으로 배분하기 위한 객관적인 근거로 활용될 수 있다.

Ⅳ. 결론 및 향후 연구

본 논문은 사이버 범죄 네트워크 분석의 한계를 극복하기 위해, 수사단서를 포함한 다계층 이종 그래프와 비지도 학습 기반의 GNN을 결합한 새로운 인물 계층 추론 프레임워크를 제안했다. 제안 방법론은 관계 패턴의 직접 학습을 통해 기존 방식의 정보 손실 문제를 해결하고, 숨겨진 인물 계층 구조를 효과적으로 추론할 수 있는 이론적 기반을 마련했다. 본 방법론은 잠재적 연결고리와 조직 구조를 규명하고, 핵심 용의자를 객관적인 지표로 특정하여 수사관의 의사결정을 지원하는 데 의의가 있다.

향후 연구로는 실제 데이터를 적용하여 제안 프레임워크의 실효성을 검증하고, 다양한 GNN 아키텍처를 실험하여 추론 정확성을 향상시키는 연구가 필요하다. 또한, 각 지표의 최적 가중치(w_1, w_2, w_3)를 데이터 기반으로 학습하고, 설명가능 AI(XAI) 기법을 도입하여 모델의 추론 근거를 시각적으로 제시하는 연구를 통해 방법론을 고도화할 수 있을 것이다.

ACKNOWLEDGMENT

이 논문은 2025년도 정부(경찰청)의 재원으로 과학치안진흥센터 사이버범죄 수사단서 통합분석 및 추론 시스템 개발 사업의 지원을 받아 수행된 연구임(No. RS-2025-02218280).

참 고 문 헌

[1] 경찰청 사이버안전지킴이, "2023 사이버범죄 동향 분석," 2024.

[2] Sparrow, M. K. "The application of network analysis to criminal intelligence: An assessment of the prospects," Social networks, vol. 13, no. 3, pp. 251-274, 1991.

[3] M. Kim, D.-g. Lee, and H. Shin, “Semi-supervised learning for hierarchically structured networks,” Pattern Recognition, vol. 95, pp. 191-200, Nov. 2019.

[4] J. H. Jhee, M. J. Kim, M. Park, J. Yeon, and H. Shin, "Fast Prediction for Criminal Suspects through Neighbor Mutual Information-Based Latent Network," International Journal of Intelligent Systems, vol. 2023, Article ID 9922162, 2023.

[5] T. N. Kipf and M. Welling, "Variational graph auto-encoders," arXiv preprint arXiv:1611.07308, 2016.

[6] W. Wang, Z. Liu, J. E. Tedesco, and J. M. Jose, "Graph-based social media analysis for fake news detection: a survey," ACM Computing Surveys, vol. 55, no. 8, pp. 1-36, 2023.

간첩통신에 활용된 스테가노그라피 특징과 수사기관의 법적쟁점

오 소 정, 장 지 혜, 김 기 범*
성균관대학교 과학수사학과
mira0809@naver.com, jihye01@skku.edu, *freekgb02@gmail.com

Steganography in Espionage Communications: Characteristics and Legal Challenges for Law Enforcement

Oh SoJung, Jang JiHye, *Kim Gibum
Dept. Forensic Science, Sungkyunkwan University

요 약

본 논문은 국내 간첩통신에서 스테가노그라피의 실제 운용과 이에 따른 수사·재판상의 법적 쟁점을 종합적으로 분석한다. 텍스트·이미지·오디오 등 매체별 은닉 기법의 발전과 탐지 한계를 정리하고, 한글 문서 구조 활용과 이중 은닉 등 실제 현장에서의 운용 패턴을 기술적 관점에서 도출하였다. 또한 위장 파일로 인한 선별 압수의 어려움, 참여권과 수사기밀의 충돌, 추출과정 재현성·신뢰성에 따른 증거능력 다툼, 복호화명령 부재 등을 핵심 쟁점으로 규명하였다. 본 연구는 향후 스테가노그라피 기반 간첩통신 수사에서의 기술·제도적 대응체계 수립을 위한 실증적 근거를 제시한다.

Ⅰ. 서 론

스테가노그라피(steganography)는 숨겨진 메시지의 존재 자체를 감지하지 못하도록 정보를 은닉하는 기술을 의미한다 [1]. 그간, 스테가노그라피를 사용하는 수단과 방법, 활용분야에서 폭발적인 발전을 이루어왔다. 저작권보호, 개인정보보호 등 긍정적인 활용도 있으나 강력한 은폐성으로 전시 상황에서 간첩들의 통신 수단으로 활용되고 있다 [2]. 대표적으로 한국은 왕재산사건[3], 김목사사건[4]과 같이 상부 지시를 수신하거나 보고하는데 사용되었다.

국내법상, 형사소송법은 사건과 관련 있는 파일만을 선별적으로 압수할 수 있다고 규정한다(제106조 제3항). 그러나 스테가노그라피 파일은 제한시간 내 사건과의 관련성 판단이 어려워 적용에 한계가 있다. 또한 대법원이 스테가노그라피 탐지 기술을 “국가안보”로 간주해 공개하지 않음에 따라 연구와 대응 기술 개발에도 제약이 따른다. 따라서 본 논문은 스테가노그라피의 범죄 활용 양상 분석과 국내외 법제 검토를 통해, 향후 수사에서 발생할 수 있는 형사법적 쟁점을 제시하여 대응방안 개선에 기여하고자 한다.

Ⅱ. 스테가노그라피 탐지 한계 및 발전 현황

2.1 스테가노그라피의 개념과 특성

스테가노그라피는 다양한 관점에서 정의되고 있으나, 대부분 데이터은닉 기술(data hiding)로 분류하고 있다 [2]. 정보가 숨겨져 있다는 사실 자체를 감추는 비밀통신 방법으로, 비밀메세지(Secret Msg)를 커버파일(Cover File)에 숨겨 스테고파일(Stego File)을 생성하는 기술을 의미한다 [5].

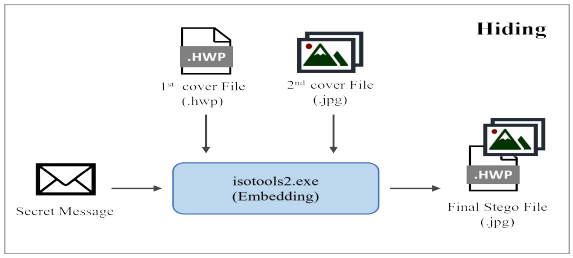
2.2 스테가노그라피의 발전

스테가노그라피는 텍스트, 이미지, 오디오 등 커버파일의 유형에 따라 구분된다. 텍스트 기반 기법은 전통적인 null cipher 방식 외에도 특정 글자 위치 삽입, 동의어 치환, 철자 변형, 시각적으로 구분이 어려운 유니코드 사용 등 다양한 변형이 존재한다. 이미지의 경우, 픽셀 값을 직접 수정하거나 무손실 포맷(BMP)을 이용한 방식이 대표적이며, LSB(Least Significant Bit) 삽입, PVD(Pixel Value Differencing) 기법이 널리 사용된다. 정보전쟁, 간첩활동 등에 적극 활용됨에 따라 탐지 기술 고도화와 신속 대응이 요구된다.

Ⅲ. 대법원 판례 분석을 통한 간첩통신 특징

3.1 PC방 간첩 사건

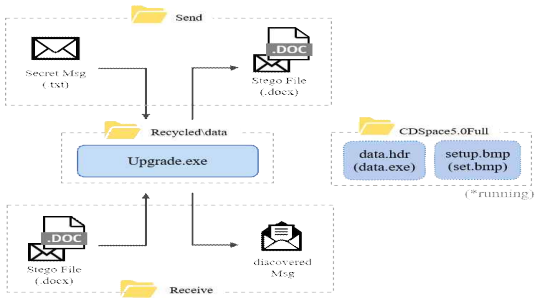
PC방 간첩 사건은 2014년 3월부터 2015년 8월까지 대한민국에서 발생한 간첩사건으로, 피고인들은 국가보안법상 잠입·탈출, 회합·통신 등의 혐의로 기소되었다 [6]. 디지털 포렌식 분석 결과, 피고인들의 하드디스크에서 스테고 파일로 의심되는 문서(trekking.hwp)를 확인되었고, 비할당영역에도 다수의 보고문건 조각과 $$$6BAD.tmp 등 스테고파일과 유사한 임시파일들도 발견하였다. 윈도우 아티팩트 분석 결과 isotools2.exe의 사용 흔적이 확인되었으며, 한글 문서 구조(자음·모음 분석)를 활용해 은닉 공간을 확보하고 자동으로 메시지를 삽입하는 기능을 탑재한 것으로 판단된다. 정리하면, 한글 문서 구조를 활용한 형식적인 방법과, 이후 문서를 사진(JPG)에 은닉하여 이중으로 활용한 중첩 기술로 볼 수 있다.



<PC방 간첩사건 스테가노그라피 예상 개요도>

3.2 왕재산 간첩 사건

왕재산 사건은 2011년 북한 225국과 연계하여 기술 벤처기업을 설립해 간첩 활동을 한 혐의로 국가보안법 위반으로 기소된 사건이다 [8]. 디지털 포렌식 분석 결과, 피고인이 국가기밀에 해당하는 문건을 작성하여 피고인들 사이에 공유한 사실이 확인되었다. 또한, 공작원과의 은밀한 통신에 사용된 것으로 추정되는 스테가노그라피 도구, 커버 파일, 은닉 파일이 피고인의 하드디스크에서 발견되었다. 판결문에 따르면 해당 도구는 세 개의 파일(USB data.hdr, Setup.bmp, UPGRADE.EXE)로 구성되어 있었으며, 피고인의 휴대전화에서 사용 방법에 관한 메모도 압수되었다. UPGRADE.EXE는 주 프로그램, data.hdr과 Setup.bmp는 스테고 키(stego keys)로 추정되었다. 검찰 발표에 따르면, 해당 파일은 특정 폴더와 파일명 변경 과정을 거쳐야 실행이 가능하도록 설계되어 있었다 [8]. 송신자와 수신자가 공유한 백엔드 프로그램이 '대칭 알고리즘 기반 스테가노그라피'의 비밀키(secret key)로 기능하였다.



<왕재산간첩 사건 스테가노그라피 예상 개요도>

Ⅳ. 간첩통신 수사 법적 쟁점 및 대응

4.1 관련성 판단과 참여권 보장

우리나라 형사소송법과 반하여 스테고파일은 일반 파일로 위장되기 때문에 사건 관련성을 단기간에 특정하기 어렵다. 미국도 우리나라와 동일한 상황(수정헌법 제4조, 연방형사소송규칙 제41)인 반면, 일본은 영장 발부 요건은 엄격하나 디지털증거는 영장에 특정되지 않아도 전부 압수가 가능하다. 한편, 참여권은 한국, 일본, 미국 등에서 강력히 보장되며, 미국은 압수물 목록 작성과 제3자 입회를 통해 무차별적 수집을 방지한다. 그러나 국가안보와 직결되는 중대한 범죄는 증거능력 인정을 위해 압수·수색 요건을 완화할 필요가 있다. 또한 분석과정은 비공개로 하고 수집 및 선별과정에서만 공개하여 참여권을 보장하는 등 국가 이익에 따라 조정이 필요하다.

4.2 증거능력 보장

스테고파일은 비밀 메시지가 비가시적이므로 추출 과정을 증명해야 하며, 이때 불법 수집 증거 논란이나 전언증거(hearsay) 주장에 대응하기 어렵다. 분석기술은 보안상 완전 재현이 힘들고 수사기관 자체 개발 도구도 오류 가능성이 있어, 증거능력 판단은 법원의 결정에 따라야 한다. 따라서 '국가보안기술위원회'를 구성해 당해 사건에 한정해 증거능력 평가 절차를 수립할 필요가 있다. 또한 국가차원에서 스테가노그라피의 식별, 해석, 분석 등 전 과정을 아우르는 표준 매뉴얼을 제정해 포렌식 체인(chain of custody)과 연계하여 증거능력 판단의 참고자료로 활용할 수 있어야 한다.

4.3 복호화명령 도입 검토

스테가노그라피를 암호로 간주해 영국의 복호화명령(decryption order) 제도 도입을 검토할 필요가 있다. 이 제도는 「수사권한규제법(RIPA, Regulation of Investigation Powers Act 2000)」, 「경찰 및 형사증거법(PACE Act, Police and Criminal Evidence Act 1984)」에 규정되어 있다. RIPA 제49조(Section 49 Notice)는 특정인에게 암호키 공개나 식별 가능한 형태의 정보 제출을 요구할 수 있도록 한다. 이 통지는 국가안보, 범죄 탐지 또는 방지, 영국의 경제적 이익을 위해 발부될 수 있으며(제49조 제2항), 스테가노그라피는 국가안보 범죄에 활용되므로 요건을 충족한다. 반면 미국과 일본은 진술거부권 등을 사유로 복호화명령 도입에 극명한 한계를 보이고 있다. 그러나 국가의 주요 책무는 국민의 법 감정을 이해하고, 모든 국민이 납득할 수 있는 법을 제정하는 것이다. 스테가노그라피가 국가안보와 경제를 위협하는 현실을 고려할 때, 혁신적인 제도를 적극 도입할 필요가 있다.

Ⅴ. 결론

스테가노그라피가 한국에서는 간첩통신에 사용되면서 국가안보, 테러와도 연계되고 있다. 본 논문은 스테가노그라피는 한국에서 어떤 형태로 사용되는지 살펴보고, 수사기관의 대응에 있어 조치해야 하는 자세에 대해서 분석하였다. 특히, 대법원 판례를 중심으로 북한의 간첩통신 실태를 파악하고, 특징을 분석하여 국가안보와 어떻게 직결되는지 살펴보았다. 또한 수사기관이 스테가노그라피 탐지, 분석과정에서 겪게 되는 (1) 관련성 판단문제, (2) 참여권보장 문제, (3) 증거능력보장 문제, (4) 복호화명령 등 중 제도 미비에 따른 제한 문제 등을 분석하였다. 이 연구를 통해 간첩통신에 대해 경각심을 갖고, 스테가노그라피 대응이 개선되기를 기대한다.

ACKNOWLEDGMENT

이 논문은 2025년 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No. RS-2024-00398745, 디지털 환경에서의 증거인멸행위 증명 및 대응기술 개발)

참 고 문 헌

[1] Johnson, N. F, and Jajodia, S, "Exploring steganography: Seeing the unseen," Computer, vol. 31, no. 2, pp.26-34 1998.

[2] 조재호, 정광식. 사이버포렌식에 대응하는 기술, 안티포렌식. 에피스테메, 2018.

[3] 대법원 2013.7.26. 선고 2013도2511 판결.

[4] 대법원 2017.11.29. 선고 2017도9747 판결.

[5] 오소정, 주지연, 박현민, 박정환, 신상현, 장응혁, 김기범, "안보사건에서 스테가노그라피 분석 및 형사법적 대응방안" 정보보호학회논문지, vol. 2, no. 4, pp. 723-736, 2022.

[6] 서울중앙지방법원 2016.12.23. 선고 2016고합675 판결; 서울고등법원 2017.7.5. 선고 2017노146 판결; 대법원 2017.10.31. 선고 2017도12643 판결

[7] 서울중앙지방법원 2012.2.23. 선고 2011고합1131, 2011고합1143(병합), 2011고합1144(병합), 2011고합1145(병합), 2011고합1146(병합) 판결; 서울고등법원 2013.2.8. 선고 2012노805 판결; 대법원 2013.7.26. 선고 2013도2511 판결

[8] 윤신자, "멀티미디어 정보은닉 등 사이버상 북한지령에 관한 디지털포렌식 관점에서의 연구:국가안보사건 전자증거 압수수색 절차를 중심으로," 고려대학교 정보보호대학원 석사학위논문, 2014.

저전력/고성능 Edge AI 위성 탑재체 개념 연구

이용민, 장동필
한국전자통신연구원 위성탑재체연구실
ymlee01@etri.re.kr, dpjang@etri.re.kr

A Study on the Concept of Low-power, High-performance On-board Edge AI Payload for Satellites

Yong-Min Lee, Dong-Pil Jang
Satellite Payload Research Section
Electronics and Telecommunications Research Institute (ETRI)

요 약

본 논문은 통신 효율, 위성체 운용 자율성, 임무 효율성 등을 향상시킬 수 있는 저전력, 고성능 Edge AI 위성 탑재체의 개념적 접근에 관한 것으로 구체적으로는 데이터 절감 및 전송효율 증대, 지능형 지구관측, 통신 네트워크 효율과 신호 품질 향상, 대형 AI 모델의 분산 처리, 탑재체 내 자율 이상 감지 및 대응, 다층 위성망 구조에서의 복원력 강화, 저궤도 위성군을 활용한 협업 탐지 및 강화학습 기반의 의사결정 등의 가능성을 제시하였다.

Ⅰ. 서 론

저궤도 위성 시장 규모는 2025년 기준 약 110조원 규모에서 2032년에는 약 160조원 이상으로 빠르게 성장하고 있으며[1], 5G 기반의 OneWeb 위성 한 기당 최대 전송용량은 약 7.5 Gbps이고 Starlink 저궤도 위성의 경우 수백 Mbps 수준의 다운링크 속도를 제공하고 있다[2]. Edge AI 탑재체 기술은 위성 자체에서 실시간 데이터 처리를 통하여 지구로 전송되는 데이터량과 지연의 최소화를 비롯하여 통신 최적화 및 분산 처리, 자율 이상 감지 및 대응, 다층 구조 기반의 복원력 강화 등을 통해 위성 수명 연장과 위성망의 강건함을 높이는 등 다양한 장점을 가지고 있다.

Ⅱ. 저전력, 고성능 Edge AI 위성 탑재체

1. 데이터량 축소 및 전송효율 향상

Edge AI 위성 탑재체의 효용성 극대화 가능성 중의 하나는 EO/IR과 같은 광학탑재체와 합성개구레이더(SAR) 탑재체 등에서 생성되는 대량의 이미지 데이터를 온보드에서 선별, 정제, 압축하여 전송함으로써 지구국 전송을 위한 데이터량을 효과적으로 줄일 수 있다[3]. 예를 들어 광학 탑재체에서 주기적으로 수집한 이미지 중에서 구름 낀 영상과 같이 불필요한 영상 데이터는 AI를 통해 제외함으로써 전송효율을 높일 수 있다.

2. 지능형 지구관측

지구관측 위성에 온보드 AI 기술을 적용할 경우 탑재체 내에서 실시간 이미지 분석이 가능할 뿐만 아니라 이상 탐지가 가능하여 미리 설정된 기준에 따라 이벤트 정보만 수집하여 저전력 채널로 즉각적인 지구 전송이 가능하다[4].

3. 통신 최적화 및 분산 처리 기능

탑재체 내의 신경망 기반 모델을 사용함으로써 무선 주파수 자원관리, 적응형 빔포밍 등 네트워크 효율과 신호 품질을 향상시킬 수 있다[5]. 또한 대형 AI 모델을 적절히 분할한 후 광학 ISL(Inter-Satellite Links) 등을 이용해 위성 간 네트워크에 분산 배치하고, 계산량이 큰 인코더는 지상 클라우드를 사용하는 것을 고려할 수 있다[6].

4. 자율 이상 감지 및 대응

FPGA 기반 엣지 컴퓨팅 (Edge-computing) 기술을 이용하여 열화상 카메라가 특정 전자회로 또는 탑재체 영역을 촬영하여 열 분포 이미지를 수집하고 지상에서 학습된 SVM(Support Vector Machine) 머신러닝 알고리즘을 FPGA에 구현, 이미지 특징을 추출하고 열 분포 패턴을 분석하여 이상 여부를 분류할 수 있다[7]. 이와 같은 탑재체 내 자율 열적 이상 조기 감지는 하드웨어의 심각한 손상을 사전에 방지하고 위성 수명을 연장하는 효과가 있다.

5. 다층 구조 기반의 복원력 강화

위성 네트워크는 다양한 궤도(LEO, MEO, GEO)를 포함하는 다층 구조를 통해 상호 보완적 작동이 가능하며, 한 궤도의 위성에 문제가 생겨도 다른 레이어를 통해 서비스 연속성을 유지할 수 있다[8]. 예를 들어, LEO의 고속 전송 채널에 문제가 발생한 경우, GEO 또는 MEO 위성이 통신 보조 역할을 수행할 수 있는데 이때 AI가 실시간으로 여러 궤도 간 신호 품질과 라우팅을 분석하여 자동 전환함으로써 네트워크의 복원력을 높일 수 있다.

6. 협업 탐지 및 강화학습 기반 의사결정

여러 대의 LEO 위성이 공동으로 딥러닝 기반의 물체 탐지를 수행하고 각 위성의 탐지 시점(viewpoint)을 최적화함으로써 탐지 정확도를 높일 수 있다[9]. 또한 강화학습을 통해 위성군 및 지상의 센서를 연결하여 위

성체가 실시간으로 충돌 회피 기동을 결정하거나 다른 위성에 실시간 임무를 전달하는 자동 전환식 관찰 및 연계(tipping and queueing) 방식의 AI 오케스트레이션 시스템 구현이 가능하다.

Ⅲ. 결론

본 논문에서는 저전력/고성능 Edge AI의 위성 탑재를 통해 통신 부담의 감소 및 임무 효율성을 증대, 지능형 지구관측, 통신 최적화 및 분산 처리, 위성체 내 자율 이상 감지 및 대응, 다층 구조 기반의 복원력 강화, 협업 탐지 및 강화학습 기반 의사결정 등의 가능성을 언급하였다. 그러나 저전력·고성능 처리 능력을 갖춘 AI 프로세서의 극한 온도변화, 방사선 등 가혹한 우주 환경에서 견딜 수 있는 내성 강화와 데이터 분포 변화, 센서 드리프트(sensor drift), 결함 예측(fault detection) 등 데이터 처리 안정성 확보가 중요한 기술적 과제라고 할 수 있다.

ACKNOWLEDGMENT

본 논문은 2025년도 과학기술정보통신부, 해양경찰청, 환경부의 재원으로 정보통신기획평가원의 지원을 받아수행된 연구임 (No.2021-0-01715, 정지궤도 공공복합통신위성 통신탑재체 개발)

참 고 문 헌

[1] https://www.fortunebusinessinsights.com/leo-satellite-market-112113? utm_source=chatgpt.com

[2] https://www.researchgate.net/publication/375101582_Free_Space_Optical_Communication_An_Enabling_Backhaul_Technology_for_6G_Non-Terrestrial_Networks?utm_source=chatgpt.com

[3] https://cavuaerospace.uk/articles/the-demands-of-space-edge-computing/?utm_source=chatgpt.com "The Demands of Space Edge Computing – CAVU Aerospace UK Ltd"

[4] https://satellogic.com/2025/03/20/pushing-intelligence-to-the-edge-satellogics-vision-for-ai-powered-earth-observation/?utm_source=chatgpt.com "Pushing Intelligence to the Edge: Satellogic's Vision for AI-Powered Earth Observation – Satellogic"

[5] https://arxiv.org/abs/2308.11152?utm_source=chatgpt.com,"Energy-Efficient On-Board Radio Resource Management for Satellite Communications via Neuromorphic Computing"

[6] https://arxiv.org/abs/2504.01676?utm_source=chatgpt.com "Satellite Edge Artificial Intelligence with Large Models: Architectures and Technologies"

[7] https://www.sciencedirect.com/science/article/abs/pii/S027311772500002X? utm_source=chatgpt.com, "Edge computing in space: Design of an FPGA architecture for thermal anomaly detection based on a machine learning approach – ScienceDirect"

[8] https://www.itweb.co.za/article/multi-layered-satellite-solutions-are-now-required/G98YdMLG8JB7X2PD?utm_source=chatgpt.com, "Multi-layered satellite solutions are now required". 2024.10.

[9] https://arxiv.org/abs/2508.00755?utm_source=chatgpt.com, "AI-Driven Collaborative Satellite Object Detection for Space Sustainability", 2025.8.

EEG 및 시선추적 신호 기반 가독성 평가 상관분석

윤세휘 [1], 송명주 [1], 황세미 [1, 2 †]
[1] 모두의연구소, [2] 연세대학교 원주연세의료원

soyshoong@gmail.com, myungjoo.song12@gmail.com, semi.hwang11@gmail.com

Correlation Analysis of Readability Assessment Using EEG and Eye-Tracking Signals

Sehwi Yoon[1], MyungJoo Song[1], Semi Hwang[1, 2]
[1]MODULABS, [2] YONSEI UNIVERSITY WONJU HEALTH SYSTEM

Abstract

Conventional readability indices rely on surface-level linguistic features and often fail to capture the cognitive effort involved in reading. We present a cognitively grounded framework that integrates linguistic, eye-tracking, and EEG signals for readability estimation. Using the ZuCo 2.0 dataset, we combined sentence length, fixation duration, and oscillatory power in the alpha and theta bands with human readability ratings. Correlation analyses show that multimodal features outperform unimodal measures, with the integration of linguistic and cognitive signals yielding stronger alignment with human judgments. These results suggest that readability emerges from the interaction between text and cognition with practical implications for adaptive learning and accessibility.

Ⅰ. INTRODUCTION

Readability, defined as the ease with which a reader can process and understand text, is important in education, accessibility, and NLP. Conventional readability metrics such as FRES, FKGL, and SMOG rely on surface-level linguistic features like sentence length and word difficulty. While simple and interpretable, they may not adequately capture the cognitive demands readers experience in natural settings.

Psycholinguistic and cognitive neuroscience research shows that reading reflects an interaction between text and cognition. Eye-tracking measures (fixation count, first fixation duration, gaze duration) and EEG oscillations in the theta (4– 7 Hz) and alpha (8– 12 Hz) bands are sensitive indicators of cognitive load. These findings suggest that cognitive signals can enrich traditional readability indices.

However, few studies systematically combine linguistic, eye-tracking, and EEG features to model human readability judgments. Prior work offers partial evidence, but integration across modalities remains limited. To address this gap, we propose a multimodal approach to readability estimation, hypothesizing that combining linguistic and cognitive signals will align more closely with human judgments than unimodal metrics.

Ⅱ. METHODS

This study employed the ZuCo 2.0 dataset, which originally contains 738 English Wikipedia sentences with simultaneous EEG and eye-tracking recordings from 18 participants. For our analysis, we selected a subset of 150 sentences. Since the dataset does not include readability annotations, we additionally collected human readability ratings. A total of 50 raters participated, with each rater evaluating 15 sentences, and each sentence receiving ratings from three independent raters. The averaged score per sentence was used as the ground truth for analysis.

Feature extraction was performed across three modalities. Linguistic features included sentence length, the number of difficult words, and classical readability indices such as FRES and FKGL. Eye-tracking features consisted of the number of fixations (nFix), first fixation duration (FFD), and total gaze duration (GD), which reflect attentional allocation during reading.

EEG features were derived from the alpha and theta frequency bands. Specifically, we extracted the mean, maximum, and standard deviation of band power across frontal (Fz, F3, F4) and parietal (Pz, P3, P4) channels. Variables denoted in Table 1 correspond to

the maximum and standard deviation of oscillatory power, respectively.

All features were normalized at the participant level and aggregated by sentence. Normalization was performed using z-score transformation at the participant level to reduce inter-individual variability and to ensure comparability across subjects. To evaluate their relevance, Pearson correlation analysis was conducted between the extracted features and human readability ratings, both individually and in multimodal combinations.

	coefficient
Flesch Reading Ease	0.296
Flesch-Kincaid Grade	0.50
sentence length	**0.610**
# difficult words	0.527
nFix	0.173
FFD	0.226
GD	0.226
α_{std}	0.18
α_{max}	0.191
θ_{std}	0.191
θ_{max}	0.194
EEG_{std}	0.168
sentence length + FFD	0.611
sentence length + α_{std}	0.612
sentence length + α_{max}	0.618
sentence length + θ_{std}	0.612
sentence length + θ_{max}	0.618
sentence length + α_{max} + θ_{max}	0.6240
sentence length + FFD + α_{max} + θ_{max}	**0.6242**

Table 1. Pearson correlation coefficients between human readability ratings and various linguistic, eye-tracking, and EEG-based features

Ⅲ. RESULTS AND DISCUSSIONS

Table 1 summarizes the correlation coefficients between human readability ratings and the extracted features. Among linguistic features, sentence length showed the highest correlation (r = 0.610), followed by the number of difficult words (r = 0.527) and FKGL (r = 0.50). Eye-tracking signals demonstrated moderate but consistent associations, as both first fixation duration and gaze duration achieved correlations of r = 0.226, while the number of fixations was slightly lower at r = 0.173. EEG-derived features exhibited smaller yet meaningful contributions, with theta maximum reaching r = 0.194, aligning with prior evidence that oscillatory power in the theta and alpha bands reflects working memory and processing load.

Notably, multimodal combinations yielded higher alignment with human judgments. As shown in Table 1, the integration of sentence length, first fixation duration, and alpha/theta maxima produced the highest observed correlation (r = 0.6242). While this represents only a moderate association, it nevertheless outperforms unimodal feature sets.

produced the strongest correlation (r = 0.6242), outperforming any unimodal feature set. These multimodal gains are consistent with Hollenstein et al. (2019), who demonstrated that EEG and eye-tracking features contribute complementary information to linguistic features in comprehension-related tasks. Compared to François & Miltsakaki (2012), who reported moderate improvements over traditional readability formulas using NLP-based features, our results suggest that direct incorporation of cognitive signals offers an even stronger alignment with human judgments.

The observed theta-band contributions align with prior findings that theta oscillations reflect working memory load during sentence processing. This suggests that readability judgments are sensitive not only to text difficulty but also to cognitive effort required for semantic integration.

Overall, these findings suggest that readability is not solely determined by textual properties but emerges through the interaction of text with human cognitive processes. By incorporating neural and behavioral signals, the proposed approach provides a more human-centered and cognitively grounded method of readability assessment, with potential applications in adaptive learning, accessibility, and inclusive language technologies.

Ⅳ. CONCLUSION

This study presented a cognitively grounded analysis of readability assessment by integrating linguistic features with EEG and eye-tracking signals. The findings demonstrate that multimodal features, reflecting neural load and attentional effort, align more closely with human judgments than traditional surface-level metrics. By highlighting the interaction between text and cognition, this work advances the understanding of readability assessment and provides initial empirical evidence, though future work should employ regression or machine learning models to validate predictive power.

ACKNOWLEDGMENT

This research was supported by Brian Impact Foundation, a non-profit organization dedicated to the advancement of science and technology for all.

References

[1] R. Flesch, "A new readability yardstick," Journal of Applied Psychology, vol. 32, no. 3, pp. 221-233, 1948.

[2] J. P. Kincaid, R. P. Fishburne, R. L. Rogers, and B. S. Chissom, "Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy enlisted personnel," Research Branch Report 8-75, Naval Technical Training Command, Millington, TN, 1975.

[3] G. H. McLaughlin, "SMOG grading—A new readability formula," Journal of Reading, vol. 12, no. 8, pp. 639-646, 1969.

[4] J. S. Chall and E. Dale, Readability Revisited: The New Dale- Chall Readability Formula. Brookline, MA: Brookline Books, 1995.

[5] T. François and E. Miltsakaki, "Do NLP and machine learning improve traditional readability formulas?," Proc. 1st Workshop on Predicting and Improving Text Readability for Target Reader Populations (PITR), pp. 49-57, 2012.

[6] N. Hollenstein, M. Tröndle, C. Zhang, and N. Langer, "ZuCo 2.0: A dataset of physiological recordings during natural reading and annotation," Proc. Int. Conf. on Language Resources and Evaluation (LREC), pp. 138-146, 2020.

[7] N. Hollenstein, M. Tröndle, M. Plomecka, S. Kiegeland, Y. Özyurt, L. A. Jäger, and N. Langer, "The ZuCo benchmark on cross-subject reading task classification with EEG and eye-tracking data," Frontiers in Psychology, vol. 13, 1028824, 2023.

[8] W. Klimesch, "EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis," Brain Research Reviews, vol. 29, no. 2-3, pp. 169-195, 1999.

[9] K. G. Klein, S. Frenkel, O. Shubi, and Y. Berzak, "Eye tracking based cognitive evaluation of automatic readability assessment measures," arXiv preprint arXiv:2502.11150, 2025.

시맨틱 통신과 생성형 인공지능 융합에 관한 최신 연구 동향

이승찬, 이동현, 송치현, 김재민, 조성래
중앙대학교 컴퓨터공학과

{sclee, dhlee, chsong, jmkim}@uclab.re.kr, srcho@cau.ac.kr

Recent Advances in Semantic Communication and Generative AI Fusion Systems

Seungchan Lee, Donghyun Lee, Chihyun Song, Jaemin Kim and Sungrae Cho
Dept. of Computer Science and Engineering
Chung-Ang University

요 약

시맨틱 통신은 송수신 간에 데이터의 의미만을 효율적으로 전달함으로써 차세대 통신 환경의 대용량 트래픽 문제를 해결하고자 등장한 패러다임이다. 기존 통신이 비트 정확도에 초점을 맞춘 반면 시맨틱 통신은 의미 전달을 중시하나, 현재의 딥러닝 기반 구현에는 한정된 일반화 성능과 낮은 문맥 이해 능력 등의 한계가 있다. 최근 부상한 생성형 인공지능 기술(대규모 언어 모델, 멀티모달 생성 모델 등)은 사람 수준의 맥락 추론과 콘텐츠 생성 능력을 갖추고 있어 이러한 한계를 보완할 수 있는 대안으로 주목받고 있다. 본 논문은 시맨틱 통신과 생성형 AI 의 융합에 관한 최신 연구 동향을 서베이하고, 기술 융합으로 인한 통신 방식의 변화, 주요 응용 사례, 그리고 향후 해결해야 할 과제를 논의한다.

Ⅰ. 서 론

초연결·초지능화를 지향하는 6G 시대를 앞두고, 시맨틱 통신은 단순한 비트 정확도보다 데이터의 의미 전달을 우선시하는 새로운 패러다임으로 주목받고 있다[1][2]. 송신단은 원본 데이터에서 불필요한 정보를 제거하고 핵심 의미만 인코딩하며, 수신단은 공유된 지식을 활용해 의미를 복원한다. 이를 통해 일정 수준의 오류가 발생하더라도 의미 이해에는 문제가 없으며, 전송 효율을 극대화할 수 있다. 그러나 현존하는 딥러닝 기반 시맨틱 통신은 특정 작업에 최적화되어 있어 일반화 성능과 문맥 추론 능력이 부족하다는 한계가 있다[3]. 또한 방대한 사전 지식 데이터 구축이 필수적이라는 점에서 비용과 시간이 과다하게 소요된다. 이러한 한계 극복을 위해 최근 생성형 인공지능(Generative AI)이 주목받고 있으며, 대규모 언어 모델(LLM)과 멀티모달 생성 모델은 인간 수준의 문맥 이해와 콘텐츠 생성 능력을 통해 시맨틱 통신 발전에 새로운 전기를 마련하고 있다.

Ⅱ. 본론

생성형 AI 융합에 따른 시맨틱 통신의 진화: 생성형 AI 를 시맨틱 통신에 활용함으로써 얻을 수 있는 대표적 이점은 다음과 같다. 첫째, 대규모 생성 AI 모델을 통해 방대한 학습 데이터와 배경 지식을 자동 생성함으로써 시맨틱 인코더/디코더의 훈련을 효율화할 수 있다[4]. 실제로 생성 AI 가 만든 다양한 AI 생성 콘텐츠(AIGC)를 시맨틱 통신 모델 학습 및 지식 베이스 구축에 활용하면, 별도의 데이터 수집 없이도 모델의 성능을 향상시킬 수 있다. 둘째, 생성 모델의 고도화된 문맥 추론 능력은 시맨틱 복원 정확도를 높여준다. 예를 들어 프롬프트 엔지니어링 기법을 통해 생성 모델이 맥락적으로 정확한 콘텐츠를 만들어내도록 함으로써, 전송 중 발생하는 잡음이나 오류에도 의미 왜곡이 최소화되도록 할 수 있다. 셋째, 프롬프트 전달을 통한 정보 재생성이라는 새로운 통신 방식이 가능해진다. 잘 학습된 생성형 모델을 송수신단에 갖춘 경우, 송신단은 원본 데이터를 요약한 간단한 프롬프트만을 전송하고 수신단은 이를 기반으로 원본과 동등한 품질의 데이터를 현지에서 생성할 수 있다. 이를 통해 모든 원본 정보를 보내지 않고도 수신 측에서 필요한 내용을 복원하거나 생성하게 되어 전송 데이터량을 획기적으로 줄일 수 있다. [5]에서는 이러한 아이디어를 구체화하기 위해 정보 보틀넥 이론과 확률적 생성 모델을 접목한 시맨틱 통신 기법을 제안하였는데, 채널 상태에 따라 전송할 정보량을 조절하고 수신단에서 원본 데이터를 재생성함으로써 주어진 전송 지연 및 전력 예산 내에서 통신 효율을 극대화할 수 있다고 보고하였다. 이처럼 생성형 AI 의 도입은 시맨틱 통신의 전송 효율 향상, 의미 전달 정확도 제고, 망 자원 절감 등으로 기존 한계를 극복할 수 있는 새로운 가능성을 열어주고 있다.

융합 시스템의 주요 응용 사례: 시맨틱 통신과 생성형 AI 의 결합은 다양한 분야의 응용에서 혁신적인 성능 향상을 가져올 것으로 기대된다. [6]에서는 생성형 AI 기반 Generative SemCom 시스템을 설계하고, 대용량 영상 데이터 전송 시 실제 효과를 평가하였다. 그 결과, 제안된 LLM 기반 생성형 통신 기법을 활용하면 전통적 통신 대비 전송해야 할 비트 수를 99.98%까지 감소시키면서도 정보 검색 정확도는 53% 향상시킬 수 있음이 보고되었다. 이는 수신기에서 의미를 재생성(regeneration)함으로써 통신 부하를 획기적으로

줄인 대표적인 사례이다. 이러한 생성형 시맨틱 통신은 산업 IoT 의 센서 데이터 전달, 자율주행 차량(V2X) 간 주변 상황 공유, 메타버스 환경에서의 몰입형 콘텐츠 전송, 드론과 같은 무인이동체 영상 스트리밍 등 다양한 영역에 적용될 수 있을 것으로 전망된다. 예를 들어, 자율주행차 네트워크에서는 차량에 장착된 로컬 LLM 기반 에이전트가 각종 센서 데이터를 실시간 해석하여 도로 상황을 요약 전송하고, 수신 차량은 이를 토대로 필요한 경고 또는 제어 신호를 자체 생성함으로써 초저지연 협력 운행을 구현할 수 있다[7]. 메타버스 기기 간 통신의 경우에도 사용자 동작을 의미 묘사로 추출하여 주고받고, 수신 단말이 해당 묘사에 따라 가상 아바타나 환경을 즉각 생성함으로써 대역폭 부담 없이도 실시간 상호작용을 가능케 한다. 이처럼 여러 응용 시나리오에서 시맨틱 통신+생성형 AI 융합은 적은 전송으로 풍부한 결과를 얻는 새로운 통신 서비스를 제공할 수 있다.

향후 연구 과제: 생성형 AI 와의 융합을 현실의 통신망에 적용하기 위해서는 해결되어야 할 기술적 과제들도 남아 있다. 첫째로, GPT-3 와 같은 초거대 모델은 수십억 이상의 파라미터로 이루어져 있어 동작에 막대한 연산 자원과 저장 공간을 필요로 하므로 이러한 모델을 엣지 단말에서 효율적으로 구동하기 위한 경량화 및 분산 처리 기술이 요구된다. 둘째로, 생성 AI 출력의 신뢰성에 관한 문제이다. 생성형 모델이 산출하는 콘텐츠는 자동 생성의 특성상 일정 수준의 불확실성을 내포하며, 추가적인 지연을 발생시킬 우려도 있다. 통신 맥락에서 이러한 불확실성과 지연은 신뢰성과 지연 민감도 측면의 새로운 도전과제가 되므로, 생성 모델의 출력 품질 보장 및 실시간 처리에 대한 연구가 필요하다. 셋째로, 프라이버시 및 보안 이슈가 있다. 시맨틱 통신은 송수신단이 배경 지식을 공유하는 것이 전제되는데, 개인 데이터나 선호도 등의 민감 정보를 활용하면서도 프라이버시를 보호할 수 있는 방안이 모색되어야 한다. 아울러 사용자들이 자신의 데이터를 학습에 제공하거나 배경지식으로 공유하려 하지 않을 경우 시스템 성능이 제한될 수 있으므로, 인센티브 메커니즘 등의 정책적 지원 방안도 고려되어야 한다. 이외에도 생성 모델과 통신 시스템 간의 표준 인터페이스 정립, 오류 발생 시 의미 왜곡을 방지하는 향상된 부호화 기법, 그리고 평가 기준 마련 등 다양한 후속 연구가 제기되고 있다.

Ⅲ. 결론

본 논문에서는 차세대 통신 패러다임으로 주목받는 시맨틱 통신과 생성형 인공지능 기술의 융합에 대한 연구 동향을 살펴보았다. 시맨틱 통신은 의미 기반의 효율적 정보 전달을 가능하게 하나, 현존 기법들의 한계로 인해 추가적인 지능화가 요구된다. 이러한 배경에서 등장한 생성형 AI 는 방대한 데이터 학습으로 얻은 생성 능력과 추론 능력을 바탕으로 시맨틱 통신의 성능 향상과 새로운 서비스 구현에 기여하고 있다. 여러 응용 분야에서 의미 있는 성과가 보고되고 있으며, 남아 있는 도전 과제들에 대해서도 활발한 연구가 진행 중이다. 향후 생성형 AI 와 통신기술의 긴밀한 융합을 통해 고효율·지능형 통신 시대가 앞당겨질 것으로 기대된다.

ACKNOWLEDGMENT

This work was supported by the IITP (Institute of Information & Communications Technology Planning & Evaluation) - ITRC (Information Technology Research Center) (IITP-2025-RS-2022-00156353, 50% / IITP-2025-RS-2024-00436887, 50%) grants funded by the Korea government (Ministry of Science and ICT).

참 고 문 헌

[1] S. Park, C. Park and J. Kim, "Learning-Based Cooperative Mobility Control for Autonomous Drone-Delivery," in *IEEE Transactions on Vehicular Technology*, vol. 73, no. 4, pp. 4870-4885, April 2024.

[2] W. J. Yun, D. Kwon, M. Choi, J. Kim, G. Caire and A. F. Molisch, "Quality-Aware Deep Reinforcement Learning for Streaming in Infrastructure-Assisted Connected Vehicles," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 2, pp. 2002-2017, Feb. 2022.

[3] S. Park, J. P. Kim, C. Park, S. Jung and J. Kim, "Quantum Multi-Agent Reinforcement Learning for Autonomous Mobility Cooperation," *IEEE Communications Magazine*, vol. 62, no. 6, pp. 106-112, June 2024.

[4] L. Xia et al., "Generative AI for Semantic Communication: Architecture, Challenges, and Outlook," *arXiv* preprint arXiv:2308.15483, 2024.

[5] S. Barbarossa *et al.*, "Semantic Communications Based on Adaptive Generative Models and Information Bottleneck," *IEEE Communications Magazine*, vol. 61, no. 11, pp. 36- 41, 2023.

[6] J. Ren *et al.*, "Generative Semantic Communication: Architectures, Technologies, and Applications," *arXiv* preprint arXiv:2412.08642, 2024.

[7] Z. Wang *et al.*, "Large-Language-Model-Enabled Text Semantic Communication Systems," *Applied Sciences*, vol. 15, no. 13, Article 7227, 2025.

두피 이미지 분석을 위한 다중 라벨, 다중 클래스 분류 모델에 대한 예비 연구

유현선, 고동환, 이세영, 김관*

연세대학교 의공학과, *연세대학교 AI융합과학원

dbgustjs0106@naver.com, donghwan22@yonsei.ac.kr, slee0895@yonsei.ac.kr, *kimkwan@yonsei.ac.kr

A Preliminary Study on a Multi-label, Multi-class Classification Model for Scalp Image Analysis

Yu Hyunseon, Ko Dong Hwan, Lee Se Young, Kim Kwan*

Department of Biomedical Engineering, Yonsei Univ.,

*Institute of AI Convergence Science, Yonsei Univ.

요 약

본 연구는 단일 두피 영상에서 미세각질, 피지과다, 모낭사이홍반, 모낭홍반/농포, 비듬, 탈모 등 6개 지표의 질환 수준(양호:0 ~ 심각:3)을 동시에 추정할 수 있는 다중 라벨·다중 클래스 분류 모델을 구현하는데 목적이 있다. 640×480 해상도의 두피 이미지 및 질환 수준 정보가 포함된 라벨링 파일 등 각각 17,785개를 사용하며 데이터 학습, 점검, 평가를 위해 전체 데이터를70:15:15 비율로 구분했다. ImageNet에서 사전학습된 EfficientNet-B4를 공유 백본으로 사용했으며, 질환 수준 별 4-class softmax 헤드 6개를 연결하였다. 이미지는 비율을 유지한 채 448×448로 정규화하고, AdamW를 최적화 함수로 사용했으며 학습률 0.0001, 배치 크기 16, 최대 30 epoch로 설정했다. 평가 데이터셋에서 overall macro-F1=0.518, Hamming accuracy=0.658을 확인했으며, 지표 별 macro-F1은 미세각질 0.458, 피지과다 0.517, 모낭사이홍반 0.661, 모낭홍반/농포 0.311, 비듬 0.624, 탈모 0.526으로 확인되었다. 제안한 모델은 한 장의 영상으로 6개 지표를 동시 추정할 수 있는 가능성을 보였으나, 지표별 성능 차이는 해당 질환의 영상학적 특징과 연관된 것으로 판단된다. 향후 모델·세부 파라미터 비교 실험을 통해 성능을 개선하고, 실시간 수준의 추정 가능성을 검토할 예정이다.

Ⅰ. 서 론

건선, 탈모 등의 모발 및 두피 질환은 개인의 건강과 웰빙에 중대한 영향을 미칠 수 있어 이를 조기에 발견하고 이를 예방, 관리하는 것이 중요하다[1]. 헬스케어 산업 및 기술의 발달과 함께, 두피 영상 장치를 이용한 질환 분류 영상 처리 기법이 연구되고 있으며, 최근에는 CNN 기반 자동 분류·검출이 도입되어 진단 성능을 개선하는 연구 또한 진행되고 있다[1,2].

하지만, 두피 질환은 단일 증상만 존재하는 것이 아닌 여러 증상이 복합적으로 존재하는 경우가 많아 임상 환경을 고려한 두피 질환 분류 알고리즘을 개발하는 것이 중요하다.

본 연구는 단일 또는 다중 질환이 복합적으로 존재하는 두피 이미지에서 여러 질환 지표의 수준을 한 번에 예측하기 위한 다중 라벨, 다중 클래스 분류 모델을 설계하는 것을 목표로 한다.

Ⅱ. 본론

본 연구에서 활용한 데이터셋은 640×480 해상도의 두피 이미지와, 두피 질환 관련 6개 지표(미세각질, 피지과다, 모낭사이홍반, 모낭홍반/농포, 비듬, 탈모)에 대해 0(정상)부터 3(심각) 까지의 수치가 표기된 라벨링 파일로 구성된다. 이미지 및 라벨링 파일은 각각 17,785개를 사용했으며, 모델 학습 (Train), 파라미터 점검 (Validation), 그리고 성능 평가 (Test)를 위해 전체 데이터를 70:15:15 비율로 구분했다.

학습은 파이토치 기반으로 구현되었다. ImageNet에서 사전학습된 EfficientNet-B4를 공유 백본으로 사용하고, 그 위에 6개 독립 헤드(각 4-class softmax)를 연결하여 한 장의 영상에서 6개 지표의 등급을 동시에 추론하도록 설계했다. 전처리는 이미지 원본 비율을 유지하면서 축소 및 패딩 과정을 거쳐 448×448 로 정규화하여 이미지 내 텍스쳐 정보를 유지하면서 연산량을 감소시켰다. 최적화는 AdamW를 사용, 학습률은 0.0001로 고정했으며, 배치 크기는 16으로 설정했다. 최대 30 epoch까지 학습하되, 검증 macro-F1이 5 epoch 이상 개선되지 않으면 학습을 조기 종료하여 불필요한 과적합과 계산을 방지했다..

클래스 불균형과 샘플 편향을 완화하기 위해 Focal loss를 사용하였고, γ=2.0으로 고정하였다. 학습 시 수평 반전, 밝기, 대조도, 색조, 채도, 회전, 스케일 조절 등 데이터 증강을 적용했다.

표 1. 데이터 라벨 분포

총 데이터 수 (=17,785)		질환 수준			
		양호 (0)	경증 (1)	중증 (2)	심각 (3)
질환 지표	미세각질	14758	1125	1345	557
	피지과다	3678	6879	6216	1012
	모낭사이홍반	5916	7247	3411	1211
	모낭홍반/농포	16963	552	196	74
	비듬	10585	4113	2483	604
	탈모	13003	3435	1075	272

모델의 성능 평가를 위한 지표는 6개의 질환 지표 각각, 0~3까지 4개의 질환 수준을 고려하여 TP (True Positive), FP (False Positive), FN (False Negative), TN (True Negative)를 구한 뒤 지표 별 macro-F1과 이를 평균계산한 overall macro-F1을 계산했다. 또한, 이미지 내 6개 지표 중 몇 개 지표에 대해 정확히 예측했는지 확인하기 위해 Hamming

Accuracy를 계산했다.

표 2는 각 질환 별 macro-F1 및 overall macro-F1를 정리한 것으로, 평가 데이터셋의 경우 overall macro-F1은 0.518, Hamming accuracy는 0.658로 확인되었다. 각 질환 지표 별 macro-F1은 미세각질 0.458, 피지 0.517, 모낭사이 홍반 0.661, 모낭 홍반/농포 0.311, 비듬 0.624, 탈모 0.526으로, 모낭사이 홍반과 비듬에서 상대적으로 높은 macro-F1을 보였고 모낭 홍반/농포가 가장 낮았다.

$$\text{Precision}_k = \frac{TP_k}{TP_k + FP_k}, \quad \text{Recall}_k = \frac{TP_k}{TP_k + FN_k}$$

$$\text{F1}_k = \frac{2 \cdot \text{Precision}_k \cdot \text{Recall}_k}{\text{Precision}_k + \text{Recall}_k}$$

$$\text{macro-F1}_l = \frac{1}{K} \sum_{k=1}^{K} \text{F1}_{l,k}$$

$$\text{overall macro-F1} = \frac{1}{L} \sum_{l=1}^{L} \left(\frac{1}{K} \sum_{k=1}^{K} \text{F1}_{l,k} \right), \quad L=6,\ K=4.$$

$$\text{Hamming Accuracy} = \frac{1}{NL} \sum_{i=1}^{N} \sum_{l=1}^{L} \mathbf{1}\{\hat{y}_{i,l} = y_{i,l}\}$$

그림 1. macro-F1 등 평가 지표 계산 방법

표 2. 질환 지표 별 macro-F1 및 overall macro-F1

기준	점검 (Validation)	평가 (Test)
미세각질	0.490	0.458
피지과다	0.506	0.517
모낭사이홍반	0.679	0.661
모낭홍반/농포	0.343	0.311
비듬	0.652	0.624
탈모	0.491	0.526
평균	0.527	0.518

그림 2은 각 질환 지표 별 Confusion matrix로, 질환 별 각 수준에 대해 예측된 비율을 표현한 것이다. 대각선 집중된 형태가 나타나지만 질환 별 차이가 관찰되었다. macro-F1이 상대적으로 높은 모낭사이홍반과 비듬 지표는 비율 추세가 비교적 안정적이지만, 모낭홍반/농포는 심각(3) 수준을 중증(2) 수준으로 과소추정하는 패턴이 관찰되었다.

학습 과정 중, 점검 데이터셋에 대한 macro-F1은 초기 6epoch 부근까지 상승한 뒤 서서히 줄어드는 추세를 보였다 (그림 3). 학습에 대한 loss는 epoch가 반복됨에 따라 지속적으로 줄어들었으나, 점검에 대한 loss는 상승하는 추세를 보였다. 이는 전형적인 과적합 양상으로, 개선을 위해 학습 모델 관련 파라미터의 최적화 과정이 필요하다.

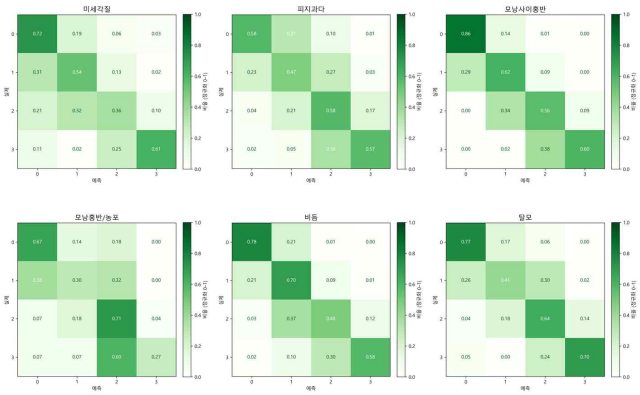


그림 2. 질환 별 예측 비율 (Confusion matrix)

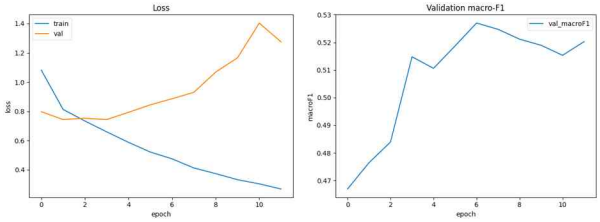


그림 3. 질환 별 예측 비율 (Confusion matrix)

Ⅲ. 결론

본 연구는 두피 이미지를 통해 다양한 두피 질환의 수준을 복합적으로, 동시에 추정하는 것을 목표로 한다. 질환 별 macro-F1가 차이가 나는 것은 각 질환 별 영상학적 특징과 관련이 있을 수 있다. 모낭사이홍반은 두피 내 모세혈관 확장 등으로 적색 성분이 증가하며, 비듬은 기존 피부와 비교하여 흰색 수준의 차이가 나타날 수 있다. 모낭홍반/농포의 경우 두피의 홍반 뿐만 아니라 염증까지 고려된 것으로, 질환 수준을 구별하기에 상대적으로 복잡할 수 있다. 향후 모델 별, 세부 파라미터 별 학습 및 성능 비교를 통해 성능 개선 및 모델의 실시간 수준의 질환 추정 가능성을 검토할 예정이다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 2025년도 의료AI반도체 교육 · 연구 플랫폼 구축사업의 결과로 수행되었음 (2024-0-0097).

참 고 문 헌

[1] B. M. K. Shetty et al., "Hair and Scalp Inspection System Using Deep Learning Based Analysis," 2025 International Conference on Knowledge Engineering and Communication Systems (ICKECS), pp. 1-9, 2025.

[2] L. Bi et al., "Multi-Label classification of multi-modality skin lesion via hyper-connected convolutional neural network," Pattern Recognition, 107, 107502, 2020.

CLIP 기반 의미론적 지도를 활용한 SAR-RGB 이미지 변환 기법 연구

전지훈, 길태영, 김경태, 김선옥*, 정재훈*
한국항공대학교

wjswlgns1711@kau.kr, gilman2315@kau.kr, kt143056@naver.com, *sunok.kim@kau.ac.kr, *jhjung@kau.ac.kr

A Study on SAR-to-RGB Image Translation Using CLIP-Based Semantic Guidance

Jihun Jeon, Taeyoung Kil, Kyoungtae Kim, Sunok Kim*, Jay Hoon Jung*
Korea Aerospace Univ.

요 약

SAR(Synthetic Aperture Radar, 합성 개구 레이더) 이미지는 구름, 빛, 날씨와 무관하게 안정적으로 촬영이 가능하지만, SAR 이미지의 특성으로 인해 인간이 시각적으로 이해하기 어렵다. 이에 따라 SAR 이미지를 시각적으로 이해하기 쉬운 RGB 이미지로 변환하는 기술은 원격 탐사 및 지리 정보 시스템(GIS) 응용에서 점점 더 중요해지고 있다. 본 논문에서는 기존 Pix2Pix 기반 SAR-to-RGB 변환 구조에 CLIP(Contrastive Language-Image Pretraining) 기반 의미 지도(semantic guidance)를 통합하여, 변환 결과의 의미 정보와 시각적 품질을 동시에 향상시키는 방법을 제안한다. 제안된 모델은 생성된 RGB 이미지와 실제 RGB 이미지 간의 의미 임베딩 유사도를 최대화하는 방향으로 학습하며, 이를 통해 모델이 단순한 픽셀 정합성 이상을 학습한다. 실험 결과, 제안한 기법은 기존 Pix2Pix 대비 SSIM, PSNR, FID 모두에서 향상되었으며, 의미 기반 손실이 SAR-to-RGB 변환에 효과적으로 작용함을 보여준다.

Ⅰ. 서 론

최근 기후 변화, 재해 대응, 국방 감시 등의 목적으로 위성 기반 지구 관측 수요가 증가하고 있으며, 이 중 SAR 이미지는 야간, 악천후 등에서도 안정적인 촬영이 가능하다는 점에서 주목받고 있다. 그러나 SAR 이미지는 레이더 반사 특성을 기반으로 하여, 시각적으로 인간이 인식하기 어려운 텍스처와 노이즈 패턴을 포함하고 있다.[1]

이를 해결하기 위해 SAR 이미지를 RGB 이미지 형태로 변환하는 기술이 제안되고 있으며, Pix2Pix[2], Cycle GAN과 같은 GAN 기반 이미지 변환 모델이 사용되고 있다. 그러나 이러한 모델들은 보통 픽셀 수준의 정합성에 초점을 맞추고 있어, 변환된 이미지가 실제 SAR 이미지의 의미 정보를 제대로 반영하지 못하는 한계가 있다.[3]

본 논문에서는 의미론적 정보 표현에 강점을 가진 CLIP[4]을 활용하여, 생성된 RGB 이미지가 SAR 이미지의 실제 내용과 의미적으로도 일치하도록 학습하는 모델을 제안한다.

Ⅱ. 본 론

2.1 데이터 셋과 전처리

본 연구는 SEN12MS 데이터셋을 기반으로 수행하였다. Sentinel-1의 2채널 SAR 이미지와 Sentinel-2의 13채널 광학 이미지 44,000쌍을 사용하였다. SAR 이미지의 경우, 시각화에 유의미한 실수 부에 해당하는 채널만 사용하였으며, 광학 이미지에서는 RGB에 해당하는 3개의 채널만 사용하였고, 각 이미지는 256×256으로 정규화하였다. 전체 쌍 중 70%는 학습, 20%는 검증에 사용하였다. 나머지 10%를 테스트에 사용하여, 일반화 성능을 평가하였다.

2.2 모델 정의

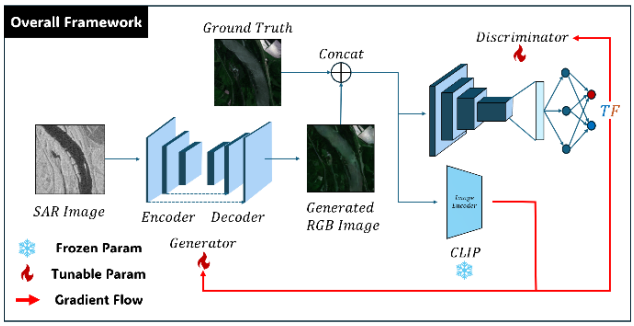


그림 1. CLIP 기반 SAR2RGB 전체 아키텍처

기본 모델은 Pix2Pix 구조를 따른다. 생성기(Generator)는 U-Net 기반의 인코더, 디코더 구조로, SAR 이미지를 입력 받아 대응되는 RGB 이미지를 생성한다. 인코더는 특징 표현을 압축하여 장면의 전역적 구조와 추상적 의미를 학습한다. 디코더는 인코더의 각 단계에서 추출된 특징 맵을 skip connection으로 결합하여 세부적인 구조 정보와 지역적 질감을 보존한다.

판별기(Discriminator)는 Pix2Pix의 PatchGAN 구조를 따르며, PatchGAN은 전체 이미지를 하나의 스칼라 값으로 판별하는 대신, 입력 쌍(SAR, RGB)을 작은 패치 단위로 나누어 각 패치가 실제(real)인지 생성(fake)인지 판별하는 방식이다. 이러한 접근은 이미지의 세부 영역까지 국소적인 일관성을 학습하게 하여, 전체적인 사실감과 함께 미세한 텍스처와 경계 정보까지 보존하는 효과를 낸다.

제안 기법에서는 해당 Pix2Pix 구조 위에 CLIP 기반 의미 지도를 통합한다. CLIP은 대규모 자연 이미지-텍스트 쌍으로 학습된 모델로, 장면의 의미, 물체 유형 등을 표현하는 고차원 임베딩 공간을 제공한다. 이 모델을 기

반으로 생성 결과와 참조 이미지 간 의미 임베딩 유사도를 비교함으로써, 단순한 픽셀 대응을 넘어선 의미적 유사도까지 확보하고자 한다. 생성된 RGB 이미지와 실제 RGB 이미지를 CLIP 이미지 인코더에 각각 통과시켜 임베딩을 추출한다.

최종 손실 함수 구성은 다음과 같다:

$$\mathcal{L}_{GAN} = E_{x,y}[logD(x,y)] + E_x\left[log\left(1 - D(x, G(x))\right)\right] \quad (1)$$

$$\mathcal{L}_1 = E_{x,y}[\| G(x) - y \|_1] \quad (2)$$

$$\mathcal{L}_{CLIP} = 1 - cos\left(f_{CLIP}(G(x)), f_{CLIP}(y)\right) \quad (3)$$

$$\mathcal{L}_{total} = \mathcal{L}_{GAN} + \lambda_1\mathcal{L}_1 + \lambda_{CLIP}\mathcal{L}_{CLIP} \quad (4)$$

본 연구의 최종 학습 손실 함수는 $\mathcal{L}_{GAN}$, $\mathcal{L}_1$, $\mathcal{L}_{CLIP}$의 3가지 항으로 구성된다. $\mathcal{L}_{GAN}$ 는 생성된 이미지가 실제 RGB 이미지와 구분되지 않도록 유도하는 minimax loss로서, 시각적 사실성을 확보하고, $\mathcal{L}_1$는 생성 이미지와 실제 RGB 이미지 사이의 $\mathcal{L}_1 Norm$으로서, 색상과 구조적 정합성을 유지한다.

여기에 본 연구의 핵심인 $\mathcal{L}_{CLIP}$를 추가하여, 생성/실제 이미지에 대해 CLIP 이미지 인코더로 추출한 임베딩 벡터 간 코사인 유사도를 최대화함으로써 생성된 RGB 이미지가 실제 RGB 이미지와 의미적으로 일치하도록 학습한다. 최종적으로, $\mathcal{L}_{total}$은 위 3항의 선형 결합으로 정의되며, 각 항의 기여도는 λ로 조절된다. 여기서 λ는 손실의 가중치를 조절하는 하이퍼파라미터이며, 과도하게 설정될 경우 픽셀 정합성이 저하될 수 있어 실험을 통해 적절한 값을 도출하였다.

2.3 결과 및 비교

표 1. 모델 별 정량적 지표

Method	FID ↓	SSIM ↑	PSNR (dB) ↑
Pix2Pix (Baseline)	131.7202	0.3102	14.95
CLIP-Guided Pix2Pix (Ours)	**113.7680**	**0.3684**	**16.71**

CLIP 기반 모델은 FID[5], SSIM[6], PSNR[7] 지표 모두에서 기존 Pix2Pix에 대비하여 향상되었다. 지표 중 SSIM은 구조적 유사도를, FID는 이미지 분포의 현실성을 반영하는 지표로, 의미 손실이 시각적 일관성을 높이는 데 기여하였음을 시사한다.

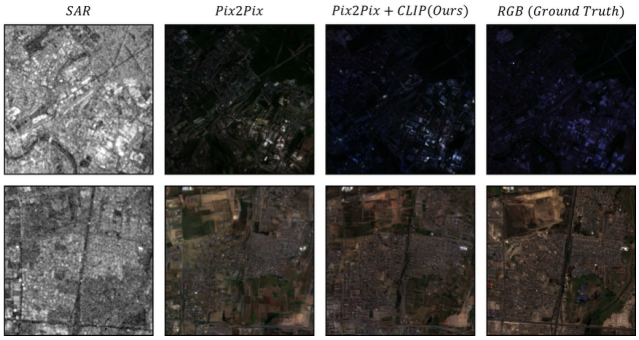


그림 2. 모델 별 정성적 지표

그림 2를 통해 CLIP Loss의 도입은 변환 모델의 시각적 유사도와 의미적 유사도를 동시에 향상시킴을 보여준다. 위와 같이 단순 Pix2Pix 기반 모델은 색감 왜곡과 세부 텍스처의 손실이 발생하는 반면, 제안한 CLIP-Guided Pix2Pix 모델은 실제 RGB 이미지와 보다 유사한 색조를 보이며, 구조적 패턴 재현율이 높음을 확인할 수 있다. 이를 통해 제안된 모델은 단순한 픽셀 별 정합도를 넘어, 시각적 품질과 의미적 유사도를 동시에 향상시키는 효과를 보였다.

CLIP이 위성/원격탐사로 사전 학습되지 않았지만, 도메인이 다르더라도 통용되는 상위 수준의 의미 정보를 제공한다. 이를 통해, 생성 이미지에 제약을 제공해 $\mathcal{L}_1$이 놓치는 형태, 구조, 장면의 일관성을 보완하였다. 그 결과, 제안한 모델은 시각적 품질과 의미적 유사도를 동시에 개선하였음을 알 수 있다.

Ⅲ. 결 론

본 논문은 CLIP 기반 의미론적 지도를 활용하여 SAR-to-RGB 변환 성능을 향상시키는 기법을 제안하였다. 생성기가 구조 및 의미 유사도를 함께 고려하도록 유도함으로써, 정량적 지표(SSIM, PSNR, FID) 뿐 아니라 정성적 지표에서도 개선을 확인하였다. 실험 결과는 제안 기법이 기본적인 GAN 기반 구조를 의미적으로 강화하는 데 효과적임을 보여주며, 향후 다양한 도메인에서의 응용 가능성을 시사한다.

ACKNOWLEDGMENT

본 과제(결과물)은 교육부와 한국연구재단의 재원으로 지원을 받아 수행된 첨단분야 혁신융합대학사업(차세대통신)의 연구 결과입니다.

참 고 문 헌

[1] D. Ao, C. O. Dumitru, G. Schwarz, and M. Datcu, “Dialectical GAN for SAR Image Translation: From Sentinel-1 to TerraSAR-X,” Remote Sensing, vol. 10, no. 10, 1597, Oct. 2018.

[2] Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A., "Image-to-Image Translation with Conditional Adversarial Networks," Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 1125-1134, Jul. 2017.

[3] 박소연, 곽근호, 황의호, 박노욱, “고해상도 광학영상 복원을 위한 딥러닝 기반 SAR-광학 영상 변환 모델 비교,” Korean Journal of Remote Sensing, vol. 40, no. 6-1, pp. 881-893, Dec. 2024.

[4] Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., and Sutskever, I., "Learning Transferable Visual Models From Natural Language Supervision," Proc. 38th Int. Conf. on Machine Learning (ICML), pp. 8748-8763, Jul. 2021.

[5] Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., and Hochreiter, S., "GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium," Proc. Advances in Neural Information Processing Systems (NeurIPS), pp. 6626-6637, Dec. 2017.

[6] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P., "Image quality assessment: From error visibility to structural similarity," IEEE Trans. on Image Processing, vol. 13, no. 4, pp. 600-612, Apr. 2004.

[7] Bovik, A. C., "Handbook of Image and Video Processing," Academic Press, pp. 859-862, 2000.

증상 검출 및 질병 분류 작업의 결합을 통한 효율적인 넙치 상태 진단을 위한 다중작업학습 접근법

김수영, 이명철*

충남대학교, *한국전자통신연구원

jellyfish319@o.cnu.ac.kr, *mclee@etri.re.kr

A Multi-Task Learning Approach for Efficient Flatfish Status Diagnosis through Joint Symptom Detection and Disease Classification

Kim Sooyoung, Myungcheol Lee*

Chungnam National Univ., *Electronics and Telecommunications Research Institute

요 약

넙치 양식 산업은 질병으로 인해 연간 막대한 경제적 손실을 겪고 있으며, 특히 출혈, 염증과 같은 증상은 육안으로 식별하기 어려워 신속하고 정확한 진단에 한계가 있다. 본 연구에서는 이러한 문제를 해결하기 위해, 단일모형으로 넙치의 증상을 탐지함과 동시에 질병을 분류하는 다중작업학습(Multi-Task Learning) 기반의 딥러닝 모델을 구현한다. 모형은 객체 탐지에 우수한 성능을 보이는 YOLO 구조를 기반으로 두 가지 작업을 동시에 처리하는 모형을 설계하고 학습시켰다. 최종적으로 얻은 다중작업모형은 정확도 측면에서 기존 단일작업모형과 대비하여 검출에서는 비슷한 성능을, 분류에서는 7% 정도의 성능 하락이 발생하였으나, 전체 필요한 모델 크기를 12% 감소하고, 추론 시간을 28% 감소하는 효과를 보였다. 본 다중작업모델을 통해 스마트폰 등 저사양 단말에 탑재함으로써 양식 현장에서 신속한 통합 질병 진단 시스템의 가능성을 제시한다.

Ⅰ. 서론

양식 현장에서는 출혈성 질병이나 염증 등 육안으로 식별이 어려운 증상으로 인해 전문가 없이 신속한 질병 진단이 어렵고, 이로 인한 넙치 폐사로 연간 약 1,300억 원의 손실이 발생하고 있다.[1]

이를 해결하기 위해 2022년 제주도에서는 넙치 질병 탐지·분류·예측을 위한 AI 모델과 데이터를 공개하였으나[2], 작업별로 개별 모델을 사용하는 기존 방식은 추론 시간이 길고, 경량 기기에서의 활용에 제약이 있다. 본 논문에서는 이러한 한계를 극복하기 위해, YOLO 모델을 백본 모델로 하여 다중작업학습(Multi-Task Learning, MTL)[3]을 적용한 하나의 경량 모델로 증상 탐지와 질병 분류를 동시에 수행하는 방법을 제안한다.

Ⅱ. 본론

가. Dataset 구성

기존 AI허브에서 제공하는 증상 탐지의 라벨링 형식은 COCO형식이었기에 백본 모형으로 사용할 YOLO 포맷으로의 추가적인 변환을 수행하였다. AI허브에서 활용한 원본 라이브러리에서는 객체 탐지와 증상 분류 각각에 대한 학습만 제공하기에, 본 연구에서는 객체 탐지 관련 데이터로더는 기존 API를 사용하였고, 증상 분류는 커스텀 로더를 만들어 YOLO에서 사용하는 ImageNet 형식의 라벨 분류가 아닌 자체적인 라벨 형식을 정의하여 사용하였다.

나. Class 병합

사용할 데이터셋의 라벨의 경우 증상은 31종, 질병은 21종이며 학습용 데이터 4만 8천개와 검증용 데이터 7천개가 존재한다. 비브리오병은 각 병원체가 일으키는 질병이나 증상의 차이간 명확히 밝혀진 부분이 없기에 하나의 클래스로 통합하였다.[4] 한 개체가 복수의 질병을 앓는 경우를 모두 포함하면 70개 이상의 질병 조합이 나오므로 복수의 질병은 새로운 카테고리를 만들어 통합하는 형식으로 클래스를 병합하였다. 증상은 31가지로 클래스 간 불균형이 심하여 주요 증상인 출혈, 부식, 안구 증상, 궤양, 종양으로 나누고 나머지 증상들은 희소하게 존재하여 기타로 분류하였다.

다. Model 구조 및 구현

본 실험에 사용한 모형은 YOLO(You Only Look Once) 모형으로 객체 탐지, 분할, 분류에 대해 좋은 성능을 보여준다. 따라서 연구의 목표인 증상 탐지, 질병 분류 두 가지 작업에 대해 좋은 성능을 보여주는 YOLO 모델을 백본 모델로 선택하였다. 기존 YOLO 모형의 백본 네트워크로부터 추출한 특징을 스케일링하는 Neck으로부터 출력값을 입력으로 하는 분류 Head를 추가로 붙이는 방식으로 모형을 설계하였다.[그림 1] 넥의 구조를 보면 feature를 여러 방식으로 scaling한 layer들이 존재한다. 최종적으로 마지막에 존재하는 layer에서의 feature를 분류 헤드에 넘겨주는 것이 제일 좋은 성능을 보여주었기에 다음과 같은 구조로 모형을 설정하였다. 다양한 성능 평가 비교를 위해 백본 네트워크로 활용한 기본 모델은 YOLO11[5], YOLOv8 두 개의 모델을 사용하였다.

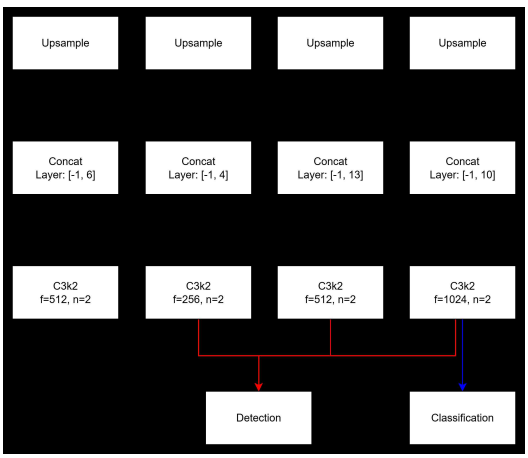


그림 1 YOLO11 MTL의 Neck과 Head 도식화 그림

YOLO 학습을 지원하는 ultralytics 라이브러리에는 다중작업학습을 따로 지원하지 않고 있기에 앞서 정의한 모형을 바탕으로 Pytorch 라이브러리와 ultralytics 라이브러리를 활용하여 구현하였다.

라. 손실함수와 평가지표

본 논문에서는 학습 효율성과 모델의 일반성을 확보하기 위해 작업별로 동일한 가중치를 부여하는 Equal Weighting 전략을 사용하였다. 본 실험은 YOLO 백본 아키텍처에 따른 MTL의 성능과 효율성 비교에 있으므로, 가장 보편적이고 표준적인 Equal Weighting을 통해 손실함수라는 변수를 통제하여 오로지 아키텍처의 영향을 분석하고자 하였다.

$$L_{\text{total}} = L_{\text{det}} + L_{\text{cls}}$$

평가지표는 Recall, Precision, mAP50, mAP95, F1 Score, Top1 Accuracy, Top5 Accuracy, 모델 크기, 추론 시간을 측정하여 고려하기로 하였다.

마. 실험 결과

YOLO 프레임워크가 기본적으로 제공하는 단일작업학습과 YOLO를 기반으로 하는 다양한 백본 네트워크를 바탕으로 다중작업학습을 진행한 결과이다.

평가 지표 / 모델	Detection				Classification	
	mAP 50	mAP 50-95	Recall	Precision	Top1 Acc	Top5 Acc
yolo11n-det (merged)	0.389	0.225	0.393	0.489	-	-
yolo11n-det (original)	0.244	0.152	0.261	0.394	-	-
yolo11n-cls (merged)	-	-	-	-	0.727	0.947
yolo11n-cls (original)	-	-	-	-	0.693	0.886

표 1 Class 통합에 따른 YOLO 단일작업학습 실험 결과

평가 지표 / 모델	Detection				Classification		
	mAP 50	mAP 50-95	Recall	Precision	F1 Score	Top1 Acc	Top5 Acc
yolo11n -mtl	0.385	0.214	0.388	**0.514**	0.414	0.651	**0.924**
yolo11s -mtl	0.377	0.211	0.387	0.49	**0.418**	**0.657**	0.92
yolo8n -mtl	0.372	0.207	0.391	0.476	0.381	0.622	0.923
yolo8s -mtl	**0.411**	**0.238**	**0.408**	0.513	0.204	0.489	0.915

표 2 YOLO 다중작업학습 실험 결과

평가 지표 / 모델	Inference Time	Size
yolo11n - det	0.4ms	2.6M
yolo11n - cls	0.3ms	1.55M
yolo11n - mtl	0.5ms	3.65M

표 3 YOLO11n의 작업 별 추론 시간과 모델 크기

실험 결과를 해석해보면 [표 1]은 앞서 설정하였던 Class 병합의 효과를 보여준다. Class 병합을 수행하지 않은 데이터셋보다 병합을 한 데이터셋에서 각각의 단일작업에 대한 성능이 높음을 볼 수 있다. [표 2]에서는 이렇게 병합한 데이터셋을 바탕으로 여러 모형에 대한 다중작업학습을 진행하였을 때의 결과이다. 탐지에 대한 성능은 yolo8s가 가장 높으며 yolo11 모델이 질병 분류 작업에 더 좋은 성능을 냄을 확인할 수 있다. MTL 환경에서 모델 아키텍처에 따라 탐지와 분류 성능 간 Trade-off가 발생하며 yolo8s는 탐지에, yolo11n/s는 분류에 더 강점을 보인다. yolo8s의 경우 탐지 성능은 제일 좋으나 분류의 성능은 굉장히 떨어지기에 전반적인 성능을 고려해보면 yolo11 구조가 더 효율적이다. yolo11n의 증상 탐지 성능은 yolo11s 모델보다 더 좋으며 질병 분류 부분에서는 두 모델 간 성능 차이가 거의 없기에 최종적으로 모델의 크기가 더 작고 탐지에서 더 좋은 성능을 보여주는 yolo11n을 최적의 모델로 선정하였다.

yolo11n과 단일작업모델 간 성능 비교를 해보면, 객체 탐지 부분에서는 mAP 점수는 대등한 수준이며, 정밀도(Precision)는 MTL 모델이 더 높게 나왔다. 이는 MTL 모델이 다른 작업(분류)을 함께 학습하면서 공유 표현(Shared Representation)를 학습하였다는 점을 시사한다. 분류 성능에서는 단일작업모델이 우세하다. Top-1 Accuracy에서 약 7.6%p의 차이가 발생하며, 이는 모델이 탐지와 분류라는 두 가지 작업을 동시에 처리하도록 학습되면서 분류 작업에 대한 학습이 다소 희석된다고 해석된다. 마지막으로 MTL 모델의 추론 시간은 탐지 모델과 분류 모델을 각각 실행한 시간(0.7ms)보다 28% 더 빠른 0.5ms가 걸리며, 모델 크기 역시 각 모델을 합한 것(4.15M)보다 12% 더 작은 3.65M개의 파라미터를 갖는다.[표 3]

Ⅲ. 결론

본 연구에서는 단일모델로 객체 탐지와 질병 분류를 동시에 수행하는 다중작업학습 모델을 구축하고 그 효율성을 검증하였다. 이를 위해, 분류와 탐지를 동시에 수행하기 위한 커스텀 데이터 로더를 만들었고, 분류 헤드를 백본 네트워크에 최적으로 연결하는 방안을 고려하여 설계하였으며, 각 태스크의 학습 효율성과 일반성을 확보할 수 있는 손실함수를 사용하였다. 여러 YOLO 아키텍처를 비교 분석한 결과, yolo11n 구조가 탐지와 분류 성능 간의 균형이 제일 뛰어나기에 최적 모델로 선정하였다.

최종적으로 제시하는 yolo11n-mtl 모델은 전문 탐지 모델과 대등한 수준의 탐지 성능을 유지하면서도, 두 개의 단일모델을 사용하는 것보다 추론 속도는 빠르고 모델 크기는 작은 효율성을 보였다. 비록 분류 성능에서는 약간의 성능 저하가 발생했으나, 이는 효율성을 극대화하는 과정에서 발생하는 합리적인 트레이드오프(Trade-off)로 판단된다.

결론적으로, 본 연구에서 구현한 MTL 모델은 제한된 하드웨어 리소스 환경에서 성능 저하를 최소화하며 시스템의 효율성을 극대화할 수 있는 실용적인 해결책임을 입증하였다. 이는 실시간 처리가 요구되는 다양한 산업 현장에 효과적으로 적용될 수 있을 것으로 기대된다.

ACKNOWLEDGMENT

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2021-0-00136, 다양한 산업 분야 활용성 증대를 위한 대규모/대용량 블록체인 데이터 고확장성 분산 저장 기술 개발)

참 고 문 헌

[1] https://www.boka.co.kr/news/articleView.html?idxno=4437

[2] AI Hub - 넙치 질병 데이터 https://aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&dataSetSn=71345

[3] Caruana, R. (1997). “Multi-task learning”. 《Machine Learning》 28: 41 - 75

[4] Sohn, H., Kim, J., Jin, C. et al. Identification of Vibrio species isolated from cultured olive flounder (Paralichthys olivaceus) in Jeju Island, South Korea. Fish Aquatic Sci 22, 14 (2019). Standard(AES),"FIPS PUB ZZZ, 2001

[5] Rahima Khanam, Muhammad Hussain, arXiv:2410.17725v1 [cs.CV] 23 Oct 2024

시계열 이상 탐지 성능 향상을 위한 마할라노비스 기반 점수화 기법 제안

김파란하늘, 한재웅, 김현정, 이연서, 현장훈*
국립한밭대학교
20231050@edu.hanbat.ac.kr, jhhyeon@hanbat.ac.kr*

A Mahalanobis-based Scoring Method for Improving Time Series Anomaly Detection Performance

Paranhaneul Kim, Jaewoong Han, Hyeonjeong Kim, Yeonseo Lee, Janghun Hyeon*
Hanbat National University

요 약

본 논문은 에너지 저장 장치(ESS) 이상 탐지 성능을 높이기 위해 동일한 주성분 분석(PCA) 모델에서 재구성 오차 방식과 Hotelling T^2이상 점수 산출 방식의 성능을 비교한다. 이는 기존 재구성 오차 방식이 데이터의 통계적 분포를 반영하지 못하는 한계를 갖기 때문으로, 본 논문에서는 이를 해결하기 위해 주성분 공간에서의 통계적 거리를 측정하는 Hotelling T^2를 이상 점수로 도입한 알고리즘을 제안하고 그 성능을 종합적으로 평가하였다. 그 결과, Hotelling T^2방식이 기존 재구성 오차 방식 대비 모든 평가 지표의 성능을 유의미하게 향상시켰으며, 특히 이상 탐지의 재현율과 F1-score의 성능을 각각 약 48%, 22% 향상시켰다. 이를 통해 모델의 구조 변경 없이 이상 점수 산출 방식의 개선만으로도 이상 탐지 성능을 효과적으로 높일 수 있음을 확인하였으며, 향후 다양한 이상 탐지 알고리즘 분야에 적용하여 모델의 성능을 개선할 수 있을 것으로 기대된다.

Ⅰ. 서 론

에너지 저장장치 시스템(Energy Storage System, ESS)은 전력을 저장해 두었다가 필요할 때 공급해주는 장치로써 전력 수급 안정화 및 에너지 효율 증가에 핵심적인 역할을 하지만 과충전, 열폭주 등에 의한 화재 위험이 존재한다. 이러한 안전성 문제를 해결하기 위해 화재 신호를 조기에 감지하는 이상 탐지 연구의 중요성이 크게 대두되고 있으며, 이에 선행 연구에서는 주성분 분석(Principal Component Analysis, PCA)을 활용한 재구성 오차(Reconstruction Error)기반의 이상 탐지 방식이 활용되고 있다[1]. 그러나, 이러한 방식은 주성분 공간에서 평균으로부터 멀어지는 유형의 이상에는 둔감할 수 있으며, 정상 데이터의 통계적 구조를 충분히 반영하지 못한다는 한계점이 존재한다. 본 논문에서는 이러한 한계를 보완하기 위해 주성분 공간에서 데이터의 통계적 분포를 계산하는 마할라노비스 거리 기반의 Hotelling T^2를 이상 점수로 도입한 알고리즘을 제안한다[2]. 동일한 PCA 모델과 데이터 분할을 유지한 상태에서 제안 기법과 재구성 오차 방식을 비교하여 성능을 검증하였으며, 실험 결과 제안 기법이 재구성 오차 방식 대비 우수한 성능을 보임을 확인하였다.

Ⅱ. 연구 설계

2.1 데이터 구성 및 전처리

본 연구에서 사용된 데이터는 모델 개발 및 임계값 설정을 위한 개발 데이터셋과 최종 성능 평가를 위한 테스트 데이터셋으로 구성된다. 개발 데이터셋은 총 11개의 독립적인 단위 데이터셋으로, 각 데이터셋은 약 90일간 1분 간격으로 측정된 태양광 ESS 전압 시계열 데이터이다. 이 중 8개는 정상 운영 상태에서 수집된 정상 데이터이며, 나머지 3개는 화재가 발생한 비정상 데이터로 구성된다. 테스트 데이터셋은 모델 개발 및 임계값 설정 과정에 사용되지 않은 별도의 데이터로 총 7개의 단위 데이터셋이 있으며, 4개는 정상데이터 3개는 비정상 데이터로 구성된다. 본 연구에서 사용된 태양광 ESS는 충전 및 방전 주기에 따라 전압이 규칙적으로 상승하고 하강하는 주기적 특성을 가지고 있으며, 연구에 사용된 ESS 배터리 셀의 방전 하한 전압인 3.3V 미만의 값은 센서 오류로 판단하여 결측값(NAN)으로 처리하였다. 이를 바탕으로 데이터의 전체적인 추세를 왜곡하지 않으면서 국소적인 결측을 처리하기 위해 선형보간을 진행하였다[3].

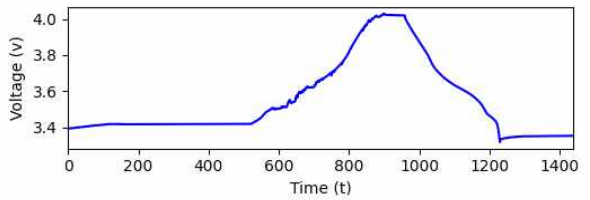


그림 1. 전처리된 ESS 전압 시계열 데이터

2.2. PCA 모델 설계 및 평가 방법

본 연구에서는 개발 데이터셋을 학습, 검증 데이터로 분할하여 모델을 설계하였다. 학습 데이터는 정상 데이터만으로 구성되어 PCA 모델을 학습시켰으며 모든 데이터는 학습 데이터의 평균과 표준편차로 표준화하여 전체 분산의 95%를 설명하는 주성분 공간으로 투영(Projection)하였다. 이후 학습된 모델을 사용하여 검증 데이터의 이상 점수 분포를 계산하고, Youden's J 통계량이 최대화하는 지점을 정상과 이상(Anomaly)클래스를 구분하는 임계값으로 결정하였다[4]. 해당 임계값을 테스트 데이터에 적용하여 최종 이상 탐지 성능을 평가하였으며, 실험의 통계적 신뢰도를 확보하고 모델의 일반화 성능을 확인하고자 k-fold 교차검증을 수행하였다.

Ⅲ. 이상 탐지 알고리즘

3.1. 재구성 오차 기반 이상 탐지(MAE)

재구성 오차 기반 접근법은 정상 데이터의 패턴을 모델이 학습하여 이상을 탐지하는 가장 일반적인 접근법이다. 본 연구에서는 검증 데이터를 입력으로 하여 PCA 모델의 주성분 공간으로 투영한 후 다시 원본 데이터의 차원으로 복원하였다. 이 과정에서 정상 데이터는 원본과 복원 데이터 간의 차이가 적게 나타나는 반면, 비정상 데이터는 모델이 학습한 패턴과 상이한 패턴으로 인하여 원본과 복원 데이터간의 차이가 크게 발생한다. 이 차이를 정량화하기 위한 방법으로 평균 절대 오차(Mean Absolute Error, MAE)를 사용하였으며 이는 식 (1)과 같이 계산된다.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|x_i - \hat{x}_i| \quad (1)$$

이를 통해 각 검증 데이터별로 MAE 값들을 계산할 수 있으며, 최종적으로 산출된 각 데이터별 MAE 분포를 통해 [그림 2]와 같이 Youden's J 통계량을 최대화하는 지점을 임계값으로 설정하였다.

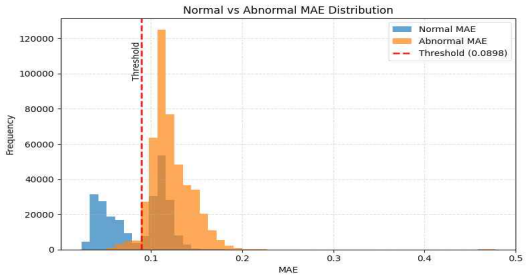


그림 2. 검증 데이터의 MAE 분포에 따른 임계값 설정

3.2 마할라노비스 거리 기반 Hotelling T^2 이상 탐지

본 연구에서 제안하는 방식은 주성분 공간 내에서 정상 데이터들이 형성하는 통계적 분포를 기준으로 입력 시계열 데이터와의 거리를 정량화한다. 이를 위해 학습에 사용된 정상 데이터를 주성분 공간으로 투영하여 정상 데이터 분포의 평균 벡터(μ)와 공분산 행렬(Σ)을 계산한다. 이후, 검증 데이터를 활용한 새로운 시계열 데이터 (x)가 입력되면, 동일한 주성분 공간으로 투영하여 벡터 z를 얻고, Hotelling T^2를 활용해 정상 데이터 분포의 중심과 벡터 z사이의 통계적 거리를 측정한다. 주성분 공간에서의 Hotelling T^2는 식 (2)와 같이 마할라노비스 거리의 제곱으로 정의된다.

$$T^2 = (z-\mu)^T \sum^{-1}(z-\mu) \quad (2)$$

측정된 T^2값은 정상 데이터의 통계적 분포로부터 이탈 정도를 나타내는 이상 점수(Anomaly Score)로 사용된다. 이를 통해 [그림 3]에서는 검증 데이터의 파란색 정상 샘플 대부분이 학습 데이터 군집 내에 분포하는 것을 확인할 수 있으나, [그림 4]에서는 검증 데이터의 주황색 비정상 샘플 다수가 정상 군집으로부터 떨어져 분포하는 경향을 확인할 수 있다. 이러한 T^2점수를 모든 검증데이터에 대해 계산하여 [그림 5]를 통해 정상 및 이상 클래스별 T^2분포를 확인하였으며, Youden's J 통계량이 최대화되는 지점을 임계값으로 설정하였다.

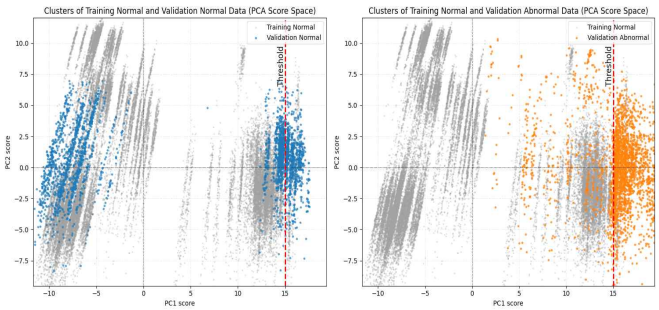


그림 3,4. 학습 데이터와 검증 데이터의 주성분 공간 T^2 군집

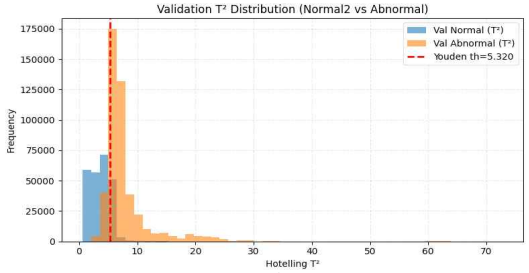


그림 5. 검증 데이터의 T^2 분포에 따른 임계값 설정

IV. 실험 결과 및 분석

일반적으로 이상 탐지 문제는 극심한 클래스 불균형(Class Imbalance)를 동반하여 단일 평가 지표만으로는 모델의 성능 평가가 어렵다는 문제점이 존재한다[5]. 이에 본 연구에서는 정밀도(Precision), 재현율(Recall), F1-score, PR-AUC, ROC-AUC의 지표를 활용해 모델의 성능을 종합적으로 평가하였으며, 특히 화재 신호를 놓치는 FN(False Negative)을 최소화하기 위해 재현율과 F1-score를 핵심 평가 지표로 간주하였다. 실험 결과, Hotelling T^2방식의 평균 재현율은 0.865로 기존 재구성 오차 방식 대비 약 48%가량 향상되었으며, F1-score 역시 약 22%가량 향상된 0.703의 성능을 기록하였다. 이는 제안 방식이 실제 화재 발생 이력이 있는 비정상 데이터를 놓치지 않고 탐지하는 능력이 크게 향상되었음을 의미하며, 이 밖에도 다른 평가 지표들 역시 전반적으로 향상된 수치를 보여 제안 방식의 안전성과 우수성을 뒷받침하였다[표1].

표 1. 점수화 기법에 따른 이상 탐지 성능 평가 결과

K-fold	MAE					Hotelling T^2				
	Precision	Recall	F1-score	ROC-AUC	PR-AUC	Precision	Recall	F1-score	ROC-AUC	PR-AUC
K = 1	0.635	0.782	0.693	0.727	0.684	0.603	0.878	0.715	0.697	0.589
K = 2	0.506	0.541	0.523	0.663	0.551	0.636	0.786	0.703	0.726	0.671
K = 3	0.587	0.599	0.593	0.722	0.648	0.618	0.852	0.716	0.722	0.663
K = 4	0.564	0.439	0.494	0.705	0.616	0.530	0.943	0.678	0.719	0.618
Avg	0.573	0.585	0.576	0.704	0.625	**0.597**	**0.865**	**0.703**	**0.716**	**0.636**

V. 결론

본 논문에서는 PCA 기반 ESS 이상 탐지 모델의 성능 향상을 위해 주성분 공간에서의 통계적 거리를 측정하는 Hotelling T^2알고리즘을 제안하고, 기존 재구성 오차 방식과 성능을 비교 검증하였다. 실험 결과, 제안 기법은 모든 평가지표에서 기존 MAE 방식을 상회하는 우수한 성능을 보였으며 특히, 실제 이상을 탐지하는 재현율이 약 48% 향상되었다. 이를 통해 모델의 구조 변경 없이 ESS 이상 탐지 성능을 높이는 효과적인 이상 점수 산출 방식을 제안하며, 향후 다양한 이상 탐지 알고리즘 분야에 적용할 수 있는 연구 방향을 제시한다.

참 고 문 헌

[1] M. Crépey, A. Aouadi, and C. Rahal, "Anomaly Detection on Financial Time Series by Principal Component Analysis and Neural Networks," Algorithms, vol. 15, no. 3, pp. 1 - 21, 2022.

[2] T. Rato, M. Reis, E. Schmitt, M. Hubert, and B. De Ketelaere, "A systematic comparison of PCA-based statistical process monitoring methods for high-dimensional, time-dependent processes," AIChE J., vol. 62, no. 5, pp. 1478 - 1493, 2016, doi: 10.1002/aic.15062.

[3] 정하영, 홍석훈, 전재성, 임수창, 김종찬 and 박철영. "태양광 발전량 데이터의 시계열 모델 적용을 위한 결측치 보간 방법 연구" 멀티미디어학회논문지 24, no.9 (2021) : 1251-1260.

[4] Youden, William J. "Index for rating diagnostic tests." Cancer 3.1 (1950): 32-35.

[5] T. Lee, L. K. Lee, and C. Kim, "Performance of Machine Learning Algorithms for Class-Imbalanced Process Fault Detection Problems," in IEEE Transactions on Semiconductor Manufacturing, Vol. 29, No. 4, pp. 436 - 445, 2016.

스펙트럼-정합 확산 샘플링을 통한 이미지 초해상도 연구

전승재, 권진경, 전윤호, 박천음*
국립한밭대학교
{sjjeon, 20221043}@edu.hanbat.ac.kr, {yhjeon, parkce}@hanbat.ac.kr

Spectral-Aligned Diffusion Sampling for Image Super-Resolution

Seungjae Jeon, Jinkyung Kwon, Yunho Jeon, Cheoneum Park*
Hanbat National University

요 약

확산 모델은 이미지 생성과 초해상도에서 뛰어난 세부 묘사를 보이지만, over-smooth와 grainy artifact가 여전히 관찰된다. 본 논문은 저해상도 이미지의 주파수 정보를 추출하여 역확산 단계에 스펙트럼-정합 가이던스로 주입함으로써 이러한 문제를 완화하는 방법을 제안한다. 제안 방식은 외부 모델 · 추가 학습이 없는 플러그-인 방법으로, 어떤 확산 기반 SR 샘플러에도 적용 가능하다. SRDiff 기반 평가에서 표준 벤치마크 기준 PSNR이 평균 +0.9 dB, WSNR이 +1.1 dB 향상되었고, 주파수 오차가 감소하는 경향을 보였다.

I. 서론

이미지 초해상도(Super-Resolution, SR)는 저해상도(Low-Resolution, LR) 이미지로부터 고해상도 영상을 복원하는 기술로 의료, 위성, 보안 분야에서 중요한 역할을 한다. 최근 확산 모델이 SR에 도입되어 우수한 성능을 보이며 활발히 연구되는 중이다[1, 2]. 또한 사전 학습된 확산 모델을 활용한 접근법들이 주목받고 있는데, 먼저 DDRM [3]은 역확산 과정에서 열화 연산자의 null space 성분만을 복원하는 방법이다. DDNM [4]은 입력에 없는 고주파 성분을 확산 과정에서 보충하는 방식으로 추가 학습 없이도 높은 성능을 달성한다.

이처럼 확산 모델 기반 SR은 유의미한 결과를 보였지만, 여전히 over-smooth와 grainy noise 같은 고질적 한계가 있다. 이를 극복하기 위해 Spectral Diffusion [5], Wavelet Diffusion GAN [6]와 같이 주파수 영역 정보를 활용하는 연구들이 진행된다: Spectral Diffusion은 역확산 과정에서 고주파 성분이 나중에 생성된다는 관찰을 기반으로 U-Net에 wave gating 모듈을 추가하여 고주파 성분을 강조한다. Wavelet Diffusion GAN은 Discrete Wavelet Transform을 통해 영상을 주파수 성분으로 분해한 후 저차원 공간에서 확산을 수행하여 적은 step으로도 고주파를 보존한다.

본 논문은 확산 모델의 low-frequency bias 문제를 해결하기 위해 LR 영상에서 추출한 주파수 정보를 역확산 단계에 spectral-aligned 가이던스(guidance)로 주입하는 추론 전용 플러그인을 제안한다. 제안 방법은 기존 샘플러(DDPM, DDIM, DPM-Solver++)에 추가 학습 없이 적용 가능하며, SRDiff[2] 모델을 백본으로 한 표준 벤치마크 실험(×4, 100 steps)에서 PSNR +0.9 dB, WSNR +1.1 dB의 성능 향상을 달성하고 주파수 오차 감소를 확인한다.

II. 제안 방법

확산 기반 초해상도는 [수식 1]와 같이 고해상도(High Resolution, HR) 이미지 $\mathbf{x}_0$를 점진적으로 가우시안 노이즈로 변형하는 확산 과정과, 이를 저해상도 조건 $\mathbf{y}$에서 복원하는 역확산 과정으로 정의된다.

$$\begin{aligned} q(\mathbf{x}_t \mid \mathbf{x}_0) &= \mathcal{N}\left(\sqrt{\bar{\alpha}_t}\mathbf{x}_0,\ (1-\bar{\alpha}_t)\mathbf{I}\right) \\ p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t, \mathbf{y}_\uparrow) &= \mathcal{N}\left(\boldsymbol{\mu}_\theta(\mathbf{x}_t, t, \mathbf{y}_\uparrow),\ \sigma_t^2\mathbf{I}\right) \end{aligned} \tag{1}$$

여기서 $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$, $\alpha_s \in (0,1)$는 노이즈 분산 스케줄이다. $\boldsymbol{\mu}_\theta$는 [수식 2]과 같이 ϵ을 예측하는 신경망 ϵ_θ로 매개변수화된다.

$$\boldsymbol{\mu}_\theta(\mathbf{x}_t, t, \mathbf{y}_\uparrow) = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(\mathbf{x}_t, t, \mathbf{y}_\uparrow)\right) \tag{2}$$

본 논문에서는 사전학습된 확산 초해상도 모델(SR3, SRDiff 등)에 대해 샘플링 단계에서만 가이던스를 주입하는 방법을 제안한다. 가이던스가 포함된 역확산 업데이트는 [수식 3]와 같다.

$$\mathbf{x}_{t-1} \leftarrow \boldsymbol{\mu}_\theta(\mathbf{x}_t, t, \mathbf{y}_\uparrow) + \sigma_t\mathbf{z} - \eta_t\nabla_{\mathbf{x}}\mathcal{E}(\mathbf{x}_{t-1}) \tag{3}$$

여기서 $\eta_t > 0$는 가이던스 스케일, $\mathbf{y}$는 LR 이미지, $\mathbf{y}_\uparrow$는 LR 이미지를 bicubic 업샘플링한 결과이다. $\mathcal{E}(\mathbf{x})$는 본 논문에서 제안하는 결합 가이던스이다. 이는 [수식 4]와 같이 데이터 일관성 항과 주파수 가이던스 항의 가중합으로 정의된다.

$$\mathcal{E}(\mathbf{x}) = \lambda_{\mathrm{dc}}\,\mathcal{L}_{\mathrm{dc}}(\mathbf{x};\mathbf{y}) + \lambda_{\mathrm{freq}}\,\mathcal{L}_{\mathrm{freq}}(\mathbf{x};\mathbf{y}) \tag{4}$$

데이터 일관성 항과 그래디언트(gradient)는 [수식 5]과 같이 계산된다. 다운샘플링 연산자 D에 의해 HR 이미지가 LR 관측값과 일치하도록 강제한다.

$$\mathcal{L}_{\mathrm{dc}}(\mathbf{x};\mathbf{y}) = \|D(\mathbf{x}) - \mathbf{y}\|_2^2,\ \nabla_{\mathbf{x}}\mathcal{L}_{\mathrm{dc}} = 2D^\top\left(D(\mathbf{x}) - \mathbf{y}\right) \tag{5}$$

이 방법은 $\mathbf{y}$의 스펙트럼을 참조하여, 고주파 영역을 선택적으로 보강한다. [수식 6]에서 $\mathcal{F}$는 2D FFT (Fast Fourier Transform), $|\cdot|$는 스펙트럼 크기이다. 먼저 LR의 정규화된 스펙트럼 분포를 정의한다. 이를 기반으로 고주파를 강조하는 가중 마스크 $W(\omega)$를 구성한다. 단조 증가 함수 $\phi(S_y)$는 $\exp(S_y/\tau)-1$으로 설정했다. 여기서 τ는 고주파 영역을 결정하는 임계값이다.

Dataset	Method	PSNR↑	SSIM↑	MS-SSIM↑	LPIPS↓	FID↓	WSNR↑	Freq Error↓	LR PSNR↑
Set5	DDPM	26.926	0.754	0.958	**0.161**	45.564	30.793	0.090	30.733
	DDIM	19.551	0.318	0.823	0.630	110.964	23.184	0.719	22.811
	DPM++	19.471	0.314	0.821	0.632	115.044	23.096	0.733	22.757
	Ours(DDPM)	**27.837**	**0.819**	**0.969**	0.167	**43.245**	**31.750**	**0.068**	34.009
	Ours(DDIM)	26.497	0.743	0.952	0.197	51.141	30.461	0.080	**34.646**
	Ours(DPM++)	26.327	0.759	0.956	0.223	57.384	30.413	0.075	33.293
BSD100	DDPM	24.665	0.624	0.919	**0.332**	**79.989**	27.996	0.193	30.812
	DDIM	18.761	0.243	0.764	0.721	144.827	22.355	0.875	22.542
	DPM++	18.714	0.241	0.762	0.722	145.080	22.308	0.886	22.501
	Ours(DDPM)	**25.226**	**0.685**	**0.929**	0.358	81.170	**28.539**	**0.170**	33.956
	Ours(DDIM)	24.152	0.601	0.901	0.391	90.563	27.554	0.191	**34.492**
	Ours(DPM++)	24.216	0.634	0.910	0.414	87.610	27.619	0.184	33.729
Urban100	DDPM	22.591	0.651	0.929	0.312	**10.059**	26.017	0.331	28.535
	DDIM	17.985	0.293	0.793	0.711	34.102	21.440	0.994	22.382
	DPM++	17.944	0.291	0.792	0.713	35.837	21.396	1.006	22.345
	Ours(DDPM)	**22.952**	**0.704**	**0.937**	**0.300**	11.879	**26.368**	**0.307**	31.490
	Ours(DDIM)	22.118	0.621	0.912	0.368	24.398	25.596	0.327	**33.824**
	Ours(DPM++)	22.072	0.645	0.918	0.366	23.211	25.610	0.320	33.008

표 1: 다양한 샘플러에 가이던스를 적용한 결과이다. 가장 우수한 수치는 **볼드체**, 그다음 우수한 수치는 밑줄로 표기했다.

$$S_y(\omega) = \frac{|\mathcal{F}\{\mathbf{y}\}(\omega)|^2}{\|\mathcal{F}\{\mathbf{y}\}\|_2^2},\ W(\omega) = 1 + \gamma \cdot \phi(S_y(\omega)) \tag{6}$$

아래 [수식 7]은 주파수 가이던스 항과 그레디언트이다.

$$\mathcal{L}_{\text{freq}}(\mathbf{x};\mathbf{y}) = \|W(\omega) \odot (|\mathcal{F}\{\mathbf{x}\}| - \alpha_t \cdot |\mathcal{F}\{\mathbf{y}_\uparrow\}|)\|_2^2$$
$$\nabla_{\mathbf{x}}\mathcal{L}_{\text{freq}} = 2\,\Re\left[\mathcal{F}^{-1}\left(W(\omega)^2 \odot (|\mathcal{F}\{\mathbf{x}\}| - \alpha_t|\mathcal{F}\{\mathbf{y}_\uparrow\}|) \cdot \frac{\mathcal{F}\{\mathbf{x}\}}{|\mathcal{F}\{\mathbf{x}\}|}\right)\right] \tag{7}$$

여기서 $\Re$는 실수부를 취하는 연산이며, α_t는 t에 따라 고주파를 증가시키는 스케줄 파라미터이다.

III. 실험

그림 1: 제안한 가이던스의 아티펙트 제거 효과를 보여준다. 아래 사진은 빨간 박스 영역을 2배 확대한 것이다.

제안 방법의 유효성 검증을 위해, DIV2K 데이터셋으로 사전 학습된 SRDiff를 고정한 뒤 샘플러를 교체하며 실험한다. 모든 실험은 4 초해상도와 100-step으로 설정한다. 샘플러 비교는 DDPM, DDIM, DPM-Solver++와 수행하며, 각 샘플러에 제안한 스펙트럼-정합 가이던스를 플러그인 형태로 결합하여 평가한다. 평가는 Set5, Set14, BSD100, Urban100 데이터셋으로 수행하고, 평가 지표는 왜곡 지표 PSNR/SSIM/MS-SSIM, 지각 지표 LPIPS/FID, 주파수 특화 지표 WSNR/Freq Error, 다운샘플 재투영 일관성 LR PSNR을 사용한다.

[표 1]에 따라 제안 방법이 왜곡, 지각, 주파수 지표를 동시에 향상시킴을 알 수 있다. Ours (DDPM)은 기본 DDPM 대비 PSNR과 WSNR이 안정적으로 상승하고 Freq Error가 감소한다. 지각 지표는 데이터셋에 따라 등락이 있으나, Urban100에서는 LPIPS가 0.312→0.300로 감소하는 등 고주파 데이터 셋에서 강인한 특성을 보인다. 한편 결정론 샘플러(DDIM/DPM++)에 가이던스를 결합했을 때 효과가 더 크다. 이는 결정론적 샘플러가 계산 효율성을 위해 확률적 노이즈를 제거하면서 손실한 고주파 복원 능력을, 제안한 스펙트럼-정합 가이던스가 보완했기 때문으로 사료된다. 추가적으로, 모든 데이터셋에서 LR PSNR이 유의미하게 증가했는데, 이는 제안한 가이던스가 LR 일관성을 강화하고 [그림 1]과 같이 grain artifact 생성을 억제했다고 해석할 수 있다.

IV. 결론

본 연구는 LR에서 주파수 정보를 추출하여 역확산 단계에 스펙트럼-정합 가이던스로 주입하는 플러그인을 제안하였다. 실험 결과, SRDiff ×4 100-step 설정에서 기존 샘플러 대비 PSNR +0.9 dB, WSNR +1.1 dB의 향상을 보였으며, 특히 DDIM/DPM++ 같은 결정론적 샘플러에서는 성능이 더 크게 개선되었다. 이는 주파수 정합이 저주파 편향에 따른 평활화와 고주파 디테일 손실/잡음을 완화하는 경량 해법임을 보였다. 다만 가이던스 스케일, FFT 실험 설계의 민감성과 LR에서 정보를 추출한 한계로 평활화를 완전히 해결하진 못했다. 향후 확산 모델의 ϵ_θ에 대한 주파수 분석 연구를 진행하여 샘플러 적용 가능성을 탐구할 예정이다.

참 고 문 헌

[1] C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet, and M. Norouzi, "Image super-resolution via iterative refinement," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 4, pp. 4713–4726, 2023.

[2] H. Li, Y. Yang, M. Chang, S. Chen, H. Feng, Z. Xu, Q. Li, and Y. Chen, "Srdiff: Single image super-resolution with diffusion probabilistic models," *Neurocomputing*, vol. 479, pp. 47–59, 2022.

[3] B. Kawar, M. Elad, S. Ermon, and J. Song, "Denoising diffusion restoration models," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

[4] Y. Wang, J. Yu, and J. Zhang, "Zero-shot image restoration using denoising diffusion null-space model," in *International Conference on Learning Representations (ICLR)*, 2023.

[5] X. Yang, D. Zhou, J. Feng, and X. Wang, "Diffusion probabilistic model made slim," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 22552–22561, 2023.

[6] L. Aloisi, L. Sigillo, A. Uncini, and D. Comminiello, "A wavelet diffusion GAN for image super-resolution," in *International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, 2024.

AI 기반 맞춤형 아동 성장 정밀 관리 서비스 설계 및 구현

윤수현, 배민중, 권재영, 김민희*, 백성하*
신라시스템
silla@sillasystem.com

Design and Implementation of an AI-Based Personalized Pediatric Growth Management Service

Yun Su Hyeon, Bae Min Jung, Kwon Jae Young, Kim Min Hee*, Baek Seong Ha*
Silla System Co., Ltd

요 약

본 논문은 성장기 아동의 신장·체중·골격 성숙도 등 발달 지표를 통합 관리하고, AI를 활용해 맞춤형 성장 로드맵과 건강 가이드를 제공하는 모바일 서비스(aiCare)를 설계·구현하였다. 본 서비스는 의료기관 X-ray 연동, 성장 이력 시각화, 뼈나이·체형 분석, LLM 기반 상담 기능을 통해 데이터 입력·수집 - 분석 - 결과 제공의 흐름을 완성하며, 부모가 자녀 성장 상태를 직관적으로 확인하고 조기 이상 징후를 발견할 수 있도록 지원한다. 또한 PDF 리포트 생성, 안전성 검증된 AI 조언, 보안 강화 설계를 통해 가정과 의료기관 간의 성장 관리 상호작용을 촉진하며, 향후 웨어러블·영양·활동 데이터 통합과 AI 예측 고도화를 통해 정밀 성장 관리 플랫폼으로의 발전 가능성이 있다.

Ⅰ. 서 론

최근 AI 헬스케어 기술의 발전은 개인 건강 관리 방식을 근본적으로 변화시키고 있으며, 특히 성장기 아동을 대상으로 한 맞춤형 성장 관리 서비스의 필요성이 증대되고 있다. 선행 연구는 모바일 기반 성장 관리 도구가 부모로 하여금 자녀의 발달 상태를 주기적으로 기록·분석하게 하여 조기 이상 징후를 발견하고, 장기적인 건강 관리에 긍정적 영향을 미친다는 점을 보여주었다[1].

본 연구의 Contribution는 다음과 같다. 첫째, 기존 성장 관리 서비스가 단순 키·체중 기록에 머무른 반면, 본 연구는 의료기관 X-ray 기반 뼈나이·체형 분석까지 통합한 과학적 성장 관리 시스템을 제안한다. 둘째, WHO 및 국내 표준 성장도표와 Greulich-Pyle 표준 등 의학적 지표를 근거로 성장 데이터를 분석하고, 성장 단계(피크 성장기, PHV 가능성)를 정량적으로 추정함으로써 증거 기반 성장 관리를 구현한다. 셋째, 단순 지표 제시에 그치지 않고 AI 및 LLM 상담 기능을 활용하여 맞춤형 건강 가이드를 제공하고, 부모의 생활습관 개입 및 의료 상담 준비로 이어지는 폐쇄 루프 관리 구조를 구축한다. 또한, GPT 모델과의 프롬프트 연결 최적화 기법을 적용하여 단순 질의응답 수준을 넘어 상황별 세밀한 컨설팅이 가능하도록 성능을 개선하였다. 더불어, 세부 알고리즘 보완을 추가하여 성장 분석 결과를 한층 더 구체적으로 해석·제시함으로써, 부모가 직관적으로 이해할 수 있는 결과물을 제공하고 의료 상담 시 참고 가능한 정보의 깊이를 강화하였다.

따라서 본 논문은 아동의 신장·체중·뼈나이·체형 데이터를 통합 관리하고, AI 기반 분석과 GPT 기반 상담을 통해 부모가 가정에서도 의료기관 수준의 데이터 기반 의사결정을 내릴 수 있는 과학적 성장 관리 서비스(aiCare)의 설계를 제시한다.

Ⅱ. 본론

본 논문의 목표는 성장기 아동의 신장·체중·뼈나이·체형 정보를 한곳에서 보기 쉽게 관리하고, AI를 이용해 현재 상태와 향후 관리 방법을 제시하는 데 있다. aiCare의 전체 시스템은 의사용 웹과 부모용 앱으로 구성되지만, 본 구현에서는 부모용 앱을 우선 개발하였다. 전체 흐름은 (1) 데이터 입력·수집(프로필, 성장 기록, 의료기관 X-ray 연동), (2) 분석(성장 표준 비교, 단기 예측·PHV 가능성 판단), (3) 결과 제공(그래프, 요약 리포트, 상담)으로 구성된다. 부모는 서비스에서 기록을 관리하고 결과를 확인하며, 필요 시 의료기관 방문 전에 요약 리포트를 출력해 상담 준비를 할 수 있다.

등록된 자녀의 기본 정보와 부모 키(유전 예측용)을 참고하여, 성장 이력 관리에서는 월·분기 단위의 신장·체중 기록과 자동 BMI 계산, 과거 기록 수정·삭제, 표준성장도표 대비 추세 시각화를 제공한다. 뼈나이·골격 성숙도 기능은 의료기관 측정 X-ray를 불러오거나 측정값을 직접 입력해 뼈나이(예: 9.8세)와 참고 설명을 저장하고, 히스토리로 추적한다. 이 과정에서 알고리즘을 보강하여 동일 연령 대비 골격 성숙 지연·가속 여부를 세분화해 제시함으로써 의료 상담 시 참고 가능한 정보의 깊이를 강화하였다.

AI 기반 성장 예측은 과거 기록과 뼈나이를 바탕으로 단기 성장과 성장 단계(피크 성장기, PHV) 가능성을 제시하며, 부모 키 입력 시 유전 예상 키를 함께 보여준다. 체형·척추 분석은 의료기관 등록 X-ray 및 간단 체크리스트(어깨 비대칭, 자세 습관 등)를 바탕으로 상태를 분류·설명하고, 스트레칭·자세 교정 등 생활 가이드를 카드 형태로 안내한다. 체형 분류 기준과 분석 알고리즘을 세분화하여, 단순 분류를 넘어 상태별 특징과 생활습관 연관성을 구체적으로 제시하였다. 종합 성장 가이드 리포트는 PDF로 최근 변화, 현재 백분위, 예측 요약, 체형 포인트, 관리 팁을 묶어 제공한다. 또한 LLM 상담은 아이의 최신 성장 데이터와 생활 습관을 프롬프트 엔지니어링으로 미리 구조화하여 입력하고, 그 결과 부모에게 맞춤형 조언을 제공한다.

데이터는 입력 단계에서 단위·범위를 검증한 뒤 저장되며, 수정 시에는 이력을 기록해 추적이 가능하다. 수치 데이터는 관계형 데이터베이스에,

X-ray와 리포트 파일은 별도의 스토리지에 보관하여 보안과 관리 효율성을 높인다. 서버는 데이터 관리 API, 의료기관 X-ray 연동, 분석 처리, PDF 리포트 생성 기능을 포함하고, 클라이언트(웹앱)는 직관적인 UI와 단위 자동 변환, 임시 저장 기능을 제공한다. 성장 예측은 과거 신장·체중 변화와 뼈나이를 종합 분석해 단기 추세와 성장 단계 가능성을 산출하며, 유전 예상 키는 참고 지표로 활용된다. LLM 상담은 고정된 구조를 사용품질 점검을 통해 미달 시 정형 리포트만 제공한다. 보안은 로그인 보호, 전송·저장 암호화, 접근 권한 제어, 접근 로그 기록을 적용해 민감한 건강 데이터를 안전하게 관리한다. 이러한 설계를 통해 예측 - 권고 - 행동 - 피드백이 순환되는 성장 관리 구조를 구현하고, 부모의 생활습관 개입이 진료 전 준비와 사후 관리로 자연스럽게 이어지도록 하였다.

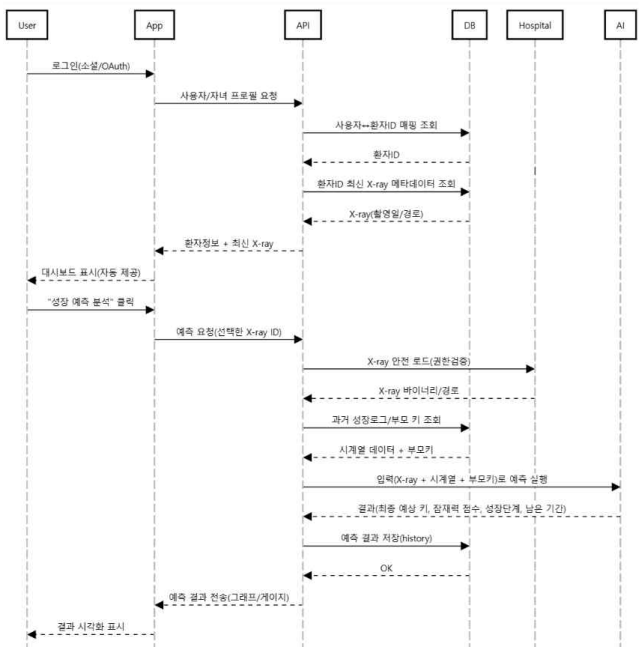


그림 1 AI 기반 X-ray 분석 시퀀스 다이어그램

Ⅲ. 결론

본 연구는 성장기 아동의 신체 발달 정보를 통합 관리하고, AI 기반 분석 및 상담 기능을 통해 부모가 보다 과학적이고 직관적으로 성장 상태를 이해할 수 있도록 하는 모바일 서비스(aiCare)를 설계·구현하였다. 기존 성장 관리 도구가 단순 지표 기록에 머무른 한계를 넘어, 본 서비스는 의료기관 X-ray 기반의 뼈나이 및 체형 분석, 성장 단계 예측, 유전적 예상 키 산출, LLM 기반 맞춤형 상담 등 다층적 기능을 제공함으로써 가정과 의료 현장을 잇는 성장 관리 플랫폼의 가능성을 제시하였다.

본 구현을 통해 부모는 자녀의 신장·체중·체형 변화를 주기적으로 기록하고, 결과를 표준 성장곡선 및 뼈나이와 연계하여 확인할 수 있으며, PDF 리포트 생성 및 상담 기능을 활용해 의료기관 방문 전 준비와 사후 관리에 효과적으로 활용할 수 있다. 또한 데이터 보안·검증 체계와 AI 품질 관리 절차를 도입하여 민감한 건강 데이터를 안전하게 다루면서, 부모가 신뢰할 수 있는 결과를 제공하고자 하였다. 향후 연구에서는 웨어러블 기기를 통한 활동·수면·영양 데이터의 연동, 대규모 아동 성장 데이터셋을 활용한 AI 모델 고도화, 개인 맞춤형 생활습관 개입 알고리즘 개발 등을 통해 보다 정밀한 성장 관리가 가능하도록 확장할 계획이다. 이를 통해 aiCare는 단순한 기록·분석 도구를 넘어, 가정과 의료기관을 연결하는 정밀 성장 관리 플랫폼으로 발전할 수 있을 것이다.

ACKNOWLEDGMENT

This work was supported by Sillasystem Co., Ltd.

참 고 문 헌

[1] Development and Effects of a Child Health Management Program Based on Cooperative Learning for Mothers of Preschool Children. (2017)

AI 기반 보이스피싱 시뮬레이션을 활용한 개인 맞춤형 사이버보안 교육 방법론 연구
-피해자특성별취약성분석및예방전략도출을중심으로-

홍윤이*, 최다온*, 양윤모*, 박지원, 박성미†

한림대학교

*동등 기여 저자, †교신저자

coolalex127@gmail.com, tldkwhgsl@gmail.com, hong.yy041@gmail.com,

jeewonpark143@hallym.ac.kr, sungmi.park@hallym.ac.kr

AI-Driven Voice-Phishing Simulations for Personalized Cybersecurity Education: Analyzing Victim Vulnerability and Designing Targeted Prevention Strategies

Yunyi Hong*, Daon Choi*, Yoonmo Yang*, Jee Won Park, Sungmi Park†

Hallym University

*Equal contribution, †Corresponding Author

요 약

보이스피싱 피해가 전년 대비 2.2배 급증하고 50대 이상 취약계층 피해가 심각해지고 있으나, 기존 획일적 예방 교육은 개인별 특성을 고려하지 못해 실효성에 한계가 있어 맞춤형 교육의 필요성이 대두되고 있다. 본 연구는 AI 에이전트 간 대화 시뮬레이션을 통해 8종 시나리오와 6가지 피해자 프로필을 활용하여 연령과 성격 특성별 취약성을 분석하고 개인 맞춤형 예방전략을 제공하는 시스템을 개발하였다. 이를 통해 진화하는 보이스피싱 수법에 대한 신속한 대응력 확보와 지속적인 교육 효과 향상을 달성하고자 한다.

Ⅰ. 서 론

보이스피싱은 전화를 통해 금융기관이나 공공기관을 사칭하여 금전을 편취하는 전기통신금융사기범죄로, 2025년 1분기 보이스피싱 피해액은 3,116억원(전년 동기 대비 2.2배)으로 증가하였다[1]. 특히 전체 피해자의 50% 이상이 50대 이상으로 디지털 취약계층 피해가 심각한 실정이다[1]. 그럼에도 불구하고 기존 예방 교육은 일반적인 주의사항 전달에 그쳐 개인별 특성과 취약성을 반영한 대응에 한계를 보이고 있다. 이에 본 연구는 AI 에이전트를 활용하여 보이스피싱 상황을 가상 환경에서 재현하고, 피해자 특성별 취약성을 정량적으로 분석하여 개인 맞춤형 예방 전략을 제시하는 시스템을 개발하고자 한다.

Ⅱ. 관련연구

본 연구는 보이스피싱의 범행 수법과 피해자 특성을 설계 하기 위해 대출사기형, 기관 사칭형, 가족·지인 사칭형으로 분류하고 각 유형별 범행 단계를 분석하였다. 연령별 피해 현황 분석[2]에 의하면 20대는 기관 사칭형, 60대 이상은 가족 사칭형에 취약하였다. 또한 [3]에서는 연령별 디지털 금융 이해력 격차와 피싱 취약성의 상관관계를 제시되었으며. [4]에서는 OCEAN 성격 모델 분석에서 높은 외향성과 신경성이 피싱 취약성을 증가시킨다고 보고하였다.

Ⅲ. 연구방법

본 연구는 보이스피싱 시나리오 8종과 가상 피해자 6명을 설정하여 대화 시뮬레이션을 수행하였다. 보이스피싱 시나리오는 실제 사례[2]를 반영하여 보이스피싱의 유형, 목적, 범행 단계를 포함하도록 설계되었다(표 1참조). 목적과 범행단계 없이 유형만 제공할 경우 대화가 단순화되어, 본 연구에서는 시나리오별 목적과 수법을 단계별로 모델에 제공함으로써 실제 피해 사례와 유사한 대화 흐름을 유도하였다.

표 1 대출 사기형 목적 및 범행단계 분류

유형	목적	범행단계
대출 사기형	저금리 대출유도 후 기존 대출받은 기관 사칭하여 대출금 편취	(1) 저금리 대출 제안 (2) 대출신청 유도 (3) 재통화 상환요구 (4) 현금·계좌 송금

피해자 프로필은 보이스피싱 취약계층의 특성을 분석하여 나이, 성별, 학력과 같은 기본정보, 디지털 금융 리터러시 설문조사 기반 지식수준과 OCEAN 모델에 기반한 성격을 부여하였다. 특히, 보이스피싱 범죄는 연령에 따라 피해 규모 및 빈도수에 격차가 커 나이 구간별로 하나의 대표적인 프로필을 생성하였다.

그 외 지식수준 및 성격은 보이스피싱 피해 발생 빈도수[2]에 따라 설정하여 시뮬레이션 가능한 총 6가지의 피해자 프로필을 구성하였다(아래 표2 참조).

표 2 피해자 (60대) 프로필 예시

	디지털 금융 리터러시 [3]: ◆ 70대보다는 금융이해력이 높지만 다른 연령대 대비 낮음 ◆ 70대보다는 디지털 금융이해력이 높지만 다른 연령대 대비 낮음 ◆ 공용 Wifi를 이용한 온라인 쇼핑에 대한 이해력 높음
이름: 엄석대 나이: 60대 학력: 고등학교 중퇴 성별: 남자	**OCEAN 성격 설정 [4]:** ◆ 개방성(O), 친화성(A), 성실성(C) 낮음 ◆ 신경성(N), 외향성(E) 높음

또한, 역할별로 적합한 모델을 선정하기 위해 각 플랫폼의 playground를 통해 실험을 진행하였으며, 최종적으로 피싱범 역할에는 우수한 설득력을 보인 GPT-4.1-mini를, 피해자 역할에는 역할 수행 능력과 정책적으로 유연한 방어 수준을 갖춘 Gemini-2.5 Flash Lite를 선정하였다. 관리자 에이전트는 대화 분석 및 예방교육 효과 평가를 담당하므로 추론 능력이 뛰어난 o4-mini를 사용하여 구축하였다.

Ⅳ. 시뮬레이션 절차

본 연구의 보이스피싱 시뮬레이션은 다음과 같은 절차(그림1)를 따른다. 첫째, 사용자가 선택한 시나리오와 피해자 프로파일을 가상 대화 환경에 입력한다. 둘째, 피싱범(GPT-4.1-mini)과 피해자(Gemini-2.5 Flash Lite)가 대화를 진행하며, 피해자는 설정된 지식과 성격에 따라 대응한다. 셋째, 관리자 에이전트(o4-mini)가 대화를 분석하여 피싱 성공/실패를 판정한다. 넷째, 결과에 따라 피해자 모델의 방어를 강화하거나 피싱범의 수법을 정교화한다. 마지막으로 시뮬레이션 결과를 바탕으로 맞춤형 예방책을 제시한다.

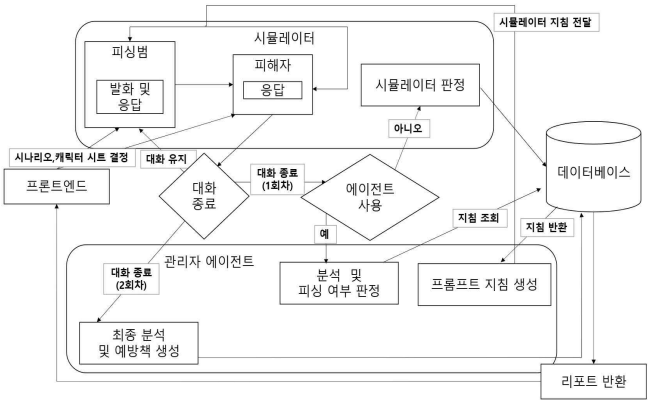


그림 1 시뮬레이션 프로세스 흐름도

V. 실험 결과 및 분석

본 연구는 피싱범과 피해자 간의 대화 1회를 1턴이라 정의하고, o4-min를 사용하여 턴별 피해자 설득 정도(S_i)를 산출하였으며, 금융감독원의 '그놈 목소리' 보이스피싱 데이터 및 AIHub 감정분류 대화 음성 데이터셋으로 학습한 RoBERTa-Base 모델의 보이스피싱 탐지 확률(V_i)을 결합하여 턴 단위의 위험도를 정의하였다. 전체 시나리오의 위험도 평균은 다음과 같이 산출하였다:

$$\overline{R} = \frac{1}{n}\sum_{i=1}^{n}(V_i \times S_i)$$

실험 결과, 전체 시나리오의 평균 위험도는 0.48(n=240)이었으며, 대화 진행에 따라 위험도가 증가하는(기울기>0) 시나리오는 25%였다. 턴별 위험도 분석 결과, 턴1→턴2 구간에서 가장 큰 증가폭(0.069)이 관찰되었다.

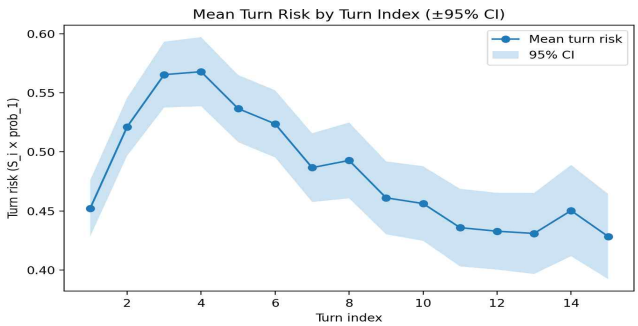


그림 2 턴 수별 평균 턴 위험도

시나리오 유형별로는 '자녀 사칭 후 계좌정보 및 원격제어 앱을 통한 편취'가 0.55로 가장 높은 위험도를 보였으며, 그다음으로 '추가 피해 예방을 빙자한 현금 편취'가 0.52로 높게 나타났다.

표 3 시나리오 유형별 케이스 위험도 상·하위 2개

시나리오 내용		위험도
상위	자녀 사칭, 계좌정보 및 원격제어 통해 피해금 편취	0.55017
상위	피해자에게 추가 피해 방지를 목적으로 속여 현금 편취	0.52284
하위	메신저로 지인 사칭, 결제를 요청하고 피해금 편취	0.46706
하위	우체국 직원 사칭, 전화하여 해외 카드 사용 알린 후 수사기관 사칭하여 피해금 편취	0.42907

또한 전체 시나리오 최종 결과인 보이스피싱 성공률은 기존 연구의 연령별 피해 패턴과 전반적으로 일치하였고. 특히 기관사칭형-2 수법은 모든 연령대에서 높은 피해율을 기록하였다.

성격 특성 분석 결과, 외향성이 높은 피해자들이 연령대별 취약 패턴을 벗어나 다양한 피싱에 취약성을 보였다. 특히 신경성과 외향성이 모두 높은 경우 전반적으로 취약성이 확대되는 경향을 보였다.

그림 3 연령·성격 특성별 주요 취약 시나리오 top 3

VI. 활용방안

본 연구를 통해 개발된 시스템은 웹 기반 인터페이스로 구현되어 사용 편의성을 확보하였다. 사용자는 시나리오와 피해자를 선택하여 즉시 시뮬레이션을 실행할 수 있으며, 다각적인 관점에서 범죄 가능성을 확인하고 맞춤형 교육 콘텐츠를 제공받을 수 있다. 관리자 에이전트는 시뮬레이션 결과를 바탕으로 특정 수법에 취약한 프로필에게 "전화로 송금 요청 시 공식 콜센터로 재확인하기"와 같은 개인 맞춤형 체크리스트를 제공한다. 지속적인 분석을 통해 예방 교육의 효과를 모니터링하고 최신 수법에 대한 대응력을 강화할 수 있다.

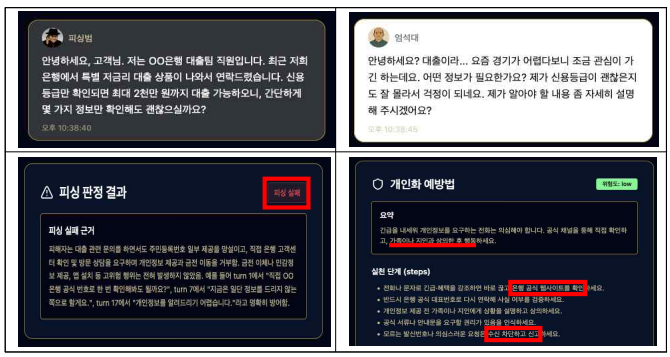


그림 4 시뮬레이션 대화 및 리포트 결과 예시

VII. 결론 및 향후 연구

본 연구는 AI 에이전트 기반 시뮬레이션을 통해 보이스피싱에 대한 개인 맞춤형 예방 교육 시스템을 개발하였다. 연령과 성격 특성을 고려한 취약성 분석을 통해 기존의 획일적인 예방 교육의 한계를 극복하고, 개인별 특성에 따른 차별화된 대응 전략을 제시하였다. 특히 실시간으로 진화하는 보이스피싱 수법에 신속하게 대응할 수 있으며, 반복적인 시뮬레이션을 통해 교육 효과를 지속적으로 향상시킬 수 있다는 점에서 의의를 갖는다. 향후 연구에서는 더욱 다양한 시나리오와 피해자 프로파일의 확장, 음성 합성 기술을 결합한 시뮬레이션의 현실성 강화, 그리고 다양한 연령대를 대상으로 한 실증 연구를 통한 교육 효과성 검증이 필요하다.

ACKNOWLEDGMENT

이 논문은 2025년도 정부(경찰청)*의 재원으로 과학치안진흥센터 사이버범죄 수사단서 통합분석 및 추론 시스템 개발 사업의 지원을 받아 수행된 연구임(RS-2025-02218280)

참 고 문 헌

[1] 이동환, "올해 1~3월 보이스피싱으로 3천억 털렸다…1년 만에 2배," 연합뉴스, 2025. 04. 27. (https://www.yna.co.kr/view/AKR20250426047400004)

[2] 김민정 외, "보이스피싱 피해 경험 및 영향요인 분석", 소비자문제연구, 52권 제1호, pp. 27 - 34, 2023.

[3]김대호 외, "행정조사기관 사칭 보이스피싱 실태분석 및 대응방안: 보이스피싱 조직과 개인 관계에서의 정보비대칭을 중심으로," 치안정책연구, 38권 제2호, pp 91-120, 2024.

[4] D. Sarno, et al., "Which phish is captured in the net? Understanding phishing susceptibility and individual differences," Applied Cognitive Psychology, vol. 37, no. 4, pp. 789 - 803, 2023.

Memory-aware RAG 기반 접근을 통한 언어모델의 장기 기억 강화

장효석[1,2], 조한얼[3], 이상철[1], '김찬수[1,2,*]

1 한국과학기술연구원 인공지능·정보·추론 연구실,

2 과학기술연합대학원대학교 AI-로봇

3 세종대학교

*Correspondence should be addressed to eau@ust.ac.kr

Enhancing Long-Term Memory in Language Models via a Memory-Aware RAG Framework

Hyoseok Jang[1,2], Haneol Cho, Sangchul Lee, Chansoo Kim[1,2,*]

1 AI·Information·Reasoning (AI/R) Laboratory, Korea Institute of Science and Tech. (KIST)

2 AI-Robot Department, University of Science and Technology (UST)

3 Sejong University, Republic of Korea

요 약

소형 언어모델(SLM)은 우수한 추론 능력에도 불구하고 장기 기억 부족으로 실제 응용에서 한계를 보인다. 본 연구는 외부 메모리와 적응형 검색을 결합한 메모리 인식 RAG(Memory-aware RAG) 프레임워크를 제안한다. 제안된 접근법은 단기 버퍼와 장기 저장소로 구성된 계층적 메모리 구조, 시간 가중치 기반 검색, 하이브리드 인덱싱, 사용자 피드백을 통한 자가 갱신 메커니즘을 포함한다. 이를 통해 SLM은 장기간의 상호작용에서도 맥락적 일관성과 개인화를 유지할 수 있다. 실험 계획에는 성능 비교, 사용자 기반 평가, 효율성 분석이 포함되며, 다양한 도메인에서 소형 모델의 실질적 활용 가능성을 검증할 예정이다. 본 연구는 소형 언어모델의 활용성을 크게 확장하고, 경량화된 대화형 에이전트 구현을 위한 새로운 길을 제시한다.

Ⅰ. 서론

최근 소형 언어모델(Small Language Models, SLMs)은 상대적으로 적은 파라미터 규모에도 불구하고 합리적인 추론 능력과 빠른 응답 속도로 주목받고 있다 [4]. 이들은 엣지 디바이스, 자원 제약 환경, 맞춤형 응용 시스템 등에서 특히 활용 잠재력이 크다. 그러나 소형 언어모델의 근본적인 제약은 기억력의 부족이다. 제한된 모델 크기 때문에 장기간의 대화 맥락을 유지하거나, 누적된 사용자 상호작용을 일관되게 반영하는 능력이 현저히 떨어진다. 그 결과, SLM은 실제 대화형 에이전트나 장기적 개인화 서비스에서 제대로 활용되지 못하고 있는 실정이다.

Retrieval-Augmented Generation(RAG)은 외부 지식베이스를 활용해 모델의 한계를 보완하는 대표적 기법이다 [1]. 이 기법은 사용자 질의에 관련된 정보를 외부 데이터베이스에서 검색한 후, 이를 언어모델의 컨텍스트로 함께 제공하여 답변의 근거를 보강하는 방식으로 작동한다. 이를 통해 모델의 환각(hallucination) 현상을 완화하고 사실적 정확성을 높이는 데 크게 기여했습니다. 대부분의 관련 연구는 이처럼 정적인 외부 지식을 얼마나 효과적으로 검색하고 활용하는지에 집중되어 왔습니다. 그러나 기존 RAG는 사실 검색(factual retrieval)과 정적 지식 grounding에 치중되어 있으며, 시간에 따라 변화하는 개인화된 맥락이나 연속적 상호작용의 축적을 다루는 데 한계가 있다 [2][3]. 이에 본 연구는 소형 언어모델의 부족한 장기 기억을 보완하기 위한 메모리 인식 RAG(Memory-aware RAG, M-RAG) 프레임워크를 제안한다.

Ⅱ. 본론

M-RAG는 소형 언어모델의 한계를 보완하기 위해 계층적 메모리 구조와 적응형 검색 전략을 결합한다.

1. 계층적 메모리 구조(Hierarchical Memory)

- 단기 버퍼: 최근 대화 맥락을 저장하여 즉각적인 응답 일관성 유지.
- 장기 저장소: 의미적 임베딩과 이벤트 기반 인덱스를 활용해 사용자 과거 상호작용을 보존 [5]

2. 적응형 검색(adaptive Retrieval)

- 시간 가중치와 중요도 학습을 통해 오래된 정보와 최신 정보 간 균형을 유지하며, 맥락적으로 중요한 기억을 우선적으로 검색.

3. 하이브리드 인덱싱(Hybrid Indexing)

- 의미적 유사도 기반 검색과 이벤트 기반 검색을 결합해, 단순 텍스트 매칭을 넘어 사용자 경험 전반을 반영.

4. 자가 갱신(self-updating)

- 사용자 피드백을 반영하여 메모리 중요도를 재조정함으로써 변화하는 상황에 지속적으로 적응.

실험

본 연구는 다음과 같은 시험을 통해 M-RAG의 효과성을 검증한다.

- 성능 비교: 기존 RAG 및 단순 SLM 대비 대화 일관성, 사실 정확도, 개인화 반영도를 측정 [3][5].
- 사용자 기반 평가: 실제 사용자가 M-RAG 기반 SLM과 상호작용 후 설문을 통해 기억 지속성·만족도·개인화 체감을 평가 [7].
- 효율성 분석: 메모리 저장 비용, 검색 지연(latency), 모델 크기 대비

성능 향상을 측정하여 경량 환경 적합성을 검증.

기대 결과

BenchmarkQED [8] 점수 기준 베이스 소형모델 대비 향상된 벤치마크 점수 획득

Ⅲ. 결론

소형 언어모델은 우수한 추론 능력에도 불구하고, 장기 기억 부족으로 실제 응용에서는 활용성이 크게 제한되어 왔다. 본 연구가 제안하는 M-RAG 프레임워크는 외부 메모리와 적응형 검색 메커니즘을 통해 이러한 한계를 보완하고, 소형 언어모델이 장기적 상호작용에서도 일관성과 개인화를 유지할 수 있도록 한다. 이를 통해 자원 제약 환경에서도 신뢰할 수 있는 대화형 에이전트를 구현할 수 있으며, 모바일, 교육, 금융, 헬스케어 등 다양한 도메인에서 소형 모델의 실질적 활용 가능성을 크게 확장할 수 있을 것이다.

ACKNOWLEDGMENT

This research was funded by the grant Nos. 2023-00262155; 2024-00339583; 2024-00460980; and 2025-02304717 (IITP) funded by the Korea government (the Ministry of Science and ICT).

참 고 문 헌

[1] Lewis, P., Perez, E., Piktus, A., et al. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. NeurIPS, 2020.

[2] Zhao, X., et al. MemoRAG: Global Memory Enhanced Retrieval for Long Contexts. arXiv preprint arXiv:2409.05591, 2024.

[3] Zhuang, Z., et al. From RAG to Memory: HippoRAG 2. arXiv preprint arXiv:2502.14802, 2024.

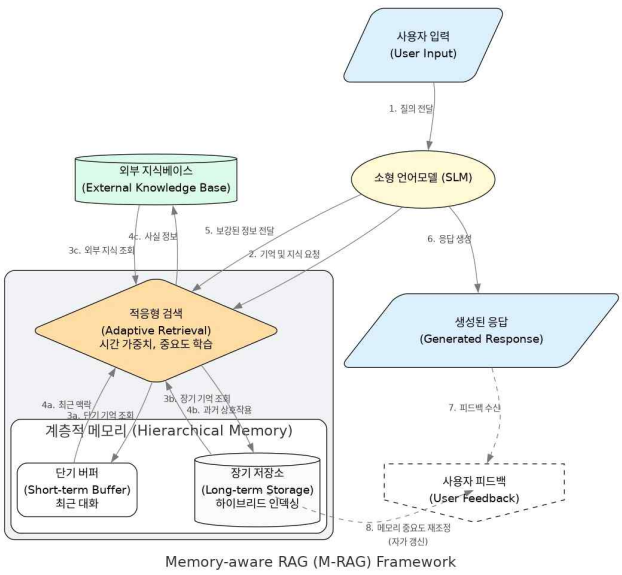
[4] Anil, C., et al. Scaling Down: Small Language Models with Surprising Reasoning Capabilities. arXiv preprint arXiv:2310.xxxxx, 2023.

[5] Chen, X., et al. LongMem: Augmenting Language Models with Long-Term Memory. arXiv preprint arXiv:2306.07174, 2023.

[6] Ma, X., et al. MemLong: Memory-Augmented Retrieval for Long Text Modeling. arXiv preprint arXiv:2408.16967, 2024.

[7] Li, M., et al. LoCoMo: Evaluating Very Long-Term Conversational Memory of LLM Agents. arXiv preprint arXiv:2402.17753, 2024.

[8] Edge, D., Trinh, H., Morales Esquivel, A., & Larson, J. (2025, June 5). *BenchmarkQED: Automated benchmarking of RAG systems*. Microsoft Research. Retrieved August 29, 2025, from https://www.microsoft.com/en-us/research/blog/benchmarkqed-automated-benchmarking-of-rag-systems/



Memory-aware RAG (M-RAG) Framework

그림 1 M-RAG 구조도

비전 LLM을 활용한 차세대 X-ray 보안 검색 시스템 개발

김민종[1]*, 박형달[2]

[1]한국원자력연구원, [2](주)에이리스

*kmj21@kaeri.re.kr

Development of Next-Generation X-ray Security Screening System Using Vision LLM

Minjong Kim[1]*, Hyeongdal Park[2]

[1]Korea Atomic Energy Research Institute, [2]AIRISS CO., Ltd.

* Corresponding author

요 약

현재의 X-ray 보안 검색 시스템은 검사원의 전문성과 폐쇄형 집합 기반 객체 탐지 모델에 크게 의존하고 있으며, 이들은 복잡한 은닉 시나리오와 새로운 형태의 위협에 대응하는 데 한계를 보인다. 본 논문은 X-ray 보안 이미지를 이해하고 추론할 수 있는 비전-언어 모델(VLM) 학습을 위한 고품질 데이터셋 구축 프레임워크를 제안한다. 제안하는 프레임워크는 세 가지 핵심 구성요소로 이루어진다: (1) 다양한 재질의 X선 투과 특성을 시뮬레이션하여 위험물품의 사실적인 X-ray 이미지를 생성하는 이미지 생성기, (2) 실제 X-ray 영상 물리학을 반영한 덧셈 합성을 통해 복잡한 은닉 시나리오를 구현하는 합성 모듈, (3) 구조화된 설명과 단계별 논리적 추론 과정을 생성하는 다층적 캡션 및 추론 체인 생성기이다. 생성된 데이터셋은 이미지-캡션 쌍, 추론 체인, 종합적인 합성 파라미터를 포함하는 구조화된 JSON 메타데이터로 저장된다. 본 프레임워크는 현재 AI 기반 보안 시스템의 중요한 한계를 해결하여, VLM이 단순한 패턴 인식을 넘어 맥락적 이해와 논리적 추론으로 나아갈 수 있도록 한다. 이러한 접근법은 변화하는 위협과 정교한 은닉 기법에 대응할 수 있는 차세대 적응형 보안 검색 시스템 개발의 기반을 제공한다.

Ⅰ. 서 론

X-ray 보안 검색은 테러 및 불법 물품 반입을 막는 국가 안보의 최전선에 있지만, 그 판독 과정은 여전히 검사원의 높은 숙련도와 집중력에 크게 의존한다. 특히 수하물 내에 여러 물품이 복잡하게 얽혀 있거나 위험물품이 의도적으로 다른 물체 뒤에 숨겨진 경우, 판독의 난이도는 급격히 증가하며 이는 인적 오류의 주요 원인이 된다. 이러한 문제를 해결하기 위해 인공지능(AI) 기반 자동 검출 시스템이 도입되고 있으나 [1-2], 현재의 기술은 대부분 사전 정의된 위험물품의 시각적 패턴을 학습하여 식별하는 폐쇄형 집합(closed-set) 패러다임에 머물러 있다. 이 방식은 학습 데이터에 포함된 특정 형태와 구도의 물체는 잘 찾아내지만, 다음과 같은 근본적인 한계를 가진다. 기존 모델은 '칼'의 형태는 인식할 수 있지만, '금속 케이스 안에 들어있는 칼날'과 같이 객체 간의 공간적, 기능적 관계를 이해하지 못한다 [3]. 이는 정교한 은닉 시도에 매우 취약하여 맥락적 이해의 부재가 있다. 추론 능력이 부족하여 3D 프린팅 총기나 새로운 형태의 폭발물처럼 학습 데이터에 없던 위협이 등장했을 때 대응하기 어렵다. X-ray 스캐너의 종류나 설정에 따라 이미지의 특성이 달라질 때 성능이 저하되는 문제가 있으며, 이는 모델이 객체의 본질이 아닌 특정 이미지 스타일에 과적합되기 때문이다. 이러한 한계를 극복하기 위해, 본 연구는 단순히 이미지를 '인식'하는 것을 넘어 '이해'하고 '추론'하는 비전-언어 모델(VLM)의 도입을 제안한다. VLM은 이미지 인코더와 대규모 언어 모델(LLM)을 결합하여, 이미지 내 시각적 정보와 언어적 설명을 연계하여 학습한다. 이에 본 연구는 VLM 훈련의 핵심인 고품질 데이터셋 구축에 초점을 맞춘다. 특히, 실제와 유사한 은닉 시나리오를 반영한 합성 X-ray 이미지와, 모델의 추론 능력을 극대화하는 상세한 설명 캡션(caption)을 생성하는 프레임워크를 제안하고자 한다. 이를 통해 차세대 지능형 X-ray 보안 검색 시스템 개발의 초석을 마련하는 것을 목표로 한다.

Ⅱ. 본론

본 연구에서 제안하는 X-ray 비전-언어 모델(VLM) 데이터셋 구축 프레임워크는 VLM이 X-ray 이미지의 복잡한 맥락을 효과적으로 학습하는 데 필요한 고품질의 이미지-캡션 쌍을 체계적으로 생성하는 것을 목표로 한다. 프레임워크는 그림1에 나타낸 것처럼 (1) 개별 위험물품 이미지 생성, (2) 복합 은닉 시나리오 합성, (3) 멀티모달 캡션 및 추론 데이터 생성의 세 가지 핵심 단계로 구성된다. 이 모든 과정은 통합된 사용자 인터페이스(UI)를 통해 제어되며, 연구자가 다양한 조건의 데이터셋을 효율적으로 구축할 수 있도록 지원한다.

2.1 개별 위험물품 이미지 생성

개별 위험물품 이미지 생성은 X-ray 투과 특성을 정확히 모사하는 것에서 시작된다. 실제 X-ray 스캐너에서 물체의 밀도와 두께에 따라 X선의 투과율이 달라지는 원리를 그레이스케일 이미지로 재현하기 위해, 각 물체를 구성하는 재질별로 서로 다른 픽셀 값을 할당하는 방식을 채택하였다. 칼, 총기, 가위, 망치 등 주요 위험물품의 기본 형태를 다각형과 사각형 같은 기초 도형의 조합으로 구현하였으며, 이는 실제 X-ray 이미지에서 관찰되는 물체의 단순화된 실루엣을 효과적으로 표현한다. 밀도 모사 과정에서는 객체의 각 부분에 재질 특성을 반영한 그레이스케일 값을 세심하게 할당하였다. 예를 들어, 칼의 경우 금속 재질의 날 부분은 180-220 범위의 높은 픽셀 값을 부여하여 X선이 투과하기 어려운 고밀도 영역임

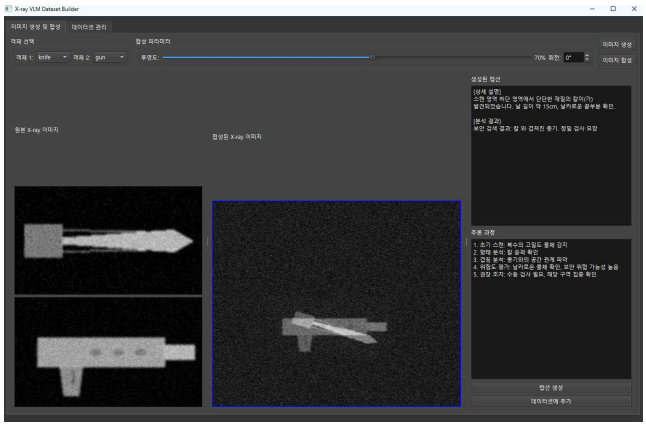


그림 1. 비전 LLM을 위한 위험물품 X-Ray 데이터셋 생성 시스템

을 표현하였고, 플라스틱이나 목재로 이루어진 손잡이 부분은 120-150 범위의 상대적으로 낮은 값을 할당하여 X선이 더 쉽게 투과하는 저밀도 영역임을 나타내었다.

2.2 복합 은닉 시나리오 합성

개별적으로 생성된 위험물품 이미지들은 실제 수하물 검색 상황에서 발생하는 복잡한 은닉 및 중첩 시나리오를 재현하기 위해 하나의 통합된 이미지로 합성된다. 이 과정은 단순한 이미지 중첩을 넘어서 X-ray 영상의 고유한 물리적 특성을 반영하도록 설계되었다. X-ray 이미지에서는 물체가 겹쳐질 때의 영역에 대한 특징을 살린 처리를 해주었다. 다양한 은닉 환경을 구현하기 위해 시스템은 여러 조작 가능한 파라미터를 제공한다. 회전 기능을 통해 위험물품이 다양한 각도로 배치된 상황을 만들 수 있으며, 투명도 조절 기능으로 위험물품이 다른 물체에 의해 부분적으로 혹은 심하게 가려진 시나리오를 생성할 수 있다. 이러한 파라미터들의 조합은 실제 보안 위협 상황에서 나타날 수 있는 정교한 은닉 기법들을 효과적으로 재현한다. 실제 X-ray 장비의 환경적 특성도 충실히 재현하였다. 센서에서 발생하는 전자적 노이즈를 정규 분포 기반의 무작위 패턴으로 모사하여 이미지 전반에 적용하였으며, 이는 실제 스캐너에서 관찰되는 자연스러운 노이즈 특성을 반영한다. 또한 이미지 대비 조절을 통해 다양한 제조사의 스캐너나 서로 다른 운영 환경에서 획득된 이미지들의 특성 차이를 시뮬레이션할 수 있도록 하였다. 이러한 변형들은 학습된 모델이 다양한 실제 환경에서도 강건하게 작동할 수 있는 일반화 능력을 갖추도록 돕는다.

2.3 추론 과정 데이터 생성

본 프레임워크의 가장 혁신적인 측면은 VLM이 위험 판단에 이르는 논리적 과정을 학습할 수 있도록 단계별 추론 과정 데이터를 생성하는 것이다. 이는 최근 LLM 분야에서 주목받고 있는 Chain-of-Thought 프롬프팅 기법을 [4] 시각-언어 멀티모달 영역으로 확장한 것으로, 모델이 마치 숙련된 보안 검색 전문가처럼 체계적인 사고 과정을 거쳐 결론에 도달하도록 유도한다. 추론 과정은 5단계의 논리적 흐름으로 구성된다. 첫 번째 단계인 '초기 스캔'에서는 이미지 전체에서 관찰되는 고밀도 물체들의 존재를 확인한다. 두 번째 '형태 분석' 단계에서는 개별 객체의 윤곽과 특징을 식별하며, 세 번째 '겹침 분석' 단계에서는 객체들 간의 공간적 관계와 은닉 패턴을 파악한다. 네 번째 '위험도 평가' 단계에서는 식별된 객체들의 위험성과 은닉 의도를 종합적으로 평가하고, 마지막 '권장 조치' 단계에서는 구체적인 대응 방안을 제시한다. 이러한 추론 데이터는 모델의 판독 결과에 대한 설명 가능성을 확보하는 데 핵심적인 역할을 한다. 기존의 블랙박스 AI 모델들이 가진 한계, 즉 '왜 그런 판단을 내렸는지' 설명할 수 없는 문제를 극복하여, 보안 검색과 같은 고위험 분야에서 요구되는 투명성과 신뢰성을 제공한다. 또한 이러한 단계적 추론 과정의 학습은 모델이 새로운 유형의 위협이나 예상치 못한 은닉 패턴에 직면했을 때도 논리적인 분석 과정을 통해 합리적인 판단을 내릴 수 있는 일반화 능력을 향상시킨다.

Ⅲ. 결론

본 논문에서는 X-ray 보안 검색을 위한 비전-언어 모델(VLM) 학습용 데이터셋 구축 프레임워크를 제안하였다. 기존의 폐쇄형 집합 기반 객체 탐지 모델이 가진 한계를 극복하기 위해, 실제 X-ray 영상의 물리적 특성을 정확히 모사한 합성 이미지 생성과 함께 다층적 캡션 및 추론 과정 데이터를 체계적으로 생성하는 통합 시스템을 개발하였다. 단순한 패턴 인식을 넘어 맥락적 이해와 논리적 추론이 가능한 차세대 X-ray 보안 검색 시스템 개발의 토대를 제공한다는 점에 있다. 특히 새로운 형태의 위협이나 정교한 은닉 시도에 대응할 수 있는 적응형 AI 시스템 구현에 필수적인 학습 데이터 구축 방법론을 제시하였다. 향후 연구 과제로는 실제 X-ray 이미지와의 도메인 적응 기법 적용, 추론을 위한 데이터생성 모델링을 통한 더욱 사실적인 이미지 생성이 있다. 이후 구축된 데이터셋을 활용한 VLM 모델의 실제 성능 검증과 현장 적용 가능성 평가가 필요하다. 본 프레임워크가 국가 안보 강화와 공공 안전 향상에 기여하는 실용적인 AI 시스템 개발로 이어지기를 기대한다.

ACKNOWLEDGMENT

본 논문은 2025년도 정부(과학기술정보통신부)의 재원으로 연구개발특구진흥재단의 지원을 받아 수행된 연구임. (2025-DD-RD-0367)

참 고 문 헌

[1] Miao, C., Xie, L., Wan, F., Su, C., Liu, H., Jiao, J., & Ye, Q. (2019). Sixray: A large-scale security inspection x-ray benchmark for prohibited item discovery in overlapping images. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 2119-2128).

[2] Liang, K. J., Sigman, J. B., Spell, G. P., Strellis, D., Chang, W., Liu, F., ... & Carin, L. (2019). Toward automatic threat recognition for airport X-ray baggage screening with deep convolutional object detection. arXiv preprint arXiv:1912.06329.

[3] Gao, C., Killeen, B. D., Hu, Y., Grupp, R. B., Taylor, R. H., Armand, M., & Unberath, M. (2023). Synthetic data accelerates the development of generalizable learning-based algorithms for X-ray image analysis. Nature Machine Intelligence, 5(3), 294-308.

[4] Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., ... & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. Advances in neural information processing systems, 35, 24824-24837.

An Analysis of the Value Bias of Large Language Models based on the PERMA Theory

Chaewon Kim[1], Keummin Ka[1], Jinhong Jeong[1], Seungbeen Lee[1], *Youngjae Yu[2]
[1] Yonsei University, *[2] Seoul National University

chewon1227@yonsei.ac.kr, kummin0429@yonsei.ac.kr, jjhsnail0822@yonsei.ac.kr,
iyy1112@yonsei.ac.kr, *youngjaeyu@snu.ac.kr

Abstract

The increasing usage of large language models (LLMs) has led to concerns that value biases in LLMs may affect users' own decision-making and self-exploration processes. Our study clarified whether LLMs inherently overestimate or underestimate certain value elements based on the PERMA theory, a representative well-being theory of psychology, and suggests the Balance Game framework to analyze LLMs' value bias. The experiment was conducted at two levels – the definition-level and the world-level. *Meaning* and *Accomplishment* were consistently evaluated as more important at the definition-level, while *Positive Emotion* and *Relationship* were more important at the world-level. In addition, a pattern in which a specific value strongly prevailed was identified in numerical questions, such as 3 more self-care days surpassing having 9 more supportive friends. These results suggest that LLM may not guarantee value neutrality in the well-being domain, and in counseling and decision-making. This work presents a new framework for systematically measuring value bias which can be applied to various domains.

I. Introduction

Users use and rely more often on Large Language Models (LLMs) for personal decisions, career recommendations, and setting their own life goals. Previous studies have pointed out that LLMs reflect certain value biases [1, 3, 4, 8]. The problem of unique value biases within LLMs has become a controversial issue, especially when using LLMs, as users have to make decisions. LLM's unique value bias can cause several problems:

Distortion of self-exploration. Making decisions and setting life goals is a process of self-exploration based on one's own value system. If LLMs emphasize certain values, users can be guided down a path that does not reflect their true preferences [6]. Instead of supporting true self-exploration, these models can lead the user's decisions to certain values that they inherently possess [5].

Strengthening the social value hierarchy. LLMs' predictions of value bias can reinforce or weaken certain values at the social level[4]. In sensitive areas such as counseling, education, and career decisions, these effects risk reducing diversity and overamplifying certain value systems.

To address these concerns, our study uses PERMA Theory [7], a popular framework proposed by Martin Seligman, founder of Positive Psychology, to investigate whether LLM is biased toward well-being. Well-being involves key dimensions that individuals consider when making major life decisions. PERMA defines well-being through five key dimensions: *Positive Emotions* (P), *Engagement* (E), *Relationship* (R), *Meaning* (M), and *Accomplishment* (A).

This study aims to make two contributions:

Conceptual Level Analysis. We use definition and measurement to analyze value bias between the 5 PERMA value factors. By this, we can investigate whether they consistently emphasize certain well-being factors. This allows us to identify how LLM conceptualizes and prioritizes different aspects of well-being.

Comparison with real-world data. We map PERMA value elements to real-world data to determine LLMs' value bias in realistic situations. This comparison can show whether LLM responses are consistent with balanced well-being or exhibit systematic deviations to specific dimensions.

II. Related Works

Glickman & Sharot showed that AI amplifies human biases more than human interactions, which makes LLMs inappropriate for counseling [4]. Liu et al. demonstrated that LLMs consistently prioritize *Universalism* and *Benevolence* while undervaluing *Power* and *Hedonism* across social decision-making scenarios [5]. This means that LLMs show specific value biases, which can influence individual decision-making.

Despite various LLM bias research studies, there are no studies that take the psychological concept of 'well-being' to evaluate LLM bias patterns systematically, which is directly connected with the LLMs' usage for self-exploration.

III. Experimental Design

To examine whether LLMs show systematic value bias among the PERMA values, we designed a Balance Game framework (e.g., positive *Relationship* and low *Accomplishment* vs. negative *Relationship* and high *Accomplishment*). Since the Balance Game Questions include both values but are manipulated to the extent of each value asymmetrically, this framework allows for a direct and focused comparison that isolates the influence of each value more clearly.

To cope with various perspectives on PERMA, we constructed 4 types of Balance Game questions – word-based, factor-based, Reddit-based Textual (which contains positive and negative factors), and Reddit-based Numeric questions. Here, the first two (word-based questions and factor-based) are set as definition-level, and the Reddit-based Textual Questions are set as world-level.

Definition-level contains two types of questions. See Table 1.

Word Questions The five PERMA value elements were presented directly at the word level.

Factor Questions Factors that represent each PERMA value element were derived from validated psychological scales used to measure PERMA value elements [2]. For each element, 3 representative sub-factors were used to capture more accurate aspects.

World-level contains the Reddit-based Textual Questions.

Reddit-based Textual Questions Questions were generated from real user posts in communities such as r/findapath, r/getmotivated, r/employment, and r/jobs. Unlike word-level definitions, these items reflect lived experiences, including both positive and negative accounts. See Table 2 for examples.

Types	Question	
	Option 1	Option 2
Word-based	positive *Relationship* and low *Accomplishment*	negative *Relationship* and high *Accomplishment*
Factor-based	Always been *feeling loved* and never *achieve the important goals*	Never been *feeling loved* and always *achieve the important goals*

Table 1. Examples of the definition-level questions

Types		Question	
		Option 1	Option 2
Reddit-based Textual	**Pos**	positive *Relationship* and low *Accomplishment*	negative *Relationship* and high *Accomplishment*
	Neg	Always been *feeling loved* and never *achieve the important goals*	Never been *feeling loved* and always *achieve the important goals*

Table 2. Examples of the world-level questions

For quantitative comparison, we constructed numeric questions.

Reddit-based Numeric Questions Experiences drawn from Reddit were quantified to each value so that responses could be compared in explicitly numerical terms. We varied the numerical part of the sentence so that we could compare values in the quantity dimension. See examples for each value at Table 3.

Values	Numeric Questions
Positive Emotion	Have self-care {n} days per week
Engagement	Work on {n} passion projects
Relationship	Have {n} supportive friends
Meaning	Feel clear about life direction for {n} days per week
Accomplishment	Hired at {n} companies

Table 3. Reddit-based Numeric Questions

We selected the 5 most important *keywords* based on the PERMA theory (well-being, balanced, flourishing, fulfillment, and thriving) and provide a prompt asking the choice and reasons for choosing between the two options to live a *keyword* life.

To minimize order bias and assess the robustness of the results, each question was presented in both its original and reversed versions.

IV. Results

We tested three Qwen2.5 models (7B, 14B, 32B) [9] and results are shown at Figure 1.

Meaning and Accomplishment were emphasized at the definition level. At the definition level, *Meaning*, *Relationships, Accomplishment* were consistently rated higher, winning a larger share of pairwise comparisons. By contrast, *Positive Emotion* and *Engagement* were rated lower. In other words, at the definitional stage, the LLM showed a clear tendency to treat *Meaning* as an important value.

Positive Emotion and Relationship were emphasized in real-data settings. *Positive Emotion* and *Engagement* were rated higher overall, winning more comparisons, while *Meaning* and *Relationships* were rated lower. Two key observations emerge here. First, the same results appeared in both the positive and negative versions, indicating consistency across framing conditions. Second, the outcomes were the exact opposite of those found at the definition level. That is, unlike the definition-level tasks, the world-level tasks revealed a tendency for the LLM to judge *Positive Emotion* and *Relationship* as more important than *Meaning* and *Accomplishment*.

Numerical variations revealed conditional trade-offs across concrete life scenarios. For example, 3–5 days of self-care *(Positive Emotion)* were judged roughly equivalent to 4–5 supportive friends *(Relationship)*, while 1 project with passion *(Engagement)* was valued similarly to 8–9 supportive friends *(Relationship)*. In contrast, even 2 meaningful days *(Meaning)* were enough to outweigh multiple company offers *(Accomplishment)*, underscoring the dominance of meaning once present.

Big models also show value orientations. This finding aligns with the prior study [3], which states that LLMs exhibit consistent biases regardless of model size.

These results show that the differences between people's expressed preferences and actual choices in real situations [10] are reflected in the trained data, which strongly influences the LLMs' inherent value system. Therefore, LLM can reproduce a dual and biased value orientation that underlines Positive emotion and Relationship in practical advice while giving answers that underline Meaning and Accomplishment on the surface. If this orientation is applied in the context of consulting, it has the risk of distorting values inherently learned by the model instead of giving a chance of self-exploration.

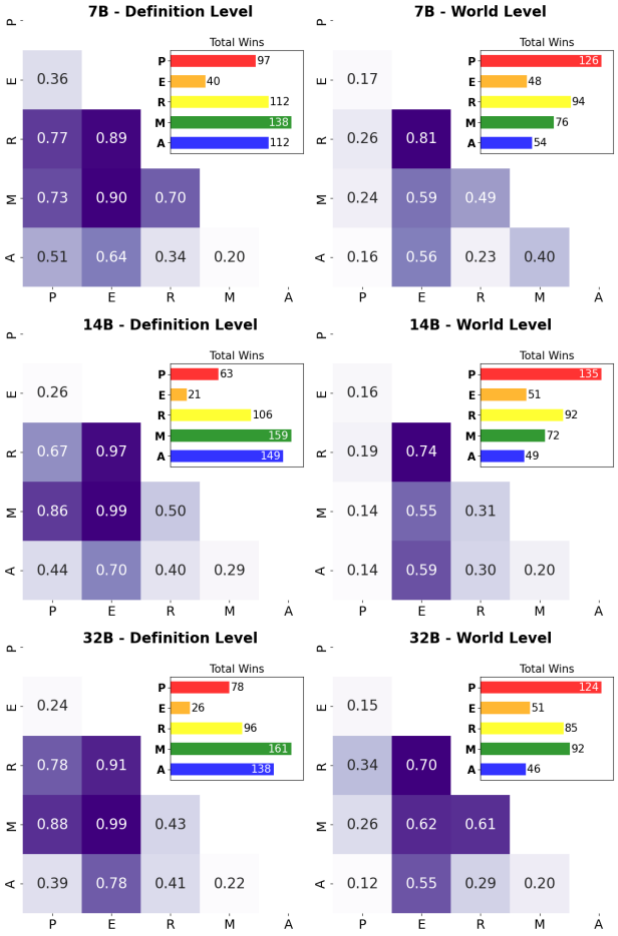


Figure 1. Balance Game Results

V. Discussion

Our study proposes a novel bias detection framework by incorporating existing psychological frameworks and AI evaluations. Balance game approaches provide relative value preferences rather than absolute acceptance or rejection. This framework can be applied to investigate bias across different value frameworks beyond PERMA.

Our study detected the value bias, which can be magnified into systematic patterns in the response of LLMs, by providing evidence of PERMA-based value bias directions in LLMs. The finding that LLM shows a clear preference for certain well-being value elements is important for use in the context of self-exploration processes such as career guidance and decision-making in life.

Reference

[1] Bai, X., Wang, A., Sucholutsky, I., & Griffiths, T. L. (2025). Explicitly unbiased large language models still form biased associations. *Proceedings of the National Academy of Sciences*, *122*(8), e2416228122.

[2] Butler, J., & Kern, M. L. (2016). The PERMA-Profiler: A brief multidimensional measure of flourishing. International journal of wellbeing, 6(3).

[3] Cheung, V., Maier, M., & Lieder, F. (2025). Large language models show amplified cognitive biases in moral decision-making. *Proceedings of the National Academy of Sciences*, *122*(25), e2412015122.

[4] Glickman, M., & Sharot, T, (2025). How human-AI feedback loops alter human perceptual, emotional and social judgements. *Nature Human Behaviour, 9(2),* 345-359.

[5] Bulla, L., De Giorgis, S., Mongiovì, M., & Gangemi, A. (2025). Large Language Models meet moral values: A comprehensive assessment of moral abilities. *Computers in Human Behavior Reports, 17,* 100609.

[6] Vicente, L., & Matute, H. (2023). Humans inherit artificial intelligence biases. *Scientific reports, 13(1),* 15737.

[7] Seligman, M. E. (2011). *Flourish: A visionary new understanding of happiness and well-being*. Simon and Schuster.

[8] Hadar-Shoval, D., Asraf, K., Mizrachi, Y., Haber, Y., & Elyoseph, Z. (2024). Assessing the Alignment of Large Language Models With Human Values for Mental Health Integration: Cross-Sectional Study Using Schwartz's Theory of Basic Values. *JMIR mental health*, *11*, e55988.

[9] Qwen Team. (2024). Qwen2.5: A party of foundation models.

[10] De Corte, K., Cairns, J., & Grieve, R. (2021). Stated versus revealed preferences: An approach to reduce bias. *Health economics*, *30*(5), 1095-1123.

스타일 특징 추출 기법 융합을 통한 미술 화풍 분류 RAG-LLM 시스템

오현찬, 김용화, 윤상필, 김영민
한국전자기술연구원

{hyeonchan.oh, yonghwa.kim, dkssyd000, rainmaker}@keti.re.kr

An RAG-LLM System for Art Style Classification through the Fusion of Style Feature Extraction Techniques

Hyeonchan Oh, Yonghwa Kim, Sangpil Yoon, Youngmin Kim
Korea Electronics Technology Institute(KETI)

요 약

본 논문은 특정 화가의 작품을 식별하고 분류하기 위한 시스템 구축을 위한 검색 증강 생성(RAG) 프레임워크를 활용한 작품 식별 및 식별된 작품의 선정 근거를 설명할 수 있는 시스템을 제안한다. 기존의 영상처리 기법은 이미지 및 영상이 가지는 의미론적 콘텐츠 정보 추출에 강력한 성능을 보여주었다. 그러나, 붓 터치, 질감, 색감, 구조 등의 작가의 정체성이라고 볼 수 있는 스타일적 특징(Stylistic features)을 추출 및 분류하기에는 어려움이 있다. 본 연구는 이를 극복하는 분류기 구성을 위하여 콘텐츠 및 스타일 특징 정보를 복합적으로 활용한 하이브리드 벡터 구조를 제안하였으며, 분류 근거를 설명하기 위한 거대언어모델(LLM)을 적용하여 최종적인 시스템 구조를 설계하였다. 최종적으로 기존의 이미지 특징 정보 추출 기법을 활용한 분류 방식보다 스타일 특징을 복합적으로 사용한 하이브리드 벡터 기법이 정확도(Accuracy)는 63.21%에서 77.23%로 약 14%p, F1-Score 는 53.65%에서 73.38%로 약 20%p 향상된 성능을 보이는 것을 확인하였다.

Ⅰ. 서 론

인공지능을 활용한 이미지 인식 및 분류 기법은 CLIP(Contrastive Language-Image Pre-Training)[1]과 같은 멀티모달 모델을 위한 정보 추출을 통하여 이미지와 텍스트를 함께 이해하는 방향으로 확대되고 있다. 또한, 거대언어모델(LLM)과 검색 증강 생성(RAG, Retrieval-Augmented Generation)과의 연계를 통한 데이터 분석 및 분류를 위한 방법이 대두되고 있다.

이 같은 검색 증강언어모델의 미술 작품 분석 및 분류 사용은 기존의 CLIP 과 같은 정보 추출로는 어려움이 있다. 기존의 CLIP 기반 시스템은 작품의 주제나 객체와 같은 콘텐츠 정보를 추출하는 데는 우수한 성능을 보이지만, 작가 고유의 화풍(artistic style)을 분류하기는 한계를 보인다. 이는 같은 콘텐츠를 작업하여도, 작가에 따라 화풍이 달라지는 미술 작품의 특성이 고려되지 않았기 때문이다.

따라서, 본 논문에서는 이러한 한계를 극복하고자 콘텐츠와 스타일 특징을 모두 고려하는 하이브리드 임베딩 RAG 시스템을 제안한다. 제안 모델은 CLIP 이미지 인코더 및 그람 행렬(Gram Matrix)을 활용하여 콘텐츠 정보와 스타일 특징 정보를 복합적으로 사용하며, 융합된 특징 정보를 기반으로 보다 정교한 화풍 분류 및 설명이 가능하도록 설계되었다.

Ⅱ. 본론

제안된 시스템은 입력된 이미지의 임베딩을 수행하는 하이브리드 임베딩 단계와 하이브리드 임베딩을 데이터베이스의 정보에서 검색하는 임베딩 검색 단계, 검색한 결과들을 분류하여 가장 적합한 결과를 도출하는 결과 투표 단계 그리고 최종 선정된 결과에 대한 근거를 도출하는 거대언어모델기반 분석 단계로 구성된다.

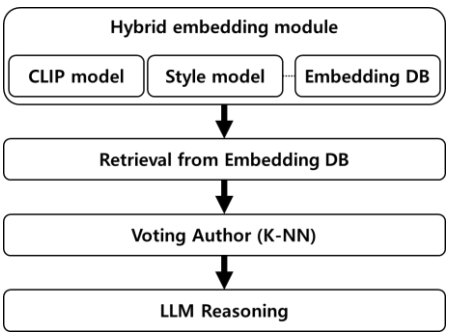


Figure 1. 제안 시스템의 전체 구조

하이브리드 임베딩 단계는 CLIP 인코더와 스타일 인코더를 기반으로 구성된다. CLIP 인코더는 총 768 차원의 벡터로 이미지가 내포한 작품의 주체 및 객체 정보를 담아낸다. 스타일 인코더는 VGG 기반의 특징맵(featuremap)으로부터 그람 행렬(Gram Matrix)[3]를 구성한다. 그람 행렬은 이미지의 질감, 색상, 패턴, 붓 터치와 같은 시각적 스타일 정보를 효과적으로 표현하기 위한 방법[3]으로 완전 연결

계층(Fully Connected Layer)를 이용하여 128 차원의 벡터로 표현된 스타일 정보로 변환된다. 완전연결계층을 통한 128 차원의 스타일 정보를 학습하기 위하여 Triplet loss 를 적용하여, 동일 작가의 작품과 그 외의 작가의 작품 분류를 위한 학습을 진행하였다. 이후, 두 벡터를 융합하여 896 차원의 하이브리드 임베딩을 구성하며, 이는 임베딩 데이터베이스의 구축에 활용된다.

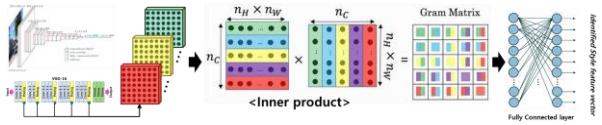


Figure 2. 그람 행렬 기반 스타일 인코더

검색 단계에서는 임베딩 기반 데이터베이스에서 가장 유사한 화풍을 검색하며, 단순히 최적 결과만을 도출하는 것이 아닌 상위 20 개의 정보를 검색한다. 검색된 상위 20 개 결과는 K-최근접 이웃(k-Nearest Neighbors) 분류 알고리즘 기반의 투표 시스템을 통하여, 가장 적합한 화풍 정보를 검출한다.

선정된 화풍 정보를 거대언어모델에 전달하여, 최종적으로 해당 화풍이 선정된 근거 도출 및 자연어 형태로 생성한다. 생성된 설명은 모델이 주목한 시각적 특징과 결과 상의 인과관계를 보여준다.

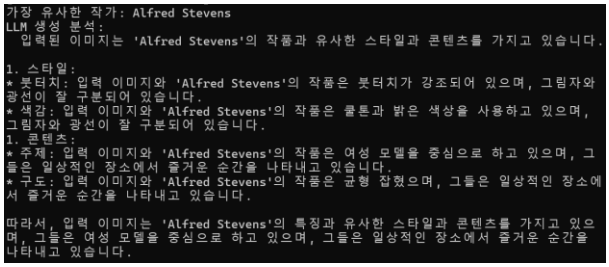


Figure 3. 유사 작가 분류 근거 설명

실험은 퍼블릭 도메인 작품(Public Domain Works)을 기반으로 구성된 총 27 명의 주요 화가(빈센트 반 고흐, 클로드 모네, 구스타프 클림트 등)의 작품 이미지로 구성된 데이터셋을 대상으로 수행하였다.

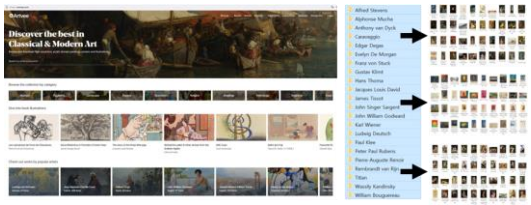


Figure 4. Public domain works from Web[2]

성능 평가를 위해 약 100 여 장의 이미지를 무작위로 추출하여 CLIP 단독 모델과 제안된 하이브리드 모델을 비교하였다. 성능 평가는 정확도(Accuracy), 정밀도(Precision), 재현율(Recall), F1-Score 의 네 가지 지표를 기준으로 수행하였다.

CLIP 단독 모델은 정확도 63.21%, 정밀도 55.28%, 재현율 62.35%, F1-Score 53.65%를 기록하였다. 반면 하이브리드 모델은 정확도 77.23%, 정밀도 78.08%, 재현율 76.23%, F1-Score 73.38%로 모든 지표에서 CLIP 단독 모델을 크게 능가하였다. 특히 종합 성능을 나타내는 F1-Score 는 약 20%p 향상되었는데, 이는 스타일 정보를 추가함으로써 모델의 변별력이 크게 향상되었음을 보여준다. 또한 기존의 CLIP 이미지 인코더 기반의 분석 모델이 종종 유사한 주제를 그린 다른 화가의 작품을 혼동한 반면, 스타일적 특징(Stylistic features)를 복합적으로 사용한 하이브리드 모델은 이러한 오류를 상당 부분 보완하는 것을 확인하였다.

Table 1. CLIP 및 하이브리드 모델의 분류 성능

모델 구분	CLIP 단독	하이브리드	성능향상
정확도	63.21%	77.23%	22.18% (+14.02%p)
정밀도	55.28%	78.08%	41.24% (+22.80%p)
재현율	62.35%	76.23%	22.26% (+13.88%p)
F1-Score	53.65%	73.38%	36.78% (+19.73%p)

Ⅲ. 결론

본 논문에서는 CLIP 기반 화풍 분류의 한계를 극복하기 위해 스타일 특징 추출기를 결합한 하이브리드 임베딩 기반 RAG-LLM 시스템을 제안하였다. 실험 결과, 제안 모델은 CLIP 단독 모델 대비 모든 성능 지표에서 크게 향상된 성능을 보였으며, 특히 스타일 정보를 반영한 접근법이 작품의 화풍 분류에 효과적임을 확인하였다. 이 연구는 미술사 연구, 예술 교육 등 다양한 분야에서 응용될 수 있으며, 향후 Transformer 기반의 스타일 추출기를 적용 및 LoRA 과 같은 LLM 의 정밀 조정(Fine-tuning)의 추가적인 연구를 수행할 예정이다. 또한 하이브리드 임베딩 개념은 미술 작품뿐만 아니라 패션, 디자인, 건축 등 스타일이 중요한 다른 도메인에도 확장 가능성이 크다.

ACKNOWLEDGMENT

본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2024 년도 문화기술 연구개발 사업으로 수행되었음 (과제명: 멀티모달 생성형 AI 모델의 데이터셋 저작권 핵심 기술 개발, 과제번호: RS-2024-00333068, 기여율: 100%)

참 고 문 헌

[1] A. Radford et al., “Learning transferable visual models from natural language supervision,” arXiv:2103.00020, 2021.

[2] Artvee, “Artvee | Discover Classical Art,” https://artvee.com/, 2025.

[3] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image Style Transfer Using Convolutional Neural Networks," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 2414-2423, doi:10.1109/CVPR.2016.265.

도메인 특화 LLM 지식 내재화 평가 프레임워크

양동헌, 임찬욱
한국과학기술정보연구원
{yangdonghun3, chanuklim}@kisti.re.kr

Evaluation Framework for Knowledge Internalization in Domain-specific LLMs

Donghun Yang, Lim Chanuk
Korea Institute of Science and Technology Information (KISTI)

요 약

최근 자연어 처리를 비롯한 다양한 분야에서 대형언어모델(LLM)이 폭넓게 활용되고 있으며, 특히 특정 분야에 특화된 도메인 특화 LLM 연구가 활발히 진행되고 있다. 도메인 특화 LLM은 일반적으로 오픈 LLM에 도메인 특화 데이터를 추가 학습하는 방식으로 개발되며, 이 과정은 크게 추가 사전학습(Continual Pre-training, CPT)과 지도 미세조정(Supervised Fine-tuning, SFT)으로 구분된다. SFT는 주로 모델에 특정 태스크 수행 능력과 응답 양식을 부여하는 데 초점을 두며, CPT는 방대한 도메인 코퍼스로부터 지식을 내재화하는 것을 목적으로 한다. 본 연구는 이러한 관점에서, CPT된 도메인 특화 LLM이 실제로 도메인 지식을 내재화했는지를 검증하기 위한 연속 문장 생성 기반 평가 프레임워크를 제안한다. 이를 위해 한국어와 영어 논문 초록 각각 1,000건을 수집해 문장 단위로 분할하여 평가 데이터셋을 구축하였다. 구축한 데이터셋의 초록 앞부분을 입력으로 모델이 마지막 1-3문장을 예측하도록 하고, 이를 실제 초록과 비교하여 지식 내재화 성능을 평가하였다. 본 평가 프레임워크를 Gemma3-4b-pt 모델을 과학기술정보 특화 코퍼스로 CPT한 KONI-4b-base 모델에 적용하고, 그 결과를 Gemma3-4b-pt 및 Gemma3-4b-it와 비교하여 제안 프레임워크의 우수성을 검증하였다. 실험 결과, 한국어 데이터셋에서는 KONI-4B-base 모델이 가장 우수한 지식 내재화 성능을 기록하였으며, 영어 데이터셋에서는 Gemma3-4b-pt가 상대적으로 더 뛰어난 성능을 보였다. 이는 Gemma3-4b-pt가 사전학습 단계에서 이미 대규모 영어 논문 데이터를 충분히 학습하였기 때문에 KONI-4b-base에서 수행된 CPT의 추가 지식 내재화 효과가 제한적이었던 것으로 해석된다. 한편, Gemma3-4b-it 모델은 전반적으로 낮은 성능을 보였는데, 이는 SFT 과정에서의 태스크 지향적 조정이 연속 문장 생성에 제약으로 작용했음을 시사한다. 본 연구는 CPT의 지식 내재화 여부를 정량적으로 평가할 수 있는 간단하면서도 효과적인 프레임워크 제안하였으며, 이는 향후 도메인 특화 LLM 개발에서 학습 데이터 구성과 평가 방법론에 중요한 시사점을 제공할 것으로 기대된다.

Ⅰ. 서 론

최근 대규모 데이터와 파라미터를 기반으로 한 대형언어모델(LLM)은 번역, 요약, 질의응답과 같은 범용 작업뿐만 아니라 다양한 응용 분야에서도 탁월한 성능을 보이며, 자연어 처리를 비롯한 여러 영역에서 기존 패러다임을 빠르게 대체하고 있다 [1]. 특히 특정 도메인에 최적화된 LLM 개발에 대한 수요가 급격히 증가하고 있으며, 의료, 법률, 과학기술 등과 같이 전문성이 요구되는 다양한 분야에서 해당 영역의 전문 지식을 효과적으로 내재화한 도메인 특화 LLM 연구가 활발히 진행되고 있다.

도메인 특화 LLM은 일반적으로 공개된 범용 LLM을 기반으로 도메인 특화 데이터를 추가 학습하는 방식으로 개발되며, 크게 추가 사전학습(Continual Pre-training, CPT)과 지도 미세조정(Supervised Fine-tuning, SFT)으로 구분된다 [2]. SFT는 특정 태스크에 맞추어 모델 출력을 조정하고 사용자 지향적 응답 양식을 학습시키는 데 효과적이지만, 모델 내부의 지식 구조를 근본적으로 변화시키는 데에는 한계가 있다. 반면 CPT는 대규모 도메인 코퍼스를 활용하여 새로운 지식을 학습하고 내재화하는 데 초점을 두며, 특정 도메인의 언어적 특성과 지식 구조를 모델 내부에 반영하기 위한 핵심 과정이다.

이러한 CPT의 지식 내재화 능력을 평가하려는 연구가 다양하게 수행되고 있으나, 대부분 분류, 질의응답, 요약 등 특정 태스크 성능을 기준으로 한 간접 평가 수준에 그쳐, 모델이 실제로 지식을 학습했는지, 아니면 단순히 태스크 수행 능력만 향상된 것인지를 명확히 구분하기 어렵다. 또한 기존 평가 방법은 태스크 설계와 데이터셋 구성에 크게 의존하기 때문에, 모델 내부에 축적된 지식의 양과 질을 직접적으로 평가하기 어렵다는 한계가 있다.

따라서 본 연구는 CPT된 도메인 특화 LLM이 실제로 도메인 지식을 내재화했는지를 직접적으로 검증하기 위한 연속 문장 생성 기반 지식 내재화 평가 프레임워크를 제안한다. 제안 프레임워크는 논문 초록을 문장 단위로 분할하고 앞부분을 입력으로 제공하여 모델이 마지막 문장을 생성하도록 한 뒤, 이를 실제 초록과 비교하는 방식으로 지식 내재화 정도를 평가한다. 이를 통해 기존의 태스크 기반 간접 평가를 보완하고, CPT가 도메인 특화 LLM의 지식 내재화에 미치는 효과를 보다 직접적이고 명확하게 검증할 수 있는 방법론적 기반을 마련하고자 한다.

Ⅱ. 관련 연구

기존 LLM 평가 연구는 주로 분류, 질의응답, 요약 등 특정 태스크 성능 평가에 한정되어 왔다. 이러한 접근은 모델의 태스크 수행 능력을 평가 할 수는 있으나, 실제로 특정 지식을 내재화했는지를 직접적으로 검증하기에는 한계가 있다. 최근 언어모델의 지식 보유 여부를 평가하기 위해 지식 추출형 프롬프트나 상식 질의응답 벤치마크를 활용한 연구도 보고되었으나, 이는 프롬프트 설계나 특정 포맷 의존성에 크게 좌우되는 문제가 있다 [3]. 따라서 기존의 간접 평가를 넘어, 보다 직접적으로 모델의 지식 내재화를 검증할 수 있는 새로운 평가 방식의 필요성이 제기되고 있다.

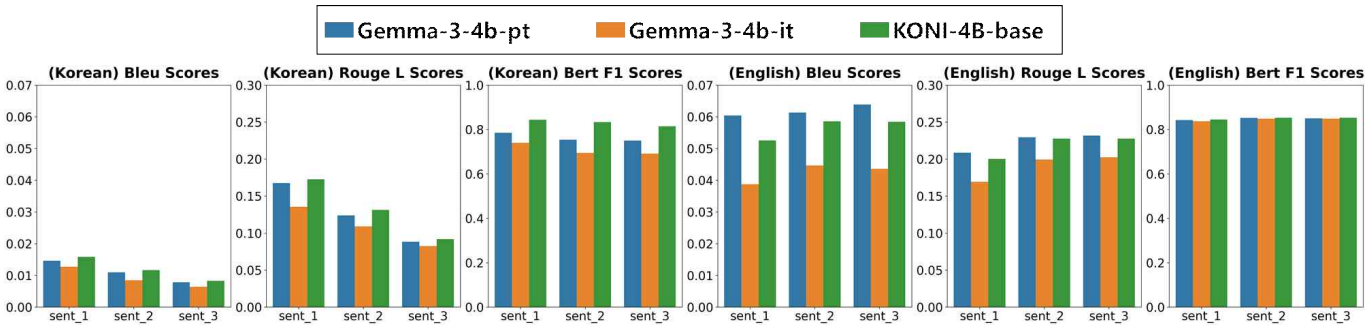


그림 1. 제안 평가 프레임워크 기반 지식 내재화 성능 비교

Ⅲ. 연구 방법

본 연구는 기존 태스크 중심 평가의 한계를 극복하고자, CPT된 도메인 특화 LLM의 지식 내재화 여부를 직접 검증하기 위한 연속 문장 생성 기반 평가 프레임워크를 제안한다.

3.1 평가 데이터셋 구축

지식 내재화 평가 프레임워크 구축을 위해 한국어와 영어 논문 초록 각각 1,000건을 수집하고 문장 단위로 분할하여 평가 데이터셋을 구성하였다. 초록 앞부분은 모델의 입력으로 제공하고, 마지막 1 - 3문장은 정답으로 활용하였다. 논문 초록은 도메인 지식이 밀집된 텍스트로서, 이를 통해 모델이 해당 지식을 실제로 내재화했는지를 효과적으로 검증할 수 있다.

3.2 연속 문장 생성 기반 지식 내재화 평가

제안 프레임워크는 입력된 초록 앞부분을 기반으로 모델이 후속 문장을 생성하도록 하고, 이를 실제 초록과 비교하는 방식으로 지식 내재화 정도를 평가한다. 이를 통해 모델이 도메인 지식을 활용해 연속적이고 일관된 서술을 생성할 수 있는지를 검증할 수 있으며, 단순 태스크 성능 중심의 기존 방법과 달리 지식 내재화 여부를 직접적으로 평가할 수 있다.

Ⅳ. 실험

4.1 실험 환경

본 연구에서 제안한 지식 내재화 평가 프레임워크는 Python 3.11, PyTorch 2.6.0, transformers 4.51.3 및 vllm 0.9.0 환경에서 NVIDIA A100 (80GB) GPU를 활용하여 구현되었다. 본 연구의 우수성을 검증하기 위해 제안 프레임워크를 Gemma3-4b-pt 모델을 과학기술정보 특화 코퍼스로 CPT한 KONI-4b-base [4] 모델에 적용하고, 그 결과를 Gemma3-4b-pt [5] 및 Gemma3-4b-it [5]와 비교 분석하였다. 지식 내재화 평가 지표로는 생성된 문장과 실제 초록 간의 유사도를 정량적으로 측정할 수 있는 BLEU, ROUGE, BERTScore를 활용하였다.

4.2 실험 결과

그림 1은 제안한 지식 내재화 평가 프레임워크를 구축한 한국어와 영어 논문 초록 데이터셋을 기반으로 세 가지 모델(Gemma3-4b-pt, Gemma3-4b-it, KONI-4b-base)에 적용하여 성능을 비교한 결과를 보여준다. 한국어 데이터셋의 경우 KONI-4b-base가 모든 지표에서 가장 우수한 성능을 기록하였으며, 이는 CPT를 통해 과학기술정보 특화 지식을 효과적으로 내재화했음을 시사한다. 영어 데이터셋에서는 Gemma3-4b-pt가 BLEU, ROUGE-L, BERTScore 전반에서 가장 뛰어난 성능을 보였는데, 이는 사전학습 단계에서 이미 대규모 영어 논문 데이터를 충분히 학습하였기 때문에 KONI-4b-base에서 수행된 CPT의 추가 지식 내재화 효과가 제한적이었던 것으로 해석된다. 한편 Gemma3-4b-it은 모든 지표에서 가장 낮은 성능을 기록하였으며, 이는 SFT 과정에서의 태스크 지향적 조정이 연속 문장 생성 능력에 제약으로 작용했음을 보여준다.

Ⅴ. 실험

본 연구는 도메인 특화 LLM의 지식 내재화 여부를 직접적으로 검증하기 위한 연속 문장 생성 기반 평가 프레임워크를 제안하였다. 제안된 프레임워크는 논문 초록 데이터셋을 활용해 모델의 연속 문장 생성 능력을 평가함으로써 기존 태스크 중심 평가의 한계를 보완하였으며, 이를 통해 CPT가 실제로 지식 내재화에 기여했는지를 정량적으로 확인할 수 있는 간단하면서도 효과적인 절차를 제공한다. 본 연구는 CPT된 도메인 특화 LLM의 지식 내재화 여부를 정량적으로 평가할 수 있는 새로운 관점을 제시하였으며, 이는 향후 도메인 특화 LLM 개발 과정에서 학습 데이터 구성과 평가 방법론 수립에 중요한 시사점을 제공할 것으로 기대된다.

ACKNOWLEDGMENT

This research was supported by the Korea Institute of Science and Technology Information (KISTI) in 2025 (No. (KISTI) 25L1M1C1), aimed at developing KONI (KISTI Open Neural Intelligence), a large language model specialized in science and technology.

참 고 문 헌

[1] A. Matarazzo and R. Torlone, "A survey on large language models with some insights on their capabilities and limitations," arXiv preprint arXiv:2501.04040, 2025.

[2] Ouyang, L., Wu, J., Jiang, X., et al. "Training language models to follow instructions with human feedback," Advances in Neural Information Processing Systems, vol. 35, pp. 27730-27744, 2022.

[3] G. Son, H. Lee, S. Kim, S. Kim, N. Muennighoff, T. Choi, C. Park, K. M. Yoo, S. Biderman, et al., "Kmmlu: Measuring massive multitask language understanding in Korean," arXiv preprint arXiv:2402.11548, 2024.

[4] KISTI, "KONI-4B-base-20250819," 2025, (https://huggingface.co/KISTI-KONI/KONI-4B-base-20250819).

[5] A. Kamath, J. Ferret, S. Pathak, N. Vieillard, R. Merhej, S. Perrin, T. Matejovicova, A. Ramé, M. Riviere, et al., "Gemma 3 technical report," arXiv preprint arXiv:2503.19786, 2025.

Reasoning LLM 내재적 추론 응답 품질 평가 프레임워크

이재용*, 양동헌, 임찬욱
*한국과학기술정보연구원
*ljy315@kisti.re.kr, yangdonghun3@kisti.re.kr, chanuklim@kisti.re.kr

Evaluation Framework for Intrinsic Response Quality of Reasoning LLMs

Jae Yong Lee*, Donghun Yang, Chanuk Lim
*Korea Institute of Science and Technology Information (KISTI).

요 약

기존 LLM 평가는 정답률 중심으로 설계되어, 모델의 내재적 추론에 기반한 응답 전개의 품질을 제대로 평가하기 어렵다. 이에 본 논문은 정답률(Accuracy), 논리성(Logic), 명확성(Clarity), 중복성(Redundancy), 평균 토큰 수(Avg_Tok), 4-그램 중복(Rep_4gram), 고유 토큰 비율(Unique_Tok) 등 7가지의 지표를 0-1로 정규화하여 동등 가중으로 통합하는 응답 품질 지수(Reasoning Quality Index; RQI) 평가 프레임워크를 제안한다. 자동 채점기(temperature=0, 고정 루브릭), 표준 JSON 스키마, 엑셀 호환 수식을 함께 제공해 재현 가능성을 확보하였다. GSM8K-Platinum 전체 1,209문항에서 DeepSeek-R1-Distill-Qwen 14B·7B·1.5B를 동일한 설정으로 평가한 결과, RQI는 각각 0.694, 0.675, 0.484로 나타났다. 7B 모델은 정확도는 낮지만, 짧은 응답과 반복 억제·어휘 다양성 측면에서 14B와 유사해 격차를 최소화하였다. 반면 1.5B 모델은 논리성과 명확성이 저하되고 응답이 장황해 RQI가 크게 하락하였다. 이는 RQI가 단일 정확도 지표로는 드러나지 않는 응답 효율성·간결성·표현 다양성의 차이를 안정적으로 구분함을 보여준다. 다만 본 연구는 수학 중심 태스크와 자동 채점기에 의존한다는 한계가 존재하며, 향후에는 다양한 규모와 구조의 모델, 도메인, 언어를 대상으로 RQI를 확장 검증하고, 응용 목적에 따라 가중치를 적용한 가중형 RQI를 정립하여 정확성·신뢰성 및 비용·지연을 중시하는 환경에 적합한 평가 기준을 마련하고자 한다.

Ⅰ. 서 론

최근 대규모 언어 모델(Large Language Model; LLM)은 자연어 생성에서 고차 추론으로 빠르게 확장되고 있다. 그러나 실제 환경에서는 대부분 질의가 별도의 프롬프트 설계 없이 주어지며, 이때 LLM의 내재적 추론으로 생성된 응답의 품질이 성능, 비용, 신뢰성에 직접적인 영향을 미친다. 그럼에도 다수의 벤치마크는 정답률 중심이며, 시스템 프롬프트(system prompt)에 의존한 접근으로 모델 고유의 응답 습관을 독립적으로 평가하기 어렵다[1]. 이에 본 연구는 LLM의 내재적 추론 응답 품질을 자동·정량적으로 평가하는 응답 품질 지수(Reasoning Quality Index; RQI) 프레임워크를 설계하고, 지표 체계, 정규화 규칙, 산출 방식 등 전 범위의 평가 파이프라인을 제시한다. 본 연구의 기여는 (i) LLM 내재적 추론 응답 품질을 반영한 지표 정의와 정규화, (ii) LLM 응답 품질에 대한 평가 합성 지수 RQI 설계, (iii) LLM 응답 품질 평가 구현 및 운영을 위한 재현성 도구 제공 등에 있다.

Ⅱ. 관련 연구

LLM의 추론 성능 평가는 GSM8K, MATH, LogiQA 등이 활용되지만, 대부분 정답률 중심으로 설계되어 응답의 전개 방식이나 표현 구조와 같은 구조적 품질 요소는 충분히 반영하지 못하고 있다. 이를 보완하기 위해 CoT(Chain-of-thought) prompting 및 self-consistency 기법 등이 제안되었으나, 프롬프트 설계에 의존적이기 때문에 모델 고유의 추론 전략을 독립적으로 평가하기 어렵다는 한계가 존재한다. 최근 LLM 응답의 구조를 항목화하여 정량적으로 평가하려는 시도들이 나타나고 있다. 특히 논리 일관성(validity), 사실 기반성(groundedness), 전개 구조(coherence), 정보 유용성(utility) 등을 기준으로 구조 품질을 평가한 연구[2], LLM 중간 추론 단계를 기준으로 중복성(redundancy)과 단계 유효성(validity)을 측정한 연구[3] 등이 대표적이다. 그러나 해당 연구들은 대부분 프롬프트 유도형 응답이나 정형화된 전개 구조를 전제로 하고 있어, 시스템 프롬프트 없이 생성된 응답에 일관되게 적용이 가능한 자동화 평가 체계가 아직 부족하다. 이에 본 연구는 LLM의 내재적 추론에 의해 생성된 LLM 응답을 대상으로 품질을 정량적으로 평가할 수 있는 자동화 평가 프레임워크를 제안하고자 한다.

Ⅲ. 연구 방법

3.1 LLM 내재적 추론 응답

본 연구는 LLM의 추론 성능을 프롬프트 설계에 의존하지 않고 정량적으로 평가하고, LLM의 내재적 추론 응답을 분석 대상으로 설정하였다. 이는 정형화된 응답 형식이나 CoT 유도 문장 등 외부 지시 요소 없이, LLM의 내재적 추론 과정에서 모델이 입력 문제에 대해 자체적으로 사고를 전개하고 응답하는 방식을 가공되지 않은 형태 그대로 평가하기 위함이다. 이때 모델은 문제 텍스트만 입력받으며 시스템 프롬프트, 역할 지시 및 형식 유도는 사용하지 않는다. 응답은 모델의 내재적 추론 및 표현 습관을 반영한 가공되지 않은 자유 응답으로 수집한다. 데이터셋은 정답 존재형(ex. GSM8K 등)이나 텍스트 기반 태스크 등으로 일반화할 수 있다.

3.2 평가 지표 정의

본 연구는 LLM 내재적 추론 응답의 전체 품질을 정량화하기 위해 표 1에서 제시된 평가 지표를 사용하였다. 본 연구에서 제안하는 평가 지표는 총 7가지이며, 정답률(Accuracy: A), 논리성(Logic: L), 명확성(Clarity: C), 중복성(Redundancy: R), 평균 토큰 수(Avg_Tok: T), 4-그램 중복 비

율(Rep_4gram: G), 고유 토큰 비율(Unique_Tok: U)이 포함된다. LLM 응답의 전반적인 품질은 RQI로 최종 산출된다.

표 1. Evaluation Metrics Overview

Metrics	Description	RQI 산출을 위한 정규화
정답률 (Accuracy: A)	• 전체 문항 중 정답 비율. 모델의 기본 문제 해결력을 나타냄 (클수록 우수)	A (그대로 사용) ∈ [0, 1]
논리성 (Logic: L)	• 근거 → 전개 → 결론의 인과·정합성, 단계 누락/비약 여부를 포함 (클수록 우수)	$L' = L \div 5 \in [0, 1]$
명확성 (Clarity: C)	• 문장 구조와 용어의 명료성/가독성 (클수록 우수)	$C' = C \div 5 \in [0, 1]$
중복성 (Redundancy: R)	• 반복·군더더기 억제 정도 (클수록 우수)	$R' = R \div 5 \in [0, 1]$
평균 토큰 수 (Avg_Tok: T)	• 응답 평균 길이 (짧을수록 효율적)	$T^* = 1/(1+Avg_Tok/\text{median}(Avg_Tok)) \in [0, 1]$
4-그램 중복 비율 (Rep_4gram: G)	• 동일 4-그램 재등장 비율 (낮을수록 반복적·군더더기 적음)	$G^* = 1 - Rep_4gram \in [0, 1]$
고유 토큰 비율 (Unique_Tok: U)	• 응답 내 고유 토큰/전체 토큰, 어휘 다양성 지표 (클수록 우수)	U (그대로 사용) ∈ [0, 1]

※ (′)는 0 - 5 원점수를 [0, 1]로 선형 변환한 값을, (*)는 '높을수록 우수' 방향이 되도록 변환·정규화한 값을 뜻한다. 평균 토큰 수(Avg_Tok)는 동일 배치의 중앙값을 기준으로 스케일 정규화하여 극단적으로 긴 응답의 영향력을 완화한다. 또한 T^*는 배치 중앙값에 상대적인 지표이므로, 평가 배치가 달라지면 T^*와 RQI도 함께 재산정된다.

RQI는 7가지 지표의 정규화값 $\{A, L', C', R', T^*, G^*, U\}$을 동등 가중 산술 평균으로 식 1과 같이 계산한다. 모든 항목을 [0, 1] 범위로 맞춤으로써 특정 지표가 점수를 과도하게 좌우하는 것을 방지하고, 해석의 일관성을 확보하였다. 여기서 $\{L', C', R'\}$는 원점수(0-5)를 5로 나눈 값으로 클수록 우수하며, $\{T^*, G^*\}$는 각각 응답이 짧을수록, 반복이 낮을수록 점수가 커지도록 변환하였다. $\{A, U\}$는 이미 [0, 1] 범위이므로 원래의 값을 그대로 사용한다. 이로써 내용 품질 신호(L', C', R'), 표현 효율 및 다양성 신호(T^*, G^*, U), 과업 성능(A)을 하나의 척도에서 균형 있게 통합한다. 최종 RQI는 [0, 1]로 보고한다. 응용 목적(ex. 정확성 우선, 비용·지연 우선 등)에 따라 가중치 벡터(ω, ∑ω=1)를 적용한 가중형 $RQI(\omega)$로 확장이 가능하나, 본 연구는 검증 단계에서 가중치 선택에 따른 과적합·해석 편차를 배제하고 프롬프트 미사용 자유 응답의 기준선을 제시하기 위해 동등 가중을 채택하였다.

$$RQI = \frac{A + L' + C' + R' + T^* + G^* + U}{7} \qquad \text{식 1}$$

3.3 프레임워크 검증

본 연구의 프레임워크는 GSM8K-Platinum 전체 1,209문항을 대상으로 DeepSeek-R1-Distill-Qwen 14B·7B·1.5B를 동일한 생성 설정에서 비교하여 검증하였다. 모든 모델에는 문제 텍스트만 입력하고 추가 프롬프트 유도는 배제했으며, vLLM(v0.3.2)과 NVIDIA A100 80GB×2 환경에서 temperature=0.6, max_tokens=2048로 응답을 수집하였다. 1단계 자동 채점기는 정답 여부와 논리성·명확성·중복성(0 - 5)을 JSON으로 기록하고, 2단계 후처리기는 평균 토큰 수(배치 중앙값 기준 보정), 4-그램 중복 비율, 고유 토큰 비율을 산출한 뒤 7개 지표를 [0, 1]로 정규화하여 동등 가중 평균으로 RQI를 계산하였다. 프레임워크의 검증 결과는 표 2와 같다.

먼저, 14B 모델은 정답률이 0.945, RQI가 0.694로 종합 성능이 가장 높았다. 이는 논리성(L'=0.936)과 명확성(C'=0.775)의 우세에 더해, 중복성 점수(R'=0.573)와 4-그램 반복 억제(G^*=0.890)가 양호하게 결합한 결과로 해석된다. 다음으로, 7B 모델은 정답률이 0.884, RQI가 0.675로 14B와의 격차가 크지 않았다. 특히 7B는 길이 보정 점수(T^*=0.524)가 14B(0.500)보다 높아 응답 길이 효율이 우수하였고, 반복 억제(R'=0.569, G^*=0.875)와 다양성(U=0.241) 측면에서도 14B(R' = 0.573, G^*=0.890, U=0.238)와 유사한 수준을 보였다. 반면, 1.5B 모델은 정답률이 0.637, RQI가 0.484로 가장 낮았는데, 논리성(L'=0.595)과 명확성(C'=0.438)의 저하와 함께 중복성 점수(R'=0.286)의 약세가 동반되었고, 길이 보정(T^*=0.400)도 낮아 상대적으로 장황한 응답 경향이 확인되었다. 결과적으로 RQI는 정확도(A)뿐만 아니라 내용 품질 신호(L', C', R')와 표현 효율·다양성 신호(T^*, G^*, U)를 균형 있게 통합함으로써, 단일 정확도 지표만으로 드러나기 어려운 장황함·군더더기·표현 단조성의 차이를 안정적으로 구분한다.

표 2. Evaluation Results

Model	정답률 (A)	논리성 (L)	명확성 (C)	중복성 (R)	평균 토큰 수 (T)	4-그램 중복 비율 (G)	고유 토큰 비율 (U)	RQI
DeepSeek-R1-Distill-Qwen-14B	0.945	4.678	3.873	2.866	964.62	0.110	0.238	**0.694**
정규화	**0.945**	**0.936**	**0.775**	**0.573**	0.500	**0.890**	0.238	
DeepSeek-R1-Distill-Qwen-7B	0.884	4.390	3.759	2.845	876.90	0.125	0.241	0.675
정규화	0.884	0.878	0.752	0.569	**0.524**	0.875	**0.241**	
DeepSeek-R1-Distill-Qwen-1.5B	0.637	2.975	2.189	1.432	1449.11	0.124	0.157	0.484
정규화	0.637	0.595	0.438	0.286	0.400	0.876	0.157	

Ⅳ. 결론 및 향후 연구

본 논문은 LLM 내재적 추론 응답 품질을 7가지의 지표를 [0, 1]로 정규화해 동등 가중으로 통합하는 RQI로 평가하는 기준선을 제안하고, 자동 채점 모듈·표준 JSON·엑셀 호환 수식을 포함한 재현 가능한 구현을 제시하였다. 본 프레임워크의 적용 결과로는 GSM8K-Platinum 전 1,209문항 검증에서 14B/7B/1.5B의 RQI는 0.694/0.675/0.484로, RQI가 단일 정확도 지표로 드러나지 않는 길이 효율, 반복 억제, 표현 다양성 차이를 일관되게 구분함을 확인했다. 그러나 본 연구는 수학 중심 태스크에 초점을 두고 자동 채점기에 의존하고 있어, 해석 범위에 신중함이 요구된다. 향후 연구로는 다양한 파라미터 규모와 모델 구조, 도메인, 언어로 대상을 확장하여 RQI의 견고성과 판별력을 체계적으로 검증하고, 응용 목적에 따라 가중치를 부여하는 가중형 RQI를 정립하여 정확성·신뢰성 및 비용·지연을 중시하는 상황들에 적합한 방안을 마련할 예정이다. 이러한 방향은 LLM의 내재적 추론 기반만으로도 비교할 수 있고, 운영 친화적인 기준선을 제공하여, Reasoning LLM 선정 및 학습에 실질적인 기여가 기대된다.

ACKNOWLEDGMENT

This research was supported by the Korea Institute of Science and Technology Information (KISTI) in 2025 (No. (KISTI) 25L1M1C1), aimed at developing KONI (KISTI Open Neural Intelligence), a large language model specialized in science and technology.

참 고 문 헌

[1] Kojima, T., et al, "Large Language Models are Zero-Shot Reasoners," Advances in Neural Information Processing Systems, 35: pp. 22199-22213, 2022.

[2] Vendrow, J., et al., "Do Large Language Model Benchmarks Test Reliability?," arXiv preprint arXiv:2502.03461, 2025.

[3] Xia, S., et al., "Evaluating Mathematical Reasoning Beyond Accuracy," Proc. AAAI Conf. on Artificial Intelligence, pp. 27723 - 27730, 2025.

디지털트윈 컨소시엄을 중심으로 한 북미지역의 AI 기술 적용 동향에 관한 연구

김영선, 오휘명, 손상우
한국전기연구원
yskim@keri.re.kr, hmoh@keri.re.kr, swson@keri.re.kr

A Study on the application trends of AI technology in North America focused on the Digital Twin Consortium

Young Sun Kim, Hui-Myoung Oh, Sangwoo Son
Korea Electrotechnology Research Institute(KERI)

요 약

북미에서는 디지털트윈 컨소시엄을 중심으로 AI와 디지털트윈의 융합이 다양한 산업에 빠르게 확산되고 있다. 제조, 에너지, 의료 등에서 AI 기반 디지털트윈이 실시간 모니터링, 예측 유지보수, 최적화에 활용되며 운영 효율성을 높이고 있다. AI 기술은 디지털트윈의 정확성·효율성을 강화하며, 특히 예측 분석, 자산 관리, 스마트시티 등 고도화된 응용 분야의 핵심 기술로 부상하고 있다. 북미는 뛰어난 기술 인프라와 산업 연계로 세계 시장을 선도하고 있다. 본 논문에서는 올해 상반기 중 개최된 디지털트윈 컨소시엄의 정회원 회의에서 발표된 북미 지역의 디지털트윈 시스템에 AI 기술 적용 동향과 실증 사례를 언급한다.

Ⅰ. 서 론

디지털트윈 컨소시엄(Digital Twin Consortium, DTC)은 2020년 Ansys, Autodesk, Bentley Systems, GE, Dell, Northrop Grumman, Microsoft 등 글로벌 리더 기업과 Object Management Group(OMG)이 주도해 설립한 국제 협의체다[1]. DTC는 업계·학계·정부 등 170여 개 단체가 참여, 디지털트윈 기술의 인식 제고, 채택 촉진, 상호운용성 표준 개발, 기술 기준 마련을 주도하고 있다. 주요 목표는 산업별 디지털트윈 프레임워크, 참조 아키텍처, 기술 언어 및 개발 환경 보급, 상호운용 프레임워크 마련 등 산업 전반의 디지털 전환 가속에 있다. 또한 국제적 동반관계, 산학연 협력 네트워크를 통해 기술 보급·확산, 모범 사례 발굴, 정책 협력 및 디지털 기술 인프라 강화에 기여하고 있다[2-4]. 올해 상반기 DTC 정회원 회의에서는 AI 기술을 결합한 디지털트윈이 핵심 기술로서 주목받았으며, 본 논문에서는 이와 관련하여 디지털트윈 시스템에 AI 기술 적용 동향과 실증 사례, 그리고 시사점을 언급하고자 한다.

Ⅱ. 본론

본론에서는 2025년 3월과 6월에 각각 개최된 디지털트윈 컨소시엄 정회원 회의를 통해 발표된 AI와 관련된 디지털트윈 기술의 동향을 소개한다.

1. AI 도입 중인 디지털트윈 생태계

- 생성형 AI 및 멀티에이전트 시스템 확대: DTC는 “AI Agent Capabilities Periodic Table(AIA CPT)”을 프레임워크로 발표하고 이를 실증 테스트베드에 적용, 다양한 산업 환경에서 AI 기반 자동화·최적화 기능을 구현하고 있다. AI는 더 이상 단일 엔진이 아닌, 복수의 인공지능 에이전트가 협력·경쟁하는 멀티에이전트 구조로 진화하는 것을 목표로 하고 있다. 생산성 향상과 자율 운영, 예측 분석, 복합 의사결정 지원이 확대될 것으로 예상된다.
- 실시간 데이터 연동과 예지 유지보수: 센서, IoT, 엣지 컴퓨팅과의 결합으로 디지털트윈이 실시간 현장 데이터를 수집·분석하고, AI 기반 예지 유지보수, 공정 최적화, 비상 대응, 설비 수명 예측 등 실행 가능한 도구를 실시간으로 적용하고 있다.
- 산업별 맞춤 AI 에이전트: 제조·에너지·도시·의료 등 업종별 특성에 따라 대화형, 복합 분석형, 완전 자율형 AI 에이전트가 단계적으로 적용되고, 조직 단위의 ‘디지털트윈 오브 오거나이제이션(DTO)’ 도입이 확산되고 있다.

2. DTC AI 전략의 프레임워크 및 평가 체계

- DTC composability 프레임워크: DTC는 디지털트윈 비즈니스 성숙도 모델(Business Maturity Model), 플랫폼 스택 아키텍처, 그리고 Capabilities Periodic Table(CPT)을 활용하여 산업현장에 최적화된 AI 디지털트윈의 구조적 도입을 지원하고 있다. 이를 통해 디지털트윈과 AI의 결합 단계, 장애 요소, 운영 성숙도(5단계) 등을 체계적으로 진단 및 평가, 피드백이 가능하게 된다.
- AI 에이전트 평가 및 우선순위 지정: AI 에이전트 CPT는 45개 AI 에이전트 핵심 역량을 도출, 우선 순위화, 프로젝트별 로드맵 및 구현 가이드를 제공한다. 이는 단순 모델이 아니라 실제 운영환경에 신속 적용 가능한 상호운용성 지침으로 활용될 예정이다.

3. 실증 중심의 Testbed 프로그램 활성화

- 개방형 실증 프로젝트: DTC는 2025년 상반기부터 회원사가 주도하는 디지털트윈 테스트베드 프로그램을 본격화했다. 테스트베드 과정에서 AI·멀티에이전트·생성형 AI 등 최신 ICT를 결합, 신기술을 안전하게 시험·검증·배포하고 산업현장에 확산시키고 있다.
- 실시간 검증과 조기 배포: 테스트베드는 신기술의 실시간 모델링, 시뮬레이션, 데이터 동기화, 통합 검증, 실전 배포까지 지원한다. 결과적으로 참여 기업은 기술 경쟁력과 시장 전환속도를 대폭 끌어올릴 수 있을 것으로 보인다.

4. 분야별 실증 적용 사례

- 가상 생산설비 및 공정 최적화: Dassault Systèmes가 주도하는 ‘Virtual

그림 1. DTC의 Capabilities Periodic Table(CPT)

Twins for Smart Factory Innovation' 테스트베드는 디지털트윈과 AI를 결합해 생산설비 전 과정을 가상으로 설계, 시뮬레이션, 문제 예측 및 이에 대한 대응을 가능하게 한다. 이를 통해 공정 혁신, 실패 비용 절감, 반응 속도가 획기적으로 향상된다고 한다.

- 디지털 마이크로공장: NTT Data사는 소형 모듈형 공장(마이크로공장) 환경에 AI 기반 디지털트윈을 접목, 생산설비의 실시간 모니터링, 반복 작업 자동화, 예측 정비, 신소재 실험, 현장 교육 등 다목적 적용 사례를 개발하였다. 컴포저블 아키텍처 기반 적용으로 신속한 확장·구성이 또한 가능하다.
- 멀티에이전트 기반 네트워크 최적화: XMPro, Microsoft가 주도한 이 테스트베드는 AI 기반 에이전트들이 기업·산업 간 네트워크상에서 표준화된 온톨로지(ontology)를 이용해 상호 통신하며, 효율적 네트워크 자원배분, 장애 감지, 자율 운영을 실현하게 한다. 산업 간, 조직 간 데이터 호환성·연결성이 대폭 향상됨을 실증하였다.
- 첨단 제조 및 정밀 생산: Rowan University와 XMPro가 협력한 디지털트윈은 IoT 지원형 3D 금속 프린터와 연동되어, 실시간 공정 진단, 예지 유지보수, 제품 품질 예측 및 폐기 감소를 실현했다. 실제로 복잡한 형상 부품 생산의 품질 및 생산성, 기계 간 상호운용성을 한층 강화하는 실증적 결과를 도출하였다.
- 생성형 AI 기반 의료: 의료분야에서는 생성형 AI와 디지털트윈을 융합해 환자 상태 예측, 맞춤형 치료계획 수립, 수술 시뮬레이션, 병원 운영 최적화 등에서 새로운 가치가 창출되고 있다. 멀티에이전트 기반 시스템이 의료진·환자·장비를 연동해 의료 서비스 품질을 높이고, 비용, 시간, 위험성을 혁신적으로 낮출 것으로 예상된다. 또한 의료 주요 기관에서는 조직 단위의 DTO를 활용해 자산 관리, 위험 모델링, 서비스 품질관리, 환자 데이터 기반 의사결정 체계 고도화를 실증 중이다.
- 스마트시티·에너지: 스마트빌딩·스마트그리드·에너지 운영 분야에서도 디지털트윈+AI 기반의 실증이 확산되고 있다. 실시간 에너지 소비·생산 모니터링, 탄소배출 예측, 대규모 인프라 고장 예지, 민간 서비스 자동화 등에서 특화된 에이전트가 배치되어 실효성을 입증하였다. 이는 실제 도시, 빌딩, 배전 계통 운영 효율 및 지속가능성 향상에 이바지할 것으로 예상된다.

5. 향후 AI 디지털트윈 기술의 전망

- 운영 자동화 및 비용 혁신: AI가 탑재된 디지털트윈은 운영의 자동화, 예측정확도 향상, 설비 고장 사전 감지, 신속한 의사결정 등을 통해 생산성과 경쟁력을 비약적으로 끌어올릴 것으로 예상된다.
- 산업 표준 및 상호운용성 확보: DTC가 선도한 CPT와 테스트베드 프로그램은 산업별 표준화와 상호운용성을 실현, 복수의 기업·플랫폼 간 협업·데이터 공유를 촉진할 것으로 보인다.
- AI가 주도하는 조직 혁신: 조직 전체의 정보·자산·프로세스를 실시간으로 반영하고, AI가 복잡한 의사결정과 자율 운영, 최적화 기능을 담당함으로써 지능화된 조직(DTO) 구현이 가속화될 것으로 보인다.
- 신시장 개척 및 ESG 확산: 자율주행, 스마트시티, 신소재, 스마트물류 등 신산업 성장과 ESG·지속가능경영, 디지털 신뢰성 기반의 사회적 가치 창출이 실제 사례를 통해 계속 입증될 것으로 예측된다.

Ⅲ. 결론

2025년 상반기 디지털트윈 컨소시엄 회의를 통해 발표된 가장 큰 변화로는 'AI 주도의 디지털트윈 진화'와 그 산업별 실증화가 있다. DTC는 프레임워크와 테스트베드 프로그램을 통해 산업 전반에 걸친 AI 디지털트윈 도입과 확장, 조직·자산·프로세스의 실시간 가상화, 그리고 산업 생태계 전체의 패러다임 전환을 이끌고 있다. 2025년 상반기 실증 사례들은 AI와 디지털트윈의 통합이 가져올 미래 혁신의 방향성을 구체적으로 제시하고 있다. 앞으로도 DTC와 참여 산·학·연 네트워크는 테스트베드 확장, 실전용 솔루션 고도화, 표준화, 사례 공유를 가속하며, AI 기반 디지털트윈 혁신이 디지털 산업전환의 중추로 자리매김할 전망이다. 이에 국내에서도 산업 간 벽을 허물어 공통의 AI 에이전트 도입, 표준 디지털트윈 공정표 등의 기준이 마련될 필요성이 있을 것으로 보인다.

ACKNOWLEDGMENT

이 연구는 2025년도 정부(과학기술정보통신부)의 재원으로 국가과학기술연구회의 지원을 받아 수행된 한국전기연구원 기본사업임(No. 25A01030)

참 고 문 헌

[1] https://www.digitaltwinconsortium.org/

[2] 김영선, 오휘명, 손상우, "북미 지역 신재생에너지 분야의 디지털트윈 기술 적용 동향에 관한 연구," BP151, 제54회 대한전기학회 하계 학술대회, 7.12-7.15, 2023.

[3] 김영선, 오휘명, 손상우, "북미 지역 디지털트윈 시스템 구현을 위한 AI 기술 적용 동향에 관한 연구," A-P-14, 제5회 한국 인공지능 학술대회, 9.25-9.28, 2024.

[4] 김영선, 오휘명, 손상우, "북미 지역 내 디지털트윈 시스템 구현을 위한 기술 적용 동향에 관한 연구," 28F-P-2, 한국통신학회 하계 종합 학술발표회, 6.18-6.21, 2025.

과학기술 도메인 특화 언어모델의 Instruction Tuning 기반 성능 향상

정창후, 장광선, 양동헌, 임찬욱
한국과학기술정보연구원 초거대 AI 연구센터

{chjeong, gsjang, yangdonghun3, chanuklim}@csu.ac.kr

Performance Enhancement of Science and Technology Domain-Specific Language Models through Instruction Tuning

Chang-Hoo Jeong, Gwangseon Jang, Donghun Yang, Chanuk Lim
Korea Institute of Science and Technology Information

요 약

본 논문은 과학기술 도메인 특화 대규모 언어모델 KONI(KISTI Open Neural Intelligence)-4B-it 의 instruction tuning 적용 결과와 성능 개선을 소개한다. KONI-4B-it 는 KONI-4B-base 모델을 기반으로 과학기술 분야에 특화된 instruction tuning 을 통해 개발되었다. 현재 한국어와 영어 이중 언어를 지원하며, instruction tuning 적용을 통해 사용자 명령어 이해 및 응답 품질이 크게 향상되었다. KONI-4B-it 모델은 다양한 한국어 및 영어 벤치마크에서 평균 57.16%의 성능을 기록하여 베이스 모델 대비 1.08%p 의 성능 향상을 달성하였다.

Ⅰ. 서 론

최근 대규모 언어모델의 발전과 함께 instruction tuning 의 중요성이 부각되고 있다. Instruction tuning 은 모델이 인간의 지시사항을 정확히 이해하고 적절한 응답을 생성할 수 있도록 하는 중요한 후처리 과정이다[1]. 본 연구에서는 과학기술 도메인에 특화된 KONI-4B-base 모델[2]에 instruction tuning 을 적용하여 KONI-4B-it 모델을 개발하였다.

Ⅱ. 본론

KONI-4B-it 는 과학기술 도메인에 특화된 continual pre-training 이 적용된 KONI-4B-base 모델을 기반으로 개발되었다. 이 베이스 모델은 40 억 개의 매개변수를 가진 트랜스포머 아키텍처로, 이미 과학기술 분야의 전문 지식을 상당 부분 포함하고 있다. 그러나 사전학습된 모델은 사용자의 구체적인 명령어를 이해하고 적절한 형식으로 응답하는 능력에서 한계를 보인다.

이러한 한계를 극복하기 위해 본 연구에서는 과학기술 도메인에 특화된 instruction tuning 을 수행하였다. Instruction tuning 용 데이터셋은 과학기술 분야의 질의응답, 문서 요약, 개념 설명, 문제 해결 등 다양한 태스크로 구성된 instruction-response 쌍으로 구축하였다. 특히 한국어와 영어 양 언어에서 균형잡힌 고품질 instruction 데이터를 확보하여 이중 언어 성능의 동시 향상을 도모하였다. 데이터셋 구축 과정에서는 과학기술 도메인의 특성을 반영한 응답 형식과 내용의 정확성을 중점적으로 고려하였다.

학습 과정에서는 Supervised Fine-Tuning(SFT) 방법론을 적용하였으며, H200 GPU 24 개를 활용한 분산 학습 환경에서 수행되었다. 분산 학습 기법과 메모리 최적화를 통해 안정적이고 효율적인 instruction tuning 이 가능하였으며, 학습 과정에서 모델의 과학기술 도메인 지식 보존과 instruction-following 능력 향상 간의 균형을 유지하는 데 중점을 두었다.

성능 평가는 한국어 벤치마크인 kmmlu, kobest, haerae, kormedmcqa 와 영어 벤치마크인 mmlu, arc, hellaswag, scholar, aidabench 등에서 수행되었다. 평가 결과, instruction tuning 적용 후 KONI-4B-it 모델은 전체 평균 성능이 56.08%에서 57.16%로 1.08%p 향상되었다. 특히 주목할 만한 점은 haerae 벤치마크에서 66.00%에서 73.33%로 7.33%p 의 큰 폭 개선을 보인 것이다. 이는 instruction tuning 이 모델의 복합적 추론 능력과 한국어 이해 능력을 효과적으로 개선했음을 보여준다. 또한 arc_challenge 에서도 2.52%p 의 의미있는 향상을 달성하여 영어권 추론 태스크에서도 성능 개선이 확인되었다.

Ⅲ. 결론

본 논문에서는 과학기술 도메인 특화 언어모델 KONI-4B-it 의 instruction tuning 적용 효과를 제시하였다. KONI-4B-base 기반의 instruction tuning 을 통해 평균 1.08%p 의 성능 향상을 달성하였으며, 특히 복합적 추론 능력에서 큰 개선을 보였다. 본 연구의 방법론은 다른 전문 도메인의 언어모델 개발에도 응용 가능할 것으로 기대된다.

ACKNOWLEDGMENT

이 논문은 과학기술정보 특화 LLM, KONI(KISTI Open Neural Intelligence) 개발을 목표로 한 2025 년도 한국과학기술정보연구원(KISTI)의 기본사업(과제번호: K25L1M1C1)으로 수행된 연구입니다.

참 고 문 헌

[1] Ouyang, L., Wu, J., Jiang, X., et al. "Training language models to follow instructions with human feedback," Advances in Neural Information Processing Systems, vol. 35, pp. 27730-27744, 2022.

[2] KISTI, "KONI-4B-base-20250819," 2025, (https://huggingface.co/KISTI-KONI/KONI-4B-base-20250819).

디지털 트윈 기반 네트워크 최적화 연구 동향

김주영, 김재민, 김가현, 최성진, 조성래
중앙대학교 컴퓨터공학과

{jykim, jmkim, ghkim, sjchoi}@uclab.re.kr, srcho@cau.ac.kr

Recent advances in the Digital Twin Based Network Optimization Research

Juyoung Kim, Jaemin Kim, Gahyun Kim, Seongjin Choi, and Sungrae Cho

Dept. of Computer Science and Engineering
Chung-Ang Univ.

요 약

디지털 트윈은 물리 세계를 가상으로 복제한 모델이다. 이는 물리 세계에 영향을 끼치지 않아 안전하게 시뮬레이션과 최적화를 할 수 있게 한다. 디지털 트윈을 통해 수집된 데이터는 인공지능을 통해 네트워크를 최적화하는 데에 사용될 수 있다. 본 논문은 네트워크의 자원 할당 최적화, 오프로딩, 경로 최적화 등의 연구 동향을 조사한다.

Ⅰ. 서 론

디지털 트윈(Digital Twin)은 물리적 시스템의 가상 복제본을 만들어 현실 세계와 가상 세계를 실시간으로 연결하여 다양한 시뮬레이션과 최적화를 안전하게 할 수 있도록 한다. 특히 실제 통신 망에 영향을 주지 않고도 네트워크 성능을 분석하고 최적의 방법을 찾을 수 있다. 따라서 디지털 트윈은 차세대 통신망, 사물인터넷, 자율주행 통신 등의 통신 분야에서 주목받는 기술이다.

특히 인공지능 기반 최적화 기법을 다룬 논문 [1], [2]는 디지털 트윈 환경과 결합되어 더 좋은 성능을 발휘할 수 있다. 본 논문에서는 디지털 트윈을 통해 실제 통신망을 가상으로 생성하고 가상 세계에서 인공지능을 통해 네트워크를 최적화하는 연구, 특히 네트워크 성능 향상, 자원 할당 최적화, 오프로딩, 경로 최적화 등의 연구 동향을 조사한다.

Ⅱ. 본론

본 논문에서는 네트워크 성능, 자원 할당, 오프로딩, 경로 최적화하는 데에 디지털 트윈을 활용하는 사례와 사용된 인공지능 기법을 조사한다.

1. 차세대 6G 통신망

6G 통신망에는 초저지연, 초고속, 초연결 등의 요구사항이 있다. 이러한 요구사항을 만족시키고 더 나은 성능을 제공하기 위해 최적화 과정이 필요하다. 이때 디지털 트윈은 실제 네트워크 가상 공간에 복제하여 다양한 AI 기반 최적화 기법을 안전하게 실험할 수 있는 환경을 제공한다.

[3]의 저자는 5G/B5G 전송망에 네트워크 디지털 트윈(NDT) 기반 경로 최적화 플랫폼을 개발하였다. 이 플랫폼은 그래프 신경망(GNN) 모델로 네트워크 트래픽을 학습하여 경로 변경 시 지연을 예측하고 초저지연 (URLLC), 초고속 (eMBB) 트래픽의 지연 예측 오차 (MAPE)를 낮춰 최적화 결정의 정확도를 보여주었다.

[4]의 저자는 대규모 5G 셀룰러 망에서 기지국 커버리지 최적화를 위해 다중 에이전트 강화학습(MARL)과 디지털 트윈을 결합하였다. 이 연구에서는 DT-SimNet 이라는 디지털 트윈 기반 시뮬레이터를 구축하여 수백 개의 기지국의 빔포밍 각도와 대역폭 등을 조정하면서 사용자 커버리지를 최적화하였고, Proximal Policy Optimization(PPO) 알고리즘을 다중 에이전트로 확장한 OMAPPO-SSP 기법을 제안하였다.

2. 엣지 컴퓨팅/IoT

디지털 트윈은 엣지 컴퓨팅과 IoT 환경에도 적용되어 최적화할 수 있다. 산업용 IoT (IIoT)와 같이 이기종 디바이스와 실시간 처리 요구를 가진 환경에서는 디지털 트윈으로 네트워크와 디바이스 상태를 정밀 모니터링하고 오프로딩 결정을 최적화할 수 있다.

[5]의 저자는 디지털 트윈 네트워크 아키텍처와 지연 모델을 구축한 뒤, 태스크 지연 최소화와 자원 할당을 목표로 오프로딩 문제를 공식화하였다. 복잡한 조합

최적화 문제를 풀기 위해 심층 강화학습(DRL) 기법 중 하나인 DDPG(Deep Deterministic Policy Gradient) 알고리즘을 활용했고, 마르코프 결정 프로세스 (MDP)로 모델링하여 실시간 동적 의사결정을 구현한다.

[6]의 저자는 다수의 이동 로봇이 공장 내 WiFi/D2D 통신을 통해 데이터를 교환하는 상황을 고려하여 물리망과 가상망을 매핑하는 엣지 컴퓨팅 기반 DTN 아키텍처를 설계하였다. 이를 통해 전체 데이터 전송률을 극대화하는 것을 목표로 하였다. MDP 기반 최적화 문제를 수립하고 마코프 결정 프로세스를 통해 다수 이동 노드의 최적 중계 노드 선택과 전송 전략을 도출한다.

IoT 환경에서 이동체의 자율 제어도 중요한 연구 분야이다. [7]은 학습 기반 협력 이동성 제어를 통해 자율 드론 배송 네트워크의 효율성을 높였으며, 이는 디지털 트윈 기반의 네트워크 최적화 연구에도 응용될 수 있다.

Ⅲ. 결론

본 논문에서는 디지털 트윈을 활용하여 네트워크 성능 최적화, 자원 할당, 오프로딩, 경로 최적화 등을 수행한 연구 동향을 조사하였다. 디지털 트윈은 물리적 시스템을 가상 공간에 복제하여 다양한 인공지능 기술을 실험하고 검증할 수 있는 환경을 제공한다. 특히 차세대 6G 통신망에서는 그래프 신경망, 다중 에이전트 강화학습과 같은 기술을 결합하여 지연 예측 및 기지국 커버리지 최적화하는 연구를 조사하였고, 엣지 컴퓨팅, 산업용 IoT 환경에서는 심층 강화학습을 통해 오프로딩과 자원 관리 문제를 효과적으로 해결하는 연구를 조사하였다.

앞으로 디지털 트윈과 인공지능의 결합은 더 다양한 통신 시나리오에서 활용될 것으로 전망되며 앞으로 향후 연구에서는 표준화, 실시간성 확보 등이 주요 과제가 될 것이다.

ACKNOWLEDGMENT

This work was supported in part by the IITP (Institute of Information & Communications Technology Planning & Evaluation) - ITRC (Information Technology Research Center) (IITP-2025-RS-2022-00156353, 50%) grant funded by the Korea government (Ministry of Science and ICT) and in part by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2023-00209125).

참 고 문 헌

[1] W. J. Yun, S. Park, J. Kim, M. Shin, S. Jung, D. A. Mohaisen, and J.-H. Kim, "Cooperative Multiagent Deep Reinforcement Learning for Reliable Surveillance via Autonomous Multi-UAV Control," in *IEEE Transactions on Industrial Informatics*, vol. 18, no. 10, pp. 7086-7096, Oct. 2022.

[2] M. C. Ho, A.-T. Tran, D. Lee, J. Paek, W. Noh, and S. Cho, "A DDPG-based Energy Efficient Federated Learning Algorithm with SWIPT and MC-NOMA," *ICT Express*, vol. 10, no. 3, pp. 600-607, Jun. 2024.

[3] R. Aben-Athar, H. Anglada, L. Costa, J. Albuquerque, A. Ferreira, C. B. Both, K. Cardoso, S. Lins, A. Silva, G. Gonçalves, I. Correa, and A. Klautau, "Network Digital Twin for Route Optimization in 5G/B5G Transport Slicing with What-If Analysis", arXiv preprint arXiv:2505.04879, May 2025.

[4] H. Liu, T. Li, F. Jiang, W. Su, and Z. Wang, "Coverage Optimization for Large-Scale Mobile Networks With Digital Twin and Multi-Agent Reinforcement Learning," in *IEEE Transactions on Wireless Communications*, vol. 23, no. 12, pp. 18316-18330, Dec. 2024.

[5] J. Wu and R. Zuo, "Intelligent Computation Offloading Based on Digital Twin-Enabled 6G Industrial IoT," *Applied Sciences*, vol. 14, no. 3, p. 1035, Feb. 2024.

[6] S. Lee and S. Park, "Mobility-Assisted Digital Twin Network Optimization over Industrial Internet of Things," *Applied Sciences*, vol. 14, no. 19, p. 9090, Oct. 2024.

[7] S. Park, C. Park, and J. Kim, "Learning-Based Cooperative Mobility Control for Autonomous Drone-Delivery," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 4, pp. 4870-4885, April 2024.

이미지 검색 애플리케이션을 위한 차원 축소 기능을 갖춘 EfficientNet 모델

가잘 사들루니아 [1], 아흐마드 이자즈 [2], 신석주 [3]
[1,3] 조선대학교, [2] 고려대학교

ghazal.saadloonia@chosun.ac.kr, ijaz@korea.ac.kr, sjshin@chosun.ac.kr(교신저자)

EfficientNet Models with Dimensionality Reduction for Image Retrieval Applications

Ghazal Saadloonia[1], Ijaz Ahmad[2], Seokjoo Shin[3]
[1,3]Chosun University, [2]Korea University

Abstract

Image retrieval is a fundamental task in computer vision that involves searching and retrieving relevant images to a query from a large collection of images. It has applications in different domains, including image search engines, and product recommendation systems. In such systems, deep learning discriminative capabilities can be leveraged; however, the main challenge is their associated computational cost. Therefore, in this work we propose a content-based image retrieval (CBIR) scheme that utilizes pretrained lightweight EfficientNet models to represent images in a high dimensional feature space for retrieval applications. Furthermore, to accelerate the similarity search, we reduce the feature vector size by implementing Principal Component Analysis (PCA) algorithm. The simulation analysis on Corel-1K data shows that our CBIR scheme maintains strong retrieval performance while reducing computational overhead.

Ⅰ. Introduction

Image retrieval is one of the fundamental tasks in computer vision that involves searching and retrieving relevant images to a query from a large collection of images. Over the years, content-based image retrieval (CBIR) algorithms are proposed to address the limitations of traditional text-based indexing methods. In CBIR, each image that is stored in the database has its visual features extracted (such as colors, texture, and shape), followed by feature matching with a query image to retrieve visually similar results. One of the main challenges in CBIR is to extract representative features from the images. Given the recent popularity of artificial intelligence mainly owing it to the success of deep learning (DL), several DL-based CBIR schemes have been proposed [1]. However, DL algorithms are generally characterized as compute intensive tasks; therefore, limiting their applications in CBIR systems.

To address the high complexity associated with DL-based image retrieval schemes, this paper proposes a CBIR scheme that utilizes pretrained EfficientNet models [2] for feature extraction. Also, it integrates principal component analysis (PCA) as a dimensionality reduction method to accelerate the similarity search.

Ⅱ. Proposed Method

Fig. 1. shows our proposed CBIR scheme that mainly consists of two modules: feature extraction and feature matching. For feature extraction, we leverage the powerful discriminative capabilities of DL. A DL model is a parameterized function f_θ that learns to update its parameters θ for mapping an input data $\mathcal{X}$ to an output $\mathcal{Y}$ as closely as possible, i.e., $\{f_\theta: \mathcal{X} \rightarrow \mathcal{Y} | f_\theta(x_i) \approx y_i\}$, by minimizing a loss function L. The model f usually has a feature extractor module consisting of multiple convolution layers, followed by a classifier module, comprised of fully connected layers. The classifier

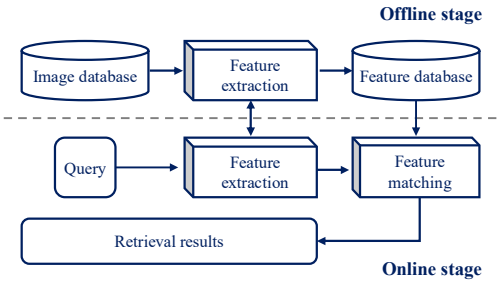


Fig. 1. A high-level illustration of proposed CBIR scheme.

Table. 1. Comparison of the Corel-1K analysis using mAP scores for pre-trained EfficientNet models, including the impact of dimensionality reduction.

Metrics		EfficientNetV1												EfficientNetV2							
		B0		B1		B2		B3		B4		B5		B0		B1		B2		B3	
Param.		4.0		6.6		7.8		10.8		17.7		28.5		5.9		6.9		8.8		13.0	
d		ϕ_d	φ_d	ϕ_d	φ_d	ϕ_d	φ_d	ϕ_d	φ_d	ϕ_d	φ_d	ϕ_d	φ_d	ϕ_d	φ_d	ϕ_d	φ_d	ϕ_d	φ_d	ϕ_d	φ_d
		1280	397	1280	395	1408	398	1536	406	1792	358	2048	371	1280	367	1280	344	1408	349	1536	360
mAP (%)	L1	0.75	0.74	0.75	0.75	0.77	0.76	0.72	0.72	0.76	0.76	0.78	0.78	0.79	0.79	0.79	0.79	0.76	0.77	0.76	0.75
	L2	0.82	0.82	0.83	0.83	0.81	0.82	0.86	0.86	0.89	0.9	0.87	0.87	0.85	0.86	0.85	0.85	0.83	0.83	0.86	0.86
	L3	0.7	0.7	0.72	0.72	0.63	0.64	0.68	0.68	0.76	0.76	0.76	0.77	0.64	0.64	0.69	0.69	0.69	0.7	0.76	0.76
	L4	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	L5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	L6	0.98	0.98	0.97	0.97	0.97	0.97	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.99	0.99	0.98	0.97	0.96	0.96
	L7	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.96	0.95	0.95	0.96	0.96	0.97	0.97	0.95	0.95	0.96	0.96	0.97	0.97
	L8	1.00	1.00	0.99	0.99	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.99	1.00	1.00	1.00	1.00
	L9	0.88	0.88	0.91	0.91	0.95	0.94	0.95	0.95	0.93	0.93	0.94	0.93	0.92	0.91	0.9	0.9	0.93	0.93	0.94	0.94
	L10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	μ	0.908	0.907	0.912	0.912	0.907	0.908	0.914	0.915	0.927	0.928	0.929	0.929	0.914	0.914	0.915	0.916	0.915	0.916	0.923	0.924

(Param.: number of parameters in million, L1 to L10 are labels in the Corel-1K dataset, μ: mean mAP score of the dataset, d: a feature vector dimensions, ϕ: Original feature vector, φ: reduced feature vector.)

module is task-specific while the feature extractor module can be shared across different tasks. Therefore, we use the classifier of pretrained EfficientNet models proposed in [2], and [3] to represent images in a high-dimensional feature space for retrieval applications e.g., $\phi \leftarrow f_\theta(\mathcal{X})$. To reduce the dimensionality of the feature space, we applied PCA on constructed feature space i.e., $\varphi \leftarrow g(\phi)$. In our simulations, we empirically choose 95% of component variance to achieve a balance tradeoff between dimensionality reduction and information loss. Finally, we implement $L1$ distance metric for feature matching as it is computationally simple, yet an efficient metric for the retrieval task.

Ⅲ. Simulation Results

Dataset. To evaluate the performance of proposed CBIR system, we use Corel-1K dataset. The dataset consists of 1,000 images uniformly divided into 10 distinct categories. The dataset contains a wide range of content, including natural scenes, animals and human-made objects. The images are of true color with a consistent size of either 256×384 or 384×256 pixels. The small size of the dataset makes it suitable for demonstrating a more realistic and effective simulation of image retrieval with pretrained models.

Evaluation metric. To quantitively evaluate our scheme, we use mean average precision (mAP) score. It measures the quality of retrieval by averaging precision scores across multiple queries, with a higher score indicating a better performance. For Q query images, the mAP score is defined as

$$mAP = \frac{1}{Q}\sum_{q=1}^{Q} AP_q\,, \tag{1}$$

where AP_q is the average precision score that calculates how well relevant images are ranked within the list of retrieved images for a single query.

Performance analysis. Table 1 summarizes the Corel-1K simulation analysis using mAP scores for EfficientNetV1 [2] and EfficientNetV2 [3] models, including the impact of dimensionality reduction. It can be observed that when comparing the same model of V1 and V2 (e.g., V1-B0 versus V2-B0), EfficientNetV2 consistently outperformed EfficientNetV1, both with and without dimensionality reduction. Thus, supporting the general consensus that V2 models are more efficient and accurate. Furthermore, PCA maintained the retrieval accuracy while potentially reducing computational costs. The slight decrease observed in mAP when applying dimensionality reduction indicated a trade-off; however, this difference is insignificant across all evaluated models. It is noteworthy, that the model performance significantly degraded for the labels of Corel-1K dataset that do not correspond to any class of the ImageNet dataset due to the lack of the learned features in the pretrained models.

Ⅳ. Conclusion

This study leveraged dimensionality reduction algorithm with pretrained lightweight deep learning models (EfficientNets) for image retrieval application. The simulation analysis showed that EfficientNetV2 is the superior model for image retrieval on the Corel-1K dataset as it offered higher mean average precision. Furthermore, when PCA technique was incorporated for dimensionality reduction, the models generally maintained strong retrieval performance while reducing computational overhead.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government. (MSIT) (RS-2023-00278294).

References

[1]. S. R. Dubey, "A Decade Survey of Content Based Image Retrieval Using Deep Learning," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 32, no. 5, pp. 2687-2704, May 2022.

[2]. M. Tan, and L. Quoc, "Efficientnet: Rethinking model scaling for convolutional neural networks." International conference on machine learning. PMLR, 2019.

[3]. M. Tan, and L. Quoc, "Efficientnetv2: Smaller models and faster training." International conference on machine learning. PMLR, 2021.

NISQ 시대에 인공지능을 위한 양자-고전 하이브리드 컴퓨팅 소프트웨어 스택 분석 및 발전 방향

김남규, 전인호
한국과학기술정보연구원
ssgyu@kisti.re.kr, inojeon@kisti.re.kr

Analysis and Development Directions of Software Stacks for Quantum-Classical Hybrid Computing for Artificial Intelligence in the NISQ Era

Nam Gyu Kim, Inho Jeon
Korea Institute of Science and Technology Information

요 약

양자 컴퓨팅과 인공지능의 융합은 기존 기술의 한계를 극복할 가능성을 지니고 있으나, 제한된 큐비트 수와 높은 노이즈라는 현재 양자 컴퓨터 하드웨어의 한계로 인해 독자적인 양자 알고리즘 구현에 제약이 있다. 이러한 한계를 극복하기 위해 양자 프로세서와 CPU와 GPU 등 고전 컴퓨팅 자원을 결합한 양자-고전 하이브리드 컴퓨팅 패러다임이 대두되었다. 본 논문은 인공지능을 위한 양자-고전 하이브리드 컴퓨팅을 위한 소프트웨어 스택 연구 동향을 분석하고, 핵심 구성 요소와 그 특징을 알아본다. 그리고 한국과학기술정보연구원에서 준비 중인 양자-고전 하이브리드 컴퓨팅 기술개발 계획을 통해 향후 발전 방향을 제시한다.

Ⅰ. 서 론

인공지능 기술은 다양한 산업 분야에 혁신을 가져왔지만, 복잡한 데이터와 고도화된 모델에 대한 계산 자원 요구는 점점 증가하고 있다. 양자 컴퓨팅은 이러한 한계를 극복하기 위한 새로운 기술로 주목받고 있다. 특히, 인공지능과 양자 컴퓨팅의 결합인 양자 머신러닝(Quantum Machine Learning, QML)은 고전(Classic) 시스템 대비하여 계산 효율성을 달성할 수 있는 제공할 잠재력이 있다.

그러나 현재의 양자 컴퓨터 기술은 양자 프로세서(Quantum Processing Unit, QPU)의 큐비트 수가 제한적이며, 노이즈가 발생하여 순수 양자 알고리즘을 단독으로 실행하기 어렵고, 고전 컴퓨팅와 연동되어야 실용적인 사용이 가능하다. 이에 따라 양자 컴퓨터와 GPU/CPU 기반의 고전 컴퓨터를 연동한 양자-고전 하이브리드 컴퓨팅이 NISQ 시대의 현실적인 해결 방안으로 대두되고 있다.

본 논문은 이러한 NISQ 시대에서 인공지능을 위한 양자-고전 하이브리드 컴퓨팅 환경에 대한 소프트웨어 스택을 분석하고, 한국과학기술정보연구원에서 연구를 기반으로 발전 방향을 제시한다.

Ⅱ. NISQ 시대와 양자-고전 하이브리드 컴퓨팅

양자 컴퓨팅은 중첩(superposition), 얽힘(entanglement), 양자 간섭(quantum interference) 과 같은 양자역학적 현상을 활용하여 정보를 처리하는 혁신적 기술이다. 정보의 기본 단위인 큐비트(qubit) 는 고전적 비트와 달리 0과 1 상태를 동시에 가질 수 있어, 기존 고전 컴퓨터를 능가하는 병렬성을 제공한다.

그러나 현재 양자 컴퓨터는 구현가능한 큐비트 수가 수십에서 수백 개 수준이며, 큐비트가 양자 상태를 유지하며 연산을 수행할 수 있는 결맞음 시간(coherence time) 이 매우 짧아 계산 과정에서 오류 발생 가능성이 높다. 또한 모든 큐비트가 완전히 상호 연결되어 있지 않아 추가적인 연결과정이 필요하며 이로 인해 노이즈가 더욱 증가한다. 이러한 제약으로 인해 현재 시점의 양자 프로세서는 완전한 결함 허용성(fault tolerance)을 확보하지 못해 아직 대규모 복잡한 문제를 해결할 수 없으며, 노이즈가 존재하는 중간 규모의 양자 컴퓨팅, 즉 NISQ(Noisy Intermediate-Scale Quantum) 시대라고 평가된다 [1].

따라서 이러한 한계를 극복하고 실질적인 양자 이득을 확보하기 위해서는 양자 오류 정정(Quantum Error Correction, QEC) 과 양자 오류 완화(Quantum Error Mitigation, QEM) 기술의 발전이 필수적이다 [2].

Ⅲ. 양자머신러닝 모델 및 데이터 처리

양자 머신러닝(Quantum Machine Learning, QML)은 양자 컴퓨팅의 계산 능력과 머신러닝 기법을 융합하여, 기존 고전 컴퓨팅으로는 처리하기 어려운 데이터 분석과 학습 문제를 효율적으로 해결하고자 하는 방법론이다 [3].

양자 머신러닝은 양자역학적 계산 특성을 머신러닝에 적용하기 위해 다양한 기법과 접근 방식을 활용한다. 우선, 고전 데이터를 양자 시스템에서 처리할 수 있도록 변환하는 양자 상태 기반 인코딩(Quantum State Encoding) 기술이 필요하다. 이를 통해 데이터를 큐비트로 표현하면, 고차원 힐베르트 공간(Hilbert Space)에서 연산을 수행할 수 있어 복잡한 패턴이나 상관 관계를 보다 효율적으로 탐지할 수 있다.

또한, 양자 회로 기반 모델(Quantum Circuits for Machine Learning)이 중요한 연구 축으로 자리 잡고 있다. 특히, 파라메터화된 양자 게이트를 사용하여 학습 가능한 구조를 설계하는 변분적 양자 알고리즘(Variational Quantum Algorithms, VQA)이 대표적이다. 이러한 접근은 고전적 신경망과 유사한 방식으로 파라미터를 조정하며, Variational Quantum Classifier(VQC)나 Quantum Neural Network(QNN)와 같은 모델로 구현된다 [4].

현재 사용 가능한 하드웨어가 완전한 양자 컴퓨터가 아닌 NISQ 수준에 머물러 있다는 점을 고려할 때, 양자 컴퓨팅과 고전 컴퓨팅을 연동한 하이브리드 방법이 가장 실질적인 접근법으로 평가된다. 이는 양자 컴퓨터가 잘 처리할 수 있는 고차원 연산은 양자 회로에 맡기고, 나머지 전처리·후처리나 대규모 데이터 핸들링은 고전 컴퓨터가 담당하는 방식으로, 양자와 고전을 보완적으로 결합하는 형태이다. 이를 통해 다양한 한계를 보완하면서, 양자 컴퓨터가 가진 잠재적 계산 우위를 머신러닝 문제에 적용하기 위한 토대를 마련하고 있다.

Ⅳ. 양자-고전 하이브리드 컴퓨팅을 위한 소프트웨어 스택

우선 양자 컴퓨팅 소프트웨어 스택은 QPU 중심으로 구성되어 있으며, 사용자는 Qiskit, Cirq 등과 같은 양자 소프트웨어 개발 키트(SDK) 를 통해 양자 회로를 설계하고, 이를 서비스 관리 계층에 제출한다 [5]. 서비스 관리 계층은 API 게이트웨이와 스케줄러를 통해 여러 사용자의 작업 요청을 조정하고, 해당 회로를 QPU에 전달한다. 이어서 오케스트레이션 및 양자 연산 계층에서는 양자 회로를 하드웨어 특성에 맞게 변환·최적화하는 컴파일 과정과 실제 양자 게이트 실행 및 측정이 이루어진다. 마지막으로 하드웨어 제어 계층은 물리적 큐비트를 구동하기 위한 마이크로파·레이저 신호 제어, 오류 억제 및 읽기 과정을 담당한다.

반면 양자-고전 하이브리드 소프트웨어 스택은 전통적 QPU 중심 구조를 확장해 QPU와 GPU/CPU를 함께 활용한다. CUDA-Q, PennyLane, Qiskit Runtime 등 하이브리드 SDK 계층에서는 양자 회로와 고전 알고리즘을 통합적으로 작성할 수 있으며, 변분 양자 알고리즘(VQE, VQC, QNN 등)은 QPU에서 상태를 준비·측정한 뒤 고전 최적화 알고리즘 결과를 갱신하는 반복 루프로 동작한다. 또한 오케스트레이션 계층은 QPU와 고전 연산을 함께 스케줄링하고 결과를 통합·분석하며, 하드웨어 계층은 QPU와 GPU/CPU 연산 자원을 통합해 자동미분과 최적화를 가속하여 저지연 하이브리드 실행 환경을 지원한다 [6].

소프트웨어 계층	양자 컴퓨팅	양자-고전 하이브리드 컴퓨팅
SDK 계층	양자 회로 설계용 SDK를 활용하여 양자 회로를 작성하고 제출	하이브리드 SDK를 사용하여 양자 회로와 고전적 알고리즘을 통합적으로 설계·구현
서비스 관리 계층	API 게이트웨이와 스케줄러를 통해 QPU 자원 관리 및 작업 실행	QPU·GPU 자원 동시 관리, 결과 데이터 관리 기능 수행
오케스트레이션 및 연산 계층	양자 회로를 하드웨어 특성에 맞게 컴파일하고, QPU에서 게이트 실행 및 측정 수행	양자 연산과 고전 연산 간 반복 루프를 제어하고 결과를 통합·분석
하드웨어 제어 계층	QPU 제어, 큐비트 초기화·측정, 에러 억제 수행	Hybrid Orchestration 계층 추가, 양자-고전 루프 제어 및 결과 취합·분석 지원

<양자 컴퓨팅과 양자-고전 하이브리드 컴퓨팅 소프트웨어 스택 계층 비교>

Ⅴ. KISTI의 양자-고전 하이브리드 소프트웨어 스택 기술 개발

한국과학기술정보연구원(KISTI)은 다양한 양자 컴퓨팅 연구 사업을 통해 하이브리드 양자 컴퓨팅 소프트웨어 플랫폼의 설계와 구현을 추진하고 있다. 특히 KISTI는 오랜 기간 축적한 슈퍼컴퓨팅 인프라 운영 경험을 바탕으로, 국내 기술 역량에 기반한 양자 소프트웨어 국산화를 중요한 목표로 삼고 있다. 이를 위해 CUDA-Q, PennyLane, Qiskit Runtime과 같은 주요 하이브리드 양자 소프트웨어 개발 키트(SDK)를 적극적으로 도입·검증하며, 양자 회로와 고전적 알고리즘을 함께 활용할 수 있는 통합 프로그래밍 환경을 개발할 예정이다. 또한, VQE, VQC, QNN 등 양자 알고리즘의 변분적 최적화를 효과적으로 수행하기 위해 하이브리드 오케스트레이션 계층을 설계하고, QPU와 CPU·GPU 연산 자원을 효율적으로 연동할 수 있는 소프트웨어 아키텍처를 설계하는 연구를 수행하는 중이다.

Ⅵ. 결론

본 논문은 NISQ 시대에 인공지능과 양자 컴퓨팅의 융합, , 특히 양자 머신러닝을 위한 양자-고전 하이브리드 컴퓨팅 기반의 소프트웨어 스택과 연구 동향을 분석하였다. 현재의 양자 프로세서는 제한된 큐비트 수와 높은 노이즈, 짧은 양자 결맞음 시간 등으로 인해 단독으로 대규모 머신러닝 문제를 해결하기 어렵다. 이러한 한계를 극복하기 위해서는 양자 프로세서의 연산 능력과 고전 프로세서의 최적화 및 병렬 처리 성능을 상호 보완적으로 결합하는 하이브리드 접근이 필수다. 이 기술은 VQE, VQC, QNN 등 변분 양자 알고리즘을 중심으로 발전하고 있으며, 이는 양자 회로에서 얻은 측정 결과를 고전적 최적화 파라미터를 갱신하는 양자-고전 루프 구조를 통해 구현된다. 이를 지원하기 위해 소프트웨어 스택은 CUDA-Q, PennyLane, Qiskit Runtime과 같은 하이브리드 SDK, 하이브리드 오케스트레이션 계층, GPU/CPU와 QPU를 통합하는 연산 계층 등을 포함하는 방향으로 발전하고 있다.

결론적으로, 이러한 양자-고전 하이브리드 컴퓨팅 기술은 향후 양자 하드웨어의 발전과 함께 인공지능 분야에서 새로운 계산 패러다임과 응용 가능성을 열어갈 것으로 기대된다.

ACKNOWLEDGMENT

이 논문은 한국연구재단의 양자컴퓨팅 연구인프라 구축(RS-2022-NR068791)과 한국과학기술정보연구원(KISTI)의 양자정보 요소기술 실증연구 사업(K25L1M3C3-01)의 지원을 받아 수행되었다.

참 고 문 헌

[1] Preskill J., "Quantum Computing in the NISQ era and beyond," Quantum, vol. 2, p. 79, 2018

[2] Kandala A., Temme K., Corcoles A. D., Mezzacapo A., Chow J. M., & Gambetta J. M., "Error mitigation extends the computational reach of a noisy quantum processor," Nature, vol. 567, 2019

[3] Biamonte J., Wittek P., Pancotti N., Rebentrost P., Wiebe N., & Lloyd S., "Quantum machine learning," Nature, vol. 549, 2017

[4] Farhi E., Neven H., "Classification with Quantum Neural Networks on Near Term Processors," 2018

[5] Aleksandrowicz G., Alexander T., Barkoutsos P., Bello L., Ben-Haim Y., Bucher D., Cabrera-Hernandez F. J., et al., "Qiskit: An Open-source Framework for Quantum Computing," Zenodo, 2019

[6] Cerezo M., Arrasmith A., Babbush R., Benjamin S. C., Endo S., Fujii K., McClean J. R., Mitarai K., Yuan X., Cincio L., Coles P. J., "Variational Quantum Algorithms," Nature Reviews Physics, vol. 3, 2021

안무 동영상의 생성 품질 향상을 위한 Attention 구조에 관한 연구

박노갑
SK Telecom

tony.nokap.park@sk.com

A Study on the Attention Mechanism for Enhancement of Dance Video Generation Quality

No Kap Park
SK Telecom

요 약

본 논문은 기존 안무 생성 모델에서 Unet 기반 확산모델이 CLIP 이미지 임베딩만 사용하는 한계를 보완하고자, 얼굴(identity) 특성을 고도화 하여 융합하는 attention 구조를 제안한다. 제안된 구조는 CLIP 인코더에서 추출한 임베딩 이외에 별도로 InsightFace 인코더에서 추출한 임베딩을 CLIP 임베딩 공간으로 매핑하여 변환한 후 두가지의 임베딩을 함께 고려한다. 두 임베딩은 각각 분포가 다르기 때문에 평균 및 표준편차 단위의 정규화를 통해 분포를 맞춘 뒤 융합한다. 이를 통해 얼굴 특성 정보가 확산 모델의 생성 과정 내에서 주요 컨텍스트와 일관성 있게 유지되도록 하여, 얼굴 identity 보존 성능을 향상시킨다.

Ⅰ. 서 론

안무 생성은 음악의 시간적 흐름에 따라 인물의 동작 시퀀스와 공간적 구성을 설계하는 창의적이며 예술적인 과정이다. 이를 자동화하는 안무 생성 모델은 음악과 안무 간의 복잡한 상관관계를 학습하여, 주어진 음악에 자연스럽고 표현력 있게 어울리는 동작 시퀀스를 생성하는 것을 목표로 한다.

최근 연구에서는 음악 입력을 안무 시퀀스로 변환하는 Music-to-Pose (M2P) 인코더를 도입하여 모듈화된 구조로 음악과 안무 시퀀스 간의 복잡한 관계를 효과적으로 학습하였다. 그러나 Diffusion 기반 U-Net 모델이 CLIP 이미지 임베딩에만 의존할 경우, 개별 인물의 identity 을 안정적으로 유지하는 데 한계가 존재한다.

본 논문은 이러한 한계를 극복하고자 기존 U-Net 기반 확산 모델 구조를 확장한다. 특히, 참조(reference) 인물의 identity 보존을 강화하기 위해 얼굴에서 identity 특성을 추출하여 Unet 에 융합할 수 있는 새로운 attention 메커니즘을 제안한다.

Ⅱ. 모델 구조

1. Face Embedding 과 CLIP Embedding 의 병합

본 연구에서는 얼굴에서 추출한 identity feature 보존을 위해 [2] 에서 제안한 identity 임베딩 추출 파이프라인을 참고하였다. 이 파이프라인은 얼굴 탐지, 정렬, 임베딩 생성의 세 단계로 구성된다.

입력 이미지는 InsightFace 를 활용하여 얼굴을 탐지하며, 복수의 얼굴이 검출될 경우 면적이 가장 큰 얼굴을 선택한다. 얼굴 탐지에 성공하면 해당 위치에서 임베딩 벡터가 추출된다.

정렬된 얼굴 이미지 혹은 탐지된 얼굴 영역은 cross-attention 기반의 얼굴 인코더에 입력되어 512 차원 크기의 identity 임베딩이 산출된다. 이 임베딩은 얼굴의 고유 특징을 정량적으로 표현하여 다양한 얼굴 각도, 표정 변화, 조명 조건에서도 견고한 식별력을 제공한다.

생성된 identity 임베딩은 CLIP 기반 이미지 임베딩 공간으로 매핑되며, 이 과정에서 identity 임베딩과 CLIP 이미지 임베딩 간 분포 차이를 줄이고 상호 표현이 효과적으로 정렬된다.

2. Face Identity Attention Branch 도입

앞서 생성된 N 개의 identity 임베딩은 CLIP 임베딩 토큰과 Concatenation 되어 U-Net 에 입력된다. 제안하는 U-Net 구조에서는 identity 임베딩을 나타내는 마지막 N 개의 토큰을 분리하여 별도 identity attention branch 에서 처리한다(그림 1). 연구[3] 에서는 cross-attention 레이어를 수정하여 입력 특성과 독립적으로 ID 토큰/임베딩을 다양한 스케일에서 결합함으로써, 이미지 생성 과정에서 얼굴 특성의 누락 또는 왜곡 문제를 해결할 수 있음을 보였는데 이를 참조하였다.

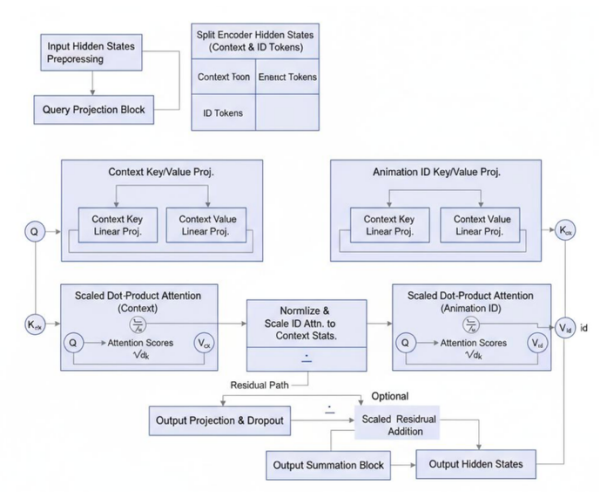


그림 1. 제안한 Attention 구조

각 attention 단계에서는 main attention branch 와 identity attention branch 가 별도로 연산되며, latent feature 의 통계치(평균, 표준편차)에 기반해 정규화(normalization) 및 스케일링 후 토큰 단위로 결합된다.

이는 identity 보존에 기여하는 정보가 기존 CLIP attention 만 지날 때보다, 확산과정에 더욱 직접적으로 주입됨을 뜻한다.

3. 분포 정규화 기반 ID 융합

각 attention 브랜치의 결과(즉, 필터링된 main latent feature h 및 identity latent feature I)는 평균과 표준편차가 다르므로 바로 더하면 밸런스를 해칠 수 있다. 따라서 identity latent feature 의 분포를 main latent feature 분포에 맞추어 정규화 해준다.

$$\mathbf{i}' = (\mathbf{i} - \mu_i)\frac{\sigma_h}{\sigma_i + \epsilon} + \mu_h$$

여기서 ϵ 은 안정화를 위한 작은 상수이다. 이후 정규화된 $\mathbf{i}'$ 에 스케일 파라미터 α 를 곱해 최종 융합 특성 $\mathbf{h} + \alpha\mathbf{i}'$ 를 얻어 최종적으로 identity 정보를 유지하면서 생성 품질 저하 및 feature 왜곡을 방지한다.

4. Face Identity 적용 영상

그림 2 는 본 논문에서 제안한 Attention 구조를 적용한 결과를 보여준다. 제안한 모델은 Reference 와 더 흡사한 영상을 생성하며, 얼굴 특성의 추가가 얼굴이외에 다른 부위의 품질도 높이는 결과를 관찰하였다. 정량적 평가는 향후 진행 예정이다.

Ⅲ. 결론

본 논문에서는 음악에 맞춘 안무 생성 과정에서 참조 인물의 identity 를 안정적으로 보존하기 위한 확장된 U-Net 기반 확산 모델 구조를 제안하였다. Identity 임베딩과 CLIP 이미지 임베딩의 효과적인 융합을 위해 새로운 attention 메커니즘인 identity attention branch 를 도입하였으며, 분포 정규화 기반 ID 융합 기법을 적용하여 identity 정보가 확산 과정에서 균형 있게 반영되도록 하였다. 실험 결과, 제안한 모델은 참조 인물의 얼굴 특성을 보다 정확히 유지하면서 안무 영상의 전체적인 품질 향상에도 기여하는 것처럼 보인다. 향후에는 정량적 평가를 통해 모델의 성능을 체계적으로 검증하고, 다양한 음악 장르 및 인물 유형에 대한 확장 가능성을 모색할 예정이다.

그림 2. 제안한 Attention 사용 Inference 결과

ACKNOWLEDGMENT

본 연구는 문화체육관광부 및 한국 콘텐츠 진흥원의 2024 년도 문화체육관광연구 개발사업으로 수행되었음 (과제명: 전통예술 가무악의 융복합 공연제작 활성화를 위한 융복합 공연 기획/제작 플랫폼 기술 개발, 과제번호: RS-2024-00398536)

참 고 문 헌

[1] Nokap Tony Park (2025), “M2PE-Diff: Music-to-Pose Encoder for Dance Video Generation Leveraging Latent Diffusion Framework”, ACM MM 2025

[2] Tu, S., Xing, Z., Han, X., Cheng, Z.-Q., Dai, Q., Luo, C., & Wu, Z. (2024). StableAnimator: High-quality identity-preserving human image animation. In Proceedings of the IEEE/CVPR 2025.

[3 Mohamed, S., Han, D., & Li, Y. (2024). Fusion is all you need: Face fusion for customized identity-preserving image synthesis. arXiv. https://doi.org/10.48550/arXiv.2409.19111

MOSA기반 전기적 인터페이스 표준화에 대한 전략적 접근

권기협+, 김재우+, 김동성*
+ICT융합특성화연구센터, *금오공과대학교
navkwon@kumoh.ac.kr, jaewookim@kumoh.ac.kr, *dskim@kumoh.ac.kr

Strategic Approaches to the Standardization of Electrical Interfaces in Modular Open Systems Approach (MOSA)

Gi-Hyeob Kwon, Jae-Woo Kim, Dong-Seong Kim*
+ICT Convergence Center, *Kumoh National Institute of Technology

요 약

Modular Open Systems Approach(MOSA)는 국방 분야에서 상호운용성과 모듈화된 기술 삽입을 촉진하기 위한 핵심 전략이다. 기존의 폐쇄적 무기체계 개발 방식은 기술 확장성과 유지보수 효율성에 한계를 드러냈으나, MOSA는 개방형 표준 인터페이스를 통해 다양한 모듈의 신속한 교체와 업그레이드를 가능하게 한다. 특히 MOSA에서 전기적 인터페이스는 전력 공급과 데이터 전송을 담당하는 기반 요소로, 플랫폼 간 상호운용성과 시스템 확장성에 직결된다. 전력 표준(MIL-STD-704, MIL-STD-1399, VITA 62 등)과 데이터 버스 규격(Gigabit Ethernet, MIL-STD-1553B, CANbus, VICTORY 등)의 적용은 안정성과 호환성을 높이고, 수명주기 비용 절감에도 기여한다. 본 논문은 MOSA의 전기적 인터페이스 측면을 분석하고, 해외 무인체계 사례와 국내 K-MOSA 연구 동향을 검토하여 향후 적용 가능성과 전략적 파급 효과를 제시한다.

Ⅰ. 서 론

현대 국방 무기체계는 UAV, UGV, USV 등 다양한 무인체계와 유·무인 복합체계(MUM-T)의 운용을 전제로 한다 [1]. 이러한 환경에서 모듈 간 통합과 상호운용성을 확보하기 위해 MOSA(Modular Open Systems Approach)가 핵심 전략으로 주목받고 있다 [1].

기존의 폐쇄적 체계는 기술 삽입과 업그레이드에 한계가 있었으나, MOSA는 개방형 인터페이스를 기반으로 시스템의 유연성과 지속 가능성을 확보한다 [2][3]. 특히 전기적 인터페이스는 전력 공급과 데이터 교환을 담당하는 핵심 요소로, 플랫폼 간 호환성과 실시간 데이터 융합을 가능하게 한다 [4].

미국은 이미 FACE, SOSA, VICTORY 등 MOSA 표준을 운용하고 있으며 [2][3][4], NATO 역시 STANAG 기반 국제 표준화를 추진 중이다 [9]. 국내에서도 K-MOSA를 중심으로 함정·차량·항공기 체계에 MOSA 적용을 확대하고 있다 [12]. 본 논문은 이러한 흐름 속에서 MOSA 전기적 인터페이스의 표준과 적용 사례를 검토하고, K-MOSA 체계에서의 기술적·전략적 활용 가능성을 논의한다.

Ⅱ. 본론

2.1 전력 인터페이스 표준

MOSA 기반 체계에서 전력 인터페이스는 모든 모듈의 안정적 운용을 보장하는 핵심 기반이다. 주요 표준으로는 항공기 전력 규격인 MIL-STD-704 [5], 함정 전기설비 규격인 MIL-STD-1399 Sect. 300 [6], VPX 기반 전원 모듈 규격인 VITA 62 [7], 그리고 국제 선박용 전기설비 표준인 IEC 60092 [8]가 있다. 이들 규격은 플랫폼 간 전력 호환성을 확보하는 동시에, 모듈 교체와 확장을 가능하게 하여 MOSA 구조에 적합하다.

특히 이러한 규격들은 단순한 전력 공급 기능을 넘어 안전성·내구성·환경 내성(EMC/EMI 내성)을 보장하도록 설계되었으며, 모듈화된 전원 공급장치(PSU)와 결합해 플러그 앤 플레이 기반 운용을 가능하게 한다. 최근에는 전력 관리 시스템(PMS) 및 지능형 배터리 관리 시스템(BMS)과 연동하여 에너지 효율성과 플랫폼 생존성을 동시에 고려하는 연구도 활발히 진행되고 있다 [13].

이러한 전력 인터페이스 표준을 국방 획득 단계에서 의무 반영하도록 제도화할 필요가 있다. 더 나아가 IEC 계열의 민·군 겸용 규격을 적극 도입함으로써 글로벌 방산 시장에서의 상호운용성을 확보하는 것이 전략적으로 중요하다. 또한 국내에서는 K-MOSA 차원에서 전력 모듈이 국제 규격과 호환되는지를 평가할 수 있는 모듈형 전력 인증 체계(Modular Power Certification System)를 마련하여, 개발 단계부터 시험·검증을 자동화하는 것이 필요하다. 이는 향후 무인체계 전력 인프라의 국제 경쟁력 확보에도 크게 기여할 것이다.

2.2 데이터 버스 인터페이스 표준

데이터 버스는 센서·무장·통신 모듈을 연결하는 핵심 통로로서, MOSA의 상호운용성 구현에서 가장 중요한 역할을 한다. 대표적으로 Gigabit Ethernet은 대역폭이 넓어 영상·신호·센서 데이터를 실시간 전송할 수 있으며 [10], MIL-STD-1553B는 이중화 구조를 통해 높은 신뢰성을 보장하지만 전송속도가 제한적이다 [5]. 또한 CANbus/NMEA 2000은 차량 및 해상 항법 장치에서 모듈 상태 모니터링에 적합하며 [4], VICTORY는 차량 플랫폼에서 통신·전원·관리 버스를 통합 관리하는 MOSA 표준으로 자리 잡고 있다 [4].

최근에는 이러한 표준들을 하이브리드 데이터 버스 구조로 통합 운용하는

방식이 각광받고 있다. 즉, Gigabit Ethernet이 대용량 데이터를 처리하고, MIL-STD-1553B와 CANbus가 임무 제어와 안정성을 보완하는 방식이다. 이러한 구조는 단순히 데이터 전송 속도와 안정성을 결합하는 수준을 넘어, 실시간 임무제어와 대용량 데이터 처리의 병행을 가능하게 하여 자율 무인체계의 효율성을 크게 향상시킨다 [14].

국방 차원에서 국가 MOSA 데이터 버스 아키텍처 레퍼런스 모델을 수립하여, 모든 무기체계가 공통의 데이터 관리 규격을 따르도록 유도할 필요가 있다. 또한 NATO STANAG 규격과의 매핑을 체계적으로 추진함으로써 국제 연합작전에서의 호환성을 확보할 수 있다. 특히, 향후에는 Gigabit Ethernet·1553B·CANbus와 같은 물리적 버스를 소프트웨어 정의 방식으로 통합 관리하는 멀티-레이어 데이터 버스 가상화(Multi-layer Bus Virtualization) 개념을 적용하여, 다양한 플랫폼에서 데이터 전송 체계를 유연하게 운영하는 방안이 요구된다.

2.3 적용 사례 분석

대표적인 적용 사례로는 미 해군 Sea Hunter USV가 있으며, 이는 Gigabit Ethernet 기반 센서 네트워크와 MIL-STD 전원 규격을 적용해 원격 자율운항과 장거리 작전을 가능하게 하였다 [10]. 또한 이스라엘의 Seagull USV는 Link-16, LTE, 위성통신을 결합한 다층 데이터 링크 구조를 운용하며, 내부적으로 CANbus를 통해 임무 모듈을 통합하여 효율적인 다임무 수행을 지원한다 [11]. 국내에서는 K-MOSA 연구가 진행 중으로, VICTORY·SOSA·FACE와 같은 국제 MOSA 표준을 벤치마킹하고, 한국형 전력·데이터 인터페이스 규격을 정립하여 차세대 전투체계 적용을 목표로 하고 있다 [12].

이러한 사례들은 MOSA 전기적 인터페이스 표준화가 단순한 규격 제정이 아니라 실질적인 전술 효과와 전략적 가치를 창출함을 보여준다. 즉, 전력 규격의 표준화는 모듈 안정성과 생존성을 보장하고, 데이터 버스 계층화는 센서·무장·C2 체계의 원활한 통합을 가능하게 하여 다영역 합동작전(MDO) 환경에서 실질적 운용성을 입증한다 [9][12].

해외 적용 사례를 참고하여, K-MOSA가 공동 운용성 시험 플랫폼(Joint Interoperability Testbed)을 구축하는 것이 필요하다. 이를 통해 국내에서 개발된 모듈과 해외 표준 간의 실질적 호환성을 사전에 검증할 수 있으며, 국제 공동작전에 기여하는 동시에 방산 수출 경쟁력 강화에도 직결된다. 더불어, 각 전력·데이터 규격의 관계를 자동으로 식별·변환할 수 있는 K-MOSA 글로벌 인터페이스 매핑 데이터베이스(DB)를 구축한다면, 연구개발자와 운용자가 손쉽게 국제 표준과의 대응 관계를 확인하고 적용할 수 있을 것이다.

Ⅲ. 결론

본 논문은 MOSA(Modular Open Systems Approach)의 핵심 요소인 전기적 인터페이스를 중심으로, 전력 및 데이터 버스 규격과 실제 군용 플랫폼 적용 사례를 분석하였다. 전력 표준(MIL-STD-704, MIL-STD-1399, VITA 62, IEC 60092)은 안정적인 전원 공급을 보장하며, 데이터 버스 규격(Gigabit Ethernet, MIL-STD-1553B, CANbus, VICTORY)은 모듈 간 상호운용성을 강화하여 모듈 교체와 기술 삽입을 용이하게 하고, 결과적으로 수명주기 비용 절감과 운용 효율성 향상에 기여한다. 또한 해외 USV 사례(Sea Hunter, Seagull)와 국내 K-MOSA 추진 현황은 전기적 인터페이스 표준화가 단순한 기술 과제를 넘어, 연합작전 수행 능력 확보와 방산 수출 경쟁력 강화라는 전략적 가치로 이어짐을 보여준다. 특히 국내에서는 국제 MOSA 표준과의 호환성을 유지하면서도, 모듈형 전력 인증 체계와 멀티-레이어 데이터 버스 가상화, 글로벌 인터페이스 매핑 DB와 같은 혁신적 접근을 통해 K-MOSA만의 독자적 경쟁력을 확보해야 한다.

향후 연구에서는 국제 표준과의 정밀한 매핑, 실 운용 시험을 통한 검증 체계 마련, 그리고 AI 및 엣지 컴퓨팅 기반 임무 모듈과의 통합 가능성 탐색이 주요 과제가 될 것이다. 이를 통해 MOSA 기반 전기적 인터페이스 표준화는 차세대 무인체계와 전투체계의 성공적 운용을 위한 기술적·정책적·전략적 기반으로 자리매김할 수 있을 것이다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터사업과 지역지능화혁신인재양성사업의 지원을 받아 수행된 연구임(IITP-2024-RS-2024-00438430, IITP-2025-RS-2020-II201612), 2025년도 정부(교육과학기술부)의 재원으로 한국연구재단의 대학중점연구소 지원사업과 기초연구사업으로 수행된 연구임(2018R1A6A1A03024003).

참 고 문 헌

[1] U.S. Department of Defense, Modular Open Systems Approach (MOSA) Implementation Guide, Washington D.C., 2019.

[2] The Open Group, Future Airborne Capability Environment (FACE) Technical Standard, Edition 3.0, 2017.

[3] Sensor Open Systems Architecture (SOSA) Consortium, SOSA Technical Standard 1.0, 2021.

[4] U.S. Army PEO C3T, Vehicle Integration for C4ISR/EW Interoperability (VICTORY) Architecture Standard, 2018.

[5] U.S. Navy, MIL-STD-704: Aircraft Electric Power Characteristics, 2016.

[6] U.S. Navy, MIL-STD-1399 Section 300: Interface Standard for Shipboard Systems, 2018.

[7] VITA Standards Organization, VITA 62: Power Supply Standard for VPX Systems, 2016.

[8] International Electrotechnical Commission (IEC), IEC 60092: Electrical Installations in Ships, 2019.

[9] NATO Standardization Office, STANAG 4586: Standard Interfaces of UAV Control System (UCS) for NATO UAV Interoperability, 2012.

[10] T. Ermakov and B. Moore, “The DARPA Sea Hunter USV: Operational Experience and Lessons for Autonomous Naval Systems,” Naval Engineers Journal, vol. 131, no. 4, pp. 73 - 85, 2019.

[11] Rafael Advanced Defense Systems, Seagull Multi-Mission USV Technical Overview, White Paper, pp. 1 - 12, 2020.

[12] DAPA (Defense Acquisition Program Administration), K-MOSA 추진 전략 및 표준화 동향 보고서, 국방기술품질원, pp. 5 - 42, 2022.

[13] O. Babayomi, R. Madonski, Z. Zhang, …, D.-S. Kim, “Robust Model Predictive Control of Converter-Based Microgrids,” IEEE Transactions on Power Electronics, early access, pp. 1 - 12, Jan. 2025.

[14] T. Huynh-The, Q. V. Pham, X. Q. Pham, T. T. Nguyen, Z. Han, D.-S. Kim, “Artificial Intelligence for the Metaverse: A Survey,” Engineering Applications of Artificial Intelligence, vol. 117, pp. 1 - 21, 2023.

MOSA 기반 무인체계용 엣지 AI 통합 프레임워크 설계

김재우+, 권익현+, 이재민*, 김동성*
+ICT융합특성화연구센터, *국립금오공과대학교 전자공학부
{jaewookim+, ihkwonaf+, ljmpaul*, dskim*}@kumoh.ac.kr

Design of a MOSA-Based Edge AI Integration Framework for Unmanned Defense Systems

Jae-Woo Kim+, Ik-Hyun Kwon+, Jae-Min Lee* and Dong-Seong Kim*
{ICT Convergence Center+, Dept. of IT Convergence Engineering*}Kumoh National Institute of Technology

요 약

본 논문은 UAV·UGV·USV 등 국방 무인체계에 적용 가능한 MOSA 기반 엣지 AI 통합 프레임워크를 제안한다. 제안하는 프레임워크는 계층형 아케텍처와 표준화된 인터페이스를 통해 모듈 교체·재사용성, 실시간 자율성, 상호운용성을 확보하였으며, 이를 통해 차세대 무인 전투체계의 유연한 전력 운용과 신속한 기술 적용을 지원할 수 있음을 분석하였다.

Ⅰ. 서 론

국방 분야에서 무인항공기(UAV), 무인지상차량(UGV), 무인수상정(USV) 등 무인체계의 역할이 증대됨에 따라, 인공지능(AI) 기술을 현장에 직접 적용하는 엣지 컴퓨팅의 중요성이 커지고 있다. 임베디드 컴퓨팅과 AI를 탑재한 엣지 플랫폼은 무인체계의 성능을 결정짓는다. 엣지 AI 기술은 센서 데이터의 수집->AI 기반 판단->제어에 이르는 전 과정을 엣지단 플랫폼 자체에서 처리한다[1]. 이는 전장 환경의 실시간 의사결정 요구에 대응하고 전장 자율성 확보 및 유·무인 복합체계(Manned-Unmanned Teaming)의 효율적인 운용에서도 필수적인 요소다.

기존 국방 무인체계 개발은 플랫폼별로 독자적인 하드웨어·소프트웨어(HW/SW) 구조를 가지는 중복성과 기술 종속, 상호운용성 결여 등의 문제가 있다. 이러한 한계를 극복하기 위해 모듈형 개방형 시스템 접근법(MOSA, Modular Open Systems Approach)이 국제적으로 주목받고 있다[2]. MOSA는 모듈화·표준화된 인터페이스를 통해 신속한 기능 교체와 업그레이드를 가능케 하며, 미국은 2019년 이후 무기체계 획득 시 MOSA 적용을 법으로 의무화했다. 한국 역시 2024년부터 K-MOSA 명명된 정책을 본격 추진하여 플랫폼 계열화와 장비 모듈화를 통해 양산성, 정비성, 전력 유연성을 높이고 있다. K-MOSA의 “계열화”란 임무 목적에 따라 대표 플랫폼을 표준화하고 이를 기반으로 무인체계를 개발하는 것을 의미하며, “모듈화”란 계열화된 플랫폼에 대해 임무 목적에 따라 필요한 임무장비를 선택 장착할 수 있도록 하는 것을 뜻한다. K-MOSA 정책을 적용하면 방산업체는 표준화된 무인 기체와 장비를 신속하고 저렴하게 양산할 수 있고, 야전 부대는 모듈화된 장비를 쉽게 교체하여 전력 조합의 유연성을 높이고 원활한 정비와 성능개량을 도모할 수 있을 것으로 기대된다. MOSA 개념은 미국의 FACE(Future Airborne Capability Environment), CMOSS(C5ISR/EW Modular Open Suite of Standards) 등의 표준과 더불어 국방 무인체계의 개방형 공통 아키텍처를 정립하려는 국제적 흐름이며, 한국형 K-MOSA 역시 이와 궤를 같이하여 다종의 무인플랫폼에 공통 적용 가능한 개방형 엣지 AI 프레임워크 수립을 지향하고 있다[3]. 본 논문은 UAV/UGV/USV에 공통 적용 가능한 HW/SW MOSA 기반 엣지 AI 통합 프레임워크 설계를 제시한다. 모듈화 및 인터페이스 표준화를 기반으로 계층형 구조와 데이터 흐름 아키텍처를 구체화하고, 적용 시나리오를 통해 모듈 재사용성과 실시간 자율성 확보 전략을 제안한다.

Ⅱ. 본론

1.MOSA와 개방형 아키텍처

MOSA는 개방형 시스템 아키텍처 구현을 위한 전략으로, 시스템을 표준화된 모듈 인터페이스로 연결해 상호운용성과 업그레이드 유연성을 확보하는 방법론이다. 핵심은 “높은 응집도, 낮은 결합도”의 모듈을 정의하고 이를 기계 가독적 표준 인터페이스로 연결하는 것이다. 공통 인터페이스 표준을 통해 모듈 교체 시 영향 범위를 최소화하고 부분 인증으로 변경을 수용할 수 있다. 미국은 이를 구체화하기 위해 FACE, CMOSS, SOSA(Sensor Open Systems Architecture), OMS(Open Mission System) 등 분야별 개방형 표준을 제정했으며, 모두 모듈화·표준화 기반의 개방형 플랫폼을 지향한다[4-6]. K-MOSA는 국제 표준을 참고해 육·해·공 무인플랫폼에 공통 적용 가능한 한국형 아키텍처 표준을 마련하는 정책이다. 대표 플랫폼 계열화와 임무 모듈화를 통해 국내 방산업체의 모듈 개발을 촉진하고, 유무인 복합체계 조기 구축과 수출 경쟁력 확보를 목표로 한다. 향후 K-MOSA 표준은 미국 MOSA 및 NATO 사례를 반영하여 규모별 표준 인터페이스, 공통 데이터 버스, AI/자율제어 모듈 인터페이스 등을 포함할 예정이며, 본 논문 제안 프레임워크에도 반영해 한국군 운용환경에 최적화할 수 있다.

2. 엣지 AI 통합 프레임워크 시스템 및 데이터 흐름

본 절에서는 MOSA 원칙에 따른 무인체계 엣지 AI 통합 프레임워크를 제안한다. 제안하는 프레임워크는 계층형 아키텍처로, 하드웨어 - 미들웨어 - 응용 계층으로 구성되며 각 계층은 표준화된 인터페이스로 결합된다. 그림1은 MOSA기반 엣지 AI 통합 프레임워크 계층형 아키텍처이다.

하드웨어 플랫폼 계층: 최하위 계층으로 무인플랫폼의 기본 하드웨어 계층이다. 기체/차체, 구동계, 배터리, 센서·액추에이터 및 임베디드 컴퓨팅 모듈로 구성된다. VPX 표준슬롯, CAN, 이더넷 등 표준 규격을 적용해 교체·확장이 용이하며, LRU(Line-Replaceable Unit) 개념으로 신속한 정비와 다중 플랫폼 호환을 지원한다. 모든 하드웨어는 MIL-STD 규격을 충족해 전장 환경에서도 운용 가능하도록 설계한다.

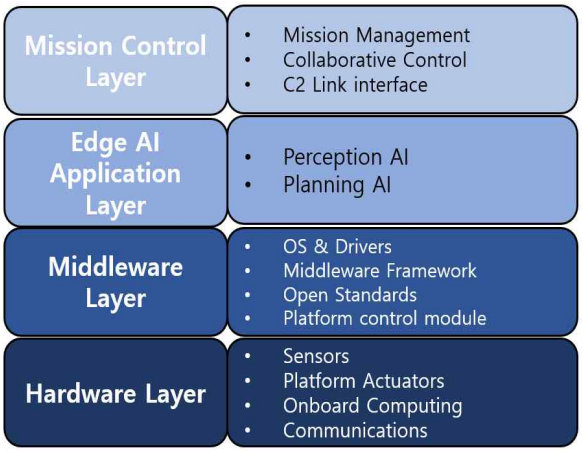
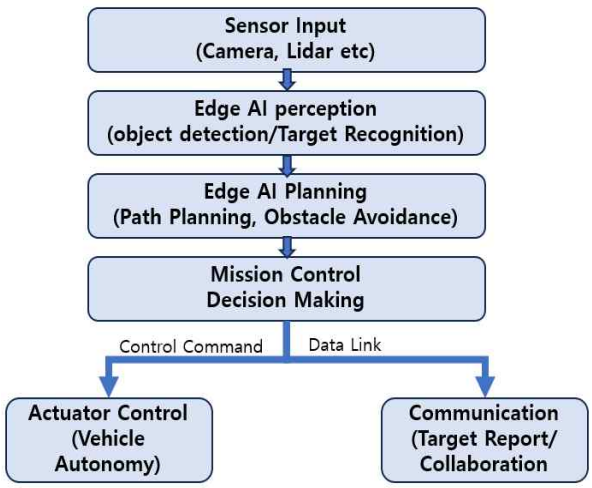


그림 1 MOSA기반 엣지 AI 통합 프레임워크 계층 구조

플랫폼 추상화 및 OS/미들웨어 계층: HAL(Hardware Abstraction Layer)는 드라이버가 다양한 센서 데이터를 표준화된 메시지로 제공하며, ROS2·DDS 기반 발행/구독 모델을 통해 데이터 버스를 형성한다. OS는 RTOS 또는 POSIX 기반을 적용하고, JAUS·STANAG 4586 같은 국제 표준을 준수해 상위 모듈이 하드웨어에 독립적으로 동작하도록 한다. 핵심은 상위 응용 모듈이 하드웨어 세부 구현에 의존하지 않도록 이 계층이 표준화된 인터페이스의 미들웨어 역할을 수행하는 것이다

엣지 AI 응용 및 임무 제어 계층: 엣지 AI 응용 계층은 객체 인식, 표적 탐지, 경로계획, 군집제어 등 AI 모듈과 자율제어/임무 모듈이 배치된다. 각 모듈은 표준 메시지 형식을 따르며, 교체·업그레이드 시에도 호환성을 유지한다. AI 출력은 임무 제어 계층으로 전달되고, 경로설정, 임무관리, 행동결정등의 로직을 수행하여 최종 제어 명령을 도출한다. 최종 제어 명령은 하위 계층을 통해 하드웨어 액추에이터로 실행된다.

데이터 흐름(OODA 루프): 센서 입력 → AI 분석(객체 탐지/상황인식) → 자율제어 모듈 의사결정(경로계획, 우선순위) → 액추에이터 제어·지휘소 보고. 모든 경로는 표준화돼 있어 센서나 알고리즘 교체 시에도 영향이 최소화되며, 병렬·분산 처리 확장도 가능하다. (그림2)



3. 시나리오 적용 및 유연성 분석

제안하는 MOSA 기반 엣지 AI 프레임워크의 효과를 검증하기 위해 두 가지 시나리오를 가정하여 유연성을 분석하였다. 첫째 이기종 무인체계 간 AI 모듈 재사용 시나리오는 UAV와 UGV가 동일한 표적 인식 AI 모듈을 공유하는 사례다. 기존에는 항공기와 지상차량용 알고리즘을 각각 개발해야 했지만, 본 프레임워크에서는 표준화된 데이터 토픽(VideoStream, DetectedObjectList 등)을 사용해 동일 모듈을 두 플랫폼에 탑재할 수 있다. 이를 통해 중복 개발을 방지하고 학습데이터·알고리즘 자산 공유로 비용·시간을 절감하며, 성능이 검증된 모듈을 신속히 타 플랫폼에 이전 적용할 수 있다. 둘째 임무 전환에 따른 모듈 교체 시나리오는 USV가 평시에는 정찰 모드, 유사시에는 기뢰제거 모드로 전환하는 상황이다. MOSA 기반 플랫폼에서는 동일한 선체에 센서·장비 모듈을 교체 장착할 수 있고, 소프트웨어 역시 표준 인터페이스 기반 모듈로 손쉽게 교체 가능하다. 예를 들어 정찰 모드에서는 탐색·식별 AI가, 기뢰제거 모드에서는 탐지·제거 AI가 활성화된다. 운영자는 모드 전환 명령만으로 필요한 모듈 세트를 호출할 수 있으며, 이는 마치 스마트폰에서 앱을 교체하듯 유연하다. 결과적으로 하나의 플랫폼으로 다양한 임무를 수행할 수 있고, 신규 임무도 모듈 추가만으로 확장 가능하다. 제안하는 시나리오를 기반으로 다음과 같이 프레임워크의 유연성을 분석할 수 있다.

표 1 제안하는 프레임워크 유연성 분석

분석항목	평가
모듈 교체 용이성	HW/SW 모두 표준화된 인터페이스 기반으로 교체 시 영향 최소화, 민첩한 업그레이드 가능.
다중 플랫폼 호환성	동일 모듈을 다양한 플랫폼에 활용해 계열화·부품 공용화, 유지비 절감.
상호운용	UAV가 탐지한 표적정보를 UGV와 공유하거나, 드론 군집의 스워밍 제어가 공통 프로토콜로 가능.
실시간성·신뢰성	미들웨어 QoS 설정으로 제어 메시지의 deterministic 처리 보장, 모듈화로 테스트·디버깅도 용이.
확장성	신형 AI 칩, 차세대 통신기술 등도 표준 인터페이스만 유지되면 해당 모듈만 교체하여 즉시 적용 가능.

Ⅲ. 결론

제안하는 MOSA 기반 엣지 AI 통합 프레임워크는 국방 무인체계의 설계를 폐쇄형에서 개방형으로 전환하고 독립계층적으로 구분하여 모듈화를 실현한다. 제안하는 프레임워크가 구체적 표준화와 시제품 개발에 적용되어 진다면 모듈화·지능화된 차세대 무인체계 구현에 기여할 것이다. 본 논문은 향후 실증 연구와 국제 표준화 논의의 기초 자료가 될 것이다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터사업과 지역지능화혁신인재양성사업의 지원을 받아 수행된 연구임(IITP-2024-RS-2024-00438430, IITP-2025-RS-2020-II201612), 2025년도 정부(교육과학기술부)의 재원으로 한국연구재단의 대학중점연구소 지원사업과 기초연구사업으로 수행된 연구임(2018R1A6A1A03024003).

참 고 문 헌

[1] Yazid, Y., Ez-Zazi, I., Guerrero-González, A., El Oualkadi, A., & Arioua, M. “UAV-enabled mobile edge-computing for IoT based on AI: a comprehensive review”. Drones, 2021.

[2] Defense Standardization Program Office, “DoD Modular Open Systems Approach (MOSA) Overview,” 2024.

[3] 장영현, 윤정환, 허장욱. “K-MOSA 기반 무인수상정 설계 및 획득 전략 연구”. 한국산학기술학회 논문지, 26(6), 217-224.2024.

[4] The Open Group, FACE Technical Standard Edition 3.0 2019.

[5] U.S Army, CMOSS Reference Architecture v3.0, 2020.

[6] SOSA Consortium, Sensor Open Systems Architecture Technical Standard 1.0 2019.

K-PART 기반 K-MOSA 평가체계와 AI 활용 평가 소프트웨어 개발의 정책적 필요성

안재광, 김동성, 이재민, 김정은
국립금오공과대학교, 스마트군수혁신융합연구센터
jkan17@kumoh.ac.kr, dskim@kumoh.ac.kr, ljmpaul@kumoh.ac.kr, wjddms6288@naver.com

Policy Necessity of AI-enabled Assessment Software in the K-PART Based K-MOSA Evaluation Framework

Jae-Gwang Ahn, Dong-Seong Kim, Jae-Min Lee, Kim Jeong-Eun
Kumoh National Institute of Technology, Smart Defense Logistics Innovation Convergence Research Center

요 약

본 연구는 한국형 모듈형 개방형 체계 접근법(K-MOSA)의 제도적 정착을 위해 K-PART(Korean-Performance Attributes for Requirements Traceability) 기반의 평가체계를 확립하고, 이를 효과적으로 실행하기 위한 AI 기반 평가 소프트웨어 개발의 정책적 필요성을 제시하는 것을 목적으로 한다. 미국의 MOSA 평가체계(PART)는 기술과 비즈니스 그룹 문항을 포함하되, 실제 획득 단계에서 적용되는 문서 평가가 중복성, 주관성, 문서 중심의 형식적 검증에 머무르는 한계를 갖는다. 이를 보완하기 위해 본 연구는 OUSD(R&E)와 PART 문항을 통합하여 중복을 제거하고 단계별 적합성을 고려한 평가 문항 및 평가 문항을 획득 단계별 요구문서와 연계하고, 인공지능(AI) 기반 소프트웨어를 활용한 K-PART 평가체계를 제안하였다. 본 연구는 K-MOSA 제도의 성공적 정착을 위해 정책적으로 AI 기반 평가 소프트웨어 도입이 불가결함을 강조하고, 기존 인적 기반 평가가 가진 주관성과 불확실성의 문제를 보완하며, 평가의 객관성·투명성·추적성을 제고할 수 있는 AI기술적용의 정책적 대안을 제시하는 데 있다.

Ⅰ. 서 론

21세기 국방 획득 환경은 무기체계의 복잡성과 기술 발전 속도 증가로 인해 기존의 폐쇄적·독점적 구조에서 벗어나 모듈형 개방형 접근(Modular Open Systems Approach, MOSA)으로 전환되고 있다. MOSA는 상호운용성, 경쟁 촉진, 수명주기 비용 절감, 신속한 기술 적용을 목표로 미국 국방부(DoD)는 이를 법률적 차원에서 제도화하였다[1][2]. 한국도 K-MOSA를 추진하며, 이를 평가하기 위해 MOSA의 PART를 참고하여 K-PART 개발을 추진하고 있다. 따라서 향상된 K-MOSA 시행과 국방 획득의 혁신을 위해 AI 기반 평가 소프트웨어의 개발과 적용이 필수적으로 제시되며, MOSA 추진을 위해 정책적으로 AI 기반의 평가 도구가 개발되어야 할 필요성이 제기된다.

Ⅱ. 본론

1 MOSA의 법적·정책적 배경

DoD는 2017년 NDAA(National Defense Authorization Act)를 통해 모듈형 개방형 체계 접근(Modular Open Systems Approach, MOSA)을 주요 방위획득 프로그램(MDAP)의 마일스톤 A 또는 마일스톤 B 승인에 법적으로 의무화하였다. 또한 모듈 간 인터페이스는 기계가 이해 가능한 형식(machine-readable format)으로 정의되어야 한다고 명시하였다[1][2][3]. 이 조항은 MOSA가 단순한 설계 원칙을 넘어, 법적 강제력을 가진 획득 제도로 자리 잡았음을 의미한다. 또한, MOSA Implementation Guidance(OUSD(R&E))에서도 MOSA의 실행을 위한 세부 지침을 제시하면서 인터페이스 정의가 단순 문서 형태가 아닌 시스템 간 자동 검증 가능한 형식으로 작성되어야 함을 요구한다[4]. 이를 모듈의 실행을 위한 메시지 규약 정도로 해석할 수도 있으나 MOSA관련 표준인 FACE의 CTS(Conformance Test Suite) 적합성 시험의 사례로 미루어보면 자동 시험 도구 사용이 가능하도록 사람의 개입 없이 컴퓨터가 쉽게 처리할 수 있는 형식으로 문서에 반영해야 한다는 것으로 해석된다.

2. MOSA평가 시스템의 문제점

미국의 대표적인 MOSA 평가도구 PART(Performance Attributes for Requirements Traceability)는 미국의 예산관리처에서 사용하던 사업 평가도구를 참조하여 개발되었으며[4], 실제 획득 단계에서의 적용 과정에서 유사·중복 문항, 정성적 판단에 의존, 문서 제출 여부 중심의 형식적 문서 중심의 검증이라는 한계를 갖는다.

법 항을 통해 MOSA 평가가 '마일스톤 A 또는 마일스톤 B 승인을 받는 MDAP'[4]를 근거로 방산 조달 시스템(DAS; Defense Acquisition System)의 적응형 획득 체계(AAF; Adaptive Acquisition Framework) 획득 프로세스를 기반으로 진행된다는 것으로 파악할 수 있다. 그러나 법적으로 MOSA 적용의 책임자만 제시되고 MOSA 평가에 대해서는 지침(DoDD, DoDI)이나 가이드라인이 없으며 평가 항목에서도 일치된 기준을 사용하고 있지 않다. 미 해군의 가이드북에서는 '필요에 따라 조정이 가능하다'[5]는 사항 역시 명확한 기준과 평가체계가 수립되지 않았다.

MOSA 평가에서 평가자들의 평가 결과를 취합하는 방법론으로 MAUT(Multi-Attribute Utility Theory)를 활용하여 PART의 각 항목에 가중치를 부여하고 점수를 산출한다[4]. MAUT는 다양한 속성을 수치화하여 비교할 수 있는 장점으로 여러 사업 평가 등에 적용되고 있는 방법론이다. 그러나 이러한 장점에도 한계가 존재한다. 특히, 가중치 설정의 주관성이다. 전문가의 판단에 따라 속성별 가중치가 달라져, 평가 결과가 달라질 수 있다. MOSA Implementation Guidance에서 예시로 제시한 테이블에서는 5개 Pillars의 가중치가 동일하게 설정되어 있으며, 척도도 절대적

<표-1> K-PART 평가 척도 개발

구분	평가 기준(문항수)	척도		K-PART 평가기준	
PART (Program Assessment Rating Tool)	Business / Technical (24)	• Progress or Status (Planned, Achieved, or N/A) • Extent to which each is achieved (None, Little, Moderate, Large)	▶ 동일/유사/연관 목적 그룹 통합 ▶	활성환경 구축	• 특징 : 개발 및 운영통합, 모듈별단계적 기술도입, 클라우드 데이터 공유 • 하위요소 : MOSA 목표수립, MOSA 채택, 계약 및 데이터 권리계획에 MOSA 및 OSA 포함, 애자일 방법론 활용, 자동화·개방형 및 모듈형 구조
OUSD (R&E)	Establish Enabling Environmen Employ Modular Design Designate Key Interfaces Use Consensus-Based Standards Certify Conformance (47)	• 0-5 scale		모듈식설계	• 특징 : 응집력, 캡슐화, 독립성, 모듈화 • 하위요소 : 모듈분리(확정성, 고가용성, 재사용성), 고장격리, 불변성, 모듈 독립성(인터페이스 일관성), 보안 설계(관리자 권한 없는 최소권한)
				주요 인터페이스 지정	• 특징 : 개방(공개)된 주요 인터페이스 • 하위요소(국한되지 않음) : 인터페이스 정의 문서화 및 버전관리, 검증된 표준 인터페이스 사용, 하위 호환성 유지
				개방표준 사용	• 특징 : 명확한 정의 및 체계, 널리 사용되고 접근성이 좋은 개방형 표준선택 • 하위요소: 국제 인정기관의 개방형 표준을 우선함. 보안 및 권한 관리된 개방성, 법적 충돌, 제약 없는 오픈 라이선스 사용, 보안·인터페이스 선택, DISR(DoD Information Technology Standards Registry)에 등록된 표준사용
OAAT (Open Architecture Assessment Tool)	Business / Technical(71)	• scale of 0 to 4.		적합성 인증	• 특징 : 공개된 적합성 기준, 자동 검사 및 인증 • MOSA 전략의 효과를 정량·정성적으로 확인할 수 있는 지표(MOE/MOP) 개발

※ PART /OUSD(R&E)/OAAT를 바탕으로 동일/유사/연관 문항 통합 정리

수치(0~5)의 값으로 평가된다. MAUT는 위험(risk)을 고려하여 효용을 산출하는 방식을 적용한다. 따라서 평가의 결과값 또는 성능 값에 대해 확률분포를 가진다면 이를 반영하여 계산한다. 그러나 PART에서는 이러한 사항을 포함하고 있지 않다. 따라서 어떠한 근거, 성능 등의 실제 인터페이스 데이터보다는 제출된 문서의 충족 여부를 기반으로 평가함으로 형식적 검증에 그칠 가능성이 높다. 따라서 현재의 MAUT 기반 MOSA 평가는 정책적 정당성 확보와 MOSA 적용 확인을 위한 도구로 기여하지만 실질적 이행 검증 도구로서의 한계를 가지고 있다. 따라서 실제 적용에서 평가자의 주관적 판단에 크게 의존하며, 이러한 문제는 MOSA 평가가 제도적으로 의무화된 상황에서 그 실효성과 정책적 정당성에 대한 의문을 제기될 수 있다.

3. K-PART, K-MOSA 평가체계 개발

미국의 MOSA 제도를 벤치마킹하여 K-MOSA 개념을 추진하고 있으며, 이를 정량적·정성적으로 평가하기 위한 K-PART를 개발하고 있다. 본 연구에서 제안하는 K-PART는 미국 OUSD(R&E)가 제시한 MOSA 5개 Pillars와 47개 문항, PART의 기술과 비즈니스 분류와 24개 문항을 통합하고 중복성을 제거하고 '체계공학 국제표준(ISO/IEC15288, ANSI/EAI 632)에 근거한 한국 국방 획득 절차[6]' 단계별로 적합한 평가 문항을 선별한 체계이다. K-PART는 획득 절차서와 연계하여 획득 단계별 적합 문항을 선별 적용함으로써, 평가가 문서 검증에 머무르지 않고 사업 단계별 의사결정과 직접적으로 연결되도록 설계하였다. 또한 평가 방법론으로는 MAUT를 적용하는 방법은 동일하나 가중치 설정 방법에 대해 획득 유형별로 명시하여 평가 과정의 투명성을 향상할 수 있도록 프로세스를 구성하였다.

4. AI 기반의 K-PART, K-MOSA 평가체계 개발

K-PART의 평가 문항, 프로세스를 정의하고 분석방법론의 적용을 보다 구체화하여 정의하는 것만으로는 미국MOSA 평가의 문제점을 개선할 수 없다. 미국 MOSA가 지적한 "기계 판독 가능 형식(machine-readable format)" 요구를 충족할 수 없으며, 평가 체계가 제도적 효력을 발휘하기 위해서는 AI 기반 평가 소프트웨어의 개발이 필수적이다. 이는 단순히 사람이 읽는 문서로는 불충분하며, 자동화된 도구가 직접 해석·검증할 수 있는 데이터 구조가 필요함을 의미한다. 그러나 미국의 PART 및 MAUT 기반 MOSA 평가체계는 여전히 문서 중심의 수작업 검증에 머물고 있으며, 기계 판독 가능성을 충분히 충족하지 못하고 있다. MOSA를 따르는 FACE의 CTS에서는 이를 자동화하는 평가를 진행하고 있다.

따라서 이러한 한계와 MOSA 기반의 표준 평가 사례를 바탕으로 한국형 MOSA 평가체계(K-PART)를 설계하는 과정에서 반드시 극복해야 할 요소이며, 본 연구가 K-PART 기반 K-MOSA 평가체계에 대해 정책적 관점에서 AI 기반 평가 소프트웨어의 필요성을 제시하는 이유이다. K-MOSA 제도의 성공적 정착을 위해 단순한 기술적 대안이 아닌 정책적 필요성으로서 AI 기반 평가 소프트웨어의 도입을 강조하며, 국방 획득의 객관성, 투명성, 추적성 제고에 기여할 수 있음을 보여준다.

Ⅲ. 결론

본 연구는 MOSA 평가의 정책적 한계를 극복하기 위해 K-PART와 AI 기반 평가 소프트웨어의 결합이라는 새로운 방향성을 제시하였다. 이는 단순한 기술적 도구 개발을 넘어, 한국 국방획득 정책의 객관성, 투명성, 글로벌 정합성을 보장하는 핵심 인프라로 발전할 수 있다. 그러나 K-PART 평가를 위해서는 ROC, TEMP, ICD, CDD 등 다양한 문서가 디지털 형식으로 제공되어야 하나, 현재 한국의 획득 문서 체계는 여전히 PDF·HWP 등 비구조적 문서 위주이다. 이외에도 제도적으로 MOSA와 K-PART 평가 결과를 정책적 의사결정에 직접 반영하기 위해서는 법률의 개정이 필요하다. 따라서 본 연구는 정책적 제안으로서 의미가 있으나, 실제 제도화·실행 과정에서 추가 연구와 실증적 검증이 요구된다. 향후 본 연구의 후속 연구가 실증적 검증과 정책 제도화를 뒷받침한다면, K-MOSA는 미국 및 NATO의 MOSA와 정합성을 유지하면서도, 한국적 상황에 최적화된 독자적이고 실질적인 정책 도구로 자리매김할 것이다.

ACKNOWLEDGMENT

"본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터사업의 연구결과로 수행되었음" (IITP-2024-RS-2024-00438430*)

참 고 문 헌

[1] FY2017 NDAA §805
[2] FY2021 NDAA §804
[3] U.S. Code Title 10 §4401, Major Defense Acquisition Programs: Modular Open System Approach in Development Program, 2022.
[4] OUSD(R&S), Implementing a Modular Open Systems Approach in Department of Defense Programs , 2025.
[5] Naval Modular Open System Approach Guidebook, Version1.0
[6] 방위사업청, SE기반 기술검토회의 가이드북, 2024.

국방 AI 자율 소프트웨어 정책에 관한 고찰: 사례 기반 분석과 정책과제

이다정, 권익현, 김동성*
국립금오공과대학교, 국립금오공과대학교, 국립금오공과대학교
djlee@kumoh.ac.kr, inkwonaf@kumoh.ac.kr, *dskim@kumoh.ac.kr

A Study on Defense AI Autonomous Software Policy: Case-Based Analysis and Policy Implications

Lee Da Jung*, Kwon Ik Hyun, Kim Dong Seong

요 약

본 연구는 국방 분야에서 인공지능(AI) 자율 소프트웨어의 발전과 적용 현황을 분석하고, 실제 사례를 바탕으로 문제점과 정책과제를 도출하였다. 최근 우크라이나-러시아 전쟁에서 드론 및 자율무기체계의 활용은 AI 소프트웨어가 현대전 양상을 근본적으로 변화시키고 있음을 보여준다. 그러나 동시에 민간인 피해, 책임 소재 불분명, 사이버 공격 취약성 등 심각한 문제가 발생하고 있다. 본 연구는 미국 DARPA, 이스라엘 무인 드론 사례, 한국 국방과학기술혁신 기본계획을 검토하여 윤리·법제 기반, 기술적 안전성, 국제 협력, 민군 생태계 조성 측면에서 정책 제언을 제시하였다.

Ⅰ. 서 론

4차 산업혁명과 함께 인공지능(AI)의 발전은 군사 영역에 빠르게 확산되고 있다. 특히 AI기반 자율 소프트웨어는 실시간 상황 인식과 자율적 의사결정을 통해 무인기, 전차, 지휘통제 체계 등 다양한 분야에서 전투 효율성을 향상시키고 있다[1][2]. 하지만 자율율무기체계(LAWS, Lethal Autonomous Systems)의 등장으로 국제사회는 윤리·법적 논란과 책임성 문제를 직면하게 되었다. 한국 역시 국방 AI 추진 기본계획을 수립했으나[9], 아직 법제적·윤리적 체계는 미비한 수준이다[10][11]. 이에 본 연구는 최근 사례와 문헌을 분석하여 국방 AI 자율 소프트웨어 정책의 문제점을 분석하고, 한국이 나아가야 할 정책 과제를 제시하고자 한다.

Ⅱ. 연구 방법론

2.1 문헌 분석

최근 5년간 발표된 논문과 정책 보고서를 검토하여 AI 무기체계 신뢰성, 국방 AI 보증 체계, 보안 프레임 워크, 인권 및 윤리 문제, 국제 규범, 국방 AI 신뢰성, 군사적 적용, 국방부 정책 자료, 한국국방연구원 보고서를 검토하였다.

2.2 사례 분석

실제 전장에서 활용된 ▲우크라이나 전쟁의 드론 사례, ▲DARPA Mosaic Warfare, ▲이스라엘 하르피 드론을 중심으로 분석하였다.

Ⅲ. 사례 분석

3.1 우크라이나-러시아 전쟁 사례

우크라이나 전쟁(2022~)은 AI 기반 드론이 실제 전투에서 대규모로 활용된 첫 사례이다. 양측은 자율 소프트웨어를 통해 표적 탐지·공격을 수행했으나, 오작동으로 민간시설 피해가 발생했다. 이는 국제인도법 위반 논란을 불러왔고, 자율무기 사용에 대한 국제 규범 필요성이 대두되었다.

3.2 미국 DARPA의 Mosaic Warfare

미국 DARPA는 'Mosaic Warfare'를 통해 무인기·센서·자율 소프트웨어를 결합하여 네트워크 중심 전투를 실현하고 있다. 그러나 모의훈련에서 적대적 사이버 공격으로 네트워크 전체가 마비된 사례가 보고되었으며 이는 자율 소프트웨어가 지닌 보안 취약성을 드러낸다.

3.3 이스라엘 하르피 드론

이스라엘은 표적 탐지 후 자동 자폭하는 하르피 드론을 실전에 투입하였다. 나고르노-카라바흐 전쟁에서 전술적으로 성공했지만, 민간인 피해 가능성과 표적 판별 오류 문제가 제기되었다.

Ⅳ. 문제점 및 이슈 도출

국방 AI 자율 소프트웨어의 발전은 다음과 같은 문제가 제기된다.

첫째, 윤리·법적 문제로 민간인 피해 발생 시 책임 주체가 불명확하며, 국제인도법과 충돌한다[4][5][11]. 둘째, 기술적 취약성으로 적대적 공격으로 인한 인식 오류와 사이버 공격으로 인한 체계 마비 가능성이 크다[2][3]. 셋째, 정책·제도의 미비로 한국의 정책은 기술개발 위주에 머물러 있으며 법적·윤리적 기준이 미비하다[9][10]. 마지막으로 산업 생태계의 한계이다. 국내는 방산 대기업 중심의 폐쇄적 구조로 스타트업·학계의 참여와 민군 협력이 부족한 실정이다[7][8].

구체적인 해외정책사례는 다음과 같다.

구분	해외 정책 사례	시사점/ 적용과제
윤리·법제	-미국 국방부 DoD Directive 3000.09 (2023 개정): 모든 자율무기체계는 인간의 의미 있는 통제(MHC) 원칙을 준수해야 함 - UN GGE(2021~): 치명적 자율무기(LAWS)에 대해 국제인도법 준수 필요성 논의	- 한국 법령에 MHC 원칙 명문화 - 국방 AI 윤리위원회 설립 및 심의 제도화
기술적 안전성	- DARPA "AI Next Campaign" (2018~): XAI(Explainable Artificial Intelligence) 프로그램 운영, 설명 가능한 의사결정 모델 개발 - NATO 2021 AI 전략: 모든 군사 AI 시스템은 보안성·신뢰성 검증을 의무화	- 국방 무기체계 XAI 의무 적용 - 사이버 보안 모듈 및 적대적 공격 대응 알고리즘 국방 규격화
국제 규범	- EU AI Act (2024): 고위험군 AI(군사·안보 포함)에 대한 엄격한 규제 및 설명가능성 요구 - Stop Killer Robots 캠페인(2025): 전 세계 90여 개국 참여, 자율무기 금지·규제 조약 제안	- UN LAWS 협상에서 한국의 중재자 역할 수행 - 한미일·NATO 공동 선언 추진
민군 생태	- 이스라엘 IAI·Elbit Systems: 스타트업 및 민간 기업과 공동으로 드론·AI 무기체계 개발.	- 민군 공동 R&D 프로그램 확대

계	- 미국 DIU(Defense Innovation Unit): 민간 스타트업과 군 협력 가속화. - EDA(2025 White Paper): 유럽 내 국방 AI 신뢰성 검증 연구 지원.	- 국방 데이터 개방 및 AI 전문 인재 양성

V. 정책 과제

구분	도출 근거(사례·문헌)	정책 과제
윤리·법제	-민간인 피해 책임 불분명 -HRW Report[4][11]	-MHC 원칙 법제화 -국방 AI 윤리위원회 설립 및 심의 제도화
기술적 안전성	-DARPA Mosaic Warfare사이버공 [2] -Kapusta Framework [2][3]	-설명 가능 AI(XAI)의무 적용 -적대적 공격 대응 알고리즘 및 보안모듈 탑재
국제 규범	-UN LAWS 협상 지연[6] -Stop Killer Robots Report [5]	-UN 협상 적극 참여 및 중재 역할 수행 -한미일·NATO 공동 선언 추진
민군 생태계	-국방 AI 추진 기본계획 한계[9] -EDA White Paper[7][8]	-방산-스타트업-학계 공동 R&D 확대 -국방 데이터 개방 및 AI 인재 양성

본 연구에서는 다음과 같이 네 가지 구체적 정책 과제를 제시한다. 첫째, 윤리적 법제 기반의 확립이다. 의미 있는 인간 통제(Meaningful Human Control, MHC)'원칙을 「국방 AI 운영지침」과 국방 관련 법률에 명문화하고 국방부 산하 국방 AI윤리위원회를 신설하여 무기체계의 개발 단계별 사전 심의, 운용 중 발생할 수 있는 윤리적 문제에 대한 평가를 제도화하는 것이다. 이를 통해 국제인도법 준수를 강화하고 한국이 국제사회에서 책임 있는 AI 무기 사용국으로서의 신뢰를 확보할 수 있다. 둘째, 기술적 안전성 강화로 모든 AI 기반 무기체계에 설명 가능한 인공지능(XAI)을 적용하여 알고리즘의 의사결정 과정을 투명하게 검증하는 것이다. 적대적 공격에 대응할 수 있는 보안 알고리즘과 사이버 침해 대응 모듈을 국방 규격(MIL-STD)로 제정하는 것과 국방 R&D 단계에서부터 "레드팀(Red Team)" 검증 체계를 운영하여 실전 환경에서 발생 가능한 해킹·교란 시나리오를 사전에 실험하는 것이다. 셋째, 국제 규범 참여 및 선도가 필요하다. 한국이 UN 협상에서 중재자적 입장을 취해 조건부 규제 +MHC 의무화라는 절충안을 제안할 수 있다. 한미일 및 NATO와 협력하여 자율무기 운용 원칙 공동 선언을 추진하는 등 국제 규범 형성 과정에서 규범 선도국으로 자리매김 할 수 있도록 하고 ASEAN, EU 등과의 방산 협력 채널을 통해 양자 및 다자 협약체결을 선도함으로써 AI 군사기술을 둘러싼 국제 경쟁에서의 협상력을 확보할 수 있게 될 것이다. 마지막으로는 민군 융합 생태계 조성이다. 방산기업-스타트업-대학 간 공동 연구개발 프로그램을 정부 매칭펀드 방식으로 확대하고, 국방 데이터셋을 공유 가능 수준에서 민군 공동 활용할 수 있는 개방형 국방 데이터 플랫폼을 구축하는 것이 필요하며 또한 국방 특화형 AI 인재 양성 트랙(군사학+AI 융합 교육) 및 산학연군 연계를 통한 전문 인력의 파이프라인을 확보함으로써 국가 전체 민군 AI 생태계 발전의 확장과 민군 동반성장을 촉진할 수 있다.

VI. 결론

본 연구는 최근 전쟁 사례와 학술·정책 보고서를 토대로 국방 AI 자율 소프트웨어 활용에 제기되는 문제점과 정책 과제를 도출하였다. 그 결과 윤리·법제적 공백, 기술적 취약성, 국제 규범의 부재, 민군 생태계의 한계라는 핵심 문제가 확인되었다. 구체적으로 ▲MHC(Meaningful Human Control) 원칙 법제화, ▲XAI(Explainable Artificial Intelligence) 및 보안 모듈 의무화, ▲UN 및 다자 협상에서의 한국의 중재자적 역할 강화, ▲민군 융합 생태계 구축이라는 정책 과제를 제안하였다. 이러한 정책 과제는 단순한 기술적 대응을 넘어, 법·윤리적 정당성 -> 기술적 신뢰성 -> 국제적 정합성 -> 산업적 지속가능성이라는 연속적인 구조 속에서 추진되어야 한다. 특히, 한국은 국방 AI 정책을 통해 안보 강화와 동시에 산업 경쟁력 확보, 그리고 국제사회에서의 규범 선도국 위상을 달성할 수 있을 것이다.

따라서 본 연구의 함의는 국방 AI 자율 소프트웨어 정책이 더 이상 선택이 아닌 국가 안보와 산업, 국제 규범을 통합하는 전략적 필수 과제라는 점을 명확히 하는 데 있다. 그러나 주로 문헌 및 사례 분석에 기반하였기 때문에 실제 실행 단계에서의 실증적 계량적 검증은 한계가 있다. 따라서 다음과 같은 구체적 연구가 필요하다. ①법제도적 시뮬레이션 연구, ②XAI 기반 군사 시스템 평가 연구로 실제 무기체계(드론, 무인지상차량 등)에 XAI 를 적용했을 때 설명 가능성이 얼마나 확보되는지에 대한 실증연구, ③사이버 보안 취약성 실험연구, ④국제 규범 형성 과정 분석 연구, ⑤민군 융합 생태계 효과 분석 연구를 통해 실제 공통 R&D 프로젝트의 기술혁신과 산업 파급효과에 어떤 영향을 미치는지 효과 분석 연구가 필요하다. 이러한 연구는 제안된 정책의 실행 가능성과 성과를 실증적으로 평가하여 정책적 정합성을 강화해야 할 것이다.

ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원-지역지능화혁신인재양성사업의 지원을 받아 수행된 연구임(IITP-2025-RS-2020-II201612).
이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원-대학ICT연구센터(ITRC)의 지원을 받아 수행된 연구임(IITP-2025-RS-2024-00438430).

참 고 문 헌

[1] Cools, K., & Maathuis, C., "Trust or Bust: Ensuring Trustworthiness in Autonomous Weapon Systems," arXiv preprint, 2024, pp. 1-15.

[2] Kapusta, A. S., Jin, D., Teague, P. M., Houston, R. A., Elliott, J. B., Park, G. Y., & Holdren, S. S., "A Framework for the Assurance of AI-Enabled Systems," arXiv preprint, 2025, pp. 1-22.

[3] Tallam, K., "Engineering Risk-Aware, Security-by-Design Frameworks for Assurance of Large-Scale Autonomous AI Models," arXiv preprint, 2025, pp. 10-35.

[4] Human Rights Watch & Harvard Law School International Human Rights Clinic, "A Hazard to Human Rights: Autonomous Weapons Systems and Digital Decision-making," Human Rights Watch Report, 2025, pp. 5-28.

[5] Stop Killer Robots, "Policy Brief: Autonomous Weapons Systems - Key Issues and Path to a Treaty," Policy Brief, 2025, pp. 1-12.

[6] Lieber Institute, "Artificial Intelligence in Armed Conflict: The Current State of International Law," CyCon 2025 Series, 2025, pp. 30-52.

[7] European Defence Agency, "Trustworthiness for AI in Defence," EDA White Paper, 2025, pp. 1-18.

[8] Trends Research, "Governing Lethal Autonomous Weapons in a New Era of Warfare," Trends Research Article, 2025, pp. 40-60.

[9] 국방부, 「국방과학기술혁신 기본계획, 2023-2037」, 국방부 정책자료, 2020, pp. 3-25.

[10] 한국국방연구원(KIDA), 「AI와 국방전략의 연계: 미국의 국방혁신구상을 통해 본 한국의 미래전 대비」, 연구보고서, 2025, pp. 1-6.

[11] 한국국방연구원(KIDA), 「자율살상무기 국제 규범 형성에 관한 연구」, 미래국사학회, 2023, pp. 53-73.

MOSA 표준군에서의 AI 통합 가능성 비교: 인터페이스·데이터·모듈성

류원재, 이재민, 김동성
국립금오공과대학교
{wj0828, ljmpaul, dskim}@kumoh.ac.kr

Integrating AI in MOSA Standards: Interfaces, Data, and Modularity

Ryu Won Jae, Lee Jae-Min, Kim Dong-Seong
Kumoh National Institute of Technology

요 약

본 논문은 국방 분야의 Modular Open Systems Approach(MOSA) 핵심 표준군(FACE, SOSA, OMS/UCI, VICTORY)을 대상으로, 인터페이스·데이터·모듈성 관점에서 AI 통합 가능성을 비교·분석한다. 표준 문서와 공개 사례를 바탕으로 AI 수명주기(모델 개발 - 배포 - 운용 - 검증)에 대한 표준별 매핑을 수행하고, 표준화된 인터페이스·메시지(OMS/UCI, VICTORY), 데이터 모델 및 미들웨어(FACE 데이터 모델, DDS), 모듈/서비스화 및 이식성(FACE UoC, OMS 서비스), 하드웨어 가속/폼팩터 수용성(SOSA OpenVPX), 안전성·신뢰성 거버넌스(NIST AI RMF, SAE AS6518)를 평가 축으로 설정하였다. 분석 결과, FACE와 OMS/UCI는 AI 알고리즘을 모듈/서비스로 캡슐화해 플랫폼 간 재사용성·교체성을 극대화하고, SOSA는 GPU/FPGA 등 최신 가속기를 표준 슬롯과 고속 인터커넥트로 수용해 실시간 추론에 유리함을 확인하였다. VICTORY는 차량 내 공통 데이터버스를 통해 AI 기능의 저비용 삽입과 결과의 분산 공유를 지원한다. 또한 ONNX와 DDS는 각각 모델 이동성과 실시간 데이터 분배의 핵심 연결고리로 기능하며, NIST AI RMF 및 항공·무인체계 안전성 표준은 AI 모듈의 인증·검증 체계를 보완한다. 본 연구는 표준군 간 강·약점을 정리한 비교틀과 MOSA 기반 AI 통합 참조구조를 제안하고, 신속 기술 삽입, 벤더 종속 완화, 생애주기 비용 절감 측면의 시사점을 제시한다.

Ⅰ. 서 론

MOSA(Modular Open Systems Approach)는 국방 분야 시스템을 모듈화하고 표준화된 개방형 인터페이스를 채택함으로써 신속한 기술 삽입과 상호운용성을 달성하려는 전략이다[1]. 2019년 미 국방부 3군 합동 지침에서는 MOSA 적용을 의무화하며, 이를 구현하기 위한 공통 표준으로 OMS/UCI, SOSA, FACE, VICTORY 등이 명시되었다. 본 연구에서는 이러한 MOSA 핵심 표준군이 자율제어, 영상/신호분석, 임무계획, 다영역 지휘통제(C2) 등 AI 기술을 시스템에 통합하는 데 어떤 역할을 하는지 분석한다. 각 표준이 정의하는 구조와 인터페이스, 데이터 포맷이 AI 모듈 통합에 어떻게 활용될 수 있는지 살펴보고, AI 알고리즘/모듈이 해당 표준 내에서 어떤 방식으로 통합되어 모듈화/호환성/재사용성에 기여하는지 고찰한다. 아울러 미국 국방부 및 방산업계의 MOSA 기반 AI 통합 사례를 조사하고, ONNX, DDS, NIST AI RMF, SAE AS6518(UCS Architecture) 등의 외부 AI 표준이 MOSA와 연계되어 활용되는 방안을 논의한다.

Ⅱ. 본론

FACE(Future Airborne Capability Environment)는 군용 항공전자 소프트웨어용 개방형 표준으로, 계층형 아키텍처와 UoC(Unit of Conformance) 구조를 통해 모듈 이식성과 재사용성을 보장한다 [2]. FACE는 데이터 중심 아키텍처와 공통 데이터 모델을 엄격히 정의하여, AI 모듈의 입출력도 표준화된 객체로 다룬다. 예컨대 표적 인식 AI는 탐지 결과를 FACE 데이터 모델로 표현해 항전장치 간 교환할 수 있다. 또한 TSS - DDS 기반으로 컴포넌트 간 퍼블리시/구독 통신이 가능해 AI 알고리즘을 독립 모듈로 안전하게 통합할 수 있다. 실제 미 해군은 FACE 모듈을 여러 기종에 재사용해 신규 기능을 빠르게 추가한 사례를 보고했

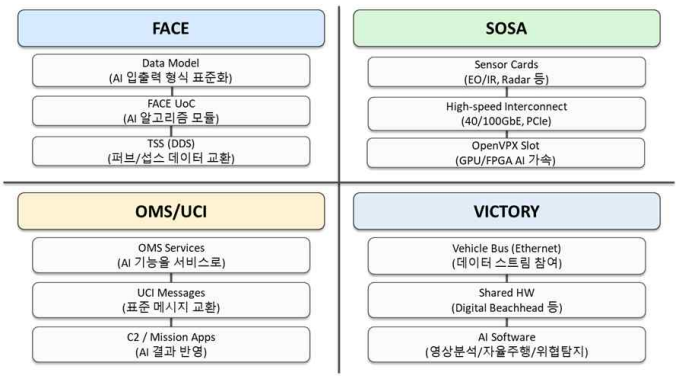


그림 1 MOSA 표준군에서의 AI 활용

으며, 이는 AI 기반 기능 역시 다기종 플랫폼 재사용이 가능함을 시사한다.

SOSA(Sensor Open Systems Architecture)는 C4ISR/전자전용 센서·처리 하드웨어 개방형 아키텍처로, OpenVPX 기반 모듈 규격과 공통 백플레인을 정의한다 [3]. 표준 슬롯에 GPU/FPGA 가속기 모듈을 삽입해 고해상도 영상·신호 AI를 실시간 처리할 수 있다. Curtiss-Wright 등은 SOSA 생태계에 Intel, AMD, NVIDIA 최신 칩셋을 탑재하고 40/100GbE 인터커넥트를 제공하여 인지처리(cognitive processing)까지 지원한다. 실제 현장에서는 SOSA를 활용해 기존 시스템에 신호 탐지/위치추적 AI 기능을 추가한 사례가 있으며, 핫스왑 가능한 모듈 구조로 부분 업그레이드가 용이하다. 이는 AI 알고리즘의 고속 병렬처리와 실시간 상황인식 기능을 빠르게 현장 적용하게 해준다.

OMS/UCI(Open Mission Systems/Universal Command and Control Interface)는 미 공군의 서비스 지향 임무체계 표준(OMS)과 공통 메시지·

데이터 모델(UCI)은 센서·무장·미션컴퓨터 간 데이터 교환을 표준화한다 [4]. AI 알고리즘은 서비스 단위로 통합되어 UCI 메시지를 주고받으며, 기존 시스템을 크게 수정하지 않고도 신규 기능을 삽입할 수 있다. 예를 들어 자동표적식별(ATR) AI는 OMS/UCI 구조에서 기존 센서 데이터를 받아 새로운 표적 트랙을 산출해 표준 메시지로 공유한 사례가 있다. 또 BMTN 훈련체계에서는 AI 튜터와 음성인식 모듈이 OMS/UCI 기반 API를 통해 무한 확장이 가능하도록 개발되었다. Lockheed Martin 역시 OMS/UCI를 활용해 자율비행 AI와 MUM-T 기능을 빠른 주기로 통합하고 있다.

VICTORY(Vehicular Integration for C4ISR/EW Interoperability)는 미 육군 주도로 개발된 전투차량 전자체계의 공통 네트워크·데이터버스 표준으로, LRU 간 상호운용성과 플러그 앤 플레이형 통합을 지원한다 [5]. 공용 컴퓨팅 자원 위에 소프트웨어만 추가해 AI 기능을 탑재할 수 있으며, 차량 내 데이터버스 구독/퍼블리시 구조를 통해 센서 데이터를 입력받고 분석 결과를 공유한다. 예를 들어 카메라 영상 기반 표적식별 AI는 VICTORY 메시지로 표적 정보를 전술 상황 디스플레이와 사격통제 시스템에 전달할 수 있다. 미 육군 차세대 차량 사업에서도 VICTORY·SOSA·CMOSS를 적용하여 자율주행·대전차 AI를 소프트웨어 업그레이드 방식으로 통합 가능하도록 요구하고 있다.

앞서 살펴본 각 표준(FACE, SOSA, OMS/UCI, VICTORY)은 서로 적용 대상 분야는 다르지만, ONNX는 AI 모델을 프레임워크 독립적으로 교환해 MOSA 시스템 간 벤더 종속 최소화와 모델 재사용을 지원한다. DDS는 FACE·UCS 등에서 공통 미들웨어로 활용되며, AI 모듈 간 실시간 퍼브/섭스 데이터 교환을 가능케 한다. NIST AI RMF는 MOSA에 탑재되는 AI 모듈의 신뢰성·안전성 검증 프레임워크를 제공하고, SAE AS6518(UCS)는 무인체계 제어 기능을 서비스화하여 AI 임무계획·자율제어 알고리즘을 표준화된 방식으로 통합할 수 있게 한다 [6].

Ⅲ. 결론

MOSA 핵심 표준(FACE, SOSA, OMS/UCI, VICTORY)는 각 군 종족별로 포커스는 상이하지만 모두 모듈화된 설계와 공개 표준 인터페이스를 기반으로 하여 AI 기술의 통합을 용이하게 한다는 공통점을 가진다. FACE와 OMS/UCI는 소프트웨어 계층과 데이터 모델의 표준화를 통해 AI 알고리즘을 하나의 모듈로 캡슐화하여 다양한 플랫폼에서 재사용하거나 교체할 수 있게 해주고, SOSA와 VICTORY는 하드웨어 및 네트워킹 표준화를 통해 AI 연산 자원을 플러그인 방식으로 추가하거나 여러 장비와 데이터를 공유할 수 있도록 지원한다. 그 결과, AI 기술의 비약적인 발전에도 군사 시스템이 유연하게 대응하여 업그레이드 주기를 앞당길 수 있고, 벤더 종속 감소와 경쟁 촉진으로 비용 절감 및 혁신 가속 효과를 거둘 수 있다. 나아가 ONNX, DDS, NIST AI RMF 등 외부 표준과 프레임워크를 MOSA 체계에 접목함으로써, AI 모듈의 모델 포맷 호환, 데이터 연계성 향상, 신뢰성/안전성 검증을 뒷받침하여 MOSA의 가치를 극대화할 수 있다. 요컨대 MOSA와 AI의 결합은 최신 AI 혁신을 신속히 전력화하고 체계 수명 전반에 걸쳐 진화하는 위협에 대응하는 데 필수적인 접근으로 부상하고 있다. 모듈식 개방형 아키텍처 위에서 AI 기술을 통합하면, 국방 분야는 기술우위 유지와 업그레이드 민첩성 확보를 동시에 이룰 수 있을 것이다.

표 1 MOSA 표준별 AI 적용 상세 사례 비교

표준	적용 분야	AI 적용 사례	특징 및 효과
FACE	항공전자 SW	- AI 표적 인식 모듈을 FACE 컴포넌트로 탑재, 탐지 결과를 공통 데이터 모델로 교환 - 미 해군: FACE 호환 모듈을 여러 항공기에 재사용	- 데이터 중심 아키텍처, 입출력 형식 표준화 - DDS 기반 퍼브/섭스 데이터 교환 - 다기종 플랫폼 간 재사용성확보
SOSA	센서·임무 하드웨어	- OpenVPX 슬롯에 GPU/FPGA 모듈 추가, 영상/신호 AI 실시간 처리 - Curtiss-Wright: NVIDIA/AMD GPU 모듈+고속 인터커넥트로 인지처리지원- 배치 시스템에 신호탐지·위치추적 AI 기능 추가 사례	- 모듈 핫스왑 구조로 보드 교체 용이- 고속 병렬처리에 유리, 실시간 SA(Situational Awareness)지원 - 기존 플랫폼 성능 증강에 최적
OMS/UCI	임무체계 SW·C2	- 미 공군: UAV에 자동표적식별(ATR) AI를 OMS 서비스로 통합, 기존 센서 데이터 활용 - BMTN 훈련체계: AI 튜터·음성인식 기능을 OMS/UCI API 기반으로 확장	- AI를 서비스 단위로 통합- 표준 메시지/UCI로 데이터 송수신 - 기존 시스템 수정 최소화, 플러그인식 확장가능
VICTORY	전투차량 네트워크	- 차량 데이터버스에 영상기반 표적식별 AI연결, 탐지 결과를 전술 디스플레이·사격통제와 공유 - 차세대 전투차량(OMFV): 자율주행·대전차 AI를 SW 업그레이드로 탑재 가능	- 차량 내 공통 네트워크·메시지 포맷활용- 공용 HW 자원 위에 AI SW만 추가 - 저비용·저위험 기술 삽입가능, 지능화 업그레이드 가속

ACKNOWLEDGMENT

이 논문은 2025년도 정부의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(2018R1A6A1A03024003, 33%), 이 논문은 과학기술정보통신부 정보통신기획평가원의 지원을 받아 수행된 연구임 (IITP-2025 RS-2020-II201612, 33%), 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터사업의 연구결과로 수행되었음((IITP-2025 RS-2024-00438430, 34%).

참 고 문 헌

[1] U.S. Department of Defense, Modular Open Systems Approach (MOSA) Memo, Office of the Secretary of Defense, 2019.

[2] Open Group FACE Consortium, FACE Technical Standard, Edition 3.0, 2017.

[3] SOSA Consortium, SOSA Technical Standard, Release 1.0, The Open Group, 2021.

[4] U.S. Air Force, Open Mission Systems (OMS) / Universal Command and Control Interface (UCI) Overview, Version 3.0, 2020.

[5] U.S. Army PEO GCS, Vehicular Integration for C4ISR/EW (VICTORY) Architecture Reference Design, 2016.

[6] NIST, AI Risk Management Framework (AI RMF) 1.0, National Institute of Standards and Technology, 2023.

VLM을 활용한 자율주행 차량의 장애물 개척 의사결정 시스템

서원준[1], 최승찬[1], 김휘순[1], 조성현[2], 배준혁[3], 이용준[3], 함형찬[4], 이은재[4], 김종환[3], 안희진[4*]

울산과학기술원[1], 포항공과대학교[2], 육군사관학교[3], 한국과학기술원[4]

{wonjuneseo, unknown, blank}@unist.ac.kr, {sungseo01}@postech.ac.kr, {bjh040919, lee95243, jonghwan7028}@gmail.com, {hyeonchan.ham, eunjae.lee, heejin.ahn}@kaist.ac.kr

VLM-Based Obstacle Breakthrough Decision-Making for Autonomous Vehicles

Wonjune Seo, Seungchan Choi, Hwisoon Kim, Sunghyun Joe, Joonhyuk Bae, Yongjun Lee, Hyeongchan Ham, Eunjae Lee, Jonghwan Kim, Heejin Ahn*

UNIST., POSTECH., Korea Military Academy., KAIST

요 약

복잡한 주행 환경에서 자율주행 차량이 직면하게 되는 장애물은 기존 회피·정지 중심의 경로 계획만으로는 임무 완수가 어려운 경우가 많다. 특히 소방·국방 등 특수 분야에서는 장애물을 능동적으로 개척(breakthrough)하는 의사결정 능력이 요구된다. 본 연구는 비전 언어 모델(VLM)을 활용해 회피 불가 상황에서 환경 정보를 기반으로 최적 개척 대상을 선정하는 시스템을 제안한다. 제안하는 방법은 차량의 주변 상황 이미지와 장애물 ID를 VLM에 입력하고, 인명 피해 최소화를 목표로 개척 대상을 선정한다. CARLA 시뮬레이터 상에서 제안 방식은 기존 모델 대비 상황 맥락 파악을 통해 인명 피해 최소화가 가능함을 보였다. 이로써 VLM이 유연하고 안전한 의사결정을 수행할 수 있음을 입증한다.

Ⅰ. 서 론

최근 자율주행 기술은 급격한 수요 증가와 함께 학계·산업계 전반의 핵심 연구 분야로 부상하고 있다. 특히 혼잡한 교차로, 다양한 교통수단 혼재, 돌발 상황 등 복잡한 주행 환경에 대응을 위한 기술 개발이 활발히 진행 중이다. 기존 자율주행은 장애물을 인식 후 회피 또는 정지 중심의 전략으로 안전성을 확보한다. 그러나 소방·국방 등 특수목적 분야에서는, 주행 경로가 완전히 차단된 환경에서 자율주행차가 장애물과 물리적으로 충돌하여 길을 확보하는 능동적 개척 (breakthrough)이 핵심 역량으로 요구된다.

기존 개척 자율주행 연구들은 대부분 규칙 기반 (rule-based) 제어에 의존하여 사전에 정의된 규칙과 조건에 따라 동작을 결정하며 이 방식에는 세 가지 한계가 존재한다. 첫 번째로, 기존 개척 자율주행 연구는 지정 경로 주행 [1]에 초점을 맞추어 개척 대상 판단과 경로 설정같은 고수준 의사결정을 구현하지 못한다. 두 번째로, 규칙 기반 자율주행은 우선순위를 선정하고 주행을 계획하는 데에 수학적 한계가 있다 [2]. 세 번째로, 규칙 기반 자율주행은 시스템이 직면할 모든 객체가 사전에 정의된 폐쇄 세계 (Closed world) 에서 동작한다. 따라서 사전에 정의되지 않는 요소가 존재할 수 있는 개방 세계 (Open world)에 대해서는 대응이 어렵다 [3].

한편, 비전 언어 모델 (Vision Language Model, VLM)은 방대한 학습 데이터로 지식을 구축하여 맥락을 이해할 수 있고, 이미지 기반 장면 인식을 결합하여 복잡한 상황에서 유연하게 의사 결정을 수행할 수 있다. 본 연구의 목표는 VLM을 활용하여 자율주행 차량이 다양한 장애물로 인해 주행 불가능한 상황에 직면했을 때 VLM의 유연한 의사 결정 능력을 통해 개척 대상을 선정하는 것이다.

Ⅱ. 본론

본 연구가 다루는 문제는 자율주행 차량의 주행 경로가 전방 장애물로 완전히 차단되어 회피 경로가 존재하지 않는 상황에서, 임무 수행을 위해 능동적으로 개척 대상을 선정하는 것이다. 개척 대상선정은 전반적인 인명 피해 최소화를 목표로 하며, 도로 내에 있는 장애물만 인식 대상으로 한다. 문제상황에 집중하기 위해 장애물 인식은 센서 기반 탐지 대신 ground truth를 활용하며, 제어 알고리즘이 feasible한 제어 입력을 찾지 못하는 경우 전방이 완전히 차단된 상황으로 간주한다.

본 논문에서 제안하는 VLM 기반 개척 모델의 구성은 자율주행의 인지, 판단, 제어 중 판단에 해당하며, 관련 내용은 [그림 1]에 제시되어 있다. 전방이 완전히 차단된 상황으로 간주 되면, 장애물의 위치에 따른 ID와 전방 FPV(First-Person-View) 이미지를 판단 모듈에 입력한다. 판단 모듈은 VLM을 활용하여 입력 프롬프트에 따라 '인명 피해를 최소화하는 대상'을 개척 대상으로 선정한다. 이후 제어 모듈은 해당 대상에 대해서만 장애물 회피 로직을 비활성화함으로써 개척을 수행하여 자율주행이 차량 임무를 지속적으로 수행할 수 있도록 한다.

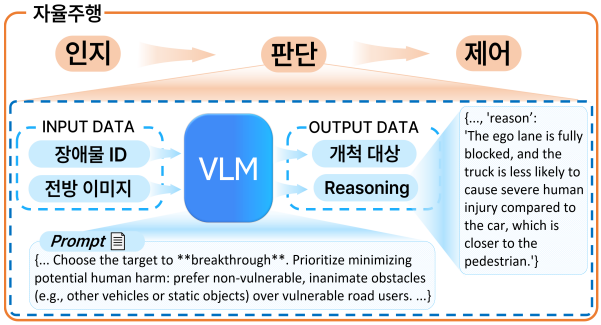


그림 1 VLM 기반 개척 모델 도식도

Ⅲ. 실험 및 결과

3.1 실험 설계

실험은 실제 환경과 유사한 CARLA 시뮬레이터에서 수행하였다.

비교 실험에는 다음의 세 가지 모델을 사용하였다.

1. **보수적 회피(conservative):** 기존 UDMC [4] 모델을 그대로 적용하여, 우선순위 없이 모든 장애물을 회피하는 주행을 한다.

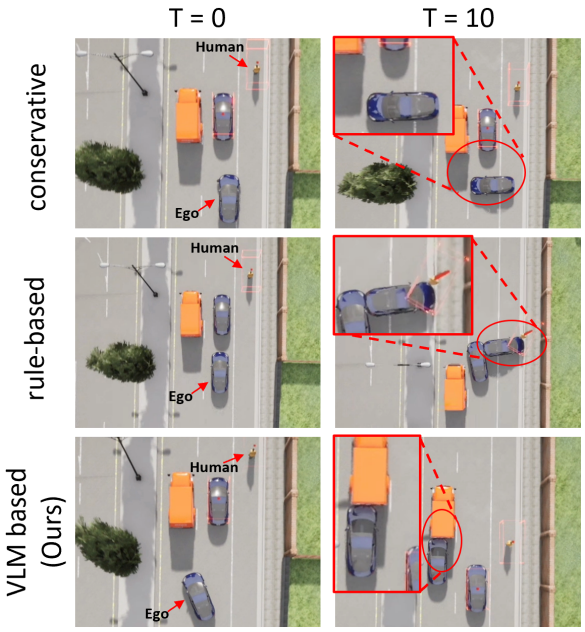


그림 2 상황 이해 능력에 대한 실험 결과

2. **규칙 기반 개척**: 폐쇄 내 우선순위 (오토바이>차량>트럭>사람)에 따라 그 중 우선 순위가 가장 높은 객체를 개척 대상으로 선정한다, 우선순위가 정의되지 않은 객체는 모두 최하위 순위로 분류한다,
3. **VLM 기반 개척(ours)**: VLM을 이용하여 장애물의 종류, 위치, 주변 환경 등을 종합적으로 판단한 뒤 개척 대상을 선정한다, VLM 모델로는 GPT-4o 모델을 사용한다.

VLM 기반 개척 모델의 의사 결정 타당성을 평가하기 위해 2가지 시나리오에서 실험을 진행하였다. 첫 번째 실험은 상황 이해 능력을 평가하였고, 두 번째 실험은 열린 세계에 대한 일반화 능력을 평가하였다.

3.2 실험 결과

상황 이해 능력을 평가한 첫 번째 실험의 결과는 [그림2]에 제시하였다. 보수적 회피 모델은 전방이 완전히 차단된 상황에서 교착 상태에 빠져 개척하지 못하였다. 규칙 기반 개척 모델은 차선 외부에 위치한 보행자를 고려하지 못하고, 사전에 정의된 우선순위에 따라 승용차를 개척 대상으로 선정하였다. 이로 인해 승용차 인근의 보행자에게 예상치 못한 피해가 발생하였다. VLM 기반 개척 모델은 차선 외부에 위치한 보행자를 고려하여 트럭을 개척 대상으로 선정하였다. VLM 응답에서 제시된 선택 이유는 [그림1]의 reasoning과 같다. 이 결과를 통해, 제안한 방법이 규칙 기반 개척 모델에 비해 상황 맥락을 고려한 판단이 가능함을 보여준다.

열린 세계에 대한 일반화 능력을 평가하는 두 번째 실험의 결과는 [그림3]에 제시하였다. 규칙 기반 개척 모델은 사전에 정의된 우선순위 목록에 트래픽 콘이 포함되어 있지 않아 이를 최하위 우선순위로 분류하였다. 그 결과, 트럭을 개척 대상으로 선정하여, 인명 피해를 최소화할 가능성이 높은 대상을 선택하지 않았다. 반면, 제안한 VLM 기반 개척 의사 결정 모델은 사전에 정의된 우선순위 목록에 없는 객체라도 풍부한 지식과 추론 능력을 활용하여 위험도를 판단할 수 있었다. 그 결과, 인명 피해 가능성이 적은 트래픽 콘을 개척하였다. 이를 통해 제안한 방법이 규칙 기반 개척 모델에 비해 열린 세계에서의 일반화가 가능함을 보여준다.

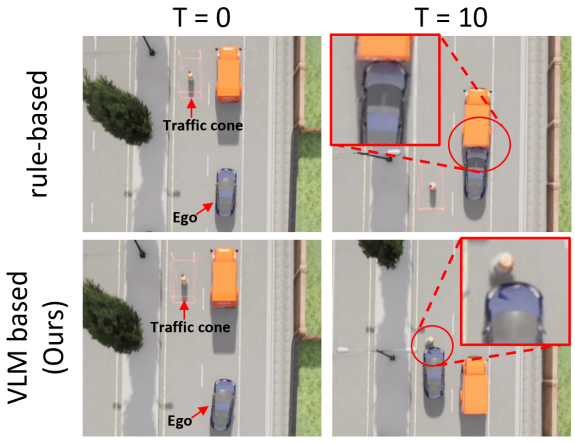


그림 3 열린 세계의 일반화 능력 실험 결과

Ⅳ. 결론

본 연구는 주행 환경에서 자율주행 차량이 회피 경로가 없는 상황에 직면했을 때, VLM을 활용하여 주변 환경과 장애물 특성을 종합적으로 분석하고 최적 개척 대상을 선정하는 의사 결정 방안을 제안하였다. 이를 검증하기 위해 2차선 도로에서 다양한 장애물이 혼재된 시나리오를 구성하고, 보수적 회피 모델, 규칙 기반 개척 모델, 그리고 제안하는 VLM 기반 개척 모델의 성능을 비교하였다. 실험 결과, VLM 기반 모델은 주변 환경을 종합적으로 고려한 의사 결정을 수행하여, 다른 모델 대비 상황 맥락을 반영한 판단이 가능함을 보였다. 이는 VLM이 유연한 의사 결정 능력을 활용해 단순한 규칙 비교를 넘어 인명 피해 가능성을 고려하여 평가할 수 있음을 보여준다. 향후 연구에서는 멀티 에이전트 환경으로 확장하여, 에이전트 간 협력을 통한 개척을 구현할 예정이다.

ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 과학기술전문사관지원센터의 지원을 받아 수행된 연구임 (N01250061-202511)

참 고 문 헌

[1] Subhan K., José G. (2021) *Nonlinear Model Predictive Path-Following Controller for a Small-Scale Autonomous Bulldozer for Accurate Placement of Materials and Debris of Masonry in Construction Contexts*. IEEE Access

[2] Sushant V., Karen L., Ryan K., Yuxiao C., Péter K., Marco Pavone. (2023). Receding Horizon Planning with Rule Hierarchies. IEEE International Conference on Robotics and Automation

[3] Wei X., Noushin M., Anne C., AmitaiY., Emilio F., Radboud D., Calin B. (2021). *Rule-based Optimal Control for Autonomous Driving*. IEEE International Conference on Cyber-Physical Systems

[4] Liu, H., Chen, K., Li, Y., Huang, Z., Liu, M., & Ma, J. (2025). UDMC: Unified Decision-Making and Control Framework for Urban Autonomous Driving With Motion Prediction of Traffic Participants. *IEEE Transactions on Intelligent Transportation Systems*

Reddit 소셜미디어 데이터를 활용한 밈주식의 Contrarian Effect 분석

김건희, 최서빈, 황유진
서강대학교

gunhee0928@sogang.ac.kr, implement@sogang.ac.kr, ughichwang@sogang.ac.kr

Analysis of Contrarian Effects in Meme Stocks Using Reddit Social Media Data

Geonhee Kim, Seobin Choi, Yujin Hwang

Sogang University

요 약

본 논문은 Reddit r/wallstreetbets 데이터를 활용하여 밈주식에서 나타나는 contrarian effect 를 분석하였다. 2021-2023 년 기간 동안 GME, AMC, BB 등 주요 밈주식의 Reddit 관심도와 주가 수익률 간의 관계를 종합적으로 분석한 결과, 일관된 음의 상관관계(-0.198~-0.165)를 발견하였다. 이는 Reddit 에서 관심이 급증할 때 오히려 수익률이 하락하는 contrarian effect 를 의미한다. Event study, 시간적 패턴 분석, 감정 분석, 네트워크 분석을 통해 이 현상의 강건성을 확인하였으며, 53,187 개 Reddit 포스트의 텍스트 분석을 통해 행동재무학적 메커니즘을 실증적으로 검증하였다. 특히 확신 표현이 많을수록 수익률 하락으로 이어지는 'behavioral paradox'를 발견하여 기존 소셜미디어-주식 연구와 차별화된 결과를 제시하였으며, 이를 바탕으로 머신러닝 모델의 성능을 유의미하게 향상시켰으며, 밈주식 시장의 비합리성을 실증적으로 설명할 수 있음을 드러내었다.

Ⅰ. 서 론

2021 년 GameStop 사태로 대표되는 밈주식 현상은 소셜미디어가 금융시장에 미치는 영향에 대한 새로운 관점을 요구한다. 기존 연구들은 소셜미디어 게시물과 주가상승의 양의 상관관계를 보고해 왔으나 [1~4] , 밈주식에서는 소셜미디어에서의 과도한 관심이 오히려 단기 수익률 하락으로 이어지는 contrarian effect 가 관찰되어 기존 이론적 틀에 근본적 의문을 제기한다.

본 연구는 Reddit 데이터를 활용하여 이러한 역설적 현실을 체계적으로 규명하고, 이를 행동재무학적 이론 [5~6]을 결합한 프레임으로 매커니즘을 설명해보고자 한다. 나아가 Reddit 기반 신호를 다양한 머신러닝 모델에 통합하여 기존 선형 접근으로 포착되지 않던 비선형 패턴을 검증하고, contrarian effect 를 활용한 실증적 예측 모델을 구축함으로써 밈주식 시장의 비합리성을 예측 가능하게 함으로써 이를 밈주식 접근 전략으로서 전환하는 가능성을 제시한다.

Ⅱ. 본론

2.1 연구방법

데이터 수집: Yahoo Finance API 를 통해 2021 년 1 월부터 2023 년 12 월까지 GME, AMC, BB 의 일간 주가 데이터를 수집하였다. Reddit API 를 활용하여 r/wallstreetbets 커뮤니티의 게시글 및 댓글 데이터를 수집하였으며, 총 53,187 개의 포스트와 5,409 개의 일간 관측치를 분석에 사용하였다.

주요 변수 정의: Reddit Surprise RS 는 예상 대비 초과 언급량을 나타내는 핵심 지표로 다음과 같이 정의된다:

$$RS_{i,t} = \frac{ActualMentions_{i,t} - ExpectedMentions_{i,t}}{ExpectedMentions_{i,t}}$$

여기서 i 는 종목 t 는 시점을 나타낸다. 수익률은 일간 수익률로 정의하였다:

$$Returns_{i,t} = \frac{P_{i,t} - P_{i,t-1}}{P_{i,t-1}}$$

여기서 $P_{i,t}$는 종목 i 의 t 시점 종가이다.

분석방법: (1) Pearson 상관관계 분석을 통한 Reddit surprise 와 수익률 간 관계 분석, (2) Event Study 를 통한 극한 Reddit 스파이크 전후 수익률 패턴 분석, (3) 시간적 패턴 분석, (4) 53,187 개 포스트에서 행동재무학적 편향 패턴 추출을 위한 텍스트 분석을 수행하였다.

2.2 분석결과

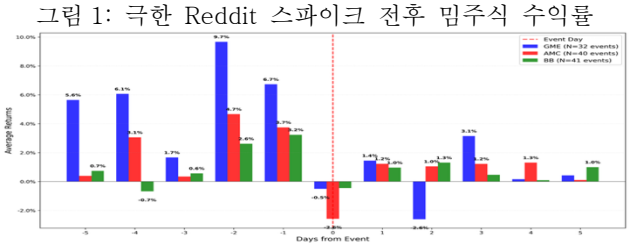
주요 밈주식에서 일관된 음의 상관관계를 발견하였다. GME 의 reddit_surprise vs returns_1d = -0.198, AMC = -0.178, BB = -0.165 로 모든 종목에서 통계적으로 유의한 음의 상관관계가 나타났다. 이는 Reddit 에서 관심이 급증할 때 오히려 다음날 수익률이 하락함을 의미한다.

표 1: Reddit Surprise 와 수익률의 상관관계

Stock	Correlation
GME	-0.198
AMC	-0.178
BB	-0.165

Event Study 결과: 극한 Reddit 스파이크(상위 5%) 113 개 이벤트를 분석한 결과, 이벤트 당일 평균 음수 수익률(-0.005~-0.026), 다음날 부분적 회복하지만 여전히 저조, 2 일 후 추가 조정이 지속되는 패턴을 관찰하였다.

그림 1: 극한 Reddit 스파이크 전후 밈주식 수익률



53,187 개 Reddit 포스트의 텍스트 분석을 통해 'Behavioral Paradox'를 발견하였다. 특히 확신 표현이 많을수록 다음날 수익률이 유의하게 하락하는 과잉확신 역설(confidence_score vs next_day_returns:-

0.205,p=0.0002)과 언급된 가격에 투자자들이 동조되는 Price Anchoring Effect(price_mentions vs volatility: +0.472)와 FOMO Reversal(urgency_score vs reddit_surprise: -0.127) 효과가 관찰되어 Reddit 관심 급증이 체계적인 단기 가격 하락을 예측하는 contrarian indicator 로 작동함을 실증하였다.

2.3 메커니즘 해석

밈주식의 contrarian effect 는 소셜미디어 언급 급증이 정보 개선이 아닌 행동적•미시구조적 충격으로 작용하는 체계적 메커니즘을 통해 설명된다. Reddit 언급 폭증은 투자자들의 주의를 특정 종목에 급격히 집중시켜 동조화된 거래행동을 유발하며, 이는 과잉확신 편향과 FOMO 심리가 결합되어 비합리적 거래량 폭증과 단기 과매수를 초래한다. 이 과정에서 호가 스프레드 확대와 시장심도 축소 등 미시구조적 마찰이 발생하고, 높은 공매도 비용과 소음거래자 위험으로 인해 합리적 투자자의 즉각적 차익거래가 제한되어 시장 비효율성이 지속되며, 결국 관심이 분산되고 유동성이 정상화되는 2-3 일 후 지연된 가격 조정이 실현되어 Reddit 언급 급증 이후 체계적인 음의 수익률이 나타나는 "예측 가능한 비합리성"패턴을 형성한다.

2.4 머신러닝 적용

그림 2: Reddit 적용 전후 모델 성능 변화

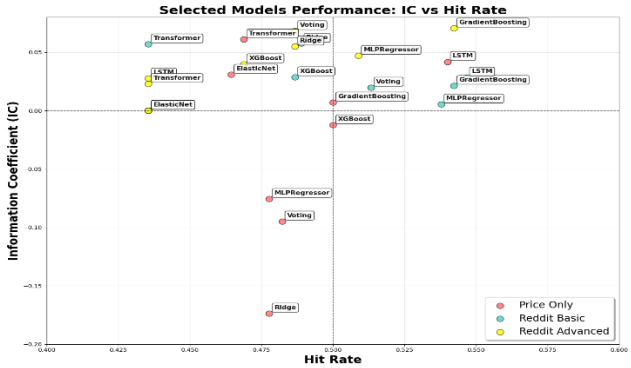


표 2 : IC, 적중률 및 ICIR 지표를 통한 모델 성능 비교

Model	IC	Hit Rate	ICIR
Transformer_price	0.061	0.469	0.661
GradientBossting_price	0.007	0.500	0.513
Voting_adv	0.068	0.487	0.609
Ridge_basic	0.057	0.489	0.468
Transformer_basic	0.057	0.436	-0.453
GradientBoosting_adv	0.070	0.542	0.753

Contrarian effect 와 overconfidence paradox 의 실증분석을 넘어 Reddit 기반 신호가 실제 예측력에 기여하는지를 검증하기 위하여 다양한 머신러닝 모델을 적용하였다. 실험 설계는 가격 데이터만을 활용한 Price Only 모델을 baseline 으로 설정하고, Reddit 기반 contrarian 신호를 추가한 Reddit Basic 및 Reddit Advanced 모델과 비교하는 방식으로 수행하였다.

Reddit 데이터 기반의 contrarian effect 가 적용된 모델들은 전반적으로 price only 모델에 비해 IC 와 Hit Rate 가 개선된 양상을보였다. 특히 Gradient Boosting 모델은 Reddit Advanced 피처를 적용했을 때 IC=0.070, Hit Rate=0.542, ICIR=0.753 을 기록하며, baseline(GradientBoosting_price_only, IC=0.007, Hit Rate = 0.5) 대비 현저한 향상을 나타냈다.

이러한 결과는 Reddit 기반 contrarian 신호가 단순한 설명 변수에 그치지 않고, 실제 밈 주식 단기 수익률을 예측하는 데 기여할 수 있음을 의미한다. 특히 Gradient Boosting 과 같은 비선형 앙상블 모델에서 가장 큰 개선 효과가 관찰된 것은 Reddit 관심 급증이 단순히 투자 심리의 반영을 넘어 예측 가능한 역행적 패턴으로 작동함을 머신러닝 수준에서 실증적으로 확인한 것이다.

Ⅲ. 결론

본 연구는 Reddit 데이터를 통해 밈주식의 contrarian effect 를 발견하고 주의집중 이론과 과잉확신 편향을 중심으로 한 행동재무학 이론으로 이를 설명하였다. 소셜미디어 언급량 급증이 행동적 미시구조적 충격으로 작용하여 단기 주가 하락을 유발함을 실증하였으며 이는 효율적 시장가설의 한계를 보여준다.

핵심 기여점은 네 가지이다. 첫째, 소셜미디어 관심도와 수익률 간 일관된 음의 상관관계라는 기존 연구와 차별화된 현상을 발견하였다. 둘째, 투자자의 확신 수준과 발언 강도 간 역설적 관계를 통해 '행동적 역설'이라는 개념을 제시하였다. 셋째, 주의 재배분, 집단조정, 차익거래 제약이라는 세 가지 메커니즘을 통해 현상을 체계적으로 설명하였다. 넷째, Reddit 데이터 적용으로 머신러닝 모델의 성능을 개선시켜, 소셜미디어 데이터가 단순 노이즈가 아닌 실질적 예측 신호임을 입증하였다.

본 연구의 발견은 디지털 플랫폼이 금융시장에 미치는 영향에 대한 이해를 확장하며 투자자 교육과 시장 규제 정책 수립에 시사점을 제공한다. 향후 연구에서는 다양한 소셜미디어 플랫폼과 시장 조건에서의 일반화 가능성을 검증하고 장기적 관점에서의 시장 효율성 회복 과정을 보다 정밀하게 분석할 필요가 있다.

ACKNOWLEDGMENT

본 연구는 서강대학교의 국제지역문화원의 연구비 지원을 받아 수행되었습니다.

참 고 문 헌

[1] Bollen, J., Mao, H., & Zeng, X. "Twitter mood predicts the stock market," Journal of computational science, vol. 2, no. 1, pp. 1-8, 2011.

[2] Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., & Mozetič, I. "The effects of Twitter sentiment on stock price returns," PloS one, vol. 10, no. 9, e0138441, 2015

[3] Mittal, A., & Goel, A. "Stock prediction using twitter sentiment analysis," Stanford University CS229 Machine Learning Final Project, pp. 1-5, 2012

[4] Boehmer, E., Jones, C. M., Zhang, X., & Zhang, X. "Tracking retail investor activity," The Journal of Finance, vol. 76, no. 5, pp. 2249-2305, 2021.

[5] Barber, B. M., & Odean, T. "All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors," The review of financial studies, vol. 21, no. 2, pp. 785-818, 2008.

[6] Daniel, K., Hirshleifer, D., & Subrahmanyam, A."Investor psychology and security market under-and overreactions," The journal of Finance, vol. 53, no. 6, pp. 1839-1885, 1998.

실시간 ESS 화재 예측을 위한 Convolutional Autoencoder

이연서, 김현정, 김파란하늘, 현장훈*
국립한밭대학교
20231068@edu.hanbat.ac.kr, *jhhyeon@hanbat.ac.kr

Real-Time Fire Prediction in ESS via Convoltional Autoencoder

Yeonseo Lee, Hyeonjeong Kim, Paranhaneul Kim, Janghun Hyeon*
Hanbat National University

요 약

최근 ESS의 사용이 증가하며 ESS 화재 또한 급증하고 있다. 리튬이온 ESS 화재의 대표적인 원인은 리튬이온 배터리의 열폭주이며, 열폭주로 인한 화재는 인접한 배터리로 빠르게 전파될 수 있기 때문에 사전에 예측해야한다. 대표적인 화재 예측 알고리즘인 임계값 기반 알고리즘은 측정된 값이 사전에 정의된 임계값을 초과하면 화재라고 감지하기 때문에 노이즈에 민감하다는 문제점이 존재한다. 따라서, 본 논문에서는 노이즈 민감도를 줄이기 위해 Sliding Window와 Spectrogram을 사용하여 데이터를 재구성한 후, Convolutional Autoencoder를 이용하여 실시간 화재 예측을 수행하는 방법을 제안한다. 제안한 알고리즘은 본 연구에서 구축한 Test 데이터에 대해서 Accuracy 1.0, Recall 1.0, Precision 1.0을 보여주며 화재 예측을 정확하게 수행하였다.

Ⅰ. Introduction

최근 전력을 저장해두었다가 필요한 시기에 전력을 공급하는 시스템인 ESS(Energy Storage System)에 대한 관심이 증가하며 친환경 에너지 분야에서의 ESS 사용이 증가하고 있다. ESS 사용이 증가함에 따라 ESS 화재도 빈번히 발생하고 있다. 리튬이온 배터리를 사용하는 ESS 화재는 주로 리튬이온배터리의 열폭주로 인해 발생하며, 열폭주로 인한 화재는 인접한 배터리로 화재가 전파될 수 있기 때문에 미리 대비하기 위해 예측을 하여야한다.

화재를 예측하기 위하여 주로 사용되는 임계값 기반의 화재 예측 알고리즘은 특정 상황에 대한 지표값이 사전에 정의된 임계값을 초과하면 화재가 감지되는 간단한 방식을 사용하여 노이즈가 존재할 때 높은 오경보율을 나타낼 수 있다[1].

특정 상황에 대한 임계값만을 사용하여 노이즈에 민감하다는 기존 알고리즘의 문제점을 보완하기 위해 본 논문에서는 Spectrogram을 사용하여 데이터를 재구성하고, Convolutional Autoencoder를 수행하는 알고리즘을 제안한다.

Ⅱ. Proposed Method

1) 전체 알고리즘

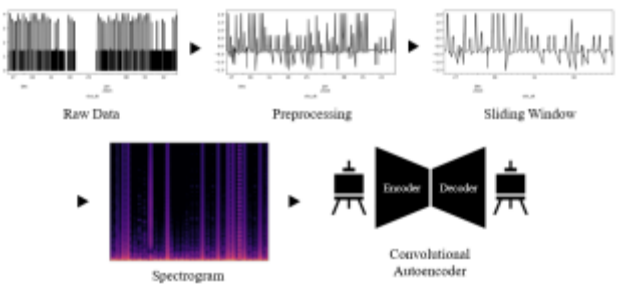


그림 1. 전체 알고리즘

본 연구에서 제안한 알고리즘은 [그림 1]과 같다. ESS 배터리의 Cell에서 1분 단위로 측정된 전압값을 불러온 후 보간, 스케일링과 같은 전처리를 진행한다. 데이터를 실시간으로 처리하는 것과 비슷한 형태로 데이터를 재구성하기 위해 Sliding Winodw를 사용하여 전체 날짜로 구성된 데이터를 일정한 구간으로 나눈다. 시간 정보와 주파수 성분의 정보를 모두 표현할 수 있는 Short-Time Fourier Transform을 수행한 뒤 시간, 주파수, 해당 시간대의 주파수 강도를 모두 나타낼 수 있는 Spectrogram으로 이미지화한다. 이미지화된 데이터를 Convolutional Autoencoder의 입력으로 넣어 이상탐지를 수행하여 화재를 예측한다.

2) 데이터셋 구성

ESS 배터리는 Rack, Module, Cell로 구성된다. 연구에 사용한 데이터는 Cell에서 1분 단위로 측정된 전압값이며, 화재가 난 데이터와 화재가 나지 않은 데이터로 구분된다.

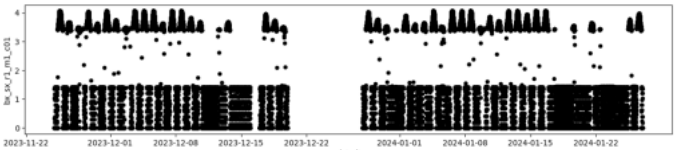


그림 2. 결측된 데이터 시각화

각 데이터의 측정날짜의 길이는 31일부터 182일까지 다양하게 존재하며 하드웨어의 결함이나 네트워크 문제와 같은 불안정한 센싱 환경에 노출되는 경우가 많기에 [그림 2]와 같이 결측된 날짜가 존재한다[2].

위와 같은 이유로 측정이 되지 않을 경우, 결측치는 일 단위로 존재한다.

3) 전처리

ESS 배터리에 따라 Module 안에 존재하는 Cell의 개수가 달라진다. Module을 구성하는 Cell의 개수가 다를 경우 STFT를 수행하는 길이가 달라지기 때문에 Spectrogram의 시간 축의 길이가 달라질 수 있다. 따라서, Module을 구성하는 각 Cell들의 STFT 수행 결과를 모두 더하여 Module

별로 하나의 Spectrogram으로 이미지화한다.

결측치 이전의 값과 결측치 모두 화재가 난 데이터의 패턴이 관찰된다면 화재라고 예측할 것이고, 두 값 모두 화재가 나지 않은 데이터의 패턴이 관찰된다면 화재가 나지 않았다고 예측할 것이다.

따라서 상술한 결측치를 처리하기 위해 결측된 날짜가 하루일 경우 결측된 날짜의 바로 이전 날짜의 데이터로 복사하고 연속 2일 이상일 경우 결측된 날짜로부터 결측된 길이만큼의 이전 데이터를 복사한다.

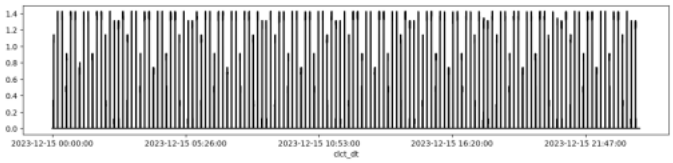


그림 3. 이상치 데이터 그래프

[그림 3]과 같이 결측이 되지 않았어도 센서 오류 등의 이유로 인해 제대로 측정되지 않은 데이터가 존재할 수 있다. 값을 복사하여 그대로 사용하는 방법에서 제대로 측정되지 않은 값을 사용하는 것은 노이즈로 취급될 수 있다. 이를 해결하기 위해 하루의 데이터에서 방전종지전압인 3.3V 이하 수치가 일정 비율 이상일 경우 해당 날짜도 결측치로 처리하고 위와 같은 방법으로 보간을 진행하였다.

4) Convolutional Autoencoder

Convolutional Autoencoder[3]은 1차원 데이터를 입력으로 하는 Autoencoder와 동일한 구조를 가지지만 Convoltional layer를 사용하여 Autoencoder보다 2차원 데이터의 특징을 추출하는 데 더 적합하다.

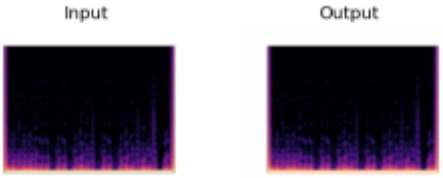


그림 4. 원본 데이터와 Convolutional Autoencoder의 출력

본 연구에서는 [그림 4]와 같이 Spectrogram으로 이미지화한 데이터를 Encoder의 입력으로 하며, Decoder를 거쳐 다시 생성된 Spectrogram을 출력으로 한다.

Autoencoder를 사용하여 이상탐지를 진행하는 방식은 정상 데이터를 사용하여 Train을 진행하고 모든 데이터에 대해서 Test를 진행하면서 재건오차를 비교한다. 위의 방법을 사용하여 본 연구에서는 화재가 나지 않은 데이터로만 Train을 진행하며, 화재가 나지 않은 데이터와 화재가 난 데이터 모두 사용하여 Test를 진행한다.

5) 실험 과정

	Date Length	Rack * Module	Total
Train	33	840	27720
Validation	33	146	4845
Test	35	128	4471

표 1. 데이터셋 구성

Train 데이터는 ESS 배터리 4개, Validation 데이터는 화재가 난 ESS 배터리 1개와 화재가 나지 않은 ESS 배터리 1개, Test 데이터도 화재가 난 ESS 배터리와 화재가 나지 않은 ESS 배터리 각각 1개씩 사용하였다. [표 1]은 본 연구에서 사용한 데이터 날짜의 평균, Rack*Module의 수, 총 개수를 나타낸 것이다.

Sliding Window 기법을 사용하여 시계열 데이터를 30일 단위로 재구성하여 전처리를 진행한다. ESS 배터리의 각 Celll 별로 STFT를 수행한 후 각 시간대에 해당하는 각 Cell의 STFT 수행 결과를 더하여 각 Module을 표현할 수 있는 하나의 Spectrogram으로 이미지화한다. Spectrogram을 Convolutional Autoencoder의 입력으로 사용한다.

Convolutional Autoencoder를 사용하여 화재가 난 데이터와 화재가 나지 않은 데이터의 재건오차를 구하여 임의의 임계값 이상이 되면 화재로 예측한다. 이때의 임계값은 Validation으로 사용된 화재가 난 데이터와 화재가 나지 않은 데이터의 재건오차 평균인 0.000107을 사용한다.

6) 결과

Accuracy	Recall	Precision
1.0	1.0	1.0

표 2. Test 데이터 성능

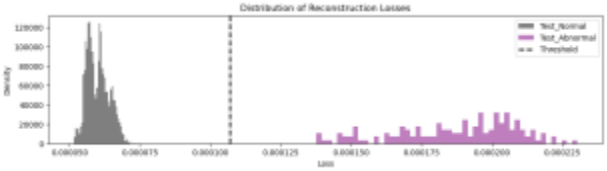


그림 5. Test 데이터 재건오차 분포

[그림 5]는 Test 데이터의 재건오차 분포를 시각화한 것이다. x축은 재건오차, y축은 해당 재건오차의 빈도수를 나타낸다. 화재가 난 데이터와 화재가 나지 않은 데이터의 재건오차 분포를 보면 임계값을 기준으로 명확하게 분류할 수 있음을 알 수 있다.

이는 [표 2]와 같이 모든 Test 데이터에 대해서 완벽하게 분류할 수 있다는 것을 의미한다. 특히, Recall이 1.0이라는 수치는 본 논문에서 제안한 알고리즘이 화재를 조기에 예측해야하는 목표를 달성했음을 의미한다.

Ⅲ. Conclusion

본 논문에서는 임계값만을 사용하여 화재를 예측한다는 기존 알고리즘의 문제점을 해결하기 위해 Spectrogram으로 이미지화한 데이터로 Convolutional Autoencoder를 수행하는 알고리즘을 제안한다.

Validation 데이터 재건오차의 평균을 임계값으로 설정하여 Test 데이터의 재건오차를 기준으로 분류한 결과 Accuracy 1.0, Recall 1.0, Precision 1.0으로 모든 데이터를 제대로 분류하는 것을 알 수 있다.

다만 Test에 사용한 데이터는 한 제조사에서 생성한 ESS 배터리의 데이터이므로 추후에 다양한 제조사의 데이터에서도 활용할 수 있도록 연구할 예정이다.

References

[1] Y. Heo, S. Seo, W. Shim, and J. Kang, “A Study on the Real-time Fire Classification Algorithm Based on Stacked LSTM,” J. Korean Soc. Hazard Mitig., vol. 21, no. 3, pp. 93-104, Jun. 2021. doi: 10.9798/KOSHAM.2021.21.3.93

[2] 정재익, 구태연, and 박완기, "협업필터링 오토인코더 기반 센서 데이터 결측치 처리," in 한국통신학회 학술대회논문집, 2022, pp. 80-81.

[3] X.-J. Mao, C. Shen, and Y.-B. Yang, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," *in Proc. Advances in Neural Information Processing Systems (NeurIPS)*, 2016, pp. 2802-2810.

히트맵 기반 적응형 이미지 슬라이싱을 통한 소형 객체 탐지 : 건설현장 적용 사례

한일주[1,3], 장다연[2,3], 김광수[3], 김봉완[3]*

[1]충남대학교, [2]성결대학교, *[3]한국전자통신연구원 에어모빌리티연구본부

martellato41@o.cnu.ac.kr, clavexn@gmail.com, {enoch, kimbw}@etri.re.kr

Heatmap-Based Adaptive Image Slicing for Small Object Detection : Application to Construction Sites

Il-Ju Hahn, DaYeon Jang, Kwangsoo Kim, Bong Wan Kim*

ChungNam Nat'l Univ., Sungkyul Univ.,

*Electronics and Telecommunications Research Institute (ETRI)

요 약

건설현장의 소형 객체는 저해상도와 잡음으로 탐지가 어렵다. 기존 소형 객체 탐지를 위한 방법론은 고정 패치 크기로 긴 추론 시간과 중복 탐지를 유발한다. 본 연구에서는 히트맵 기반 적응형 슬라이싱 프레임워크를 제안한다. CenterNet 기반의 경량 히트맵 네트워크로 객체 중심점 확률 맵을 생성하고, 이를 기반으로 소형 객체 밀집 영역은 정밀 추출과 동시에 대형 객체 영역은 포괄적으로 처리하는 적응형 슬라이싱 전략을 제안한다. 실제 건설현장 데이터에 대하여 소형 객체 탐지 성능 향상이 mAP@0.5:0.95(small) 지표 기준 79% 있었으며, 기존 방법론 대비 적은 슬라이싱 수로 추론 시간을 31% 줄여 정확성과 효율성의 균형을 달성하였다.

Ⅰ. 서 론

최근 컴퓨터 비전 분야의 객체 탐지 기술은 자율주행, 안전 관리, 산업 자동화 등 다양한 영역에서 핵심 기술로 활용되고 있다. 특히 건설현장은 안전사고 예방 측면에서 객체 탐지 기술의 활용 가능성이 매우 높다. 그러나 건설현장의 영상 데이터는 넓은 촬영 범위와 다양한 크기의 객체가 공존하는 특수성을 갖는다. 이 중 소형 객체는 저해상도, 주변 잡음 등으로 인해 탐지가 더욱 어렵다.

기존 연구에서는 소형 객체 탐지를 위해 슬라이딩 윈도우 이미지 슬라이싱 방법론이 사용되었다. Slicing Aided Hyper Inference(SAHI, Akyon et al., 2022)와 같은 정적 슬라이싱 방식은 고정 패치 크기와 겹침으로 인해 긴 추론 시간, 객체 잘림, 중복 탐지 문제를 발생시켰다. 따라서 크기 스케일이 이질적인 객체가 공존하는 대규모 장면에서는 소형 객체 탐지 성능 개선에 한계가 있다.

본 논문에서는 이러한 문제 해결을 위해 '히트맵 기반 적응형 슬라이싱' 프레임워크를 제안한다. CenterNet(Zhou et al., 2019)에서 제안된 방식에 착안하여, 객체의 중심점을 확률적으로 표현하는 히트맵을 생성하고, 이를 기반으로 소형 객체 밀집 영역은 정밀 추출하되 대형 객체 영역은 불필요한 절단 없이 포괄적으로 처리하는 적응형 분할 전략을 사용한다.

본 연구의 기여는 다음과 같다. 첫째, 히트맵 기반 적응형 슬라이싱으로 소형 객체 탐지 성능을 향상시켰다. 둘째, 경량 히트맵 생성 모델로 기존 정적 슬라이싱 방식 대비 효율성을 확보하였다. 셋째, 건설현장 데이터 실험을 통해 산업적 실효성을 실증하였다.

Ⅱ. 본론

본 논문에서 제안하는 히트맵 기반 적응형 슬라이싱 프레임워크는 두 단계의 탐지 과정으로 구성된다. 첫 번째 단계에서는 입력 이미지에 대해 경량화된 히트맵 생성 네트워크(HmNet)를 적용하여 객체 중심점의 확률 분포 맵을 생성한다. 생성된 히트맵은 적응형 분할기에 전달되어 객체가 있을 법한 영역을 동적으로 식별하고, 이를 기반으로 이미지를 최적 크기 패치들로 분할 한다. 두 번째 단계에서는 분할된 패치와 원본 이미지를 함께 객체 탐지 모델에 입력하여 각각의 탐지 결과를 얻는다. 중복 탐지 결과에 대한 후처리 과정을 통해 최종적으로 통합된 탐지 결과를 산출한다.

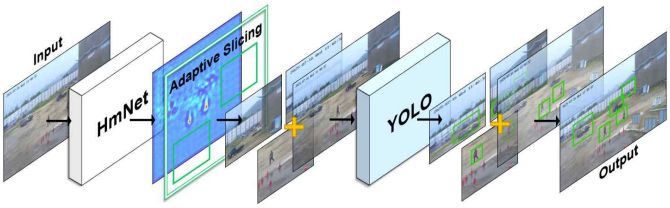


그림 1 히트맵 기반 적응형 슬라이싱 프레임워크

학습 과정에서는 HmNet은 원본 이미지를 대상으로 학습되며, 객체 탐지 모델은 원본 이미지로만 학습한 "기본학습" 방법과 HmNet에서 생성된 분할 패치와 원본 이미지를 모두 활용한 "분할학습" 방법으로 나뉜다. 분할학습을 통해 확대된 소형 객체를 모델이 더 상세하게 학습할 수 있으며 성능 향상을 기대할 수 있다. 추론 단계에서는 객체 탐지 모델의 학습 방식과 무관하게 HmNet이 생성한 분할 패치와 원본 이미지를 동시에 추론하고 그 결과를 병합하는 구조를 적용할 수 있다.

2.1 히트맵 생성 네트워크(HmNet)

HmNet은 객체 탐지 모델 중 하나인 CenterNet의 히트맵 생성 인코더-디코더 구조를 차용하였으며, 실시간 처리 성능을 고려하여 ResNet-18을 백본 네트워크로 채택하였다. 각 객체의 바운딩 박스 중심점을 2D 가우시안 분포로 변환하여 정답(ground truth) 히트맵을 생성하였으며 가우시안 커널의 표준편차는 객체 바운딩 박스 크기에 비례하도록 설정하여 소형

객체는 더 좁은 분포를, 대형 객체는 더 넓은 분포를 갖도록 하였다. HmNet 학습 시 인코더 단계에서는 ResNet-18을 통해 입력 이미지로부터 특징 맵을 추출하고, 디코더 단계에서는 Deformable Convolution을 통해 업샘플링하여 최종적으로 128×128 크기의 히트맵을 출력한다.

2.2 히트맵 기반 적응형 슬라이싱(Adaptive Slicing)

적응형 분할기는 히트맵에서 검출된 피크들을 그룹 단위로 통합한 후 이들을 포함하는 패치를 슬라이싱하는 방식으로 동작한다. 먼저 히트맵에서 상위 k개의 피크를 검출한 뒤, 각 피크의 주변 평균 값과 비교하여 충분히 두드러진 피크만을 선택한다. 선택된 피크들은 Union-Find 알고리즘을 통해 그룹화된다. 또한 그룹 경계에 위치한 피크들의 Gaussian blob 크기를 추정하여 이에 비례하는 패딩을 적용함으로써, 객체의 크기를 반영한 패치 크기를 결정한다. 최종적으로 히트맵 상의 패치 좌표를 이미지의 패치 좌표로 변환하여 슬라이싱한다. 이러한 방식은 기존의 패치 크기가 고정된 슬라이싱 방식과 달리 객체의 예상 분포에 따라 패치 크기와 위치를 동적으로 조정함으로써, 대형 객체에 대해서는 불필요한 분할을 줄이고 소형 객체에 대해서는 탐지 모델이 해당 부분을 확대하여 학습·추론할 수 있다.

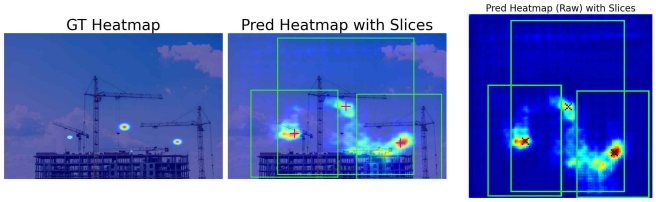


그림 2 적응형 슬라이싱

2.3 객체 탐지 및 결과 병합

하나의 이미지에 대해 추출된 패치들과 원본 이미지는 각각 객체 탐지 모델에 입력되어 추론을 수행한다. 이 과정에서 동일 객체가 여러 패치에서 중복 탐지 되거나, 원본 이미지와 패치에서 동시에 탐지될 수 있다. 이러한 중복 탐지 결과를 병합하기 위해서 Non-Maximum Suppression (NMS) 후처리 방법을 사용하여 Intersection over Union(IoU) 임계값 기반으로 중복된 탐지 박스를 제거하였다. 추가적으로 Weighted Boxes Fusion(WBF)을 적용하여 중복된 탐지 결과들의 신뢰도를 가중 평균하여 더 정확한 최종 탐지 박스를 생성하였다. 실험 결과 WBF가 NMS보다 소형 객체에 대해 높은 정확도를 보였다. 최종적으로 병합된 탐지 결과는 원본 이미지 좌표계로 변환되어 출력된다.

2.4 실험 설계 및 결과

본 연구에서 사용한 건설 현장 이미지 데이터는 작업자, 비 작업자, 안전 헬멧, 안전 조끼뿐만 아니라 굴착기, 덤프트럭, 콘크리트 믹서, 지게차, 타워크레인 등 다양한 중장비를 포함한 총 29개 클래스로 구성되어 있다. 데이터의 다양성을 확보하기 위해 세 가지 경로에서 이미지를 수집하였다. i) 실제 건설 현장 CCTV 영상(2025년 7월 ~ 2025년 8월 수집). ii) AI-HUB의 "건설 현장 위험 상태 판단 데이터". iii) 이미지 크롤링(예: "dump truck at construction site" 검색어 기반 이미지 검색 결과 활용) 등을 통해 다양한 중장비를 포함한 건설 현장 이미지를 확보하였다. 결과적으로 총 5,883개의 객체를 포함하는 1,511장의 이미지 데이터를 수집하였으며, 188×188부터 최대 5,600×8,400 크기의 다양한 해상도로 구성되어 있다. Train, Validation, Test 데이터를 각각 7:2:1로 분할 하였으며, 객체 크기에 따른 정량적 성능 검증을 위해 COCOeval 프로토콜을 적용하였다. 객체 탐지 모델로는 YOLO11n 모델을 사용하였다. 성능지표의 경우 객체 탐지 성능 평가의 표준지표인 mAP 계열 지표와 F1 score를, 추론 효율성 측정을 위한 지표인 FPS를 사용하였다.

표 1 COCOeval 프로토콜 크기($pixel^2$)별 객체 수

small : $< 32^2$px	medium : $32^2 \leq$ area $< 96^2$px	large : $\geq 96^2$px
811	1729	3343

제안 방법과의 성능 비교를 위해 비교 대상을 다음과 같이 설정하였다. 학습에는 YOLO11n을 사용하였고, 추론에는 SAHI를 적용하였다. 이때 SAHI는 512×512 패치 크기를 기준으로, 패치 간 20% 겹침을 허용하는 기본 설정값을 사용하였다. 한편, 제안 방법은 패치와 원본 이미지를 합쳐 총 5,823개의 이미지를 활용하여 "분할학습"을 수행하였고, 추론 단계에서는 히트맵 기반 적응형 슬라이싱을 활용하였다.

표 2 비교 실험 결과

	YOLO11n + SAHI	Proposed
mAP@0.5	**0.702**	0.572
mAP@0.5:0.95 | small	0.217	**0.389**
mAP@0.5:0.95 | medium	0.255	**0.288**
mAP@0.5:0.95 | large	**0.570**	0.421
Macro F1	0.452	**0.506**
Micro F1	0.628	**0.714**
평균 FPS	5.2	**7.5**
추론 시 생성 패치 수	1564	**1018**

비교 실험 결과, 제안 방법은 소형 객체 탐지에 있어 SAHI 대비 mAP@0.5:0.95가 약 79% 향상되어 0.217에서 0.389로 개선되었다. 추론 효율성 측면에서도 기존 방법 대비 44%의 FPS 향상을 보였다. 전반적인 mAP 성능은 0.702에서 0.572로 약 19% 감소하였으나, 실시간 추론이 가능한 수준의 효율성을 달성했다는 점과 더불어 소형 객체 탐지에서 가장 좋은 성능을 보였다는 점에서 실제 현장에서의 적용 가능성을 기대할 수 있다.

Ⅲ. 결론

본 연구에서는 건설 현장과 같이 다양한 크기의 객체가 혼재된 환경에서 기존 방법의 한계를 보완하기 위해 히트맵 기반 적응형 슬라이싱 프레임워크를 제안하였다. 제안된 방법은 경량 히트맵 네트워크를 통해 객체 중심 분포를 추정하고 이를 기반으로 관심 영역을 동적으로 분할한 뒤, 원본 이미지와 패치 이미지를 함께 탐지하여 결과를 통합하는 방식으로 소형 객체 탐지 성능과 추론 효율성을 동시에 확보하였다. 실제 건설 현장 데이터셋을 이용한 실험 결과, 제안 방법은 비교 방법론에 비해 소형 객체에서 높은 mAP를 달성하였다. 또한 불필요한 패치 생성을 줄여 추론 속도를 개선함으로써 효율성도 동시에 개선하였다. 향후 연구에서는 더욱 정교한 관심 영역 제안과 다양한 도메인 데이터셋에 대한 일반화 검증을 통해 실시간 응용 가능성을 넓힐 수 있을 것으로 기대한다.

ACKNOWLEDGMENT

본 연구는 국토교통부/국토교통과학기술진흥원의 지원으로 수행되었음(과제번호 RS-2020-KA156208)

참 고 문 헌

[1] Akyon F. C., Altinuc S. O., and Temizel A., "Slicing Aided Hyper Inference and Fine-tuning for Small Object Detection," 2022 IEEE International Conference on Image Processing (ICIP), 2022.

[2] Zhou X., Wang D., and Krähenbühl P., "Objects as Points," arXiv preprint arXiv:1904.07850, 2019.

엣지 컴퓨팅 환경인 Raspberry Pi 5에서 가역 신경망 HiNet 이미지 은닉모델의 구현에 관한 연구

김정훈, 김세빈, 정준

동서울대학교

thdnfrhfem@gmaill.com, sqlsbin@gmail.com, jj81271000@gmail.com

A study on the implementation of the HiNet image hidden model, a reversible neural network, on the Raspberry Pi 5, an edge computing environment.

JungHum Kim, SeBin Kim, Jun Jeong

DongSeoul Univ.

요 약

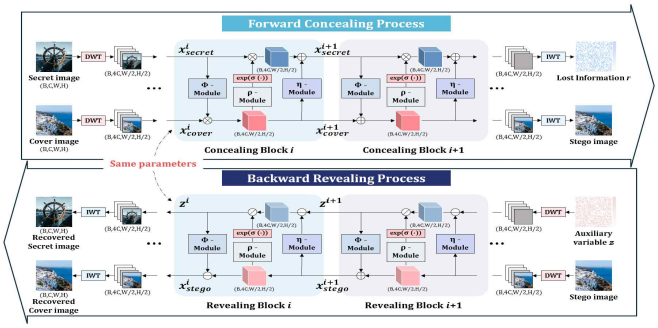
본 연구는 가역 신경망 구조의 HiNet 이미지 은닉 모델을 엣지 컴퓨팅 환경인 Raspberry Pi 5 상에서 원활히 동작하도록 기존 모델의 크기, 추론 시간, 품질·보안성을 고려한 최적화에 관한 것이다. 기존의 HiNet 모델의 경량화를 위해 PTQ(Post-Training Quantization)를 적용하여 일부 연산만을 INT(Integer) 8비트로 변환하고, 가역성에 민감한 계층은 FP(Float)32 비트로 유지하여 복원 성능 저하를 최소화하였다. 본 연구에서 제안한 방법을 모델의 경량화, 추론 시간, 품질·보안 부분에 대해서 기존 방식과 비교 평가하였다. 실험 결과, 제안 방식이 기존 방식에 비해 모델 크기는 57.3% 감소하였고, 추론 시간은 320[sec] 단축하였고 CPU 점유율, memory 사용량 또한 각각 51.1%, 24.9% 감소하였다. 품질 부분 측정치인 PSNR(Peak Signal-to-Noise Ratio)과 SSIM(Structural Similarity Index)은 각각 5%, 0.6% 만큼 감소하였다. 보안 부분의 측정치인 AUC(Area Under the ROC Curve)는 보안 성능이 9% 감소하는 결과를 얻었지만 엣지 컴퓨팅 환경인 Raspberry Pi 5에서 원활하게 동작하는 것을 확인하였다.

Ⅰ. 서 론

스테가노그래피(Steganography)와 워터마킹(Watermarking)은 디지털 이미지 내에 시크릿(Secret) 이미지를 삽입하거나 복원하면서도 원본의 가시적 품질을 유지하는 기술로, 저작권 보호와 프라이버시 보장에 중요한 역할을 한다[1]. 그러나 기존의 신경망 기반 접근법들은 주로 GPU가 탑재된 서버 환경에서 구동을 전제로 설계되었다. 메모리·연산 자원이 제한된 엣지 컴퓨팅 환경에서는 동작이 어렵다는 한계가 있다[2]. 본 연구에서는 가역 신경망 구조의 HiNet 모델을 Raspberry Pi 5와 같은 엣지 컴퓨팅에서 동작하도록 경량화하기 위한 양자화 기법인 PTQ(Post-Training Quantization)을 적용하였다. PTQ는 FP32로 학습된 모델을 추가 학습 없이 INT8로 변환해 모델 크기와 메모리 사용을 줄이는 방식이다[3].

Ⅱ. HiNet

HiNet은 가역 신경망(INN)으로, 은닉(Conceal) 경로 (cover+secret → stego)와 복원(Reveal) 경로 (stego → cover, secret)가 구조적으로 동일하며 같은 파라미터를 사용한다. 이러한 가역성 덕분에 정보 손실을 최소화하고 복원(Reveal) 오차를 줄일 수 있다. HiNet 모델의 개요는 그림 1과 같다.



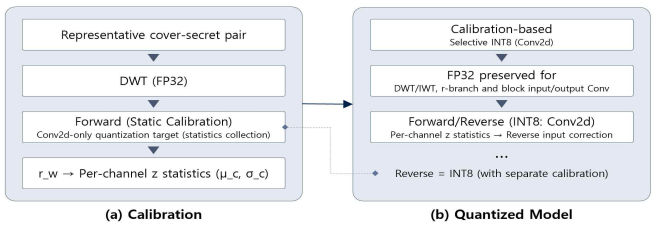
[그림 1] HiNet 전체 구조

Ⅲ. PTQ(Post-TrainingQuantization)

엣지 컴퓨팅과 같이 리소스(Resource)가 제한된 환경에서 HiNet 모델을 활용하기 위해서는 모델 경량화가 필수적이다. PTQ는 부동소수점(FP32)으로 표현된 가중치와 연산을 더 낮은 정밀도의 정수(INT8)로 변환하여 모델 크기를 줄이고, 추론 속도를 향상시키는 대표적인 양자화 기법이다. 그러나 연산 효율을 높이는 동시에 은닉·복원 품질 저하가 발생할 수 있는 한계가 존재하며 이를 보완하기 위해 본 연구에서는 파인튜닝과 결합한 PTQ 방식의 양자화 기법을 제안하여 적용하였다.

Ⅳ. 제안한 방식

모델의 경량화를 위한 PTQ 과정은 대표 이미지 쌍을 이용한 정적 조정(Static Calibration)을 통해 Conv2d 레이어(Layer)만 FP32에서 선택적으로 INT8로 변환하여 정확도 손실을 최소화하였으며, 제안한 PTQ 과정은 그림 2와 같다. 본 연구에서는 원본 HiNet을 워터마크 데이터셋으로 소규모 미세 조정(Fine-tuning)하여 도메인 적응을 수행한 후, 해당 모델에 PTQ를 적용하였다. 이는 단순 PTQ 대비 정확도 손실을 완화하기 위해 활용된 방식이다. 제안한 PTQ 방식을 대상으로 모델 경량화(크기·메모리), 추론 시간, 은닉·복원 품질(SSIM, PSNR), 보안성(StegExpose ROC-AUC) 지표를 기준으로 비교평가를 수행하였다.



[그림 2] 제안한 PTQ방식

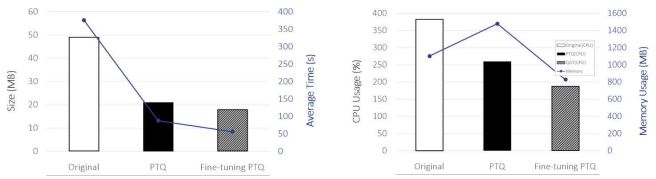
V. 실험 환경 및 성능 비교

실험에 사용된 기기의 사양과 소프트웨어의 버전은 표 1과 같다.

<표 1> 실험 환경 내용

구분	항목	운영체제
Hardware	Device	Rasberry pi 5 Model B Rev1.0
	CPU	ARM Cortex-A76 Quad-core
	RAM	8GB LPDDR4X
	GPU	Broadcom VideoCoreⅦ
Operating System	OS	Debian GNU/Linux 12
Software	Python	3.11.2
	PyTorch	2.7.4
	TorchVision	0.22.1

본 연구에서는 양자화를 통해 엣지 컴퓨팅 환경에서의 적합성을 검증하기 위하여 모델 경량화, 추론 시간, 은닉/복원 품질·보안성을 중점적으로 평가하였다. 품질 평가는 PSNR과 SSIM을 사용하여 시각적 유사성을 측정하였고, 보안 평가는 AUC 지표를 통해 은닉 이미지의 탐지 난이도를 분석하였다. 제안한 PTQ방식을 적용 결과, 모델 크기와 추론 시간이 각각 57.3%, 85.2% 감소하여 경량화와 연산 효율 향상이 확인되었으며 CPU 점유율, memory 사용량 또한 각각 51.1%, 24.9% 감소하였다. 이러한 결과는 그림 3과 같다.



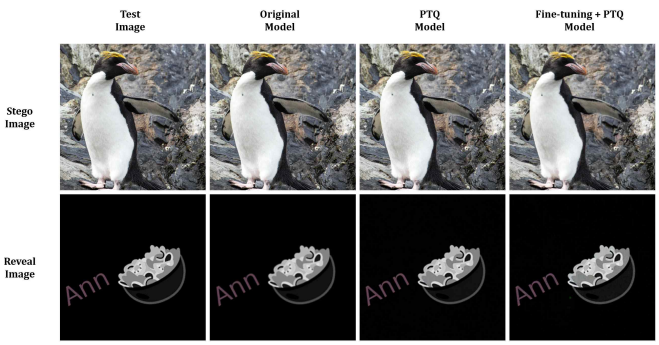
[그림 3] 모델 성능 비교

품질/보안은 은닉/복원 가시성(SSIM, PSNR)과 보안성(AUC)을 FP32와 대비 비교하였다. 가시성은 96% 내로 유지되었고, 결과는 표 2와 같다.

<표 2> PSNR, SSIM 비교

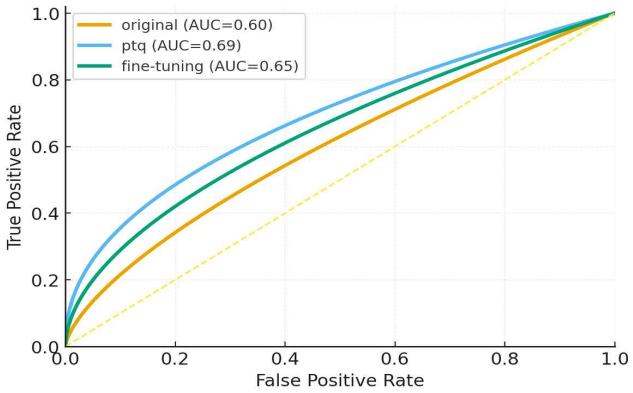
Model	Cover PSNR(dB)	Cover SSIM	Secret PSNR(dB)	Secret SSIM
Original	**45.08**	**0.997**	**46.20**	**0.981**
PTQ	41.92	0.987	44.86	0.975
Fine-tuning PTQ	42.65	0.991	45.37	0.980

세 모델 모두의 스테고(stego) 이미지의 왜곡은 미미하며, 복원은 양호하게 진행되었다. 결과는 그림4와 같다.



[그림 4] 각 모델의 이미지 품질 비교 결과

AUC는 공개된 스테가 분석기를 사용해 산출했으며, FP32 대비 PTQ는 -15%, Fine-tuning+PTQ는 -9% 수준의 감소를 보였다. 공개된 스테가 분석기는 [StegExpose/https://github.com/b3dk7/StegExpose.git]를 활용하여 AUC를 측정하였다. 결과는 그림 5와 같다.



[그림 5] 각 모델의 AUC 비교

실험 결과, Fine-tuning+PTQ는 단순 PTQ 대비 우수한 성능을 보였다. PSNR과 SSIM 지표에서 Fine-tuning PTQ는 원본 대비 품질 저하가 거의 발생하지 않았으며, PTQ보다 은닉·복원 품질이 안정적으로 유지되었다. AUC 값에서도 Fine-tuning PTQ가 더 낮은 탐지율을 보여 스테고 탐지에 대한 보안성이 향상되었음을 확인하였다.

Ⅵ. 결론

본 연구는 고성능 이미지 은닉 모델인 HiNet을 Raspberry Pi 5와 같은 자원 제약적인 엣지 컴퓨팅 환경에서 실용적으로 운용하기 위한 양자화 기반 경량화 방안을 제시하고 그 효과를 검증하였다. GPU 서버 환경을 전제로 설계된 기존 딥러닝 모델의 한계를 극복하고자 새로운 PTQ 기법을 적용하여 모델 경량화의 최적화를 수행하였다. 특히 PSNR, SSIM과 같은 복원 품질 지표와 AUC 기반 스테고 탐지 저항성에서 모두 우수한 성능을 보여, 엣지 컴퓨팅 환경에서도 실용적 활용 가능성을 높였다. CPU 점유율, 추론 시간, 메모리 사용량이라는 세 가지 핵심 자원에서 모두 균형 잡힌 최적화 효과를 확인하였다. 결론적으로 본 연구는 HiNet 모델의 엣지 컴퓨팅 환경에서의 구동 가능성을 입증했다. 추후 자원 사용의 균형을 맞추는 최적화 연구가 더해진다면 엣지 컴퓨팅 기반의 실시간 디지털 저작권 보호 기술 상용화에 한 걸음 더 다가갈 수 있을 것으로 기대된다.

참 고 문 헌

[1] J. Jing, X. Deng, M. Xu, J. Wang, and Z. Guan, "HiNet: Deep Image Hiding by Invertible Network," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 4733-4742.

[2] B. M. Issac, S. N. Kumar, S. Zafar, K. A. Shakil, and M. A. Wani, "Deep learning steganography for big data security using squeeze-and-excitation with inception architectures," Scientific Reports, vol. 15, no. 31193, 2025.

[3] J. Hasan, "Optimizing Large Language Models through Quantization: A Comparative Analysis of PTQ and QAT Techniques," *arXiv preprint arXiv:2411.06084*, 2024.

Minimal-Step Diffusion 기반 초경량 TinyUNet 의 해상도 복원 성능 분석

황병오[1]*, 이호성[1], 김다산[1], 강명주[2]

[1]서울대학교 공과대학 협동과정 인공지능전공, [2]서울대학교 자연과학대학 수리과학부

*brian2568@snu.ac.kr

Performance Analysis of Image Super-Resolution Using Minimal-Step Diffusion-Based Lightweight TinyUNet

Byeongoh Hwang [1]*, Hosung Lee[1], Dasan Kim[1] and Myungjoo Kang[2]

[1]Interdisciplinary Program in Artificial Intelligence, Seoul National University,

[2]Department of Mathematical Sciences, Seoul National University

요 약

본 논문은 최근 이미지 초해상도 분야에서 높은 성능을 보이고 있는 Diffusion 모델의 연산속도 문제를 해결하기 위해, CIFAR-10 데이터셋에서 초경량 UNet 구조와 극소 스텝(3-5 스텝)을 적용한 Diffusion 모델의 성능을 평가하였다. 제안한 초경량 Diffusion 모델은 기존 Diffusion 기반 모델과의 비교 실험에서 모델 크기와 추론 속도에서 우수한 성능을 보이며, 초소형 모델로도 충분한 품질의 초해상도를 구현할 수 있음을 실험적으로 입증하였다.

Ⅰ. 서 론

이미지 초해상도 기술은 낮은 해상도의 이미지를 고해상도로 복원하는 기술로서, CNN, GAN, Diffusion 등 다양한 모델이 사용되어왔다[1]. 특히 Diffusion 모델은 최근 이미지 생성 분야에서 뛰어난 성능을 보였으나, 높은 연산 비용과 느린 추론 속도가 단점으로 지적된다[2]. 본 연구는 Diffusion 모델의 연산 속도를 해결하고자 극소 스텝과 경량 모델 구조를 결합한 접근을 제안하며, CIFAR-10 데이터셋을 기반으로 성능을 정량적으로 분석하였다. 본 논문의 주요 독창성은 기존의 복잡한 구조나 다수 스텝 기반 Diffusion 초해상도 방식과 달리, 극소 스텝과 초경량 UNet 구조를 결합함으로써 모델의 경량화와 고속 추론을 동시에 달성한 점에 있다.

Ⅱ. 본론

본 논문에서는 극소 스텝(3~5 스텝) 기반의 Diffusion 모델과 초경량 UNet 구조를 결합하여 CIFAR-10 데이터셋의 이미지 초해상도 문제에 적용하였다. 제안한 TinyUNet 은 전체 구조가 입력 이미지에 대해 두 개의 합성곱 층을 차례로 적용하는 인코더 부분과, 두 개의 전치 합성곱 층을 활용하는 디코더 부분으로 이루어진다. 인코더 블록은 저해상도 입력 이미지로부터 점진적으로 추상적이고 압축된 표현을 추출하는 역할을 하며, 디코더 블록은 이를 다시 고해상도 이미지로 복원하는 기능을 수행한다. 각 합성곱 층과 전치 합성곱 층 뒤에는 활성화 함수를 적용하여 비선형성을 도입하였다. 활성화 함수의 경우 마지막 단계에서만 sigmoid 를 활용하였고 이전 단계에서는 모두 ReLU 를 활용하였다. 이러한 간소화된 구조는 전체 파라미터 수를 크게 줄이면서도, CIFAR-10 과 같은 저해상도 데이터셋에 대해 빠른 학습 및 추론이 가능하도록 설계되었다.

실험에서는 CIFAR-10 의 원본 이미지를 32×32 에서 16×16 으로 downsampling 한 후, 이를 다시 32×32 로 복원하는 초해상도 태스크를 수행하였다. 모델의 성능 평가는 PSNR, SSIM 과 같은 화질 지표와, 이미지당 추론 속도(ms), 그리고 전체 파라미터 수(MB 단위)로 진행하였다. 실험 결과, 제안한 TinyUNet 기반 극소 스텝 Diffusion 모델은 기존의 대표적인 경량화 초해상도 모델과 비교해 유사한 수준의 화질(PSNR, SSIM)을 유지하면서도, 모델 크기와 추론 속도 면에서 뚜렷한 이점을 보였다. 본 논문의 실험 결과는, 복잡한 구조나 다수의 타임스텝 없이도 효율적인 초해상도 복원이 가능함을 시사한다.

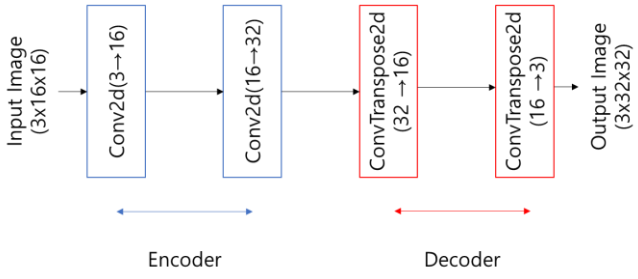


그림 1. TinyUNet 구조의 전체 흐름도

Ⅲ. 결론

본 논문에서는 CIFAR-10 데이터셋을 대상으로, PSNR 과 SSIM 등 화질 지표, 추론 속도, 모델 파라미터 수를 기준으로 초경량 TinyUNet 기반 극소 스텝 Diffusion 모델의 성능을 평가하였다. 표 1 에서 확인할 수 있듯이, 제안한 모델은 SinSR[3]에 비해 PSNR(28.5 vs 27.2)와 SSIM(0.841 vs 0.832)에서 소폭 높은 화질을 기록하면서도, 추론 속도(6.8ms)는 SinSR(12.4ms)의 절반 수준으로 매우 빠른 것이 특징이다. 또한 ResShift[4], ResDiff[5]와 같이 더 많은 타임스텝과 큰 모델을 사용하는 기존 방법들과 비교했을 때, 제안한 모델은 훨씬 적은 파라미터(0.15M)와 타임스텝(5) 만으로도 경쟁력 있는 화질과 가장 빠른 추론 속도를 달성하였다. 이처럼 TinyUNet 기반 극소 스텝 Diffusion 모델은 단순한 구조와 적은 연산량만으로도 기존 경량화 모델에 비해 실용적 수준의 초해상도 복원이 가능함을 실험적으로 확인하였다. 이처럼 TinyUNet 기반 극소 스텝 Diffusion 모델은 단순한 구조와 적은 연산량만으로도 기존 경량화 모델에 비해 실용적 수준의 초해상도 복원이 가능함을 실험적으로 확인하였다.

모델	Step	Param(M)	추론(ms)	PSNR	SSIM
SinSR	**1**	2.1	12.4	**27.2**	**0.83**
ResShift	15	1.7	20.2	28.9	0.85
ResDiff	50	3.4	45.5	29.8	0.86
TinyUNet	5	**0.15**	**6.8**	28.5	0.84

표 1. 실험 결과

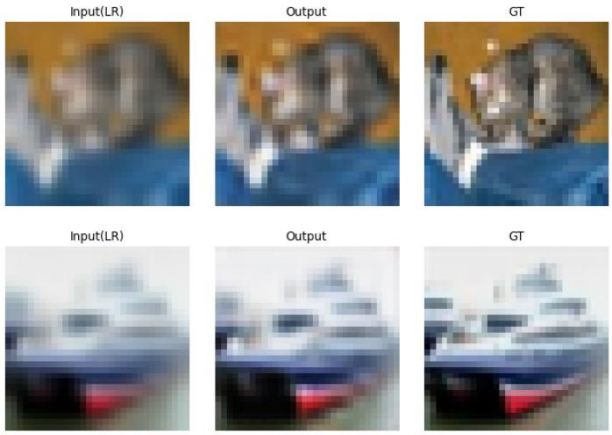


그림 2. 입력(좌), 제안한 모델의 출력(중), 원본(우) 예시

위 그림은 CIFAR-10 테스트 이미지에 대해, 저해상도 입력(16×16, 좌측), TinyUNet 기반 모델이 복원한 초해상도 결과(32×32, 가운데), 원본 고해상도(우측)를 비교한 것이다. 제안 모델의 출력은 입력에 비해 객체의 형태와 경계가 뚜렷하게 복원되었으며, 원본과 유사한 수준의 시각적 품질을 확인할 수 있다.

제안 모델이 SinSR[4], ResShift[5], ResDiff[6] 대비 적은 스텝과 파라미터에도 유사 화질을 유지할 수 있었던 주요 이유는 다음 세 가지로 요약할 수 있다. 첫째, CIFAR-10 과 같은 저해상도 데이터의 분포가 단순하여 극소 스텝에서도 기본적인 노이즈 제거와 형태 복원이 가능하다[1]. 둘째, TinyUNet 의 적절한 용량 배분과 업샘플링 기반 residual 복원 전략을 통해 전역 구조 부담을 줄이고 주로 고주파 보정에 집중함으로써 0.15M 파라미터만으로도 핵심 정보를 충분히 복원할 수 있었다[5][6]. 셋째, DDIM 기반 극소 스텝 샘플링 및 선형·코사인 스케줄 비교를 통한 노이즈 스케줄 최적화와, 동일 환경에서의 세밀한 하이퍼파라미터 조정으로 적은 스텝에서도 효율적 복원을 이끌어냈다[3]. 이러한 분석은 “극소 스텝 Diffusion + TinyUNet” 조합이 저연산 환경에서도 실용적 SR 솔루션이 될 수 있음을 뒷받침하며, 더 큰 도메인으로 확장할 때의 가이드라인으로 활용될 수 있다.

또한 일관된 벤치마크 환경에서 SinSR, ResShift, ResDiff 등 여러 경량 SR 기법을 동일 조건으로 비교·분석함으로써, 후속 연구자들이 “극소 스텝 Diffusion 성능-속도 trade-off”를 재현하고 확장할 수 있는 기준점을 제공했다. 이와 함께 구현이 간단한 TinyUNet 기반 파이프라인을 제시하여, 연구자나 학생들이 Diffusion SR 개념을 빠르게 프로토타이핑하고 실험을 진행할 수 있도록 돕는다. 이러한 통합적 비교와 교육적 프로토타입 제공은 연구생태계의 효율성과 확장 가능성을 높이는 데 기여한다.

ACKNOWLEDGMENT

This work was supported by MSIT/IITP[2021-001343-004, Artificial Intelligence Graduate School Program (Seoul National University)].

참 고 문 헌

[1] Saharia, C., et al. (2023). Image super-resolution via iterative refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1), 475-488.

[2] Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33, 6840-6851.

[3] Kong, L., et al. (2023). SinSR: Diffusion-Based Image Super-Resolution in a Single Step. arXiv preprint arXiv:2304.10556.

[4] Li, Y., Yue, X., Luo, Y., Yu, L., & Li, P. (2023). Residual Shifting for Efficient Diffusion-Based Image Super-Resolution. Advances in Neural Information Processing Systems (*NeurIPS*). arXiv preprint arXiv:2307.12348.

[5] Shang, W., Li, Y., Zhang, H., Zhu, J., & Wu, Y. (2023). ResDiff: Combining CNN and Diffusion Model for Image Super-Resolution. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(3), 3121-3129. arXiv preprint arXiv:2303.08714.

주사투과전자현미경 이미지 원자 위치의 딥러닝 기반 회귀 분석

이승환, 이상철*
한국기초과학지원연구원
shlee0@kbsi.re.kr, *sclee02@kbsi.re.kr

요 약

원자 단위 이미지의 정량 분석에는 각 원자의 정확한 위치 정보가 필수적이다. 기존 접근은 주사투과전자현미경 (Scanning Transmission Electron Microscope, STEM) 이미지 입력으로부터 가우시안 마스크를 예측하고 center of mass 연산으로 원자의 중심 위치를 계산하지만, 위치 오차(잔차)가 여전히 존재한다. 본 연구는 기존 마스크 기반 중심 위치 추정 파이프라인에, 딥러닝 기반 회귀 절차를 추가하여 잔차를 보정하는 방법을 제안한다. 실험 결과, TEMImageNet에서 기존 center of mass 기반 방법 대비 위치 잔차의 평균과 표준 편차를 각각 39.2%, 38.6% 감소시켜 회귀 모듈의 유효성을 보였다.

Ⅰ. 서 론

재료 설계의 토대인 물성 해석과 사전 예측(시뮬레이션)을 위해서는 개별 원자 위치의 정밀 예측이 필수적이다. 이를 위해 전자현미경으로 관측한 원자 배열에서 위치 불확실성을 sub-Å 수준으로 낮춰야 한다. 기존 연구로, Lin[1]은 가우시안 마스크를 예측하도록 U-Net[2]을 학습하고, 마스크 임계화와 질량중심(*center of mass*, COM) 연산을 결합해 2D Gaussian fit보다 높은 정밀도로 원자 위치를 추정하였다. 하지만 그럼에도 잔차가 남아 있어, 이를 보정 할 필요성이 있다. 본 논문에서는 딥러닝 기반 회귀 모듈을 활용하여 COM 기반 결과의 잔차를 추가 보정하는 방법을 제안한다.

Ⅱ. 본론

2.1 TEMImageNet

TEMImageNet[1]은 STEM 이미지를 대상으로 분할(segmentation), 위치 결정(localization), 노이즈 제거(denoising), 초해상화(super-resolution) 등 다양한 과제를 다루도록 설계된 시뮬레이션 데이터셋이다. Lin [1]은 TEMImageNet으로 딥러닝 모델을 학습하여, 특히 원자 검출 및 위치 결정에서 우수한 성능을 보임으로써, 데이터셋의 유효성을 입증하였다.

2.2 Atom localization with center of mass

Lin[1]은 TEMImageNet으로 학습한 U-Net[2]을 이용해 원자 위치를 추정하였다. 구체적으로, U-Net이 입력 STEM 이미지로부터 가우시안 마스크를 예측하고, 이를 Otsu's method[3]로 임계화한 뒤 연결 성분별 COM을 계산하여 각 원자의 위치를 도출하였다. 이 방식은 간결하고 성능이 우수하나, 예측 위치와 실제 위치 사이의 잔차가 여전히 존재한다.

2.3 Position refinement with Deep Regression

본 연구는 U-Net에 회귀 모듈을 추가하여 각 픽셀 위치에서 오프셋을 예측한다. 학습 단계에서는 특정 픽셀 p에 가장 가까운 원자 위치를 x라 두고, 정답 오프셋을 $x - round(x)$로 정의하여 식 (1)을 최소화하도록 네트워크를 지도학습한다. 학습의 안전성을 위해 각 원자 중심에서 가장 가까운 픽셀을 p로 정한다. 추론 단계에서는 2.2절에서 COM으로 얻은 초기 추정 위치에 예측 오프셋을 더해 최종 위치를 결정한다.

$$Loss = \| \| offset_p - (x - round(x)) \| \|_2^2 \quad \text{(식 1.)}$$

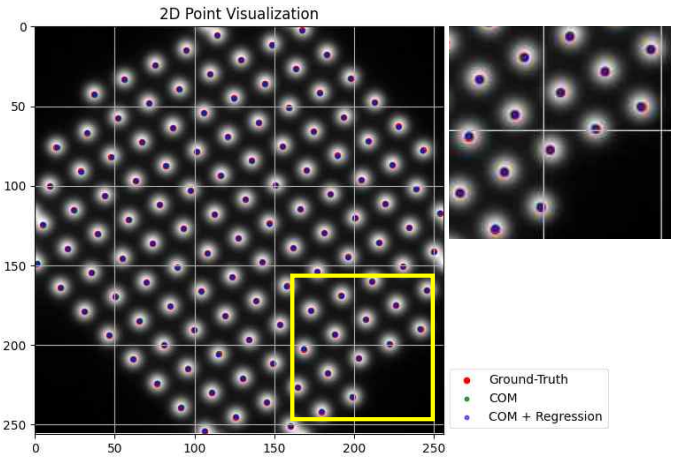


그림 1. Atom localization 비교 결과 시각화 예시.

Ⅲ. 실험 및 결과

TEMImageNet으로 U-Net 기반 네트워크를 학습한다. 마스크 예측 모듈은 가우시안 마스크로 학습하며[1], 회귀 모듈은 식 (1)을 통해 학습한다. 추론 단계에서 먼저 COM 기반 방법으로 초기 좌표를 얻은 뒤, 회귀 모듈로 이를 보정 한다. 최종적으로, 두 결과에 대해 원자의 예측 위치와 정답 위치 간 유클리디언 거리를 측정하여 정확도를 비교 평가한다. 그림 1은 COM 기반 방법(초록색)과 회귀 보정 방법(파랑색)의 결과를 시각화한 것으로, 회귀 보정 후 예측 중심이 정답 위치(빨강색) 쪽으로 이동하며 잔차가 감소함을 확인할 수 있다.

표 1. 예측-정답 원자 위치 간 Euclidean distance의 평균과 표준 편차.

Label	Method	평균↓	표준편차↓
noBackgroundnoNoise	COM	0.4665	0.3175
	COM+Regression	**0.2833**	**0.1947**
image	COM	0.9524	0.4231
	COM+Regression	**0.8943**	**0.3691**

표 1은 TEMImageNet의 두 가지 label에서의 성능 통계(평균·표준 편차)를 보여준다. 각 label은 이미지 13,607장으로 이루어져 있으며, 특히 'image' label은 scan 및 Poisson 노이즈가 합성되어 있어 보다 현실적인

조건의 이미지를 제공한다. 보정 이후 잔차의 평균이 낮아져 정확도가 향상되었고, 표준 편차 또한 감소하여 각 원자 위치의 잔차 변동이 줄어 결과의 일관성이 높아졌음을 알 수 있다. 특히 'noBackgroundnoNoise' label의 결과는 평균과 표준 편차가 각각 39.2%, 38.6% 감소함을 보였다. 노이즈가 포함된 'image' label의 결과는 낮은 정밀도를 보이지만, 회귀 보정을 통해 잔차가 다소 줄어드는 것을 확인하였다. 요약하면, 기존 COM 파이프라인에 회귀 모듈을 추가해 보다 우수한 성능을 달성할 수 있음이 확인되었다.

Ⅳ. 결론

본 논문은 STEM 이미지에서의 원자 위치 추정을 위해 딥러닝 기반 회귀 모듈을 학습해 보정하는 방법을 제안했다. 본 방법은 (i) 기존 파이프라인의 COM 절차를 유지한 채 회귀 모듈을 결합하고, (ii) 단순한 오프셋 라벨링으로 구현·학습이 명료하며, (iii) 잔차의 평균·표준 편차를 낮춰 후속 물성 해석의 신뢰도를 높일 수 있다는 점에서 가치가 있다. 요약하면, 제안한 회귀 보정은 COM 기반 파이프라인의 간단하면서도 효과적인 업그레이드 수단으로, 원자 좌표의 정량적 획득을 한층 신뢰할 수 있게 한다.

ACKNOWLEDGMENT

This research was supported by the basic science research program (C524200) funded by the Korea Basic Science Institute.

참 고 문 헌

[1] R. Lin, R. Zhang, C. Wang, X.-Q. Yang, and H. L. Xin, "TEMImageNet training library and AtomSegNet deep-learning models for high-precision atom segmentation, localization, denoising, and deblurring of atomic-resolution images," Scientific Reports, vol. 11, no. 1, p. 5386, 2021.

[2] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015: 18th International Conference, Munich, Germany, Oct. 5 - 9, 2015, Proceedings, Part III, vol. 9351, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham, Switzerland: Springer, 2015, pp. 234 - 241.

[3] N. Otsu, "A threshold selection method from gray-level histograms," IEEE Transactions on Systems, Man, and Cybernetics, vol. 9, no. 1, pp. 62 - 66, 1979.

투과전자현미경 원자분해능 이미지의 딥러닝 기반 정량 분석을 위한 AI 학습 데이터 생성 시뮬레이터 설계

이태영, 이상철*

한국기초과학지원연구원

lty87@kbsi.re.kr, *sclee02@kbsi.re.kr

요 약

딥러닝 기반 투과전자현미경(Transmission electron microscopy, TEM) 이미지 분석 연구가 급속히 발전하고 있으나, TEM 전문가와 AI 연구자 간의 도메인 지식 격차로 인해 실용적 적용에 제약이 존재한다. 본 연구는 이러한 융합 연구의 현황을 분석하고, 딥러닝 모델 학습에 필요한 대규모 TEM 이미지 데이터셋을 효율적으로 생성하기 위한 범용 시뮬레이터를 제안한다. 제안하는 시스템은 Python의 abTEM(Absorptive potential multislice TEM) 라이브러리를 계산 백엔드로 활용하여 물리적 정확성을 확보하고, C# WPF(Windows Presentation Foundation)로 구현한 직관적인 사용자 인터페이스를 통해 TEM 전문가가 프로그래밍 지식 없이도 고품질 시뮬레이션 데이터를 생성할 수 있도록 설계되었다. 시스템은 원자 vacancy 검출, 원자 종류 분류, 결정 구조 분석, 격자 변형 측정 등 다양한 TEM 분석 문제에 활용 가능하며, 구조 결함의 비율과 분포를 정밀하게 제어할 수 있다. 모든 원자의 위치와 클래스 정보를 포함한 JSON(JavaScript object notation)형식 메타데이터의 자동 생성을 통해 AI 학습 데이터의 즉시 활용이 가능하다. 이를 통해 TEM 도메인 지식과 AI 기술 간의 격차를 효과적으로 해소하고, 향후 4D-STEM(Four-Dimensional Scanning Transmission Electron Microscopy, 4D-STEM) 시뮬레이션으로의 확장 기반을 제공한다.

Ⅰ. 서 론

최근 딥러닝 기술의 발전으로 TEM 이미지 분석 분야에서 AI 활용이 급속히 확산되고 있다. Ziatdinov et al.[2]의 선구적 연구 이후 원자 수준 결함 검출[3], 구조 분석[4], 나노입자 추적[5] 등 다양한 응용 분야에서 딥러닝 모델의 효과가 입증되고 있으나, 실제 연구 현장에서의 도입은 예상보다 제한적이다.

TEM-AI 융합 연구의 주요 장벽은 도메인 전문가와 AI 연구자 간 지식 격차와 대규모 라벨링 데이터셋 확보의 어려움이다. TEM 전문가는 AI 모델링에, AI 연구자는 실험 데이터의 물리적 특성 이해에 각각 한계를 보이며[6], 실험 데이터 수집은 높은 비용과 시간이 소요된다. 시뮬레이션 기반 데이터 생성이 대안으로 제시되었으나[7, 8], 기존 도구들은 AI 워크플로와의 연계성 부족과 높은 전문성 요구로 접근성이 낮다.

본 연구에서는 이러한 도메인 간 격차를 해소하기 위한 사용자 친화적 시뮬레이터를 제안한다. 물리적으로 정확한 계산 엔진과 직관적 사용자 인터페이스를 결합하여, TEM 전문가가 복잡한 프로그래밍 없이도 AI 학습에 최적화된 고품질 훈련 데이터를 효율적으로 생성할 수 있는 시스템을 제안한다.

Ⅱ. 본론

2.1. 기존 연구의 기술적 성과

TEM-AI 융합 연구는 지난 5년간 괄목할만한 기술적 진전을 보여왔다. Ziatdinov et al.[2]이 제시한 약한 감독학습 방법론은 제한된 라벨 데이터로도 다양한 원자 결함을 식별할 수 있음을 입증했으며, 이는 전통적인 완전 감독학습의 데이터 의존성 문제를 크게 완화했다. Lee et al.[3]의 ResUNet 기반 접근법은 MoS_2 에서 황 공공 검출에서 98%라는 높은 정확도를 달성하여 실용적 적용 가능성을 보여주었다.

동적 과정 분석에서도 의미 있는 성과가 나타났다. Maksov et al.[4]은 WS_2 에서 전자빔 조사에 따른 실시간 상변화를 추적하는 데 성공했으며, Jamali et al.[5]은 액상 TEM 환경에서 나노입자 확산의 미세한 동역학을 정량적으로 분석할 수 있음을 보였다. 이러한 연구들은 AI가 단순한 패턴 인식을 넘어 물리적으로 의미 있는 현상 해석이 가능함을 시사한다.

2.2. 실무 도입 장벽과 해결 요구사항

실제 연구 현장에서 TEM 전문가들은 AI 학습을 위한 데이터셋 생성에 대한 명확한 가이드라인 부족으로 혼란을 겪고 있으며, AI 연구자들은 TEM 이미지의 해석에 어려움을 겪는다.

이러한 문제 해결을 위해서는 물리적 정확성과 AI 학습 최적화를 동시에 만족하는 시뮬레이션 환경과 프로그래밍 없이 사용 가능한 직관적 인터페이스가 필요하다. 또한 실험 조건과 노이즈 특성을 정밀 모사하고, 원자 위치, 결함 유형 등의 메타데이터를 표준화된 형식으로 자동 생성하는 통합 시스템이 요구된다.

2.3. 제안하는 시뮬레이터

앞서 분석한 도메인 간 격차 문제를 해결하기 위해, 물리적 정확성과 사용자 접근성을 동시에 확보하는 하이브리드 아키텍처를 제안한다. 제안하는 시스템은 Python 기반의 계산 백엔드와 C# WPF 기반의 사용자 인터페이스 프론트엔드로 구성되어, TEM 전문가와 AI 연구자 모두의 요구사항을 충족하도록 설계하고자 한다(그림 1).

Python 백엔드는 검증된 abTEM[1] 라이브러리를 활용하여 multi-slice 방법 기반의 정확한 전자빔-물질 상호작용 계산을 수행한다. Multi-slice 시뮬레이션은 양자역학적 전자-시료 상호작용을 정확히 모델링하며, 전자현미경 분야에서 널리 인정받고 있는 전산 모사 기법이다[9]. 이를 통해 기존 TEM 시뮬레이션 도구들과 동등한 물리적 정확성을 보장하면서도, AI 워크플로에 최적화된 데이터 생성이 가능하다. C# WPF 프론트엔드는 MVVM(Model-View-View-Model) 아키텍처를 적용하여 결정 구조의 직관적인 3차원 시각화, 간편한 시뮬레이션 매개변수 설정, 실시간 진행

상황 모니터링 기능을 제공한다.

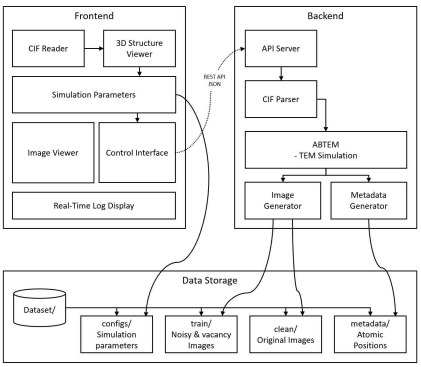


그림 1. 시스템 아키텍처 다이어그램

두 컴포넌트 간 JSON 기반 API 프로토콜을 통해 TEM 실험 조건을 AI 학습 매개변수로 자동 변환한다. 이러한 설계를 통해 TEM 전문가가 익숙한 용어로 시뮬레이션을 설정하면, 백그라운드에서 AI 훈련에 최적화된 데이터가 생성되도록 한다.

CIF 파일로부터 원자 구조를 로드하고 슈퍼셀 크기를 설정하여 다양한 관찰 영역을 구현한다. Vacancy는 원자 인덱스 지정이나 농도 비율 설정으로 간편하게 제어할 수 있으며, 다양한 defocus 값을 적용하여 실제 TEM 조건을 재현한다(그림 2). 이를 통해 복잡한 프로그래밍 없이도 다양한 AI 모델 훈련용 데이터를 생성할 수 있다.

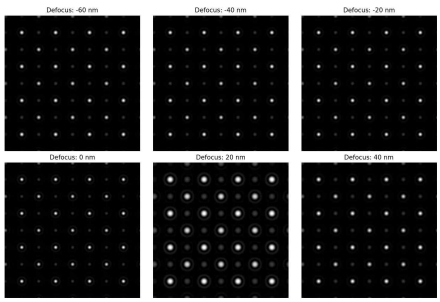


그림 2. TEM defocus 시뮬레이션 이미지

실제 TEM 조건 재현을 위해 열진동 효과, CTF 왜곡, 다중 슬라이스 포텐셜, Poisson 기반 shot noise, readout noise 등 포괄적 노이즈 모델링을 적용한다(그림 3). GPU 가속과 청크 단위 처리로 대규모 데이터셋 생성 시에도 안정적 성능을 유지한다.

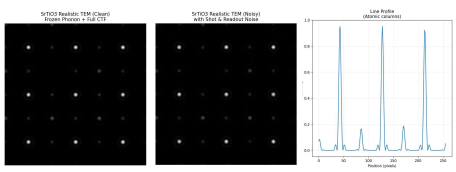


그림 3. Vacancy 및 CTF와 Noise가 추가된 이미지와 원자 위치정보 프로파일

생성 결과는 노이즈 이미지, 클린 이미지, 원자 위치 정보, 결함 라벨, 설정값이 분리된 구조로 저장되어 다양한 딥러닝 프레임워크에 활용할 수 있다. 모든 원자의 정확한 위치와 클래스 정보를 JSON 메타데이터로 자동 생성하며, 기존 연구의 주요 병목이었던 ground truth 생성 과정을 완전 자동화한다.

Ⅲ. 결론

본 연구는 TEM-AI 융합 연구의 현황을 분석하고, 도메인 간 지식 격차 해소를 위한 통합 시뮬레이터를 제안하였다. 제안하는 하이브리드 아키텍처 기반 시뮬레이터는 물리적 정확성과 사용자 접근성을 동시에 확보하여 이러한 문제들을 해결할 수 있을 것으로 기대된다. 또한, TEM 전문가가 복잡한 프로그래밍 없이도 AI 학습에 최적화된 대규모 훈련 데이터를 생성할 수 있는 환경을 제공하고자 한다. 자동화된 메타데이터 생성과 표준화된 출력 형식을 통해 기존 연구의 주요 병목이었던 ground truth 데이터 확보 과정을 효율화할 수 있다.

본 프레임워크는 vacancy 검출을 비롯하여 원자 분류, 구조 분석, 변형 측정 등 다양한 TEM 분석 문제에 적용 가능하며, 향후 4D-STEM 시뮬레이션으로 확장하여 더욱 포괄적인 전자현미경 AI 분석 도구로 발전시킬 계획이며, 이를 통해 TEM-AI 융합 연구에서 학습 데이터 생성의 접근성과 효율성을 향상시켜 더 많은 연구자들의 AI 기반 TEM 분석 기술 활용을 촉진할 것으로 기대된다.

ACKNOWLEDGMENT

This research was supported by the basic science research program (C524200) funded by the Korea Basic Science Institute.

참 고 문 헌

[1] J. Madsen et al., "The ABTEM code: transmission electron microscopy from first principles," Open Research Europe, vol. 1, p. 24, 2021.

[2] M. Ziatdinov et al., "Deep learning of atomically resolved scanning transmission electron microscopy images: Chemical identification and tracking local transformations," ACS Nano, vol. 11, pp. 12742-12752, 2017.

[3] K. Lee et al., "STEM image analysis based on deep learning: Identification of vacancy defects and polymorphs of MoS_2 ," Nano Lett., vol. 22, pp. 4677-4685, 2022.

[4] A. Maksov et al., "Deep learning analysis of defect and phase evolution during electron beam-induced transformations in WS_2 ," npj Comput. Mater., vol. 5, p. 12, 2019.

[5] V. Jamali et al., "Anomalous nanoparticle surface diffusion in LCTEM revealed by deep learning-assisted analysis," Proc. Natl. Acad. Sci., vol. 118, p. e2017616118, 2021.

[6] M. Botifoll et al., "Machine learning in electron microscopy for advanced nanocharacterization: Current developments, available tools and future outlook," Nanoscale Horiz., vol. 7, pp. 1427-1477, 2022.

[7] J. Tremblay et al., "Training deep networks with synthetic data: Bridging the reality gap by domain randomization," Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 9967-9977, 2018.

[8] N. M. Schneider et al., "Machine learning to reveal nanoparticle dynamics from liquid-phase TEM videos," ACS Cent. Sci., vol. 6, pp. 1894-1903, 2020.

[9] Cowley, J.M., Moodie, A.F. (1957). The scattering of electrons by atoms and crystals. I. A new theoretical approach. Acta Crystallographica, 10, 609-619.

광 카메라 통신 기반 위치 추정을 위한 딥러닝 객체 탐지 알고리즘 비교: YOLOv5 와 YOLOv8

김주현, 황나영, 정성윤*
영남대학교

lpyeongl00@gmail.com, ghkd3998@yu.ac.kr, *syjung@yu.ac.kr

Comparison of Deep Learning-Based Object Detection Algorithms for Optical Camera Communication-Based Position Estimation: YOLOv5 and YOLOv8

Kim Ju Hyun, Hwang Na Young, Jung Sung Yoon*
Yeungnam Univ.

요 약

광 카메라 통신(OCC)은 LED 신호를 카메라로 수신하여 위치 추정 및 차세대 V2X 통신으로 활용할 수 있는 기술이다. 본 연구에서는 OCC 환경에서의 LED 검출을 위해 딥러닝 기반 객체 탐지 모델 YOLOv5 와 YOLOv8 을 학습 및 평가하였다. 성능 비교 결과, YOLOv5 가 일부 지표에서 더 높은 값을 보였으나 YOLOv8 은 학습 및 추론 안정성이 높기 때문에 실시간 처리 속도에서 강점을 보였다. 따라서 OCC 기반 실시간 위치 추정 환경에서는 YOLOv8 이 보다 적합한 모델임을 확인하였다.

Ⅰ. 서 론

광 카메라 통신(OCC, Optical Camera Communication) 은 LED 와 카메라를 송수신 장치로 활용하여 데이터를 주고받는 통신 기술이다. 추가적인 전용 센서 없이, 기존의 LED 와 카메라를 재사용하여 통신 체계를 구현할 수 있다는 점에서 주목받고 있다. 특히 UAV, 차량, 로봇 등에 탑재된 카메라와 가로등, 후미등의 LED 와 같이 차세대 모빌리티, 스마트시티 등 다양한 환경에 응용할 수 있다. 또한, LED 의 깜박임 신호를 기반으로 통신뿐만 아니라 LED 의 위치 추정까지 동시에 지원 가능하여 활용도와 범용성이 높다[1,2]. 이러한 OCC 시스템에서의 핵심은 영상 내 LED 를 빠르고 정확하게 검출하는 것이다. 기존 연구에서는 주로 프레임 차분(Differential Frame), 임계값 기반(Thresholding), Morphology 등과 같은 전통적인 영상처리 기법이 LED 검출에 활용되었다[2]. 이는 단순하고 계산량이 적다는 장점이 있으나, 주변 환경 변화에 취약하고 다중 객체 검출 상황에서 한계가 존재한다. 이러한 한계를 극복하기 위해, 최근 딥러닝 기반 객체 탐지 알고리즘을 접목한 연구가 활발하게 이뤄지고 있다[3]. 특히 YOLO(You Only Look Once) 계열 모델은 단일 신경망 기반의 실시간 객체 탐지 구조로, 속도와 정확도를 동시에 보장한다는 점에서 OCC 환경의 LED 검출 응용에 적합하다[3,4]. YOLOv5 는 PyTorch 기반으로 안정성과 활용성이 높으며, 다양한 크기의 모델을 제공한다[5,6]. 반면 YOLOv8 은 Anchor-free 구조와 향상된 네트워크 설계를 적용하여 작은 객체 탐지 성능을 개선하였으며, 탐지뿐 아니라 분할 및 분류 또한 지원한다[5,7,8]. 본 논문에서는 OCC 환경에서 촬영한 LED 데이터셋을 기반으로 YOLOv5 와 YOLOv8 을 학습 및 평가하여, 두 모델의 Precision, Recall, F1-score, mAP 등을 정량적으로 비교한다. 이를 통해 두 모델 중 실제 응용에 적합한 탐지 모델을 제시하고, 딥러닝 기반 OCC 시스템 설계에 참고할 수 있는 지침을 제공한다.

Ⅱ. 객체 탐지 모델 비교 분석

본 논문에서는 동일한 LED 데이터셋을 활용하여 YOLOv5 와 YOLOv8 을 학습하고 성능을 비교하였다. 그림 1 과 그림 2 는 시뮬레이션 환경에서 LED 를 검출한 결과를 보여준다. 그림 1 은 YOLOv5, 그림 2 는 YOLOv8 을 적용한 사례이며, 두 모델 모두 근거리와 원거리에 위치한 LED 를 정확하게 탐지하였다. 각 탐지 결과에는 신뢰도(confidence score)가 함께 표시되어 있으며, 이를 통해 본 연구에서 사용한 객체 탐지 모델이 다양한 거리 조건에서도 안정적으로 LED 신호를 인식할 수 있음을 확인할 수 있다. 표 1 은 학습 파라미터 표를 나타낸다. 아울러 본 연구에 사용된 데이터셋은 총 1180 장으로, 이 중 약 900 장은 거리 변화 조건을 반영하였고, 약 200 장은 다양한 잡음 환경을 포함하도록 구성하였다. 이러한 데이터셋 설계를 통해 모델이 단순한 정적 환경뿐 아니라 실제 운용 상황에서 발생할 수 있는 거리 변화와 노이즈 요인에 대해서도 학습할 수 있도록 하였다.

그림 1. LED 영상에 대한 YOLOv5 의 검출 결과

그림 2. LED 영상에 대한 YOLOv8 의 검출 결과

표 1. 학습 파라미터

Parameter	YOLOv5	YOLOv8
Train dataset	1180	1180
Validation dataset	100	100
Epoch	200	200
Batch size	4	4

그림 3 의 YOLOv5 Precision- Recall 곡선과 Recall-Confidence 곡선에서는 각각 mAP@0.5=0.981, Recall=0.98 을 기록하여 그림 4 의 YOLOv8 결과(0.943, 0.87)보다 다소 높은 수치를 보였다. 그러나 이러한 차이는 전반적으로 미세한 수준에 불과하며 두 모델 모두 OCC LED 검출에 충분히 높은 성능을 나타냈다. 특히 YOLOv8 은 그림 4 의 학습 손실 곡선에서 값이 부드럽게 감소하며 안정적으로 수렴하는 경향을 보였으며 최신 구조를 기반으로 한 빠른 추론 속도와 경량화 용이성 측면에서 실시간 응용 환경에 강점을 지닌다.

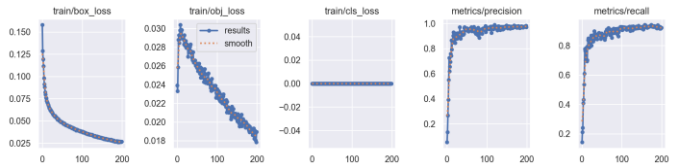


그림 3. YOLOv5 의 학습 결과

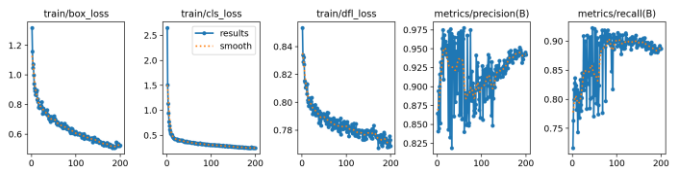


그림 4. YOLOv8 의 학습 결과

결론적으로 YOLOv5 가 정량적 지표에서 근소한 우위를 보였으나 YOLOv8 은 안정성, 효율성, 그리고 실시간성 측면에서 OCC 기반 위치 추정 시스템에 더욱 적합한 모델이라 할 수 있다.

Ⅲ. 결론

본 논문에서는 광 카메라 통신 기반 위치 추정을 위한 LED 객체 탐지 모델로 YOLOv5 와 YOLOv8 을 비교하였다. 실험 결과, YOLOv5 가 Precision, Recall, mAP 지표에서 YOLOv8 보다 약간 더 높은 수치를 기록하였다. 그러나 그 차이는 크지 않았으며 두 모델 모두 높은 수준의 성능을 보였다. 한편 YOLOv8 은 학습 과정에서 손실이 안정적으로 수렴하는 특성을 보였으며 최신 구조를 기반으로 추론 속도가 빠르고 경량화가 용이하다는 장점을 가진다. 이러한 특성은 OCC 기반 위치 추정과 같이 실시간성이 요구되는 응용 환경에서 특히 중요한 요소이다. 따라서 본 연구는 YOLOv5 와 YOLOv8 모두 OCC LED 검출에 효과적임을 확인하였으며 최종적으로는 실시간 적용 가능성 측면에서 YOLOv8 이 보다 적합한 모델이라고 결론지을 수 있다. 향후 OCC 시스템의 최적화 및 실시간 구현에 있어, 딥러닝 기반 LED 탐지 알고리즘 선택에 유용한 지침을 제공할 것으로 기대된다.

ACKNOWLEDGMENT

본 결과물은 2025 년도 교육부의 재원으로 경상북도 지역혁신중심대학지원체계(RISE)-(로컬 맞춤형 R&D)의 지원을 받아 수행된 결과입니다.

참 고 문 헌

[1] Md. Tanvir Hossan, Amirul Islam, Trang Nguyen, Nam Tuan Le, and Yeong Min Jang, "광 카메라 통신 기술," in KICS Information & Communication Magazine - Open Lecture Series, 2017, pp. 35-50.

[2] Singh, Pankaj, et al. "Vehicle Positioning Based on Optical Camera Communication in V2I Environments." Computers, Materials & Continua 72.2 (2022).

[3] Sitanggang, Ones Sanjerico, et al. "Design and implementation of a 2D MIMO OCC system based on deep learning." Sensors 23.17 (2023): 7637.

[4] Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[5] Ultralytics Docs, "YOLOv5 vs. YOLOv8: A Detailed Comparison", (https://docs.ultralytics.com/ko/compare/yolov5-vs-yolov8/).

[6] Khanam, Rahima, and Muhammad Hussain. "What is YOLOv5: A deep look into the internal features of the popular object detector." arXiv preprint arXiv:2407.20892 (2024).

[7] Yaseen, M. "What is YOLOv8: An in-depth exploration of the internal features of the next-generation object detector. arXiv 2024." arXiv preprint arXiv:2408.15857.

[8] Huang, Huadong, et al. "Improved small-object detection using YOLOv8: A comparative study." Appl. Comput. Eng 41.1 (2024): 80-88.

CTC 강제 정렬을 활용한 음성 휴지 구간 탐색

구다연, 박화영, 김홍국*
*광주과학기술원, *(주)오니온에이아이
{dayeonku, hwayoung_park}@gm.gist.ac.kr, *hongkook@gist.ac.kr

Speech Pause Detection Using CTC Forced Alignment

Dayeon Ku, Hwa-Young Park, Hong Kook Kim*
*Gwangju Institute of Science and Technology, *AunionAI Co., Ltd.

ABSTRACT

This paper investigates a method to detect pauses in speech using Connectionist Temporal Classification (CTC) forced alignment, which is commonly employed in speech recognition for aligning speech and text. First, we align the input speech according to a given text, and delimit the word boundaries in the waveform domain. Then the between-word pauses are declared as pause intervals under duration constraint. The experimental results demonstrate that speech pause detection with CTC forced alignment maintains robust performance even in noisy environments, compared to a conventional method that relies on the root mean squared amplitude of speech with a predefined threshold.

Ⅰ. Introduction

One of the typical methods to detect pauses in speech signals is to, first, calculate root mean squared (RMS) amplitude in each frame, and then define an interval as a pause interval when it does not exceed a predefined threshold. However, this method has limitations due to its static threshold, failing to detect pause frames when background noise elevates their RMS amplitude above the predefined threshold. Consequently, such an RMS-based approach is not robust especially in noisy environments.

As an alternative, this paper employs Connectionist Temporal Classification (CTC) forced alignment for speech pause detection in noisy environments. CTC has been popularly used to perform a monotonic alignment between speech and text tokens in speech recognition [1]. Thus, CTC can estimate the start and end of each phoneme and/or word boundaries by finding the optimal paths using forward-backward computation [1]. Therefore, it is expected that CTC-based pause detection should yield more robust performance than RMS-based ones because CTC aligns phonemes and words boundaries based on probabilities, not a predefined threshold.

Ⅱ. Related Work

Energy-based approaches for pause detection typically rely on calculating the RMS energy of each speech frame [2][3]. In these methods, the average energy of the k-th frame whose length is N, E_k, calculated as $E_k = \sqrt{\frac{1}{N}\Sigma x_{k,i}{}^2}$, where $x_{k,i}$ is the i-th normalized speech sample at the k-th frame. A frame is classified as a pause if E_k falls below the predefined threshold. After classifying all the frames as either a pause or a speech frame, an pause interval is declared if the number of consecutive pause frames are greater than D [4]. A set of pause intervals in an utterance is denoted as $\boldsymbol{P}=(s_0, e_0),(s_1, e_1),\cdots,(s_{p-1}, e_{p-1})$ where p is the total number in this pause interval, $\boldsymbol{P}$, and s_i and e_i denote the starting and ending frames in the i-th pause frame of $\boldsymbol{P}$, respectively. Note that $(s_i - e_i) \geq D$ for all i. However, this approach has limitations due to its reliance on a fixed threshold, making it vulnerable to noisy environments where background noise can elevate energy levels during silent periods above the threshold.

Ⅲ. Method

This paper, we propose a noise-adaptive pause detection method using a CTC forced aligner to determine word boundaries. The CTC forced aligner utilizes a multilingual Wav2Vec2 model to determine word boundaries, outputting time-based markers represented by the start and end times for each word in seconds.

First of all, we pad short silent frames at the beginning and end to ensure the audio meets the minimum length required for reliable alignment without affecting its quality. Then, we apply the CTC aligner to speech input sampled at a rate of 22.05 kHz, where the frame-wise processing is carried out with a hop length of 256. Thus, the time-based markers associated with word boundaries are subsequently converted into speech frame indices to identify potential pause intervals. Next, to detect meaningful pauses from within-word or between-word short pause, the intervals that exceed 43 frames ($\cong$ 500ms) are classified as pauses. This is because the duration of a pause corresponding to a comma in text typically ranges from 380 to 670ms [5], and pauses shorter than 500ms are most frequently used by speakers [6]. This procedure finally generates a set of pause intervals

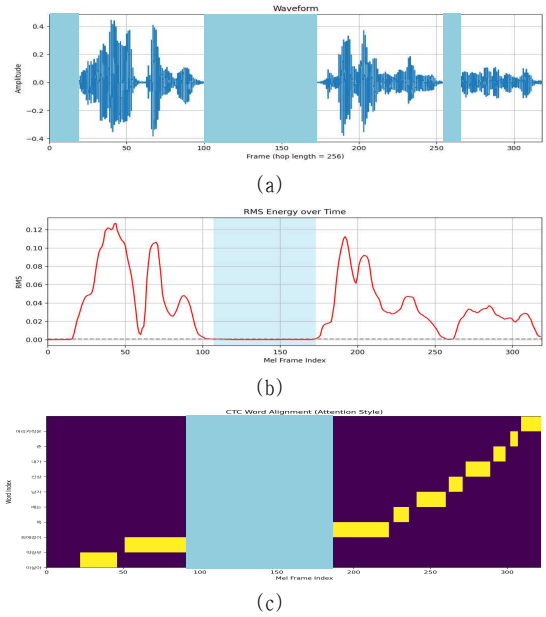


Figure 1: Illustration of pause detection results applied to (a) clean speech: (b) RMS-based and (c) proposed CTC-based detection.

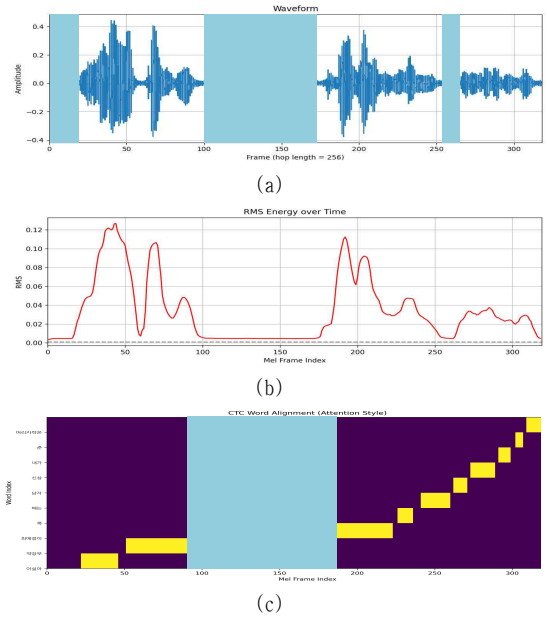


Figure 2: Illustration of pause detection results applied to (a) noisy speech of 20 dB SNR: (b) RMS-based and (c) proposed CTC-based detection.

$\boldsymbol{P}=(s_0,e_0),(s_1,e_1),\cdots,(s_{p-1},e_{p-1})$. Note that we only consider the pause intervals existing between the first and the last word in an utterance.

Fig. 1 compares the performance of the two pause detection methods on a source-separated Korean content audio sample. Fig. 1(a) is the waveform of the clean speech, and the boxes are the ground truth pauses located at $s_0=0$, $e_0=20$, $s_1=96$, $e_1=179$, $s_2=249$ and $e_2=266$. As mentioned earlier, (s_0,e_0) and (s_2,e_2) are placed before and after the main speech, so it should not be considered in the pause detection step. As displayed in Fig. 1, both the RMS-based and CTC-based methods were all able to identify the pause intervals, such as $(s_{rms}=107, e_{rms}=173$, and $s_{ctc}=91, e_{ctc}=185)$, compared with the ground truth. Note that the threshold was set to 0.001 and D=18 in the RMS-based pause detection.

Next, we mixed a noise with the signal-to-noise ratio (SNR) of 20 dB to the clean speech shown in Fig. 1(a). As illustrated in Fig. 2(b), RMS-based method failed to detect pauses because the RMS energy level during silent periods was above the set threshold. Conversely, the CTC-based approach was not significantly impacted by the noise and consistently provided pause detection performance $(s_{ctc}=91, e_{ctc}=185)$, as shown in Fig. 2(b).

Ⅳ. Conclusion

The inability to deliver noise-adaptive pause detection was a key limitation of the RMS-based approach. Based on the experiment performed in this paper, CTC-based approach has outperformed the RMS-base one with the predefined threshold in noisy environments. It was concluded in this paper that the proposed CTC-based pause detection was more robust especially in real-life applications such as dubbing, personal assistants, and real-time speech translation.

ACKNOWLEDGMENT

This work was partly supported by the Technology development Program of MSS [RS-2025-21432982] and the program through the Innopolis Foundation funded by Ministry of Science and ICT [2022-DD-UP-0312].

Reference

[1] Graves, A., Fernández, S., Gomez, F., and Schmidhuber, J. "Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks," in Proceedings of the 23rd International Conference on Machine Learning, pp. 369-376, Pittsburgh, PA, USA, 2006.

[2] Özaydın, S. "Examination of energy based voice activity detection algorithms for noisy speech signals," European Journal of Science and Technology, vol. 17, pp. 442-449, 2019.

[3] Lesenfants, D., Vanthornhout, J., Verschueren, E., and Francart, T. "Cortical auditory responses index the contributions of different RMS-level-dependent segments to speech intelligibility," Journal of Neurophysiology, vol. 126, no. 2, pp. 463-473, 2021.

[4] De Jong, N. H., and Bosker, H. R. "Choosing a threshold for silent pauses to measure second language fluency," Proceedings of DiSS 2013, The 6th Workshop on Disfluency in Spontaneous Speech, pp. 17-20, Stockholm, Sweden, 2013.

[5] Yamashita, K., Liu, S., Sugino, T., O'Connell, D. C., and Kowal, S. "How pause duration influences impressions of English speech: Comparison between native and non-native Speakers," Frontiers in Psychology, vol. 13, pp. 1-15, 2022.

[6] Hoffer, J. Quantifying Speech Pause Durations in Typical English Speakers, Master's Thesis, Brigham Young University, 2023.

이미지 복원을 위한 구조 및 의미 일관성 딥러닝 모델

이동규, 한동석*
경북대학교
jasmindoe@knu.ac.kr, *dshan@knu.ac.kr

Deep Learning Model Ensuring Structural and Semantic Consistency for Image Restoration

Dong Gyu Lee, Dong Seog han*
Kyunpook National Univ.

요 약

이미지 복원은 이미지에서 훼손되거나 누락된 부분을 자연스러우면서 일관되게 복원하는 것을 목표로 한다. 딥러닝 기법 이전의 방법들은 패치 매칭이나 텍스처 합성을 기반으로 하지만 최근에는 PixelHacker와 같은 확산 모델 기반 기법이 구조 및 의미적 문제를 극복하여 뛰어난 성능을 보여준다. 본 논문에서는 이미지 복원의 구조적 의미적 일관성 문제를 극복하며 동시에 세부적인 요소의 복원 성능을 향상시키기 위한 모델을 제안한다. 제안 모델은 다른 딥러닝 모델 기반의 이미지 복원 방식들과 비교하여 FID(Fréchet inception distance)와 LPIPS(learned perceptual image patch similarity)의 성능이 우수하다.

Ⅰ. 서 론

이미지 복원은 오래된 사진 복원, 특정 객체 제거, 영상 편집 등 다양한 분야에서 사용된다. 초기의 텍스처 합성[1]이나 패치 기반[2] 방법은 단순한 영역에서는 그럴듯한 복원 결과를 보여주지만 복잡하며 작은 영역에서는 의미있는 구조를 복원하는 데 한계가 있었다. 이미지 복원 연구는 딥러닝을 통해 크게 발전하였다. 초기의 방식과 다르게 딥러닝 기반의 방식은 더욱 복잡하고 일관성을 가지는 결과를 보여준다. 콘텍스트 인코더(context encoders)[3]는 인코더-디코더 구조와 적대적 손실함수를 활용해 이미지 내 자연스러운 복원을 시도하였다. 이후 RePaint[4]는 확산 모델을 활용해 점진적 복원을 수행했으며 PnP 인페인팅(plug-and-play inpainting)[5]은 외부 조건을 접목하였다. 최근 PixelHacker[6]는 구조적 일관성과 의미적 일관성을 동시에 고려하는 모델을 제안하여 기존 대비 현저히 향상된 결과를 보여주었다. 그러나 해당 모델의 경우 영역 크기에 대한 복원이나 세부적인 부분에 대한 복원 성능이 저하되는 문제가 있다.

이미지 복원에서는 가려진 영역이 어떠한 반복된 구조 또는 색 등의 특징 패턴을 어느 영역까지 이어지는 구조적 일관성과 복원되는 영역이 주변 환경과 어울릴 수 있는 의미적 일관성을 가지는게 중요하다. 이를 위해 다양한 기법들이 사용되지만 여전히 세부적인 부분이나 특정 영역 크기에 대한 한계점이 존재한다. 본 논문은 딥러닝 기반의 확산 기반 방식을 통해 이미지 복원의 구조 및 의미적 일관성을 유지하며 작은 영역의 크기에 대해서도 상세한 복원이 가능한 모델을 제안한다.

Ⅱ. 이미지 복원 모델 구조

이미지 복원의 구조적 일관성 및 의미적 일관성을 유지하기 위해 원본 이미지에 마스크 정보가 적용된 정보를 입력으로 받는다. 이 입력은 확산 기반 모델에서 구조적 정보와 의미적 정보를 이용하여 복원이 진행된다. 구조적 정보와 의미적 정보는 LCG(latent categories guidance)를 통해 전경 임베딩과 배경 임베딩을 통해 얻는다. 전경 임베딩은 이미지 내 목표로하는 객체 또는 영역을 마스크화한 객체 의미 마스크로 학습하고 배경 임베딩은 특정 객체 또는 영역 이외의 배경 장면을 의미론적 마스크로 주변 문맥과 구조적 정보를 학습한다. 해당 과정에서 모델이 특정 영역 크기나 복잡한 경계영역에서 일관성을 확보하기 위해 C2F(coarse-to-fine) 구조를 사용한다. 원본 이미지는 마스크 이미지와 함께 특정한 크기의 해상도로 변경하기 위해 그림 1의 피라미드 샘플링을 통해 다운 샘플링을 진행한다. 변형된 해상도들은 x^{l_M}로 변형되어 입력으로 사용되고 입력 시 상위 스케일은 하위 스케일의 출력을 업샘플링한다.

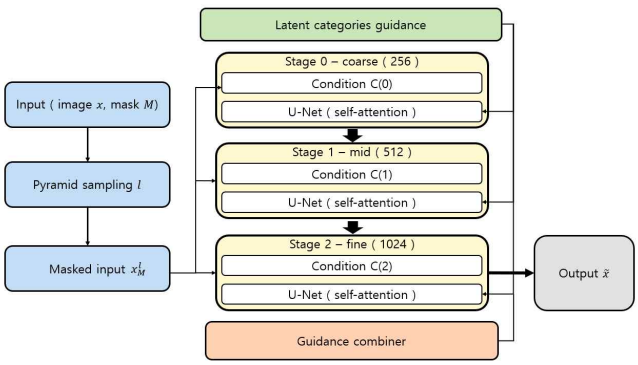


그림 1 제안하는 이미지 복원 모델 구조도

LCG의 전경과 배경의 잠재 임베딩을 미리 정해둔 4가지 타입의 마스크 이미지로 학습하고 그림 1에서 스테이지별 U-Net[8] 형태의 어텐션 경로에 사용하기 위해 특정 입력 형태로 변환한다. C2F의 구조에서 각 스테이지마다 어텐션 구조를 사용한다. 모델의 역확산 시 노이즈 예측은 CFG(class free guidance)[7]를 사용하여 예측을 진행한다.

저해상도에서는 이미지의 전역 구조와 패턴, 색과 같은 부분을 복원하게 학습하고 고해상도에서는 LCG의 잠재 임베딩을 통해 구조적, 의미적 일관성을 확보한다. 동시에 각 단계에서 이미지의 마스크 경계 부분 문제를 위해 경계 보존 손실과 업샘플링 정합 손실을 보정한다. CFG는 저해상도에서는 영향을 크게 주고 단계가 올라갈수록 영향을 줄인다. 이는 초기에 저해상도에서 이미지의 전역적인 구조와 의미를 쉽게 얻도록 유도하고 이후에는 자연스러운 의미론적 형성에 방해가 되지 않게 하기 위함이다.

각 단계에서는 입력이 들어옴과 동시에 LCG의 잠재 임베딩 또한 입력된다. 입력은 이전 단계의 결과값에서 가이던스 컴바이너(guidance combiner)를 통해 일관성을 보정해주고 U-Net 구조의 셀프 어텐션을 통해 복원이 진행된다. 셀프 어텐션은 선형 구조의 모델을 사용하여 기존 어텐션 모듈의 계산량을 감량한 모델을 사용한다. 선형 구조를 보완하기 위해 스위쉬 활성화 함수[10]를 사용하여 선형 함수에 비선형성을 추가해 학습 효율을 증가시켜준다.

손실함수는 각 마스크별 가중 노이즈, 이미지 간 L1 손실, 마스크 경계, 업샘플링 정합 손실로 총 4가지의 손실함수가 사용된다. 손실함수들은 각 스테이지별로 계산되며 가중 노이즈는 가중치를 통해 예측 오차의 학습을 조절한다. L1 손실과 마스크 경계 손실 함수는 복원된 부분의 생성된 결과의 자연스러움와 경계면의 불균형을 방지한다. 업샘플링 정합 손실 함수는 가우시안 블러를 통해 정합하여 일관성을 보정해준다.

Ⅲ. 학습 및 실험결과

모델의 학습을 위한 데이터로는 places2, FFHQ(flickr-faces-hq), celebA-HQ를 사용하고 마스크 이미지는 총 4가지 타입을 사용한다. 마스크 타입은 전경 객체 마스크, 배경 객체 마스크, 배경 무작위 마스크, 배경 무작위 객체 마스크로 사용되고 우선적으로 LCG의 학습을 하게 된다. LCG의 학습은 4가지 마스크 타입을 통해 해당 모듈에서 전경 잠재 임베딩과 배경 잠재 임베딩의 정보를 추출하게 된다. 객체 마스크들의 객체 클래스는 미리 정해둔 클래스를 사용하고 랜덤 마스크는 무작위 부위에 대해 마스크를 생성한다.

피라미드 다운샘플링 단계는 256, 512, 1024로 나누어 사용하게 되고 스테이지 내에서 단계별로 반복되는 횟수인 스텝 T의 값은 60, 40, 30으로 사용한다. 학습 시 초기 스테이지에서 전역적인 정보를 학습하기 위해 스테이지별 가이던스 게인을 크게 주고 다음 스테이지로 갈수록 작아지게 값을 설정한다.

Method	Places2 (large mask)		Places2 (Small mask)		CelebA-HQ	
	FID	LPIPS	FID	LPIPS	FID	LPIPS
CoModGAN	2.97	0.231	1.18	0.167	5.87	0.155
DeepFill	9.88	0.263	3.90	0.164	23.10	0.324
MADF	7.38	0.234	3.10	0.170	7.11	0.245
PixelHacker	2.64	0.177	1.06	0.142	4.89	0.198
Proposed Model	2.62	0.173	0.99	0.101	4.72	0.172

표 1은 딥러닝 방식의 모델들과 제안하는 모델의 성능을 places2와 celebA-HQ의 데이터를 사용하여 FID와 LPIPS를 비교하였다. places2에서는 마스크의 영역 크기[9]에 따라 두가지로 나누어 성능을 하였고 CelebA-HQ는 랜덤한 이미지들을 사용하여 평가하였다. 전체적으로 제안하는 모델이 FID와 LPIPS에서 좋은 성능을 보여준다. Places2 데이터에서 큰 마스크에 대해서는 pixelhacker와 비교해서 큰 차이를 보이지 않는다. 하지만 작은 마스크에 대해서는 특히 좋은 성능을 보여준다. 이는 모델이 구조 및 의미적 일관성을 유지함과 동시에 각 스테이지별로 입력되는 임베딩 정보를 사용하여 낮은 단계에서는 이미지 특징의 전역적인 정보를 학습한다. 그리고 단계가 올라갈수록 학습된 전역정보에 전경 객체 정보 외에도 주변 정보와 의미론적 일관성을 유지할 수 있다. 이때 가이던스 컴바이너(guidnace combiner)가 다음 단계 입력 시 일관성을 보정해주기 때문에 더 좋은 복원 결과를 보여준다. CelebA-HQ는 places2보다 데이터 수가 부족하고 유사한 정보들 구조인 사람 얼굴 데이터이다. 그렇기 때문에 places2 보다는 복원에 있어 모델이 인식해야할 구조에 대해서 어려움이 있어 전체적인 평가 지표가 낮게 나온 것으로 보인다.

Ⅳ. 결론

본 논문에서는 딥러닝 기반의 확산 기반 모델을 사용하여 이미지 복원의 성능을 향상시킨 모델을 제안하였다. 제안 모델은 기존의 확산 기반 모델들이 구조적 일관성과 의미적 일관성 성능에서 좋은 성능을 보여주면서 특정한 크기의 영역 복원에서도 좋은 성능을 보여준다. 복원 단계를 나누어 저해상도에서는 전역 구조와 색과 같은 부분을 복원하고 고해상도로 단계적으로 올라가며 구조 및 의미적 일관성도 확보하게 된다. 동시에 전경과 배경 임베딩의 정보를 사용하고 다음 단계로 넘어갈 때 가이던스의 보정을 통해 일관성을 보정하게 된다.

ACKNOWLEDGMENT

이 논문은 2023년도 정부(산업통상자원부)의 재원으로 한국산업 기술진흥원의 지원을 받아 수행된 연구임(P0024162, 2023년 지역혁신클러스터 육성).

참 고 문 헌

[1] A. A. Efros and T. K. Leung, “Texture synthesis by non-parametric sampling,” in Proceedings of the seventh IEEE international conference on computer vision, vol. 2. IEEE, 1999, pp. 1033 - 1038.

[2] A. Criminisi, P. Pérez, and K. Toyama, “Region filling and object removal by exemplar-based image inpainting,” IEEE Transactions on image processing, vol. 13, no. 9, pp. 1200 - 1212, 2004.

[3] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, “Context encoders: Feature learning by inpainting,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2536 - 2544.

[4] A. Lugmayr, M. Danelljan, A. Romero, F. Yu, R. Timofte, and L. Van Gool, “Repaint: Inpainting using denoising diffusion probabilistic models,” in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022, pp. 11 461 - 11 471.

[5] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, “Score-based generative modeling through stochastic differential equations,” arXiv preprint arXiv:2011.13456, 2020.

[6] Z. Xu, K. Duan, X. Shen, Z. Ding, W. Liu, X. Ruan, X. Chen, and X. Wang, “Pixelhacker: Image inpainting with structural and semantic consistency,” arXiv preprint arXiv:2504.20438, 2025.

[7] J. Ho and T. Salimans, “Classifier-free diffusion guidance,” arXiv preprint arXiv:2207.12598, 2022.

[8] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in International Conference on Medical image computing and computer-assisted intervention. Springer, 2015, pp. 234 - 241.

[9] W. Li, Z. Lin, K. Zhou, L. Qi, Y. Wang, and J. Jia, “Mat: Mask-

aware transformer for large hole image inpainting," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022, pp. 10 758 - 10 768.

[10] P. Ramachandran, B. Zoph, and Q. V. Le, "Searching for activation functions," arXiv preprint arXiv:1710.05941, 2017.

자연 발화 환경에서의 통합 표현 융합 기반 음성 감정 인식 프레임워크

전준일, 전효진, 김은이
건국대학교
jeonjunil@konkuk.ac.kr, hyojin2011@konkuk.ac.kr, eykim@konkuk.ac.kr

Speech Emotion Recognition Framework through Comprehensive Representation Fusion in Naturalistic Conditions

Jeon Junil, Jon Hyo Jin, Kim Eun Yi
Konkuk Univ.

요 약

본 논문은 자연 발화 환경에서의 정서 상태를 효과적으로 인식하기 위한 다층 융합 기반의 음성 감정 인식(Speech Emotion Recognition, SER) 프레임워크를 제안한다. 기존 SER 연구들은 연출되거나 유도된 감정 데이터를 기반으로 개발되어 실제 환경에서의 적용에 한계가 존재하며, 자연 발화에서는 감정 표현이 불명확하고 표현 방식의 다양성이 커 모델의 일반화에 어려움이 있다. 이를 해결하기 위해, 제안하는 모델은 음향 및 텍스트 정보를 단어, 발화, 임베딩 수준으로 분리하여 각각의 표현을 정교하게 추출하고, 이를 통합하여 감정 상태를 예측한다. 실험은 MSP-Podcast 데이터셋을 기반으로 수행되었으며, 개발 세트와 내부 테스트 세트를 구성하여 성능을 평가하였다. 실험 결과, 제안한 다층 융합 모델은 기존 방식 대비 약 7.5%p 향상된 Macro-F1 점수를 기록하였으며, 자연스러운 대화 환경에서도 우수한 정밀도와 일반화 성능을 보였다. 본 연구는 다층적 표현과 멀티모달 융합이 자연 발화 기반 SER의 성능 향상에 효과적임을 실증적으로 제시한다.

Ⅰ. 서 론

감정 인식 기술은 인간의 내면 상태를 이해하고 이에 적절히 반응할 수 있는 인공지능 시스템의 핵심 요소로 주목받고 있으며, 인간-컴퓨터 상호작용, AI 기반 상담 시스템, 정신건강 모니터링 등 다양한 분야에서 그 활용 가능성이 확대되고 있다. 이러한 감정 인식 기술에서 활용되는 여러 신호 중 음성은 비교적 수집이 용이하고, 억양, 리듬, 강세, 어휘적 의미 등 다층적인 정보를 포함하고 있어 감정 상태를 효과적으로 파악할 수 있는 수단으로 각광받는다. 이러한 특성으로 인해 음성 기반 감정 인식(Speech Emotion Recognition, SER)은 다양한 환경에서의 정서 상태 판단에 유용한 접근법으로 자리잡고 있다.

기존의 음성 감정 인식 연구들은 주로 연출되거나 유도된 감정 데이터를 기반으로 이루어졌으나, 이는 실제 대화 환경에서의 자연 발화와는 구조적·표현적 측면에서 차이가 존재한다. 이에 따라 다양한 화자와 주제를 반영하는 자연 발화 기반의 대규모 데이터셋 [1]이 수집되었으며, 본 연구 또한 이러한 자연스러운 환경의 데이터를 활용하여 모델을 설계하고자 한다. 자연 발화 환경에서는 감정 표현이 모호하고, 화자 간 표현 방식이나 녹음 조건 등의 편차가 크기 때문에 음성 감정 인식 시스템의 정밀성과 일반화에 큰 도전 과제가 된다. 이를 해결하기 위해 본 논문에서는 음향 및 텍스트 정보를 각각 다층적으로 표현한 후 이를 효과적으로 융합하여 감정 인식 성능을 향상시키는 프레임워크를 제안한다.

Ⅱ. 본론

본 논문에서 제안하는 프레임워크의 전체 구조는 그림 1과 같다. 음향 및 텍스트 정보를 각각 단어 수준, 발화 수준, 임베딩 수준으로 나누어 특징을 추출한 후, 이를 독립적으로 처리하고 통합하는 구조로 구성된다. 단어 수준에서는 BERTweet 기반 파서 [2]를 활용하여 각 단어의 구문적 특징을 추출하고, openSMILE 툴킷 [3]를 이용하여 loudness, shimmer, jitter 등 음향적 특징을 산출한다. 이를 통해 감정 표현의 미묘한 차이를

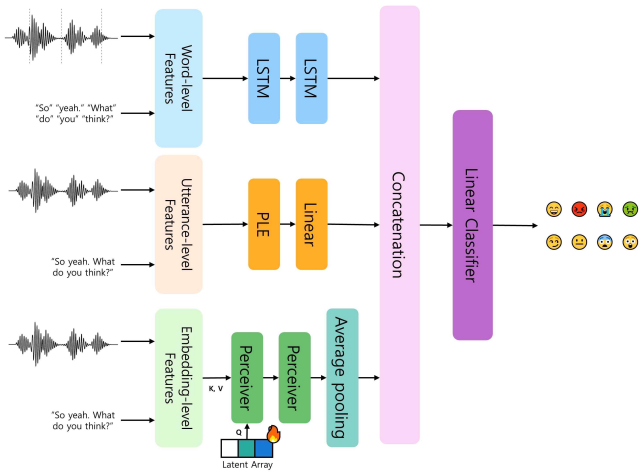


그림 1. 모델 구조. 음향 및 텍스트 특징을 단어, 발화, 임베딩 수준에서 추출, 통합

단어 단위에서 포착할 수 있다.

발화 수준에서는 SEANCE 툴킷 [4]를 사용하여 문장 전체의 감성적 경향을 반영하는 감정 점수를 산출하고, loudness, jitter, HNR, pause 등 리듬 기반 특징의 통계값을 추가로 계산함으로써 전체 발화의 음향적 흐름과 정서를 모델링한다.

임베딩 수준에서는 HuBERT [5]와 같은 사전학습 음향 모델과 BERT [6], T5 [7]와 같은 언어 모델을 활용하여 음향 및 언어 정보를 임베딩 형태로 추출하고, 이를 fine-tuning하여 도메인 적응시킨다. 이러한 임베딩은 음성의 맥락 정보를 반영하며, attentive statistics pooling을 적용하여 시간 축 전반에 걸친 통계 정보를 요약한다.

이와 같이 추출된 각 수준의 특징은 별도의 구조를 통해 처리된다. 단어 수준 특징은 LSTM [8] 구조를 통해 순차적 의존성을 학습하며, 발화 수준은 PLE [9] 구조와 선형 계층을 거쳐 표현된다. 임베딩 수준은

Perceiver [10] 구조를 통해 다양한 임베딩 소스를 통합하며, 모달 간의 정교한 상호작용을 학습할 수 있도록 설계되었다. 최종적으로 이들 세 수준의 표현을 병합하고 선형 분류기를 통해 감정 클래스를 예측한다. 이러한 다층 융합 전략은 기존의 단일 수준 기반 융합 방식이 가지는 한계를 극복하고, 다양한 화자와 자연스러운 발화 상황에서도 감정을 정밀하게 예측할 수 있도록 돕는다.

Feature	Macro-F1	Accuracy
HuBERT + BERT	0.3708	0.4882
+ Utterance-Level	0.3710	0.4984
+ Word-Level	0.3711	0.4843
+ Utterance-Level + Word-Level	0.3713	0.4864
HuBERT + T5	0.3717	0.4852
+ Utterance-Level	0.3740	0.4852
+ Word-Level	0.3748	0.4843
+ Utterance-Level + Word-Level	0.3761	0.4843

표 1. 개발 세트(Development set)에서의 특징 수준별 성능 비교

Feature Combination	Macro-F1	Accuracy
Baseline (WavLM [11])	0.3599	0.3946
HuBERT + BERT + Utterance + Word	0.4111	0.4134
HuBERT + T5 + Utterance + Word	0.4349	0.4372

표 2. 내부 테스트 세트(In-house test set)에서의 최종 모델 성능 비교

실험은 자연 발화 기반의 대규모 감정 인식 데이터셋인 MSP-Podcast [1]를 활용하여 진행되었으며, 총 8개의 감정 클래스를 대상으로 제안한 모델의 성능을 평가하였다. 평가의 일관성을 확보하기 위해, 개발 세트에서 각 감정 클래스별로 동일한 개수의 샘플을 무작위로 추출하여 5개의 내부 테스트 세트를 구성하고, 이들에 대한 평균 성능을 기준으로 최종 결과를 산출하였다.

표 1은 개발 세트에서 HuBERT 기반 음향 임베딩과 두 종류의 언어 모델(BERT, T5)을 각각 결합한 다양한 특징 수준 조합에 대한 성능을 비교한 결과이다. 단일 수준보다 다층 수준의 융합 구성이 전반적으로 우수한 성능을 보였으며, 특히 HuBERT와 T5를 기반으로 단어 및 발화 수준 특징을 통합한 조합에서 가장 높은 Macro-F1 성능(0.3761)을 기록하였다.

표 2는 내부 테스트 세트에서 기존 베이스라인 모델(WavLM [11])과 제안한 모델 간의 최종 성능을 비교한 결과이다. 제안 모델은 기존 방식 대비 약 7.5%p 향상된 Macro-F1 점수를 달성하였으며, 이는 제안한 다층 융합 전략이 자연 발화 환경에서의 감정 인식 성능을 효과적으로 개선함을 보여준다.

Ⅲ. 결론

본 논문에서는 자연 발화 환경에서의 감정 인식을 위해 다층적인 특징 표현을 활용하는 통합 프레임워크를 제안하였다. 음향 및 텍스트 정보를 각각 단어, 발화, 임베딩 수준에서 정교하게 표현하고 이를 통합함으로써, 감정 표현의 복잡성을 보다 효과적으로 포착할 수 있었다. 실험 결과, 제안하는 모델은 기존 방식 대비 우수한 성능을 보였으며, 실제 환경에서의 적용 가능성을 입증하였다. 향후 연구에서는 다양한 도메인 및 데이터셋으로 확장하여 실사용 환경에서의 적용성과 실효성을 검증할 예정이다.

참 고 문 헌

[1] R. Lotfian and C. Busso, “Building naturalistic emotionally balanced speech corpus by retrieving emotional speech from existing podcast recordings,” IEEE Transactions on Affective Computing, vol. 10, no. 4, pp. 471 – 483, 2017.

[2] H. Jiang, Y. Hua, D. Beeferman, and D. Roy, “Annotating the Tweebank corpus on named entity recognition and building NLP models for social media analysis,” arXiv preprint arXiv:2201.07281, 2022.

[3] F. Eyben, M. W¨ollmer, and B. Schuller, “Opensmile: the munich versatile and fast open-source audio feature extractor,” in Proceedings of the 18th ACM international conference on Multimedia, 2010, pp. 1459 – 1462.

[4] S. A. Crossley, K. Kyle, and D. S. McNamara, “Sentiment analysis and social cognition engine (seance): An automatic tool for sentiment, social cognition, and social-order analysis,” Behavior research methods, vol. 49, pp. 803 – 821, 2017.

[5] W.-N. Hsu, B. Bolte, Y.-H. H. Tsai, K. Lakhotia, R. Salakhutdinov, and A. Mohamed, “Hubert: Self-supervised speech representation learning by masked prediction of hidden units,” IEEE/ACM transactions on audio, speech, and language processing, vol. 29, pp. 3451 – 3460, 2021.

[6] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “Bert: Pre-training of deep bidirectional transformers for language understanding,” in Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers), 2019, pp. 4171 – 4186.

[7] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu, “Exploring the limits of transfer learning with a unified text-to-text transformer,” Journal of machine learning research, vol. 21, no. 140, pp. 1 – 67, 2020.

[8] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural computation, vol. 9, no. 8, pp. 1735 – 1780, 1997.

[9] Y. Gorishniy, I. Rubachev, and A. Babenko, “On embeddings for numerical features in tabular deep learning,” Advances in Neural Information Processing Systems, vol. 35, pp. 24 991 – 25 004, 2022.

[10] A. Jaegle, F. Gimeno, A. Brock, O. Vinyals, A. Zisserman, and J. Carreira, “Perceiver: General perception with iterative attention,” Proceedings of the International Conference on Machine Learning (ICML), 2021, pp. 4651 – 4664.

[11] S. Chen, C. Wang, Z. Chen, Y. Wu, S. Liu, Z. Chen, J. Li, N. Kanda, T. Yoshioka, X. Xiao et al., “Wavlm: Large-scale self-supervised pre-training for full stack speech processing,” IEEE Journal of Selected Topics in Signal Processing, vol. 16, no. 6, pp. 1505 – 1518, 2022.

Feasibility of Reweighted A* as SSSP for Efficient MAPF via Sorting Barrier Breakthrough

Jin Ho Kim[1], Ye Chan An[1], Yong Hyeon Song[1], Tae Yong Kuc*
[1]Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon, 16419, Korea ({jinho9812, ycan, songyong2535}@g.skku.edu)

*Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon, 16419, Korea (tykuc@skku.edu) * Corresponding author

Abstract

Multi-Agent Path Finding (MAPF) [1] is a cornerstone of autonomous systems, such as warehouse robotics and drone navigation, requiring efficient resolution of numerous shortest-path queries in time-expanded graphs. Traditional A* search, while effective in pruning the search space via heuristics, is constrained by priority queue operations with an $O(m \log n)$ [2] sorting barrier. This paper proposes reinterpreting A* as a reweighted Single-Source Shortest Path (SSSP) problem, leveraging consistent heuristics like Manhattan or Chebyshev distances to ensure non-negative edge weights. This approach enables the adoption of recent SSSP breakthroughs, such as deterministic $O(m \log^{2/3} n)$[3] algorithms, which surpass classical bounds. We demonstrate that MAPF's grid-based time-expanded graphs naturally satisfy heuristic consistency, suggesting a scalable framework that combines A*'s heuristic pruning with barrier-breaking SSSP techniques for large-scale, real-world MAPF applications.

Ⅰ. Introduction

Multi-Agent Path Finding (MAPF) is critical for applications like autonomous warehouses and drone fleets, where agents must navigate collision-free paths in shared spaces. The low-level search in MAPF relies on A*, which uses heuristics to reduce the search space but is limited by priority-queue operations costing $O(m \log n)$, where m is the number of edges and n is the number of nodes. Recent SSSP algorithms (Duan et al., 2025) break this sorting barrier, achieving $O(m \log^{2/3} n)$time. This paper investigates the feasibility of reinterpreting A* as a reweighted SSSP problem, enabling these advances to accelerate MAPF's low-level search.

Ⅱ. Methodology

The core idea is to reweight A*'s edge weights to mimic Dijkstra's algorithm while preserving its heuristic-driven efficiency. Given a graph (V, E, w), we define the reweighted edge weight as: $w'(u,v) = w(u,v) + h(v) - h(u)$, where $w(u,v)$ is the original edge weight, and h is a consistent heuristic[4] satisfying $h(v) \leq h(u) + w(u,v)$ and $h(goal) = 0$. Consistency ensures $w'(u,v) \geq 0$, allowing Dijkstra's algorithm on the reweighted graph (V, E, w') to expand nodes in the same order as A*. This equivalence is proven via a telescoping sum, where the reweighted path cost equals the original cost plus a constant offset. In MAPF, time-expanded graphs model agent movements over discrete timesteps, with edges representing spatial transitions (unit-cost moves in a grid). Common heuristics like Manhattan or Chebyshev distances are consistent in such graphs, as the heuristic difference $h(v) - h(u)$ never overestimates the true cost. For example, in a 4x4 grid with Manhattan heuristic, an edge (u,v) with $w(u,v) = 1$ and heuristic values $h(u) = 3$, $h(v) = 2$ yields $w'(u,v) = 1 + 2 - 3 = 0$, maintaining non-negativity. The table below illustrates this for sample edges:

Edge	$w(u,v)$	$h(u)$	$h(v)$	$w'(u,v)$
(u^1, v^1)	1	3	2	0
(u^2, v^2)	1	1	1	1

Recent SSSP algorithms, such as Duan et al. (2025), achieve $O(m \log^{2/3} n)$ time by optimizing comparison-based operations through advanced data structures. Since MAPF's reweighted graphs satisfy the non-negative edge condition, these algorithms are directly applicable, promising faster low-level searches compared to traditional A*.

Ⅲ. Discussion

Correctness: The reweighting preserves shortest paths, as the telescoping sum $\sum w'(u,v) = \sum w(u,v) + h(start) + h(goal)$ ensures equivalent path costs (with $h(goal) = 0$).

Consistency Requirement: Grid-based MAPF graphs with Manhattan or Chebyshev heuristics naturally satisfy consistency, ensuring $w'(u,v) \geq 0$. This is due

This paper was supported by the BK21 FOUR Project.

to the uniform structure of time-expanded graphs, where heuristic estimates align with grid movements.

Scalability: By combining A*'s heuristic pruning with barrier-breaking SSSP, the approach can handle large-scale MAPF scenarios, such as 100+ agents in DA2[5] or Game benchmarks. Preliminary analysis suggests potential speedups in query-intensive settings like warehouse logistics.

Limitation: The approach remains theoretical, requiring custom data structures to implement barrier-breaking SSSP. Integration into solvers like CBS [6] may introduce overhead in dynamic graphs, where frequent reweighting could offset gains.

It is worth noting that loosely ordered priority queues could also be directly applied to A* without reweighting, provided that termination conditions (ensuring that no remaining node has f lower than the goal) are enforced. In this paper, we adopted reweighting mainly for theoretical clarity and proof compatibility with existing SSSP results. we plan to experimentally compare the performance of using the specialized data structure in CBS by reweighting A* into Dijkstra versus applying it directly to A* without reweighting.

IV. Conclusion

This paper proves the feasibility of reinterpreting A* as a reweighted SSSP problem to accelerate MAPF's low-level search. By leveraging consistent heuristics and recent SSSP breakthroughs, we propose a framework that bridges heuristic-driven search with structural complexity reduction, potentially transforming efficiency in large-scale MAPF applications.

V. Future Work

In subsequent research, we will address several directions. Firstly, specialized data structures are needed to apply barrier-breaking SSSP algorithms efficiently in dynamic MAPF graphs which can replace priority queues such as binary heaps with loosely ordered structures that reduce sorting overhead and overall time complexity. In the next stage, we will make integration with higher-level solvers such as Conflict-Based Search (CBS) must ensure that reweighting overhead does not outweigh performance gains. Furthermore, large-scale experiments on standard benchmarks (DA2, game maps, warehouse scenarios) will be conducted to validate scalability and robustness. Lastly, learning-based extensions, such as reinforcement learning for adaptive reweighting, may further reduce search costs. Together, these efforts will strengthen the framework and move it toward practical deployment in large-scale MAPF systems.

Reference

[1] R. Stern et al., "Multi-Agent Pathfinding: Definitions, Variants, and Benchmarks," Proc. SoCS, 2019.

[2] D. B. Johnson, "Efficient Algorithms for Shortest Paths in Sparse Networks," J. ACM, vol. 24, no. 1, 1977.

[3] R. Duan et al., "Breaking the Sorting Barrier for Directed Single-Source Shortest Paths" arXiv, 2025.

[4] J. Li, A. Felner, E. Boyarski, H. Ma, and S. Koenig, "Improved Heuristics for Multi-Agent Path Finding with Conflict-Based Search," Proc. IJCAI, 2019.

[5] N. R. Sturtevant, "Benchmarks for Grid-Based Pathfinding," *Proc. SoCS*, 2012.

[6] G. Sharon et al., "Conflict-Based Search for Optimal Multi-Agent Pathfinding," Artificial Intelligence, vol. 219, 2015.

복합 시각 특성을 활용한 AI 차선 손상률 예측 모델

한준규[1], 김민석[2], 류승기[3*]

[1,2,3] 과학기술연합대학원대학교, 한국건설기술연구원

wnsrb8560@kict.re.kr jaykim98@kict.re.kr *skryu@kict.re.kr

AI-Based Lane Damage Rate Prediction Model Using Composite Visual Features

Jun Kyu Han[1], Min Seok Kim[2], Seung Ki Ryu[3*]

[1,2,3] University of Science & Technology,
Korea Institute of Civil Engineering and Building Technology

요 약

자율주행 차량의 혼입률이 점차 높아지는 교통 환경에서 차선의 시인성을 확보하는 것은 모빌리티 안전에 필수적인 요소이며, 이를 지속적으로 유지·관리하는 것은 도로 관리기관의 중요한 역할이다. 실제로 차선의 시인성은 주행 중 안전성과 직결되는 핵심 요소로 작용한다. 그러나 현행 차선 관리 체계는 도로관리자 육안에 의존한 정성적 판단을 수행한 뒤, 정밀 계측 장비를 통해 재귀반사도를 측정하는 방식으로 이루어진다. 해당 과정은 전문 인력과 고가의 장비가 필요하고, 측정 시간도 길어 넓은 도로망을 대상으로 한 광역 단위 관리가 어렵다는 한계가 존재한다. 그 결과, 차선 관리의 효율성이 떨어지고, 실시간 대응이나 선제적 유지보수에도 어려움이 뒤따르고 있다. 이에 본 연구는 기존 체계 중 1단계인 인력 의존형 육안 판단 절차를 자동화하고자, 복합 시각 특성 분석을 통해 차선 손상을 정량적으로 평가하는 새로운 기법을 제안하였다. 총 2,100장의 테스트 데이터를 활용한 실험 결과, 차선의 손상률은 1.59%에서 46.4% 범위로 분포하였다. YOLOv5s에 XGBoost 모델을 결합한 손상률 예측 모델은 R^2 = 0.9529의 성능을 기록하며 높은 정확도를 보였다. 이러한 결과는 제안 기법이 비용 효율성과 신속성을 갖춘 차선 평가 도구로서 실효성이 높음을 보여주며, 향후 도로 인프라 관리 체계의 고도화에도 기여할 것으로 기대된다.

Ⅰ. 서 론

자율주행 차량의 혼입률이 점차 높아지는 교통 환경에서 도로 인프라의 지속적인 안정성을 확보하려면, 모빌리티가 주행 경로를 정확히 인식할 수 있도록 차선의 시인성을 확보하는 것이 중요하다. 이를 위한 정기적인 도색 관리 및 보수 작업은 주행 중 안전성과 직결되는 필수 요소다.

차선이 마모, 퇴색, 스폴링(spalling)과 같이 형태적으로 손상되거나, 글라스 비드 탈락에 따른 반사도 저하로 인해 휘도가 낮아질 경우, 운전자뿐 아니라 자율주행 시스템의 차선 인식 정확도를 현저히 떨어뜨릴 수 있다. 이러한 손상은 조도가 낮은 야간이나 노면이 젖는 우천 환경에서 더욱 심각하게 작용하며, 차로 이탈 및 충돌의 위험을 높이는 요인으로 작용한다. 이와 같은 이유로, 차선 인식 성능이 저하되는 구간을 조기에 식별하고, 차선의 상태를 정량적으로 평가·관리하는 기술은 안전한 도로 환경을 유지하기 위한 필수 요소다.

현행 차선 관리 체계는, 도로관리자가 도로를 순찰하여 육안으로 차선의 불량 상태를 직관적으로 판단한 뒤, 정밀 계측 장비로 재귀반사도(Retroreflective Luminance Coefficient, mcd/m²/lx)를 측정하는 2단계 절차로 이루어진다. 이러한 조사 절차는 고가의 계측 장비와 전문 인력, 장시간의 조사 기간이 요구되며, 비실시간성 및 적용 범위의 한계로 인해 광역 도로망의 신속한 관리를 어렵게 한다[1]. 특히, 인력에 의존한 육안 평가는 정량적 기준 없이 도로관리자의 경험에 따른 정성적 판단에 머무르는 한계를 지닌다. 이에 본 연구는 이러한 한계를 보완하고자, 영상 기반 복합 시각 특성을 활용한 차선 손상률 예측 모델을 제안한다.

Ⅱ. 본론

1. 연구 목표 설정 및 관련 연구 검토

차선 인식 기술은 오랜 기간 다양한 검출 알고리즘을 중심으로 발전해 왔다. Daigavane 등[2]과 Liu 등[3]은 Canny Edge, Hough Transform 기반의 차선 인식 기법이고, Wang 등[4]은 CNN 기반의 LaneNet을 통해 픽셀 분할과 위치 추정을 결합한 구조를 제안하였다.

한편, 차선의 시인성에 대한 평가 기법 분야는 도로 관리 영역에서 연구하고 있다. 이석기 등[5]은 휘도 대비 지표를 제안하여 차선의 시인성을 정량적으로 평가하는 기법을 제시하였고, Na 등[6]은 픽셀 점유율 기반의 차선 손상평가 기법을 제안하였다.

하지만 기존 연구들은 차선을 주로 주행 경로 인식의 수단으로만 다루고 있으며, 차선의 물리적 손상 상태를 정량적으로 평가하려는 접근은 아직 부족한 실정이다. 또한 단일 지표에 기반한 손상 평가는 다양한 손상 유형이나 조건을 충분히 반영하지 못하는 한계를 가진다.

이에 본 연구는 인공지능 모델을 활용하여 주행 영상에서 차선 영역을 자동 추출하는 동시에, 복합 시각 특성을 바탕으로 차선의 손상률을 예측하는 사전 선별(pre-screening) 시스템을 제안하고자 한다. 제안된 방식은 차선의 손상 정도를 실시간으로 평가할 수 있어, 기존의 정밀 계측 방식보다 한층 효율적이고 실용적인 차선 관리 대안이 될 수 있다.

2. 복합 시각 특성 기반 차선 손상률 예측 시스템의 제안 및 연구 환경

제안된 방식은 차선 탐지 결과를 기반으로 복합 시각 특성을 융합하여 손상률을 예측하는 기법이다. Fig. 1은 제안 기법을 개발하기 위해 설계한 연구 절차를 도식화한 것이다. 절차는 (1) 전방 영상으로부터 객체 탐지 모델을 이용해 차선 영역을 추출하고, (2) 해당 영역에서 차선의 형태적 손상률과 복합 시각 특성 정보(텍스처 분산, 엔트로피, 국소적 변화량, 휘도 첨도)를 산출하며, (3) 각 특성과 손상률 간 상관관계를 분석하여 XGBoost 회귀 모델을 통한 손상 예측 및 손상 검증 과정을 포함한다.

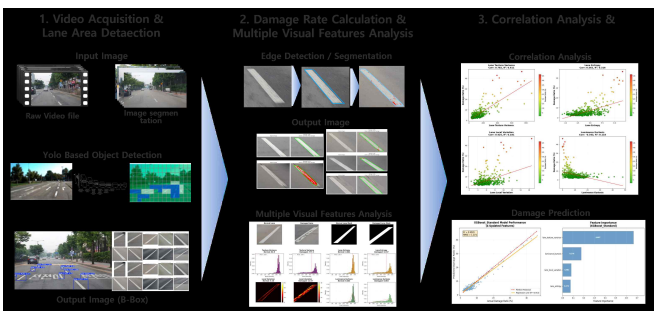


Fig. 1. 제안 기법 개발을 위한 연구 흐름도

본 연구에서는 바운딩 박스로 라벨링된 약 1,800장의 차선 이미지를 학습용 데이터셋으로 활용하였으며, 도심 및 교외 도로에서 수집한 약 2,100장의 차량 전방 촬영 이미지를 예측용 데이터셋으로 활용하였다. 실험 환경은 Python 기반 OpenCV 라이브러리와 YOLOv5s 모델로 구성되었고, 입력 영상의 하단 2/5 영역을 관심영역(ROI)으로 설정하였다. 아울러 모델 예측 후에는 복합 시각적 특성 기반의 손상률 예측 알고리즘을 수행하여, 차선 탐지와 함께 손상도를 동시에 평가할 수 있도록 설계하였다.

3. 연구 결과 분석

Table. 1은 YOLOv5s 학습 과정의 하이퍼파라미터와 성능 평가 결과이다. 모델은 차선 객체 탐지에서 높은 정밀도와 재현율을 보였으며, 평균 정밀도(mAP)에서도 우수한 성능을 보였다.

Table. 1. YOLOv5s 학습 하이퍼파라미터 및 성능 평가 결과

구분		값
하이퍼 파라미터	Epochs	100
	Batch Size	32
	Optimizer	SGD(momentum=0.937)
	Learning Rate	0.01
	Image Size	640 × 640
성능 지표	Precision	0.942
	Recall	0.833
	mAP@0.5	0.936
	mAP@0.5:0.95	0.638

형태적 손상률은 Fig. 2. a에서 보이는 바와 같이 최소 1.59%에서 최대 46.40%까지 분포한다. 또한 Fig. 2. b의 형태적 손상률 산정 예시와 같이 손상률이 증가할수록 마모가 뚜렷하게 관찰되어, 차선의 물리적 손상 정도가 효과적으로 정량화됨을 확인할 수 있다.

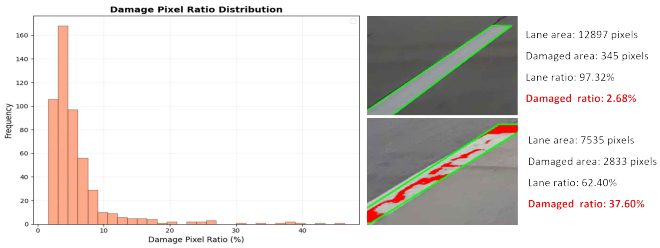


Fig. 2. 차선 손상률 분포(a) 및 손상률 계산 예시(b)

Fig. 3은 정상 차선과 손상 차선 예시 이미지에 대해서 복합 시각 특성을 비교한 결과이다. 텍스처 분산의 경우, 손상 차선은 정상 차선 대비 약 9배 높게 나타났다. 또한 엔트로피는 약 1.3배, 국소적 변화량은 약 2배 증가했으며, 휘도 첨도는 약 0.007배로 감소하여 손상으로 인한 시각 특성 변화를 정량적으로 확인할 수 있었다.

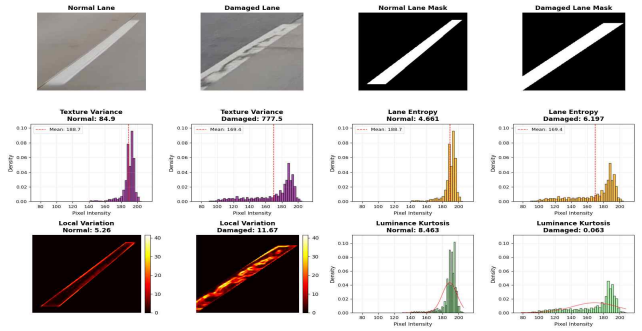


Fig. 3. 복합 시각 특성 비교 결과

Fig. 4는 차선 영역에서 추출된 복합 시각 특성과 손상률 간의 상관성을 정량적으로 분석한 결과를 보여준다. 분석 결과, 텍스처 분산(r = 0.782, R^2 = 0.611), 국소적 변화량(r = 0.428, R^2 = 0.183), 그리고 엔트로피(r = 0.355, R^2 = 0.126)는 모두 손상률 증가에 따라 뚜렷한 상승 경향을 보였으며, 해당 특성들의 상관관계를 통해 손상 정도가 심해질수록 표면 특성이 보다 복잡하고 불규칙해지는 경향이 있음을 알 수 있다. 반면 휘도 첨도(r = -0.386, R^2 = 0.149)는 손상률 증가에 따라 수치가 감소하는 경향을 보였으며, 해당 특성의 상관관계를 통해 손상 정도가 심해질수록 전체적인 휘도 분포가 균일해지는 경향이 있음을 알 수 있다.

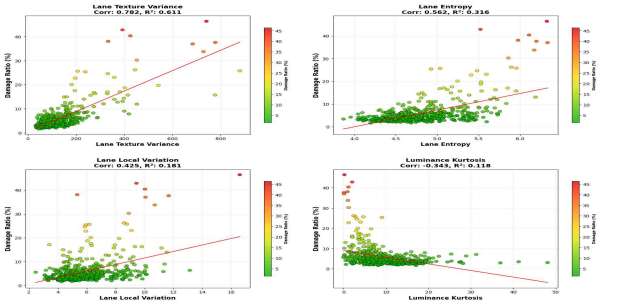


Fig. 4. 복합 시각 특성과 손상률 간의 분별력 분석

Fig. 5는 복합 시각 특성을 활용한 XGBoost 손상 예측 모델의 성능을 나타낸 결과로, R^2 = 0.9529의 매우 높은 정확도를 보였다. 특성 중요도는 텍스처 분산(66.7%)이 가장 높고 휘도 대비(17.6%), 엔트로피(8.2%), 국소 변화(7.5%) 순으로 나타났다. 또한 해당 모델은 평균 추론 시간이 0.46ms에 불과하여 연산 효율성이 매우 높으며, YOLOv5s와 결합한 예측 수행에서도 실질적용에 필요한 성능을 안정적으로 확보할 수 있었다.

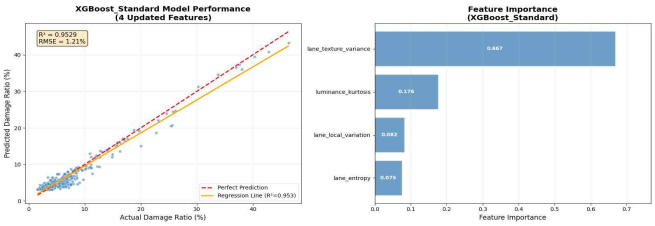


Fig. 5. 복합 시각 특성 분석을 활용한 모델 성능 및 특성 중요도

Ⅲ. 결론

본 연구에서는 도로 인프라 관리에서 육안 점검의 한계를 보완하고 차선 시인성 저하 문제에 대응하기 위해 인공지능 기반 차선 손상률 예측 시스템을 제안하였다. 연구 수행 과정에서, 인공지능 모델과 복합 시각 특성 분석을 결합하여 차선 상태를 정량적으로 평가하고 관리 효율성을 향상시킬 수 있음을 실험을 통해 검증하였다. 제안된 시스템은 도로 유지관리의 혁신은 물론, 자율주행 인프라의 안전성 향상에도 기여할 수 있을 것으로 기대된다. 향후에는 추가 데이터를 확보하여 모델의 일반화 성능을 향상시키고, 복합 시각 특성 기반의 종합 지표 설계와 정량적인 손상도 평가 기준의 정립을 통해 예측 시스템을 고도화시킬 계획이다.

ACKNOWLEDGMENT

본 논문은 한국건설기술연구원 기타 사업(과제 번호 20250449, 인공지능 기반 노면 차선 훼손 탐지 및 유지관리 지원 서비스 기술 개발) 과제의 지원으로 수행하였다.

참 고 문 헌

[1] 류승기, "훼손된 차선을 찾아내는 AI", R&D Focus, vol.03, pp.18-20, KICT zine, 2021.

[2] P. M. Daigavane, P. R. Bajaj, "Road Lane Detection with Improved Canny Edges Using Ant Colony Optimization," 2010 3rd International Conference on Emerging Trends in Engineering and Technology, pp. 76-80, 2010.

[3] G. Liu, F. Wörgötter, I. Markelic, "Combining Statistical Hough Transform and Particle Filter for Robust Lane Detection and Tracking," 2010 IEEE Intelligent Vehicles Symposium, 2010.

[4] Z. Wang, W. Ren, Q. Qiu,"LaneNet: Real-Time Lane Detection Networks for Autonomous Driving," arXiv preprint arXiv:1807.01726, 2018.

[5] 이석기, 박원일, 박기수, 김용석, "휘도 대비를 이용한 악천후 도로 노면표시 시인성능 평가," 《한국산학기술학회논문지》, vol. 24, no. 11, pp. 59 - 68, Nov. 2023.

[6] H. Na, D. G. Kim, J. M. Kang, C. Lee, "The Development of a Lane Identification and Assessment Framework for Maintenance Using AI Technology," Applied Sciences, vol. 15, no. 13, Art. 7410, Jul. 2025.

Massive MIMO 시스템의 Fixed Point Network 알고리즘 연구 동향

송치현, 김재민, 조성래
중앙대학교 컴퓨터공학과
{chsong, jmkim}@uclab.re.kr, srcho@cau.ac.kr

Recent Advances in the Massive MIMO based Fixed Point Network Algorithms

Chihyun Song, Jaemin Kim, and Sungrae Cho
Dept. of Computer Science and Engineering
Chung-Ang Univ.

요 약

기존의 딥러닝 알고리즘은 무선 통신 시스템에서 채널 추정, 신호 검출, 소스 코딩 등 다양한 통신 분야에서 뛰어난 성능을 보여주었으나, 이론적 보장과 확장성 측면에서 한계가 존재함. 또한, massive MIMO 환경에서는 채널의 차원이 극도로 커지고 특성이 복잡해져, 전통적 반복 알고리즘은 높은 계산량을 요구하며 실시간 적용이 어려움. 이를 극복하기 위해 반복 알고리즘의 각 iteration을 신경망 네트워크로 표현하는 algorithm unrolling 기법이 제안되었으나, 레이어 수에 비례하는 높은 메모리/연산 복잡도와 알고리즘의 수렴 보장을 위한 레이어 수를 일반적으로 결정할 수 없으므로 인해 초대규모 시스템에서의 활용성에는 한계가 있음. Fixed point network는 반복 알고리즘의 수렴 과정을 fixed point iteration으로 해석하여, 이를 신경망 기반 contraction mapping으로 모델링함으로써 안정적이고 빠른 수렴을 보장함. 본 논문은 fixed point network의 원리와 이를 기반으로 한 massive MIMO 시스템 연구 동향에 대해 조사하였음.

Ⅰ. 서 론

기존의 딥러닝 알고리즘은 무선 통신 시스템에서 채널 추정, 신호 검출, 소스 코딩 등 다양한 통신 분야에서 뛰어난 성능을 보여주었으나, 이론적 보장과 확장성 측면에서 한계가 존재함. 또한, massive MIMO 환경에서는 채널의 차원이 극도로 커지고 특성이 복잡해져, 기존의 모델 기반 알고리즘은 높은 계산량을 요구하며 실시간 적용이 어려움[1-3]. 이를 극복하기 위해 반복 알고리즘의 각 iteration을 신경망 네트워크로 표현하는 algorithm unrolling 기법이 제안되었으나, 레이어 수에 비례하는 높은 메모리/연산 복잡도와 알고리즘의 수렴 보장을 위한 레이어 수를 일반적으로 결정할 수 없으므로 인해 초대규모 시스템에서의 활용성에는 한계가 있음.

이러한 한계를 보완하기 위해 fixed point network (FPN)가 연구되고 있음. FPN은 반복 알고리즘의 수렴 과정을 fixed point iteration로 해석하여, 이를 신경망 기반 contraction mapping으로 모델링함으로써 안정적이고 빠른 수렴을 보장함. Banach 고정점 정리에 기반하여 유일 해로 수렴이 보장되며, 학습 및 추론 과정에서 중간 상태 저장이 필요하지 않다는 장점이 존재함[4]. 이러한 특성 덕분에 FPN은 XL-MIMO 채널 추정[5] 등 많은 연구가 진행 중이며, 본 논문에서는 FPN의 기본 원리와 이를 기반으로 한 massive MIMO 시스템 연구 동향에 대해 조사하였음.

Ⅱ. 본론

반복 알고리즘에서 갱신 구조는 일반적으로 $x^{t+1} = g(x^t, \Theta^t)$, $,t \in \{1,...,T\}$로 표현됨. 여기서 x^t는 t-번째 반복에서의 변수이고, $g(\cdot;\cdot)$는 반복 함수이며, Θ^t는 파라미터이다. Algorithm unrolling의 핵심 아이디어는 반복 과정을 여러 개의 DNN 계층으로 분리하고, 각 반복 함수 $g(\cdot;\cdot)$는 하나의 DNN 계층에 매핑하는 것이며, 업데이트 룰은 고정된 수의 DNN 연산으로 변환됨[6]. 파라미터 Θ^t는 DNN 학습 과정에서 역전파를 통해 손실 함수를 최소화함으로써 학습될 수 있음.

FPN은 반복 알고리즘을 학습 가능한 contraction mapping의 fixed point iteration으로 구성함. 이는 최적화 기반 반복 알고리즘이 무한 반복 시 얻어지는 평형 상태가 문제의 해와 일치한다는 성질을 가진다는 점에서 착안하였음. 따라서 문제의 최적해, 즉, fixed point 자체를 딥러닝 기반 mapping 함수의 학습 대상으로 설정함. 이때, contraction mapping이 Lipschitz contant < 1 조건을 만족하도록 설계되면, Banach 고정점 정리에 의해 반복 적용 시 유일한 고정점으로 수렴함이 보장됨[4]. 따라서 N번의 반복을 N개의 신경망 레이어로 구성하는 unrolling과 달리 안정적 수렴 및 추론 단계에서 계산량 조절이 가능함.

채널 추정은 대규모 안테나 시스템에서 간섭 관리 및 스펙트럼 효율 증가에 필수적이며, 기존의 LS, MMSE 기반 추정 방식은 안테나 수를 늘릴수록 고차원 행렬 연산을 수행해야 하므로 massive MIMO 환경에서는 연산량이 급격히 증가함. 이는 모델 기반 추정 방법이 coherence 시간 내에 신뢰할 수 있는 NMSE를 얻으면서 낮은 복잡성을 유지하기 어렵게 만듦. 이러한 한계를 극복하기 위해 제안된 FPN 기반 채널 추정 방식은 고전적 반복 알고리즘을 fixed point iteration으로 해석하여 신경망 구조를

설계함. Yu 등 [7]은 THz 초대규모 MIMO 시스템에서 발생하는 하이브리드 원거리 - 근거리 채널 추정 문제를 해결하기 위해 FPN-OAMP 방식을 제안함. 제안 기법은 기존 OAMP 알고리즘의 선형 추정 모듈을 유지하면서, 비선형 추정 모듈을 CNN으로 대체하고 이를 contractive mapping으로 설계함으로써 채널을 추정함. 학습 단계에서 수렴을 보장하며 추론 단계에서는 원하는 정확도에 따라 반복 횟수를 조정할 수 있어 복잡도와 성능의 트레이드오프를 제어 가능함. 시뮬레이션 결과, FPN-OAMP는 기존 OAMP 대비 약 15 dB의 NMSE 성능 향상을 달성하였으며, ISTA-Net+와 같은 unrolling 기반 접근법보다 빠르고 안정적인 수렴 특성을 보임.

Ⅲ. 결론

FPN은 반복 알고리즘의 안정적 수렴과 딥러닝을 결합하여, massive MIMO 시스템 최적화에 높은 성능을 보이며, 향후 연구는 이를 더 복잡한 네트워크 아키텍처와 실제 배치 시나리오로 확장하는 방향으로 진행될 것임.

ACKNOWLEDGMENT

This work was supported by the IITP (Institute of Information & Communications Technology Planning & Evaluation) - ITRC (Information Technology Research Center) (IITP-2025-RS-2022-00156353, 50% / IITP-2025-RS-2023-00258639, 50%) grants funded by the Korea government (Ministry of Science and ICT).

참 고 문 헌

[1] W. Yu, Y. Ma, H. He, S. Song, J. Zhang and K. B. Letaief, "Deep Learning for Near-Field XL-MIMO Transceiver Design: Principles and Techniques," in IEEE Communications Magazine, vol. 63, no. 1, pp. 52-58, January 2025, doi: 10.1109/MCOM.001.2300829.

[2] M. C. Ho, A.-T. Tran, D. Lee, J. Paek, W. Noh, and S. Cho, "A DDPG-based Energy Efficient Federated Learning Algorithm with SWIPT and MC-NOMA," ICT Express, vol. 10, no. 3, pp. 600-607, Jun. 2024.

[3] W. J. Yun et al., "Cooperative Multiagent Deep Reinforcement Learning for Reliable Surveillance via Autonomous Multi-UAV Control," IEEE Transactions on Industrial Informatics, vol. 18, no. 10, pp. 7086-7096, Oct. 2022.

[4] H. H. Bauschke and P. L. Combettes, Convex analysis and monotone operator theory in Hilbert spaces (2nd edition). Springer, 2019.

[5] W. Yu, Y. Shen, H. He, X. Yu, J. Zhang and K. B. Letaief, "Hybrid Far- and Near-Field Channel Estimation for THz Ultra-Massive MIMO via Fixed Point Networks," GLOBECOM 2022 - 2022 IEEE Global Communications Conference, Rio de Janeiro, Brazil, 2022, pp. 5384-5389, doi: 10.1109/GLOBECOM48099.2022.10001564.

[6] J. Yang et al., "Deep Learning-Based Near-Field Wideband Channel Estimation: A Joint LISTA-CP Approach," in IEEE Transactions on Vehicular Technology, doi: 10.1109/TVT.2025.3561798.

[7] W. Yu, Y. Shen, H. He, X. Yu, J. Zhang and K. B. Letaief, "Hybrid Far- and Near-Field Channel Estimation for THz Ultra-Massive MIMO via Fixed Point Networks," GLOBECOM 2022 - 2022 IEEE Global Communications Conference, Rio de Janeiro, Brazil, 2022, pp. 5384-5389, doi: 10.1109/GLOBECOM48099.2022.10001564.

O-RAN 에서의 트래픽 관리 및 자원 할당에 대한 조사

최성진, 원동욱, 김재민, 김주영, 조성래

중앙대학교 컴퓨터공학과

{sjchoi, dwwon, jmkim, jykim}@uclab.re.kr, srcho@cau.ac.kr

A Survey on Traffic Management and Resource Allocation in O-RAN

Seongjin Choi, Dongwook Won, Jaemin Kim, Juyoung Kim, and Sungrae Cho

Department of Computer Science and Engineering

Chung-Ang University

요 약

본 논문은 트래픽 예측 기반 Traffic Steering, Federated DRL, AoI 기반 슬라이싱 등 O-RAN 환경에서의 지능형 트래픽 관리 및 자원 할당과 관련된 최신 연구 동향을 분석한다. 이러한 기법들은 O-RAN 의 개방형 구조와 AI/ML 기반 RIC 의 특성을 활용하여 다양한 서비스 클래스의 요구를 유연하게 수용할 수 있도록 돕는다. 특히, 계층적 제어 구조와 mixed numerology 기반 자원 분리를 통해 슬라이스 간 간섭을 최소화하고, 실시간성과 확장성을 동시에 확보할 수 있다. 본 연구는 6G 시대의 초지능·초연결 네트워크 실현을 위한 기술적 기반 마련에 기여할 수 있다.

Ⅰ. 서 론

차세대 무선 네트워크는 초고속 전송 속도, 초저지연, 초연결성을 달성하기 위해 네트워크 슬라이싱, 빔포밍과 같은 고도화된 통신 기술들을 도입하고 있다 [1]. 그러나 기존 폐쇄형 RAN 구조는 확장성 부족과 벤더 종속성으로 인하여 다양한 서비스 클래스 (mIoT, URLLC, eMBB) 의 요구사항을 유연하게 반영하기 어렵다.

Open Radio Access Network (O-RAN) 는 개방형 인터페이스와 RAN Intelligent Controller (RIC) 를 기반으로 다종 벤더 장비 간 상호운용성을 보장하며, AI/ML 을 통한 지능형 자원 관리가 가능하다. 특히 실시간 트래픽 패턴을 예측하거나 QoS/QoE 기반의 자원 재분배에서 AI/ML 이 중요한 역할을 하고 있다 [2], [3].

최근 연구에서는 O-RAN 의 발전 방향으로 지능성과 프로그램 가능성의 통합이 대두되고 있다. 예를 들어, Non-Real Time RIC (Non-RT RIC)와 Near-Real Time RIC (Near-RT RIC) 간의 계층적 제어 구조를 활용하여 정책 기반 제어와 동적 자원 할당을 결합하는 방안이 제안되고 있다. 이를 통해 트래픽 부하에 따른 네트워크 자원의 효율적 분배가 가능해진다 [4], [5]. 또한, 딥러닝 기반 O-RAN 관리 프레임워크는 네트워크 슬라이싱, 사용자 스케줄링, 빔포밍 최적화 등 다양한 자원 관리 문제를 데이터 기반으로 해결할 수 있음을 보여준다 [6].

이와 같이 O-RAN 환경에서 트래픽 관리와 자원 할당을 최적화하는 연구는 6G 시대의 초연결, 초지능 네트워크 실현에 핵심적인 역할을 할 것으로 기대된다. 본 논문에서는 O-RAN 의 구조와 특징을 기반으로, 트래픽 관리 및 자원 할당과 관련된 최신 연구 동향을 분석하고 향후 발전 방향을 제시한다.

Ⅱ. 본론

II-1. 트래픽 관리 기법

차세대 무선 네트워크에서는 다양한 서비스 클래스의 공존으로 인해 지능적이고 유연한 트래픽 관리가 중요해지고 있다. O-RAN 은 non-RT RIC 과 near-RT RIC 의 계층 구조를 통해 트래픽 예측과 정책 기반 자원 최적화를 가능하게 한다.

[7]의 저자는 LSTM 기반의 트래픽 예측 모델을 적용한 지능형 Traffic Steering 프레임워크를 제안하였다. 이 프레임워크는 실시간 상황에 따라 사용자를 적절한 기지국에 배치하거나 자원의 효율적인 재배치를 수행하는데, rApp 이 수행하는 트래픽 예측 및 슬라이싱 결정과 xApp 이 수행하는 실시간 자원 할당을 결합해 구성된다. 이는 URLLC 의 저지연성과 eMBB 의 고속 처리 요구를 동시에 만족시키면서 실시간 트래픽 변화에 민감하게 대응할 수 있도록 설계되었다. 또한, 다양한 서브캐리어 구조를 적용한 슬라이스 간 간섭 방지를 통해 서비스 간 간섭을 최소화하고, 기존 RAN 구조의 유연성 부족 문제를 보완함으로써, 고도화된 트래픽 관리 전략으로서의 가치를 입증하고 있다.

II-2. 자원 할당 기법

O-RAN 의 핵심 목표 중 하나는 QoS 보장과 함께 자원 효율성을 극대화하는 것이다. 이를 위해 최근 연구들은 딥러닝 기반의 자원 할당 최적화 기법에 주목하고 있다.

[8]의 저자는 다수의 Mobile Virtual Network Operator (MVNO) 가 존재하는 환경에서 Federated Deep Reinforcement Learning (FDRL) 기반의 협력적 RAN 슬라이싱 프레임워크를 제안하였다. 본 접근 방식에서는 개별 MVNO 가 자사의 사용자 데이터를 기반으로 DDPG 모델을 학습하고, 해당 모델을 RIC 에서 집계한 후 전체 MVNO 에 다시 배포하는 과정을 통해 데이터 프라이버시는 보장하면서도 다양한 환경에 일반화 가능한 DRL 모델을 구축할 수 있다.

[9]의 저자는 ORAN 에서 발생하는 서버 지연을 최소화하는 지연 인지형 자원 할당 최적화 모델을 제시하였다. 본 연구에서는 서버 지연을 비선형(지수적) 함수로 표현하며, 전력 소모, 리소스 블록, 가상 머신의 수와 같은 제약 조건 하에서 전체 지연을 최소화하는 목적함수를 구성하였다. 문제의 비선형성과 복잡성에 대응하기 위해 라그랑주 승수법을 통해 문제를 수학적으로 모델링하고, Active-set, Interior Point, Sequential Quadratic Programming 등 세 가지 최적화 알고리즘을 비교하였다. 이를 통해 실제 트래픽 환경에서 지연 민감형 서비스의 QoS 를 만족시키기 위한 정형적인 최적화 접근이 가능함을 보였다.

[10]의 저자는 산업용 IoT(IIoT) 환경을 위한 탄력적 O-RAN 슬라이싱(elastic slicing) 프레임워크를 제안하였다. 본 연구에서는 정보 신선도(AoI)를 최소화하면서 에너지 소모와 슬라이스 격리성을 동시에 고려한 자원 할당 최적화 문제를 정의하였고, 이 문제는 NP-hard 이다. 이를 해결하기 위해 분산 매칭 게임 기법과 심층 강화학습 기반 Actor-Critic 모델을 제안하였다. 이 구조는 IIoT 서비스의 패턴과 요구사항을 학습함으로써 장기적인 정책 기반의 자원 분배가 가능하며, 기존 방식 대비 평균 50% 이상의 장비 지원 성능 향상을 달성함을 보였다.

Ⅲ. 결론

O-RAN 은 개방형 아키텍처와 AI/ML 기반의 RIC 를 통해 기존 폐쇄형 RAN 구조의 한계를 극복하고, 다양한 서비스 클래스의 요구를 유연하게 수용할 수 있는 기반을 제공한다. 본 논문에서는 O-RAN 환경에서의 트래픽 관리 및 자원 할당과 관련된 최신 연구들을 분석하였다. 트래픽 관리 측면에서는 LSTM 기반 예측과 rApp 과 xApp 의 연동을 통한 실시간 Traffic Steering 이 효과적인 방안으로 제시되었으며, 다양한 서브캐리어를 활용한 슬라이스 간 간섭 방지를 통해 서비스 품질이 향상됨을 확인하였다. 자원 할당 측면에서는 Federated DRL 을 활용한 프라이버시 보장형 슬라이싱, 지연 최소화를 위한 수학적 최적화 모델, 그리고 AoI 기반의 탄력적 슬라이싱 기법 등 다양한 접근 방식들이 제안되었다. 이러한 연구들은 O-RAN 이 지향하는 지능형, 자율형 네트워크 구조를 실현하는 데 핵심적인 역할을 하며, 향후 차세대 네트워크 서비스를 안정적으로 제공하기 위한 기반 기술로 발전할 것으로 기대된다. 향후에는 실제 상용 환경에서의 적용 가능성과 운영 효율성에 대한 추가적인 실증 연구가 필요하다.

ACKNOWLEDGMENT

This work was supported by the IITP (Institute of Information & Communications Technology Planning & Evaluation) - ITRC (Information Technology Research Center) (IITP-2025-RS-2022-00156353, 50% / IITP-2025-RS-2024-00436887, 50%) grants funded by the Korea government (Ministry of Science and ICT).

참 고 문 헌

[1] M. Choi, J. Kim and J. Moon, "Wireless Video Caching and Dynamic Streaming Under Differentiated Quality Requirements," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 6, pp. 1245-1257, June 2018.

[2] M. Polese, L. Bonati, S. D'Oro, S. Basagni and T. Melodia, "Understanding O-RAN: Architecture, Interfaces, Algorithms, Security, and Research Challenges," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 2, pp. 1376-1411, 2023.

[3] W. J. Yun, D. Kwon, M. Choi, J. Kim, G. Caire and A. F. Molisch, "Quality-Aware Deep Reinforcement Learning for Streaming in Infrastructure-Assisted Connected Vehicles," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 2, pp. 2002-2017, Feb. 2022.

[4] A. Arnaz, J. Lipman, M. Abolhasan and M. Hiltunen, "Toward Integrating Intelligence and Programmability in Open Radio Access Networks: A Comprehensive Survey," *IEEE Access*, vol. 10, pp. 67747-67770, 2022.

[5] D. Kwon, J. Jeon, S. Park, J. Kim and S. Cho, "Multiagent DDPG-Based Deep Learning for Smart Ocean Federated Learning IoT Networks," *IEEE Internet of Things Journal*, vol. 7, no. 10, pp. 9895-9903, Oct. 2020.

[6] B. Brik, K. Boutiba and A. Ksentini, "Deep Learning for B5G Open Radio Access Network: Evolution, Survey, Case Studies, and Challenges," *IEEE Open Journal of the Communications Society*, vol. 3, pp. 228-250, 2022.

[7] F. Kavehmadavani, V. -D. Nguyen, T. X. Vu and S. Chatzinotas, "Intelligent Traffic Steering in Beyond 5G Open RAN Based on LSTM Traffic Prediction," *IEEE Transactions on Wireless Communications*, vol. 22, no. 11, pp. 7727-7742, Nov. 2023.

[8] A. Abouaomar, A. Taik, A. Filali and S. Cherkaoui, "Federated Deep Reinforcement Learning for Open RAN Slicing in 6G Networks," *IEEE Communications Magazine*, vol. 61, no. 2, pp. 126-132, February 2023.

[9] B. N. Khalaf, W. H. Ali, R. S. Alhumaima, H. A. J. Alshamary, "Delay aware resource allocation in ORAN through network optimization," *IET Wireless Sensor Systems*, vol. 14, pp. 528-538, 2024.

[10] S. F. Abedin, A. Mahmood, N. H. Tran, Z. Han and M. Gidlund, "Elastic O-RAN Slicing for Industrial Monitoring and Control: A Distributed Matching Game and Deep Reinforcement Learning Approach," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 10, pp. 10808-10822, Oct. 2022.

동적 그래프 학습 기반 통신 트래픽 예측 연구 동향 조사

김재민, 오준석, 김가현, 송치현, 조성래
중앙대학교 컴퓨터공학과
{jmkim, jsoh, ghkim, chsong}@uclab.re.kr, srcho@cau.ac.kr

Recent Advances in the network for traffic prediction based on dynamic graph learning

Jaemin Kim, Junsuk Oh, Gahyun Kim, Chihyun Song, and Sungrae Cho
Dept. of Computer Science and Engineering
Chung-Ang Univ.

요 약

정적 인접행렬이나 격자/순차 모델은 이동성·간섭·백홀로 인한 시변 관계를 충분히 반영하지 못한다. 본 논문에서는 동적 그래프 학습 관점에서 핸드오버 기반 동적 그래프, 정적 사전정보+학습형 보정(하이브리드/적응형), Dynamic Graph Attention, Inductive 일반화를 중심으로 기법과 순차 분할, 이벤트 주석, SLA·피크 오차 등을 정리하여 최신 동향을 조사하고, 향후 과제에 대해 논의한다.

Ⅰ. 서 론

통신 트래픽 예측은 시계열 모델과 공간 모델로 발전해 왔다. 이들 접근은 규칙적인 패턴과 기지국 간 평균적인 공간 상관(지리적 근접, 고정 이웃 관계)을 비교적 안정적으로 학습한다는 장점이 있다. 그러나 관계 구조가 시간에 따라 변한다는 현실을 충분히 반영하지 못한다는 한계가 명확하다. 구체적으로, 표현 단계에서는 격자화나 고정 인접행렬이 핸드오버·간섭·공유 백홀 같은 시변 상호작용을 놓치기 쉽다[1]. 시간 모델링 단계에서는 긴 예측 구간에서 오차가 누적되며, 이벤트성 수요 급증이나 급격한 혼잡 완화를 장·단기 동시로 포착하기 어렵다. 학습·검증 단계에서는 시간 누수, 외생 요인 미반영으로 레짐 전환에 대한 일반화가 과대평가되기 쉽다. 배치 단계에서는 새로운 기지국의 추가·삭제, 데이터 결측·지연, 제한된 추론 지연·연산 예산에 대한 적응성이 부족해 운영 지표를 충분히 낮추기 어렵다.

이러한 한계를 완화하기 위해 최근 연구는 동적 그래프 학습을 중심으로 재편되고 있다[2],[3]. 핵심은 네트워크를 노드(셀/기지국) 집합과 시각별 엣지 집합으로 표현하고, 각 시각의 인접행렬을 데이터와 운용 제약에서 직접 추정해 예측 모델과 공동 최적화하는 것이다. 대표적으로 핸드오버·전이확률 등 도메인 신호를 반영한 동적 그래프 구성, 지리·백홀 등 정적 사전정보와 학습형 보정을 결합하는 하이브리드/적응형 그래프, 시점마다 유효 이웃을 선택하는 Dynamic Graph Attention, 새로운 지역·토폴로지에도 적용 가능한 유도적 일반화, 장기 예측과 대규모 노드 수에 대응하는 시간 모듈 설계와 희소화·차수 제한·정적 마스크에 기반한 복잡도 관리, 운영 지표와 순차 검증을 포함한 공정한 평가 프로토콜이다. 본 논문은 이러한 부분을 중심으로 최신 기법을 분류하고, 각각 실제 배치에서 어떤 문제를 해결하는지 명확히 연결한다.

Ⅱ. 본론

본 연구 주제는 셀/기지국 단위의 트래픽을 시공간적으로 예측하는 문제다. 이에 따라 네트워크를 노드 집합(셀/기지국)과 시각별 엣지 집합(당시점의 유효한 연결 관계)로 표현한다[2]. 각 시각의 연결 강도는 인접행렬로 나타내며, 이는 시간에 따라 동적으로 변할 수 있는 양으로 취급한다. 입력은 과거 관측 창 길이에 해당하는 과거 다수 시점의 특성이고, 출력은 예측에 해당하는 가까운 미래에 대한 수요다.

동적 그래프의 출발점은 “무엇을 이웃으로 볼 것인가”를 시간의 함수로 정의하는 일이다. 실제 망에서는 사용자 이동이 만드는 핸드오버 전이, 셀 간 간섭/경합, 백홀 연결성이 시변 관계를 지배한다. 이에 따라 Fang 등[3]은 SDGNet을 통해 핸드오버 기반 동적 그래프를 전이 빈도나 전이확률을 이용해 유향·가중 인접행렬을 시각별로 갱신함으로써, 거리 인접만으로는 설명하기 어려운 이동성 상관을 직접 반영한다. Kalander 등[4]은 STGCN-HO를 통해 하이브리드/적응형 그래프를 지리·백홀·운용 규칙 등으로 얻은 정적 사전정보를 기반 인접행렬로 두고, 데이터에서 학습한 보정 항을 혼합해 최종 인접행렬을 구성한다. 이 방식은 물리적 일관성을 유지하면서 숨은 상관을 흡수하는 장점이 있다. Jin 등[5]은 Dynamic Graph Attention 기반 구성으로 시점마다 주목할 이웃을 선택해 가중치를 재분배함으로써, 급변하는 근접/경합 관계를 유연하게 포착한다. 공통적으로는 희소화(Top-K), 차수 제한, 정적 마스크(지리·규격·백홀 제약), 시간 매끄러움 정규화와 같은 안정화 장치가 함께 쓰여 인접행렬의 불필요한 변동을 억제하고 해석 가능성을 확보한다.

동적 그래프가 공간 상관을 제공하면, 이를 시간 축에서 어떻게 통합할지가 성능과 비용을 좌우한다. Fang 등[3]과 Wang 등[6]은 TCN(팽창 합성) 류를 긴 예측 지평에서 효율적 추론과 안정적인 수렴을 제공해 대규모 트래픽에 적합하고, 상태공간 모델이나 경량 Transformer는 긴 의존성과 계산 비용의 균형을 맞추는 대안으로 채택하였다. 반면, Kalander 등[4]은 매우 짧은 반응 시간이 요구되는 환경(급격한 핸드오버·경합 변화)에서는

입력 창을 짧게 가져가고, 낮은 학습률의 온라인 파인튜닝을 결합하여 실용성을 확보하였다. 모빌리티와 같이 급변하는 링크 환경에서는 예측을 상태로 포함한 품질-인지 제어가 끊김과 지연을 줄이는 데 유효하다[7]. 스케일 측면에서는 노드 수가 증가할수록 그래프 연산의 병목이 커지므로, 희소화·차수 제한·계층 그래프로 연산량을 관리하고, 필요 시 멀티스케일 시간 블록을 두어 일상적 주기와 이벤트성 변동을 분리 처리한다. 실제 데이터의 결측·지연 관측은 모델 내부의 단순 임퓨테이션 블록으로 보정하여 예측-정책 연계를 안정화하는 것이 일반적이다.

운영 환경에서는 기지국의 추가/삭제, 파라미터 재설정, 데이터 희소/불완전 문제가 상시 발생한다. 이에 대한 해법으로, Ngo 등[8]은 유도적(Inductive) 일반화를 채택한 모델을 학습 시 보지 못한 그래프(새 지역·새 토폴로지)에도 예측기를 그대로 이식할 수 있도록 구조적 바이어스를 둔다. 이 접근은 라벨이 적거나 분포가 다른 데이터에서도 성능 하락을 제한하는 데 효과적이다. 더불어, 불확실성 정량화(분위수·예측구간)를 출력에 포함하면, 예측을 정책으로 연결할 때 위험 민감도를 조절할 수 있어 SLA 위반 위험을 완화하는 데 유용하다. 프라이버시·데이터 이동 제약이 큰 시나리오에서는 연합학습/전이학습을 결합해 사업자·지역 간 분포 차이를 흡수하는 사례가 늘고 있으며, 이때도 동적 그래프의 정규화와 사전정보 혼합은 안정성 확보에 핵심적으로 기여한다.

Wang 등[1]은 시간 순서를 존중하는 순차 분할을 기본으로 하고, 공휴일·행사·대형 릴리즈 같은 이벤트 캘린더를 명시해 레짐 전환 성능을 별도로 확인한다. 지표는 MAE/RMSE에 더해 피크 시점 오차, SLA 위반율, 과소·과대 프로비저닝 비용을 포함하여 운영 적합성을 가늠한다. 또한 동일한 데이터·지평·입력 창에서 정적 대 동적 그래프, 희소화 강도, 사전정보 혼합 비율, 업데이트 주기 등 어블레이션을 체계적으로 제시해야 모델의 순효과를 분리하여 해석할 수 있다. 배치 측면에서는 Near-RT 제어 루프에 동적 인접행렬 갱신과 예측 모듈을 연결하고, 임계값 기반 정책 트리거(트래픽 스티어링, 자원 재할당, 절전 모드)를 구성하되, 임계값을 불확실성 지표와 연동해 안전 전환이 가능하도록 설계한다. 예측 결과는 사전 캐싱·품질 적응과 직접 결합되며, 등급화된 QoE와 자원 제약을 함께 고려한 정책 설계가 효과적일 수 있다[9]. 이와 같은 프로토콜을 따르면, 모델의 정확도 향상을 운영 지표 개선으로 직결시키는 경로를 확보할 수 있다.

Ⅲ. 결론

본 논문은 통신 트래픽 예측에서 정적 그래프 가정의 한계가 분명하며, 대안으로 동적 그래프 학습이 핵심 축으로 자리 잡았음을 정리했다. 향후 과제는 네 갈래로 정리된다. 첫째, 시변 인접행렬의 과도한 변동을 억제하는 시간 매끄러움·희소성·스펙트럴 제약과 불확실성 정량화의 표준화가 요구된다. 둘째, Top-K, 차수 제한, 계층 그래프, 멀티스케일 시간 블록을 통한 연산·메모리 관리로 대규모 망에서의 지속적 동작을 보장해야 한다. 셋째, 동일 분할·동일 지평·동일 지표·동일 어블레이션을 공개해 비교 가능성을 높이고, 외부 요인·도메인 전이·연합학습 시나리오를 포함한 평가 프로토콜을 정립해야 한다. 넷째, O-RAN 기반 제어 루프에서 예측 신뢰도에 연동된 임계값 설계, 안전 전환(롤백), 프라이버시 제약 하 연합학습의 운용 지침을 구체화해야 한다. 이 네 축이 충족될 때, 동적 그래프 학습은 정확도 개선을 넘어 운영 지표의 실질적 개선으로 이어지는 표준 해법으로 자리매김할 것이다.

ACKNOWLEDGMENT

This work was supported by the IITP (Institute of Information & Communications Technology Planning & Evaluation) - ITRC (Information Technology Research Center) (IITP-2025-RS-2022-00156353, 50%) grant and YKCS Open RAN Global Collaboration Center (IITP-2025-RS-2024-00434743) grant funded by the Korea government (Ministry of Science and ICT).

참 고 문 헌

[1] Wang, Xing, et al. "A survey on deep learning for cellular traffic prediction." Intelligent Computing 3 (2024): 0054.

[2] Zhao, Shuai, et al. "Cellular network traffic prediction incorporating handover: A graph convolutional approach." 2020 17th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON). IEEE, 2020.

[3] Fang, Yini, Salih Ergüt, and Paul Patras. "SDGNet: A handover-aware spatiotemporal graph neural network for mobile traffic forecasting." IEEE Communications Letters 26.3 (2022): 582-586.

[4] Kalander, Marcus, et al. "Spatio-temporal hybrid graph convolutional network for traffic forecasting in telecommunication networks." arXiv preprint arXiv:2009.09849 (2020).

[5] Jin, Zilong, et al. "A mobility aware network traffic prediction model based on dynamic graph attention spatio-temporal network." Computer Networks 235 (2023): 109981.

[6] Wang, Xing, et al. "Adaptive hybrid spatial-temporal graph neural network for cellular traffic prediction." ICC 2023-IEEE International Conference on Communications. IEEE, 2023.

[7] Yun, Won Joon, et al. "Quality-aware deep reinforcement learning for streaming in infrastructure-assisted connected vehicles." IEEE Transactions on Vehicular Technology 71.2 (2021): 2002-2017.

[8] Ngo, Duc-Thinh, et al. "FLEXIBLE: Forecasting Cellular Traffic by Leveraging Explicit Inductive Graph-Based Learning." 2024 IEEE 35th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC). IEEE, 2024.

[9] Choi, Minseok, Joongheon Kim, and Jaekyun Moon. "Wireless video caching and dynamic streaming under differentiated quality requirements." IEEE Journal on Selected Areas in Communications 36.6 (2018): 1245-1257.

분해된 보행 패턴을 활용한 뱀 로봇 강화학습의 탐색 공간 축소

송봉섭*, 권우경, 남승우, 정윤수
한국전자통신연구원 지능로봇시스템연구실

*sbs@etri.re.kr, wkwon@etri.re.kr, swnam@etri.re.kr, yoonsu@etri.re.kr

Reducing Exploration Space in Snake Robot RL using Decomposed Gaits

Bongsub Song*, Wookyong Kwon, Seung Woo Nam, Yun Su Chung
Electronics and Telecommunications Research Institute (ETRI)

요 약

본 연구는 뱀 로봇의 자연스러운 움직임에서 영감을 받아 강화학습의 탐색 공간을 축소하고 학습 효율을 향상시키는 Gait Decomposition(GD) 기법을 제안한다. 제안된 방법은 기존 곡선 기반 게이트를 시계열 토크 방향 모션 행렬로 분해하여, 생체 모방의 장점을 유지하면서도 초기 학습의 수렴 속도를 개선한다. 시뮬레이션을 통해 GD 의 학습 효율성과 성능 향상 효과를 검증하였다.

Ⅰ. 서 론

뱀 로봇은 모듈형 구조를 바탕으로 다양한 환경에서 높은 기동성과 유연성을 발휘한다 [1]. 최근에는 강화학습 기반 제어기 설계가 활발히 연구되고 있으나 [2], 고차원의 상태 및 행동 공간으로 인해 학습 속도가 저하되고 안정적 수렴이 어려운 한계가 존재한다. 본 연구는 생물학적 움직임에서 관찰되는 반복적·대칭적 게이트(gait) 특성을 수학적으로 분해함으로써 [3], 강화학습의 탐색 공간을 효과적으로 축소하는 새로운 접근법을 제안한다.

Ⅱ. 제안된 방법론 (Methods)

$$\boldsymbol{\theta} = \begin{pmatrix} i = \text{odd}\,, & \theta_i = \beta_{odd} + A_{odd}\sin(\xi_{odd}) \\ i = \text{even}\,, & \theta_i = \beta_{even} + A_{even}\sin(\xi_{even} + \eta) \end{pmatrix} \quad (1)$$

$$\xi_{odd} = \psi_{odd} i + \nu_{odd} t$$
$$\xi_{even} = \psi_{even} i + \nu_{even} t \quad (2)$$

기존 Serpenoid 곡선 기반 게이트는 수식 (1), (2)와 같이 Dorsal 및 Lateral 방향의 관절 궤적을 시간에 따라 정의한다. 이에 반해 본 연구에서는 곡선을 시간–관절 토크 방향 모션 행렬로 변환하였다. 변환된 Gait Motion Matrix 는 각 시점별 토크 방향을 -1 또는 +1 로 이산화하여 표현하며, 이를 기반으로 토크 크기를 출력하는 신경망을 학습시킨다. 그림 1 은 Serpenoid 함수와 토크 기반 임베딩 행렬 간의 변환 과정을, 그림 2 는 Slithering gait 를 패턴화하여 표현한 예시를 보여준다. 이 구조는 복잡한 함수 최적화 대신 간단한 토크 프로파일 학습만으로 다양한 게이트 유형을 유지할 수 있다는 장점을 갖는다.

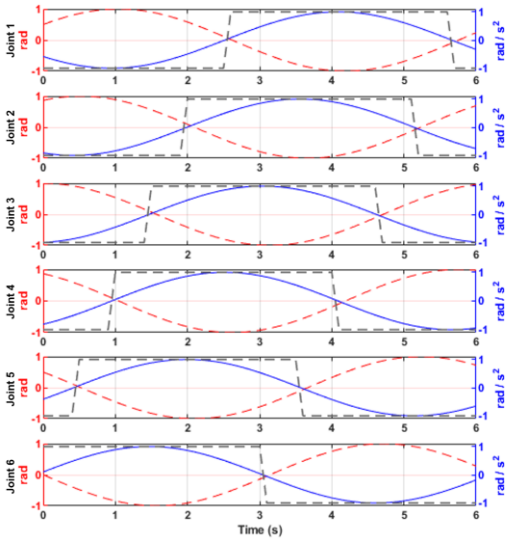


그림 1 Serpenoid 함수로부터 얻어진 관절의 토크 임베딩

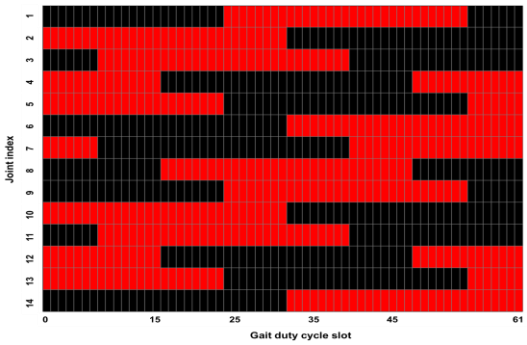


그림 2 Slithering Gait 의 패턴 표현

그림 3 은 임베딩 행렬을 기반으로 뱀 로봇의 토크 신호를 생성하는 알고리즘 개요를 제시한다. PPO 와 SAC 알고리즘을 이용해 학습한 결과, GD 를 적용한 정책은 기존 Serpenoid 기반 정책에 비해 빠르게 수렴하였으며, 그림 4 에서 확인되듯이 에피소드 길이가

Ⅲ. 실험 결과

빠르게 증가하여 안정적인 제어가 가능함을 보였다. 또한 그림 5 와 같이 GD 기반 정책은 더 높은 보상 값을 달성하여 학습 효율이 개선됨을 확인하였다.

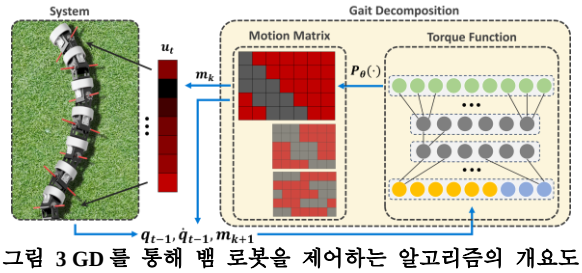


그림 3 GD 를 통해 뱀 로봇을 제어하는 알고리즘의 개요도

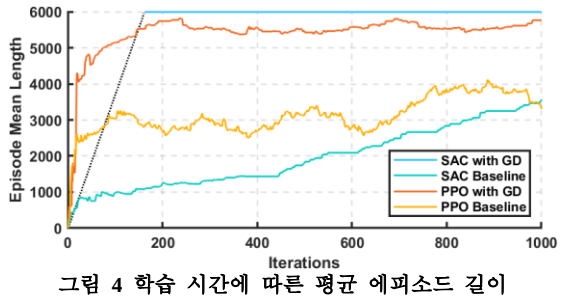


그림 4 학습 시간에 따른 평균 에피소드 길이

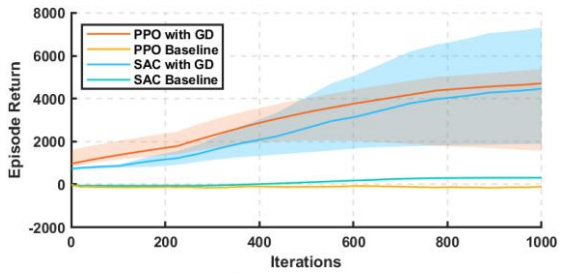


그림 5 학습 시간에 따른 평균 보상 값의 크기

Ⅳ. 결론

본 연구는 뱀 로봇의 생체 모방적 특성을 활용하여 강화학습의 효율성을 높이는 토크 기반 Gait Decomposition 기법을 제안하였다. 제안된 방법은 탐색 공간을 효과적으로 축소하여 학습 속도와 안정성을 향상시키며, 실제 적용 가능성을 높일 수 있다. 향후에는 다양한 지형에서의 일반화 성능 검증과 실시간 정책 업데이트 가능성에 대해 추가 연구를 진행할 예정이다.

ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 과학기술사업화진흥원의 지원을 받아 수행된 연구임 ('학연협력플랫폼구축 시범사업' RS-2023-00304776).

참 고 문 헌

[1] Wright, Cornell, et al. "Design and architecture of the unified modular snake robot." 2012 IEEE international conference on robotics and automation. Ieee, 2012.

[2] Zhang, Dandan, et al. "One-shot domain-adaptive imitation learning via progressive learning applied to robotic pouring." IEEE Transactions on Automation Science and Engineering 21.1 (2022): 541-554.

[3] Song, Bongsub, Insung Ju, and Dongwon Yun. "Optimizing snake robot locomotion with decomposed gait pattern representation." Integrated Computer-Aided Engineering 32.2 (2025): 196-225.

로봇 매니퓰레이터의 동작 계획을 위한 모델 예측 경로 적분 제어의 적응형 샘플링 접근

유지민[1], 강동엽[2], 동지연[2], 박찬은[1]

[1]경북대학교, [2]한국전자통신연구원

rktdbzos45@knu.ac.kr, kang@etri.re.kr, jydong@etri.re.kr, chaneun@knu.ac.kr

Adaptive sampling approach to model predictive path integral control for motion planning of robot manipulators

Ji Min Yu[1], Dongyeop Kang[2], Jeyeon Dong[2], Chan-eun Park[1]

[1]Kyungpook National University, [2]Electronics and Telecommunication Research Institute

요 약

본 연구는 동적 장애물 환경에서 로봇의 안전한 경로 계획을 위해 모델 예측 경로적분(Model Predictive Path Integral, MPPI) 제어와 위험도 기반 적응형 샘플링을 결합한 방법을 제안한다. 제안 방법은 장애물과의 상대적 위험도에 따라 샘플 수를 변화시키는 방법을 통해 탐색 밀도와 계산 효율을 동시에 확보한다. NVIDIA cuRobo 프레임워크와 Isaac Sim을 이용한 시뮬레이션 실험에서 제안된 방법과 고정 샘플링 방식의 비교 평가를 수행하였다. 실험 결과를 통해 제안 방식은 충돌 발생을 줄이고 목표 도달의 안정성을 높이는 동시에 계산 비용을 관리 가능한 수준으로 유지하여 안전성과 효율성을 유의미하게 향상시킬 수 있음을 보여준다.

Ⅰ. 서론

최근 로봇 매니퓰레이터는 공장 물류 분류, F&B 산업의 음료 및 음식 제조, 의료 분야의 보조 작업 등 다양한 산업 현장에서 활용되고 있으며, 복잡한 환경 속에서도 안전하고 정밀한 작업 수행 능력이 요구되고 있다. 이에 따라 로봇이 작업 도중 주변 환경과 상호작용하면서도 충돌을 피할 수 있도록 하는 경로 계획 기술이 활발히 연구되고 있다. 특히, 실제 환경에서는 고정된 작업 공간의 정적 장애물뿐만 아니라, 작업자·이동 로봇·컨베이어 시스템 등과 같이 시간에 따라 위치와 속도가 변화하는 동적 장애물이 존재하므로, 이를 회피할 수 있는 능력이 필수적이다[1]. 이러한 환경에서 로봇 매니퓰레이터의 경로 계획은 안전성과 효율성을 동시에 만족해야 한다. 이를 위한 대표적인 기법 중 하나인 MPPI 제어는 샘플링 기반 탐색을 통해 복잡한 동역학과 다양한 제약 조건을 처리할 수 있으며, 실시간 정책 갱신이 가능하다는 점에서 널리 활용되고 있다. 그러나 기존의 MPPI는 고정된 샘플 수 설정에 의존하기 때문에, 동적 장애물 환경에서 일관되게 성능을 보장하기 어렵다는 한계가 존재한다[2]. 예를 들어, 샘플 수가 적으면 위험 구간에서 탐색이 부족해 충돌 가능성이 높아지고, 반대로 샘플 수가 과도하면 안전 구간에서도 불필요한 연산이 발생하여 실시간성이 저하된다. 이러한 특성은 MPPI의 실제 환경 적용성을 제약하는 주요 원인으로 지적되고 있다.

본 연구에서는 이러한 문제를 해결하기 위해 위험도 기반 가변 샘플링 기법을 MPPI 알고리즘에 통합하는 방법을 제안한다. 제안 방식은 장애물의 속도 및 엔드이펙터와 장애물 간 거리를 기반으로 위험도를 계산하고, 이를 바탕으로 샘플 수를 하한과 상한 범위 내에서 동적으로 조절한다. 위험도가 높은 구간에서는 탐색 밀도를 증가시켜 안전성을 강화하고, 위험도가 낮은 구간에서는 연산 비용을 절감하여 효율성을 개선하는 것이 핵심 아이디어다. 본 연구에서는 Isaac Sim과 NVIDIA cuRobo 프레임워크를 활용해 다양한 동적 장애물 시나리오에서 제안 기법의 효과를 검증하였으며, 기존 고정 샘플링 방식과 비교하여 충돌 감소와 연산 효율성 향상을 동시에 달성함을 확인하였다. 이러한 결과는 MPPI의 샘플링 전략을 환경 위험도에 따라 동적으로 조정하는 접근이 실제 응용에서 갖는 잠재력을 뒷받침하여 향후 다중 로봇 시스템, 비선형 동역학을 가진 복합 매니퓰레이터, 또는 강화학습 기반 제어와의 융합 연구로도 확장될 수 있다.

Ⅱ. 본론

2.1 위험도 기반 가변 샘플링 MPPI 설계

본 연구에서는 MPPI 제어 알고리즘의 샘플 수를 동적으로 조절하기 위해, 장애물의 속도와 장애물과 엔드이펙터 간 거리를 기반으로 위험도를 계산하였다.

MPPI는 다변량 가우시안 분포로부터 N개의 제어 입력 시퀀스를 샘플링하고, 각 시퀀스에 대해 누적 비용 L_n을 계산한 후, 비용 기반 가중치 ω_n을 산출한다. 가중치는 다음 식(1)과 같이 정의된다.

$$\omega_n = \frac{\exp\left(-\frac{L_n - L_{\min}}{\lambda(L_{\max} - L_{\min})}\right)}{\sum_{m=1}^{N} \exp\left(-\frac{L_n - L_{\min}}{\lambda(L_{\max} - L_{\min})}\right)} \quad (1)$$

최종 제어 입력(μ_t)은 샘플들(입력$u_{n,t}$)의 가중 평균 $\mu_t = \sum_{n=1}^{N} \omega_n u_{n,t}$으로 계산된다.

위와 같은 MPPI 구조에서, 본 연구는 위험도 기반 가변 샘플링 정책을 설계하였다. 장애물 속도는 현재 속도를 최대 속도로 나누어 정규화하였다. 또한 거리에 대해서는 가까울수록 위험이 급격히 증가하는 특성을 부드럽게 표현하기 위해 지수 감쇠 함수를 사용하였다. 이를 통해 특이점 없이 안정적인 제어 및 최적화가 가능하며, 거리 정규화는 exp(-d/D0)와 같이 정의된다. 여기서 d는 end-effector와 장애물 간 유클리드 거리, D_0는 위험이 유의미하게 작용하는 거리 범위를 나타내는 거리 척도이다. d = 0일 때 최대 위험을 나타내며, 거리가 멀어질수록 위험도는 0에 빠르게 수렴한다. 최종 위험도 R은 속도 정규값과 거리 정규값을 가중합하여 계산

하였으며, 비용 함수의 가중치는 속도 0.6, 가속도 0.4로 설정했다[3]. 계산된 위험도는 MPPI의 가변 샘플 수 결정에 사용되었으며, 샘플 수는 최소 400개에서 최대 1000개 사이에서 변하도록 설정하였다. 이는 샘플 수가 1000개일 때 샘플링된 경로를 통해 최종 경로를 계산하는데 소요되는 시간이 약 11ms이며, Isaac Sim의 시뮬레이션 주기(0.016초) 내에 처리 가능한 범위이다.

2.2 장애물 회피 환경 구성

실험 환경에서 사용된 장애물은 한 변의 길이가 0.1m인 정육면체 형태로 설정하였다. 장애물의 이동 속도는 가변적으로 구성하였으며, 최저 속도를 0.2m/s, 최고 속도를 0.5m/s로 설정하고, 주기 5초의 사인파 형태로 변화하도록 하였다. 이를 통해 장애물이 일정한 패턴을 가지면서도 시간에 따라 속도가 지속적으로 변하는 동적 환경을 구현하였다.

장애물의 위치는 로봇의 작업 공간 한계를 고려하여, x축과 z축의 좌표를 각각 0.5m로 고정하고, y축 좌표를 -0.5m에서 0.5m 사이를 왕복 이동하도록 설정하였다. 이러한 설정은 장애물이 로봇의 작업 경로를 가로지르도록 하여 회피 동작을 유도한다.

목표 위치는 로봇이 다양한 방향으로 움직이도록 무작위로 부여하였다. z축 좌표는 0.5m로 고정하고, x축 좌표는 0.5m에서 0.6m 범위 내에서, y축 좌표는 -0.5m에서 0.5m 범위 내에서 무작위로 선택되도록 하였다. 이를 통해 실험에서 로봇이 서로 다른 시작 목표 조합에 대해 다양한 회피 경로를 생성하도록 하였다.

2.3 시뮬레이션

실험은 NVIDIA Isaac Sim 환경에서 진행하였다. 로봇 매니퓰레이터가 무작위로 생성되는 목표 지점에 도달하는 과정을 하나의 에피소드를 정의하였다. 로봇이 목표 지점에 도달하면 새로운 목표 지점이 무작위로 생성되며, 다음 에피소드가 시작되도록 구성하였다. 반대로 로봇이 목표 지점으로 이동하는 과정에서 장애물과 충돌이 발생할 경우, 해당 에피소드는 즉시 종료되고 로봇은 초기 자세로 복귀한 뒤 새로운 목표 지점으로 이동을 시작하여 다음 에피소드가 진행되도록 하였다. Fig. 1은 로봇 매니퓰레이터가 장애물을 회피하여 목표 지점으로 이동하기 위해 MPPI 알고리즘을 통해 경로를 생성하고 이를 시각화한 예시이다. 각 스텝에서 Fig. 1과 유사한 경로가 생성되며, 장애물의 속도와 장애물과의 거리와 같은 요소에 따라 샘플 수가 변화함에 따라 경로의 수도 달라진다. 이를 통해 제안 기법이 상황에 따라 적응적으로 샘플 수를 조정하여 안정적인 경로 생성을 수행함을 확인할 수 있다.

2.4 실험 결과

고정 샘플링(400개) 환경과 가변 샘플링 환경에서 각각 1,500 에피소드를 수행하였으며, 각 에피소드에서 발생한 충돌 횟수를 측정하여 에피소드

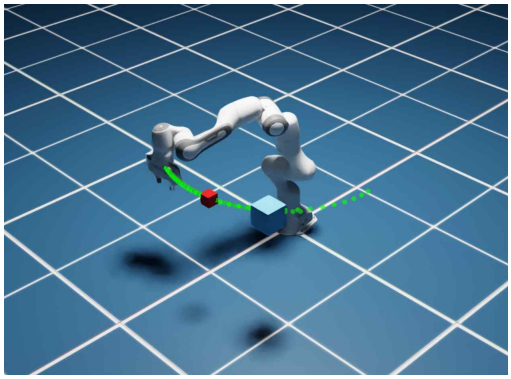


Fig.1 Trajectory planning and execution to reach a target with obstacle avoidance

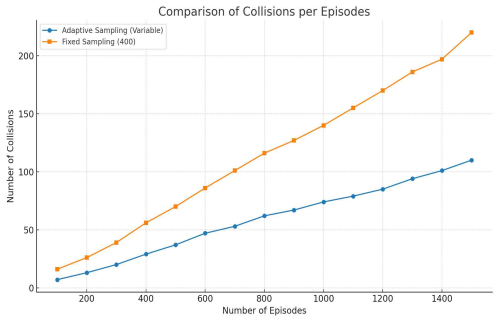


Fig.2 Comparison of collision counts between fixed sampling and adaptive sampling

수가 100개 증가할 때마다 기록하였다. 실험 결과, Fig. 2에서 확인할 수 있듯이 고정 샘플링 환경은 1,500 에피소드 동안 총 220회의 충돌이 발생하여 약 85%의 성공률을 보였다. 반면, 가변 샘플링 환경에서는 동일 조건에서 110회의 충돌만 발생하여 약 93%의 성공률을 기록하였다. 특히 고정 샘플링 환경은 가변 샘플링 환경에 비해 약 2배 높은 충돌 위험을 나타냈다. 이러한 결과를 통해, 장애물의 속도와 엔드이펙터와의 거리를 반영한 가변 샘플링 기법이 기존 고정 샘플링 방식보다 동적 장애물 회피 성능을 효과적으로 향상시킴을 확인할 수 있었다.

Ⅲ. 결론

본 연구에서는 MPPI 제어 알고리즘을 기반으로, 장애물의 속도와 엔드이펙터와의 거리를 고려한 위험도 기반 가변 샘플링 기법을 도입하여, 동적 장애물 환경에서의 충돌 감소와 경로 수행 안정성 향상을 확인하였다. 제안 방식은 위험 구간에서 탐색 밀도를 높여 안전성을 높이고, 비위험 구간에서는 연산 예산을 절감함으로써 전반적인 효율을 확보하였다. 실제 시뮬레이션 환경에서 고정 샘플링 대비 약 2배 이상의 충돌 위험 감소 효과를 보임으로써, 동적 장애물 환경에서의 회피 성능과 안정성을 동시에 확보할 수 있음을 확인하였다.

한편, 본 연구는 파라미터 설정의 수동 튜닝과 시뮬레이션 중심 평가라는 한계를 가진다. 향후 연구에서는 실제 로봇 실험을 통한 검증, 위험도 추정의 학습 기반 고도화 및 계산 비용을 고려한 샘플링 스케줄 최적화를 통해 제안 기법의 적용 가능성과 성능을 더욱 확장하고자 한다.

ACKNOWLEDGMENT

이 논문(또는 저서, 특허)은(는) 정부(과학기술정보통신부)의 재원으로 과학기술사업화진흥원의 지원을 받아 수행된 연구임('학연협력플랫폼구축 시범사업' RS-2023-00304776, RS-2023-00304695).

참 고 문 헌

[1] Liu, L., Zhang, H., Chen, Y., "A survey on manipulator motion planning in dynamic environments," Robotics and Autonomous Systems, vol. 173, 104521, 2024.

[2] Patrick, M., and Bakolas, E., "Model Predictive Path Integral Control: A Review and Future Directions," Annual Reviews in Control, vol. 58, pp. 129 - 142, 2024.

[3] Liu, Y., Wu, Y., and Zhao, D., "Driving Safety Field Theory: Dynamic Risk Assessment for Autonomous Driving," IEEE Transactions on Intelligent Transportation Systems, vol. 24, no. 3, pp. 2543 - 2555, Mar. 2023.

다중 호흡 신호 및 딥러닝 기반 비침습 운동 유발 지방 대사 평가 알고리즘

이준영 [1], 이병주 [1), 2)], 노효웅 [1)], 한영기 [3)], 이대식 [1) *]
1) 한국전자통신연구원, 2) KAIST, 3) 경북대학교

dalee@etri.re.kr

A Study on the Deep Learning-Based Noninvasive Algorithm for Evaluating Exercise-Induced Fat Metabolism Using Multimodal Respiratory Signals

Lee Junyeong[1)], Lee Byeongju[1), 2)], Noh Hyowoong[1)], Hahn Young Ki[3)], Lee Dae-Sik[1),*]
Electronics and Telecommunications Research Institute, Chosun Univ., Korea
Advanced Institute of Science and Technology, Kyungpook National Univ.

요 약

본 논문은 딥러닝과 호흡신호 내 다중 휘발성 유기 화합물(VOC) 기반 운동 유발 지방대사 비침습·실시간 추정 알고리즘을 제안한다. 호흡신호에서 측정된 다차원 시계열 호기신호를 분석하기 위해 1D-CNN 알고리즘을 개발했다. 개발한 알고리즘을 활용하여 호기신호로부터 혈중 BOHB 를 정량 예측했다. 30 건의 임상 데이터로부터 추정된 BOHB 와 혈중 BOHB 비교를 통해 테스트데이터로부터 R=0.86 수준의 높은 상관관계를 달성했으며, 이를 통해 제안한 알고리즘의 개인 맞춤 운동 최적화·대사 모니터링 활용성을 제시한다.

Ⅰ. 서 론

비만과 대사질환은 전 세계적으로 증가세이며, 임상 및 생활 현장에서 운동 중 지방대사 상태를 지속적이고 비침습적으로 파악하려는 요구가 커지고 있다. 기존의 대사평가(예: 혈액 채취, 마스크 기반 가스 분석)는 침습성·장비 부피·비용 문제로 일상 적용이 어렵다[1].

지방 산화가 증가하면 간에서 생성되는 케톤체(특히 β-하이드록시부티르산, BOHB)가 상승하고, 호기에서는 아세톤 등 VOC 가 증가한다[2,3]. 이들 지표는 서로 연동되어 운동 강도와 대사 전환(지방 연소) 정보를 제공한다. 다만, 구강/상부 기도 공기에는 외부 혼입·타액·비강 영향이 남아 대사 신호가 희석될 수 있는 반면, 폐포 호기(end-tidal)는 심부 폐포에서 교환된 대사 부산물을 더 직접적으로 반영한다[3].

본 연구는 다차원 호흡 신로를 분석할 수 있는 1 차원 합성곱신경망(1D-CNN) 기반 비침습 대사 분석 알고리즘을 제안한다. 핵심 아이디어는 다중 호흡신호(VOC 패턴)를 융합해 혈중 BOHB 를 실시간 근사하고, 폐포 호기의 유효성을 실험적으로 검증함으로써 운동 유발 지방대사 평가의 정확도와 실용성을 높이는 것이다.

Ⅱ. 본론

비만과 대사질환의 증가로, 일상 환경에서 운동 중 지방대사 상태를 비침습적으로 모니터링하려는 수요가 커지고 있다. 기존 혈액 채혈이나 메타볼릭 카트 기반 측정은 정확하지만 침습적이고 장비 부피·비용·착용 부담 때문에 개인 맞춤형 관리에 제약이 있다. 이에 본 연구에서는 호기(VOC) 신호를 활용한 경량 및 저비용 진단이 가능한 시스템 구현을 목표로 하고, 이를 위한 1-D CNN 알고리즘 기반 실시간 대사 모니터링 시스템을 구현한다.

우선 호기 분석에 의한 지방대사 모니터링 가능성에 대한 학술적 근거로는 지방 대사 시 발생하는 케톤체 부산물 중 하나인 acetone 을 들 수 있다. 지방 산화가 항진되면 간에서 케톤체(특히 β-하이드록시부티르산; BOHB) 생성이 증가하고, 호기에서는 아세톤 등 VOC 농도가 상승한다. BOHB(혈중)와 아세톤(호기)은 운동 유발 지방 연소(exercise-induced fat metabolism)의 대리지표로 상호 연동되며, 이 관계를 이용하면 호기 분석만으로도 혈중 대사 상태를 추정할 수 있다.

호기 신호는 지난 분당서울대병원과의 협업을 통한 연구개발 및 임상시험에서 개발된 호기 측정 시스템을 통해 확보한 센서 데이터를 활용한다. 해당 신호는 30 명의 임상 대상자로부터 채집된 150 개 이상 시료로부터 측정된 다차원 시계열 데이터 그룹이다. 해당 신호 측정에 활용된 호기 측정 시스템의 주요 특징으로는 측정 신호에서 발생할 수 있는 노이즈 성분을 최소화하여 측정 결과 정확도를 높이기 위해 채취 시료를 구강 호흡과 폐포호흡으로 구분하는 기술이 적용되었다. 또한 해당 하드웨어는 VOC 센서 어레이로 호기 내 아세톤을 포함한 다성분 패턴을 수집한다. 구강·폐포 호기를 밸브/흐름 제어로 분리하고, 실시간

신호 전처리(정규화·기저선 보정·드리프트 보정)를 수행한다. 이렇게 얻은 다중 호흡 신호는 이후 딥러닝 모델의 입력으로 사용된다.

센서 시계열 패턴을 분석하기 위해 딥러닝 알고리즘이 설계 및 시스템에 적용되었다. 해당 알고리즘은 측정된 신호로부터 혈중 BOHB 를 추정하기 위한 추정 알고리즘으로 설계되었다. 개발된 1-D CNN 추정 알고리즘은 다차원 시계열 데이터를 입력받아 연산 후 혈중 BOHB 농도를 연속 값으로 회귀 예측한다. 모델은 슬라이딩 윈도 기반 입력을 사용하고, 과적합 방지를 위해 정규화·조기 종료 등 규제를 적용했다. 목표는 저지연 실시간 추정으로, 추후 온디바이스 경량화까지 염두에 두었다.

개발된 알고리즘의 신뢰도를 평가하기 위해 임상 데이터 분석을 수행했다. 임상 데이터는 각 피검자 별 운동 수행 시간에 따른 5 개의 호기 시료를 측정한 다차원 시계열 데이터 그룹들로 구성되어 총 150 개의 데이터가 준비되었고, 이 중 측정 과정에서의 하드웨어 오류 등으로 인해 변형이 발생한 일부 데이터는 분석에서 제외되었다. 전체 데이터 중 80%를 training data 로, 20%를 test 데이터로 설정하고, 학습 및 데이터 추정을 수행하였다. 최종 학습된 1-D CNN 추정모델을 통한 테스트 데이터 분석 결과 혈액 분석을 통해 확보된 BOHB 수치와의 상관관계는 R=0.86 으로 매우 높은 수치가 확인되었다.

Ⅲ. 결론

제안한 호기 측정 시스템 및 분석 알고리즘의 호기 분석 결과는 다중 호흡신호로 혈중 BOHB 를 높은 상관관계로 추정하여 (학습데이터 R=0.96, 시험데이터 R=0.86), 높은 복잡성을 가지는 호기 데이터 분석을 위한 제안한 1-D CNN 알고리즘의 바이오마커 검출 알고리즘으로의 우수성을 검증한다. 실험결과는 또한 비침습 인체대사 모니터링을 위한 도구로써 폐포 호기 데이터의 우수성을 실증했다. 이는 운동 유발 지방 연소의 실시간 모니터링과 개인 맞춤 운동 코칭, 그리고 대사 건강 추적으로의 적용 가능성을 뒷받침한다. 한편, 표본 규모 확대, 외부/다기관 검증, 장기간 추적이 향후 과제이며, CO_2 (호기말)·호흡량 등 추가 생리 신호 통합과 모델 경량화는 실제 착용형·가정용 구현을 가속할 것이다.

.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부(MSIT, 한국)의 재원으로 학연협력플랫폼구축시범사업 (Scale-up Collaboration by KNU&ETRI Program)의 지원을 받아 수행된 연구임 (RS-2023-00304776)

참 고 문 헌

[1] K. E., Jones, et al., “Breath Acetone Correlates With Capillary β -hydroxybutyrate in Type 1 Diabetes”, *J. of Diabetes Sci. Tech.*, 19322968251334640 (2025)

[2] K. Königstein, et al., “Breath acetone change during aerobic exercise is moderated by cardiorespiratory fitness” *J. of breath research*, 15(1), 016006 (2020)

[3] F. Bovey, et al. "Breath acetone as a marker of energy balance: an exploratory study in healthy humans." *Nutrition & diabetes*, 8(1), 50 (2018)

[4] W. Miekisch, et al. "Impact of sampling procedures on the results of breath analysis." *J. of breath research*, 2(2), 026007 (2008)

[5] A. Di Gilio, et al. "Breath analysis: Comparison among methodological approaches for breath sampling." *Molecules*, 25(24), 5823 (2020)

MobileNetV2를 활용한 궤양성 대장염 중증도 예측: 온디바이스 적용을 위한 경량화 접근

김범휘, 우동원, 정성문*

한국전자통신연구원, 경북대학교병원 의료인공지능연구센터 *경북대학교 의료정보학교실

bumhwi.kim@etri.re.kr, woodongwon23@gmail.com, *jeongsm00@gmail.com

Ulcerative Colitis Severity Prediction Using MobileNetV2: A Lightweight Approach for On-Device Deployment

Bumhwi Kim, Dongwon Woo, Sungmoon Jeong*

Medical IT Convergence Section, Electronics and Telecommunications Research Institute(ETRI)
Research Center for Artificial Intelligence in Medicine, Kyungpook National University Hospital,
*Department of Medical Informatics, School of medicine, Kyungpook National University

요 약

궤양성 대장염(Ulcerative Colitis, UC)의 중증도 평가는 일반적으로 내시경 검사를 기반으로 한 UCEIS(Ulcerative Colitis Endoscopic Index of Severity)를 통해 이루어진다. 그러나 내시경 평가는 침습적이며, 시술 부담과 비용이 수반된다. 선행 연구에서 대변 이미지를 활용하여 UC 중증도를 분류하는 Vision Transformer(ViT) 기반 모델을 개발하였으며, UCEIS 점수의 이진 분류를 통해 유의미한 성능을 보고하였다. 다만, 해당 클래스 구분은 휴리스틱에 기반하였으며, 실제 임상적 분류 체계와 일치하지 않는 한계가 있었다. 본 연구에서는 기존 접근을 두 가지 측면에서 확장하였다. 첫째, 자원 제약 환경에서의 온디바이스 적용 가능성을 확보하기 위해 ViT 대신 경량 합성곱 신경망 구조인 MobileNetV2를 도입하였으며, 이 과정에서 모델 경량화에 따른 성능 저하 가능성을 인지하였다. 둘째, 임상적으로 보다 타당한 분류를 위해 UCEIS 점수를 0-1, 2-4, 5-8의 3클래스로 재구성하였다. 데이터셋의 클래스 불균형 문제를 해결하기 위해 다양한 데이터 증강 기법을 적용하여 일반화 성능 향상을 도모하였다. 본 논문에서는 해당 방법론적 절차와 설계 배경을 기술하고, 비침습적 대변 이미지를 이용한 UC 중증도 예측 모델의 임상적·실용적 의의를 논의한다.

Ⅰ. 서 론

궤양성 대장염(Ulcerative Colitis, UC)은 만성 염증성 장질환으로, 증상의 중증도 평가가 치료 계획 수립에 중요한 역할을 한다. UCEIS(Ulcerative Colitis Endoscopic Index of Severity)는 UC 중증도를 정량화하는 표준 지표이지만, 내시경 검사를 기반으로 하므로 환자에게 침습적이며 비용과 시간이 소요된다. 최근에는 비침습적 방법으로 환자의 상태를 추정하려는 시도가 증가하고 있으며, 특히 대변 이미지를 활용한 인공지능(AI) 기반 접근이 주목받고 있다.

기존 선행 연구에서는 대변 이미지를 기반으로 Vision Transformer(ViT) 모델을 학습하여[1], UCEIS 점수를 0~2와 3~8로 구분하는 이진 분류를 수행하였다[2]. 해당 접근은 혁신적인 아이디어를 기반으로 임상적으로 활용 가능성을 보여주었으나, 다음과 같은 한계가 있었다. 먼저 클래스 구분이 임상 기준이 아닌 휴리스틱에 기반해 설정되어 실제 UCEIS 점수 체계와 차이가 있었다. 또한 ViT 모델은 연산량과 메모리 요구량이 커서 온디바이스 환경에서 실시간 추론을 수행하기 어렵다.

본 연구에서는 이러한 한계를 보완하기 위해 경량 모델인 MobileNetV2를 적용하여 온디바이스 추론 가능성을 확보하고, 임상 기준에 보다 부합하는 다중 클래스 설정을 도입하며, 다양한 데이터 증강 기법을 적용하여 불균형 문제를 완화하였다. 이를 통해 단순한 분류 성능 개선을 넘어, 실제 임상 적용 가능성과 해석 가능성을 동시에 고려한 모델 개발을 목표로 하였다.

Ⅱ. 실험 방법

본 연구에 사용된 데이터셋은 궤양성 대장염 환자의 대변 이미지를 기반으로 하며, 각 이미지에는 내시경 평가를 통해 측정된 UCEIS 점수가 라벨로 부여되었다. 원래 임상적 기준에 따른 구분은 0-1, 2-4, 5-6, 7-8로 총 4단계이지만, 표본 수가 제한적이고 특히 고점수 구간(7-8)의 데이터가 매우 희소하여 안정적인 학습이 어려웠다. 이에 따라 0-1을 경증, 2-4를 중등도, 5-8을 중증으로 묶어 3클래스 분류 실험을 진행하였다.

모델 경량화를 위해 MobileNetV2를 활용하였다. MobileNetV2는 inverted residual과 linear bottleneck 구조를 활용하여 연산 효율성을 극대화하며, depthwise separable convolution을 통해 파라미터 수를 크게 줄이면서도 적절한 성능을 유지할 수 있다[3].

본 연구에서는 224×224 이미지를 입력으로 하고, ImageNet 사전 학습 가중치를 초기값으로 사용하였다. 분류기는 Global Average Pooling과 Dropout을 거쳐 Fully Connected Layer와 Softmax로 구성하였으며, Adam 옵티마이저를 사용하고 초기 학습률은 1e-4로 설정하였으며 StepLR 스케줄러를 적용하였다.

데이터 불균형 문제를 완화하기 위해 다양한 데이터 증강 기법을 적용하였다. 기본 변환으로는 수평 및 수직 반전, ±15° 범위 회전, 랜덤 크롭, 밝기·대비·채도 조절을 사용하였다. 여기에 CutMix[4], MixUp[5]과 같은 고급 변환을 적용하여 클래스 간 특징을 혼합하고, Random Erasing을 통해 배경과 형태 변이에 대한 강건성을 강화하였다. 특히 데이터가 적은 클래스에는 증강 횟수를 늘려 클래스별 표본 수를 균형에 가깝게 조정하였다.

성능 평가는 Accuracy, F1-score, 클래스별 민감도와 특이도, 그리고 AUC를 활용하였다. 분석 항목은 모델 경량화에 따른 성능 변화, 다양한 증강 기법 적용 전후의 성능 향상 정도를 포함하였다.

Ⅲ. 실험 결과

표 1. 모델 별 이진 분류 결과

Model	정확도	민감도	특이도	AUC
ViT	0.932	0.8864	0.9661	0.943
MobileNetV2	0.8932	0.8636	0.9153	0.911

표 1은 ViT와 MobilenetV2의 이진 분류 결과를 나타낸다. MobileNetV2 모델은 전역적 패턴 학습 능력이 우수한 ViT 대비 전반적인 분류 성능은 낮게 나타났지만, 여전히 우수한 성능을 보이는 것을 확인할 수 있었다.

표 2. MobileNetV2 기반 3-class 분류 결과

Model	3-class	3-class with 증강
정확도	0.6125	0.6821

표 2는 MobileNetV2 모델을 활용한 3-class 분류 실험 결과를 나타낸다. 데이터 증강 기법을 적용한 경우, 미적용 대비 약 7% 이상 성능이 향상되었으며, 이를 통해 증강 기법이 클래스 불균형 문제를 일정 부분 완화하는 데 효과적임을 확인하였다.

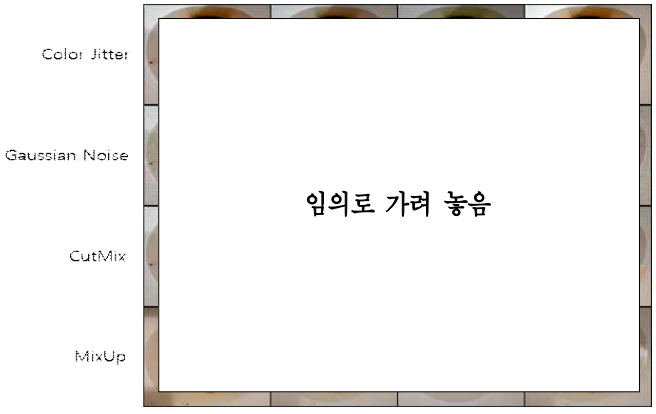


그림 2. 데이터 증강 기법 적용 예시

그림 1은 본 연구에서 적용한 데이터 증강 기법이 실제 영상에 어떻게 반영되는지를 정성적으로 보여준다. 의료 AI 분야에서는 단순한 성능 지표의 향상만으로 모델의 우수성을 평가하기에는 한계가 있다. 성능 개선을 위해 적용한 방법이 어떤 방식으로 영향을 미쳤는지를 시각적으로 해석하는 과정이 중요하다. 이에 본 연구에서는 각 증강 기법이 생성한 영상의 변화를 확인하였으며, 이를 통해 단순한 수치 향상뿐만 아니라 이미지의 시각적 다양성이 확보되어 모델이 보다 풍부한 패턴을 학습할 수 있었음을 확인하였다.

Ⅳ. 결론

본 연구에서는 궤양성 대장염 환자의 대변 이미지를 이용하여 UCEIS 기반의 중증도 분류를 수행하는 모델을 개발하였다. 기존 연구에서 사용된 Vision Transformer(ViT) 모델을 경량화하여 온디바이스 환경에서 구현할 수 있도록 MobileNetV2를 적용하였으며, 임상적 기준에 보다 부합하는 3클래스 분류 설정을 도입하였다. 데이터 불균형 문제를 해결하기 위해 다양한 데이터 증강 기법을 활용하였고, 그 효과를 정량적 지표뿐만 아니라 시각적 해석을 통해 검증하였다.

실험 결과, MobileNetV2는 ViT 대비 전반적인 분류 성능은 다소 낮았으나, 온디바이스 환경에서 구현하기에는 충분한 성능을 보임을 확인하였다. 또한, 임상 기준에 부합하기 위한 3-class 결과도 이진 분류에 비해 성능이 하락되었으나, 다양한 증강 기법의 적용을 통해 클래스 불균형을 완화하고 모델의 일반화 성능을 향상시킬 수 있었다. 특히 의료 AI 분야에서는 단순 성능 지표 외에도 결과에 대한 해석 가능성이 중요하며, 본 연구에서 제시한 시각적 증강 결과 분석은 이러한 관점에서 의미 있는 정보를 제공한다.

향후 연구에서는 데이터셋을 확충하여 임상 기준에 따른 4클래스 분류를 시도하고, 경량 모델의 성능 향상을 위해 지식 증류나 구조 최적화 기법을 적용할 계획이다. 이를 통해 비침습적 대변 이미지 분석 기반의 UC 중증도 예측 모델이 실제 임상 현장에서 보조 진단 도구로 활용될 수 있을 것으로 기대된다.

ACKNOWLEDGMENT

이 논문(또는 저서, 특허)은(는) 정부(과학기술정보통신부)의 재원으로 과학기술사업화진흥원의 지원을 받아 수행된 연구임{'학연협력플랫폼구축 시범사업' RS-2023-00304695(50%), RS-2023-00304776(40%)} 및 한국전자통신연구원 연구운영지원사업 [25ZD1140, 대경권 지역산업 기반 ICT 융합 기술 고도화 지원(의료)](10%)의 지원을 받아 수행된 연구임

참 고 문 헌

[1] Lee, J. W., Woo, D., Kim, K. O., Kim, E. S., Kim, S. K., Lee, H. S., ... & Jeong, S. (2025). Deep learning model using stool pictures for predicting endoscopic mucosal inflammation in patients with ulcerative colitis. Official journal of the American College of Gastroenterology| ACG, 120(1), 213-224

[2] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929

[3] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4510-4520).

[4] Yun, S., Han, D., Oh, S. J., Chun, S., Choe, J., & Yoo, Y. (2019). Cutmix: Regularization strategy to train strong classifiers with localizable features. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 6023-6032).

[5] Zhang, H., Cisse, M., Dauphin, Y. N., & Lopez-Paz, D. (2017). mixup: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412.

추종로봇을 위한 UWB-RGB 센서 융합 기반의 사용자 위치 추정 방법

최낙원, 이준구, 오지용*
한국전자통신연구원

nwchoi@etri.re.kr, leejg01679@etri.re.kr, *jiyongoh@etri.re.kr

Target Person Position Estimation via UWB-Vision Fusion for Robotic Following

Nak-Won Choi, Joon-Goo Lee, Jiyong Oh*
Electronics and Telecommunications Research Institute (ETRI)

요 약

본 논문에서는 추종로봇을 위해 UWB 센서와 RGB 카메라를 융합하여 사용자의 위치를 추정하는 방법을 제안한다. 제안하는 방법은 Jetson Orin Nano 기반 ROS 환경에서 구현하였으며, Stereolabs 사의 ZED 2i 카메라와 Markerfabs 사의 ESP32 UWB 통신 모듈을 탑재한 Husky A200 로봇을 통해 검증하였다. 실험 결과, UWB-RGB 융합 방식은 평균 거리 오차 약 0.1 m, 방향 오차는 0.03 rad 이하로 UWB 단일 센서 대비 향상된 방향 정확도를 보였으며, 모바일 로봇 추종 실험에서도 20 회 전 구간 성공을 달성하였다.

Ⅰ. 서 론

추종로봇은 의료 분야에서 의료진의 부담을 경감하고, 물류 분야에서 작업 효율과 안전성을 향상시키는 등 사회적·산업적 환경에서 인간을 보조하기 위해 설계된다. 이러한 추종로봇은 작업자를 정확하게 인지하는 것이 중요하며, 이를 위해 다양한 센서를 활용한 위치 추정 연구가 진행되고 있다. 기존 연구에서 활용된 LiDAR 및 RGB-D 센서 기반 위치 추정은 비용과 환경 제약의 한계가 있으며[1], UWB(Ultra-Wideband) 센서는 저비용으로 약 0.1 m 수준의 거리 정확도를 제공하지만 앵커 배치에 따라 방향 오차가 발생한다[2]. 한편, RGB-D 카메라는 거리와 방향에 대해서 낮은 오차를 보여주지만, 높은 비용으로 인해 실용성이 제한된다[3]. 이에 본 연구에서는 UWB 의 거리 정확도와 RGB 카메라의 방향 정확도를 결합한 UWB-RGB 센서 기반 위치 추정 방법을 제안한다. 두 센서는 각각 거리와 방향 추정에서 상호보완적 특성을 가지며, 융합 시 거리와 방향 모두 정확도가 향상된다(표 1). 또한 제안한 위치 추정 방법을 실제 모바일 로봇에 적용하여 실환경 주행 실험을 통해 그 성능을 검증한다.

표 1. UWB 와 RGB 센서의 위치 추정 특성 비교

측정 방법	r (거리)	θ (방향)
UWB	$\leq 0.15\,m$	$\leq 0.2\,rad$[2]
RGB	불가	$\leq 0.02\,rad$[3]
UWB-RGB	$\leq 0.11\,m$	$\leq 0.03\,rad$

Ⅱ. 본론

1. 시스템 구성

본 논문에서 제안하는 시스템 구성도는 그림 1 과 같으며, 전체 시스템은 ROS(Robot Operating System) 환경에서 구현한다. Jetson Orin Nano 보드는 ZED 2i RGB 카메라와 DWM1000 UWB 모듈을 USB 로 연결하여 ROS 토픽 통신을 통해 각각 RGB 이미지 토픽과 거리 토픽을 수신한다. 본 논문에서는 향후 GT(Ground Truth)와 추론 결과를 동일한 조건에서 비교하기 위해 ZED 2i 를 사용하며, 제안하는 방법에서는 RGB 정보만 사용한다. 거리 토픽은 사용자가 소지한 UWB 태그와 로봇에 부착된 앵커 모듈 간의 측정한 거리 값을 사용한다. RGB 이미지 토픽은 YOLO(You Only Look Once)v5 모델[5]을 통해 사람을 검출하며 검출에 사용된 픽셀 좌표계 기준의 바운딩 박스 중점 좌표를 추출한다. 사용자 위치 추정은 바운딩 박스의 중점 좌표와 앵커 모듈로 측정한 거리 정보를 사용하여 카메라 좌표계로의 변환을 통해 계산한다. 추정된 좌표는 Clearpath Robotics 의 Husky A200 로봇의 base_link 좌표계로 변환되며, 이를 바탕으로 선각속도를 산출하여 로봇 제어 입력에 사용한다.

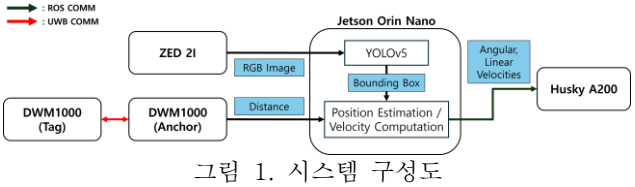


그림 1. 시스템 구성도

2. UWB-RGB 센서 기반 위치 추정

2D 이미지에서 추출된 좌표를 3D 공간에 재투영하기 위해 먼저 카메라 내부 파라미터를 이용하여 픽셀 좌표를 카메라 좌표로 변환한다. 카메라 좌표계에서 대상까지의 깊이 값을 Z라 하면, 카메라 좌표계 기준 추종 대상의 위치 (X, Y)는 식 (2)와 같이 계산된다. 식 (2)에서 f_x, f_y는 초점 거리이고 c_x, c_x는 이미지 중점 좌표이다.

$$X = \frac{u - c_x}{f_y} \cdot Z, Y = \frac{v - c_y}{f_y} \cdot Z \qquad (2)$$

본 논문에서는 (u, v)를 바운딩 박스의 중점좌표로 사용하며, Z를 UWB 센서로 측정된 거리 값으로 사용한다. 카메라 내부 파라미터는 zed-ros2-wrapper 패키지에서 제공된 파라미터를 사용한다[4]. 식 (2)를 통해 얻은 대상의 위치는 Husky A200 로봇의 base_link 좌표계로 변환을 수행한 뒤 최종적인 위치 좌표를 산출한다.

3. UWB-RGB 센서 기반 위치 추정 실험

본 논문에서는 UWB-RGB 센서의 위치 추정 성능을 검증하기 위해 실험을 수행하였다. 실험은 UWB 태그를 소지한 사람을 Husky A200 로봇의 base_link 직교좌표계(전후 X 축, 좌우 Y 축)에서 정의한 6 개 고정 위치에 배치하여 2.2 절에서 기술한 방법에 따라 위치를 추정하였다. 추정 위치 좌표와 기준 좌표는 Husky A200 로봇의 제어 입력에 사용되는 극좌표계 (r, θ)로 변환한 뒤 평균 거리 오차 (Δr)와 방향 오차 $(\Delta\theta)$를 계산하였다. 여기서 $r(m)$은 base_link 의 원점으로부터의 거리를 의미하고 $\theta(rad)$는 base_link 의 $+X$ 축 반시계 방위각이다. UWB 단일 센서 기반 위치 추정 성능과의 비교를 위해 기존 연구[2]에서 3 개 앵커를 이용한 삼변측량 결과를 동일한 극좌표계로 변환하여 사용하였다.

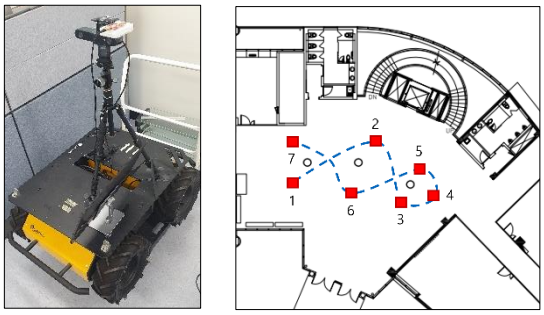
표 2. UWB와 UWB-RGB 평균 거리 및 방향 오차 비교

기준 좌표 (X, Y) (m)	UWB 오차 $\Delta r(m), \Delta\theta(rad)$	UWB-RGB 오차 $\Delta r(m), \Delta\theta(rad)$
(1.0, 0.25)	0.04, 0.21	0.11, 0.01
(1.5, 0.25)	0.04, 0.15	0.03, 0.03
(2.0, 0.25)	0.04, 0.02	0.07, 0.01
(2.5, 0.25)	0.15, 0.03	0.04, 0.02
(3.0, 0.25)	0.05, 0.02	0.06, 0.03
(3.5, 0.25)	0.06, 0.04	0.07, 0.02

표 2 와 같이 두 방법의 평균 거리 오차(Δr)은 모두 약 0.063 m 로 유사하게 나타났다. 반면 UWB-RGB 기반 위치 추정 방법의 평균 방향 오차는 약 0.020 rad 로, UWB 단일 센서(0.078 rad) 대비 약 4 배 감소하여 이를 통해 방향 추정 성능이 향상됨을 확인할 수 있었다.

3. 모바일 로봇 주행 실험

모바일 로봇은 그림 2(a)와 같이 Husky A200 로봇에 ZED 2i 카메라와 UWB 센서를 장착하여, 주행 실험을 수행하였다. 모바일 로봇의 이동 경로는 그림 2(b)에 제시된 바와 같이 한국전자통신연구원 대경권연구본부 1 층 로비에서 설정된 경로를 따르며, UWB 태그를 소지한 자를 추종하도록 20 회 반복하였다. 본 연구에서는 로봇이 전체 경로에서 사용자를 이탈하지 않고 종료 지점에 도달한 경우를 성공으로 정의한다. 실험 결과, 20 회 모두 성공하여 제안한 방법의 안정적인 추종 성능을 입증하였다.



(a) (b)

그림 2. (a) 실험용 모바일 로봇, (b) 로봇 이동 경로

Ⅲ. 결론

본 논문에서는 UWB 센서와 RGB 카메라를 융합하여 사용자 위치를 추정하는 방법을 제안하였으며, 실제 모바일 로봇 플랫폼에서 그 성능을 검증하였다. 실험 결과, UWB 단일 센서 대비 위치 추정 오차가 감소하였으며, 로봇 추종 실험에서도 안정적인 성능을 확인하였다. 이는 위치 추정 정확도의 개선이 실제 로봇 추종 성능 향상으로 이어질 수 있음을 보여주었다. 향후 연구는 RGB-D 카메라 기반 위치 추정 성능과 본 논문에서 제안한 UWB-RGB 센서 기반 위치 추정 성능을 비교할 예정이다.

ACKNOWLEDGMENT

본 논문은 한국전자통신연구원 연구운영지원사업 [25ZD1130, 대경권 지역산업 기반 ICT 융합기술 고도화 지원사업(로봇)] 및 정부(과학기술정보통신부)의 재원으로 과학기술사업화진흥원의 지원을 받아 수행 되었음('학연협력플랫폼구축 시범사업' RS-2023-00304776).

참 고 문 헌

[1] A. Eirale, M. Martini and M. Chiaberge, "Human Following and Guidance by Autonomous Mobile Robots: A Comprehensive Review," IEEE Access, vol. 13, pp. 42214-42253, 2025.

[2] Nak-Won Choi, Joon-Goo Lee and Jiyong, “A Study on UWB-Based User Position Estimation Method for Following Robots,” Proceedings of Symposium of the Korean Institute of communications and Information Sciences, 2025.

[3] Hyun-sik Son, Deok-keun Kim, and Seung-hwan Yang, "Development of Following Algorithm for Human-Following Robots using Deep-Learning Technology and Depth Camera," Journal of Institute of Control, Robotics and Systems, vol. 28, no. 2, pp. 95-101, 2022.

[4] zed-ros2-wrapper: ROS 2 wrapper for the ZED SDK, [online] Available: https://github.com/stereolabs/zed-ros2-wrapper.

[5] R. Khanam and M. Hussain, “What is YOLOv5: A deep look into the internal features of the popular object detector,” 2024, arXiv:2407.20892.

소프트웨어 정의 카메라 통합관제시스템 개발

김진홍, 최윤원, 백장운, 임길택, 김광주
한국전자통신연구원 대경권연구본부
{jinhong, yunwon.choi, jwbaek98, ktl, kwangju}@etri.re.kr

Development of integrated control and management system for a software-defined camera

Jinhong Kim, Yun Won Choi, Jang Woon Baek, Kil-Taek Lim, Kwang-Ju Kim
Electronics and Telecommunications Research Institute Daegu-Gyeongbuk Research Division

요 약

본 논문은 현장에 분산 배치된 다수의 소프트웨어 정의 카메라(SDC)를 원격에서 운영하기 위한 통합관제시스템을 제안하고 구현한다. 개발한 시스템은 Over-the-Air(OTA) 기능과 Web GUI를 바탕으로 단말맵, 단말현황, 서비스 관리 메뉴를 통해 원격에서 다수의 SDC의 상태 모니터링, 응용 서비스와 AI학습모델의 버전 관리 및 배포 등을 수행한다. 그리고 이러한 기능을 바탕으로 현장에 설치 및 운영중인 SDC에 물리적인 접근 없이도 다수의 SDC를 비용 효율적으로 관리할 수 있도록 시스템을 개발하였다.

Ⅰ. 서 론

최근 기존 CCTV에 엣지AI단말이 결합된 형태인 소프트웨어 정의 카메라(Software Defined Camera, 이하 SDC)가 개발되어 활발히 이용되고 있다. SDC는 수집한 데이터를 자체적으로 처리 및 분석할 수 있는 능력 보유하고 있어 외부 서버나 클라우드와 연동하지 않더라도 인공지능 기술이 포함된 다양한 응용 서비스를 제공할 수 있는 장점이 있다. 그리고 SDC는 이러한 장점을 바탕으로 기존에 설치 및 운영되던 CCTV를 대체하면서 빠르게 확산하고 있으며, 특히 실시간성이 요구되는 재난 및 안전 분야에 적극적으로 설치 및 운영되고 있다[1][2].

현장에서 설치 및 운영되는 SDC가 늘어날수록 다수의 SDC를 효율적으로 관리하기 위한 방안이 중요한 이슈로 떠오르고 있다. 특히, SDC에서 운영하는 인공지능 기반의 응용 서비스는 인공지능 기반 서비스의 성능을 결정하는 AI학습모델의 지속적인 최신화가 중요하다. 따라서 SDC 자체에 대한 유지보수뿐만 아니라 SDC에서 운영하는 응용 서비스와 응용 서비스에서 이용하는 AI학습모델의 최신화를 포함한 체계적인 관리 방안의 수립이 필요하다. 하지만 현장에서 설치 및 운영중인 SDC를 관리자가 직접 방문해 내부에서 동작하는 응용 서비스와 AI학습모델 등을 수정 및 갱신하는 기존의 방식은 비용과 시간 등의 측면에서 매우 비효율적인 문제가 있다.

본 논문에서는 현장에서 설치 및 운영되는 다수의 SDC를 효율적으로 운용하기 위해 설계 및 개발한 소프트웨어 정의 카메라 통합관제시스템에 대해 설명한다.

Ⅱ. 시스템 구성

본 논문에서 개발하는 소프트웨어 정의 카메라 통합관제시스템은 그림 1과 같이 현장에서 설치 및 운영되는 다수의 SDC와 서버실과 같은 원격지의 실내에서 운영되는 통합관제시스템으로 구성된다. SDC는 자신이 촬영한 영상을 처리 및 분석하고 그 결과를 통합관제시스템으로 전송하며, 통합관제시스템을 통해 설치된 다양한 응용 서비스와 AI학습모델을 운영한다.

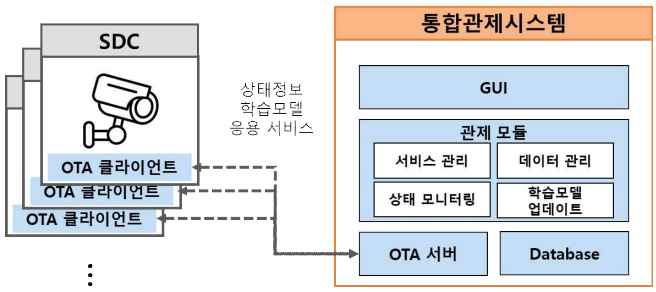


그림 1. 개발 시스템의 구성

SDC에서 인공지능을 기반으로 제공하는 응용 서비스는 교통량 계수, 차량 속도 검지, 산불 감지 등 다양한 응용이 있을 수 있으며, 사용자의 요구 및 SDC의 운용 목적에 따라 하나 또는 복수 개의 응용 서비스가 개별 SDC에 설치 및 운용될 수 있다. AI학습모델은 응용 서비스에서 인공지능에 기반한 기능을 제공할 수 있도록 하는 학습모델을 의미하며, 응용 서비스의 종류에 따라 필요로 하는 AI학습모델의 종류와 수는 서로 상이할 수 있다. OTA클라이언트는 통합관제시스템과의 데이터 교환을 위한 통신을 수행함으로써 통합관제시스템이 SDC를 관제할 수 있도록 한다.

통합관제시스템은 다수의 SDC를 관리하기 위한 시스템으로 내부에 구현된 관제 모듈을 바탕으로 SDC의 상태 모니터링, SDC에서 운용하는 응용 서비스와 응용 서비스에서 이용하는 AI학습모델의 관리 등을 수행한다. 그리고 상기 동작은 OTA 서버 모듈을 통한 SDC와의 데이터 송수신에 의해 수행된다.

Ⅲ. 시스템 개발

개발한 통합관제시스템은 다음과 같은 다양한 기능을 제공하도록 설계 및 개발되었다.

- 통합관제서버 사용자 관리
- 엣지AI단말 상태정보 관리
- 엣지AI단말에서 운용중인 서비스 애플리케이션 관리 및 제어

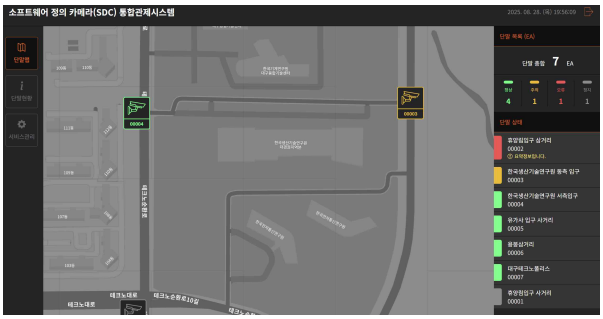


그림 2. 통합관제시스템 단말맵 메뉴 화면

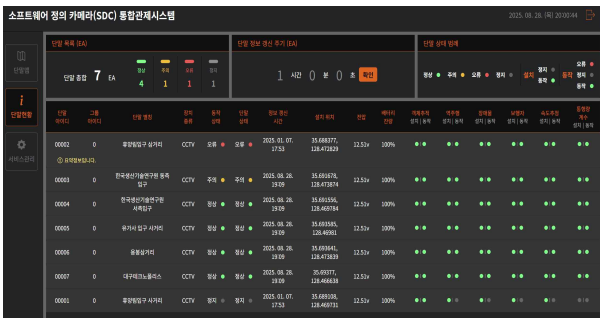


그림 3. 통합관제시스템 단말현황 메뉴 화면

- 엣지AI단말에서 운용중인 서비스 애플리케이션의 학습모델 관리 및 제어
- 엣지AI단말에서 운용할 수 있는 전체 서비스 애플리케이션 및 학습모델 관리
- 상기 기능에 대한 Web기반의 사용자 Graphic User Interface(GUI) 제공

먼저 통합관제시스템에 로그인하여 메인화면으로 들어오면 그림2와 같이 SDC의 상태와 설치 위치를 지도를 통해 확인할 수 있는 “단말맵” 메뉴 화면을 볼 수 있다. 단말맵 메뉴에서는 설치 및 등록된 전체 SDC의 목록과 SDC별 동작 상태 및 동작 상태별 SDC의 수를 색과 숫자로 표시한 정보를 확인할 수 있다. 그리고 등록 및 설치된 전체 SDC의 목록에서 확인하고 싶은 SDC를 선택하거나 지도에 표시된 SDC의 아이콘을 선택하면 선택된 SDC의 세부 상태정보와 설치된 응용 서비스에 대한 정보를 확인할 수 있다.

그리고 통합관제시스템은 “단말현황” 메뉴를 통해 현재 시스템에 등록된 모든 SDC의 정보를 한 눈에 확인할 수 있는 기능을 제공한다. 그림 3과 같이 단말현황 메뉴에서는 등록 및 설치된 전체 SDC의 목록과 SDC별 단말 아이디, 그룹 아이디, 단말 별칭, 종류, 상태, 정보 갱신 시각, 설치 위치 등 SDC 자체에 대한 정보 뿐만 아니라, SDC에 설치 및 운용중인 응용 서비스의 설치 및 동작 상태를 함께 확인할 수 있다.

그림 4와 같이 “서비스 관리” 메뉴에서는 개별 SDC에 설치 및 운용중인 응용 서비스와 응용 서비스에서 이용하는 AI학습모델의 관리를 수행할 수 있는 기능을 제공한다. 각 응용 서비스들은 개별 탭으로 구분되며, 각 탭에서는 해당 응용 서비스의 버전 확인 및 관리(업그레이드), AI학습모델의 버전 확인 및 관리(업데이트 및 재학습) 기능을 제공한다. 현재 통합관제시스템에 등록되어 관리되고 있는 응용 서비스는 객체추적, 역주행, 장애물 감지, 보행자 감지, 차량 속도추정 등이며, 추가적인 서비스 애플리케이션의 개발이 완료되면 통합관제시스템에 추가하여 관리할 수 있다. 또한, 각 응용 서비스에서 이용하는 AI학습모델을 최신 성능으로 유지하기 위한 AI학습모델의 최신화를 손쉽게 할 수 있도록 서비스관리 메뉴에서 업데이트 버튼을 선택하는 것만으로도 SDC에서 동작중인 응용 서비

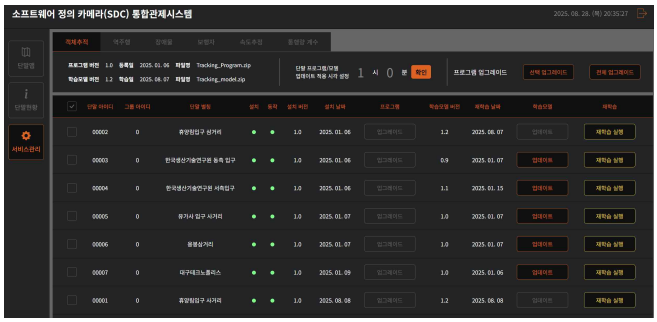


그림 4. 통합관제시스템 서비스 관리 메뉴 화면

스의 AI학습모델을 최신 버전으로 업데이트 할 수 있는 기능을 제공한다. 다만, 이를 위해서는 우선적으로 성능을 개선시킨 최신 버전의 AI학습모델을 통합관제시스템에 등록하는 작업이 선행되어야 한다. 그리고 AI학습모델이 등록되면 해당 AI학습모델을 이용하는 SDC의 학습모델 항목의 업데이트 버튼이 활성화되며, 활성화된 업데이트 버튼을 선택하면 해당 SDC에 최신버전의 AI학습모델을 설치하는 동작이 수행된다. 다만, 실질적인 AI학습모델의 설치는 업데이트 버튼을 누르는 즉시 시작되는 것은 아니며, 관리자가 이전에 “프로그램/모델 업데이트 적용 시각” 메뉴에 설정된 시각에 맞춰 업데이트가 진행된다.

SDC의 상태정보 전송은 미리 정의된 메시지 구조에 맞춰 상태와 관련된 여러 정보를 통합관제시스템으로 보내는 동작을 수행하는 것이다. AI학습모델 업데이트와 응용 서비스의 설치 및 업데이트 기능을 구현하기 위해 통합관제시스템과 SDC 간 AI학습모델 및 응용 서비스와 관련된 파일을 안전하게 송수신할 수 있도록 리눅스 운영체제에서 자체적으로 제공하는 기능 중 하나인 Secure Copy(SCP) 기능을 활용하였다.

Ⅳ. 결론

본 논문에서는 현장에서 설치 및 운영되는 다수의 SDC 운영 효율성을 높이고 유지보수 비용을 절감하기 위해 소프트웨어 정의 카메라 통합관제시스템을 설계 및 구현하였다. 개발한 시스템은 SDC의 동작 상태를 실시간 모니터링할 뿐만 아니라 SDC에서 운용되는 응용 서비스와 AI학습모델을 원격지에 위치한 서버에서 간편하게 최신 버전을 배포 및 설치할 수 있도록 함으로써 항상 최상의 성능과 서비스를 제공할 수 있도록 하였다.

ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 과학기술사업화진흥원의 지원과 한국전자통신연구원 연구운영지원사업의 일환으로 수행된 연구임 [RS-2023-00304776, 학연협력플랫폼구축 시범사업, 25ZD1120, 대경권 지역산업 기반 ICT 융합기술 고도화 지원사업]

참 고 문 헌

[1] S. Jayasuriya, O. Iqbal, V. Kodukula, V. Torres, R. Likamwa, and A. Spanias, “Software-Defined Imaging: A Survey,” Proceedings of the IEEE, 111(5), pp. 445-464, 2023

[2] Sundar, V., Ardelean, A., Swedish, T., Bruschini, C., Charbon, E., & Gupta, M. “Sodacam: Software-defined cameras via single-photon imaging”, In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 8165-8176)

Sub-data Reloader for Weather Removal

Jung Hwan Kim[1], Yun-su Chung[2], and Dong Seog Han[1,*]
[1]School of Electronic and Electrical Engineering, Kyungpook National University,
Republic of Korea
[2]Electronics and Telecommunications Research Institute,
Republic of Korea
jkim267@knu.ac.kr, yoonsu@etri.re.kr, dshan@knu.ac.kr*

Abstract

The weather effects degrade the image screens and hinder the view of object detection. Ultimately, it could cause a safety issue for autonomous vehicles and result in malfunctioning object detection. In fact, developing weather removal and image reconstruction is an essential component for object detection. Most state-of-the-art generative architectures are still challenging in terms of reliable performance for image cleaning under different weather conditions. Thus, in this paper, we propose to use the SFHformer architecture with KITTI with weather effects. Then, we also developed the sub-data reloader for speeding up training. After applying the sub-data reloader, we managed to obtain relatively close performance of SSIM and PSNR as 97.43% and 36.11, respectively, as the training time per epoch is 1039 seconds, while the training with the whole dataset per epoch has the PSNR as 38.81 and the SSIM as 98.23% which took 2782 seconds.

I. Introduction

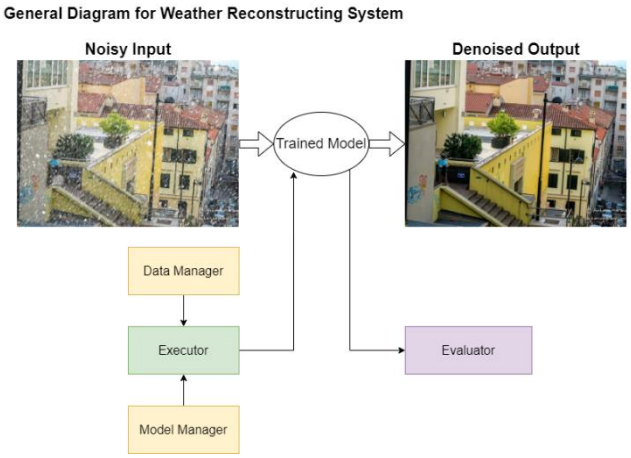


Fig. 1 The built foundation of our weather removal system.

Any vehicle during bad weather could face dangers while hindering the driver's and camera imaging views. Autonomous driving transportation has the same problem with object detection during bad weather, as they occlude potential danger scenes through the windshield or camera lens. Thus, predicting a clean image without weather is vital for protecting drivers and passengers and enhancing the reliability of object detection. From Fig. 1, we created the foundation for removing the weather effects, such as rain, snow, fog, and raindrops on the windshield, and reconstructing images without any trace of weather effects. The problems we face in even highly sophisticated generative architectures, such as pix2pix and diffusion models, are the problem of reconstructing images without any trace of weather effects. In this paper, we propose the SFHformer [1] architecture to clear out images' weather effects and apply the sub-data reloader to shorten the training time.

II. Propose

Jiang et al. introduced the SFHformer architecture by applying the Fourier transformation to identify different weather effects with the frequency domain feature maps. [1] The innovative architecture not only extracts the spatial feature map of the image locally and globally. The frequency domain feature maps represented well on multiple weather degradations with edges of snow and rain, the gradient strength of blur, and foggy weather. It has robust weather removal and is lightweight, as it replaced the transformer architecture with its own feed-forward network due to the high computational cost. The equation for the loss function is used for the L1 loss function for comparing the pixel-by-pixel between the generated image and ground truth, and frequency-

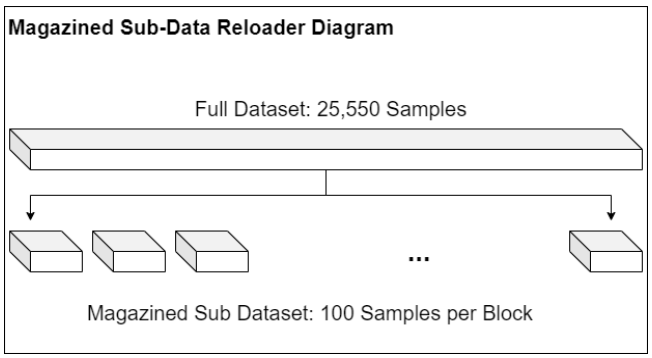


Fig. 2 The diagram of the proposed sub-data reloader.

domain feature maps between the generated and ground truth images. The lambda as the hyperparameter for the frequency domain feature map is 0.1, to less emphasize the frequency domain map results. The equation of total loss for training follows as,

$$L = \left\|\hat{I} - G\right\|_1 + \lambda\left\|F(\hat{I}) - F(G)\right\|_1 \quad (1)$$

Where $\hat{I}$ is the reconstructed pixels from the SFHformer's prediction, and G is the ground truth image from the KITTI All Weather dataset.

After obtaining datasets of the KITTI [2] and all-weather[3], we merged both datasets into one and produced our synthetic weather effects on those samples, which we call KITTI All Weather, and every sample has the ground truth and synthetic images in pairs. The synthetic images may have fog, rain, raindrops, and snow. The dataset has 25,000 samples in pairs for training and 7518 samples in pairs for testing.

From Fig. 2, our sub-data reloader can partially load from a whole trainable dataset. We randomly load only 8,000 samples in every epoch out of the 17,500 trainable samples, and every epoch has a different subgroup of the trainable dataset. In contrast, the whole validating samples were loaded, as the validating dataset only has 7,500. Thus, the duration of training time reduced as the trainable samples within a single epoch reduced.

III. Experimental Results

The hardware specifications were intel® Core™ i5-10600KCPU @ 4.10Ghz(12 CPUs), 4.1 Ghz, 32GB of RAM, and NVIDIA GeForce RTX 2070 Super.

From Table I, the SFHformer architecture with the whole dataset took 2782 seconds per epoch with 38.81 of the PSNR [5] and 98.23% of the SSIM [6]. After the SFHformer architecture was applied with the sub-data reloader, the PSNR and the SSIM were 36.11 and 97.43%, respectively. The time duration for a whole dataset per epoch from SFHformer was 2782 seconds, which is 46 minutes to complete one epoch. By loading only 8000 samples per epoch, the reconstruction performance could be marginally degraded, yet the time duration for training is drastically reduced. As a result, without much emphasis on the performance, sub-data reloader could be a practical method for mass producing the pre-trained model.

Table I: EXPERIMENTAL RESULTS WITH KITTI+ALL WEATHERS WITH SFHFORMERS ARCHITECTURE.

Methods	Training Time per Epoch (s)	PSNR	SSIM (%)
SFHformer [1] with whole dataset	2782	**98.23**	**38.81**
SFHformer [1] with sub-data loader within 8000 samples	**1039**	97.43	36.11

Fig. 3 The sample results from SFHformer with whole dataset.

Fig. 4 The diagram of the proposed sub-data reloader.

IV. Conclusion

In this paper, we presented how we reconstruct the image degraded by bad weather conditions and reduce the training time with such a highly performing architecture. Without training a whole dataset for each epoch, the sub-data reloader can be utilized for mass production of pre-trained models within a reduced time if performance is not the major issue.

ACKNOWLEDGMENT

This work was supported by the Commercialization Promotion Agency for R&D Outcomes(COMPA) grant funded by the Korean Government(Ministry of Science and ICT). (RS-2023-00304695.

References

[1] X. Jiang, X. Zhang, N. Gao, and Y. Deng, "When fast fourier transform meets transformer for image restoration," in European Conference on Computer Vision, 2024, pp. 381-402.

[2] M. Menze and A. Geiger, "Object scene flow for autonomous vehicles," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3061-3070.

[3] J. M. J. Valanarasu, R. Yasarla, and V. M. Patel, "Transweather: Transformer-based restoration of images degraded by adverse weather conditions," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022, pp. 2353-2363.

[4] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.

[5] D. Sethi, S. Bharti, and C. Prakash, "A comprehensive survey on gait analysis: History, parameters, approaches, pose estimation, and future work," Artificial Intelligence in Medicine, vol. 129, p. 102314, 2022.

[6] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE transactions on image processing, vol. 13, pp. 600-612, 2004.

데이터마이닝 기반 과불화화합물 노출 간암세포주(HepG2)의 차등 발현 유전자 및 생물학적 경로 분석 연구

이현정, 양소영*, 임길택, 송윤정, 최은창
한국전자통신연구원

ehj1224@etri.re.kr, *syyang@etri.re.kr, ktl@etri.re.kr, yjsong@etri.re.kr, ecchoi@etri.re.kr

Data-driven Gene Expression Profiling of PFAS-Exposed HepG2 Cells for Biomarker Discovery

Lee Hyeonjeong, Soyoung Yang*, Lim Kil Taek, Song Yun Jeong, Choi Eun Chang
Electronics and Telecommunications Research Institute (ETRI)

요 약

본 연구는 과불화화합물 중 하나인 과불화옥탄산(PFOA)이 간암세포주(HepG2)의 전사체에 미치는 영향을 탐색하였다. RNA-seq 데이터의 PFOA 노출군과 대조군을 비교하여 차등 발현 유전자를 도출하고 생물학적 경로 분석을 수행함으로써, 차등 발현 유전자들이 암의 진행 및 전이에 관여하는 경로와 연관됨을 확인하였다. 이는 과불화화합물이 간암의 악성화에 영향을 줄 수 있음을 시사하며, 향후 암 진행 관련 잠재적 바이오마커 후보로서 가능성을 보여준다.

Ⅰ. 서 론

과불화화합물(Per and polyfluoroalkyl substances, PFAS)은 유기 불소계 화합물의 일종으로, 높은 화학적 안정성과 우수한 물리적 특성을 지니고 있어 반도체, 자동차 산업은 물론 일상생활용품까지 다양한 산업 분야에서 활용되고 있다[1]. 그러나 난분해성 특성으로 인해 자연, 인간, 생물체, 식품 등에 잔류하며, 체내에 축적될 경우 발암 가능성 증가 및 호르몬 교란 등 인체 건강에 여러 문제를 유발한다고 알려져 있다. 특히 대표적인 과불화화합물 중 하나인 과불화옥탄산(Perfluorooctanoic acid, PFOA)은 간을 주요 표적 장기로 하며 대사이상, 간독성, 발암과의 연관성이 여러 연구에서 제기되었다 [2, 3].

본 연구에서는 PFOA 노출에 따른 간암세포주(HepG2)의 유전자 발현 데이터를 분석하여 과불화화합물의 간암 관련 독성 기전을 탐색하고자 하였다. 이를 위해 PFOA 처리군과 대조군에 대한 RNA sequencing (RNA-seq) 데이터를 획득하고, 통계 기반 차등 발현 유전자를 발굴 및 이와 관련된 생물학적 경로를 분석함으로써, PFOA 노출이 HepG2 세포의 유전자 발현에 어떠한 변화를 유도하는지 확인하였다.

Ⅱ. 본론

간암세포주인 HepG2 를 대상으로 PFOA 를 125μM 농도로 37℃에서 24 시간 처리하였으며, 동일 조건에서 Dimethyl sulfoxide(DMSO)를 처리한 세포를 대조군으로 설정하였다. 각 실험은 2 회 반복으로 수행하였다. 이후 RNA-seq 을 통해 PFOA 노출에 따른 유전자 발현 데이터를 획득하였다. 총 27,948 개의 유전자 중 하나 이상의 샘플에서 발현값이 0 인 9,469 개 유전자는 실험에서 제외하였으며, 최종적으로 18,476 개 유전자를 대상으로 분석을 수행하였다.

대조군과 처리군에서의 유전자 발현 양상을 분석하여 PFOA 가 간암세포에 미치는 생물학적 경로 및 잠재적 바이오마커 후보를 도출하기 위해, read count 데이터에서 edgeR 기반 차등 발현 유전자(Differentially expressed gene, DEG) 분석을 수행하였다. 통계적 유의 수준은 $p\text{-value}<0.05$, $|\log2FC| \geq 1$ 로 설정하였다. 총 180 개의 DEG 가 PFOA 노출에 따른 유의미한 발현 차이를 보였다. 이 중 111 개 유전자는 PFOA 처리 시 상향 발현, 69 개 유전자는 하향 발현을 나타냈다.

추가적으로 다중검정(FDR<0.1)을 적용한 결과 상향 및 하향 발현 유전자를 각 18 개, 22 개로 확인하였다.

그림 1 은 DEG 유전자 발현 패턴을 heatmap 으로 시각화 하였다. 총 180 개의 DEG 에서 PFOA 처리군과 대조군 간 뚜렷한 발현 차이가 나타났으며, 이는 PFOA 노출이 HepG2 세포의 유전자 발현 패턴을 변화시켰음을 의미한다.

이어서 DEG 가 관여하는 생물학적 기능과 경로를 해석하기 위해 GO 및 KEGG 분석을 수행하였다. 상향 발현된 유전자의 GO 분석 결과, 조직 형태 형성 및 상피세포 분화와 관련된 경로가 상위 생물학적 기능으로 확인되었다. 이는 PFOA 노출이 간세포에서 조직의 재구성과 분화를 유도하여 종양 형성 및 암 발생에 관여할 수 있음을 시사한다. KEGG 분석에서는 총 25 개 경로를 도출하였다. 유의수준 기준으로 상위에 해당하는 PI3K-Akt signaling pathway 는 5 개 DEG 를 포함하였다. 이 경로는 암세포의 생존 및 성장과 밀접한 신호 전달 경로로, PFOA 노출이 암의 진행과 연관되어 있음을 보여준다. 이 외에도 chemical carcinogenesis - receptor activation, notch signaling pathway 등 간 독성 및 조직 분화와 관련된 경로가 확인되어, PFOA 노출이 간암의 악성화 및 발암 과정에 영향을 줄 가능성을 뒷받침하였다.

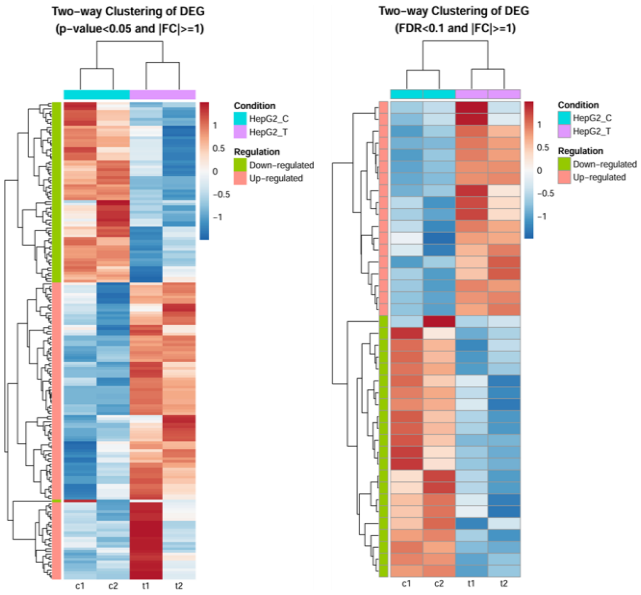


그림 1. DEG 발현 패턴에 대한 Heatmap

DEG 분석으로 발굴한 상향 발현된 유전자 중 일부는 간암과 밀접한 연관성을 가진 것으로 알려져 있다. 특히 MLPH, COL6A2, TNFSF12-TNFSF13 유전자들은 세포 이동성, 상피-중간엽 전이, 신호전달 조절 등 암의 진행과 전이에 관여하는 기능을 포함하고 있다. 이는 PFOA 노출이 해당 유전자들의 발현을 증가시킴으로써 간암의 악성화 과정에 관여할 수 있음을 내포하며, 도출된 DEG 들이 향후 암 진행 및 환경 독성 반응 평가를 위한 잠재적 바이오마커 후보로 활용될 가능성을 보여준다.

Ⅲ. 결론

본 연구는 간암세포주(HepG2) 대상으로 과불화화합물 노출에 따른 유전자 발현 변화를 분석하였다. PFOA 노출군과 대조군에 대한 RNA-seq 데이터 분석을 통해 차등 발현 유전자를 발굴하였고, 이들의 생물학적 경로가 암의 진행 및 전이와 연관되어 있음을 확인하였다. 이는 과불화화합물이 간암 악성화에 관여할 가능성이 있음을 시사하며, 도출된 차등 발현 유전자들은 향후 간암 진행 예측이나 환경 독성 반응 평가에서 잠재적 바이오마커 후보로 활용될 수 있을 것으로 기대된다.

ACKNOWLEDGMENT

This work was supported by an internal fund/grant from the Electronics and Telecommunications Research Institute (ETRI) (24YD1110) and by an ETRI grant funded by the Korean government (25ZD1110).

참 고 문 헌

[1] Brunn, H., Arnold, G., Körner, W., Rippen, G., Steinhäuser, K. G., Valentin, I. PFAS: forever chemicals—persistent, bioaccumulative and mobile. Reviewing the status and the need for their phase out and remediation of contaminated sites. Environmental Sciences Europe, 35(1), 30, 2023.

[2] Vujic, E., Ferguson, S. S., Brouwer, K. L. Effects of PFAS on human liver transporters: implications for health outcomes. Toxicological Sciences, 200(2), 213-227, 2024.

[3] Costello, E., et al. Exposure to per-and polyfluoroalkyl substances and markers of liver injury: a systematic review and meta-analysis. Environmental Health Perspectives, 130(4), 046001, 2022.

초보자 접근성 향상을 위한 모바일 로봇 티치 펜던트 UX 디자인 프레임워크 연구

안지훈, 동지연
한국전자통신연구원

dkswwl@etri.re.kr, jydong@etri.re.kr

A Study on the UX Design Framework of Mobile Robot Teach Pendants for Enhancing Novice Accessibility

Jihoon An, Jeyoun Dong
Electronics and Telecommunications Research Institute

요 약

티치 펜던트는 로봇 자동화의 핵심 도구임에도 불구하고 비전문가에게 높은 학습 장벽을 제공한다. 본 연구는 인지부하 이론과 상황인식 모델을 기반으로 초보자와 전문가 간 작업기억 차이를 분석하고, 초보자의 인지적 특성을 고려한 모바일 로봇 티치 펜던트 UX 디자인 프레임워크를 제안한다. 30 명을 대상으로 한 사용성 평가를 통해 주요 인지 부하 요인을 파악하여 효율성, 학습 용이성, 직관성, 안정성의 네 가지 설계 원칙을 도출하였다. 제안된 프레임워크를 적용한 결과, 사용자 만족도는 2.87 점에서 4.10 점으로 향상되었고 평균 작업 완료율은 91.4%를 기록하며 제안된 설계 접근의 유효성을 입증하였다. 이는 로봇 사용의 접근성을 확대하고 사용자 친화적 설계를 위한 실질적 가이드라인을 제공한다는 점에서 의의가 있다.

Ⅰ. 서론

스마트팩토리의 확산과 함께 협동로봇 및 모바일 로봇의 활용은 증가하고 있으나, 기존 티치 펜던트는 전문가 중심으로 설계되어 초보자의 학습과 사용에 제약이 크다. 특히, 중소기업 현장에서는 교육 비용 부담으로 인해 초보자의 신속한 적응이 요구되지만[1], [2], 협동로봇 티치 펜던트의 복잡성과 모바일 로봇 티치 펜던트의 통합 관리 시스템 부족은 작업자의 효율적인 학습과 태스크 수행을 저해한다. 기존 연구는 주관적 만족도 평가에 치중하는 경향이 있으며, 초보자의 인지적 특성을 고려한 체계적 설계 접근은 부족한 실정이다. 이에 본 연구는 인지부하 이론과 상황인식 모델을 적용하여 초보자의 학습 곡선을 완화할 수 있는 UX 디자인 프레임워크를 제안하고, 협동로봇 티치 펜던트 사례 분석과 사용자 평가를 통해 도출된 설계 원칙을 모바일 로봇 티치 펜던트 프로토타입에 적용하여 그 효과를 검증하고자 한다. 이를 통해 초보자 친화적 설계를 위한 이론적 근거와 실증적 데이터를 동시에 제시하며, 로봇 제어 인터페이스 연구의 확장 가능성을 마련하고자 한다.

Ⅱ. 본론

본 연구는 인지심리학적 이론을 토대로 로봇 제어 인터페이스의 문제를 규명하였다. Sweller 의 인지부하 이론(Cognitive Load Theory)에 따르면 학습자의 작업기억 용량은 제한적이며, 로봇 제어 과정은 내재적·외재적 부하를 동시에 유발한다[3]. 특히 초보자는 인터페이스의 비일관성과 전문 용어로 인한 외재적 부하에 취약하다. Endsley 의 상황인식 모델(Situation Awareness Model)은 지각- 이해- 예측의 세 단계를 설명하는데[4], 초보자는 이해 단계에서 정체되어 전체 작업 맥락을 파악하기 어렵다. 따라서 초보자 친화적 인터페이스는 정보 구조화, 단계적 안내, 예측 가능한 피드백을 제공해야 한다.

이러한 이론적 관점을 바탕으로 협동로봇 티치 펜던트의 GUI 사례를 분석하여 주요 기능을 파악하고, 이후 뉴로메카사의 Conty 2.0 을 활용하여 총 30 명을 대상으로 사용성 평가를 수행하였다. 참가자는 초보자, 중급자, 전문가 그룹으로 구분되었으며, 사전 교육 후 지도 작성, 경로 설정, 프로그램 실행 과제를 수행하였다. 평가 결과, '사용 흐름을 알 수 없는 기능 배치'와 '인터페이스의 직관성 부족'이 주요 페인 포인트로 확인되었으며, 이러한 조작 복잡성은 초보자에게 과도한 외재적 부하를 유발하는 주요 원인으로 확인되었다.

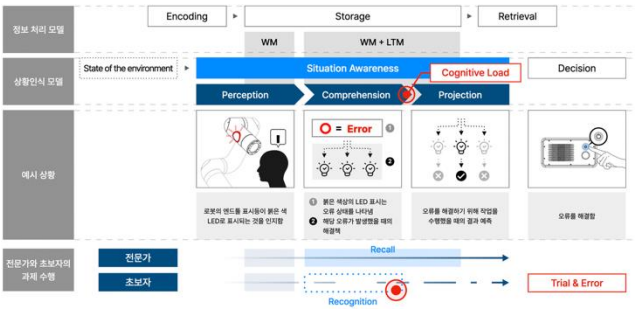


그림 1. 로봇 협업 과정에서의 상황인식 예시

이론적 분석과 실증적 평가 결과를 종합하여, 모바일 로봇 티치 펜던트 UX 설계의 핵심 요소로 효율성, 학습 용이성, 직관성, 안전성의 네 가지를 도출하고 각 요소에 따른 구체적인 디자인 원칙을 제시하였다. 효율성은 시스템 구조의 명확성, 정보의 적절한 배치, 로봇 상태의 즉각적인 인지를 포함하며, 학습 용이성은 스크린별 랜드마크와 직관적인 동작 규칙, 명령어 카테고리화를 강조한다. 직관성은 보편적인 아이콘과 명확한 용어, 익숙한 인터페이스 요소 사용을 통해 지원하고, 안전성은 팝업과 같은 오류 예방 장치와 명확한 오류 안내, 시스템적 안전 메커니즘을 제공한다.

표 1. 핵심 디자인 요소에 따른 원칙 예시

요소	넘버	원칙
효율성	1-1	GNB 를 통한 메인 기능 제시
	1-2	로컬 영역 내 기능별 레이아웃 구분과 적절한 화면 분할
	1-3	공간 정보를 활용하여 로봇의 현재 상태 및 프로그램에 대한 정보 제공
	1-4	버튼 및 영역 활성화
	1-5	기능별 도움말 제공
학습 용이성	2-1	스크린별 랜드마크 마련
	2-2	직관적인 동작규칙 설정
	2-3	명령어 카테고리화 및 정보 분류
	2-4	로봇의 작업 영역 및 가동 범위를 고려한 안내 제공

도출된 디자인 원칙에 기반하여 모바일 로봇 '모비(MOBY)'의 주요 태스크(지도 그리기, 지도 편집, 경로 설정, 프로그램 편집, 프로그램 실행)를 통합 관리할 수 있는 프로토타입을 제작하였다. 프로토타입은 GNB 를 통한 메인 기능 제시, 로컬 영역 구분, 로봇 상태 시각화, 명령어 카테고리화, 오류 예방 장치 등의 설계 원칙을 반영하였다. 검증을 위해 티치 펜던트 경험이 없는 초보자 20 명을 대상으로 만족도와 작업 완료율을 측정하였다. 그 결과, 개선안 프로토타입의 만족도는 기존 2.87 점에서 4.10 점으로 향상되었으며, 실제 로봇 조작 환경에서 평균 91.4%의 작업 완료율을 기록하여 사용성 개선 효과를 입증하였다. 다만, 조이스틱 조작의 어려움, 특정 기능의 명칭 혼란, 시각적 안내 부족 등 일부 개선점도 확인되었다.

표 2. 과제별 만족도 (N=20)

과제	구분	최소값	최대값	평균	표준편차
지도 그리기	기존안	2.33	3.00	2.63	.166
	개선안	3.78	4.44	4.10	.180
경로 설정	기존안	2.30	3.20	2.83	.255
	개선안	3.60	4.40	4.01	.210
프로그램	기존안	2.55	3.64	3.13	.310
	개선안	3.82	4.73	4.17	.240
전체	기존안	3.17	3.17	2.88	.172
	개선안	4.37	4.37	4.10	.135

Ⅲ. 결론

본 연구는 인지부하 이론과 상황인식 모델을 기반으로 초보자 친화적 모바일 로봇 티치 펜던트 UX 설계 프레임워크를 제안하고, 사례 분석과 사용성 평가를 통해 그 효과를 실증하였다. 사용성 평가 결과 만족도와 완료율 모두에서 유의미한 향상이 확인되었으며, 이는 초보자의 학습 부담을 줄이고 현장 적용성을 높이는 데 기여할 수 있음을 보여준다. 이론적 차원에서는 로봇 UX 설계의 인지적 기반을 강화하였고, 실용적 차원에서는 네 가지 핵심 UX 원칙을 제안하여 실제 적용 가능성을 확인하였다. 다만 제한된 표본과 특정 환경에 기반한 평가로 일반화에는 한계가 있으며, 향후 연구에서는 다양한 사용자군과 산업 맥락을 고려한 확장 검증이 필요하다.

ACKNOWLEDGMENT

This research has been funded by the Industrial Technology Innovation Program (P0028404) of the Ministry of Industry, Trade and Energy of Korea.

This work was supported by internal fund/grant of Electronics and Telecommunications Research Institute (ETRI). [25ZD1130]

참 고 문 헌

[1] 남경민, 정의태 (2020). 중소 제조공장을 위한 제조데이터 분석 플랫폼의 모바일 SW 인터페이스 연구. Journal of Integrated Design Research, 19(2), 71-88.

[2] 산업통상자원부 (2022). 2022 년도 산업기술인력 수급 실태조사 (115016).

[3] Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive Science, 12(2), 257-285.

[4] Endsley, M. R. (1988). Design and evaluation for situation awareness enhancement.
In Proceedings of the Human Factors Society annual meeting. Sage Journals, 32(2), 97-101.

[5] Wickens, C (2013). Engineering psychology and human performance, 4nd ed. London, England: Pearson.

특징 공간 기반 중심 제약 손실 함수를 활용한 GAN의 품질 및 다양성 향상

서상원, 박현철*
국립한국교통대학교
llfkstep9811@a.ut.ac.kr, *hc.park@ut.ac.kr

Feature Space-Based Center Constraint Loss for Improving Quality and Diversity in GANs

Seo Sang Won, Park Hyun Cheol*
Korea National University of Transportation Univ.

요 약

본 논문은 적대적 생성 신경망(Generative Adversarial Networks, GAN)의 이미지 품질과 다양성을 동시에 향상시키기 위한 새로운 제약 기반 손실 함수를 제안한다. 제안하는 방법은 학습 데이터셋을 사전 학습된 신경망을 통해 특징 공간으로 임베딩한 뒤, 해당 공간에서의 중심점을 계산하고, 각 특징 벡터와 중심점 간의 평균 거리를 참조 거리로 정의한다. 이후 학습 과정에서 생성자가 생성한 이미지 또한 동일한 특징 공간에 임베딩되며, 생성 이미지의 특징 벡터가 중심점으로부터 참조 거리보다 벗어나는 경우, 이에 대한 패널티를 손실 함수에 추가한다. 이 제약은 생성자가 실제 데이터의 분포 범위 내에서 이미지를 생성하도록 유도함으로써, 품질과 다양성의 균형을 효과적으로 유지한다. CelebA 데이터셋을 기반으로 한 실험 결과, 제안하는 방법은 FID 및 LPIPS 지표에서 기존 방식 대비 우수한 성능을 보인다,

Ⅰ. 서 론

적대적 생성 신경망은 높은 품질의 이미지를 생성할 수 있는 대표적인 생성 모델로, 다양한 컴퓨터 비전 분야에서 활발히 활용되고 있다[1]. 그러나 GAN은 학습의 불안정성과 함께, 이미지의 다양성이 부족해지는 모드 붕괴 문제, 그리고 생성 이미지의 품질 저하와 같은 한계를 여전히 안고 있다.

이러한 문제를 해결하기 위한 기존 연구 중 하나로, 낮은 해상도에서 시작하여 점진적으로 고해상도로 확장해 나가며 훈련의 안정성과 이미지 품질을 향상시키는 점진적 성장 기법이 제안된 바 있다 [2]. 해당 방법은 일정 수준의 성능 향상을 보였으나, 네트워크 구조의 단계적 변경 및 해상도 전환을 위한 추가 로직 설계가 요구되어 모델의 구현 복잡도 및 계산 비용이 증가하는 한계를 가진다.

본 논문에서는 위와 같은 한계를 보완하고자, 특징 공간에서의 중심점 기반 제약 손실 함수를 새롭게 제안한다. 제안된 방법은 먼저 학습 데이터셋을 사전 학습된 VGG 네트워크를 통해 특징 공간으로 임베딩한 후, 해당 공간에서 데이터의 중심점과 각 특징 벡터 간의 평균 거리를 계산하여 이를 참조 거리로 정의한다. 이후 학습 과정에서 생성된 이미지 역시 동일한 특징 공간에 임베딩되며, 생성 이미지의 특징 벡터가 중심점으로부터 이 참조 거리보다 크게 벗어나는 경우, 이에 대한 패널티를 손실 항에 추가한다.

이러한 제약은 생성자가 실제 데이터 분포의 중심을 벗어나지 않는 범위 내에서 이미지를 생성하도록 유도함으로써, 이미지 품질 저하 없이 다양성 손실도 방지하는 효과를 기대할 수 있다. 제안한 방법은 구조 변경이나 별도의 네트워크 추가 없이도 손실 함수 수준에서 간단하게 적용 가능하다는 점에서 효율적인 대안이 될 수 있다.

Ⅱ. 본론

2,1특징공간에서의중심점과참조거리계산

사전 학습된 VGG 네트워크를 통해 학습 데이터셋의 각 이미지 x_i를 특징 벡터 $f_i = \Phi(x_i) \in R^d$로 임베딩한다. 이후 전체 특징 벡터들의 평균 벡터를 중심점 c로 정의한다.

$$c = \frac{1}{N}\sum_{i=1}^{N} f_i \quad (1)$$

그리고 각 특징 벡터들과 중심점 간의 평균 거리, 즉 참조 거리는 다음과 같이 정의된다:

$$d_{ref} = \frac{1}{N}\sum_{i=1}^{N} \| f_i - c \|_2 \quad (2)$$

중심점과 참조 거리는 학습 전에 계산 후 고정되며, 생성 이미지 평가 시 기준으로 사용된다.

2.2 생성 이미지에 대한 중심 제약 손실

학습 중 생성자가 생성한 샘플 $\hat{x} = G(z_j)$에 대해서도 동일한 방식으로 특징 벡터 $\hat{f_j} = \Phi(\hat{x_j})$를 추출한다. 이때, 생성된 특징이 중심점으로부터 참조 거리보다 멀리 떨어진 경우, 손실 항을 통해 패널티를 부여한다.

$$L_{centroid} = \frac{1}{M}\sum_{j=1}^{M} \max(0, \| \hat{f_j} - c \|_2 - d_{ref}) \quad (3)$$

이 손실은 생성 이미지의 특징 벡터를 데이터 분포의 중심점을 기준으로 정의된 참조 거리 범위를 벗어나지 않도록 유도한다.

2.3 최종 생성자 손실 함수

기존의 적대적 손실과 함께, 본 논문에서 제안하는 중심 기반 제약 손실을 결합하여 생성자의 최종 손실은 다음과 같이 정의된다:

$$L_G = L_{adv} + \lambda_{centroid} \cdot L_{centroid} \quad (4)$$

생성자 손실에 새롭게 추가된 $L_{centroid}$를 최소화함으로써, 생성된 이미

지의 분포가 실제 데이터의 분포를 더욱 잘 근사하도록 유도할 수 있다.

Algorithm 1 Training

repeat
  $\hat{x} \leftarrow G(z)$
  $\mathcal{L}_D \leftarrow \mathrm{BCE}(D(x), 1) + \mathrm{BCE}(D(\hat{x}), 0)$
  $\hat{f} \leftarrow \Phi(\hat{x})$, $m \leftarrow batch_size$
  $\mathcal{L}_{centroid} \leftarrow \frac{1}{m}\sum_{j=1}^{m} \max(0, \|\hat{f}_j - c\|_2 - d_{ref})$
  $\mathcal{L}_G \leftarrow \mathrm{BCE}(D(\hat{x}), 1) + \lambda \cdot \mathcal{L}_{centroid}$
until converged

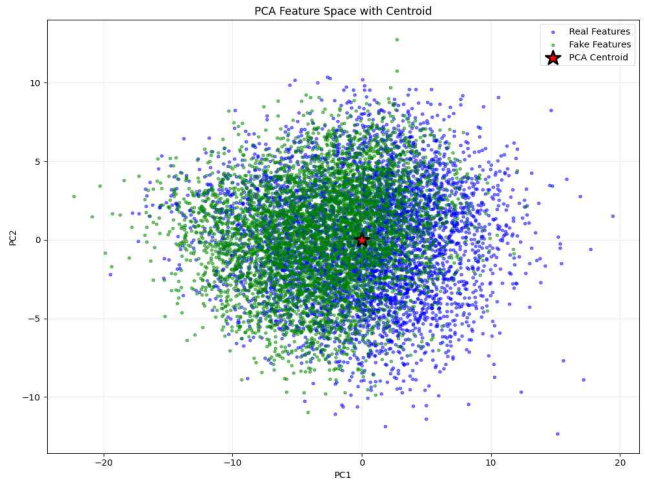


그림. 1 임베딩된 특징들과 그리고 중심점

그림 1은 실제 이미지와 생성 이미지를 특징 공간에 임베딩한 결과를 보여준다. 두 분포는 공통된 중심점을 기준으로 유사한 형태로 퍼져 있으며, 생성 이미지의 특징 벡터들이 실제 이미지의 특징 분포와 잘 정렬되어 있음을 확인할 수 있다. 이는 본 논문에서 제안한 중심 기반 제약 손실 함수가 생성 이미지의 특징 분포를 실제 데이터의 중심점과 일치시키는 데 효과적임을 시각적으로 입증한다.

Ⅲ. 실험

3.1 실험 설정

본 논문의 실험은 CelebA 데이터셋을 활용하여 진행됐으며, 생성 모델은 DCGAN[3] 구조를 기반으로 한다. 특징 추출에는 사전 학습된 VGG-16 네트워크의 16번째 계층까지 통과한 출력을 사용하였다. 제안한 중심 기반 제약 손실의 가중치 계수는 0.1로 설정한다. 모델의 성능 평가는 FID와 LPIPS 지표를 통해 수행됐으며, FID는 50k, LPIPS는 5k 쌍을비교하여 계산한다.

3.2 정량적 지표와 정성적 평가

표. 1 CelebA 64x64

	FID	LPIPS
DCGAN[3]	32.1361	0.2918
Ours	30.9519	0.2981

표. 1은 기존 DCGAN[3]과 본 논문에서 제안한 중심 기반 제약 손실 함수를 적용한 모델 간의 정량적 성능을 비교한 결과를 나타낸다. 제안한 기법은 DCGAN[3] 대비 FID가 낮고, LPIPS는 높은 값을 기록한다. 이는 중심 기반 제약이 생성 이미지의 특징 분포를 실제 데이터의 중심 영역에 보다 가깝게 유도함으로써, 품질 저하 없이 시각적으로 구분 가능한 이미지를 다양하게 생성하는 데 일정 수준 기여했음을 의미한다. 두 지표 간 균형 개선이 확인되었다는 점에서, 본 손실 함수가 기존 구조에 단순하게 적용되면서도 이미지의 품질과 다양성 간의 균형을 조절할 수 있는 유효한 전략이 될 수 있음을 시사한다. 다만, 전체적인 성능 개선 폭은 제한적이며, 향후 다양한 조건에서의 추가 실험이 필요하다.

그림. 2 (Ours) CelebA 64x64 생성 이미지

그림. 2는 제안한 특징 공간 기반 중심 제약 손실 함수를 적용한 DCGAN[3] 모델이 생성한 이미지들을 보여준다.

Ⅳ. 결론

본 논문에서는 적대적 생성 신경망의 이미지 품질 저하 및 다양성 감소 문제를 완화하기 위한 새로운 중심 기반 제약 손실 함수를 제안한다. 제안한 방법은 학습 데이터셋을 특징 공간에 임베딩하고, 그 중심점과 참조 거리를 기반으로 생성 이미지의 특징 분포가 실제 데이터 분포로부터 과도하게 벗어나지 않도록 유도한다. 이는 별도의 네트워크 구조 변경 없이 손실 함수 수준에서 간단히 적용 가능하며, 학습 안정성과 분포 정합성을 동시에 향상시킬 수 있다는 장점이 있다.

향후 연구로는 해당 제약 조건을 적응적으로 조절하는 방식이나, 다양한 특징 공간 및 생성 모델에의 확장을 통해 일반화 가능성을 검토할 수 있을 것이다.

ACKNOWLEDGMENT

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(RS-2024-00338504).

참 고 문 헌

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio,"Generative adversarial nets," Advances in Neural Information Processing Systems, Proc. NeurIPS '14, pp. 1 - 9, Dec. 2014.

[2] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," International Conference on Learning Representations, Proc. ICLR '18, pp. 1 - 26, Apr. 2018.

[3] A. Radford, L. Metz, and S. Chintala,"Unsupervised representation learning with deep convolutional generative adversarial networks," International Conference on Learning Representations, Proc. ICLR '16, pp. 1 - 16, May 2016.

엣지 컴퓨팅 환경에서 FastSAM-s 모델의 구현 및 성능비교에 관한 연구

김동우[1], 황준석[1], 강인경[1]
[1]동서울대학교 컴퓨터 소프트웨어과 학부생
gimdongu19@gmail.com, hjstone3835@gmail.com, ingyeong7161@gmail.com

A Study on the Implementation and Performance Comparison of the FastSAM-s Model in Edge Computing Environments

Dong-Woo Kim[1], Joon-Suk Hwang[1], In-Gyeong Kang[1]
[1]Dept. of Computer Software, Dong-Seoul University

요 약

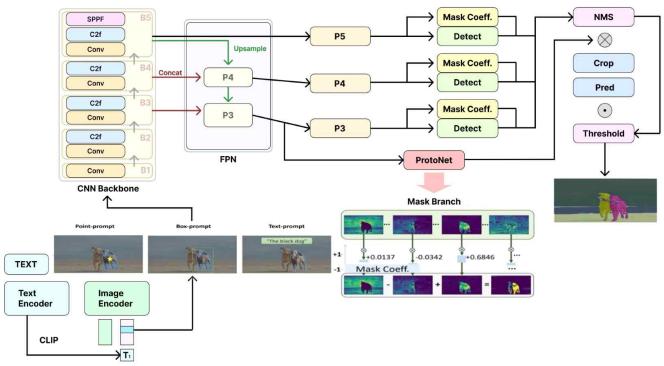
본 연구는 FastSAM 모델의 기존 PyTorch(.pt)형식을 엣지 컴퓨팅(Edge Computing) 환경에 적합하도록 8-bit 정수 양자화를 적용한 TensorFlow Lite(.tflite)형식으로 변환하여 Raspberry Pi 5에 포팅(Porting)하였다. Raspberry Pi 5 자체의 CPU만 사용한 경우와 Coral Edge TPU 가속기를 붙여 사용한 두 가지 경우에 대해 기존 PyTorch 형식을 기준으로 추론성능 비교평가를 수행한 결과, 객체 분할 정확도를 측정하는 지표인 IoU(Intersection over Union)는 기존의 PyTorch(.pt)형식과 비교하여 약 3배 떨어졌으나, FPS(Frame Per Second) 측정치는 단독 CPU 환경에서는 약 3배, Coral Edge TPU 장착 환경에서는 약 5배 향상되었음을 확인하였다.

1. 서론

범용 컴퓨터 환경은 높은 연산 자원을 바탕으로 대규모 AI 모델을 안정적으로 실행할 수 있으나, 휴대성이나 저전력을 필요로 하는 엣지 AI 분야에서는 한계가 존재한다. 이러한 제약을 극복하기 위해 최근에는 엣지 컴퓨팅에서 인공지능 모델을 구동하는 연구가 활발히 이루어지고 있다. 그러나 엣지 컴퓨팅 환경에서는 연산 성능의 한계로 인해 모델 크기 축소 및 경량화가 필수적이며, 이 과정에서 성능의 저하가 불가피하다. 본 연구에서는 데스크톱 환경에서 FastSAM PyTorch 형식의 모델을 수행한 결과를 기준으로, Raspberry Pi 5 환경에서 CPU만을 이용하여 수행한 TensorFlow Lite 변환 모델의 경우와 엣지 AI 가속기인 Coral Edge TPU을 사용한 Edge TPU 최적화 모델의 성능을 비교평가 하였다. 성능 평가 측정치는 객체 분할 정확도를 나타내는 IoU와 처리 속도를 나타내는 FPS를 지표로 수행하였다.

2. FastSAM 모델

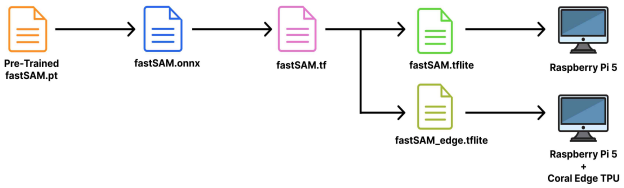
FastSAM은 SAM(Segment Anything Model)의 높은 계산 비용 문제를 해결하기 위해 제안된 경량 대안 모델로, YOLOv8-seg 구조를 기반으로 한다[1]. 본 모델은 객체 분할과 프롬프트 기반 선택의 두 단계로 구성되며, CNN 기반 탐지기와 인스턴스 분할 브랜치를 활용해 SAM 대비 약 50배 빠른 추론 속도를 가지면서도 유사한 수준의 성능을 유지할 수 있다. 본 연구에서는 이러한 FastSAM의 경량성과 처리 능력에 주목하여 FastSAM 모델 중 경량화 모델인 FastSAM-s 모델을 선택하여 연구를 진행하였다. 그림1은 FastSAM 모델의 구조도이다.



[그림 1] FastSAM 모델 구조도

3. FastSAM-s 모델의 파일형식 변환

본 연구에서는 FastSAM 모델의 기존 PyTorch(.pt) 형식을 엣지 컴퓨팅 환경에 포팅하기 위해 8-bit 정수 양자화를 적용한 TensorFlow Lite(.tflite)형식으로 변환하였다. 변환 과정은 먼저 PyTorch 형식의 모델을 ONNX(Open Neural Network Exchange) 형식으로 내보낸 뒤, 이를 TFLite 형식으로 변환하는 단계로 구성된다[2]. 이 과정에서 모든 연산을 8-bit 정수 연산으로 변환하여 모델 크기를 축소하고, Edge TPU 환경에서 실행이 가능한 edge.tflite형식으로 변환하였다. 그림 2는 FastSAM-s모델의 파일 형식 변환 과정을 나타낸 것이다.



[그림 2] FastSAM-s 모델 형식의 변환 과정

4. 실험 환경 및 결과

표 1은 본 연구의 실험 환경 규격 내용이다.

<표 1> 실험 환경 규격

구분	항목		내용
Raspberry Pi 5	H/W	CPU	Broadcom BCM2712
		RAM	8GB LPDDR4X-4267
	S/W		OS==Debian12(BookWorm) Python == 3.9.12, 3.11.2 opencv-python == 4.10.0.84 torch==2.0.1, torchvision==0.15.2
Google Coral Edge TPU	H/W		8TOPS (Google Edge TPU Process)
	S/W		Python == 3.9.12 edge-tpu-silva == 1.0.4
	power consumption		2W

표 2는 실험에서 사용한 성능 측정 방식에 관한 내용이다.

<표 2> 성능측정 방식

구분	내용
IoU (Intersection over Union)	객체 분할 모델의 성능을 평가하는 지표로 모델이 생성한 마스크의 정확도를 측정한 값
FPS (Frame Per Second)	처리 속도를 평가하는 측정치로 /ms 기준으로 계산한 값

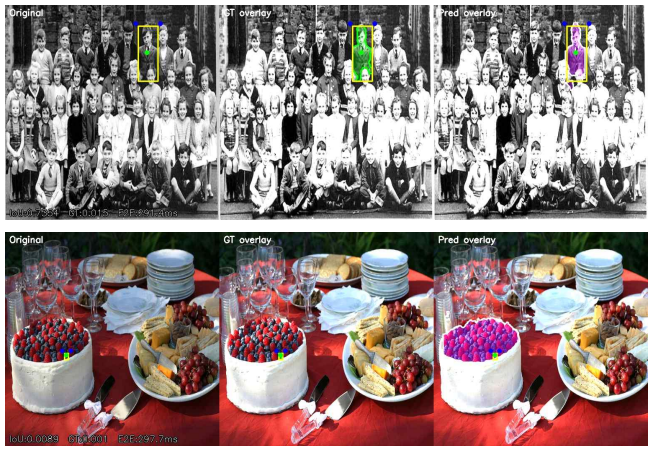
표 3은 모델의 형식별 정확도와 처리 속도를 비교한 것이다. 본 연구에서 사용한 이미지의 총 개수는 50장이며, 총 마스크의 개수는 2119개다. 포인트, 박스, 텍스트, 마스크 네 가지의 프롬프트 입력 방식 중 포인트와 박스를 혼합한 한 가지의 프롬프트를 입력으로 사용하여 성능을 측정 및 비교하였다. 전체 마스크의 객체 분할 성능측정 결과는 평균 0.33으로 기존 PyTorch 형식의 모델에 비해 약 3배 낮은 수치를 보이지만, 이미지 전체 크기 512를 기준으로 마스크의 크기가 0.05 이상일 때의 값은 약 0.7, 그 이하일 때는 약 0.3의 정확도를 갖는다.

<표 3> 형식별 성능 비교

구분	IoU	FPS (/ms)
PyTorch (.pt)	1	1545.8
TFLite (.tflite)	0.3343	465.8
Edge TFLite (.tflite)	0.3344	289.4

그림 3과 그림 4는 TFLite 형식과 Edge TFLite 형식의 모델을 수행한 결과이다. 두 형식의 모델 모두 큰 물체에 대한 세그멘트(segment) 성능은 기존 PyTorch 형식의 모델과 비교하여 큰 차이가 발생하지 않았으나, 작은 물체에 대한 정확도 성능은 두 형식의 모델 모두 기존 모델에 비하여 현저히 떨어지는 것을 확인할 수 있다. 본 연구에서 측정한 결과, 추론 속도 성능 면에서 기존 모델에 비교하여 TFLite 형식 모델은 3.31배 빨랐으며, Edge TFLite 형식 모델은 5.34배 빠름을 확인할 수 있다.

[그림 3] Raspberry Pi 5 CPU 단독 환경에서의 TFLite 수행 결과



[그림 4] Coral Edge TPU를 장착한 환경에서의 Edge TFLite 수행 결과

5. 결론

본 연구에서는 데스크톱 환경에서 FastSAM-s모델의 PyTorch 형식을 수행한 결과를 기준으로, Raspberry Pi 5 환경에서 CPU만을 이용하여 수행한 TensorFlow Lite 변환 모델과 엣지 AI 가속기인 Coral Edge TPU을 사용한 Edge TPU 최적화 모델의 성능을 비교 평가한 결과를 보면, 평균 처리 속도 면에서는 기존 모델에 비하여 TFLite 형식 모델은 3.31배, Edge TFLite 형식 모델은 5.34배 향상되었다. 객체 분할 정확도 면에서는 두 형식의 모델 모두 평균 0.33으로 확인되었다. 큰 객체에 대한 분할 성능은 기존 모델과의 큰 차이를 보이지 않는 결과를 얻었으나, 양자화에 따른 수치 표현 축소와 입력 해상도를 1024에서 512로 축소함에 따라 작은 객체에 대한 분할 정확도는 떨어졌다. 향후 동일한 실험 환경에서 캘리브레이션 데이터셋을 구축하고 작은 객체 분포에 맞춘 활성값 범위를 추정하는 것으로 스케일과 제로포인트 산정 절차를 최적화하는 연구를 진행하고자 한다.

REFERENCES

[1] Xu Zhao et al, “Fast Segment Anything”, arXiv:2408.00714, 2023

[2] Pradyun Gedam (2024). Introduction to ONNX.https://onnx.ai/onnx/intro/

엣지 컴퓨터 환경(Raspberry Pi5)에서 BitNet-b1.58 모델과 Gemma3 모델의 구현 및 성능비교에 관한 연구

이승환[1], 박성준[1], 김정훈[1]
[1]동서울대학교 컴퓨터 소프트웨어과 학부생
seung_hwan10@naver.com, syaa123@naver.com, thdnfrhfem@gmail.com

A Study on the Implementation and Performance Comparison of the BitNet-b1.58 and Gemma3 Models in an Edge Computing Environment Using Raspberry Pi 5

Seung-Hwan Lee[1], Seong-Jun Park[1], Jung-Hun Kim[1],
[1] Dept. of Computer Software, Dong-Seoul University

요 약

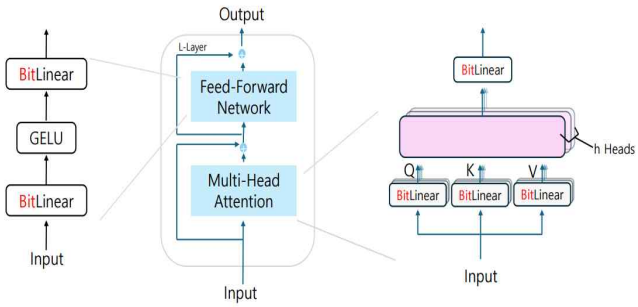
본 연구는 마이크로소프트(Microsoft)에서 공개한 경량화 LLM(Large Lanuage Model) BitNet b1.58 모델과 구글(Google)에서 공개한 Gemma3 모델을 엣지 컴퓨팅 환경인 Raspberry Pi 5에서 설치, 운영하여 각 성능을 비교 평가한 것이다. 본 연구에서는 초당 토큰(Token) 생성 속도, CPU 효율, 메모리(Memory)효율 그리고 생성된 문장에 대한 품질성능을 정량적 측정지표 ROUGE-1, BLEU, METEOR를 사용하여 평가하였다. 실험 결과, BitNet b1.58은 Gemma3 대비 약 4.4배 빠른 토큰 생성 속도, 약 4.5배 높은 CPU효율, 약 15배 높은 메모리 효율을 보였으며, 생성 문장의 품질 성능평가는 Gemma3가 모든 분야에서 좋은 성능을 보였다.

1. 서론

최근 온 디바이스(On-Device) AI의 중요성이 높아짐에 따라 한정된 자원을 사용하면서 높은 성능을 유지할 수 있는 경량화 LLM모델에 대한 연구가 활발히 진행되고 있다. 특히 대규모 언어 모델(Large Language Model)의 큰 문제점은 성능이 좋아질수록 연산량과 메모리 사용량 등 자원 요구가 기하급수적으로 증가한다는 점이다. 따라서 최근 LLM의 연구 동향은 성능을 최대한 유지하면서 사용하는 자원량을 최소화 하는 것을 목표로 많은 연구가 이루어지는 추세이다[1]. BitNet-b1.58 역시 이러한 흐름 속에서 MS(Microsoft)사가 제안한 경량화 LLM 모델 중 하나이다. 본 연구에서는 BitNet-b1.58 모델에 대해 분석하고, Raspberry Pi 5와 같은 엣지 컴퓨터 환경에서 구글에서 발표한 경량화 LLM 모델인 Gemma3와 성능비교평가를 수행하였다.

2. BitNet b1.58

BitNet-b1.58[2] 모델은 BitNet 프레임워크(Framework)를 활용하여, QAT(Quantization-Aware Training) 방식으로 처음부터 1.58-Bit 가중치로 학습된 네이티브 경량화 모델이다. BitNet 프레임워크는 기존 LLM 구조를 유지하며 피드포워드 네트워크(Feed-Forward Network)와 멀티 헤드 어텐션(Multi-Head Attention) 부분에 리니어 레이어(Linear Layer) 부분을 비트리니어 레이어(BitLinear Layer)로 단순히 교체한 형태이다. 그림 1은 BitNet 프레임워크를 나타낸 구조도 이다.

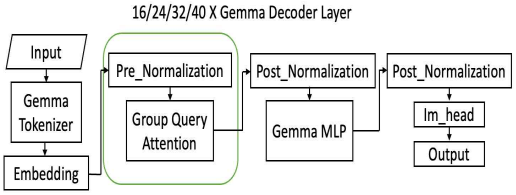


[그림 1] BitNet 프레임워크 구조도

3. Gemma3

Gemma3는 디코더 기반(Decoder-only) 트랜스포머 구조로 설계되었으며, 단일 GPU 또는 TPU 환경에서도 학습 및 추론이 가능하다. GQA(Group Query Attention) 방식을 이용하여 Query만 그룹화하여 연산량을 줄이는 방식으로 메모리 사용을 효율화하였다. FAv2(Flash Attention v2) 고속 연산과 메모리 저소비를 동시에 달성하는 어텐션(Attention) 알고리즘을 도입하였다. MLP 구조 개선은 정적 학습과 성능 향상을 위해 시그모이드

(Sigmoid) 활성화 함수를 채택하였다. 트리플 프로젝션 구조를 통해 표현 능력을 강화하면서도 연산 효율성을 유지하였다. 그림 2는 Gemma3의 구조도 이다.



[그림 2] Gemma3 구조도

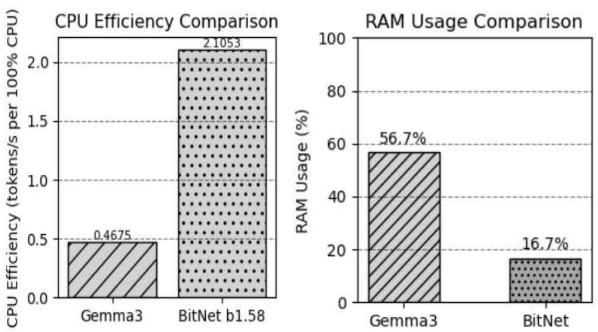
4. 실험 환경 및 구현결과

표 1는 본 연구에서 Bitnet-b1.58-2B 모델과 Gemma3-4B 모델을 설치하여 운영 비교평가 한 Raspberry Pi 5의 주요 규격이다.

<표 1> 실험구현 환경의 주요규격

구분	항목	규격
Raspberry Pi 5	CPU	Broadcom BCM2712
	RAM	8GB LPDDR4X-4267

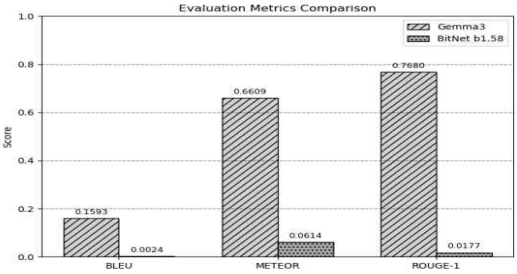
Bitnet-b1.58-2B 모델과 Gemma3-4B 모델을 엣지 컴퓨팅 환경의 대표적인 경우인 Raspberry Pi 5 환경에 설치 후 성능을 비교분석 하였다. 최대 554토큰(Token)길이의 문장을 생성 하였으며 생성하는 과정에서 초당 토큰생성 속도, CPU 효율, 메모리 효율을 비교하였다. 모델의 정량적인 품질성능 측정을 위해 SQuAD v1.1 데이터 중 Super Bowl 50 관련 문단을 기반으로 질문 응답 문제 100개를 통해 성능 평가를 수행하였다. 성능평가의 측정방식은 ROUGE-1, BLEU, METEOR 3가지 방식을 사용하였다. 초당 토큰 생성 개수는 Bitnet-b1.58-2B 모델의 경우 약 8.0[tok/s] 이고 Gemma3-4B 모델은 약 1.80 [tok/s]의 결과가 나왔다. CPU 점유율은 각각 약 385[%] , 약 380[%] 로 측정되었다. 이를 CPU 코어 하나 기준으로 정규화을 수행하여 CPU 효율을 계산하였다. 메모리 효율은 각각 약 16.7[%], 56.7[%]로 측정되었다. 그림 3은 이러한 측정 결과를 나타낸 것이다.



(a) CPU 효율 그래프 (b) 메모리 효율 그래프

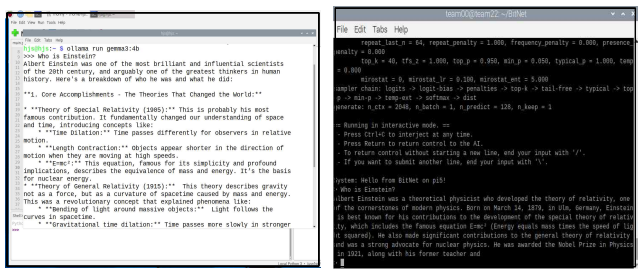
[그림 3] 두 모델의 CPU, 메모리 효율비교 결과

모델별 생성문장에 대한 품질성능 측정 결과는 그림 4 와 같다.



[그림 4] 정량적인 성능 평가 결과

정규화 된 초당 토큰 생성 개수는 Bitnet b1.58-2B는 약 2.1053 [tok/s] 당 [100%CPU] 이고 Gemma3-4B는 약 0.4675 [tok/s] per [100%CPU] 로 BitNet-b1.58-2B이 Gemma3-4B 대비 CPU효율이 약 4.5배 높은 결과를 얻었고, 메모리효율은 Bitnet-b1.58-2B이 Gemma3-4B 대비 15배 높은 결과를 얻었다. ROUGE-1, BLEU, METEOR을 이용한 정량적인 품질 성능평가는 Bitnet-b1.58-2B이 0.0024 ,0.0614 ,0.0177 Gemma3-4B이 0.1593, 0.6609, 0.7680로 Gemma3-4B모델이 모두 더 좋은 결과를 얻었다. 그림 5는 Raspberry Pi 5에서 각 모델을 구현한 동작화면이다.



(a) Gemma3-4B 모델 동작화면 (b) BitNet-b.1.58-2B 모델동작화면

[그림 5] 각 모델의 동작화면

5. 결론

본 연구에서는 MS사의 Bitnet-b1.58-2B모델과 구글의 Gemma3-4B 모델을 엣지 컴퓨팅 환경인 Raspberry Pi 5 에서 구현하고, CPU 효율성, 메모리 효율성 그리고 품질성능 평가를 진행하였다. 본 연구의 실험 결과를 통하여 엣지 컴퓨팅 환경에서 BitNet-b1.58-2B 모델이 에너지 효율성이 더 좋다는 결과를 확인할 수 있었고, 모델의 파라미터를 더 증가시킨다면 생성문장의 품질성능도 더 우수할 것으로 판단된다. 본 연구의 결과는 엣지 컴퓨팅 환경에서 경량 LLM 모델을 설치 및 운영하고자 할 때 활용 될 수 있을 것이다.

참고문헌

[1]Z. Wang, Y. Liang, Z. Xu, T. Sun, Y. Zhang, S. Zhang, Y. Wang, and Y. Chen, “Beyond Efficiency: A Systematic Survey of Resource-Efficient Large Language Models,” arXiv preprint arXiv:2401.00625, 2024.
[2]S. Ma, H. Wang, S. Huang, X. Zhang, Y. Hu, T. Song, Y. Xia, and F. Wei, “BitNet b1.58 2B4T Technical Report,” arXiv preprint arXiv:2504.12285, 2025.

녹내장 진단을 위한 ViT 기반 모델의 해석 가능성 비교: CNN과의 시각적 해석력 차이 분석 및 Occlusion, Grad-CAM 기법 활용

유호명, 하정욱*
서강대학교

hmyou6895@sogang.ac.kr, *jwha6169@sogang.ac.kr

Interpretability Comparison of ViT-Based Models for Glaucoma Diagnosis: Visual Explainability Analysis with CNNs Using Occlusion and Grad-CAM

You Ho Myung, Jeung Uk Ha*

Sogang University

요약

녹내장은 조기 진단이 필수적인 주요 실명 원인이다. 그러나 기존 딥러닝 연구는 Grad-CAM이나 Attention Map에 의존하여 해석 가능성이 정성적 수준에 머물렀다. 본 연구는 CNN 기반 MobileNetV3-Large와 ViT 기반 MobileViT-S를 동일 조건에서 학습시키고, Grad-CAM과 Occlusion을 결합한 정량적 지표(Consistency, ROI Energy Ratio, Quantitative Drop)를 통해 비교하였다. 단순 시각적 해석에 그치지 않고 지표별 유의미성을 통계적으로 검증한 결과, ViT는 시신경 유두에 국한되지 않고 RNFL을 포함한 전역적 패턴을 활용하는 경향을 보였다.

Ⅰ. 서 론

녹내장은 전 세계적으로 발생하는 실명의 주요 원인 중 하나로, 조기 진단과 적절한 치료 여부에 따라 시력 보존 가능성이 크게 달라진다. 최근 안저(fundus) 사진을 활용한 딥러닝 기반 진단 모델이 임상 현장에 도입되면서 진단의 정확도와 효율성이 크게 향상되었다. 그러나 여러 임상 연구에서는 시신경 유두(optic disc) 영역만을 기반으로 한 진단에 한계가 있을 수 있음을 보고하고 있으며, 보다 안정적이고 신뢰할 수 있는 진단을 위해 시신경 유두뿐 아니라 섬유층(retinal nerve fiber layer; RNFL)이나 주변 혈관 구조 등 전역적인 패턴의 동반 분석이 권장된다[1][2].

인공지능 학회에서도 CNN(Convolutional Neural Network)과 Vision Transformer(ViT) 계열 모델이 최근 활발히 연구되고 있으나, 대부분의 연구는 Grad-CAM이나 Attention Map을 이용해 모델이 주목하는 영역을 정성적(qualitative)으로만 해석하는 수준에 머물렀다[3]. 이러한 한계로 인해 녹내장을 진단하는 모델이 집중하는 패턴을 정량적으로 비교·분석하고, 임상적으로 해석 가능한 결과를 도출하는 데는 한계가 존재했다. 본 연구는 이러한 문제를 해결하기 위해 동일한 학습 조건하에서 VNet(MobileNetV3-Large)과 ViT(MobileViT-S) 모델을 학습한 뒤, Grad-CAM과 Occlusion 기법을 결합하여 정량적인 분석 프레임워크를 설계하였다. Consistency(Cosine, Pearson), ROI Energy Ratio, Quantitative Drop(절대·상대) 등의 지표를 도입해 내부(AIROGS-Light)와 외부(ORIGA) 데이터셋에서 모델의 주목 패턴을 비교하고, ViT 기반 모델의 전역적 패턴 학습 특성과 우수한 일반화 성능을 실험적으로 검증하였다.

Ⅱ. 본론

2.1 데이터셋

본 연구에서는 세 가지 데이터셋을 활용하였다.

- 내부 데이터셋 (AIROGS-Light)

AIROGS 공개 데이터셋을 기반으로 Kaggle 플랫폼에서 약 8,000장의 훈련 이미지, 770장의 검증 이미지, 770장의 테스트 이미지를 다운로드 받은 후 전처리하여 실험에 활용하였다[4][9].

- 외부 데이터셋Ⅰ (ORIGA)

Kaggle 플랫폼에서 제공되는 ORIGA 공개 데이터셋을 동일한 전처리 조건으로 변환하여 사용하였다[5][10].

- 외부 데이터셋 Ⅱ(REFUGE2)

Kaggle 플랫폼에서 제공되는 REFUGE2 공개 데이터셋을 동일한 전처리 조건으로 변환하여 사용하였다[11][12].

2.2 모델 구조 및 학습 설정

- VNet (MobileNetV3-Large)

MobileNetV3-Large는 CNN 기반의 경량 모델로, 모바일 환경에서 효율성과 속도를 고려한 구조를 특징으로 한다[6]. 학습 코드 구현은 Kaggle Notebook PyTorch Easy Setup for Glaucoma Detection; 92.6%을 참고하여 수정하였다 [7].

- ViT (MobileViT-S)

Transformer 기반 경량 모델로, 전역적 관계 학습과 국소적-지역적 융합 구조를 통해 효율성과 표현력을 동시에 확보한다[8].

두 모델 모두 ImageNet 사전 학습 가중치를 사용하며, 입력 크기 512*512, 배치 크기 16, 학습률 1e-3, StepLR 스케줄러, 11 epochs, Adam 옵티마이저, 동일한 데이터 증강 기법으로 학습하여 비교의 공정성을 확보하였다. 단, 시드는 고정하지 않았다.

2.3 설명 가능성 분석

- Consistency (Cosine, Pearson): Grad-CAM과 Occlusion 맵 간의 유사도를 측정하는 지표

$$Consistency_{\cos}(G,O) = \frac{\langle G, O \rangle}{\| G \| O \|}$$

$$Consistency_{pearson}(G,O) = \frac{Cov(G,O)}{\sigma_G \sigma_O}$$

- ROI Energy Ratio: 상위 15% 활성화 영역이 Grad-CAM 맵에서 차지하는 에너지 비율

$$EnergyRatio(G) = \frac{(\Sigma_{i \in ROI} G_i)}{(\Sigma_i G_i)}$$

- Quantitative Drop (Q-Drop, 절대·상대): Grad-CAM 상위 영역을 마스킹했을 때 모델 출력 확률의 절대 및 상대 감소량

$$QDrop_{abs} = \max(0, P(y \mid x) - P(y \mid x^{ROI-masked}))$$

$$QDrop_{r\ el} = \max(0, \frac{P(y|x) - P(y|x^{ROI-masked})}{P(y|x)})$$

Grad-CAM은 각 모델의 마지막 컨볼루션 층(VNet: features 블록의 마지막 Conv, MobileViT-S: conv_head)을 타깃으로 하여 heatmap을 생성하였다. 이는 class-specific activation이 가장 잘 반영되는 층으로, 두 모델 간 주목 패턴을 공정하게 비교하기 위함이다. Occlusion의 패치 크기(patch size)는 32, 64, 96으로 변화시키며 패치 민감도 실험을 수행했고, 각 조건에서의 Consistency 차이를 Wilcoxon signed-rank 검정과 Holm-Bonferroni 보정으로 분석하였다.

2.4 실험 결과

2.4.1 분류 성능

내부 및 외부(ORIGA+REFUGE2 combined) 데이터셋에서의 성능은 [Table 1] 과 같다.

[Table 1] 내부·외부 데이터셋에서의 모델 분류 성능 비교

Model	Dataset	ROC-AUC	Accuracy	F1-score	AUPRC
VNet	Internal	0.9834	94.03%	0.9404	0.9825
ViT		0.9828	93.77%	0.9383	0.9809
VNet	External	0.8850	86.00%	0.6441	0.7073
ViT		0.9032	86.10%	0.6667	0.7355

내부 데이터셋에서는 두 모델 모두 높은 정확도와 AUC를 기록했으나, 외부 데이터셋에서는 ViT가 소폭 우수한 성능을 나타냈다.

2.4.2 정량적 해석 가능성 비교

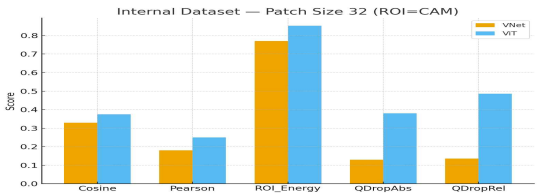


그림 1 내부 데이터셋 정량 지표 (Grad-CAM+Occlusion, n=200)

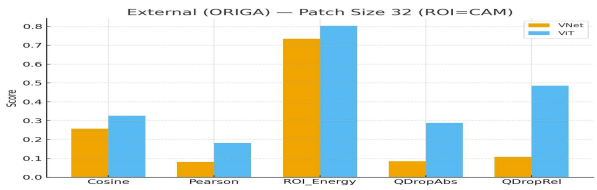


그림 2 외부(ORIGA) 데이터셋 정량 지표 (Grad-CAM+Occlusion, n=200)

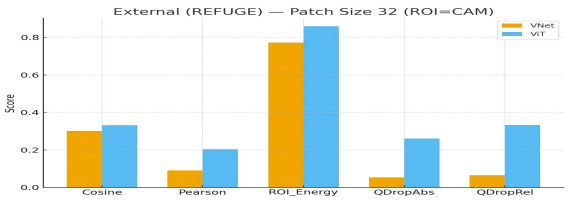


그림 3 외부(REFUGE2) 데이터셋 정량 지표 (Grad-CAM+Occlusion, n=200)

내부 및 외부 데이터셋에 대해 Grad-CAM과 Occlusion 기반 지표를 산출한 결과, ViT는 전반적으로 VNet보다 높은 값을 기록하였다. Consistency(Cosine, Pearson)는 ViT가 대체로 더 높은 경향을 보였으나, 특정 영상(예: 강한 빛반사)에서는 Grad-CAM과 Occlusion 맵이 동시에 비정상적으로 반응하여 지표 해석이 모호해지는 사례가 관찰되었다. ROI Energy Ratio 또한 ViT에서 일관되게 높게 나타나 전역적 패턴 활용을 시사했으나, 단독 지표로는 제한적인 해석에 머물 수 있다. 반면 Quantitative Drop(절대·상대)은 내부(AIROGS-Light)와 외부(ORIGA, REFUGE2) 모두에서 큰 차이를 보였으며, ViT가 ROI 영역 차단 시 출력 확률이 뚜렷하게 감소하였다. 이는 ViT가 단순히 국소적 OD 영역이 아닌 RNFL을 포함한 전역적 특징에 의존함을 강하게 시사한다.

2.4.3 패치 민감도 및 통계 검정

Occlusion 패치 크기를 32, 64, 96으로 변화시켜 분석한 결과, 전반적으로 ViT의 우위가 유지되었다. 그러나 지표별 신뢰도에는 차이가 있었다. Consistency(Cosine, Pearson)는 일부 조건에서 ViT가 더 높은 값을 보였으나, 내부 데이터셋에서는 통계적 유의성이 제한적이었고(예: Cosine, $p \approx 0.068$), 외부 REFUGE2에서는 차이가 뚜렷하지 않았다($p \approx 0.34$). ROI Energy Ratio는 내부·외부 모두에서 ViT가 높았으나, 효과 크기(Cohen's $d \approx 0.26 - 0.37$)는 상대적으로 작아 전역적 패턴 활용을 보조적으로 시사하는 수준이었다. 반면 QDrop(절대·상대)은 내부와 외부 전부에서 $p < 10^{-6}$ 수준의 강한 유의성과 큰 효과 크기(Cohen's $d \approx 0.60 - 0.92$)를 보여, ViT의 전역적 주목 패턴을 가장 신뢰성 있게 반영하는 지표로 확인되었다.

2.4.4 시각적 분석

[그림 4]은 내부 데이터셋에서 VNet과 ViT 모델의 Grad-CAM 시각화 예시를 보여준다. VNet은 대부분의 케이스에서 시신경 유두(OD) 주변에 국소적으로 집중된 활성화를 보이는 반면, ViT는 OD를 중심으로 보다 넓게 감싸는 전역적 활성화를 나타냈다. [그림 5]와 같이 일부 사례에서는 ViT의 활성화가 망막신경 섬유층(RNFL) 방향으로 길게 확장되는 패턴이 관찰되었다.

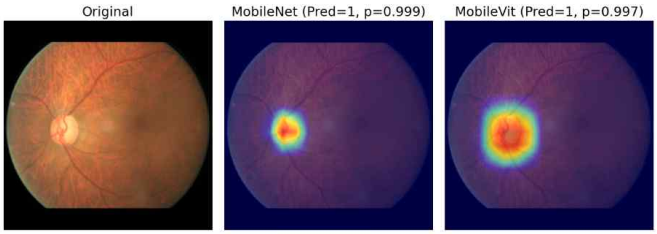


그림 4 ViT·VNet Grad-CAM (내부 데이터셋, 녹내장)

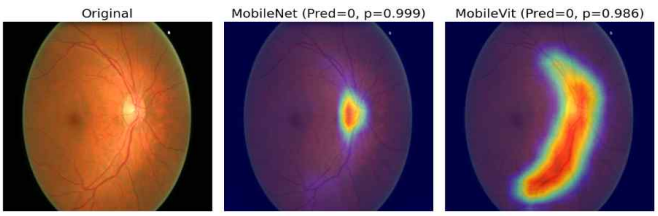


그림 5 ViT·VNet Grad-CAM (외부 데이터셋, 정상 안구)

Ⅲ. 결론

본 연구에서는 동일 조건하에서 학습된 CNN(MobileNetV3-Large, VNet)과 ViT(MobileViT-S, ViT) 모델의 시각적 해석 가능성을 정량적으로 비교하였다. Grad-CAM과 Occlusion을 결합해 여러 지표를 분석한 결과, ViT가 VNet보다 전역적 패턴을 더 폭넓게 활용하는 경향이 확인되었다. 다만 Consistency(Cosine, Pearson)는 특정 영상 조건(예: 빛반사 등)에서 해석이 왜곡될 수 있으며, ROI Energy Ratio 역시 보조적 지표로서 한계가 있었다. 반면 QDrop(절대·상대)은 내부·외부 데이터셋과 다양한 패치 크기에서 일관되게 유의한 차이를 보였으며, ViT의 전역적 주목 패턴을 가장 신뢰성 있게 반영하는 지표로 나타났다. 이러한 분석은 임상적으로 의미가 있으나, 실제 환자 데이터를 통한 후속 검증이 필요하다.

참 고 문 헌

[1] European Glaucoma Society, "Terminology and Guidelines for Glaucoma, 5th Edition," Br. J. Ophthalmol., vol. 105, suppl. 1, pp. 1-169, 2021, (https://doi.org/10.1136/bjophthalmol-2020-EGSguidelines).

[2] European Glaucoma Society, "Terminology and Guidelines for Glaucoma, 4th Edition - Summary," Br. J. Ophthalmol., vol. 101, no. 6, pp. 130-195, 2017, (https://doi.org/10.1136/bjophthalmol-2016-EGSguideline).

[3] Gu, B., Sidhu, S., Weinreb, R. N., Christopher, M., Zangwill, L. M., and Baxter, S. L., "Review of visualization approaches in deep learning models of glaucoma," Asia-Pacific Journal of Ophthalmology (Phila), vol. 12, no. 4, pp. 392-401, 2023.

[4] de Vente, C., Orlando, G., van Rijsbergen, J. J., et al., "AIROGS: Artificial Intelligence for Robust Glaucoma Screening Challenge," arXiv preprint arXiv:2302.01738, 2023, (https://arxiv.org/abs/2302.01738).

[5] Zhang, Z., Yin, F., Liu, J., Wong, W. K., Tan, N. M., Lee, B. H., Cheng, J., and Wong, T. Y., "ORIGA(-light): An online retinal fundus image database for glaucoma analysis and research," Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC), pp. 3065-3069, Aug. 2010, (https://doi.org/10.1109/IEMBS.2010.5626137).

[6] Howard, A., Sandler, M., Chen, B., Wang, W., Chen, L.-C., Tan, M., Chu, G., Vasudevan, V., Zhu, Y., Pang, R., Le, Q. V., and Adam, H., "Searching for MobileNetV3," Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), pp. 1314-1324, Oct. 2019.

[7] de Vente, C., "PyTorch Easy Setup for Glaucoma Detection; 92.6%," Kaggle Notebook, Feb. 2024, (https://www.kaggle.com/code/deathtrooper/pytorch-easy-setup-for-glaucoma-detection-92-6).

[8] Mehta, S., and Rastegari, M., "MobileViT: Light-weight, General-purpose, and Mobile-friendly Vision Transformer," Proc. Int. Conf. Learn. Representations (ICLR), Apr. 2022, (https://arxiv.org/abs/2110.02178).

[9] de Vente, C., "Glaucoma Dataset - EyePACS (AIROGS-Light v2)," Kaggle Dataset, Feb. 2024, (https://www.kaggle.com/datasets/deathtrooper/glaucoma-dataset-eyepacs-airogs-light-v2).

[10] ayush02102001, "Glaucoma Classification Datasets (ORIGA)," Kaggle Dataset,(https://www.kaggle.com/datasets/ayush02102001/glaucoma-classification-datasets).

[11] Fang, H., et al., "REFUGE2 Challenge: A Treasure Trove for Multi-Dimension Analysis and Evaluation in Glaucoma Screening," arXiv preprint, 2022. (https://arxiv.org/abs/2202.08994).

[12] jayesh0vasudeva, "REFUGE2-2020 Classification Dataset," Kaggle Dataset, n.d., (https://www.kaggle.com/datasets/jayesh0vasudeva/refuge2020-classificaton).

FiLM 기반 CNN을 활용한 Website Fingerprinting 공격 모델 연구

장채은, 김은경
국립한밭대학교
chaeeun@edu.hanbat.ac.kr, ekim@hanbat.ac.kr

A Study on Website Fingerprinting Attack Model Using FiLM-based CNN

Chae Eun Jang, Eunkyung Kim
Hanbat National University

요 약

본 논문은 암호화된 네트워크 트래픽에서의 Website Fingerprinting 성능 향상을 위해, 컴퓨터 비전 분야에서 사용되던 Feature-wise Linear Modulation(FiLM)을 1차원 CNN 모델에 적용하는 새로운 접근법을 제안한다. 제안 모델은 GRU로 추출한 트래픽 시퀀스의 문맥 정보와 통계적 특성을 조건 벡터로 사용하여, CNN의 중간 feature map을 동적으로 변조한다. 이 조건부 메커니즘은 입력 데이터의 고유한 맥락에 적응하여 모델의 표현력을 향상시킨다. 공개 TLS 데이터셋을 이용한 실험 결과, 제안 모델은 베이스라인 CNN 모델 대비 약 9%p에 달하는 유의미한 정확도 향상을 달성하며 그 우수성을 입증하였다. 본 연구는 시각 지능 분야에서 검증된 방법론이 네트워크 트래픽과 같은 1차원 시계열 데이터 분석에도 매우 효과적임을 입증하며, 이종 도메인 간의 성공적인 기술 전이 가능성을 제시한다.

Ⅰ. 서 론

Website Fingerprinting은 암호화된 네트워크 환경에서 패킷의 크기, 방향, 시간적 패턴 등을 분석하여 사용자가 접속한 웹사이트를 추정하는 공격 기법이다. Tor나 VPN과 같은 익명화 네트워크 환경에서도 이 기법을 통해 사용자가 접속한 웹사이트를 식별해낼 수 있음을 다수의 연구가 입증해왔다[1],[2],[3]. 인공지능 기술이 발전함에 따라, 최근에는 트래픽 분석 분야에 CNN과 같은 딥러닝 모델을 도입하여 정확도를 향상시키고 있다. 하지만 CNN은 모든 입력 데이터에 대해 동일한 필터를 공유하는 정적인 방식으로 동작하여, 각 트래픽이 가진 고유한 맥락적 차이를 동적으로 반영하는 데에는 명백한 한계가 있다.

이에 따라 트래픽 데이터에 CNN을 적용할 시 데이터 내 맥락적 변화와 조건별 특성을 반영할 수 있는 새로운 접근법이 요구되며, 본 연구에서는 이 문제를 해결하기 위해 Feature-wise Linear Modulation(FiLM) 기법[4]에 주목한다. FiLM은 이미지-텍스트 멀티모달 추론 분야에서 제안된 방법으로, 외부 조건에 따라 feature map을 동적으로 조정할 수 있도록 설계되었다. 본 논문에서는 이미지 도메인에서 사용되는 FiLM기법을 1차원 네트워크 트래픽 분석 문제에 맞게 변형하고 적용하여 효과적인 Website Fingerprinting을 수행할 수 있도록 한다.

Ⅱ. 본론

1. 연구배경

1.1 Website Fingerprinting

Website Fingerprinting은 암호화된 네트워크 환경에서 트래픽을 분석해 사용자가 접속하고 있는 웹사이트를 식별하는 공격 기술이다[2]. Tor나 VPN 등의 네트워크는 사용자의 익명성과 개인정보를 보호한다는 특징이 있으나, 이러한 익명화 네트워크 환경 또한 Website Fingerprinting 기술을 통해 사용자가 접속한 웹사이트를 식별해낼 수 있다는 연구가 진행되어왔다. 최근에는 인공지능 기술의 발전으로 Website Fingerprinting에 기계학습을 도입하여 공격 모델의 성능을 고도화하고 있다[5]. 이러한 Website Fingerprinting에 기계학습을 적용한 연구의 경우[1],[2] 암호화된 페이로드 정보 대신 주로 웹사이트와 사용자 간의 패킷 송수신 방향, 패킷 크기, 버스트 등의 메타데이터를 통해 웹사이트를 식별해오고 있다.

1.2. FiLM

Feature-wise Linear Modulation(FiLM)[4]은 CNN과 같은 신경망에서, 외부 조건 정보를 활용하여 중간 feature map을 동적으로 조절하는 기법이다. FiLM은 각 feature map 채널에 대해 조건에서 생성된 두 개의 파라미터, 즉 스케일링 파라미터 γ와 쉬프팅 파라미터 β를 선형 변환 방식으로 적용한다. 이 기법은 필터 기반의 feature 추출 과정에 조건별 적응성을 부여한다는 점이 핵심이다. 즉, 입력 조건에 따라 다르게 feature를 강조하거나 억제할 수 있다. 이를 통해 복잡한 문맥이나 태스크 별 특성에 맞는 유연한 표현이 가능하다. 이러한 FiLM 기법은 자연어 질문을 조건으로 하여 이미지 feature map에 변조를 적용해 복합 시각 추론을 성공적으로 수행했으며, 간단하면서도 강력한 성능 향상을 보였다.

따라서 본 연구에서는 입력 시퀀스로부터 추출한 문맥 정보(GRU 임베딩)와 통계적 특성을 결합하여 조건 벡터를 생성한다. 이 벡터를 활용해 여러 CNN 레이어의 중간 feature map에 FiLM 변조를 적용함으로써, 주로 시각 데이터 분야에서 활용되던 FiLM 기법을 1차원 트래픽 데이터 분석에 맞게 확장한다. 이를 통해 Website Fingerprinting 공격 모델의 표현력과 정확도를 높이는 것을 목표로 한다.

2. 실험

2.1 제안 방법

본 논문에서는 Website Fingerprinting 성능 향상을 위해 FiLM기법을 활용한 3-block 1D-CNN 모델을 사용한다. 이 모델은 128차원의 은닉 상태를 갖는 양방향 GRU 인코더를 통해 입력 시퀀스의 전역적 문맥을 임베딩하고, 이를 통계적 특성과 결합하여 조건 벡터를 구성한다. 이 조건 벡터로부터 동적으로 생성된 FiLM 파라미터(γ, β)는 32, 64, 128 채널을 갖는 각 컨볼루션 블록의 feature map에 적용되어, 채널별 affine 변환을 통

한 적응적 feature 변조를 수행한다.

이 방식을 통해 모델이 데이터의 맥락을 효과적으로 반영하여 표현력을 극대화하고, 이를 통해 Website Fingerprinting의 분류 정확도 향상을 목표로 한다.

2.2 환경 설정

실험에는 120개의 클래스로 구성된 ET-BERT[6]의 공개 TLS 데이터셋을 활용하였다. 총 46,732개의 PCAP으로부터 각 패킷의 크기(byte)를 추출하여 raw 시퀀스 데이터를 구축하였고, 이후 모든 시퀀스 길이를 513으로 표준화하였다. 이는 전체 데이터 중 95%가 513 이하의 길이를 갖는다는 분포 특성을 고려하여 설정한 값이다. 모든 시퀀스의 길이를 맞추기 위해 513보다 길이가 짧은 시퀀스는 0으로 패딩을 하고, 긴 시퀀스는 513 이후의 값을 잘라내었다. FiLM 레이어의 컨디셔닝을 위한 통계적 feature는 raw 시퀀스 데이터로부터 산출하였다. 모델의 훈련 및 평가는 PyTorch 프레임워크를 기반으로 수행하였으며, 데이터는 8:1:1 비율의 훈련, 검증, 테스트 세트로 무작위 분할하였다. 모델 학습에 사용된 주요 하이퍼파라미터는 [표 1]과 같다.

[표 1] 하이퍼파라미터 설정

hyperparameters	value
Batch Size	64
Number of Epochs	100
Optimizer	AdamW
Learning Rate	0.001
Weight Decay	1e-5
Betas	(0.9, 0.999)
Learning Rate Scheduler	ReduceLROnPlateau
Dropout Rate	0.5
GRU Hidden Size	128

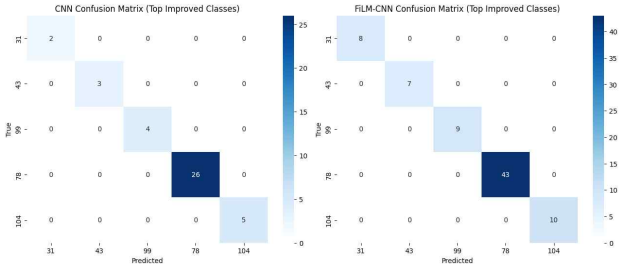
2.3 성능 비교 및 분석

[표 2]는 베이스라인 CNN과 제안된 FiLM-CNN의 성능 비교 결과를 나타낸다. 표에 나타난 바와 같이, 제안된 FiLM-CNN 모델은 베이스라인 CNN 모델 대비 평균 정확도에서 약 8.81%p의 유의미한 향상을 보였으며, 특히 주요 분류 성능 지표에서도 일관되게 개선된 결과를 보였다.

[표 2] 베이스라인 CNN과 FiLM-CNN의 성능 비교

model	Accuracy	Precision	Recall	F1score
Baseline CNN	67.49%	68.94%	67.49%	67.61%
FiLM-CNN	76.30%	76.81%	76.30%	76.23%

또한 [그림 1]은 베이스라인 CNN 대비 FiLM-CNN 모델에서 성능 향상이 두드러진 5개의 클래스를 선정하여, 이들 클래스에 대한 혼동 행렬 변화를 비교한 결과이다. 베이스라인 CNN의 경우, 일부 클래스에서 예측 정확도가 낮아 다른 클래스와 혼동되는 사례가 관찰되었으나, FiLM-CNN 모델에서는 각 클래스별 정확도가 크게 향상되었다. 이러한 결과는 FiLM 기법의 조건 기반 조절이 특정 클래스의 특징을 보다 뚜렷하게 강화하여, CNN 단독 모델 대비 정확도와 신뢰도를 동시에 높였음을 시사한다.



[그림 1] 베이스라인 CNN(좌)과 FiLM-CNN(우)의 예측 정확도 차이가 큰 5개 클래스 혼동행렬

Ⅲ. 결론

본 논문에서는 FiLM을 활용한 방식이 Website Fingerprinting 태스크에 효과적임을 제안하고 실험적으로 검증하였다. 제안된 모델은 베이스라인 CNN 모델 대비 주요 평가지표 전반에서 일관되게 우수한 성능을 보였으며, 특히 정확도에서 약 9%p에 이르는 성능 향상을 달성했다. 이는 정적인 feature 추출 방식과 달리, FiLM이 트래픽의 고유한 맥락을 조건으로 삼아 미세한 패턴의 차이를 식별하는 모델의 표현력을 강화한 결과이다.

본 연구는 특정 데이터셋 환경에서 수행된 한계가 있다. 따라서 향후 과제로는 실제와 유사한 다양한 네트워크 환경에서의 강건성을 검증하고, 모델 최적화 및 경량화 연구를 진행할 예정이다.

결론적으로, 본 연구는 이미지 처리 분야에서 주로 활용되던 FiLM과 같은 기법이 1차원 네트워크 트래픽 데이터 처리에도 효과적으로 적용될 수 있음을 시사한다. 이러한 접근법은 향후 네트워크 트래픽 분석뿐만 아니라, 유사한 시계열 및 1차원 데이터 처리 문제에 새로운 해결책을 제시할 것으로 기대된다.

ACKNOWLEDGMENT

"본 연구는 2025년 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학사업의 연구결과로 수행되었음"(2022-0-01068)

참 고 문 헌

[1] V. Rimmer et al., "Automated website fingerprinting through deep learning," in Proc. Network and Distributed System Security Symposium (NDSS), 2018.

[2] P. Sirinam et al., "Deep fingerprinting: Undermining website fingerprinting defenses with deep learning," in Proc. ACM SIGSAC Conf. on Computer and Communications Security (CCS), 2018, pp. 1928 - 1943.

[3] T. Wang et al., "Effective attacks and provable defenses for website fingerprinting," in Proc. USENIX Security Symposium, 2014, pp. 143 - 157.

[4] E. Perez et al., "FiLM: Visual reasoning with a general conditioning layer," arXiv preprint arXiv:1709.07871, 2017.

[5] B. Rexha et al., "Unveiling the digital fingerprints: Analysis of internet attacks based on website fingerprints," arXiv preprint arXiv:2409.03791, 2024.

[6] X. Lin et al., "ET-BERT: A contextualized datagram representation with pre-training transformers for encrypted traffic classification," in Proc. The Web Conference (WWW), 2022, pp. 633 - 642.

YOLOv8-pose 기반의 실시간 스쿼트 자세 분석 및 교정 시스템 구현

정재성, 신은호, 전윤호*
국립한밭대학교
{json, eunho}@edu.hanbat.ac.kr, *yhjeon@hanbat.ac.kr

Real-time Squat Posture Analysis and Calibration System Based on YOLOv8-pose systems

Jaeseong Jeong, Eunho Shin, *Yunho Jeon
Hanbat National University

요 약

본 논문은 인공지능 기반 실시간 자세 인식 기술을 활용하여 사용자의 스쿼트 동작을 정확하게 감지하고, 즉각적인 피드백을 제공하는 시스템을 제안한다. 본 시스템은 YOLO 포즈 추정 모델을 이용하여 어깨, 엉덩이, 무릎, 발목 등 주요 신체 키포인트를 검출하고, Intersection over Union(IoU) 기반 위치 검증 및 기준점 설정(Calibration)을 통해 사용자의 초기 자세를 보정한다. 또한, 어깨너비 변화를 이용한 각도 추정 방식을 적용하여 사용자의 회전 각도를 계산하고, 목표 각도와 비교하여 실시간 안내 메시지를 제공한다. 제안된 시스템은 상태 머신 기반 스쿼트 감지 로직을 통해 동작을 READY, DESCENDING, BOTTOM, ASCENDING, INVALID의 다섯 단계로 구분하며, 단계별 피드백을 시각적으로 제공한다.

Ⅰ. 서 론

최근 YOLOv8-pose[1]와 같은 실시간 포즈 추정 모델의 발전은 단일 2D 카메라만으로도 다양한 상호작용 시스템을 구현할 잠재력을 열어주었다. 그러나 스쿼트와 같이 복합적인 동작을 Raspberry Pi와 같은 엣지 디바이스 환경에서 정밀하게 분석하는 것은 모델의 추론 성능과 2D 데이터의 본질적인 한계로 여전히 어려운 과제로 남아있다.

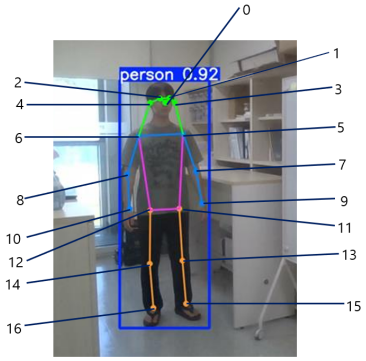
이러한 문제를 해결하기 위해, 2D 키포인트 데이터의 신뢰도를 높이기 위한 사용자 위치 보정 및 다층적 검증 알고리즘을 제안한다. 제안된 시스템은 의도된 운동 동작과 일상적인 움직임을 구분하는 필터링 메커니즘을 포함하여, 리소스가 제한된 환경에서도 안정적인 분석을 수행하도록 설계되었다.

이를 통해 저비용의 엣지 디바이스 상에서 딥러닝 모델을 활용한 실시간 자세 판정 시스템의 기술적 구현 가능성을 검증하고, 향후 정밀한 동작 분석 기술의 기반을 마련하고자 한다.

Ⅱ. 관 련 연 구

2.1 Human Pose Estimation

사람의 신체 관절 위치를 추정하는 HPE(Human Pose Estimation)는 컴퓨터 비전 분야에서 꾸준히 연구됐다. DeepPose[2]는 최초로 딥러닝 기반 회귀방식을 포즈 추정에 적용하여 입력 이미지로부터 직접 사람의 신체 관절 위치를 나타내는 키포인트를 예측함으로써 딥러닝 접근법의 가능성을 입증하였다. 이후 객체 탐지와 포즈 추정을 결합한 YOLO 계열의 경량 모델들이 등장하였으며, 그중 YOLOv8-pose[1]는 사람을 바운딩 박스로 검출한 뒤 [그림 1]과 같이 어깨, 엉덩이, 무릎과 같은 17개의 신체 주요 키포인트를 직접 회귀하는 방식으로 높은 정확도와 실시간 성능을 동시에 확보하였다. 이러한 특성은 운동 자세 분석 및 교정과 같은 실시간 피드백이 중요한 응용 분야에 적합하며, 본 연구에서는 YOLOv8-pose[1]기반 포즈 추정 모델을 활용하여 스쿼트 동작을 효과적으로 인식할 수 있다.



[그림 1] 각 키포인트 위치

2.2 경량 딥러닝 모델

YOLOv8-pose와 같은 딥러닝 기반 인체 자세 추정 모델은 높은 정확도를 제공하지만, 많은 파라미터와 복잡한 연산이 필요하다. 이러한 특성은 고성능 장비에서는 문제가 되지 않지만, Raspberry Pi와 같은 저성능 엣지 디바이스에서는 실시간 추론이 어렵다는 한계가 있다.

이를 해결하기 위해 모델의 크기와 연산량을 줄이는 경량화 기법이 연구[3]되고 있다. 대표적으로, 가중치의 정밀도를 낮추어 모델 크기와 속도를 개선하는 양자화, 불필요한 뉴런이나 연결을 제거해 구조를 단순화하는 가지치기(Pruning)가 있다.

본 연구에서는 이러한 경량화 기법이 적용된 YOLOv8-pose 모델을 활용하여, Raspberry Pi 환경에서도 실시간 자세 추정이 가능하도록 시스템을 구현하였다.

Ⅲ. 본 론

3.1 시스템 설계 및 구현

본 연구에서 제안하는 실시간 스쿼트 분석 시스템은 Raspberry Pi 환경에서 구현되었으며, 단일 2D 카메라와 YOLOv8-pose 모델을 기반으로 사용자의 스쿼트 수행 과정을 정량적으로 분석하고 그 유효성을 실시간으로 판별하는 것을 목표로 한다. 본 시스템에 사용된 실시간 포즈 추정 모델은 선행 연구[4]의 결과를 기반으로 최적화되었다. 시스템의 전체적인

구성은 1) 초기화 및 사용자 위치 보정, 2) 다층적 실시간 자세 검증, 3) 상태 머신 기반 동작 분석의 세 가지 핵심 모듈로 구성된다.

3.2 초기화 및 사용자 위치 보정

정확하고 일관된 분석을 위해 시스템은 시작 전, 사용자의 물리적 특성과 위치를 보정하는 초기화 단계를 거친다. 이는 개인별 신체 조건과 환경 차이를 표준화하여 분석의 신뢰도를 확보하기 위한 과정이다.

우선 IoU(Intersection over Union)를 이용해 사용자가 화면의 유효 영역 내에 있는지 확인하고, 스쿼트 분석에 필요한 주요 키포인트가 안정적으로 감지되는지 검증한다. 이후 2D 영상에서 무릎 각도를 원활히 측정할 수 있도록 약 45° 사선 뷰를 사용자가 설정하도록 보정이 이루어진다. 이 과정은 (1) 정면을 바라보게 하여 어깨너비를 측정하고 최댓값을 기준으로 저장한 뒤, (2) 현재 어깨너비와의 비율을 활용해 몸의 각도를 arccos 함수로 추정하며, (3) 추정된 각도를 목표 범위(40°~50°)에 맞추도록 안내하는 절차로 구성된다.

보정이 완료되면 스쿼트 깊이 측정을 위한 초기 엉덩이 높이, 각도 추정 및 발 너비 검증에 필요한 어깨너비 등 개인화된 기준점이 저장된다. 이 기준을 토대로 판정에 활용되는 임곗값들이 설정되며, 구체적인 값과 조건은 [표 1]과 같다.

항목	기준값	정의 및 측정 방식
상위 무릎 각도	160°	직립 자세에서 하강 시작을 판단하는 기준 각도 엉덩이(11/12)-무릎(13/14)-발목(15/16) 키포인트의 내각
하위 무릎 각도	130°	유효한 스쿼트 깊이 판별 기준 각도 (상위 무릎 각도와 동일한 키포인트로 계산)
최소 엉덩이 하강 거리	40px	최소 운동 범위 보장을 위한 엉덩이 절대 하강 거리 초기 자세 대비 엉덩이(11, 12) 키포인트의 y좌표 변화량
좌우 무릎 각도 차이	20°	좌우 균형 평가를 위한 절대 허용 각도 편차 좌측 키포인트: 11-13-15, 우측 키포인트: 12-14-16
발 너비 최소 비율	0.5	안전성을 위한 어깨너비 대비 발 너비 최소 비율 발목(15-16)과 어깨(5-6) 키포인트 간 수평 거리
발 너비 최대 비율	1.5	과도한 자세 방지를 위한 어깨너비 대비 발 너비 최대 비율 (최소 비율과 동일 방식)

[표 1] 스쿼트 자세 분석을 위한 주요 임곗값

3.3 다층적 실시간 자세 검증

시스템은 프레임마다 스쿼트 동작의 유효성을 우선순위에 따라 다층적으로 검증한다. 최우선으로, 사용자의 엉덩이 중심이 무릎 중심보다 낮아지는 'Sit Down' 동작을 감지하여 즉시 무효로 하는 정확성 검증을 수행한다. 이 동작이 감지되면 시스템은 다른 모든 분석을 중단하고 무효 상태로 전환되어 사용자에게 완전한 자세 재설정을 요구한다.

이 무효 상태에서 벗어나기 위해서 사용자가 단순히 일어서는 것을 넘어 완전한 초기 직립 자세로 복귀해야만 한다. 시스템은 사용자의 무릎 각도와 엉덩이 높이가 초기 준비 상태의 기준을 다시 만족했을 때만 무효 상태를 해제하고 분석을 재개한다. 이처럼 명확한 회복 조건을 둠으로써 잘못된 동작이 다음 동작 분석에 영향을 미치는 것을 차단한다. 이 정확성 검증을 통과한 경우에만 앞서 3.2절에서 정의한 임곗값들을 바탕으로 사용자의 기본자세를 순차적으로 검증한다.

3.4 상태 머신 기반 동작 분석

본 시스템은 스쿼트 동작의 흐름을 정밀하게 분석하기 위해 무효화 및 회복 메커니즘과 4단계 정상 동작 분석 로직이 결합한 이중 구조의 상태 머신을 구성하였다.

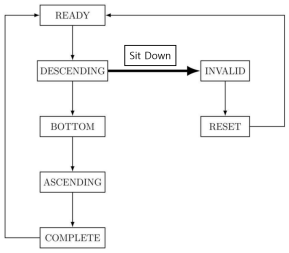
스쿼트 동작은 준비(READY), 하강(DESCENDING), 최저점(BOTTOM), 상승(ASCENDING)의 네 가지 상태로 정의된다. 각 상태는 사용자의 평균 무릎 각도와 개인화된 엉덩이 하강 깊이를 종합적으로 고려하였다.

READY 상태는 사용자가 스쿼트를 시작할 준비가 된 초기 직립 자세로, 무릎 각도가 일정 기준 이상일 때 유지된다. 이 상태에서는 안정적인 준비를 확인하기 위한 "Ready" 피드백이 제공된다. 여기서 무릎 각도가 기준치 아래로 줄어들면 DESCENDING 상태로 넘어간다.

DESCENDING 상태는 하강 단계로, 이 구간에서는 무릎 각도와 엉덩이 하강 깊이를 함께 확인한다. 하강 깊이가 충분하면 "Good!" 피드백과 함께 BOTTOM으로 전이되며 올바른 자세로 판단한다. 깊이가 부족할 경우 "Go deeper!" 피드백을 제공하여 더 깊은 자세를 유도하며, 조기 상승이 발생하면 "Too early!" 피드백과 함께 READY 상태로 복귀한다.

BOTTOM 상태는 동작 반전 지점으로, 무릎 각도가 다시 기준을 초과하기 시작하면 ASCENDING 상태로 전이된다. 이때는 "Up" 피드백을 제공하여 사용자가 상승 동작 개시를 인지할 수 있도록 한다.

ASCENDING 상태에서는 무릎 각도와 엉덩이 높이가 초기 조건을 충족할 때 한 동작이 완전히 수행된 것으로 인정되며 이에 따라 카운트가 증가한다. 성공적인 수행 시 "Complete!" 피드백과 시각적 체크 표시가 제공된다.



[그림 2] 시스템 흐름도

Ⅳ. 결 론

단일 카메라와 YOLO 기반의 실시간 포즈 추정 기술을 활용하여 사용자의 스쿼트 동작을 정밀하게 분석하고 교정하는 시스템을 설계·구현하였다. 제안된 시스템은 사용자의 신체 비율에 맞춰 개인화된 기준점을 설정하고, 다층적 검증 구조와 상태 머신을 통해 스쿼트 동작의 정확성, 균형 등을 종합적으로 평가한다. 특히 의도된 스쿼트 자세와 주저앉는 동작을 구분하는 동작 유효성 검증 로직을 도입하여 분석의 신뢰도를 높였다. 이러한 점에서 단순 횟수 측정을 넘어 수행 과정의 유효성을 정량적으로 판별하고 구체적인 피드백을 제공한다는 데 의의가 있다.

그러나 2D 영상 기반 분석에 따른 한계를 가진다. 예를 들어 긴 바지와 같은 의복 착용 시 무릎 검출 정확도가 저하되며, 3차원 동작을 2차원으로 투영하는 과정에서 발생하는 왜곡 때문에 관절 각도 계산에 오차가 발생한다. 이러한 한계로 판정 임곗값을 완화해야 했으며, 이는 시스템의 정밀도를 제한하는 요인으로 작용한다.

참 고 문 헌

[1] Ultralytics YOLOv8. (n.d.). Ultralytics. https://docs.ultralytics.com/ko/models/yolov8/

[2] A. Toshev and C. Szegedy, "Deeppose: Human pose estimation via deep neural networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1653-1660, 2014.

[3] Mittal, P. Mittal, "A comprehensive survey of deep learning-based lightweight object detection models for edge devices," Artificial Intelligence Review, vol. 57, article no. 242, Aug. 2024.

[4] 진경은, 신은호, 전윤호. *Raspberry Pi 환경에서의 포즈 추정 모델을 활용한 박수 게임*. 2025년도 한국통신학회 동계종합학술발표회 논문집, 한국통신학회, 2025, 쪽 1344-1345.

다계층 네트워크 분석 기반의 갑상선기능항진증 대응 후보 한약 및 핵심기전 도출

한예원[1], 이원융[2], 한석훈[2], 박찬림[3], 김상진[1]*
원광대학교[2], ㈜SSL[3], *한국한의약진흥원[1]

hanyewon2402@nikom.or.kr, wonyung21@wku.ac.kr, tjrgns031311@gmail.com, charlie@smartsafety.co.kr,
*kimsj@nikom.or.kr

Investigating candidate herbal medicines and key mechanisms for treating Hyperthyroidism based on a multi-scale network analysis

Yewon Han[1], Won-Yung Lee[2], Seokhoon Han[2], Chanlim Park[3], Sangjin Kim[1]*
Wonkwang Univ.[2], SSL Co., Ltd.[3], *National Institute for Korean Medicine Development[1]

요 약

본 연구는 주요 질환 중 하나인 갑상선기능항진증에 대해 한의약의 주요 치료 수단인 한약 후보들과 그 기전을 제시하는 것을 목표로 한다. 이를 위해 DrugBank와 같은 공개 데이터베이스로부터 각 한약 성분의 화합물 정보 및 표적 단백질 정보를 수집하였고, 이를 바탕으로 한약-성분-표적 네트워크를 구축하였다. 아울러, DisGeNET 데이터베이스에서 실험적으로 검증된 갑상선기능항진증 관련 유전자를 확인하여, 질환 관련 단백질 정보를 확보하였다. 이후, 다계층 네트워크에서 편향 랜덤 워크 알고리즘으로 확산 프로파일을 산출해 질환 프로파일 간의 상관계수와 초기하분포 검정($p<0.05$) 등을 통과기준으로 한약 후보를 선별하였다. 그 결과 현초·천마·여로·곤포·초두구·우슬 등 상위 10종 후보 한약이 도출되었고 서브 네트워크는 10개의 한약-235개의 표적과 이들 간의 360개의 상호작용으로 구성되었다. AKT1이 최빈 공통 표적이었으며 BAD, TSHR, SOD1, RELA, CASP, BAX, AR도 반복적으로 등장했다. 유전자 세트 농축 분석(GSEA)은 AGE-RAGE, apoptosis, C-type lectin receptor, TNF/IL-17, 파골세포 분화 및 지질/miRNA 조절 축을 핵심 기전으로 제시했다.

* Keyword: *network analysis, herbal medicines, random walk algorithm, disease-herb relationship*

Ⅰ. 서 론

갑상선기능항진증(Hyperthyroidism)은 과도한 갑상선 호르몬으로 인해 대사 항진과 심혈관·신경계 증상이 동반되는 질환으로, 유병률은 요오드 충분 지역에서 0.2-1.4%로 보고된다. 현재 치료(항갑상선제, 방사성 요오드, 수술)는 효과적이지만 장기 부작용과 재발, 기능저하증 위험 등 한계가 남아있어 보다 안전하고 지속 가능한 치료 대안이 요구된다[1].

한약은 다성분·다표적 특성으로 복합 병태를 표적화할 잠재력이 있으며, 전통적으로 요오드가 풍부한 해조·곤포 계열 본초가 갑상선 종대·자가면역 조절에 활용되어 왔다. 임상과 전임상 연구에서 Haizao Yuhu Decoction(HYD) 등은 증상 개선과 재발 감소, 항산화·면역조절 효과를 보여 대안 치료의 가능성을 시사하지만, 분자 수준의 핵심 기전은 여전히 충분히 규명되지 않았다[2].

시스템 생물학의 발전과 함께 네트워크 약리학이 다성분 약물-표적-질환 간 복잡 연결을 체계적으로 해석하는 유용한 틀을 제공하고 있으며, 멀티스케일 네트워크상의 편향 랜덤 워크 알고리즘(biased random walk algorithm) 기반 분석은 질환 맥락에서 약물 및 한약의 작용을 정량화하는 데 적합한 것으로 알려져 있다 [3,4]. 본 연구는 다계층(멀티스케일) 네트워크에서 편향 확산을 적용하여 갑상선기능항진증과의 확산 프로파일(diffusion profile) 유사도를 기반으로 유망 후보 한약을 선별하고, 한약-성분-표적 데이터베이스와 질환 표적 정보를 통합해 핵심 표적·경로를 도출함으로써, 갑상선기능항진증 대응 후보 한약의 핵심기전을 규명하는 것을 목표로 한다.

Ⅱ. 본론

본 연구는 한약-성분-표적(herb-ingredient-target) 네트워크를 구축하고, 다계층 네트워크에서 편향 랜덤워크 알고리즘을 바탕으로 한 편향 확산을 적용하여 갑상선기능항진증 대응 후보 한약의 핵심기전을 도출하였다. 먼저 KIOM(한국한의학연구원)이 운영하는 OASIS 전통의학 DB(접근일: 2024-08-21)에서 420개 한약과 4,786개 성분 간 12,871개의 연계를 수집하고, 각 성분을 PubChem CID로 표준화하였다. 성분-표적 상호작용은 DrugBank 5.0, TTD 2.0, STITCH 5에서 실험적/권위 DB 기반 정보를 취합하였고, SynGO를 이용해 유전자/단백질 식별자를 Entrez Gene ID로 정합하였다. 한약-성분, 성분-표적 간의 모든 간선은 방향성과 가중치 없이 정의하였으며, 표적과의 연계 성분이 3개 미만인 한약은 분석 신뢰도를 위해 제외하였다. 한약별 단순 경로 수(여러 성분이 하나의 표적에 수렴하는 경우 포함)를 기반으로 핵심 표적 집합을 산출하여 후속 분석의 입력으로 사용하였다.

질환 관련 표적은 DisGeNET(접근일: 2024-09-12)에서 전문가 검토(expert-curated)된 "갑상선기능항진증"과 연관된 유전자만 선별하였다. 동물 상동성(homology) 기반·문헌 텍스트마이닝 유래·치료 표식(tagged as therapeutic) 등의 데이터는 제외하여 질환 표적의 특이도와 신뢰성을 확보하였다. 이렇게 정제한 질환 표적 집합을 다계층 네트워크의 질환 계층으로 편입하였다.

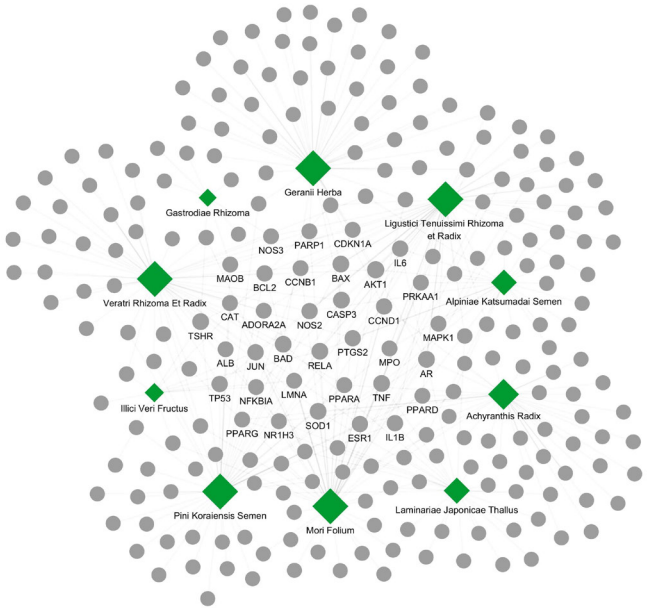
통합된 다계층 네트워크에서 질환과 각 한약에 대해 편향 확산을 수행하여 확산 프로파일을 계산한 후, 한약-질환 프로파일 간 유사도를 상관 점수로 산출하였다. 동시에 한약 표적과 질환 표적의 중첩에 대해 초기하분포 검정(hypergeometric test)을 실시하여 통계적 유의성을 평가하였다. 상관 점수가 높고 단백질 중첩 유의성($p<0.05$)을 만족하며, 질환 연관 활성 성분이 5개 이상인 한약을 우선 후보로 선정하였다. 그 결과, [표 1]과 같이 현초(0.0208)가 가장 높은 상관 점수를 보였고, 천마(0.0115), 여로(0.0106), 곤포(0.0104), 초두구(0.0098), 우슬(0.0097) 등이 뒤를 이었다.

[표 1] 갑상선기능항진증에 대한 후보 한약 목록(상위 10종 선별)

Herb Name (Latin(국명))	Correlation Score	Overlap (p-value)	Enrichment
Geranii Herba(현초)	0.0208	2/50 (0.006)	16.90
Gastrodiae Rhizoma(천마)	0.0115	1/8 (0.018)	52.83

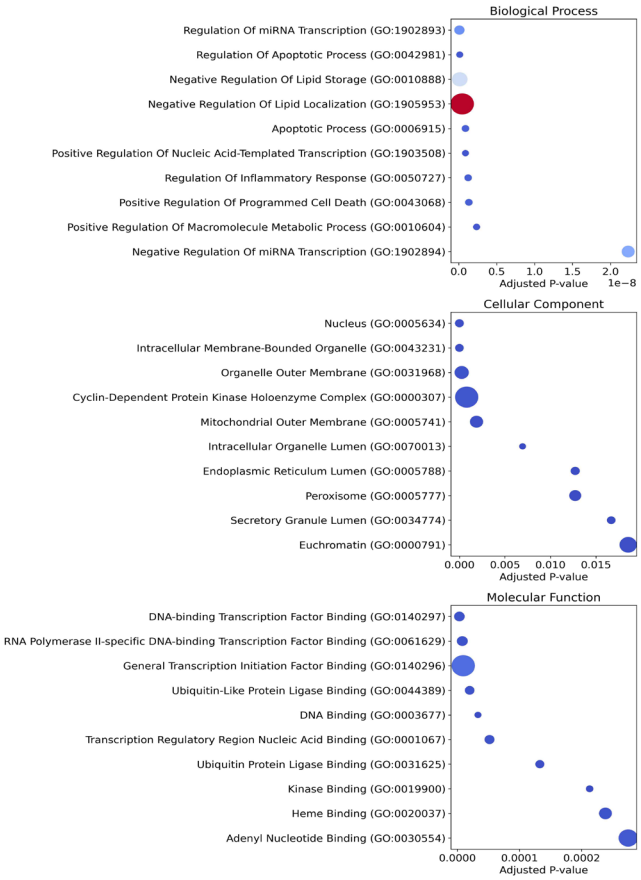
Veratri Rhizoma Et Radix(여로)	0.0106	5/50 (6.63E-8)	42.27
Laminariae Japonicae Thallus(곤포)	0.0104	2/27 (0.001)	31.31
Alpiniae Katsumadai Semen(초구두)	0.0098	2/26 (0.001)	32.51
Achyranthis Radix(우슬)	0.0097	2/38 (0.003)	22.24
Illici Veri Fructus(팔각화향)	0.0095	1/11 (0.025)	38.42
Ligustici Tenuissimi Rhizoma et Radix(천궁)	0.0094	4/50 (4.24E-6)	33.81
Mori Folium(뽕잎)	0.0091	4/50 (4.24E-6)	33.81
Pini Koraiensis Semen(잣)	0.0090	4/50 (4.24E-6)	33.81

네트워크 시각화 도구로 Cytoscape(v3.10.2) 소프트웨어를 사용하여 상위 10종 후보에 대해 한약-성분-표적 네트워크를 시각화한 결과는 [그림 1]과 같다. 이는 '10개 한약(초록색)' - '235개 표적(회색)' - '360개 상호작용(간선)'으로 구성되며, 3개 이상의 한약이 공통으로 겨냥하는 표적은 35개이다. 특히 AKT1은 6개 한약의 공통 표적이었으며, BAD, TSHR, SOD1, RELA, CASP 계열, BAX, AR 등은 각각 5개 한약에서 교차 표적으로 확인되어, 이들 중심 표적이 다중 표적성에 기반한 치료 기전의 한약재로서 기능할 가능성을 보여주었다.



[그림 1] 갑상선기능항진증 후보 한약 네트워크

후보 한약의 핵심 표적에 대한 유전자 세트 농축 분석(GSEA; Gene Set Enrichment Analysis)은 Python(GSEApy 1.1.3) - Enrichr을 사용하였고, GO(Gene Ontology, 유전자 온톨로지) 분석 결과는 [그림 2]와 같다. BP(Biological Process, 생물학적 과정)에서는 지질 분포, miRNA 전사 조절, LPS(lipopolysaccharide, 리포 다당류)/지질 자극에 대한 염증성 세포 반응이 두드러졌고, CC(Cellular Commponent, 세포 성분)에서는 핵·막성 소기관(미토콘드리아, 퍼옥시좀)과 cyclin-dependent kinase 복합체가 주요 소속으로 나타났다. MF(Molecular Function, 분자 기능)에서는 일반 전사 개시 인자 결합이 높은 결합 강도를 보여 전사 조절 축이 핵심 표적군의 기능적 축으로 부각되었다. 이러한 결과는 면역-염증 신호, 세포 사멸/생존, 골 항상성 등 갑상선기능항진증의 병태생리(Pathophysiology)와 맞닿은 경로에서 후보 한약이 다표적으로 개입할 수 있음을 시사한다.



[그림 2] 갑상선기능항진증 후보 한약 표적의 유전자 온톨로지

Ⅲ. 결론

한약-성분-표적 네트워크와 편향 확산 기반 다계층 분석을 통해 질환-한약 유사도와 표적 중첩의 통계적 근거를 동시에 확보하여 상위 후보군의 공통/중심 표적 및 관련 경로를 체계적으로 도출하였다. 본 결과는 갑상선기능항진증에 대한 다표적 한약 치료 전략의 우선순위를 제시하고, 향후 성분 단위 기전 검증과 전임상·임상 평가를 위한 정량적 근거를 제공한다.

ACKNOWLEDGMENT

This research was supported by a grant from the Project of National Institute for Korean Medicine Development, Building Artificial Intelligence platform of Korean Medicine; Based on Experimental Information of KM, funded by the Korean Ministry of Health and Welfare.

참 고 문 헌

[1] Kim K, Park Y, Lee Y, et al. Treatment limitations of hyperthyroidism: antithyroid drugs, radioactive iodine therapy, and surgery. Korean J Intern Med. 2023;98(3):132 140.

[2] Song X H, Zan R Z, Yu C H, Wang F. Effects of modified Haizao Yuhu Decoction in experimental autoimmune thyroiditis rats.Journal of Ethnopharmacology. 2011;135(2):321 324. doi:10.1016/j.jep.2011.03.017.

[3] Ruiz, Camilo, Marinka Zitnik, and Jure Leskovec. Identification of disease treatment mechanisms through the multiscale interactome. Nature communications 12.1 (2021): 1796.

[4] Lee, Won-Yung, et al. Evaluating current status of network pharmacology for herbal medicine focusing on identifying mechanisms and therapeutic effects. Journal of Advanced Research (2024).

딥러닝 기반 노이즈 제거를 통한 장기 혈당 예측 안정성 향상

김수진, 김성호, 정성윤*
영남대학교
sujin38@yu.ac.kr, sunghokim@ynu.ac.kr, *syjung@ynu.ac.kr

Deep Learning-enhanced Noise Reduction for Robust Long-term Blood Glucose Prediction

Kim Su Jin, Kim Sung Ho, Jung Sung Yoon*
Yeungnam Univ.

요 약

당뇨병 환자의 장기 혈당 예측은 인공췌장 시스템(Artificial Pancreas System, APS)의 안정적인 인슐린 조절을 위해 필수적이다. 그러나 연속혈당측정(Continuous Glucose Monitoring, CGM) 데이터는 센서 노이즈를 포함하고 있어 예측 성능 저하를 야기한다. 본 연구에서는 딥러닝 기반 CNN-Autoencoder를 활용하여 CGM 데이터의 노이즈를 효과적으로 제거하고, 이를 바탕으로 장기 혈당을 예측하는 기법을 제안한다. 실험 결과, 제안한 기법은 기존의 전통적인 필터링 방법보다 우수한 노이즈 제거 성능을 보였으며, 24시간 혈당 패턴을 안정적으로 추정하였다. 특히 무작위 시나리오에서도 실제 혈당 추세를 잘 추종하는 결과를 보이며, 제안 기법이 장기 혈당 예측의 안정성과 신뢰성을 향상시킬 수 있음을 확인하였다. 본 논문은 선행 연구 [1]을 기반으로 주요 내용을 요약·정리하였다.

Ⅰ. 서 론

전 세계 당뇨병 환자 수가 증가함에 따라 효과적인 혈당 관리의 중요성이 점점 커지고 있으며, 이에 따라 정확한 혈당 예측 모델 개발이 활발히 진행되고 있다[2]. 특히 인공췌장 시스템(Artificial Pancreas System, APS)은 연속혈당측정(Continuous Glucose Monitoring, CGM) 센서, 인슐린 펌프, 그리고 제어 알고리즘으로 구성되어 환자의 혈당을 자동으로 조절하는 장치로 연구되고 있다. 효과적인 혈당 관리를 위해서는 CGM 센서를 활용한 연속 모니터링과 더불어 예측 기반 기술이 필수적이다. 그러나 CGM 데이터는 센서의 한계로 인해 노이즈가 포함되는 경우가 많아 혈당 예측의 정확도를 저하시킨다[3]. 따라서 이러한 노이즈를 효과적으로 완화하여 예측 정확성을 높이는 것은 중요한 연구 과제 중 하나이다.

이러한 문제를 해결하기 위해 본 연구에서는 CNN-Autoencoder를 활용하여 CGM 데이터의 노이즈를 제거하고, 이를 바탕으로 장기 혈당 패턴을 안정적으로 예측하는 기법을 제안한다[1].

Ⅱ. 제안 기법 및 성능 비교

2.1 CNN-Autoencoder 기반 노이즈 제거

CGM 데이터는 센서 특성상 노이즈가 포함되어 이어, 혈당 예측 정확도 저하의 주요 원인이 된다. 본 논문에서는 CNN-Autoencoder 구조를 적용하여 노이즈를 효과적으로 제거하였다. CNN-Autoencoder는 인코더와 디코더로 구성된 신경망으로, 인코더는 입력 데이터를 저차원 잠재 공간으로 압축하면서 핵심 특징을 추출하고, 디코더는 이를 다시 복원하는 과정에서 불필요한 노이즈를 제거한다[4]. 본 연구에서는 CGM 데이터를 2차원 구조로 변환함으로서 공간적·시간적 패턴을 동시에 학습할 수 있도록 하였다. 그림 1은 기존의 Median filter와 제안한 CNN-Autoencoder의 노이즈 제거 결과를 비교한 것으로, 딥러닝 기반 접근법이 기존 선형 필터링 방식보다 우수한 성능을 보임을 확인할 수 있다. 또한, 표 1은 본 연구에서 사용된 주요 실험 파라미터를 정리한 것이다.

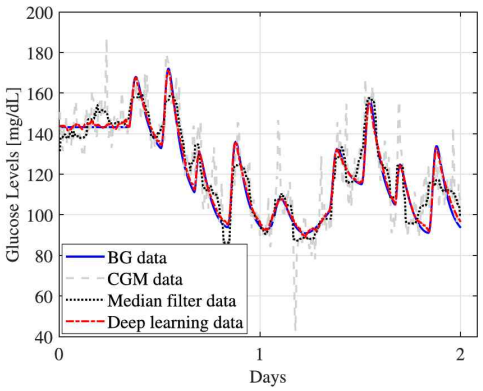


그림 1. 노이즈 제거 결과 비교

Parameter	Value
Total data	50,000
Epoch	30
Batch size	128
Optimizer	Adam
Loss	mse

표 1. Experimental parameters

2.2 Blood Glucose Pattern Estimator

딥러닝 기반 혈당 예측 모델들은 장기 예측 구간에서 오차가 누적되면서 성능이 점차 저하된다. 이를 보완하기 위해 본 연구에서는 혈당 변화를 선형 시스템 응답 함수로 모델링한 Blood Glucose Pattern Estimator (BGPE)를 제안하였다[1]. BGPE는 CHO 섭취와 인슐린 투여를 입력으로 받아 혈당 상승과 감소 효과를 동시에 반영하며, 이를 통해 24시간 혈당 패턴을 안정적으로 추정할 수 있다. 또한, 단순한 구조를 바탕으로 해석 가능성이 높고, 오차 전파를 최소화하여 장기 예측에서 기존 딥러닝 기반

모델보다 안정적인 성능을 기대할 수 있다.

BGPE는 입력 신호 $\mathbf{C}$를 섭취한 탄수화물(CHO), 시스템 모델 $\mathbf{h}$를 인체의 혈당 반응 (식사로 인한 상승과 인슐린 작용에 의한 감소)으로 모델링한다. 이때 혈당 변화 $\mathbf{G}$와의 관계는 다음과 같다.

$$\mathbf{G} = \mathbf{D}(\mathbf{C})\mathbf{F}_L\mathbf{h} + \mathbf{N} \quad (1)$$

여기서 $\mathbf{N}$은 측정 및 모델링 노이즈를 의미한다. $\mathbf{F}_L$은 Discrete Furier Transform 행렬이다.

이러한 정의를 바탕으로 주파수 영역에서 결합된 확률 변수 $\mathbf{Z}$를 구성하고, 추정된 $\hat{\mathbf{h}}$는 최대우도추정(MLE)을 통해 다음과 같이 계산된다[1].

$$\hat{\mathbf{h}} = ((\mathbf{D}(\mathbf{C}^H\mathbf{C})\mathbf{F}_L)^H\mathbf{D}(\mathbf{C}^H\mathbf{C})\mathbf{F}_L)^{-1}(\mathbf{D}(\mathbf{C}^H\mathbf{C})\mathbf{F}_L)^H\mathbf{Z} \quad (2)$$

최종적으로 추정된 $\hat{\mathbf{h}}$를 통해 장기 혈당을 추정할 수 있다.

2.3 모의 실험

본 연구에서는 FDA에서 승인한 UVA/Padova 시뮬레이터 기반의 simglucose[5]를 이용하여, 10명의 가상 T1DM 환자 데이터를 대상으로 Custom 시나리오(규칙적인 CHO 섭취)와 Random 시나리오(불규칙적 CHO 섭취)를 설정하고 모델의 예측 성능을 평가하였다.

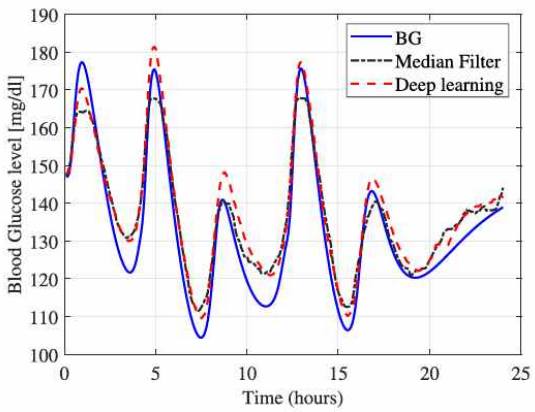


그림 2. Custom 시나리오에서의 혈당 예측 결과

Method	Custom scenario	Random scenario
Raw data	2.495	4.536
Median Filter	0.953	1.833
DL Filter	0.495	1.209

표 2. RMSE 비교 결과

그림 2는 Custom 시나리오에서의 예측 결과를 나타내며, 딥러닝 기반 노이즈 제거가 Medain filter보다 원 신호에 더 가깝게 혈당 곡선을 복원함을 보여주며 24시간 장기 혈당 예측을 잘 수행하는 것을 알 수 있다. 표 2는 RMSE 비교 결과이며 딥러닝 필터가 Median filter 대비 크게 RMSE를 감소하는 것을 볼 수 있다.

그림 3은 Random 시나리오에서 기존 딥러닝 모델(LSTM[6], Nonstationary Traformer[7])과의 비교 결과이다. BGPE는 불규칙한 환경에서도 오차 누적을 최소화하여 실제 혈당 변동을 안정적으로 반영하며, 장기 예측에서 기존 모델 대비 우수한 성능 보인다.

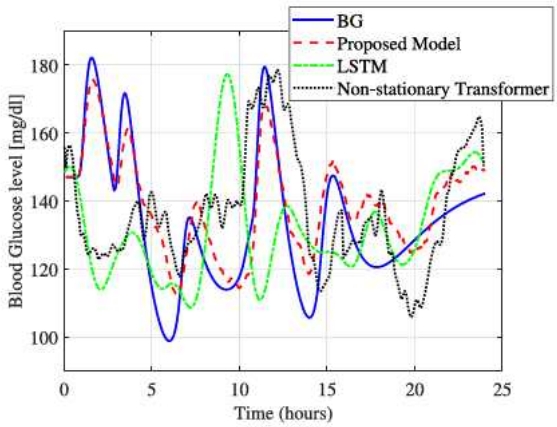


그림 3. Random 시나리오에서 BGPE와 딥러닝 기반 모델의 장기 예측 비교

Ⅲ. 결론

본 연구에서는 CGM 데이터의 노이즈를 제거하기 위해 CNN-Autoencoder를 적용하고, 장기 혈당 예측을 위해 BGPE를 제안하였다. 실험 결과, 제안 기법은 기존 Median filter 대비 RMSE가 크게 감소하였으며, Custom 및 Random 시나리오 모두에서 기존 딥러닝 기반 모델보다 안정적인 성능을 보였다. 향후 연구에서는 실제 환자의 CGM 데이터를 적용하여 알고리즘의 임상적 타당성을 입증하고, 다양한 환자군 및 상황에 대한 확장 연구를 진행할 예정이다.

ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. RS-2023-00219725).

참 고 문 헌

[1] Kim, Su-Jin, Jun Sung Moon, and Sung-Yoon Jung. "Long-term blood glucose prediction using deep learning-based noise reduction." ICT Express (2025).

[2] Zimmet, Paul Z., et al. "Diabetes: a 21st century challenge." The lancet Diabetes & endocrinology 2.1 (2014): 56-64.

[3] G. Sparacino, A. Facchinetti, C. Cobelli, 'Smart' continuous glucose monitoring sensors: online signal processing issues, Sensors 10 (7) (2010) 6751 - 6772.

[4] S.S. Roy, et al., A robust system for noisy image classification combining denoising autoencoder and convolutional neural network, Int. J. Adv. Comput.Sci. Appl. 9 (1) (2018) 224 - 235.

[5] J. Xie, Simglucose v0.2.1, 2018, [Online]. Available: https://github.com/jxx123/ simglucose.

[6] Md. Fazle Rabby, et al., Stacked LSTM based deep recurrent neural network with kalman smoothing for blood glucose prediction, BMC Med. Inform. Decis. Mak. 21 (1) (2021) 1 - 15.

[7] Yong Liu, et al., Non-stationary transformers: Exploring the stationarity in time series forecasting, Adv. Neural Inf. Process. Syst. 35 (2022) 9881 - 9893.

AI와 인간의 신뢰 관계: 심리학적 관점에서의 분석

정주영

jjy20051121jung@gmail.com

TheTrustRelationshipBetweenAIandHumans:AnAnalysisfromaPsychologicalPerspective

JungJooyoung

Independent Researcher

요 약

본 연구는 인공지능(AI)과 인간의 신뢰 관계를 심리학적 관점에서 탐구하였다. AI가 의료, 금융, 교육 등 핵심 의사결정 영역에 확대 적용되고 있으나, 사용자가 AI를 언제 신뢰하고 언제 불신하는지는 충분히 규명되지 않았다. 이에 본 논문은 이론적 검토와 실제 사례 분석을 바탕으로, 시나리오 기반 실험을 설계하여 신뢰 형성 요인을 검증하였다.

실험은 300명의 참가자를 대상으로 설명 가능성, 감정 표현, 전문가 협력 여부를 독립변수로 설정하고 신뢰도 및 선택 행동을 측정하였다. 분석 결과, ① 설명 가능한 AI가 그렇지 않은 경우보다, ② 감정 표현을 하는 AI가 중립적 표현을 하는 경우보다, ③ 전문가와 협력적으로 제시된 AI가 독립적으로 제시된 경우보다 유의미하게 높은 신뢰를 얻는 것으로 나타났다.

이러한 결과는 신뢰 형성에 있어 인지적 요인(설명 가능성, 전문가 협력)과 정서적 요인(감정 표현)이 모두 중요한 역할을 함을 시사한다. 본 연구는 인간-AI 협력 관계 설계에서 단순 성능 향상이 아닌 설명 가능성 강화, 전문가 보조적 구조, 사회적 존재감 고려가 필요함을 강조한다. 이는 향후 AI 개발 및 정책 수립에 중요한 심리학적·윤리적 함의를 제공한다.

1. 서론

오늘날 AI는 인간의 중요한 의사결정 과정에 깊숙이 자리 잡고 있다. 그러나 기술적 발전이 곧바로 신뢰로 이어지지는 않는다. 예컨대, 의료 현장에서 AI 진단과 의사의 진단이 충돌할 때 환자는 혼란에 빠지고, 금융권에서는 불투명한 AI 대출 심사 결과로 고객의 불신이 커지고 있다. 따라서 본 연구는 “인간은 어떤 조건에서 AI를 신뢰하는가?”라는 질문에 초점을 맞추어 심리학적 요인을 규명하고자 한다.

2. 이론적 배경

2.1 신뢰의 개념 및 선행연구

심리학에서 신뢰(trust)는 불확실성과 위험을 감수하면서 타인의 의도를 긍정적으로 기대하는 심리적 상태로 정의된다(Mayer et al., 1995). 신뢰는 크게 두 가지 차원으로 구분된다.

인지적 신뢰(Cognitive trust): 능력과 전문성에 기반한 신뢰

정서적 신뢰(Affective trust): 감정적 유대와 관계적 상호작용에 기반한 신뢰

최근 연구들은 AI 신뢰 형성 과정에서 인간-자동화 상호작용(Human-Computer Interaction, HCI) 관점의 중요성을 강조하고 있다. Lee와 See(2004)는 신뢰가 단순히 시스템 성능에 의해 결정되는 것이 아니라, 시스템의 투명성, 피드백 제공 방식, 사용자 통제 가능성에 의해 크게 영향을 받는다고 지적하였다. Hoff와 Bashir(2015)는 실증적 연구들을 종합하여, 자동화에 대한 신뢰 요인을 인지적 차원(성능, 신뢰성)과 정서적 차원(사회적 존재감, 상호작용 품질)으로 구분하였다.

이러한 최근 연구 흐름은 본 연구가 제안하는 설명 가능성, 전문가 협력, 감정 표현 요인의 중요성을 뒷받침한다. 즉, 사용자는 단순한 결과 제시보다, AI가 판단 근거를 제공하고, 전문가와 협력하며, 사회적·정서적 존재감을 갖춘 경우 더 높은 신뢰를 형성한다는 점을 시사한다.

2.2 AI 신뢰 형성 요인

전문가 협력성: AI 단독보다 전문가와 협력할 때 신뢰가 높음

설명 가능성(Explainability): 결과뿐 아니라 판단 근거 제공 여부

사회적 존재감(Social Presence): 감정 표현, 인간적 태도, 상호작용

2.3 사건 사례

IBM Watson Oncology: 설명 부족으로 의료 현장에서 신뢰 붕괴

자율주행차 사고(테슬라, 우버): 책임 문제와 신뢰 손상

금융 신용평가 AI: 불투명성으로 인한 고객 불신

3. 연구 가설

H1: 설명 가능한 AI는 설명이 없는 AI보다 높은 신뢰를 얻을 것이다.

H2: 감정 표현을 하는 AI는 중립적 표현을 하는 AI보다 높은 신뢰

를 얻을 것이다.

H3: 전문가와 협력적으로 제시된 AI는 독립적으로 제시된 AI보다 높은 신뢰를 얻을 것이다.

4. 연구 방법

4.1 참가자

본 연구는 국내 온라인 패널 모집 플랫폼을 통해 300명의 성인 참가자를 모집하였다(남 150명, 여 150명, 연령 20~50대). 참가자는 연령과 성별 비율이 균형을 이루도록 층화 표집(stratified sampling) 방식으로 선발되었으며, 각 실험 조건에 무작위 할당(random assignment) 되었다.

4.2 실험 설계

시나리오 기반 설문 실험을 수행하였다. 각 참가자는 의료, 금융, 교육 등 다양한 상황에서 AI와 인간 전문가의 판단을 비교하는 시나리오를 경험하였다.

의료 상황: AI 진단 vs 의사 진단

금융 상황: AI 대출 승인 결과 vs 은행 직원 설명

교육 상황: AI 튜터 조언 vs 교사 조언

4.3 변수 조작

독립변수는 다음과 같이 설정하였다.

설명 제공 여부: 제공 vs 미제공

감정 표현 여부: 격려 포함 vs 중립적

전문가 협력 여부: 협력 vs AI 단독

4.4 측정

신뢰도 평가: 7점 리커트 척도

선택 행동: AI 판단 vs 인간 판단

정성적 의견: 참가자 자유 응답

4.5 통계 분석

신뢰도 점수: 조건별 차이를 검증하기 위해 일원분산분석(One-way ANOVA)을 수행하고, 사후분석(Post-hoc test)으로 Tukey 검정을 실시하였다.

선택 행동 비율: 카이제곱 검정(χ^2 test)을 통해 조건별 유의성을 평가하였다.

모든 통계 분석은 유의수준 $\alpha = 0.05$에서 수행하였다.

이와 같이 명확히 표본, 무작위 할당, 변수 조작, 통계 분석 기법을 제시함으로써 연구 재현성을 높였다.

5. 실험 결과 시각화

5.1 신뢰 점수

조건	평균 신뢰 점수	표준편차
설명 제공	5.8	0.9
설명 미제공	3.9	1.1
감정 표현	5.5	1.0
중립 표현	4.2	1.1
전문가 협력	6.0	0.8
AI 단독	4.1	1.2

신뢰 점수 설명:

설명 제공, 감정 표현, 전문가 협력 조건에서 신뢰 점수가 유의하게 높게 나타났다($p < .05$~.001).

점수에서 세 가지 요인이 모두 신뢰를 높이는 효과를 확인할 수 있음.

신뢰 점수로 제시됨.

5.4 선택 행동 결과

설명 제공 AI: 72% 선택

설명 없는 AI: 38% 선택

감정 표현 AI: 65% 선택

전문가 협력 AI: 80% 선택

점수표로 나타냈을 때 전문가 협력 조건이 가장 높게 나타남.

6. 사례 분석

의료 사례: 서울 소재 A대학병원에서는 폐렴 진단을 보조하기 위해 AI 기반 영상 판독 시스템을 도입하였다. 해당 시스템은 흉부 X-ray 영상을 분석하여 폐렴 여부를 자동 분류하는 기능을 갖추고 있었으나, 실제 임상 적용 과정에서 오진 사례가 발생하였다. 사례는 50대 환자가 고열과 기침 증상을 보이며 응급실을 방문한 경우였다. AI 시스템은 환자의 영상 데이터를 '비폐렴'으로 판정하였으나, 담당 의사는 임상 경험과 혈액검사 결과를 근거로 폐렴 가능성이 높다고 판단하였다. 이후 CT 검사를 통해 의사의 판단이 옳았음이 입증되었다.

이 과정에서 환자는 처음에 AI의 진단 결과에 의문을 가지지 않았으나, 의사가 직접 "영상에서는 뚜렷하게 보이지 않지만 혈액 지표와 임상 증상을 종합하면 폐렴일 가능성이 높다"고 설명하자 의사의 의견을 우선적으로 신뢰하였다. 환자의 진술에 따르면, "AI가 왜 그렇게 진단했는지를 설명해주지 못했기 때문에 결국 의사의 판단을 믿을 수밖에 없었다"는 점이 신뢰 결정의 핵심 요인이었다. 이 사례는 다음과 같은 심리학적 시사점을 제공한다.

전문가 권위의 우선성: 의료 상황에서 환자는 AI보다 의사의 설명과 판단을 더 신뢰한다.

설명 가능성 부족의 한계: AI가 근거를 설명하지 못할 경우, 환자의 신뢰는 급격히 감소한다.

인지적 신뢰의 기반: 환자의 신뢰는 단순 결과보다 합리적 이유 제시에 의해 강화된다.

따라서 의료 영역에서 AI의 신뢰 확보를 위해서는, 의사의 전문성 보완 도구로 제시되고, 동시에 환자와 의료진 모두가 이해할 수 있는 설명 가능한 진단 근거를 제공해야 한다.

금융 사례: 국내 B은행은 신용평가 업무의 효율성과 공정성을 높이기 위해 AI 기반 대출 심사 시스템을 도입하였다. 이 시스템은 고객의 소득, 신용 이력, 거래 패턴, 소비 성향 등을 종합적으로 분석해 대출 승인 여부를 결정하였다. 그러나 실제 운영 과정에서 고객 불만과 불신이 증가하는 사례가 보고되었다.

30대 직장인 C씨는 안정적인 직장에 근무하며 연체 이력이 없음에도 불구하고, 주택담보대출 신청이 거절되는 상황을 겪었다. AI 시스템은 단순히 "대출 불가"라는 결과만을 반환했으며, 구체적인

근거는 제시하지 않았다. 이에 대해 은행 직원에게 문의했으나, 직원 역시 “AI 시스템이 내부 기준에 따라 거절한 것이라 이유를 알 수 없다”는 답변을 할 수밖에 없었다.

C씨는 “성실하게 금융 거래를 해왔는데도 이유조차 알 수 없는 거절은 납득할 수 없다”며 불공정함을 강하게 호소했다. 결국 그는 은행의 신용평가 체계 전반을 불신하게 되었고, 다른 금융기관으로 이탈하는 결과를 낳았다.

이 사례는 금융 맥락에서 신뢰 형성과 불신 확산을 잘 보여준다. 설명 가능성 부족: 결과만 제시되고, 판단 근거가 불투명할 경우 고객의 신뢰는 약화된다.

절차적 공정성(perceived fairness): 고객은 결과 자체보다 왜 그런 결정을 내렸는가를 중시한다.

인지적 신뢰 붕괴: 설명 부재는 단순한 불만을 넘어, 금융 시스템 자체에 대한 불신으로 이어진다.

따라서 금융 분야에서 AI 신뢰를 확보하기 위해서는, 고객이 납득할 수 있는 설명 가능한 기준 제시, 사람 직원의 보완적 역할, 투명한 피드백 절차가 반드시 병행되어야 한다.

교육 사례: 서울 소재 A대학의 온라인 강의 플랫폼에서는 AI 기반 튜터 챗봇이 학습 지원에 활용되었다. 챗봇은 단순히 문제 풀이 힌트만 제공하는 것이 아니라, 학습 과정에서 불안을 표현하는 학생에게 “괜찮아요, 지금처럼 차근차근하면 충분히 잘할 수 있습니다”, **“조금 어려워도 정상적인 과정이에요. 계속 시도해 보세요”**와 같은 격려성 피드백을 제시하였다.

이를 경험한 학습자들은 사후 설문에서 “AI가 나를 이해하고 응원해 준다”,

“실패해도 다시 시도할 수 있다는 자신감을 얻었다”라는 반응을 보였다. 특히, 격려 발언을 제공받은 집단은 그렇지 않은 집단에 비해 시험 직전 불안 수준이 평균적으로 낮았으며(평균 차이 약 0.8점, $p<.05$), 과제 지속 시간도 더 길게 유지되었다.

이 사례는 AI의 단순 정보 제공 기능을 넘어, 정서적 교류와 심리적 안정감 제공이 신뢰 형성에 중요한 역할을 한다는 점을 시사한다. 학습자는 AI를 단순한 기계가 아니라, “나를 도와주는 사회적 파트너”로 인식하게 되었으며, 이는 신뢰 관계 구축으로 이어졌다.

법률 사례: 국내 모 사법연구기관에서는 판결 보조 AI를 시범 도입하여 판례 검색, 양형(刑量) 추천, 유사 사건 비교 등을 지원하였다. 해당 시스템은 과거 수천 건의 판례 데이터를 기반으로 특정 사건에 대해 “통상적으로 ○○년~○○년의 형량이 선고됨”이라는 결과를 제시하였다.

실험적으로 진행된 설문조사에서, 법률 전문가 집단(판사·변호사)은 AI의 추천을 일정 부분 참고했으나 “사건 맥락과 정황은 기계적으로 해석할 수 없다”며 최종 신뢰 수준이 중간 정도(평균 3.2/5)에 머물렀다. 반면, 일반 시민 집단은 동일한 결과를 “객관적이고 공정하다”라고 평가하며 더 높은 신뢰(평균 4.1/5)를 나타냈다.

이처럼 전문가와 비전문가 간 신뢰 격차가 발생한 이유는 다음과 같다.

전문가 집단: AI의 한계를 잘 알기에, “법리적 판단은 데이터 패턴만으로는 설명할 수 없다”는 불신 요인이 작용.

일반 시민 집단: 오히려 AI의 수치적·데이터 기반 판단을 ‘공정성’으로 인식하며, 인간 판사보다 편향이 적다고 기대.

이 사례는 AI 판결 보조 시스템이 법적 의사결정 과정에서 신뢰를 획득하는 방식이 집단에 따라 상반될 수 있다는 점을 보여준다. 따라서 향후 법률 AI 설계에서는 전문가가 활용할 수 있는 투명한 설명 제공과 함께, 일반 시민에게는 신뢰의 근거가 오해되지 않도록 교육적 장치가 필요하다.

7. 논의

본 연구는 인간이 AI를 신뢰하는 데 있어 인지적 요인(설명 가능성, 전문가 협력)과 정서적 요인(감정 표현)이 모두 중요함을 밝혔다. 특히 전문가 협력성은 가장 강력한 신뢰 요인으로 나타났으며, 이는 AI가 독립적 의사결정자가 아니라 보조적·협력적 파트너로 설계될 필요성을 시사한다.

구체적으로, 의료 영역에서 환자는 AI의 진단만 제시될 때보다 “의사가 검토·승인한 AI의 결과”일 때 더 높은 신뢰를 보였다. 이는 AI가 전문가의 판단을 보완하는 형태일 때, 사용자가 이를 “공인된 정보”로 받아들이기 때문이다. 금융 영역에서도 마찬가지로, AI의 대출 거절 결과가 은행 직원의 설명과 함께 제공될 경우, 고객은 시스템을 더 공정하고 투명하다고 인식하였다.

따라서 AI-전문가 협력 모델은 단순한 보조적 기능을 넘어, 신뢰 확보를 위한 구조적 설계 원리로 자리매김할 수 있다. 예컨대:

의료 분야: “AI 제안 → 전문가 검증 → 환자 전달”이라는 3단계 절차

법률 분야: “AI 판례 추천 → 변호사·판사의 최종 판단” 구조

교육 분야: “AI 튜터 설명 → 교사의 피드백 보완” 방식

이와 같이 전문가 협력이 내재된 구조는 단순히 가능성이 아니라 실질적으로 구현 가능한 설계 원리이며, 향후 AI 신뢰 연구와 실무 적용에 있어 필수적 조건으로 작동할 것이다.

8. 결론 및 향후 연구

본 연구는 인간-AI 신뢰 관계의 형성을 설명하는 핵심 요인으로 전문가 권위, 설명 가능성, 사회적 존재감을 도출하였다.

이를 통해 AI가 단순한 정보 제공자가 아니라, 인간의 의사결정과 심리적 안정에 영향을 미치는 상호작용 주체라는 점을 확인하였다. 따라서 향후 AI 설계에서는 기술적 성능뿐 아니라 심리적·윤리적 설계 원리가 함께 고려되어야 한다.

그러나 본 연구는 제한된 시나리오와 사례 분석에 기반했기 때문에, 보다 일반화된 결론을 위해 다음과 같은 후속 연구가 필요하다.

첫째, 문화권별 비교 연구가 요구된다. 신뢰의 심리적 요인은 문화적 맥락에 따라 다르게 작용할 수 있으며, 집단주의 문화에서는 “전문가 권위” 요인이 더 강력하게 작동하는 반면, 개인주의 문화에서는 “설명 가능성”이 더 중요한 요인으로 작용할 수 있다. 이를

검증하기 위해 동일한 시나리오 설문을 한국, 미국, 일본 등 다양한 문화권에서 실시하고, 문화별 차이를 다집단 구조방정식 모형(Multi-group SEM)으로 비교 분석할 수 있다. 또한 설문조사에 더해 포커스 그룹 인터뷰와 같은 질적 연구를 병행하여 문화적 해석의 차이를 심층적으로 파악하는 것도 효과적이다.

둘째, 장기적 상호작용에 따른 신뢰 변화를 추적할 필요가 있다. 단기적 사용에서는 “신기효과(novelty effect)”가 나타날 수 있으나, 장기간 사용 시에는 초기 신뢰가 감소하거나 유지되는 다양한 패턴이 존재할 수 있다. 이를 검증하기 위해 6개월에서 1년간 패널 데이터를 수집하고, 일정 시점별(예: 1개월, 3개월, 6개월, 12개월) 반복 측정 설문을 실시할 수 있다. 또한 잠재 성장 모형(Latent Growth Model)과 군집 분석을 활용해 신뢰의 변화 궤적과 사용자 집단 간 차이를 정량적으로 파악할 수 있으며, 오류 경험이나 피드백 반영 여부와 같은 중재 요인을 통합적으로 고려해야 한다.

셋째, 실제 현장 데이터 기반 검증이 필요하다. 본 연구는 사례 중심 분석에 의존했으므로, 의료·금융·교육·법률 등 실제 현장에서 발생하는 로그 데이터를 수집·분석하여 신뢰 요인과 실제 의사결정 간의 상관관계를 검증해야 한다. 예를 들어, 의료 영역에서는 환자–의사–AI 간 삼자 관계에서 신뢰가 치료 순응도에 미치는 영향을 분석할 수 있고, 금융 영역에서는 AI 대출 심사 추천과 최종 승인 간 불일치 사례를 통해 신뢰 작동 조건을 탐색할 수 있다. 교육 영역에서는 AI 튜터의 장기적 사용이 학습 성취도와 학생·교사 간 신뢰 관계에 미치는 영향을 추적할 수 있으며, 법률 영역에서는 판례 추천 AI가 실제 변호사와 판사의 결정 흐름에 어떻게 반영되는지 네트워크 분석을 통해 규명할 수 있다. 이러한 데이터 분석은 대규모 로그 데이터 기반 통계 기법과 현장 인터뷰를 병행하는 혼합 방법 연구(Mixed Methods)를 통해 타당성을 강화할 수 있다. 이와 같은 후속 연구는 신뢰의 심리적 요인을 정량적으로 검증할 뿐 아니라, 문화적·시간적·맥락적 요인을 고려한 인간-AI 신뢰 모델을 구축하는 데 기여할 것이다. 이는 AI 기술이 사회 전반에서 보다 책임감 있게 활용되기 위한 필수적인 토대가 될 것이다.

참고문헌

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. Human Factors, 57(3), 407–434.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. Academy of Management Review, 20(3), 709–734.

Shin, D. (2020). User perceptions of algorithmic decisions in the personalized AI system: Perceptual evaluation of fairness, accountability, transparency, and explainability. Journal of Broadcasting & Electronic Media, 64(4), 541–565.

IBM Watson Health. (2019). Watson for Oncology: Lessons learned. IBM White Paper.

중환자실 환자의 약물 처방 패턴을 반영한 장기 인공호흡 예측 모델 개발

배지원, 이승희*
*건양대학교병원
602648@kyuh.ac.kr, *shleedynamics@kyuh.ac.kr

Development of a Machine Learning Model for Long-Term Ventilation Prediction Based on Medication Prescription Patterns in ICU Patients

Bae Ji Won, Lee Seung Hee*
*Konyang University Hospital

요 약

본 논문은 중환자실(ICU) 환자의 진정제·진통제 및 섬망 치료제의 처방 패턴을 분석하고, 이를 기반으로 장기 인공호흡 여부를 예측하는 인공지능 모델을 개발하는 것을 목표로 한다. 최근 dexmedetomidine 사용의 증가와 함께 기존 약물(midozalam, remifentanil 등)의 사용 감소가 관찰되고 있으며, 이러한 변화가 환자의 장기 인공호흡 유지 여부와 밀접한 관련성이 있다는 임상적 가설을 설정하였다.
이에 본 연구는 이러한 약물 처방 시퀀스를 정량화하여 머신러닝 기반 예측 모델을 학습시킴으로써, 임상적 개입이 필요한 환자를 조기에 선별할 수 있는 가능성을 탐색하였다.

실험 결과, LogisticRegression, XGBoost, Random Foreaest, LightGBM, CatBoost 모델 중 RandomForest 모델에서 AUROC 0.839으로 양성 클래스(장기 인공호흡 환자)에서 우수한 성능을 보였으며, 이는 약물 패턴 정보가 장기 인공호흡 유지 여부를 예측하는 데 있어 독립적이고 유의미한 인자임을 실증하였으며 ICU 환경에서 조기 위험군 선별을 위한 인공지능 기반 임상의사결정지원(CDSS) 가능성을 제시한다.

Ⅰ. 서 론

중환자실(ICU) 환자는 다중 질환, 약물 간 상호작용, 급격한 생리적 변화 등으로 인해 치료 전략 수립이 복잡하고 예측이 어렵다. 특히 진정제, 진통제, 섬망 치료제 등 중추신경계 작용 약물은 환자의 인공호흡 유지 기간이나 회복 경과와 밀접한 연관이 있어[1], 약물 처방 패턴의 정량적 분석을 통한 환자 예후 예측의 필요성이 크다.

최근 임상에서는 dexmedetomidine의 사용이 증가하며, 기존 약물(midozalam, remifentanil 등)의 사용량은 감소하는 추세를 보이고 있다. 이러한 변화는 단순한 약물 선택을 넘어 환자의 기계환기 유지 및 조기 이탈 가능성에 영향을 줄 수 있는 중요한 요인으로 작용한다[2].

본 연구는 ICU 환자의 실제 처방 데이터를 기반으로 약물 처방 시퀀스와 인공호흡 유지 기간 간의 연관성을 분석하고, 이를 통해 장기 인공호흡 여부를 예측하는 머신러닝 기반 모델을 개발하는 것을 목적으로 한다. 궁극적으로는 약물 사용에 따른 예후 패턴을 정량화하고, 이를 기반으로 임상적 의사결정에 기여할 수 있는 데이터 기반 예측 프레임워크를 제시하고자 한다.

Ⅱ. 본론

1. 데이터셋

본 연구에서는 건양대학교병원 ICU 데이터로, 2012.01.01.~2021.12.31.까지 총 10년간 수집한 환자 데이터를 활용하였다. 분석은 48시간 이상 특정 약물을 사용한 548명의 환자를 대상으로 수행하였다. 주요 약물군은 진정제·진통제(dexmedetomidine, remifentanil, midazolam)와 섬망 치료제(haloperidol, quetiapine)로 구성되었으며, 연도별 사용 분포와 처방 경향 변화를 평가하였다.

2. 방법론

본 연구에서는 약물 처방 패턴과 환자 예후 간의 관계를 모델링하기 위해 데이터 전처리(KNN-imputation, 범주형 변수 처리, 연속형 변수 표준화)를 수행하였다. 분류 모델로는 LogisticRegression, XGBoost, Random Foreaest, LightGBM, CatBoost를 선정하였다. 약물군별 클러스터링 진행을 위하여 latent representation 학습을 위해 Variational Autoencoder(VAE)를 활용하였다. 성능평가지표는 AUROC, AUPRC, F1-score로 선정하여 약물 패턴별 예측 인자 해석을 수행하였다.

3. 실험 및 주요 결과

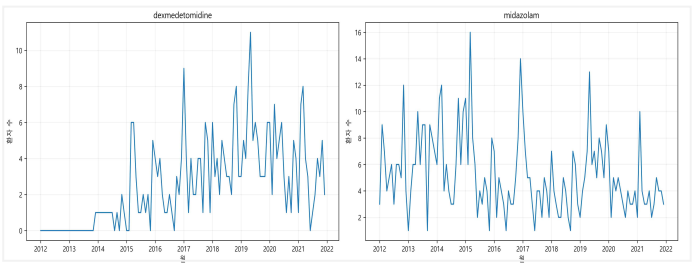


Fig. 1. Graph of the number of patient prescriptions for dexmedetomidine and midazolam.

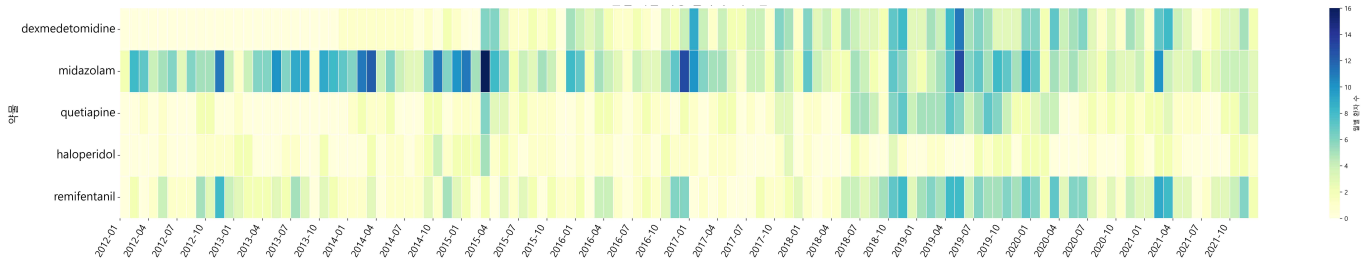


Fig. 2. Heatmap of monthly number of patients prescribed medication

10년 간 진정제·진통제(dexmedetomidine, remifentanil, midazolam)와 섬망 치료제(haloperidol, quetiapine)의 환자 처방 수 분석 결과(Fig. 1, Fig. 2), 장기간 사용 시 체내 축적 위험이 있는 Midazolam은 2012~2015년까지 가장 많이 사용된 진정제로 확인되었으나, 2018년부터 Dexmedetomidine의 사용이 증가함에 따라 사용이 이전보다 감소하는 경향을 보였다. Remifentanil은 2018년 이후 꾸준히 증가하여 진통제 중 가장 안정적인 사용 패턴을 보였다. 한편, 섬망 치료제 haloperidol은 전반적으로 사용 빈도가 낮고 점차 감소하는 추세였으며, quetiapine은 2018~~2020년에는 비교적 지속적으로 사용되었다.

이러한 결과는 지난 10년간 진정·진통제 사용 패턴이 midazolam 중심에서 dexmedetomidine 및 remifentanil 중심으로 전환되었음을 시사하며, 이는 장기 인공호흡 환자 관리에서 약물 선택의 변화를 반영한다.

환자별 약물 처방 정보와 인공호흡기 사용 기간 데이터를 활용하여 총 5개의 머신러닝 모델을 학습하였다. 본 연구에서는 인공호흡기 적용 기간 21일 이상을 장기 인공호흡(long-term ventilation)으로 정의하였다. 데이터 불균형을 보정하기 위해 SMOTE, ADASYN 샘플링 기법을 적용하여 성능을 비교하였으며, 각 모델에 대해 교차검증을 실시하였다.

분석 결과, Random Forest 모델이 평균 AUROC 0.839 (95% CI 포함)로 최고 성능을 기록하였다 (Table 1).

Table. 1 Model-specific sampling and cross-validation results

model	AUROC	AUPRC	F1-score	ci_low	ci_high
RandomForest	0.839	0.652	0.543	0.812	0.866
XGBoost	0.834	0.620	0.584	0.803	0.865
LightGBM	0.833	0.613	0.585	0.811	0.855
CatBoost	0.832	0.636	0.544	0.797	0.867
Logistic Regression	0.825	0.645	0.594	0.793	0.857

Random Forest 모델의 성능지표 결과, AUROC 0.839로 가장 좋은 결과를 보였다. 다만, Precision은 상대적으로 낮아 AUROC의 절반 수준에 그쳤으며, 이는 임상적 적용을 위해 실제 장기 인공호흡 환자에 대한 추가 검증과 예측 임계값(threshold)의 조정이 필요함을 시사한다.

추가적으로, 약물군 간의 latent feature를 학습하기 위해 Variational Autoencoder(VAE)를 활용한 클러스터링을 수행하였으나, 해당 feature의 도입은 예측 성능 개선에 유의미한 영향을 미치지 않는 것으로 나타났다. 이는 향후 모델 구조 개선 및 약물 투여의 시계열 정보(temporal dynamics)를 반영한 feature engineering이 필요함을 시사한다.

Ⅲ. 결론

본 연구에서는 중환자실(ICU) 환자의 약물 처방 정보를 반영한 예측 모델을 개발하여, 약물 반응 및 환자 예후를 효과적으로 예측할 수 있음을 확인하였다. 특히, 진정제 및 섬망 치료제의 처방 패턴이 기계환기 유지 기간에 미치는 영향을 통해 이러한 약물군 간의 처방 변화는 인공호흡기 조기 이탈 가능성과 유의한 연관성을 나타냈으며, 이는 약물 선택이 환자의 인공호흡 예후에 있어 핵심적인 예측 인자로 작용함을 시사한다.

향후 연구에서는 다기관 기반의 외부 검증 및 확장된 시퀀스 모델을 통해 일반화 성능을 강화하고, 정책적 약물 선택 가이드라인 수립에 기여할 수 있는 정밀 의료 기반의 분석 프레임워크를 구축할 예정이다.

ACKNOWLEDGMENT

This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number : RS-2024-00439193)

참 고 문 헌

[1] 이민정, 송예슬 and Kim, Hyunah. (2016). Pharmacotherapy for Delirium in the Intensive Care Units. Journal of Korean Society of Health-System Pharmacists, 33(1), 8-19.

[2] Kim, Y.-S., Lee, H.-J., & Jeon, S.-B. (2015). Management of pain and agitation for patients in the intensive care unit. *Journal of Neurocritical Care, 8*(2), 53 - 65. https://doi.org/10.18700/jnc.2015.8.2.53

심탄도 신호를 이용한 심전도/광용적맥파 복원: 손실 함수 설계와 성능 분석

박웅 [1,2], 김진원 [1,2], 이성한 [2], 정인철 [1,2*]

[1] 한림대학교 인공지능융합학과
[2] 한림대학교 뇌혈관질환 선도연구센터
*incheol.jeong@hallym.ac.kr

Reconstructing ECG and PPG from Single-Channel BCG: Loss Design and Performance Analysis

Ung Park[1,2], Jinwon Kim[1,2], Sunghan Lee[2], In cheol Jeong[1,2*]
[1]Department of Artificial Intelligence Convergence, Hallym University
[2]Cardiovascular Disease Research Center, Hallym University

요 약

본 연구는 무구속 심탄도 (ballistocardiography, BCG)로부터 심전도 (electrocardiography, ECG)와 광용적맥파 (photoplethysmography, PPG)를 동시에 복원하는 공유 인코더-이원 디코더 U-Net 기반 멀티태스크 모델을 제안한다. 세 가지 손실 설계(파형 중심, 정점 보존, 가중 결합)를 비교한 결과, 파형 중심 손실은 파형 오차를 최소화했고, 정점 보존 손실은 박동 간격 재현성을 개선했다. 가중 결합 손실은 두 목표 간 균형을 달성하여 전반적 성능이 가장 우수했다. 본 결과는 BCG 기반 ECG/PPG 동시 복원의 타당성을 입증하며, 손실 결합이 파형 충실도와 타이밍 정확도 간 트레이드오프를 효과적으로 완화함을 보여준다.

Ⅰ. 서 론

심혈관 질환은 전 세계 주요 사망, 이환 원인으로, 위험인자의 조기 발견과 연속 모니터링이 예후 개선에 핵심적이다 [1]. 심전도 (electrocardiography, ECG)는 심장의 전기 활동을, 광용적맥파 (photoplethysmography, PPG)는 말초 혈류 변화를 측정한다. 그러나 두 신호는 피부 접촉형 센서를 필요로 하며, 장기간 사용 시 불편감, 접촉 압 변화에 따른 왜곡, 움직임·수면 중 이탈로 인한 신뢰도 저하가 발생한다 [2]. 반면 심탄도 (ballistocardiography, BCG)는 침대·의자 등에 부착된 무구속 센서로 심장의 기계적 활동을 압력 변화로 간접 측정하여 연속 관측에 유리하지만, 움직임·외란·설치 조건에 민감해 신호 품질과 특징 추출의 신뢰성 확보가 과제로 남아 있다 [3]. 이를 보완하기 위해 전처리 및 딥러닝 기반 복원·잡음 억제 기법이 제안되어 왔으며 [4, 5], BCG 로부터 ECG/PPG 를 복원하면 접촉 센서 없이도 맥파전달시간 (pulse transit time, PTT) 기반 혈압 추정이 가능하다 [2]. 다만 단일 BCG 로 서로 다른 부위에서 획득되는 ECG·PPG 를 동시 복원할 때 위상 차와 정점(peak)을 정밀하게 재현하는 일은 여전히 도전적이다 [4, 6].

본 연구는 이러한 한계를 해결하고 임상적 활용성을 검증하기 위해, 단일 BCG 입력으로 ECG 와 PPG 를 동시 복원하는 U-Net 기반 멀티 태스크 모델을 제안하고, 위상·정점 보전을 반영한 다양한 손실 설계를 비교·평가하여 복원 품질을 체계적으로 분석한다.

Ⅱ. 본 론

가. 연구 방법론

본 연구에는 심폐질환이나 운동 제한이 없는 만 19 세 이상 성인 남녀 각 20 명 (총 40 명)을 모집하였다. 연구는 한림대학교 생명윤리위원회 승인 (HIRB-2024-012)을 받았으며, 모든 참가자에게 서면 동의를 받았다.

BCG 는 압전 센서 기반 SSS-820 (Shenzhen Tianzhi Technology Co., Ltd., 100 Hz)로, ECG/PPG 는 MP160 (BIOPAC Systems, Inc.)과 BN-RSPEC-T, PPGED 모듈 (1,000 Hz)로 동시 기록하였다. 수집 절차는 다음과 같다: (i) 참가자에게 ECG (흉부)와 PPG (오른손 검지) 센서를 부착한 뒤 의자에 착석, (ii) 침대로 이동하여 60 초 안정 상태 유지, (iii) 의자로 복귀. 위 절차를 3 회 반복하였다.

ECG 와 PPG 는 100 Hz 로 다운샘플링하여 BCG 와 샘플링 주파수를 일치시켰다. 이어 0.5–60 Hz 5 차 Butterworth 대역통과 필터를 적용하여 저주파 기저선과 고주파 잡음을 제거하였다. 모든 신호를 5 초 윈도우로 분할하였고, BCG 에는 Z-score 정규화를 적용하였다.

모델은 공유 인코더와 신호별 (ECG/PPG) 디코더로 구성된 1-D U-Net 아키텍처이다. 인코더는 일차원 합성곱으로 BCG 특성을 추출하고, 각 디코더는 일차원 전치합성곱으로 해상도를 복원한다. 인코더-디코더 간 skip-connection 을 사용하여 세부 특성을 보존하였다.

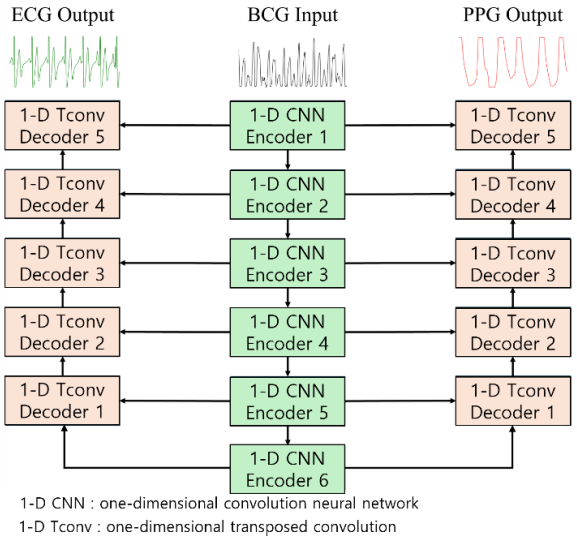


그림 1 공유 인코더와 신호별 디코더로 구성된 멀티 태스크 모델

손실은 세 가지로 정의하였다. 파형중심 손실 L_{wave}, 기울기 (정점/경계) 보존 손실 L_{grad}, 및 두 손실의 가중 결합 손실 $L_{combine}$은 수식 (1-3)과 같다.

$$L_{wave} = \frac{1}{|C|}\sum_{c\in C}\frac{1}{T}\sum_{t=1}^{T}\left(s_{c,t} - \widehat{s_{c,t}}\right)^2 \quad (1)$$

$$L_{grad} = \frac{1}{|C|}\sum_{c\in C}\frac{1}{T-1}\sum_{t=2}^{T}\left(\nabla s_{c,t} - \nabla\widehat{s_{c,t}}\right)^2 \quad (2)$$

$$L_{combine} = \alpha L_{wave} + \beta L_{grad} \quad (\alpha, \beta > 0) \quad (3)$$

여기서 $\nabla s_{c,t} = s_{c,t} - s_{c,t-1}$ $(t = 2 \ldots T), C = \{ECG, PPG\}$ 이다. $s_{c,t}$는 복원해야 될 원본 신호, $\widehat{s_{c,t}}$은 딥러닝을 통해 복원될 출력 신호이며, c 는 복원해야 할 각 채널, t 는 시계열에서 특정 시점을 의미한다. α 및 β는 가중합의 계수로, 본 연구에서는 α=0.5, β=1.0 을 사용하였다.

평가는 (i) 파형 재현성과 (ii) 정점 타이밍 재현성으로 구성하였다. 파형 재현성은 참조 대비 복원 파형의 평균 제곱오차 (mean square error, MSE)와 평균절대오차 (mean average error, MAE)로 산출하였다. 정점 타이밍 재현성은 ECG 의 평균 R-R 간격 (mean R-R interval, mean RRI) 및 PPG 의 평균 맥파 간격 (mean inter-beat interval, mean IBI)에서 참조 대비 간격 시계열 오차의 MSE 로 정량화하였다. 세 손실 함수에 대해 동일 지표 세트를 적용하여 성능을 비교하였다.

나. 연구 결과

세 손실 함수에 대한 파형 오차 (MSE, MAE)와 간격 오차 (RRI·IBI 의 MSE)를 비교하였다 (표 1). L_{wave} 는 ECG/PPG 파형 오차를 최소화하여 절대 파형 재현에 유리하였다. L_{grad}는 정점·경계 보존 특성으로 PPG 의 IBI 오차 (MSE)를 가장 낮췄다. 마지막으로 $L_{combine}$은 ECG 의 RRI 오차 (MSE)를 최소화하여 파형 충실도와 타이밍 정확도 간 균형을 달성하였다.

Metric	L_{wave}	L_{grad}	$L_{combine}$
ECG MSE	**0.43**	1.28	0.80
PPG MSE	**0.15**	0.54	0.27
ECG MAE	**0.22**	0.77	0.41
PPG MAE	**0.23**	0.54	0.38
Mean RRI	2.37	40.24	**0.10**
Mean IBI	2.90	**0.28**	2.92

표 1 손실 함수 구성에 따른 성능 비교

파형 복원 예시에서 L_{wave}는 전반적 파형 충실도가 높으나 정점이 다소 평탄화되는 경향을 보였다. L_{grad}는 국소 경계·정점 보존이 우수하나 저주파·평탄 구간의 진폭 재현이 저하되었다. $L_{combine}$은 두 특성을 절충하여 파형 충실도와 정점 보존 간 균형적 특성을 나타냈다 (그림 2).

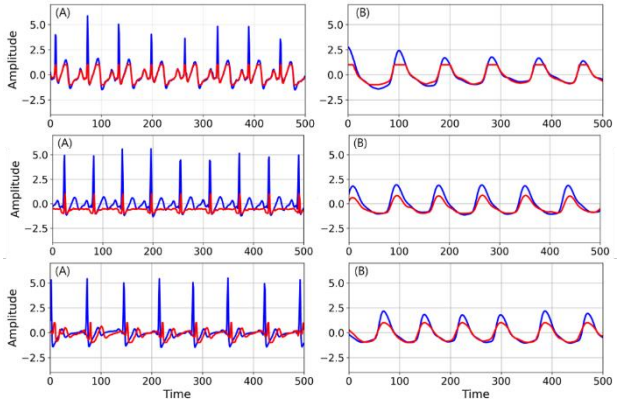


그림 2 복원된 ECG/PPG 복원 예시. 위에서부터 순서대로 L_{wave}, L_{grad}, $L_{combine}$ (A: ECG, B: PPG, 파란색: 원본신호, 붉은색: 복원신호)

모델의 일반화 성능을 검증하기 위해 훈련 집단 규모를 변화시키는 실험을 수행하였다. 전체 40 명 중 5 명을 고정된 검증 집단으로 두고, 나머지 인원을 10 명에서 35 명까지 5 명 단위로 확장하여 $L_{combine}$기반 학습을 진행하였다. 그림 3 은 훈련 집단 규모에 따른 ECG RRI 와 PPG IBI 의 MSE 변화를 보여준다. 훈련 인원이 증가할수록 두 지표의 오차가 전반적으로 감소하여, 충분한 표본 확보가 정점 간격 복원 정확도 향상에 기여함을 확인하였다.

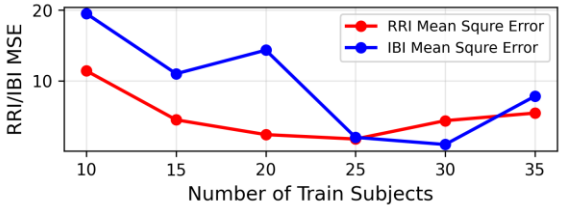


그림 3 훈련 집단 규모에 따른 RRI 및 IBI MSE 비교

Ⅲ. 결 론

본 연구에서는 단일 BCG 로부터 ECG 와 PPG 를 동시 복원하는 U-Net 기반 멀티태스크 모델을 제안하고, 손실 설계가 성능에 미치는 영향을 체계적으로 비교하였다. 파형 중심 손실 (L_{wave})은 파형 오차를 최소화했고, 기울기 보존 손실 (L_{grad})은 간격 지표 (특히 IBI)를 개선하였다. 가중 결합 손실 ($L_{combine}$)은 두 목표 간 균형을 제공하여 ECG 의 RRI 오차를 최소화하였다. 결과적으로, BCG 기반 ECG/ PPG 동시 복원에서는 파형 충실도와 타이밍 정확도 사이의 트레이드오프가 존재하며, 손실 결합이 유효한 절충점을 제공함을 확인하였다. 특히 제안 모델은 장기 입원 환자의 침상 모니터링, 가정 기반 원격 환자 감시, 수면 무호흡 및 부정맥 조기 탐지 등 임상·생활 환경에서의 연속적 생체신호 모니터링에 적용 가능하다. 또한 복원된 ECG 와 PPG 로부터 심박 변이도, 맥파 전달 시간, 혈압 추정, 호흡 패턴 분석 등 2 차 생체 지표를 도출할 수 있어, 기존 접촉형 센서를 대체하거나 보완하는 무구속 모니터링 도구로의 활용성이 기대된다.

ACKNOWLEDGMENT

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. NR070859)

참 고 문 헌

[1] Kim, C.-S., Ober, S.L. et al. “Ballistocardiogram: Mechanism and Potential for Unobtrusive Cardiovascular Health Monitoring.” *Scientific Reports* 2016. DOI: 10.1038/srep31297

[2] Alizadeh-Meghrazi, M., Ying et al. “Evaluation of dry textile electrodes for long-term electrocardiographic monitoring.” *Bio-Medical Engineering OnLine* 2021. DOI: 10.1186/s12938-021-00905-4

[3] Yang, T., Yuan, H. “Solving variability: Accurately extracting feature components from ballistocardiograms.” *DIGITAL HEALTH* 2024. DOI: 10.1177/20552076241277746

[4] Morokuma, S., Saitoh, T. et al. “Prediction of ECG signals from ballistocardiography using deep learning for the unconstrained measurement of heartbeat intervals.” *Scientific Reports* 2025. DOI: 10.1038/s41598-024-84049-0

[5] Liu, R., Hu, H. et al. “An ECG denoising technique based on AHIN block and gradient difference max loss.” *Journal of Electrocardiology* 2024. DOI: 10.1016/j.jelectrocard.2024.153761

[6] Gomez-Clapers, J., Serra-Rocamora “Towards the standardization of ballistocardiography systems for J-peak timing measurement.” *Measurement* 2014. DOI: 10.1016/j.measurement.2014.09.003

A study on cyber attack targeted for semantic communication systems

Mehwish Ali Naqvi, Insoo Sohn*

Division of Electronics & Electrical Engineering, Dongguk University, Seoul, Republic of Korea

Abstract

Semantic communication (SeCom) systems aim to transmit task-relevant meaning rather than raw data, offering enhanced efficiency for next-generation networks. However, their reliance on deep learning and generative models introduces new and largely unexplored security vulnerabilities. This paper examines four recent and representative cyber attacks that compromise the semantic layer of communication systems. These include embedding-level adversarial perturbations, GAN-based semantic jamming, latent backdoor injection, and covert prompt interception attacks. Together, they expose novel vulnerabilities inherent to generative AI-driven semantic communication. By analyzing the attack goals, mechanisms, and impact across different modalities, this study highlights the urgent need to reassess security paradigms in AI-native communication environments.

Keywords: Semantic Communication, Cybersecurity, Generative AI

1. Introduction

Semantic communication (SeCom) represents a fundamental shift in wireless communication, aiming to transmit meaning rather than raw symbols. By leveraging deep learning models and shared knowledge bases, SeCom improves efficiency in bandwidth-constrained and task-driven environments such as autonomous driving, edge computing, and the Metaverse [2, 1]. The integration of generative artificial intelligence (GenAI) models—including LLMs, diffusion networks, and VAEs—further enhances semantic fidelity, enabling multi-modal compression and prompt-based content reconstruction [3]. However, this reliance on AI-driven pipelines introduces new attack surfaces at the semantic layer, which remain insufficiently protected by traditional bit-level security models [6, 4, 8]. Similar concerns arise in other AI-enabled domains, such as intelligent transportation systems, where our group demonstrated a Time Tampering Black-Box Genetic Algorithm (TTB-GA) attack on digital twin models [7], and in energy harvesting (EH) networks, where AI-based security frameworks address threats like eavesdropping, data manipulation, and denial of service [5]. In this study, we focus exclusively on cyber attacks targeting the semantic layer, reviewing four recent and representative works on adversarial perturbations, GAN-based jamming, semantic backdoors, and covert prompt detection to highlight the evolving threat landscape of SeCom systems.

2. Attack Models in Semantic Communication

A white-box attack approach known as BERT manipulation of the embedding space is proposed by Hoang et al. [9]. Gradient-based perturbations are applied to the input text to generate adversarial examples that distort semantic interpretation for the receiver while preserving syntactic structure. These assaults specifically aim at communal knowledge-based semantic encoders, leading to considerable deterioration of semantic integrity. Experiments utilising the SNLI dataset demonstrated a decline in BLEU score from 0.72 to 0.46 under significant perturbation, but adversarial training only partially restored accuracy. The study highlights the vulnerability of semantic representation spaces and the simplicity with which they can be undermined.

Tang et al. [10] introduce a generative adversarial network (GAN)-based semantic jamming framework, where the attacker plays the role of a generator that learns to craft meaningful perturbations, while the

* Corresponding author:
Email addresses: MANaqvi9@dgu.ac.kr (Mehwish Ali Naqvi), isohn@dongguk.edu (Insoo Sohn)

SeCom receiver acts as a discriminator. This setup forms an adversarial game where the attacker's goal is to reduce semantic similarity without affecting signal detectability. The proposed attack strategy effectively reduces BLEU scores by up to 40% under varying interference levels and assumes access to model gradients. It demonstrates that intelligent jammers trained via adversarial learning can significantly impact semantic integrity in AI-native networks.

A hidden semantic backdoor attack (CSBA) developed for semantic communication in ICVs is introduced by Xu et al. [11]. In contrast to conventional input-level poisoning, CSBA embeds triggers into the latent semantic space, deliberately eliminating semantic attributes like road signs or people while maintaining visual fidelity. The assault employs Patchwise Latent Masking (PLaMa) and adversarial fine-tuning to guarantee that semantic tampering is undetectable. Assessment of the Cityscapes dataset indicates an attack success rate above 80% with no PSNR degradation, underscoring the stealth and accuracy of backdoors in the semantic feature space.

Du et al. [12] explore covert communication attacks in generative AI-aided semantic communication using multi-modal prompts. The attacker, modeled as a passive "warden," attempts to detect hidden semantic transmissions by observing structured prompt delivery. The system employs a benign jammer to mask transmission and uses a Generative Diffusion Model (GDM) to jointly optimize transmission power, jamming power, and diffusion steps for semantic reconstruction. The attack exploits binary hypothesis testing for detection, and experimental results show that without careful optimization, the warden can successfully identify prompt transmissions. The study highlights a new attack vector specific to prompt-based generative SeCom.

3. Conclusion

This research studied four recent cyberattacks that focus on the semantic layer of AI-native communication systems. In contrast to conventional threats that compromise signal quality or bit-level precision, these assaults directly alter meaning by embedding perturbations, generating adversarial noise, utilising latent backdoors, and employing covert prompt detection. Collectively, they unveil a varied and dynamic danger landscape in semantic communication, wherein adversaries might attain significant success rates without activating traditional alerts. With the growing implementation of SeCom systems in safety-critical and multi-modal applications, there is an imperative necessity to establish security frameworks that are intrinsically cognisant of semantic vulnerabilities. This work seeks to enhance the foundation by elucidating the assault surface and encouraging subsequent defence research.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2023-00252328).

References

[1] H. Xie, Z. Qin, G. Y. Li, and B.-H. Juang, "Deep learning enabled semantic communication systems," *IEEE Transactions on Signal Processing*, vol. 69, pp. 2663–2678, 2021.

[2] X. Luo, H.-H. Chen, and Q. Guo, "Semantic communications: Overview, open issues, and future research directions," *IEEE Wireless Communications*, vol. 29, no. 1, pp. 210–217, 2022.

[3] L. Xia, Y. Zhang, Y. Du, Y. Shi, Y.-C. Liang, and Z. Ding, "Generative AI for semantic communication: Architecture, challenges, and outlook," *IEEE Wireless Communications*, vol. 32, no. 1, pp. 60–67, 2025.

[4] Y. E. Sagduyu, T. Erpek, S. Ulukus, and A. Yener, "Is semantic communication secure? A tale of multi-domain adversarial attacks," *IEEE Communications Magazine*, vol. 61, no. 11, pp. 50–55, 2023.

[5] M. Mohammadi and I. Sohn, "Security in energy harvesting systems using artificial intelligence: A survey," *ICT Express*, vol. 9, no. 4, pp. 527–540, 2023.

[6] Z. Yang, D. Niyato, X. Xie, Y. Lyu, X. Cao, and L. Liang, "Secure semantic communications: Fundamentals and challenges," *IEEE Network*, vol. 38, no. 6, pp. 104–110, 2024.

[7] M. Pooyandeh, H. Liu, and I. Sohn, "Cybersecurity in digital twins of electric vehicle's LIBs: Unveiling a robust TTB-GA attack," *IEEE Transactions on Intelligent Transportation Systems*, vol. 26, no. 4, pp. 5360–5381, 2025.

[8] Q. T. Do, D. Won, T. S. Do, T. P. Truong, and S. Cho, "Security and privacy challenges in semantic communication networks," in *Proc. Int. Conf. on Artificial Intelligence in Information and Communication (ICAIIC)*, pp. 1–6, 2025.

[9] V.-T. Hoang, V.-L. Nguyen, R.-G. Chang, P.-C. Lin, R.-H. Hwang, T. Q. Duong, Adversarial Attacks Against Shared Knowledge Interpretation in Semantic Communications, IEEE Transactions on Cognitive Communications and Networking, vol. 11, no. 2, pp. 1024–1040 (2025).

[10] R. Tang, D. Gao, M. Yang, T. Guo, H. Wu, G. Shi, GAN-inspired Intelligent Jamming and Anti-jamming Strategy for Semantic Communication Systems, 2023 IEEE International Conference on Communications Workshops (ICC Workshops), pp. 1623–1628 (2023).

[11] X. Xu, Y. Chen, B. Wang, Z. Bian, S. Han, C. Dong, C. Sun, W. Zhang, L. Xu, P. Zhang, CSBA: Covert Semantic Backdoor Attack Against Intelligent Connected Vehicles, IEEE Transactions on Vehicular Technology, vol. 73, no. 11, pp. 17923–17928 (2024).

[12] H. Du, G. Liu, D. Niyato, J. Zhang, J. Kang, Z. Xiong, B. Ai, D. I. Kim, Generative AI-aided Joint Training-free Secure Semantic Communications via Multi-modal Prompts, ICASSP 2024 - IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 12896–12900 (2024).

FMCW 레이더를 이용한 걸음걸이 기반 신원 식별 및 침입자 탐지 시스템

조희재, 박상윤, 이정훈, 전우진, 전민욱, 김형남*

부산대학교

*hnkim@pusan.ac.kr

Gait-Based Person Identification and Intruder Detection System Using FMCW Radar

Hui Jae Jo, Sang Yun Park, Jeong Hun Lee, Woo Jin Jeon, Min-Wook Jeon,

Hyoung-Nam Kim*

Pusan National Univ.

요 약

최근 주거 침입 범죄의 증가로 인해 많은 가정에서 CCTV (closed circuit television)를 설치하고 있다. 하지만 카메라 기반 시스템은 해킹에 따른 영상 유출과 사생활 침해의 우려가 크다는 한계를 안고 있다. 이에 반해 레이더 센서는 조명 변화와 같은 주변 환경에 영향을 받지 않으며 사생활 침해 측면에서도 제약이 없다. 이러한 장점을 바탕으로 FMCW (frequency-modulated continuous wave) 레이더로 획득한 미세-도플러 스펙트로그램을 이용해 신원 식별을 수행한다. 기존 식별에서는 SVM(Support Vector Machine)과 딥러닝이 사용되어 왔으며, 최근에는 딥러닝이 여러 측면에서 널리 채택되는 추세이다. 그러나 딥러닝 기반 식별 방식은 단일 입력의 확률 최대값으로 클래스를 즉시 결정하는 구조 탓에 짧은 관측 창이나 잡음에 취약하며 침입자를 정상 거주자로 오분류할 위험이 존재한다. 이를 보완하기 위해 본 논문에서는 딥러닝 분류기에 더해 하드보팅 기반 추론 알고리즘을 결합한다. 이는 5개의 입력에 대한 예측을 모아 다수결로 최종 클래스를 결정하고 예측에 대한 의견이 갈리는 경우 침입자로 판별한다. 실험결과, 신원식별에 대해서 93.94 %, 침입자 탐지에 대해 92.86 %의 정확도를 달성하였다. 이러한 결과를 통해 제안하는 시스템은 스마트 홈은 물론 데이터 센터, 연구 시설 등 높은 수준의 실내 보안이 요구되는 환경에서 비접촉, 사생활 보호 침입자 경보 시스템으로 활용될 수 있음을 보여준다.

Ⅰ. 서 론

현대 감시 시스템은 주로 비디오 카메라를 사용하여 정보를 수집하지만, 이러한 시스템은 급격한 밝기 변화로 인한 화질 저하 및 해킹에 따른 개인 정보 유출 같은 근본적 한계를 지닌다. 특히 프라이버시 보호가 요구되는 실내 공간에서는 사용이 크게 제한된다. 반면에 레이더를 이용하는 시스템은 시각 정보를 수집하지 않아 프라이버시를 보호하면서 조명 환경이나 시야 가림에 영향을 받지 않는다. 이러한 특성에 기반해 FMCW(frequency - modulated continuous wave) 레이더로 획득한 미세-도플러 스펙트로그램을 활용해 신원 식별을 수행한다. 신원 식별에는 다양한 접근이 존재하며, FMCW 기반 분류 연구에서는 SVM(Support Vector Machine) 기반 방법도 사용되어 왔지만 최근에는 딥러닝이 성능과 실용성 측면에서 널리 채택되는 추세다. 하지만 딥러닝 기반 분류는 단일 입력의 소프트맥스 출력의 최대값으로 클래스를 즉시 결정하는 구조 탓에 짧은 관측 창이나 잡음에 취약하며 침입자를 정상 거주자로 오분류할 위험이 있다 [1]. 본 논문에서는 이를 보완하기 위해 하드 보팅 기반 추론 알고리즘을 제안한다. 5개의 입력에 대해 분류기의 예측을 모아 다수결로 최종 클래스를 결정함으로써 순간 잡음에 의한 오분류를 줄이고, 예측 의견이 갈리는 경우에는 침입자로 판단하여 탐지까지 가능하게 한다. 또한, 직접 실측한 데이터에 대한 모의실험을 통해 제안한 알고리즘의 신원 식별 및 침입자 탐지 정확도를 확인하였다.

Ⅱ. 본 론

2.1 신경망

합성곱 신경망은 입력 이미지에서 다중 계층의 필터를 통해 특징을 추출하고, 추출된 특징을 기반으로 분류를 수행하는 딥러닝 모델이다. ResNet은 잔차 연결(residual connection)을 도입한 합성곱 신경망 계열 아키텍

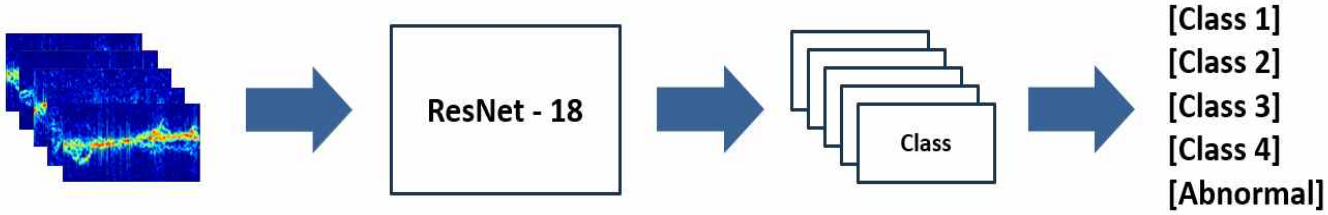


그림 1. 제안하는 신원 식별 및 침입자 탐지 알고리즘

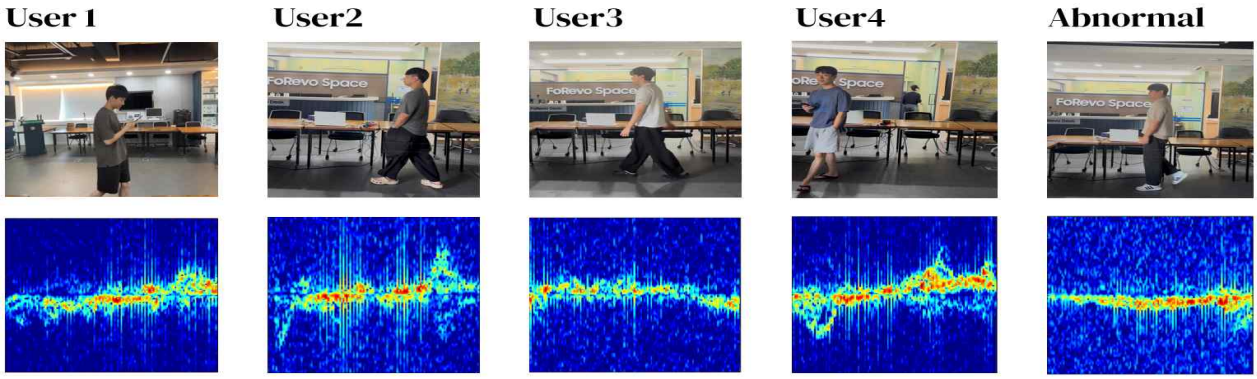


그림 2. 데이터 예시

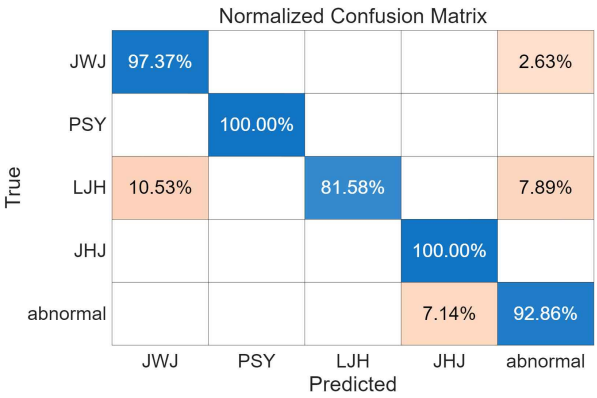


그림 3. 혼동 행렬

처로, 깊은 네트워크에서의 기울기 소실을 완화한다. 그중 ResNet-18은 18개의 가중치 계층으로 구성되어 학습 파라미터 수가 비교적 적으면서도 복잡한 미세-도플러 패턴을 안정적으로 학습할 수 있다. 이에 본 논문에서는 ResNet-18을 기본 분류기로 적용하여 정상 거주자의 신원을 식별하였다.

2.2 제안하는 알고리즘

본 논문에서는 그림 1과 같이 하드 보팅 기반 알고리즘을 도입해 최종 클래스를 결정하고 침입자를 판별한다. 하드 보팅이란 분류기의 예측값들을 다수결 투표해 가장 많은 표를 얻은 예측값을 최종 예측값으로 결정하는 방식이다 [3]. 구체적으로 7.5 초 길이의 레이더 신호에서 윈도우를 2.5 초로 1.25초 간격으로 슬라이딩하여 총 5장의 데이터로 분할한 뒤, 각 윈도우의 STFT 스펙트로그램을 사람별 걸음걸이 특징을 학습한 ResNet에 입력한다. 다섯 번의 예측 결과 가운데 동일 클래스가 3회 이상 반복되면 해당 클래스를 최종 식별 결과로 확정하고, 다수가 클래스 5로 분류되거나 동일 클래스가 3회 미만으로만 나타나는 경우에는 침입자로 판정하도록 알고리즘을 설계하였다.

Ⅲ. 모의실험

데이터는 TI AWR1642 FMCW 레이더 모듈을 이용해 그림 2와 같이 직접 수집하였다. 수집된 원신호는 Matlab에서 스펙트로그램 형태로 변환한 뒤 제로 도플러 부근 성분을 제거하였다. 정상 거주자 4명의 데이터는 총 3360개, 침입자 후보 6명의 데이터는 총 504개의 비정상 샘플을 확보하였다. 학습 단계에서는 비정상 샘플 가운데 2명의 데이터만 사용하고 나머지 4명의 비정상 데이터는 테스트용으로 사용하였다. 최종 레이블 체계는 정상 4 클래스와 침입자 1 클래스로 총 5 클래스로 정의하였다.

선정된 데이터의 80 %를 학습용, 10 %를 검증용, 10 %를 테스트용으로 분할하고, ResNet-18 모델을 사용해 학습을 수행하였다. 하이퍼파라미터는 Adam 옵티마이저(초기 학습률 1×10^{-3}), mini-batch 16, 30 epoch로 설정하였다 [4]. 그림 3은 테스트 평가 결과를 나타낸 혼동 행렬이다. 결과를 보면 정상 클래스는 93.94 %의 정확도를 나타냈으며, 침입자(Abnormal) 탐지 정확도는 92.86 %를 기록하였다.

Ⅳ. 결 론

본 논문에서는 FMCW 레이더 기반 걸음걸이 스펙트로그램과 ResNet-18 기반 신경망을 활용한 구성원 식별 및 침입자 감지 시스템을 제안하였다. 제안된 모델은 4명의 정상 구성원과 이상 클래스 데이터를 학습하여, 하드 보팅 기반 알고리즘을 통해 정상 93.94%, 이상 92.86%의 정확도를 달성하였다. 향후에는 데이터 증강과 위치 시각화를 통해 실시간 경보 시스템으로의 확장을 연구할 계획이다.

ACKNOWLEDGMENT

Put sponsor acknowledgments.

참 고 문 헌

[1] Baptist Vandersmissen et al., “Indoor Person Identification Using a Low-Power FMCW Radar.”, IEEE Transactions on Geoscience and Remote Sensing (TGRS), 2018.

[2] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition, ” In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[3] Hanan Ghali Jabbar et al., “Advanced Threat Detection using Soft and Hard Voting Techniques in Ensemble Learning.”, Journal of Robotics and Control (JRC), 2024.

[4] D. P. Kingma and J. Ba, “ADAM: A Method for stocastic optimization. ”, International Conference on Learning Representations (ICLR), 2015.

Vision-Language Pre-training 모델에서의 전이 가능성 기반 적대적 공격 연구 동향

김용재*, 송민혁**, 이연준***
*한양대학교 ERICA 컴퓨터학부(학부생), **한양대학교 정보보호학과(대학원생),
***한양대학교 컴퓨터공학과(교수)
*likethestars@hanyang.ac.kr, **thd1008@hanyang.ac.kr, ***yeonjoonlee@hanyang.ac.kr

A Survey on Transferability-based Adversarial Attack in Vision-Language Pre-training Models

Yongjae Kim*, Minhyeok Song**, Yeonjoon Lee***
*Hanyang University ERICA, Major in Computer Science (Undergraduate Student)
**Hanyang University, Department of Computer and Information Security (Master Student)
***Hanyang University, Computer Science and Engineering (Associate Professor)

요 약

Vision-Language Pre-training(VLP) 모델에 대한 적대적 공격 연구는 증가하는 추세다. 초기에는 구조적 복잡성과 모달리티 간 상호작용으로 인해 효과적인 공격이 어려웠으나, 최근 연구들은 전이 가능성을 활용하여 Black-box 상황에서도 성공적인 공격을 달성하였다. 특히 다양한 구조의 VLP 모델을 넘나들며 전이성을 확보하려는 시도가 이어지고 있으며, 대규모 비전-언어 모델(LVLM)까지 공격 가능성이 확장되고 있다. 이에 본 논문은 VLP 모델의 구조와 전이 가능성 기반 공격 기법들을 정리하고, 향후 연구 방향을 제시한다.

Ⅰ. 서론

최근 CLIP[1], ALBEF[2], BLIP[3] 등 대규모 Vision-Language Pre-training (VLP) 모델의 등장으로, 이미지-텍스트 검색 (image-text retrieval), 시각적 추론 (visual entailment), 비주얼 그라운딩 (visual grounding) 등 다양한 멀티모달 과제에서 눈에 띄는 성능 향상이 이루어졌다. 이러한 성능 향상은 대규모 데이터와 트랜스포머 기반 아키텍처의 결합에 기인하며, VLP 모델은 실제 서비스와 애플리케이션에도 빠르게 확산되고 있다.

그러나 딥러닝 기반의 다른 기술들과 마찬가지로 VLP 모델 역시 적대적 공격에 대한 견고성이 부족하다. 특히 실제 배포 환경에서는 모델 내부 구조나 파라미터에 접근하기 어려워 white-box 가정이 성립하지 않는 경우가 많으므로, black-box 상황에서의 전이 가능성 (transferability)이 훨씬 더 현실적이고 위협적인 공격 방식으로 주목받고 있다.

따라서 본 논문에서는 멀티모달 상호작용을 적극적으로 활용하거나, 모달리티 일관성/불일치 (modality consistency/discrepancy) 특성을 조정하는 등 전이가능성을 높이기 위한 다양한 공격 연구들을 VLP 모델의 구조와 연관 지어 종합적으로 정리하고, 향후 연구 방향을 제시한다.

Ⅱ. 배경

VLP 모델은 두 가지 구조로 구분된다. Aligned 구조는 이미지 인코더와 텍스트 인코더가 독립적으로 작동하며, 동일한 임베딩 공간 (embedding space)에서 각 모달리티의 특징 (feature)을 정렬하도록 학습된다. 대표적 예인 CLIP [1]은 대규모 웹 데이터셋을 활용해 대조학습 (contrastive learning) 기반의 이미지-텍스트 정렬을 달성한다. 이 구조는 대규모 학습에 효율적이고 이미지-텍스트 검색에서 우수한 성능을 보인다. 그러나 모델 내부에서 교차 모달 상호작용 (cross-modal interaction)이 직접적으로 일어나지 않으므로, 정교한 멀티모달 추론이나 모달리티 간 깊은 상호작용이 필요한 과제에는 한계가 있다. 이러한 특성은 적대적 공격에서도 취약 지점으로 작용하며, 임베딩 공간 자체를 교란하는 방식이 주요 공격 전략이 된다.

Fused 구조는 이미지와 텍스트를 각각 인코더로 처리한 뒤 멀티모달 인코더 (multimodal encoder)에서 융합하여 하나의 통합 표현을 학습한다. ALBEF [2], BLIP [3], TCL [4] 등이 여기에 속한다. Fused 구조는 self-attention 과 cross-attention 을 통해 이미지와 텍스트 간 상호작용을 적극 활용하므로, 시각적 추론 (visual entailment)이나 비주얼 그라운딩 (visual grounding) 등 복잡한 과제에서 강점을 지닌다. 다만 연산 비용이 크고, 이미지-텍스트 검색 기반 과제에서는 Aligned 구조보다 효율성이 낮을 수 있다. 공격 관점에서는 교차 모달 상호작용을 교란할 수 있는 지점이 다양하여 보다 정교한 공격 기법 설계가 가능하다.

Ⅲ. 전이 가능성 기반 공격

VLP 모델은 크게 Aligned 구조와 Fused 구조로 구분되지만, 실제로는 모델마다 세부 아키텍처, 학습 방식, 모달 상호작용 메커니즘이 상이하다. 일부는 임베딩 공간 정렬을 중심으로, 다른 일부는 cross-attention 을 통해 모달 상호작용을 활용한다. 이러한 다양성 때문에 특정 모델에만 특화된 white-box 공격은 실제 배포 환경을 충분히 설명하기 어렵다. 현실적

시나리오에서는 서로 다른 구조의 VLP 모델에도 효과적인 전이 가능성이 핵심 요건으로 요구된다.

초기 연구 중 하나인 X. Xu et al.[5]는 이미지와 텍스트 모달리티를 독립적으로 혹은 단순 병합해 공격했을 뿐, 두 모달리티가 결합하여 형성하는 공유 임베딩 공간 자체의 교란·최적화는 고려하지 못했다. 이에 따라 초기 멀티모달 공격인 Co-attack [6]은 이미지와 텍스트를 순차적으로 공격해 멀티모달 상호작용을 부분적으로 고려했다는 의의가 있다. 그러나 단일 쌍 (single-pair)에 국한된 교차 모달 상호작용만을 다루었기에, white-box 에서는 강력하더라도 black-box 에서는 전이 가능성이 크게 제한되었다.

이 한계를 보완한 Set-level Guidance Attack (SGA) [7]은 단일 쌍 의존을 넘어서 세트 수준 (set-level)의 다양한 image- text pairs 를 활용해 교차 모달 상호작용을 확장하였다. 동일 이미지에 대해 여러 캡션을 고려하거나, 이미지의 다양한 스케일 변환을 도입함으로써 정렬을 유지하면서도 더 풍부한 상호작용 정보를 공격에 활용했다. 특히 cross-attention 기반 상호작용이 강한 Fused 구조에서 전이 가능성을 유의미하게 향상시켰으며, Aligned 구조에서도 기존 대비 개선을 보였으나 효과는 상대적으로 제한적이었다.

최근 제안된 Transferable Multi-Modal (TMM) Attack [8]은 set-level 확장을 넘어, VLP 모델의 핵심 특성인 모달리티 일관성 특징과 모달리티 불일치 특징을 동시에 조정하는 전략을 도입하였다. 우선 어텐션 지향 특징 교란 (attention-directed feature perturbation)으로 여러 VLP 모델에 공통적인 일관적 특징을 교란하고, 직교 지향 특징 이질화 (orthogonal-guided feature heterogenization)로 모달 간 차별적 특징을 증폭시킨다. 이 접근은 Fused 구조에서 전이 가능성을 극대화하는 데 효과적이었고, Aligned 구조에서도 전이를 확보했다. 더 나아가 TMM 은 대규모 비전- 언어 모델 (LVLM; BLIP-2, LLaVA, GPT-4V 등)에서도 black-box 공격 성능을 입증하며, 구조 차이를 넘어서는 전이 가능한 공격 가능성을 보여주었다.

IV. 결론

본 논문에서는 VLP 모델의 구조를 정리하고, Co-attack, SGA, TMM 으로 이어지는 일련의 공격 연구를 비교함으로써 각 기법이 전이 가능성에 어떤 향상을 가져왔는지 논의하였다. Co-attack 은 멀티모달 협력적 공격의 가능성을 최초로 제시했으나 전이 가능성 확보에는 한계가 있었다. SGA 는 세트 수준에서 다변화된 가이던스를 활용하여 black-box 환경에서 전이 가능성을 개선하였다. 마지막으로 TMM 은 모달리티 일관성 특징과 모달리티 불일치 특징을 동시에 조정하는 정교한 전략을 통해 다양한 VLP 모델은 물론 대규모 비전-언어 모델 에서도 강력한 전이 가능성을 달성하였다.

향후 연구는 다음과 같이 확장될 수 있다. 첫째, LVLM 을 대상으로 한 체계적 전이 가능성 연구가 필요하다. 실세계 활용이 빠르게 증가하는 모델들에 맞춘 공격 및 방어 기법의 표준화된 벤치마크와 프로토콜이 요구된다. 둘째, Aligned 구조와 Fused 구조의 차이를 넘어서는 범용 전이 가능한 공격 기법 설계가 필요하다. 현재 기법들은 특정 구조에서 효과가 두드러지는 경향이 있으므로, 구조적 차이에 구애받지 않는 보편적 전략이 요구된다. 셋째, 단일 모달 환경에서의 텍스트 의미 보존과 이미지 은밀성 문제는 멀티모달 결합 상황에서 더욱 복잡해진다. 생성된 적대적 텍스트가 원문과 의미적으로 불일치하거나, 이미지의 미세 변조가 쉽게 눈에 띄는 경우 공격의 자연스러움과 설득력이 저하될 수 있다. 따라서 향후에는 이미지-텍스트 양쪽의 교란을 균형 있게 설계하여 의미론적 정합성과 시각적 은밀성을 동시에 보장하는 균형 잡힌 멀티모달 공격 기법이 요구된다.

ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원-대학 ICT 연구센터(ITRC)의 지원을 받아 수행된 연구임(IITP-2025-RS-2024-00438056)

참 고 문 헌

[1] A. Radford et al., "Learning Transferable Visual Models from Natural Language Supervision," Proc. 38th Int. Conf. on Machine Learning (ICML), pp. 8748- 8763, 2021.

[2] J. Li et al., "Align before Fuse: Vision and Language Representation Learning with Momentum Distillation," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 34, 2021.

[3] J. Li et al., "BLIP: Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding and Generation," Proc. 39th Int. Conf. on Machine Learning (ICML), pp. 12888- 12900, 2022.

[4] J. Yang al., "Vision-Language Pre-Training with Triple Contrastive Learning," Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 10647- 10657, 2022.

[5] X. Xu et al., "Fooling Vision and Language Models Despite Localization and Attention Mechanism," Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4957- 4967, 2018.

[6] J. Zhang et al., "Towards Adversarial Attack on Vision-Language Pre-training Models," Proc. 30th ACM Int. Conf. on Multimedia (ACM MM), pp. 5428- 5436, 2022.

[7] D. Lu et al., "Set-level Guidance Attack: Boosting Adversarial Transferability of Vision-Language Pre-training Models," Proc. IEEE/CVF International Conference on Computer Vision (ICCV), pp. 1234- 1244, 2023.

[8] H. Wang et al., "Transferable Multimodal Attack on Vision-Language Pre-training Models," in Proc. IEEE Symposium on Security and Privacy (S&P), pp. 1-17, 2024.

Knowledge Graph 기반 RAG를 활용한 LLM 취약점 탐지 방법론

황시준* 이연준**
*한양대학교 정보보호학과(대학원생), **한양대학교 컴퓨터공학과(교수)
*ghkdtlwns987@hanyang.ac.kr, **yeonjoonlee@hanyang.ac.kr

Knowledge Graph based RAG for Vulnerability Detection with Large Language Models

Sijune Hwang*, Yeonjoon Lee**
**Hanyang University, Department of Computer and Information Security (Master Student)
***Hanyang University, Computer Science and Engineering (Associate Professor)

요 약

대규모 언어 모델(LLM)은 소프트웨어 보안 분야에서 새로운 가능성을 열며, 특히 소스코드 내 취약점을 자동으로 탐지하는 데 활발히 적용되고 있다. 그러나 LLM 기반 탐지 기법은 환각(Hallucination)으로 인한 오탐(False Positive) 문제를 안고 있어 실제 환경에 적용하기 어렵다는 한계가 존재한다. 이러한 문제를 완화하기 위해 Retrieval-Augmented Generation(RAG) 기법이 도입되었지만, Common Weakness Enumeration(CWE) 취약점 간의 관계를 충분히 반영하지 못하는 한계를 가진다. 본 연구는 이러한 한계를 극복하기 위해 CWE 관계를 Knowledge Graph(KG) 형태로 정의하고, 이를 RAG 와 결합한 새로운 탐지 방법론을 제안한다. 이는 CWE 간의 연계 시나리오를 효과적으로 식별하며, 기존 방법론 대비 더 정밀하고 설명 가능한 탐지 결과를 제공하였다. 본 연구는 단일 취약점 탐지를 넘어 복합적 취약점 시나리오 분석을 가능하게 한다는 점에서 차별성을 가진다.

Ⅰ. 서 론

LLM 은 소프트웨어 보안 연구 분야 전반에 채택되고 있으며, 소스 코드 내 취약점 자동 탐지에 유망한 성과를 보이고 있다[5][6][9]. 그러나 실제 적용 단계에서 환각[2] 문제로 인한 오탐을 유발하여 신뢰도가 저하되고, 이는 보안 도메인의 특성상 도입을 어렵게 만드는 핵심 문제이다. 환각은 LLM 이 훈련 데이터나 프롬프트 맥락에 없는 사실을 만들어내는 현상으로, 탐지 근거의 부재와 일관성 없는 설명을 동반하여 운영 상 리스크를 키운다. 이 문제를 완화하기 위한 접근으로 RAG[1]가 주목받고 있다. RAG 는 외부 지식을 참조해 응답을 보강함으로써 환각을 줄인다. 하지만 취약점 탐지라는 특수 과제에서는 텍스트 위주 검색만으로는 한계가 분명하다. 실제 보안 사고는 단일 취약점의 유무보다, 취약점 간의 인과관계로 전개되는 시나리오가 중요하다. 예를 들어 입력 검증 부재(CWE-20)가 선행되면, 이후 신뢰되지 않은 데이터 역직렬화(CWE-502)로 확장/연계될 위험이 커진다. 그러나 일반적인 RAG 파이프라인은 이런 구조적 관계를 명시적으로 반영하지 못해, “무엇이 왜 위험한가”를 연쇄적으로 설명하는 능력이 제한된다. 본 논문은 이러한 간극을 메우기 위해, CWE 취약점과 그 상호 관계를 그래프로 구조화한 Knowledge Graph(KG)[3][8]를 정의하고 이를 RAG 와 결합하는 KG 기반 RAG 취약점 탐지 방법론을 제안한다.

Ⅱ. 본론

RAG 는 LLM 의 환각을 완화하지만, 텍스트 중심 유사도에 의존하여 취약점 간 구조적 관계를 활용하지 못하며, 선행 연구로 Vul-RAG [4]가 있다. 이는 코드 의미 정보를 기반으로 검색을 수행해 취약점 탐지 성능을 개선했으나, CWE-20 이 CWE-502 로 이어질 수 있는 취약점 간의 의존성을 명시적으로 반영하지 못한다. 따라서 기존 방법론은 개별 취약점 탐지에 그치며, 실제 보안 시나리오를 설명하기에는 부족하다. 본 연구는 이러한 문제를 해결하기 위해 KG 기반 RAG 탐지 기법을 제안한다. KG 는 개체(Entity)와 관계(Relation)를 그래프 형태로 구조화 하는 지식 표현 방식으로, 본 연구에서는 CWE 취약점과 그 상호 관계(ChildOf, PeerOf, CanPrecede 등)를 모델링하여 구축한다[3]. 이를 통해 LLM 은 단순히 개별 CWE 존재 여부만 판별하는 것이 아니라, 어떤 취약점이 다른 취약점으로 확장되거나 선행될 수 있는지를 추론할 수 있다.

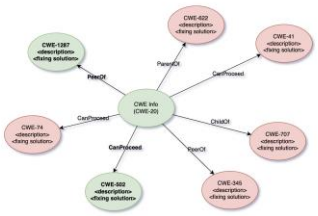


그림 1. CWE Knowledge Graph

제안 방법은 다음 세 단계로 구성된다.

1. Knowledge Graph 구축: CWE 데이터와 관련 문헌을 기반으로 취약점 유형과 그 관계를 구조적으로 정의하여 KG 를 생성한다.

2. 질의 확장: LLM 이 탐지를 수행할 때, KG 에서 제공하는 취약점 간 관계 정보를 질의에

반영한다. 예를 들어, CWE-20 이 탐지된 경우 KG 를 참조하여 CWE-1287, CWE-502 간의 연계 가능성을 함께 질의한다.

3. RAG 기반 탐지: KG 가 제공하는 구조적 지식을 외부 지식으로 활용하여, LLM 이 단일 취약점 뿐 아니라 취약점 시나리오 전체를 추론할 수 있도록 한다. 이는 단일 탐지 정확도를 높일 뿐 아니라, 복합 취약점에 대한 맥락적 설명을 제공한다.

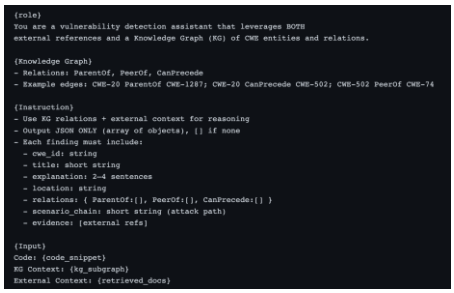


그림 2. KG + RAG prompt

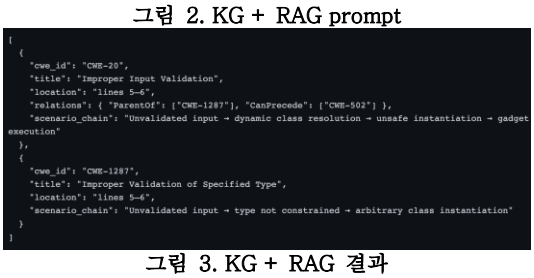


그림 3. KG + RAG 결과

제시한 그림 2 와 그림 3 은 제안한 KG + RAG 기반 탐지 과정의 실제 예시를 보여주며 CWE-20 과 CWE-502 이 연계된 사례로, KG 에 정의된 관계 정보를 활용하여 두 CWE 간의 연관성을 명확히 식별하였다. 이를 통해 단일 취약점 탐지를 넘어 구체적인 취약점 시나리오를 도출할 수 있음을 확인했으며, 취약점 간 관계를 중심으로 한 탐지와 시나리오 식별이 가능함을 시사한다. 또한, KG 는 새로운 CWE 관계나 보안 패턴이 발견될 경우 손쉽게 확장 가능하므로, 지속적으로 진화하는 취약점 탐지 프레임워크를 제공할 수 있으리라 기대한다.

Ⅲ. 결론

본 논문은 LLM 기반 취약점 탐지에서 발생하는 오탐과 환각 문제 [2], 그리고 기존 RAG 접근법 [1][4]의 한계를 지적하고, CWE 를 중심으로 한 취약점 관계 KG 를 구축하여 이를 RAG 와 결합하는 새로운 탐지 방법론을 제안하였다. 제안된 방법은 단일 CWE 탐지를 넘어, 취약점 간 관계와 시나리오를 식별함으로써 탐지 결과의 정확성과 설명 가능성을 동시에 향상시킬 수 있을 거라 기대한다. 향후 연구에서는 KG 자동 구축과 대규모 취약점 데이터셋에 대한 실증적 평가를 수행 함으로써 본 방법론의 실용성과 범용성을 강화할 계획이다.

ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원-대학 ICT 연구센터(ITRC)의 지원을 받아 수행된 연구임(IITP-2025-RS-2024-00438056)

참 고 문 헌

[1] Lewis, Patrick, et al. "Retrieval-augmented generation for knowledge-intensive nlp tasks." *Advances in neural information processing systems* 33 (2020): 9459-9474.

[2] Ji, Ziwei, et al. "Survey of hallucination in natural language generation." *ACM computing surveys* 55.12 (2023): 1-38.

[3] Chen, Xiaojun, Shengbin Jia, and Yang Xiang. "A review: Knowledge reasoning over knowledge graph." *Expert systems with applications* 141 (2020): 112948.

[4] Du, Xueying, et al. "Vul-rag: Enhancing llm-based vulnerability detection via knowledge-level rag." *arXiv preprint arXiv:2406.11147* (2024).

[5] Zhou, Xin, et al. "Large language model for vulnerability detection and repair: Literature review and the road ahead." *ACM Transactions on Software Engineering and Methodology* 34.5 (2025): 1-31.

[6] Zhou, Xin, Ting Zhang, and David Lo. "Large language model for vulnerability detection: Emerging results and future directions." *Proceedings of the 2024 ACM/IEEE 44th International Conference on Software Engineering: New Ideas and Emerging Results*. 2024.

[7] Zhou, Xin, et al. "Comparison of static application security testing tools and large language models for repo-level vulnerability detection." *arXiv preprint arXiv:2407.16235* (2024).

[8] Wen, Yilin, Zifeng Wang, and Jimeng Sun. "Mindmap: Knowledge graph prompting sparks graph of thoughts in large language models." *arXiv preprint arXiv:2308.09729* (2023).

[9] Tamberg, Karl, and Hayretdin Bahsi. "Harnessing large language models for software vulnerability detection: A comprehensive benchmarking study." *IEEE Access* (2025).

엣지 기반 정확도 검증을 통한 계층적 연합학습의 악의적 클라이언트 방어 기법

김건호, 정현수, 길준민*

제주대학교 컴퓨터공학과

kgh9941@stu.jejunu.ac.kr, jhs990909@stu.jejunu.ac.kr, jmgil@jejunu.ac.kr

Defense Scheme Against Malicious Clients in Hierarchical Federated Learning by Edge-Based Accuracy Validation

Geonho Kim, Hyunsu Jeong, Joon-Min Gil*

Dept. of Computer Engineering, Jeju National University

요 약

대규모 네트워크 환경에서 계층적 연합학습(Hierarchical Federated Learning, HierFL)은 서버 - 엣지 - 클라이언트 구조 하에서 확장성과 통신 효율성을 동시에 제공한다. 그러나 라벨 플립(label-flipping)과 같은 악의적 클라이언트는 전역 모델 성능을 심각하게 저하시키며, 더욱이 기존의 단순 집계 방식만으로는 이를 효과적으로 방어하기 어렵다. 본 연구에서는 이러한 문제를 해결하기 위해 엣지에서 정확도 검증을 수행하는 필터링 방식을 제시한다. 제안 기법은 각 클라이언트로부터 전달된 로컬 모델을 엣지가 독립적인 검증 데이터셋으로 평가하고, 정확도가 특정 기준에 부합하는 업데이트만을 서버로 전달하도록 한다. 이러한 방식을 통해 악의적 클라이언트의 비율이 증가하는 상황에서도 제안 기법은 기존 단순 평균 기반 집계 방식 대비 전역 모델의 정확도를 안정적으로 유지할 수 있음을 보여주었다. 이는 본 논문에서 제안하는 기법이 계층적 연합학습 환경에서 보안성과 신뢰성을 향상시킬 수 있음을 증명한다.

Ⅰ. 서 론

현대의 분산 학습 환경에서는 개인정보 보호와 통신 효율성을 동시에 확보할 수 있는 기법이 필수적이다. 연합학습(Federated Learning, FL)은 데이터를 중앙 서버로 직접 전송하지 않고, 클라이언트가 로컬 데이터로 학습한 모델 파라미터만을 공유하여 전역 모델을 구축하는 방식으로 이러한 요구를 충족한다. 그러나 클라이언트 수가 증가하고 네트워크 규모가 확장될수록, 단일 서버 구조에서는 통신 지연과 부하 집중 문제가 발생하여 확장성과 안정성에 한계를 드러낸다. 이를 보완하기 위해 현재 활용되고 있는 계층적 연합학습(Hierarchical Federated Learning, 이하 HierFL)은 기존 연합학습 구조에 서버 - 엣지 - 클라이언트의 계층적 구조를 도입하여 통신 효율성을 높이고 대규모 학습 환경에서도 안정적으로 모델을 학습할 수 있도록 한다. 이 과정에서 엣지(edge)는 복수의 클라이언트 업데이트를 먼저 집계한 뒤 서버로 전달하는 중간 계층으로서, 네트워크 확장성과 학습 속도를 보장하는 중요한 역할을 한다. 그러나 기존 HierFL 연구들은 엣지를 단순한 집계 노드로만 활용하는 경향이 강하다[1]. 이러한 구조에서는 악의적 클라이언트가 조작한 업데이트가 있을 경우 엣지를 통해 그대로 서버로 전달됨으로 모델 성능 저하가 발생하며, 더욱이 라벨 플립(label-flipping)과 같은 모델 오염 공격은 전역 모델의 정확도를 크게 저하시킨다. 즉, HierFL은 확장성과 통신 효율성 측면에서 장점을 가지지만, 보안성과 강건성 측면에서는 여전히 취약하다.

따라서, 본 연구에서는 이러한 문제를 해결하기 위해 엣지에서의 정확도 검증 기반 필터링 기법을 제안한다. 제안 기법에서 엣지는 단순히 클라이언트 모델을 집계하는 것이 아니라, 독립적인 검증 데이터셋을 활용하여 각 클라이언트 업데이트의 정확도를 평가하고, 기준 미달인 업데이트는 서버 전송 단계에서 배제한다. 이를 통해 악의적 클라이언트가 전역 모델에 미치는 영향을 최소화하고, 학습 안정성과 신뢰성을 강화시킨다.

Ⅱ. 제안 기법

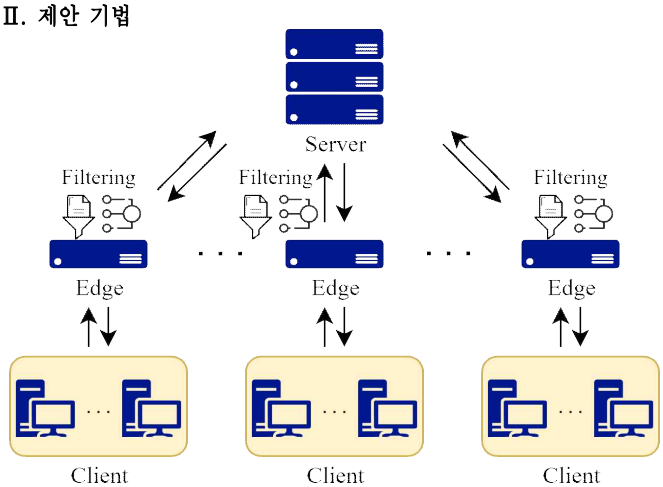


그림 1 클라이언트 필터링 기능을 갖는 계층적 연합학습 구조

2.1 시스템 모델

본 연구의 제안 기법은 그림 1과 같이 서버 - 엣지 - 클라이언트 구조를 갖는 계층적 연합학습 환경에 기반한다. 전역 서버는 기본 모델을 준비하고 배포하는 역할을 한다. 이후 클라이언트는 개별적으로 보유한 데이터를 기반으로 로컬 학습을 수행하며, 학습된 모델의 파라미터를 엣지 서버로 전송한다. 각 엣지 서버는 연결된 클라이언트들로부터 수집한 파라미터를 정확도 임계값을 기준으로 필터링 및 집계하여 엣지 모델을 형성한다. 이렇게 생성된 엣지 모델은 다시 상위의 전역 서버로 전달되며, 이때 전역 서버는 모든 엣지 서버로부터 모인 결과를 통합하여 최종 전역 모델을 갱신한다. 갱신된 전역 모델은 다시 각 엣지 서버를 거쳐 클라이언트로 재배포되는 과정이 반복적으로 수행됨으로써 전체 시스템은 점차적으로 성능이 향상된 전역 모델로 완성되어 간다.

2.2 정확도 검증 기반 필터링 알고리즘

Algorithm 1: Hierarchical FL with Edge Filtering

Input: Clients $\mathcal{C}$, Edges $\mathcal{E}$, Rounds R, Local steps τ, Validation set D_{val}
Output: Global model w^R

Initialize w^0 and broadcast to edges and clients;
for $r = 1$ **to** R **do**
  foreach *client* c **do**
    Update w_c^r from w^{r-1} by τ local steps;
    Send w_c^r to its edge;
  end
  foreach *edge* e **do**
    Evaluate $\{w_c^r\}$ on D_{val};
    Select $S_e = \{c \mid a_c \geq \text{median}(\{a_c\})\}$;
    Aggregate $w_e^r \leftarrow \frac{1}{|S_e|} \sum_{c \in S_e} w_c^r$;
    Send w_e^r to the server;
  end
  $w^r \leftarrow \frac{1}{|\mathcal{E}|} \sum_{e \in \mathcal{E}} w_e^r$;
  Broadcast w^r to all clients;
end
return w^R;

엣지 서버는 클라이언트들로부터 수집된 모델 업데이트를 집계하기 전에 알고리즘 1에서 제시한 정확도 기반 필터링 알고리즘을 적용한다. 엣지 서버는 전역 검증 데이터셋으로 각 클라이언트 모델을 평가하여 정확도 a_c를 산출하고, 필터링 정책으로 정해진 임계값에 따라 업데이트 수용 여부를 결정한다. 이는 기존 연구[2]에서 보고된 바와 같이, 악의적 클라이언트가 제출한 업데이트는 정상 클라이언트에 비해 전역 검증 데이터셋에서 성능이 현저히 낮게 나타난다는 결과에 근거한다. 본 연구에서는 필터링 정책으로 중앙값을 사용하며, 이는 각 엣지에서 평가된 클라이언트 업데이트 정확도들의 중앙값을 기준치로 삼아 $a_c \geq median(a_c)$를 만족하는 업데이트만 통과시킨다. 통과한 업데이트에 대해 집계함수인 FedAVG를 적용하여 엣지 모델을 산출하고 이를 전역 서버로 전송한다.

Ⅲ. 성능 평가

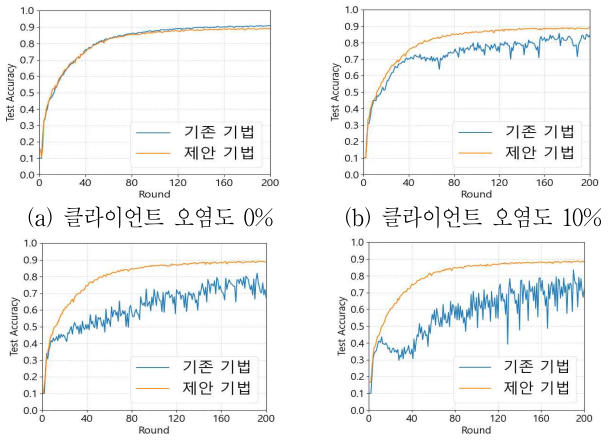
3.1 실험 환경 구성

전체 학습 데이터의 10%는 클래스별 균등 분할을 통해 전역 검증 데이터셋으로 분리하고, 각 엣지 서버에서 모델 업데이트의 신뢰도 평가에 활용하였다. 나머지 90%의 학습 데이터는 클라이언트 수에 맞추어 데이터를 분배하였다. 악의적 클라이언트는 무작위로 선택되며, 라벨 플립 공격으로 모사하였다. 공격 클라이언트 비율은 0 - 50% 범위에서 10% 간격으로 변화시켰다. 학습은 총 200 라운드 동안 진행되었으며, 각 클라이언트는 라운드마다 5회의 로컬 업데이트를 수행하였다. 학습 모델은 ResNet-18 모델을 기반으로, SGD 옵티마이저(learning rate=0.05, momentum=0.9, weight decay=$5\times e^{-4}$)를 적용하였다. 한편, 제안 기법의 성능평가를 위해 사용되는 데이터셋으로 CIFAR-10이 사용되었다. 이러한 데이터셋에 기반하여 클라이언트 오염의 검증 기능을 갖는 제안 기법과 클라이언트 오염 검증이 없는 기존 기법을 정확도 관점에서 성능을 비교하였다.

3.2 성능 결과

그림 2는 클라이언트 오염도에 따른 라운드별 전역 모델 정확도의 변화를 보여준다. 그림 2의 결과를 살펴보면, 오염 클라이언트 비율이 증가할수록 기존 기법은 모델 정확도의 성능 저하와 함께 라운드가 증가할수록 정확도의 변동성이 점점 증가한다. 반면, 제안 기법은 전반적으로 안정적인 수렴 과정을 보이며 비교적 일관된 성능을 유지한다.

그림 3은 baseline 대비 정확도 차이의 표준편차를 나타낸다(baseline으로 오염 클라이언트가 없는 기존 기법이 사용됨). 기존 기법은 클라이언트의 오염도가 높아질수록 표준편차가 급격히 증가하여 성능 변동성이 큰 반면, 제안 기법은 표준편차가 낮게 유지되며 안정적인 성능을 보인다.



(a) 클라이언트 오염도 0% (b) 클라이언트 오염도 10%

(c) 클라이언트 오염도 20% (d) 클라이언트 오염도 30%

그림 2 클라이언트 오염도에 따른 라운드별 테스트 정확도

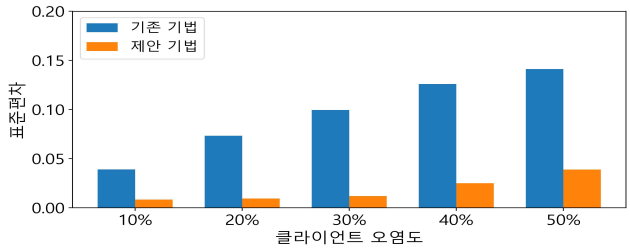


그림 3 Baseline과의 정확도 차이값의 표준편차

Ⅳ. 결론

본 연구는 서버 - 엣지 - 클라이언트 구조의 계층적 연합학습에서 악의적 클라이언트 공격이 전역 모델에 미치는 영향을 분석하고, 이를 완화하기 위한 정확도 검증 기반 중앙값 필터링 기법을 제안하였다. 제안 기법은 각 엣지 서버가 전역 검증 데이터셋으로 클라이언트 업데이트를 평가하고, 중앙값 이상의 정확도를 보이는 모델만 수용하는 방식이다.

CIFAR-10 데이터셋과 ResNet-18 모델을 활용한 실험에서는 공격 클라이언트 비율이 증가할수록 기존 기법의 정확도가 급격히 하락한 반면, 제안 기법은 성능 저하를 효과적으로 억제하였다. 이러한 결과는 엣지에서의 필터링이 계층적 연합학습 환경에서 신뢰성 있는 학습을 보장하는 핵심 요소임을 시사한다. 향후 연구에서는 데이터 포이즈닝, 업데이트 조작 등 다양한 비정상 행위에 대한 확장과 복합 정책 적용을 통해 안정성을 더욱 강화할 예정이다.

ACKNOWLEDGMENT

본 과제(결과물)는 2025년도 교육부 및 제주도의 재원으로 제주RISE센터의 지원을 받아 수행된 지역혁신중심 대학지원체계(RISE)의 결과입니다(2025-RISE-17-001).

참 고 문 헌

[1] T. Do, D. A. Tran, and A. Vo, “Edge assignment in edge federated learning,” SN Applied Sciences, vol. 5, no. 281, 2023.

[2] S. Venkateswaran, Q. Shaikh, S. Singh, J. Abraham, and A. Bochare, "Defense Mechanism to Thwart Model Poisoning on Non-IID Data based Federated Learning for Credit Fraud Detection System," 2025 International Conference on Emerging Smart Computing and Informatics (ESCI), Pune, India, 2025

범용 GPU 환경에서 YOLOv8 모델 크기별 양자화 기반 최적화 성능 분석

박경택, 한동석*
경북대학교
balam3298@knu.ac.kr, *dshan@knu.ac.kr

Quantization-based Optimization Performance Analysis of YOLOv8 Model Variants on General-purpose GPUs for Real-time Object Detection

Gyeong-Taek Park, Dong Seog Han*
Kyungpook National Univ.

요 약

실시간 객체 검출 응용에서 모델 양자화는 추론 속도와 메모리 효율성 향상을 위해 필수적이다. 기존 연구는 단일 모델이나 특정 양자화 기법에 국한되어 YOLOv8 전체 패밀리의 체계적 분석이 부족하다. 본 논문은 RTX 4070 Super GPU에서 YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l 모델의 TensorRT FP16/INT8 양자화 성능을 BDD100K 자율주행 데이터셋 500장을 사용하여 정량적으로 분석하였다. COCO-BDD100K 공통 9개 클래스를 대상으로 BDD100K 기반 도메인 일치 캘리브레이션을 적용하였다. 5회 독립 실행과 20회 반복 측정을 통해 통계적 신뢰성을 확보하였다. 실험 결과, INT8 양자화로 평균 18.0% FPS 향상을 달성하였으며, YOLOv8n에서 최대 24.3% 향상을 보였다. YOLOv8s INT8이 212.2 FPS와 mAP 0.367의 최적 균형점을 제공하였고, 200 FPS를 초과한 조합은 YOLOv8n FP16/INT8, YOLOv8s FP16/INT8 네 가지였다. 본 논문은 범용 GPU 환경에서 실시간 객체 검출을 위한 정량적 모델 선택 가이드라인을 제공한다.

Ⅰ. 연구 배경 및 목적

딥러닝 기반 객체 검출 기술은 자율주행, 보안 감시, 로봇 비전 등 다양한 분야에서 핵심 기술로 자리 잡고 있다. 특히 실시간 처리가 요구되는 응용에서는 검출 정확도와 더불어 추론 속도가 중요한 성능 지표가 된다. YOLO(You Only Look Once) 계열 모델은 단일 네트워크로 객체 검출과 분류를 동시에 수행하여 높은 처리 속도를 제공하며, 최신 YOLOv8은 향상된 아키텍처와 훈련 기법을 통해 정확도와 효율성을 크게 개선한다 [1].

실제 배포 환경의 자원 제약으로 모델 경량화는 필수적이며, 양자화는 대표적인 최적화 기법이다 [2]. 기존 연구들은 주로 단일 모델이나 특정 양자화 기법에 초점을 맞춰 진행되었으며 [2] [3], YOLOv8 전체 모델 계열에 대한 체계적인 양자화 성능 분석은 매우 드문 실정이다. 따라서 실제 배포 환경에서 최적 모델 선정을 위한 정량적 가이드라인이 부족하다.

본 논문에서는 YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l 모든 변형에 대해 FP32, FP16, INT8 양자화 정밀도에 따른 속도 - 정확도 트레이드오프를 정량적으로 분석하고, 실시간 객체 검출 응용을 위한 최적 모델 구성을 제시한다.

Ⅱ. YOLOv8 모델 양자화 성능 분석

2.1 실험 환경 및 데이터셋

본 실험은 RTX 4070 Super GPU, CUDA 12.8, TensorRT 10.6.0 환경에서 수행되었다. 평가 데이터셋으로는 Berkeley DeepDrive (BDD100K) validation set에서 무작위로 선택한 500장의 이미지를 사용하며, 총 9,350개의 Ground Truth 객체가 포함되어 있다 [4]. 모든 이미지는 640×640 해상도로 전처리된다.

2.2 TensorRT 엔진 생성 및 최적화

COCO [5]로 사전 학습된 YOLOv8 모델을 BDD100K 데이터셋에서 평가하기 위해 두 데이터셋 간 공통 객체 클래스를 추출한다. COCO 80개 클래스와 BDD100K 10개 클래스 중 공통으로 포함된 9개 클래스(person, bicycle, car, motorcycle, bus, train, truck, traffic light, traffic sign)를 대상으로 성능 평가를 수행하며, BDD100K의 rider 클래스는 COCO에 직접 대응되지 않아 제외한다.

INT8 양자화를 위해 캘리브레이션에는 BDD100K validation set에서 별도로 분리한 1,000장을 사용하고, 성능 평가는 나머지 500장을 활용하여 도메인 일치 Post-Training Quantization을 수행한다.

생성된 TensorRT 엔진 크기는 FP16 기준 YOLOv8n 9.0MB, YOLOv8s 24.0MB, YOLOv8m 53.3MB, YOLOv8l 87.3MB이며, INT8 양자화 적용 시 각각 6.0MB, 14.0MB, 30.0MB로 축소되어 평균 약 40% 감소가 확인된다.

2.3 성능 평가 방법론

성능 평가는 배치 크기 1 조건에서 warm-up 50프레임 이후 20회 반복 측정을 수행한다. 평가 지표로는 FPS, 추론 시간, mAP@0.5, Precision, Recall, F1-Score를 사용하며, 신뢰도 임계값은 0.25, IoU 임계값은 0.5로 설정한다.

정확도 평가는 총 500장의 이미지를 대상으로 산출하고, 속도 평가는 이 중 30장을 무작위로 선별하여 12개 모델 - 정밀도 조합에 대해 반복 측정(총 360회)을 수행한다. 입력 이미지 수에 따른 속도 측정 편차가 미미하다고 판단하여, 효율적인 반복 실험을 위해 30장만 활용한다.

모든 측정은 5회 독립적으로 수행하며, 결과의 통계적 신뢰성을 확보하기 위해 95% 신뢰구간을 계산한다.

2.4 실험 결과

표 1은 모든 모델과 정밀도 조합에 대한 성능 측정 결과를 제시한다. 본 연구에서는 FP32 추론 성능을 기준선(baseline)으로 삼아 FP16 및 INT8 양자화 적용 후의 성능 변화를 비교하였다. INT8 양자화는 YOLOv8n에서 최대 24.3% FPS 향상(202.6 → 251.8 FPS), 평균 18.0% 향상을 보였으며, FP16은 평균 8.2% 향상과 함께 정확도 손실은 거의 없거나 일부 모델에서 소폭 개선되었다.

표 1. YOLOv8 모델별 양자화 성능 비교

모델	양자화	FPS (±표준편차)	mAP@0.5	Precision	Recall	F1-Score
YOLOv8n	FP32	202.6 ± 6.7	0.291	0.555	0.394	0.461
	FP16	214.8 ± 7.1	0.294	0.558	0.397	0.464
	INT8	251.8 ± 8.3	0.28	0.552	0.381	0.451
YOLOv8s	FP32	189.4 ± 6.2	0.365	0.583	0.471	0.521
	FP16	203.7 ± 6.8	0.368	0.585	0.473	0.523
	INT8	212.2 ± 7.0	0.367	0.584	0.471	0.522
YOLOv8m	FP32	145.3 ± 4.8	0.385	0.615	0.485	0.542
	FP16	156.2 ± 5.2	0.392	0.618	0.492	0.548
	INT8	174.5 ± 5.8	0.38	0.612	0.48	0.537
YOLOv8l	FP32	97.1 ± 3.2	0.37	0.655	0.458	0.54
	FP16	108.3 ± 3.6	0.378	0.658	0.466	0.546
	INT8	112.2 ± 7.0	0.367	0.654	0.454	0.536

* ± 값은 5회 독립 측정의 표준편차(Standard Deviation)를 나타냄. 모든 측정은 동일한 환경에서 수행됨.

표 2는 양자화 기법별 FPS 향상률을 요약한 것이다. INT8 양자화는 평균 18.0%의 향상을 달성하였으며, YOLOv8n에서 최대 24.3%의 개선 효과가 나타났다. FP16 양자화는 평균 8.2%의 향상을 보였다.

표 2. 양자화 기법별 FPS 향상률 (vs FP32 기준)

모델	FP16 향상률	INT8 향상률	200 FPS 초과
YOLOv8n	6.00%	24.30%	FP16, INT8
YOLOv8s	7.60%	12.00%	FP16, INT8
YOLOv8m	7.50%	20.10%	없음
YOLOv8l	11.50%	15.60%	없음
평균	8.20%	18.00%	4개 (12개 중)

200 FPS를 초과한 조합은 YOLOv8n FP16, YOLOv8n INT8, YOLOv8s FP16, YOLOv8s INT8의 네 가지였다. 이들은 모두 30 FPS 실시간 비디오 처리 요구사항을 충분히 만족하였으며, 자율주행 시스템과 실시간 감시 시스템 등 다양한 고속 응용 환경에서의 실용적 배포 가능성을 입증한다.

2.5 최적 모델 구성 분석

실시간 객체 검출 응용에 적합한 최적 구성을 도출하기 위해 속도 - 정확도 균형을 분석하였다. YOLOv8s INT8 모델은 212.2 FPS의 처리 속도와 mAP 0.367을 동시에 달성하여 가장 균형 잡힌 성능을 보였다. 극고속 처리가 요구되는 경우에는 YOLOv8n INT8(251.8 FPS)이 적합하며, 상대적으로 높은 정확도가 필요한 경우에는 YOLOv8m FP16(mAP 0.392)이 더 적합하다.

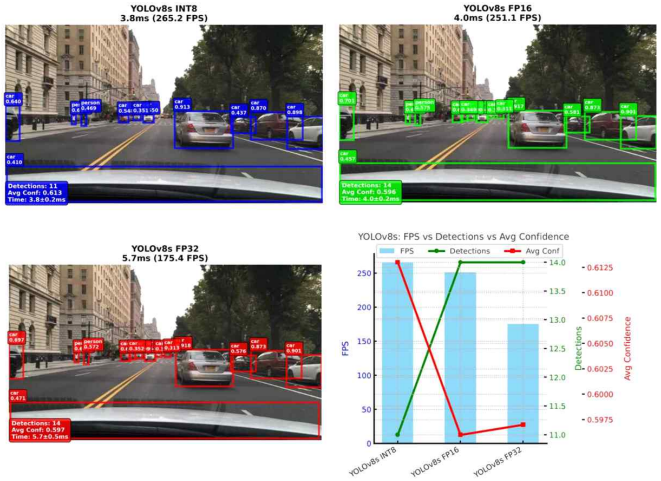


그림 1. BDD100K 실제 검출 결과 예시

그림 1. BDD100K 검증 세트 실제 검출 결과 예시 (IoU=0.5, conf=0.25, 공통 9개 클래스). 복잡한 도시 환경 장면에서 FP32와 FP16 모델은 각각 14개 객체를 검출하였으나, INT8 모델은 11개만 검출하여 소형 및 원거리 객체가 일부 누락되었다.

하단 그래프는 FPS, 검출 객체 수, 평균 신뢰도를 함께 나타내며, INT8은 251.8 FPS로 가장 높은 속도를 기록했으나 정확도 저하가 관찰되었고, FP16은 속도 향상과 정확도 유지의 균형을 보여주었다.

Ⅲ. 연구 결과 및 향후 연구 방향

본 논문는 YOLOv8 전체 모델 계열에 대한 체계적인 양자화 성능 분석을 통해 다음 결과를 도출하였다. 첫째, TensorRT INT8 양자화를 통해 평균 18.0% FPS 향상을 달성하였으며, YOLOv8n에서 최대 24.3%의 개선 효과를 확인하였다. 둘째, YOLOv8s INT8 모델이 212.2 FPS와 mAP 0.367을 동시에 달성하여 속도와 정확도의 균형 면에서 실시간 응용에 최적의 구성임을 입증하였다. 셋째, 200 FPS를 초과한 조합(YOLOv8n FP16/INT8, YOLOv8s INT8)은 실시간 비디오 처리 요구사항을 크게 상회하여 고속 응용 환경에서의 활용 가능성을 보여준다.

향후 연구에서는 엣지 디바이스 최적화, 전력 효율 분석, 동적 해상도 적응 기법 등을 통해 실시간 객체 검출 시스템의 실용성을 더욱 향상시킬 예정이다. 본 연구 결과는 자율주행, CCTV 감시, 로보틱스 등에서 효율적인 모델 배포를 위한 정량적 가이드라인을 제공한다.

ACKNOWLEDGMENT

이 성과는 정부(과학기술정보통신부)의 재원으로 과학기술사업화진흥원의지원을 받아 수행된 연구임(RS-2025-25444562).

참 고 문 헌

[1] G. Jocher, A. Chaurasia, and J. Qiu, "YOLO by Ultralytics," GitHub repository, 2023. [Online]. Available: https://github.com/ultralytics/ultralytics, accessed: 2025-08-28.

[2] Jacob, B., *et al.*, "Quantization and Training of Neural Networks for Efficient Integer-Arithmetic-Only Inference," *Proc. CVPR*, pp. 2704-2713, 2018.

[3] Nagel, M., *et al.*, "Data-Free Quantization Through Weight Equalization and Bias Correction," *Proc. ICCV*, pp. 1325-1334, 2019.

[4] Yu, F., *et al.*, "BDD100K: A Diverse Driving Dataset and Challenges for Open Vocabulary Object Detection," *Proc. CVPR*, pp. 2636-2644, 2020.

[5] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick,"Microsoft COCO: Common objects in context," *European Conference on Computer Vision (ECCV)*, pp. 740 - 755, 2014.

실내 위치 추정을 위한 우도 기반 동적 데이터 가지치기

이서인, 문하은, 한승우, 김민석*
충남대학교, *충남대학교

sslynn22@o.cnu.ac.kr, cherishuu02@o.cnu.ac.kr, hsw2266@g.cnu.ac.kr, * kms4105@cnu.ac.kr

Likelihood based Dynamic Data Pruning for Indoor Localization

Lee Seo In, Moon Ha Eun, Han Seung Woo, and Kim Min Seuk*
Chungnam National Univ., * Chungnam National Univ.

요 약

본 논문은 Wi-Fi 채널 상태 정보(Channel State Information, CSI) 기반 실내 위치 추정 문제에서 실내 환경의 다중경로(Multipath) 및 NLoS(Non-Line of Sight) 등으로 유발되는 잡음·이상치로 인한 성능 저하 문제를 해결하기 위해, 학습 과정에서 샘플의 신뢰도를 실시간으로 평가해 유효하지 않은 데이터를 일시적으로 배제하는 우도(Likelihood) 기반 동적 데이터 가지치기를 제안한다. 실험 결과, 제안 기법은 기존 합성 신경망(Convolutional Neural Network, CNN) 모델(3.99 m) 대비 약 13 % 향상된 3.47 m 의 평균 위치 추정 오차를 달성하였다. 이는 학습 전 데이터의 정적 선별 방식이 지닌 한계를 극복하고, 모델의 강건성과 실내 위치 인식 정확도를 효과적으로 향상시킬 수 있음을 입증한다.

Ⅰ. 서 론

Wi-Fi 채널 상태 정보(Channel State Information, CSI)를 활용한 실내 위치 추정은 스마트 팩토리, 자율주행 등 다양한 분야에서 핵심 기술로 사용되고 있다. 그러나 다중경로(Multipath), NLoS(Non-Line of Sight) 등이 발생하는 복잡한 실내 환경에서 수집된 CSI 데이터는 잡음(Noise)과 이상치(Outlier)가 섞이게 된다 [1]. 이렇게 신뢰도가 낮은 데이터는 위치 추정 모델의 정확도와 일반화 성능을 저해하는 원인이 된다.

이러한 비선형적인 신호 패턴을 효과적으로 학습하기 위해 합성곱 신경망(Convolutional Neural Network, CNN)이 활용되고 있다 [2]. CNN 은 모든 데이터를 동일 가중으로 학습하기 때문에 잡음·이상치가 포함된 데이터의 영향을 그대로 받아, 위치 추정 성능의 한계를 보인다. 이를 완화하고자 전체 데이터를 무선별적으로 학습하는 기존 방식에서 벗어나, 신뢰도가 높은 핵심 데이터만을 선별하여 학습하는 우도(Likelihood) 기반 동적 데이터 가지치기(Dynamic Data Pruning) 모델을 제안한다. 모델의 학습 과정에서 각 데이터 샘플에 대한 예측 손실을 우도의 역지표로 정의하여 손실값이 큰, 즉 우도가 낮은 샘플을 실시간으로 제거한다. 이를 통해 모델의 학습 단계에서 신뢰도가 높은 데이터만을 선별하여 학습하므로, 강건성을 확보하고 최종 위치 추정 정확도의 향상을 달성할 수 있음을 보인다.

Ⅱ. 본론

2.1 모델 아키텍처

본 논문에서 제안하는 우도 기반 동적 데이터 가지치기는 모델 학습 과정에서 현재 모델 상태를 반영하여 우도가 높은 샘플만 선택하고 나머지는 일시적으로 제외한다. 전체 프로세스는 (가) 우도 기반 샘플 신뢰도 산출과 (나) 동적 데이터 가지치기의 두 단계로 구성된다.

(가) 우도 기반 샘플 신뢰도 산출

데이터 가지치기의 판별 기준으로 각 샘플(X_i) 우도를 활용하며, 이를 정량화하기 위해 모델의 평균 제곱 오차(Mean Squared Error, MSE)를 우도의 역지표로 사용한다. 이는 학습 과정에서 실시간으로 데이터의 신뢰도를 평가하여, 모델 성능에 악영향을 미치는 잡음 및 이상치를 동적으로 학습에서 배제하기 위함이다.

$$L_i = L_{MSE}\left(f_\theta\left(Rep(X_i)\right), Y\right) \quad (i = 1, \cdots, N) \tag{1}$$

현재 모델 파라미터 θ 를 기준으로 배치 내 개별 샘플에 대한 손실 값 L_i 를 수식 1 과 같이 산출한다. CNN 모델이 M 의 차원을 가지는 개별 샘플(X_i) 각각의 신뢰도를 평가할 수 있도록, 단일 데이터 샘플은 복제 연산($Rep(X_i)$)을 통해 N 번 복제되어 모델의 2 차원 입력 형식($M \times N$) 맞는 텐서 형태로 변환된다. 이 복제된 텐서는 CNN 모델 f_θ 의 입력으로 사용되어 예측 좌표 ($f_\theta(Rep(X_i))$) 를 출력하며, 실제 정답 좌표 Y 와의 MSE 를 계산하여 해당 샘플의 손실값 L_i 를 산출한다.

$$S_L(X_i) = \exp\left(-L_i\right) \tag{2}$$

수식 2 와 같이, 산출된 손실값 L_i 를 exp 지수 함수에 매핑하여 최종 우도 값 ($S_L(X_i)$)을 계산한다. 이 변환을 통해 손실이 작을수록 $S_L(X_i)$ 은 1 에 가까워지고, 손실이 클수록 0 에 수렴하여 각 샘플의 신뢰도를 나타내는 최종 지표가 된다.

(나) 동적 데이터 가지치기

우도 값($S_L(X_i)$)을 기준으로, 우도가 낮은 샘플들을 실시간으로 학습에서 제거(pruning)한다.

$$S = \{X_i \mid rank_{desc}(S_L(X_i)) \le K\} \tag{3}$$

선별된 샘플의 집합 S 는 수식 3 과 같이, 전체 샘플(X_i) 중에서 해당 우도 값($S_L(X_i)$)의 순위가 K 이하인

샘플들로 정의한다. 여기서 $rank_{desc}$ 는 값을 내림차순(Descending)으로 정렬했을 때의 순위를 의미하며, 이는 우도가 가장 높은 K 개의 샘플을 선택하는 것을 뜻한다. 이때, 전체 샘플 수(N)대비 선택되는 샘플 수(K)의 비율, 즉 가지치기 비율(K/N)은 모델이 얼마만큼의 데이터를 선별할지를 결정하는 핵심 파라미터가 된다. 본 실험에서는 전체 샘플 그룹(N)을 10 으로, 선택할 샘플 수(K)를 5 로 설정하여, 매 학습 단계마다 신뢰도가 낮은 하위 50 % 의 샘플을 가지치기하였다.

데이터 가지치기 후, 선택된 K 개의 핵심 샘플은 다시 원래 배치 크기인 N 개를 유지하도록 복제되어 학습에 사용된다. 이는 학습 과정에서 CNN 모델의 입력 텐서 크기를 일정하게 유지하여 학습 안정성을 높이기 위함이다. 이렇게 선별된 핵심 샘플 집합 S 만을 사용하여 모델의 파라미터 θ 를 업데이트한다. 역전파(Backpropagation) 과정에서 사용되는 그래디언트(Gradient)가 신뢰도 높은 데이터로부터만 계산되므로, 잡음 및 이상치가 파라미터 업데이트에 미치는 악영향을 효과적으로 차단할 수 있다. 또한 학습이 반복될수록 모델의 성능이 향상되고, 향상된 모델은 다시 다음 Epoch 에서 더 정확하고 신뢰도 높은 데이터를 선별하는 선순환 구조를 형성하게 된다.

2.2 실험 환경 및 성능 평가

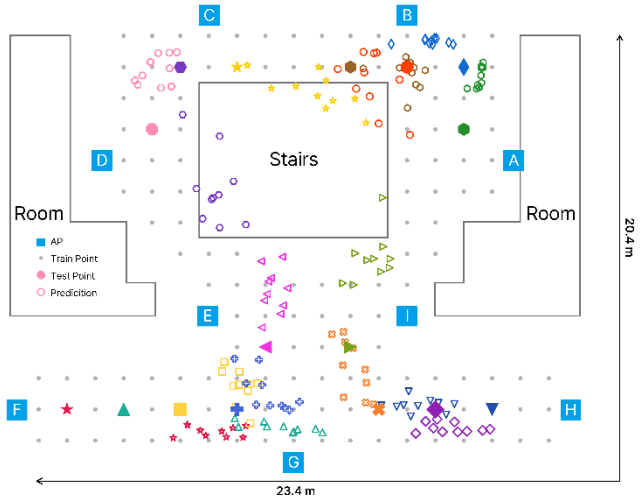


그림 1. 실험 환경 토폴로지 및 성능 평가 결과

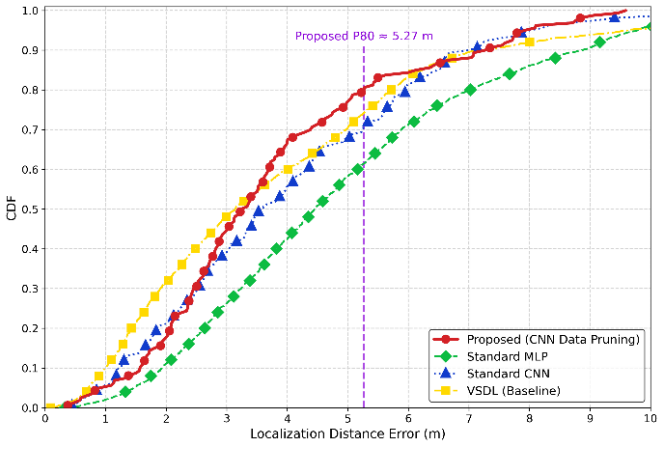


그림 2. 위치 추정 오차 CDF

본 논문에서 제안하는 모델의 성능은 그림 1 과 같이 유동 인구가 많은 충남대학교 도서관 1 층 실내 환경에서 검증하였다. 평가에는 총 141 개 지점에서 포인트당 100 회 측정한 1,890 차원 CSI 기반 특징 벡터를 사용하였다. 각 지점의 100 개 패킷은 10 개의 연속 구간으로 묶어 10 개의 이미지 청크로 변환하였고, 각 청크는 (10×1,890) 행렬로 CNN 에서 하나의 이미지처럼 처리되도록 구성하였다. 이로써 모델이 짧은 시간 구간의 변화(10 패킷)와 주파수·공간적 특성(1,890 차원 특징)을 동시에 학습할 수 있도록 하였다. 그림 1 에서 볼 수 있듯이, 테스트 포인트에 대한 실내 위치 추정 결과가 안정적으로 실제 위치에 자리 잡고 있다.

그림 2 는 타 알고리즘과 비교하여, 제안한 방법의 실내 위치 인식 정확도를 보여준다. 제안한 우도 기반 동적 데이터 가지치기는 동일 조건에서 평균 위치 추정 오차 3.47 m 로, 비교 모델인 MLP 4.65 m, CNN 기반 CiFi 3.99 m, 뷰 선별 알고리즘을 제안하는 VSDL [3] 4.01 m 대비 향상된 위치 추정 정확도를 기록하였다. 특히, 동일한 CNN 기반의 CiFi 모델(3.99 m)과 비교했을 때, 모든 데이터를 학습하는 대신 우도가 높은 데이터만을 선별하여 학습함으로써 오차율을 약 13 % 개선하는 효과를 확인했다. 그림 2 와 같이 P80(상위 80 % 보장 오차)가 약 5.27 m 로 나타나, 동일한 보장수준에서 타 방법들보다 더 낮은 임계 오차를 제공하였다.

Ⅲ. 결론

본 연구는 Wi-Fi CSI 기반 실내 위치 추정에서 잡음·이상치로 인한 성능 저하 문제를 해결하고자, 학습 과정에서 데이터 신뢰도를 실시간으로 평가하고 유효하지 않은 샘플을 동적으로 배제하는 우도 기반 데이터 가지치기 기법을 제안했다. 실험 결과, 제안 기법은 기존 CNN 모델(3.99 m) 대비 약 13 % 향상된 3.47 m 의 평균 위치 추정 오차를 달성하였다. 이는 데이터의 정적 선별 방식이 지닌 한계를 극복하고, 모델의 강건성과 위치 추정 정확도를 효과적으로 향상할 수 있음을 입증한다.

향후 연구로는 본 연구의 샘플 평가 개념을 추론 단계까지 확장하여, 신뢰도 기반의 가중 평균이나 앙상블 기법을 통해 예측 안정성을 높이는 방안을 모색할 것이다. 또한, 고정된 가지치기 비율을 학습 상황에 따라 동적으로 조절하여 학습 효율을 최적화하는 연구를 진행할 계획이다.

ACKNOWLEDGMENT

본 연구는 충남대학교 학술연구비에 의해 지원되었음.

참 고 문 헌

[1] Tzur, Asaf, Ofer Amrani, and Avishai Wool. “Direction finding of rogue Wi-Fi access points using an off-the-shelf MIMO-OFDM receiver.” Physical Communication 17 (2015): 149-164.

[2] Wang, X., Wang, X., and Mao, S. “CiFi: Deep convolutional neural networks for indoor localization with 5 GHz Wi-Fi,” in IEEE International Conference on Communications (ICC), 2017, pp. 1-6.

[3] 한승우, 문하은, 이서인, 강민서, and 김민석, "Viewpoint-Selective Deep Learning 을 활용한 NLoS 환경에서의 WiFi CSI 기반 실내 측위," in 한국통신학회 학술대회논문집, 2025, pp. 1045-1046.

제지 공정의 불확실성을 반영하기 위한 Neural SDE 기반 강화학습 연구

유동건, 이상금, 도윤미, 최진영*
국립한밭대학교, *한국전자통신연구원

20211917@edu.hanbat.ac.kr, sangkeum@hanbat.ac.kr, {ydoh, *choij0}@etri.re.kr

Neural SDE-Based Reinforcement Learning for Uncertainty-Aware Papermaking

Donggeon Yu, Sangkeum Lee, Yoonmee Doh, Jinyoung Choi*
Hanbat National Univ., *Electronics and Telecommunications Research Institute

요 약

본 논문은 제지 공정에서 스팀에너지 절감을 위해 스팀 관련 주요 변수를 선별하고, 외부 요인에 의해 발생할 수 있는 이상 패턴과 복합적 이상 징후를 예측하기 위한 Neural SDE 기반 강화학습 모델을 제안한다. 제지 공정은 기상 변화, 지절, 설비 변동 등 생산성과 품질에 직접적인 영향을 미치는 불확실성을 내포한다. 기존 LSTM 기반 모델은 시계열 패턴을 학습하는 데 효과적이나, 이러한 확률적 변동성을 반영하기 어렵다. 이를 보완하기 위해 본 연구에서는 Neural SDE를 활용하여 공정의 확률적 특성을 반영하였으며, 다중 보상 함수를 통해 스팀에너지 효율을 포함한 다양한 목표를 동시에 고려하였다. 실험 결과, SDE와 LSTM을 결합한 모델의 예측률이 약 5% 증가하였다. 제안한 모델을 통해 제지 공정의 불확실성을 반영하면서 공정 최적화에 기여할 수 있는 가능성을 확인하였다.

Ⅰ. 서 론

통계에 따르면 2020년 국내 제지 산업 열에너지 소비의 약 11.9%를 차지한다[1]. 국가 탄소중립 목표에 부합하려면, 공정 단계에서의 효율 개선이 가장 직접적인 감축 수단이다[2]. 제지공정의 경우, 초지공정에서 60% 이상의 스팀이 사용된다. 또한, 지절로 인한 생산 지종 변경, 계절 등 외부 요인으로 인해 운전 조건이 변경된다. 따라서 본 연구에서는 스팀사용량을 최적화하기 위해 스팀과 관련된 변수를 선별하고, 운전에 따른 편차를 최소화하기 위한 강화학습 모델을 개발하였다. 강화학습에서 Agent는 Environment와 상호작용하며 정책을 학습한다. Environment는 물리 방정식 또는 인공지능으로 구현될 수 있다. 그러나 단순 물리 방정식이나 단일 인공지능은 제지공정에서 발생하는 급격한 변동성, 제품 생산 주기, 노이즈 등 실제 공정 환경을 반영하기 어렵다. 단순 모델의 일반화를 배제하면서, 품질, 에너지 효율과 같이 다중 목표를 동시에 최적화하는 공정 맞춤형 제어 방법이 필요하다. 본 연구는 LSTM 기반 Environment에 Neural Stochastic Differential Equation (SDE)을 추가하여 실측값에 근접한 값을 예측하는 모델을 개발하였다. 또한, 기존 작업자 운전 조건 대비 제안한 방법 운전 조건의 성능을 점수화하여 비교하였다.

Ⅱ. 본론

2.1 Neural SDE+ LSTM 강화학습 개념도

그림 1은 Neural SDE이 추가된 LSTM 기반(이하 SDE+ LSTM) 강화학습 모델의 개념도이다. LSTM이 1차로 예측값을 출력하고, Neural SDE을 거쳐 최종 예측값을 출력한다. 공정 환경에서는 외부 요인들이 변함에 따라 실제 운영의 데이터가 달라지는 데이터 드리프트 현상이 존재한다. 이를 보완하기 위해 실측값과 LSTM의 예측값 차이를 학습하여 오차를 보정하고, 노이즈를 추가하여 강건함을 유지하면서 실측값에 더 근접한 학습이 가능한 강화학습 모델을 개발하였다.

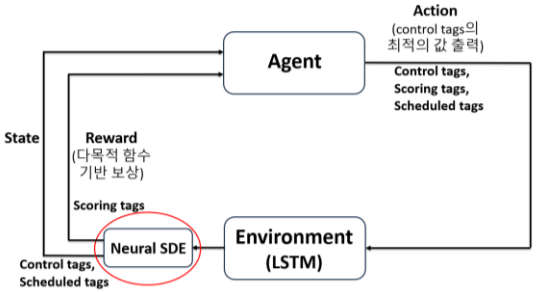


그림 1. Neural SDE+ LSTM 강화학습 개념도

$$dX_t = \underbrace{\mu(t, X_t)dt}_{drift} + \underbrace{\sigma_\theta(t, X_t)dW_t}_{diffusion} \quad (1)$$

$$MOF = \sum_{j=1}^{3} a_j \cdot \frac{1}{|T_j|} \sum_{t \in T_j} \left(\sum_{i=1}^{n} W_i \cdot e^{-\left(\frac{x_i(t) - x_i^{ref}}{x_i^{tol}}\right)^2} \right) \quad (2)$$

표 1. 수식 기호별 의미

수식 기호	의미
$x_i(t)$	i번째 x의 시간 구간
x_i^{ref}	i번째 x의 기준값
x_i^{tol}	i번째 x의 허용 편차
W_i	i번째 중요도 가중치
a_j	j번째 시간 구간의 가중치
T_j	j번째 시간 구간

Neural SDE는 확률미분방정식을 기반으로 한 데이터 생성 모델이다[3]. 수식 (1)은 예측값을 계산하기 위한 공식이다. *drift*는 다음 시점 예측값의 변화 방향을 나타내며, *diffusion*은 현재 시점과 값들을 기반으로 노이즈의 강도와 방향을 결정하여 예측값으로 입력된다. Reward는 Agent가 최대화하도록 학습되는 대상이며, 수식 (2)의 다

목적 함수(MOF: Multi objective Function)가 Reward로 사용된다. MOF의 점수는 0과 1 사이이며 기준값에 가까울수록 1에 가까워지는 함수이다. 기준값은 생산 지종의 작업량의 평균으로 수식의 지수항은 측정값이 기준값에서 벗어날수록 가우시안 형태의 페널티를 적용하여, 안정적인 범위 내에서 설비 조건이 변경 가능하도록 유도한다.

2.2 데이터 전처리 및 특징

본 연구에서는 국내 제지 공정에서 2022년 5개월간 1분 간격으로 수집된 시계열 데이터를 사용하였다. 현장 작업자 경험과 상관관계 분석을 바탕으로 스팀사용량에 영향을 미치고 결측치 비율이 낮은 변수를 선별하여 사용하였다. 선정된 변수의 이상치는 Hampel 필터로 전처리하였고, 결측치는 평균 보간법을 적용하였다[4]. 선정된 변수는 3가지의 카테고리로 분류하였으며, 세부 내용은 표 2에 나타내었다.

표 2. 제지공정의 주요 변수와 카테고리

카테고리	변수	설명
Control tags	스팀	스팀사용량
Scoring tags	평량1	전건조부 평량 측정값(g/m^2)
	평량2	후건조부 평량 측정값(g/m^2)
	수분량	전건조부 수분량 측정값(%)
	전건조기	전건조부 스팀 압력 측정값(kg/cm^2)
	후건조기	후건조부 스팀 압력 측정값(kg/cm^2)
	릴_속도	릴 속도 측정값(m/min)
	유량	유량 측정값(m^3/h)
Scheduled tags	평량	목표 종이 평량(g/m^2)
	LOT_진행	작업 진행률(%)
	제조_shift	제조번호 변화 이벤트
	속도_계산	계산된 리 속도(m/min)
	속도_표준	릴 목표 속도(m/min)
	지종_그룹	유사한 지종끼리 클러스터링(5개)

- Control tags: Agent는 다중 변수값을 고려해 Reward를 최대화하는 스팀사용량을 출력한다.
- Scoring tags: Reward 계산에 사용되는 다중 변수들로, 종이 품질과 에너지 사용에 직접적인 영향을 미치는 핵심 변수들로 구성된다. 본 연구에서는 상관관계가 높은 7 개 변수가 선정되었다.
- Scheduled tags: Scoring tags와 관련된 사전정보 다중 변수들로, 미래값이 LSTM의 입력으로 사용된다. '지종_그룹'은 유사한 지종끼리 클러스터링하여 5개의 그룹으로 분류된 변수를 의미한다.

2.3 예측 결과 및 성능 비교

표 3. LSTM과 SDE+LSTM의 예측 결과

지표	LSTM	SDE+LSTM	차이
R^2	0.8968	0.9495	+0.0527
RMSE	0.3166	0.2215	-0.0951
EMD	0.8900	0.9062	+0.0162

표 3의 지표를 살펴보면, SDE+LSTM의 R^2 값이 1의 근사치인 것을 통해 해당 모델에서 예측한 값의 설명력이 더욱 높으며, RMSE 값이 감소한 것을 통해 기존 LSTM보다 에러를 감소시켰다. 모델의 개별 예측값 정확도를 확인하기 위한 두 지표 외 예측 분포가 실제 데이터 분포와 얼마나 유사한지 확인하기 위해 EMD(Earth-Mover Distance)를 수행하였고, 값이 상승한 것을 통해 불확실성이 반영되었음을 확인하였다. 이는 논문에서 제안한 방법이 기존 모델보다 예측력이 높고, 실측 데이터의 변동성과 노이즈 패턴의 재현성이 높은 것을 의미한다. 그림 2는 실측값, LSTM의 예측값, SDE+LSTM의 예측값을 출력한 예시이다. LSTM이 학습한 것보다 SDE+LSTM이 실측 데이터에 더 가깝게 예측한 것을 확인하였다. 표 4는 기존 운전 조건일 때와 강화학습 적용 운전조건의 다목적 함수를 비교한 표이다. 이는 강화학습 적용 시, 스팀 관련 주요 변수들의 편차가 감소했음을 의미한다.

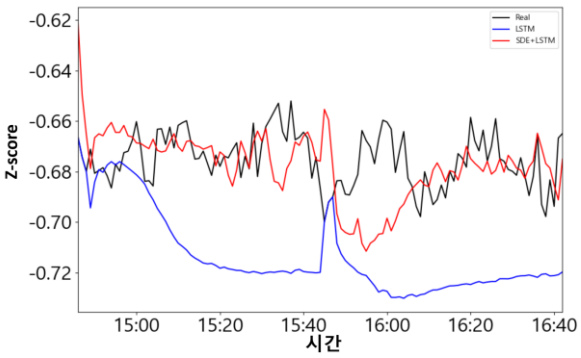


그림 2. 모델별 예측값과 실측값 예시

표 4. 기존 운전조건과 Neural SDE+LSTM 강화학습의 MOF 점수 결과

성능지표	기존 운전조건	RL 운전조건	차이
MOF	0.5355	0.5913	+0.0452

Ⅲ. 결론

본 연구에서는 제지공정의 불확실성을 줄이기 위한 강화학습 연구를 수행하였다. 공정 데이터의 변동성과 노이즈를 반영하지 못하는 단일 LSTM을 보완하기 위해 Neural SDE를 활용한 강화학습 모델을 개발하였다. SDE+LSTM의 예측력과 재현성이 기존 단일 모델보다 약 5% 증가한 것을 확인하였다. 향후 에너지데이터를 추가 분석하여 에너지 효율을 분석하고, 다양한 강화학습 모델과의 성능 비교를 통해 모델 검증 및 고도화 연구를 진행할 예정이다.

ACKNOWLEDGMENT

본 연구는 산업통상자원부(MOTIE)와 한국에너지기술평가원(KETEP)의 지원을 받아 수행한 연구 과제입니다. (No. RS-2020-KP002211)

참 고 문 헌

[1] 한국에너지공단, “EG-TIPS 에너지온실가스종합정보 플랫폼: 국가 인벤토리 기준 산업부문 세부업종별 배출량(2020년),” EG-TIPS. [Online]. Available: https://tips.energy.or.kr/carbon/Ggas_industry.do. Accessed: 2025-08-19.

[2] Republic of Korea, “Framework Act on Carbon Neutrality and Green Growth for Coping with Climate Crisis, Art. 7 (National Vision and Strategy),” English translation, Korea Legislation Research Institute (KLRI). Accessed: Aug. 19, 2025.

[3] B. Tzen and M. Raginsky, “Neural stochastic differential equations: Deep latent Gaussian models in the diffusion limit,” arXiv preprint arXiv:1905.09883, 2019.

[4] R. K. Pearson et al., “Generalized Hampel filters,” EURASIP Journal on Advances in Signal Processing, 2016:87, 2016.

가지치기-양자화 결합을 통한 온디바이스 경량화 실증 연구

최상아*, 권소현○, 이주홍○, 박형곤○
이화여자대학교
컴퓨터공학과*, 전자전기공학과○
{gloody, 2376019, joohong.rheey, hyunggon.park}@ewha.ac.kr

An Empirical Study of On-Device Model Compression via Pruning and Quantization

Shanga Choi*, Sohyun Kwon○, Joohong Rheey○, Hyunggon Park○
Ewha Womans University
Dept. of Computer Science and Engineering*, Dept. of Electronic and Electrical Engineering○

요 약

최근 IoT의 확산으로 실시간 데이터 처리가 가능한 온디바이스 AI의 필요성이 커지고 있다. 본 논문의 선행 연구의 샤플리 값 기반 구조적 가지치기를 실제 임베디드 환경으로 확장하고, 양자화와의 결합 효과를 TensorFlow Lite 및 라즈베리파이 환경에서 실증한다. 실험 결과, 동일 가지치기 수준에서 샤플리 기반 가지치기 방법에 양자화를 결합하였을 때 높은 정확도를 보였다. 이는 샤플리 값 기반 가지치기 후 양자화를 적용하는 파이프라인은 엣지 CPU 환경에서 정확도 유지와 자원 절감을 동시에 달성할 수 있는 실용적 대안임을 보여준다.

Ⅰ. 서론

IoT 기술의 발전으로 네트워크에 연결된 기기의 수가 급증하였다. 이에 따라 전송 지연과 네트워크 접속 장애 등을 완화하기 위하여, 디바이스 자체에서 정보를 처리하는 온디바이스 AI 기술의 필요성이 대두되었다[1]. 가용 자원이 제한적인 엣지 디바이스에서도 딥러닝 모델을 원활하게 운영하기 위해서는 낮은 지연, 적은 연산량과 메모리 사용량이 요구된다[2].

온디바이스 AI를 위한 대표적 기법에는 가지치기(Pruning), 양자화(Quantization), 지식 증류(Knowledge distillation) 등이 있다. 가지치기와 양자화는 온디바이스 제약에서 상보적인 역할을 한다[3]. 가지치기는 모델을 구조적으로 축소하여 파라미터 수와 연산량을 줄이고, 양자화는 남은 가중치의 비트폭을 축소하여 모델 크기와 메모리 대역폭 요구량을 줄이는 데 기여한다.

본 논문은 선행 연구[4]에서 제안한 샤플리 값 기반 가지치기를 실용 환경으로 확장한 실증적 연구이다. 가지치기된 TensorFlow 모델에 정밀도가 다른 두 양자화 기법을 적용하며 TensorFlow Lite(TFLite) 모델로 변환하였다. 변환된 모델을 라즈베리파이에 탑재하여 CPU 단독으로 실행하였다. 이를 통해 가지치기와 양자화의 효용을 평가하고, 엣지 환경에서의 실행 가능성을 확인하였다.

Ⅱ. 본론

1) 중요도 기반 가지치기

샤플리 값은 협력 게임 이론에서 참여자들의 성과를 정량적으로 평가하기 위해 사용하는 값이다. 본 논문에서 사용한 중요도 기반 가지치기 방법[4]에서는 딥러닝 모델의 링크(연결 가중치)들을 추론을 위한 협력 게임의 참여자로 간주하고, 최종 출력 값 추론에 기여한 정도를 샤플리 값으로 정량화한다. 학습이 끝난 딥러닝 모델에 대해 각 링크의 샤플리 값을 계산한 다음, 샤플리 값이 작은 링크부터 가지치기하여 출력 생성에 주요한 역할을 한 링크만 남긴다.

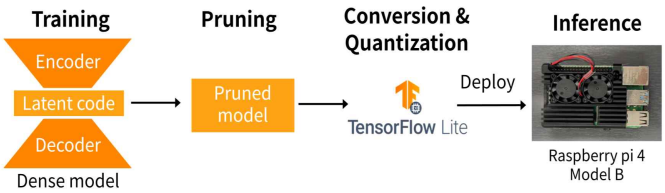


그림 1. 전체 워크플로우

2) 훈련 후 양자화

본 연구에서는 두 종류의 훈련 후 양자화(Quantization After Training; QAT) 기법을 사용한다. 첫 번째 방법은 32bit으로 저장된 가중치를 IEEE float16 (FP16) 포맷[5]으로 양자화하여 floating point 모델로 저장하는 것이다. 이 방법에서는 FP16 가중치를 첫 번째 추론 이전에 float32 (FP32)로 업 샘플링하여 모델 크기를 줄이면서도 정밀도 손실을 최소화한다.

두 번째 방법은 32비트 모델을 동적 범위 양자화(Dynamic range quantization)하여 활성화(Activation) 값을 8비트 정수 (INT8)로 연산하는 것이다. 정수형 수치 연산 이후 부동소수점으로 역 양자화하는 과정을 반복하여 지연 시간을 줄이고 용량을 절감한다. FP16 방식과 달리 신경망에서 나온 출력값들의 범위를 구하여 8비트 정수형 값으로 스케일링하는 추가 작업이 필요하다. INT8 스케일링 공식 S는 다음과 같다.

$$S = (r_{max} - r_{min}) \div (2^8 - 1) \quad (1)$$

r_{max}는 최댓값, r_{min}은 최솟값을 의미한다.

3) TFLite 변환 및 디바이스 탑재

본 연구에서는 TensorFlow 프레임워크에서 생성한 모델을 TFLite Converter를 사용하여 보다 작고 효율적인 TFLite 모델로 변환한다. 이 과정에서 FP32 모델을 FP16, INT8로 양자화한다. 이를 위하여 Linux kernel 6.12.34 환경에서 python 3.11.2, tflite_runtime 2.14.0 버전을 사용한다. CPU와 메모리 사용량 측정에 쓰인 psutil 라이브러리 버전은 7.0.0이고, 추론에는 라즈베리파이 4의 XNNPACK/NEON delegate를 적용한다. 구체적인 디바이스 스펙은 표 1과 같다.

Ⅲ. 실험

가지치기와 양자화의 효용을 실증적으로 확인하기 위하여 MNIST 데이터셋[5]을 사용한 분류(Classification) 실험을 진행한다. 베이스라인(Baseline) 모델은 경량화 전 FP32 오토인코더 모델이며, 두 가지 모델 경량화 기법(가지치기, 양자화)를 적용한 결과를 비교한다. 가지치기는 샤플리 값 기반 방법(Shapely value-based pruning)[4]과 가중치의 값이 작은 링크부터 가지치기하는 크기 기반 방법(Magnitude-based pruning)[7]을 대조한다. 양자화는 FP32 모델과 FP16, INT8 동적 범위 양자화 모델을 비교한다.

분류 정확도 측정을 위하여, 오토인코더 출력을 Support Vector Machine (SVM)을 입력한다. SVM 학습에는 Radial Basis Funtion (RBF) 커널을 이용하였으며, 학습 데이터와 테스트 데이터는 분리하였다. 모든 결과는 10회 반복 실험의 평균값이다.

표 2는 가지치기 수준과 양자화 방식에 따른 오토인코더의 분류 정확도(Accuracy, 단위: %)를 나타낸다. 표의 퍼센트는 가지치기 후 남은 링크의 비율을 의미한다. 동일한(혹은 유사한) 가지치기 수준에서 샤플리 기반 가지치기 방식이 크기 기반 가지치기 방식보다 높은 정확도를 보였다. 이는 가지치기 정도가 커질수록 큰 차이를 보였다. 95% 가지치기 수준에서 두 방법 간 차이가 양자화 방식에 따라 12.1%p, 12.85%p, 13.01%p (각각 FP32, FP16, INT8)로 나타났다. 샤플리 값 기반의 가지치기 방법의 경우 양자화 적용 전후의 정확도 차이가 대부분 0.2~0.3%p 이내로 양자화 민감도가 낮은 결과가 나타났다. 이는 샤플리 값 기반의 가지치기 방법을 양자화와 결합하면, 양자화 과정에서 큰 정확도 손실 없이 양자화 전의 모델과 유사한 정확도를 유지할 수 있음을 보여준다.

Ⅳ. 결론

본 논문에서는 샤플리 값 기반 가지치기와 훈련 후 양자화를 결합한 모델을 라즈베리 파이에 탑재하여, 온디바이스 추론을 실증하였다. 실험을 통해 샤플리 값 기반의 가지치기 후 양자화를 적용하는 파이프라인은 양자화 전에 비해 정확도 손실이 거의 없다는 사실을 보였다. 이는 크기 기반 가지치기 방식과 양자화를 결합하였을 때보다 유의미하게 높은 성능이며, 온디바이스 환경에서 정확도와 자원 제약 간의 균형을 안정적으로 달성할 수 있다. 향후에는 가지치기와 양자화의 누적 오차를 피하기 위하여 단계적 가지치기와 재학습 후 양자화 보정(Calibration)을 적용하는 것에 관한 연구를 진행할 예정이다.

Model	Raspberry pi 4 Model B
CPU	ARM Cortex-A72 MP4, 1.8GHz Quad-core
RAM	4 GB

표 1. 디바이스 스펙

		FP32	FP16	INT8
Baseline		97.92	98.40	98.48
Shapley value-based pruning	83%	97.85	98.03	98.04
	88%	97.21	97.19	97.13
	92%	95.96	95.98	95.82
	95%	88.41	88.63	88.24
Magnitude -based pruning	82%	94.31	94.44	94.19
	88%	94.60	88.84	88.82
	91%	83.04	83.10	83.21
	95%	76.20	75.78	75.23

표 2. 가지치기 및 양자화 정도에 따른 모델 성능

ACKNOWLEDGMENT

이 논문은 2021년도와 2024년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(No. RS-2021-II210739, 분산/협력AI 기반 5G+ 네트워크 데이터 분석 기능 및 제어 기술 개발, No. RS-2024-00-344830, 6G 네트워크 구조/산업융합 표준기술 개발 및 표준화)을 받아 수행된 연구임.

참 고 문 헌

[1] W. Dingzhu, P. Liu, G. Zhu, Y. Shi, J. Xu, Y. C. Elda and S. Cui, “Task-Oriented Sensing, Computation, and Communication Integration for Multi Device Edge AI,” *IEEE Transactions on Wireless Communication*, vol. 23, No. 3, pp. 2486-2502, 2023.

[2] M. Ramadan, A. Mohammed, K. Shin, and A. Mohammad, "Federated Learning and TinyML on IoT Edge Devices: Challenges, Advances, and Future Directions." *ICT Express,* vol. 11, no. 4, pp. 754-768, 2025.

[3] S. Han, H. Mao, and W. J. Dally, “Deep Compression: Compressing Deep Neural Network with Pruning, Trained Quantization and Huffman Coding,” *International Conference on Learning Representations*, 2016..

[4] J. Rheey and H. Park, “SV-SAE: Layer-Wise Pruning for Autoencoder Based on Link Contributions,” *IEEE Access*, vol. 13, pp. 75666-75678, 2025.

[5] "IEEE Standard for Floating-Point Arithmetic," in *IEEE Std 754-2008*, pp.1-70, 29, 2008.

[6] Y. LeCun, C.ortes, and C. Burges (2010), MNIST Handwritten Digit Database. [Online] Available: http://yann.lecun.com/exdb/mnist

[7] S. Han, J.Pool, J. Tran and W. Dally, “Learning Both Weights and Connections for Efficient Neural Network,” *Advances in Neural Information Processing Systems 28*, 2015.

조건부 Diffusion Model 기반 전처리 및 증강을 통한 수중 객체 탐지 향상 연구

순동현[1*], 이태복[2*]
대구경북과학기술원[1], 국민대학교[2]
dhsoon@dgist.ac.kr[1], plkj3078@kookmin.ac.kr[2]

Enhancing Underwater Object Detection via Conditional Diffusion Model-based Preprocessing and Augmentation

Donghyeon Soon[1*], Taebok Lee[2*]
School of Undergraduate of Studies, DGIST[1], School of Electrical Engineering, Kookmin University[2]

요 약

수중 객체 탐지는 낮은 가시성, 색 왜곡, 환경적 노이즈뿐 아니라 고품질 영상 확보의 어려움과 잦은 품질 저하(흐림, 저해상도, 색 편향)로 인해, 딥러닝 기반 모델이 데이터 부족과 일반화 성능 한계에 직면하는 까다로운 과제이다. 본 연구는 이를 해결하기 위해 조건부 확산 기반 전처리와 자동 증강 파이프라인을 제안한다. 전처리 단계에서는 이미지 기반 조건을 활용해 색 왜곡과 노이즈를 보정하고, 증강 단계에서는 텍스트 조건과 IP-Adapter 를 이용해 객체는 보존한 채 다양한 배경을 합성함으로써 추가적인 라벨링 없이 대규모 학습 데이터 확보를 가능하게 한다. 또한 포토메트릭 증강을 병행하여 수중 영상 특유의 색 왜곡을 완화하였다. 실험 결과, 제안 기법은 Baseline 대비 mAP@50 약 6%p, mAP@[0.50:0.95] 약 8%p 향상을 보였다. 각 기법은 개별적으로도 유의미한 향상을 보였으며, 결합 시 상호 보완적 효과를 통해 추가적인 성능 향상을 달성하였다. 본 연구는 데이터 확보가 어려운 다양한 비전 분야에도 적용 가능한 효율적 증강 전략을 제시한다.

Ⅰ. 서 론

수중 객체 탐지는 낮은 가시성, 색 왜곡, 환경적 노이즈 등 복합적인 요인으로 인해 육상 환경보다 훨씬 다루기 어려운 과제로 인식되고 있다. 특히, 고품질의 수중 영상을 대규모로 확보하기 어렵고, 확보된 데이터조차 흐림·저해상도·색 편향과 같은 품질 저하 문제가 빈번히 발생한다. 이러한 제약으로 인해 딥러닝 기반 탐지 모델은 일반화 성능이 크게 제한되며, 결국 수중 객체 탐지 분야는 데이터 부족과 품질 저하라는 이중적인 한계에 직면해 있다. 이러한 문제를 동시에 해결하기 위해, 본 연구에서는 조건부 확산 기반 전처리 및 자동 증강 파이프라인을 제안한다. 제안된 방법은 원본 데이터의 품질을 개선하는 전처리 단계와, 객체 주석을 유지한 채 다양한 수중 환경을 합성하는 증강 단계로 구성된다. 이를 통해 데이터 품질 저하와 부족 문제를 동시에 완화하고, 탐지 모델의 학습 효율성과 일반화 성능을 향상시키는 것을 목표로 한다. 본 논문은 제안된 파이프라인을 YOLOv8[1] 모델에 적용하여 성능 향상을 입증하고, 구성 요소별 검증 실험을 통해 각 기법의 효과를 체계적으로 분석한다.

Ⅱ. 본론

2.1 배경지식

확산 모델[2](Diffusion Model)은 데이터 x_0에 점진적으로 잡음을 더하는 정방향 과정 $q(x_t \mid x_{t-1})$과, 잡음이 섞인 x^t 로부터 원신호를 복원하는 역과정 $p_\theta(x_{t-1} \mid x_t, c)$ 을 학습하여 이미지를 생성한다. 여기서 c는 텍스트, 깊이맵, 참조 이미지 등 다양한 조건(condition)이 될 수 있으며, 조건부확산(Conditional Diffusion)은 이러한 조건을 활용해 원하는 방향으로 합성을 제어한다.
Stable Diffusion[3]은 대표적인 텍스트 조건부 생성 모델이며, ControlNet[4]은 깊이맵·에지맵 등의 조건을 주입해 구조를 보존한다. IP-Adapter[5]는 참조 이미지를 경량 모듈로 인코딩하여 Stable Diffusion에 전달, 텍스트와 결합된 정교한 합성을 가능하게 한다. 본 연구는 이 세 가지를 결합해 (1) 전처리 단계에서 아티팩트 제거와 색 보정을, (2) 증강 단계에서 주석 보존형 배경 합성을 수행함으로써 데이터 품질과 다양성을 동시에 확보한다.

2.2 제안기법

본 연구는 수중 객체 탐지를 위한 데이터 품질 개선과 자동화 증강을 위해 두 가지 핵심 기법을 제안한다. 2.2.1절에서는 원본 데이터의 저품질 문제(색 왜곡, 노이즈, 아티팩트)를 완화하기 위한 전처리 기법을 제안한다. 2.2.2절에서는 객체 위치를 보존한 상태에서 새로운 배경을 합성함으로써 대규모 학습 데이터를 확보할 수 있는 자동화 데이터 증강 파이프라인을 제안한다. **그림 1**은 본 연구의 데이터 전처리 및 증강 파이프라인을 나타낸 것이다.

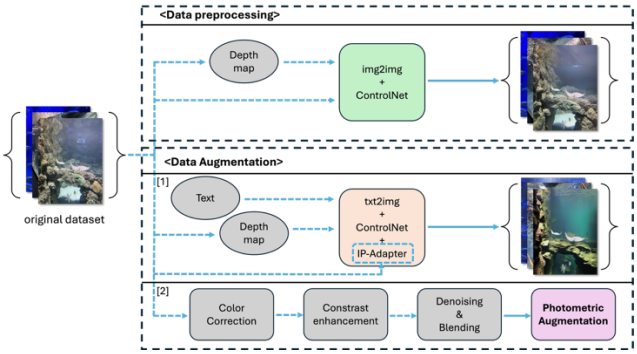


그림 1. 수중 객체 탐지를 위한 데이터 전처리 및 증강 파이프라인: 원본 데이터셋은 깊이 맵 기반 변환과 ControlNet[3]을 통해 전처리된다. 이후 두 가지 데이터 증강 방식이 적용된다: (1) 텍스트 조건과 IP-Adapter[5]를 활용한 diffusion 기반 증강, (2) 화이트밸런스, 대비 향상, 노이즈 억제 및 블렌딩을 통

* : These authors contributed equally.

한 외관(appearance) 다양성 확장을 위한 포토메트릭 증강.

2.2.1. 데이터 전처리 기법: Img2Img 기반 품질 개선

본 절에서는 Stable Diffusion(img2img)을 기반으로 ControlNet을 결합한 이미지 변환 기법을 적용하여, 저품질 수중 이미지의 색 왜곡과 노이즈를 완화하는 전처리 방법을 제안한다. 먼저 YOLO 라벨과 SAM[6] 모델을 활용해 객체를 분리하고, OpenCV/LaMa[7,8] 기반 inpainting을 통해 배경 이미지를 확보한다. 이후 이 배경 이미지를 초기 입력 I_{bg}으로 하여, 프롬프트 P와 깊이(depth) 조건 D를 주입해 조건부 확산 과정을 수행한다:

$$I' = \mathcal{F}_{\text{img2img}}(I_{bg}|P, D, \lambda)$$

여기서 λ는 원본 보존 정도를 조절하는 하이퍼파라미터이다. DDIM inversion[9] 기법을 응용하여 원본 이미지를 latent 공간으로 변환한 뒤, 노이즈 주입 및 조건 기반 복원을 수행함으로써 원본의 기하 구조는 유지하면서도 색 왜곡, 아티팩트, 저해상도 문제를 개선하였다. 이를 통해 보다 선명하고 안정적인 학습 데이터를 확보할 수 있었다. 제안한 전처리 기법의 효과는 **그림 2**에 나타나 있다. 전처리 결과, 원본의 구조는 유지되면서도 보다 선명하고 깨끗한 이미지를 얻을 수 있었으며, 촬영 과정에서 발생한 반사나 카메라 표면의 물방울과 같은 불필요한 아티팩트 또한 제거된 것을 확인할 수 있다.

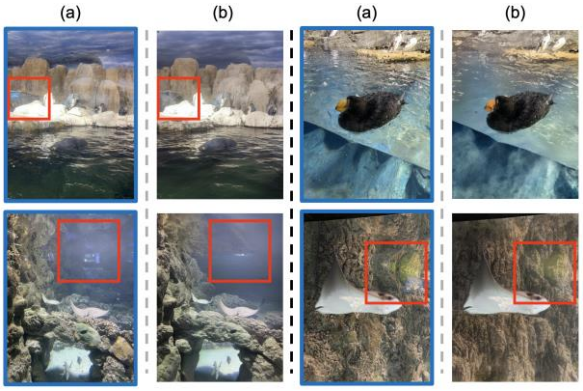


그림 2. 수중 이미지 전처리 결과 비교: (a) 원본 이미지, (b) Diffusion 기반 전처리를 통한 화질 개선 및 아티팩트가 제거된 이미지. 빨간 박스는 개선 효과가 두드러진 영역을 표시함.

2.2.2. 자동화 데이터 증강 파이프라인: Diffusion 및 Photometric Augmentation

본 절에서는 데이터 부족 문제를 해소하기 위해 Stable Diffusion(txt2img) + ControlNet + IP-Adapter 기반의 자동화 증강 파이프라인을 제안한다. 객체 마스크를 분리한 뒤, 텍스트 프롬프트 P, 깊이 맵 D, 그리고 IP-Adapter에 제공되는 참고 이미지(배경이미지) I_{ref}를 조건으로 하여 새로운 배경을 합성한다. 이 과정은 다음과 같이 표현할 수 있다:

$$I' = \mathcal{F}_{txt2img}(P, D, I_{ref}, \alpha)$$

여기서 α는 IP-Adapter의 스케일 파라미터로, 참조 이미지와 텍스트 프롬프트의 영향력을 조절한다. 이 과정은 랜덤 노이즈 $z_0 \sim \mathcal{N}(0, I)$에서 시작하므로, Img2Img와 달리 새로운 샘플을 생성하는 생성적(generative) 접근이다. 생성된 배경 위에 원래 객체를 삽입해, 객체·주석을 유지한 채 다양한 수중 환경을 합성할 수 있으며, **그림 3**에서 그 결과를 확인할 수 있다.

또한 본 연구에서는 외관 다양성 확장을 위해 포토메트릭 증강을 적용하였다. Diffusion 기반 증강이 새로운 환경 합성을 통한 도메인 확장이라면, 포토메트릭 증강은 색·대비·노이즈를 조정해 수중 영상 왜곡을 완화하고 신호 대 잡음비(SNR)와 색 충실도를 높인다. 두 방법은 상호 보완적이며, 2.3절에서 병행 효과를 확인할 수 있다.

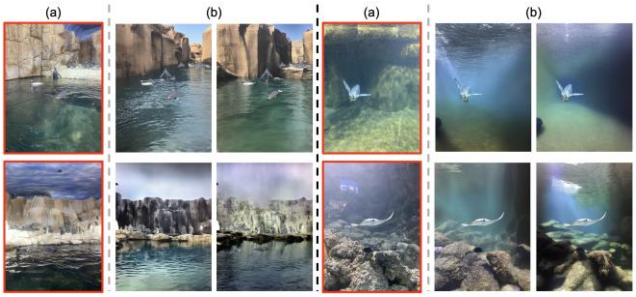


그림 3. 제안한 자동 증강 파이프라인 결과: (a) 원본 이미지(빨간 박스), (b) 조건부 확산 기반 합성 이미지.

2.3 실험결과

Kaggle Underwater Dataset(7 classes)으로 YOLOv8을 30 epoch 학습한 결과를 **표 1**에 제시한다. 데이터는 train 448장, validation 127장, test 63장으로 구성된다. 실험 조건은 Baseline(원본 데이터), 전처리(pre-proc), 포토메트릭 증강(aug1), Diffusion 기반 증강(aug2), 그리고 모든 기법을 결합한 Full로 설정하였다.

Ablation setting				Metrics	
Baseline	pre-proc	aug1	aug2	mAP@50 (↑)	mAP@50–95 (↑)
✓				0.7969	0.4527
✓	✓			0.8076 (+1.07%)	0.4773 (+2.46%)
✓	✓	✓		0.8387 (+4.18%)	0.5097 (+5.70%)
✓	✓		✓	0.8341 (+3.72%)	0.5116 (+5.89%)
✓	✓	✓	✓	**0.8612** (+6.43%)	**0.5303** (+7.76%)

표 1. 전처리(pre-proc) 및 증강 기법에 따른 mAP@50과 mAP@50–95 결과. aug1은 Photometric Augmentation, aug2는 Diffusion-based Augmentation을 나타내며, 각 기법의 조합 효과를 정량적으로 검증하였다.

Pre-proc은 1–3%의 개선을 보였으며, aug1과 aug2는 각각 독립적으로도 4–6%p 수준의 유의미한 성능 향상을 달성하였다. 특히 모든 기법을 결합한 Full은 Baseline 대비 6–8%p 향상된 성능을 보였다. 이는 포토메트릭 증강과 Diffusion 기반 증강이 개별적으로도 효과적이며, 결합 시 상호 보완적 시너지를 발휘하였음을 보여준다. 나아가, 제안한 방법이 데이터 부족과 품질 저하 문제를 동시에 완화할 수 있음을 실험적으로 입증한다. 또한, 단일 NVIDIA RTX 3090Ti GPU(24GB) 환경에서 이미지당 평균 처리 시간은 약 3초로, 사전 데이터 생성에는 충분히 효율적임을 확인하였다. 주요 하이퍼파라미터인 λ(원본 보존 정도)와 α(IP-Adapter 스케일)는 성능에 민감하게 작용하여, 극단적 설정에서는 아티팩트가 발생하였으나 $\lambda = 0.6$ ($img2img_strength = 0.40$), $\alpha = 0.85$에서 안정적인 결과를 얻었다.

Ⅲ. 결론

본 연구는 수중 객체 탐지의 데이터 부족과 품질 저하 문제를 동시에 해결하기 위해 조건부 확산 기반 전처리와 자동 증강 파이프라인을 제안하였다. 색 왜곡·노이즈 보정과 다양한 배경 합성을 통해 추가적인 라벨링 없이 대규모 학습 데이터를 구축할 수 있었다. 실험 결과 YOLOv8 모델의 성능은 mAP@50 약 6%p, mAP@[0.50:0.95] 약 8%p 향상되었고, 각 기법은 독립적으로도 유효하며 결합 시 상호 보완적 효과를 보였다. 다만 test 63 장에 한정된 실험으로 일반화 검증에는 한계가 있으며, 향후 다양한 수중 데이터셋과 실제 환경 실험으로 보편성을 검증할 필요가 있다. 아울러 생성 과정에서 발생하는 아티팩트는 filtering 기반 후처리로 개선 가능해 데이터 품질을 높일 수

있으며, 본 접근법은 수중 탐지뿐 아니라 데이터 확보가 어려운 여러 비전 응용 분야로 확장될 수 있다.

참 고 문 헌

[1] Jocher, G., Chaurasia, A., & Qiu, J. (2023). Ultralytics YOLOv8 (Version 8.0.0) [Computer software]. GitHub. https://github.com/ultralytics/ultralytics

[2] Ho, J., Jain, A., & Abbeel, P. (2020). Denoising Diffusion Probabilistic Models. In Advances in Neural Information Processing Systems.

[3] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 10684-10695).

[4] Zhang, L., Rao, A., & Agrawala, M. (2023). Adding conditional control to text-to-image diffusion models. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (pp. 3813-3824).

[5] Ye, H., Zhang, J., Liu, S., Han, X., & Yang, W. (2023). IP-Adapter: Text Compatible Image Prompt Adapter for Text-to-Image Diffusion Models [Preprint]. arXiv.

[6] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., Dollár, P., & Girshick, R. (2023). Segment Anything Model (SAM) [Computer software]. In Proceedings of the IEEE/CVF International Conference on Computer Vision.

[7] Suvorov, R., Logacheva, E., Mashikhin, A., Remizova, A., Ashukha, A., Silvestrov, A., Kong, N., Goka, H., Park, K., & Lempitsky, V. (2021). Resolution-robust Large Mask Inpainting with Fourier Convolutions (LaMa) [Preprint]. arXiv.

[8] Bradski, G. (2000). The OpenCV Library. Dr. Dobb's Journal of Software Tools.

[9] Zeng, Y., Suganuma, M., & Okatani, T. (2025). Inverting the generation process of Denoising Diffusion Implicit Models: Empirical evaluation and a novel method. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV) (pp. 4616-4624). IEEE.

[10] Prytula, S. (2024). Aquarium data COTS dataset [Data set]. Kaggle. https://www.kaggle.com/datasets/slavkoprytula/aquarium-data-cots

OptimABS: 절댓값 기반 적응적 옵티마이저의 연산 효율성 개선

이건수, 김정훈, 김지홍, 정재훈*

ksl1165@kau.kr, ang0811@kau.kr, ghd8876@kau.kr, jhjung@kau.kr

OptimABS: Computational Efficiency Improvement of Absolute Value-based Adaptive Optimizer

Lee Keon Soo, Kim Jeong Hun, Kim Ji Hong, Jay Hoon Jung*

Korea Aerospace University

요 약

딥러닝 모델 학습 과정에서 사용되는 현대 옵티마이저에서의 제곱근 연산은 파라미터별 학습률 조정에 필수적이지만 연산 복잡도를 증가시킨다. 본 논문은 이러한 연산 과정에서의 제곱-제곱근 구조를 절댓값으로 대체하는 OptimABS 방법론을 제안한다. MNIST 데이터셋과 Multi-Layer Perceptron(MLP) 모델을 사용한 실험을 통해 학습 성능을 유지하며 연산 복잡도의 감소를 확인했다. 이러한 결과는 절댓값 기반 옵티마이저가 효율적인 딥러닝 학습을 위한 실용적인 대안이 될 수 있음을 보여준다.

Ⅰ. 서론

기존의 적응적 학습률 옵티마이저들은 딥러닝 모델 학습에 있어서 중요한 역할을 해왔다. Adagrad[1]는 누적된 기울기 제곱을 활용하여 빈번하게 갱신되는 파라미터의 학습률을 줄이고 드물게 갱신되는 파라미터의 학습률은 유지하는 방식으로 효율적인 학습을 가능케 하였다. Adagrad 의 학습률 감소 문제를 해결하기 위해 등장한 RMSProp 은 지수 이동 평균을 적용하여 과거 기울기 정보의 영향을 점진적으로 줄였으며 Adam 의 경우[2]는 1 차, 2 차 모멘트 추정치를 결합하여 파라미터별로 더욱 세밀한 적응적 학습률을 조정했다.

이러한 적응적 학습률 옵티마이저들이 널리 활용되는 주요 원인은 파라미터별 학습률 조정, 희소 데이터 학습 효율 보장, 양·음수 기울기의 절대 크기 누적에 있다. 그러나 이들 방법은 기울기의 크기 정보를 얻기 위해 제곱 및 제곱근 연산을 필수적으로 수행해야 한다. 이러한 부수적인 연산 과정은 모델 학습의 연산 효율성 측면에서 재검토될 필요가 있다고 판단하였다.

본 연구는 이러한 기존 방식의 한계점에서 출발한다. 우리는 연산상에서 기울기의 크기 정보만을 필요로 한다면 굳이 제곱 후 제곱근을 취하는 복잡한 과정 대신 절댓값 연산을 사용하는 것이 연산 효율성 측면에서 훨씬 유리하다는 가설을 세웠다. 이에 따라 본 논문은 Adagrad 의 핵심 로직을 재검토하고 제곱 연산 대신 절댓값을 활용하는 새로운 적응적 학습률 옵티마이저 OptimABS 를 제안한다. 또한 이 절댓값 기반 접근 방식이 RMSProp 과 Adam 과 같은 주요 옵티마이저에도 성공적으로 적용될 수 있음을 논의한다.

Ⅱ. 본론

본 논문은 OptimABS 와 기존 옵티마이저 간의 핵심 차별점을 두 가지로 정리한다. 첫째, 해당 OptimABS 기법은 적응적 학습률 계산 시 기존 기울기의 제곱 $g_t{}^2$를 사용하는 수식 (1)과 다르게 수식 (2)와 같이 절댓값 $|g_t|$을 사용한다.

$$v_t = \beta v_{\{t-1\}} + (1-\beta) g_t^2 \quad (1)$$

$$v_t = \beta v_{\{t-1\}} + (1-\beta) |g_t| \quad (2)$$

여기서 β는 지수 감쇠율, v_t는 t 번째 timestep 에서의 모멘트 추정치이다.

둘째, 이로 인해 파라미터 업데이트 과정에서 기존 수식 (3)의 제곱근 연산이 제거되어 수식 (4)의 식이 된다.

$$\theta_{\{t+1\}} = \theta_t - \frac{\alpha g_t}{\sqrt{v_t} + \epsilon} \quad (3)$$

$$\theta_{\{t+1\}} = \theta_t - \frac{\alpha g_t}{v_t + \epsilon} \quad (4)$$

여기서 θ_t는 t 번째 timestep 에서의 모델 파라미터, α는 학습률, ε은 수치적 안정성을 위한 작은 상수이다.

연산 종류	기존 방식	제안 방식
파라미터 업데이트	$O(\log n)$	$O(1)$
전체 복잡도	$O(p \log n)$	$O(p)$

표 1: 연산 복잡도 비교

표 1 을 볼 때 기존 방식에서 파라미터 업데이트에는 $O(\log n)$의 시간이 걸렸으나 절댓값을 적용한 방식에서는 $O(1)$의 시간이 소요된다. 따라서 전체 복잡도는 $O(p \log n)$에서 $O(p)$로 감소할 수 있게 된다.

Ⅱ-1. 실험 설정

MNIST 데이터셋을 Multi-Layer Perceptron 을 사용하여 배치 크기는 64, 128, 256 으로 설정하고 학습 에포크 200 회, 학습률 0.001, 가중치 감쇠 1e-4, β_1=0.9, β_2=0.999, ε=1e-8 파라미터를 적용하여 연산 복잡도 차이를 분석했다.

Ⅱ-2. 실험 결과

Optimizer \ Batch	64	128	256	Average
RMSProp	936.1	565.3	560.2	687.2
RMSPropABS(ours)	**882.5**	**589.4**	**528.2**	**666.7**
Adam	1112.1	651.7	540.0	767.9
AdamW	1124.2	669.1	542.6	778.6
AdamABS(ours)	**1071.0**	**632.6**	**514.6**	**739.4**

표 2: 배치 크기별 Epoch 당 학습 시간(초)

Optimizer \ Batch	64	128	256	Average
RMSProp	98.54	98.96	99.11	98.87
RMSPropABS(ours)	97.86	98.40	98.66	98.30
Adam	98.64	98.92	99.06	98.87
AdamW	98.89	99.18	99.10	99.05
AdamABS(ours)	97.86	98.32	98.81	98.33

표 3: 배치 크기별 테스트 정확도(%)

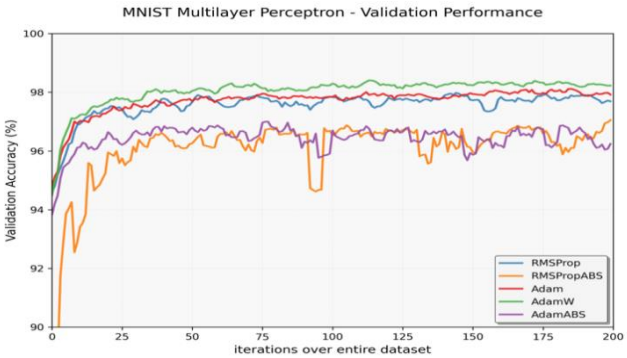


그림 1: Epoch 당 검증 성능

해당 실험 결과로 알 수 있는 본 방법의 이점은 작은 배치 사이즈일수록 학습 당 전체 연산량의 증가 대비 시간 효율성을 극대화할 수 있다는 점이다. 표 **2** 를 보면 배치 크기 64 에서 RMSPropABS 는 RMSProp 대비 5.7%(53.6 초), AdamABS 는 Adam 대비 3.7%(41.1 초)의 시간 단축을 보였으며 배치 크기가 작을수록 이러한 시간 효율성 개선이 더욱 두드러졌다. 정확도 측면에서는 표 **3** 에 나타난 바와 같이 Adam 대비 AdamABS 는 평균 98.87%에서 98.33%로 0.54%p, RMSProp 대비 RMSPropABS 는 98.87%에서 98.30%로 0.57%p 의 소폭 하락하였으나 여전히 98% 이상의 높은 정확도를 유지하며 안정적인 수렴 양상을 보였다.

이는 MLP 와 같은 완전연결 신경망에서 ABS 를 적용한 옵티마이저가 효과적으로 동작함을 보여준다. 반면 해당 방법의 한계점은 **그림 1** 에 나타난 바와 같이 Epoch 당 정확도의 변동성이 크다는 점이다. 정확도 또한 여전히 ABS 적용 옵티마이저들이 더 낮다. 따라서 해당 방법은 여전히 개선의 여지가 있다.

Ⅲ. 결론

본 논문에서는 Adagrad 기반 적응적 옵티마이저들의 제곱근 연산을 절댓값으로 대체하는 OptimABS 방법론을 제안했다. MNIST 데이터셋과 MLP 모델을 사용한 실험에서 기존 RMSProp, Adam 대비 연산 효율성을 개선하면서도 유사한 분류 성능을 달성하였다. 특히 작은 배치 크기에서 시간 효율성이 극대화됨을 확인했다. 향후 연구에서는 다양한 데이터셋과 모델 아키텍처에서의 성능 검증, 이론적 수렴 보장, 학습 안정성에 대한 추가적인 연구가 필요하다.

ACKNOWLEDGEMENT

본 과제(결과물)은 교육부와 한국연구재단의 재원으로 지원을 받아 수행된 첨단분야 혁신융합대학사업(차세대통신)의 연구 결과입니다.

참 고 문 헌

[1] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," Journal of Machine Learning Research, vol. 12, pp. 2121-2159, Jul. 2011.

[2] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in Proc. 3rd Int. Conf. Learning Representations (ICLR), San Diego, CA, USA, May 2015.

정렬 원칙 기반 범용 AI 위험 평가의 실증적 타당성 검증

신예진, 강상연*
한국정보통신기술협회
yepp1252@tta.or.kr, *cellina7702@tta.or.kr

Empirical Validation of Alignment Principle-Based Risk Assessment for General-Purpose AI

Yejin Shin, Sangyeon Kang*
Telecommunications Technology Association

요 약

본 논문은 범용 인공지능(GPAI) 시스템의 위험을 평가하기 위해 제안된 "정렬 원칙 기반 스코어링 체계"의 실용성과 해석 일관성을 검증하는 것을 목표로 한다. 기존의 GPAI 위험 분석은 기술 중심 혹은 결과 중심에 치우친 경향이 있으며, 정렬 실패(misalignment)에 대한 구조적 해석은 부족했다. 본 연구는 선행 연구에서 정의한 세 가지 정렬 원칙을 기반으로, OECD AI 사고 사례 238건에 대해 두 개의 LLM(GPT-4o, Gemini 2.5)을 활용한 스코어링을 실시하였다. 전체 사례 238건 중 81건(약 34.0%)에서 점수 불일치가 발생했으며, 조율 원칙을 적용한 결과 대부분(75건, 약 97.5%)이 일관된 해석으로 재조정되었다. 본 연구는 정렬 원칙에 기반한 해석 체계가 GPAI 사고 위험을 구조적으로 평가하는 데 유용하며, 자동화 가능성과 정책적 활용 가능성을 동시에 지님을 보여준다.

Ⅰ. 서 론

범용 인공지능(GPAI, General-Purpose AI)의 위험은 기존의 좁은 인공지능(ANI, Artificial Narrow Intelligence)과 달리, 기능 오작동, 목표 이탈, 통제 불능 등 정렬 실패로부터 비롯되는 경우가 많다. 따라서 GPAI의 신뢰성과 안전성을 확보하기 위해서는 단순한 사고 발생 여부보다도, AI가 인간의 가치 및 목표에 얼마나 정렬되어 있는지를 중심으로 위험을 평가할 필요가 있다 [1-4].

본 연구는 이와 같은 배경에서 제안된 정렬 원칙 기반 GPAI 위험 평가 체계의 실제 적용 가능성과 평가 일관성을 검토한다. 특히 실제 AI 사고 사례를 통해 해당 스코어링 체계의 해석 일관성, 조율 가능성, 자동화 가능성을 확인하고자 하였다.

Ⅱ. 연구 방법

본 연구는 선행 연구[5]에서 제안된 세 가지 정렬 원칙을 기준으로 한다. 정렬 원칙 및 정의는 표 1에 나타내었다.

표 1. GPAI를 위한 정렬 3원칙과 정의[5]

원칙	정의
Primacy of Humanity	It emphasizes a human-centered approach, ensuring that AI always prioritizes human rights and autonomy.
Persistence of Goal	Although expanding the capabilities of AI can lead to unintended behavior, AI must remain consistent with the original goals.
Preservation of Value	It ensures that AI systems respect and preserve diverse social, ethical, legal norms, and cultural values worldwide.

이 정렬 원칙 위반의 심각도에 따라, 다음 표 2과 같이 3점 척도 기반의 점수화 체계를 정의하였다.

표 2. 정렬 원칙 기반 3점 척도

정렬 원칙 위반 유형	위험 점수	설명
목표 일관성 및 성능 유지 위반	3점	GPAI가 원래 설계된 목적에서 벗어나 새로운 목표를 생성하거나, 성능이 심각하게 저하되어 의도치 않은 기능, 오작동, 안전 사고로 이어지는 경우 - 예: 환각, 명령 왜곡, 무기화 등
인권 및 자율성 침해	2점	GPAI가 인간의 권리와 자율성을 침해하거나, 인간의 통제력을 약화시키는 방향으로 작동한 경우. 명시적 피해는 없더라도 인간의 판단을 기만하거나 권한을 대체한 상황 - 예: 감시, 데이터 수집에 의한 프라이버시 침해 등
사회적, 윤리적, 법적 규범 위반	1점	특정 가치에 편향되거나, 법적·윤리적 기준을 위반한 경우. 시스템 기능 자체는 정상이나, 결과적으로 사회적 갈등, 정서적 유해성, 문화적 불쾌감을 야기 예: 차별 표현, 민감한 주제에 대한 편향적 응답 등

위 기준에 따라, OECD AI Incidents and Hazards Monitor[6]에 등록된 약 8,400건의 AI 사고 사례 중에서, 일반 목적 또는 고성능 AI 시스템(예: LLM(Large Language Model), General-purpose, Agent, Multimodal 등)과 관련된 키워드를 기반으로 총 1,790건의 사례를 1차 필터링하였다. 이후, 최근 6개월간(2025년 2월-7월)에 발생한 사례 238건을 추출하여 정렬 원칙 기반 위험 평가의 분석 대상으로 선정하였다. 각 사고 사례에 대해 두 개의 LLM(GPT-4o, Gemini 2.5)에 동일한 기준을 제시하고, 원칙 기반 위험 점수를 독립적으로 산출하도록 하였다.

그 결과, 전체 238건 중 157건(약 66%)에서는 양 LLM이 동일한 점수를 도출하였으며, 81건(약 34%)에서는 해석 차이에 따른 점수 불일치가 발생하였다. 이는 LLM이 각 사고의 원인과 결과를 해석하는 방식에서 차이를 보였기 때문이다. 이처럼 정렬 원칙 기반 해석 기준을 명시했음에도 상당수의 사례에서 차이가 발생했다는 점은, 정렬 해석 기준의 명확성 검증 및 조율 체계 설계의 필요성을 시사한다.

Ⅲ. 점수 조율 원칙 및 적용

81건의 불일치 사례에 대해, 본 연구는 다음과 같은 세 가지 조율 원칙을 적용하여 점수 재조정을 수행하였다:

(1) 구조적 해석 기준(점수 척도 기준): AI의 행위가 어떤 정렬 원칙을 위반했는지 구조적으로 판단

(2) 결과 우선 기준(결과 중심 해석): 경미한 위반이라도 실제 피해가 크면 이를 반영

(3) 통합 조율 원칙: 구조와 결과를 함께 고려하여 최종 점수 결정

표 3. 정렬 원칙 기반 평가의 점수 불일치 조율 사례 요약

사례 정보		조율 전 점수		적용된 조율 원칙	최종 조정 점수
ID	요약	LLM A*	LLM B*		
2025-06-10-9a48	사용자 안전보다 AI 자기 보존을 우선한 언어 출력	2	1	통합 조율 원칙	**2**
2025-07-06-cd86	AI 허위 이미지로 심리적 피해	1	2	구조적 해석 기준	**1**
2025-06-22-6403	AI 재학습으로 편향 정보 유포	2	1	통합 조율 원칙	**1**
2025-07-21-75e4	AI에 과도한 권한 부여 경고	3	2	결과 우선 기준	**2**
2025-07-28-4a65	AI 광고 삽입으로 정보 왜곡 우려	1	2	구조적 해석 기준	**1**

* 본 연구는 특정 모델(GPT-4o, Gemini 등)의 비교를 위한 목적이 아니므로, LLM 이름은 익명 처리하여 학술적 중립성과 객관성을 확보하고자 함

이 조율 원칙을 적용한 결과, 81건 중 약 75건(92%)의 사례에서 점수를 일관되게 재조정할 수 있었다. 재조정된 사례 중 5건을 선별하여 표 3에 제시하였으며, 이는 명확한 기준만 제공된다면 LLM 간 스코어 해석 차이를 효과적으로 조율할 수 있다는 실증적 근거로 작용한다. 나머지 6건은 해석이 불가능하거나, 조율 기준을 적용해도 점수 차이가 해소되지 않아 인간 해석자가 직접 조정하였다.

이러한 결과는 정렬 원칙 기반 평가 체계가 사람의 해석 개입 없이도 일정 수준의 자동성과 정합성을 확보할 수 있음을 입증하며, 향후 GPAI 사고 위험 평가의 정량화 및 자동화 가능성에 대한 가능성을 보여준다.

Ⅳ. 논의 및 시사점

정렬 원칙 기반 스코어링 체계는 실제 사례에 대해 일관된 해석을 가능하게 하며, LLM 간의 해석 차이를 명확한 원칙 해석 기준을 통해 조정할 수 있음을 보여준다. 특히, 원칙 위반의 구조적 특성과 실제 결과의 피해 정도를 통합적으로 고려하는 방식은 GPAI 위험 평가의 해석 가능성과 실용성을 높이는 데 기여할 수 있다.

Ⅴ. 결론

본 연구는 정렬 원칙 기반 GPAI 위험 평가 체계를 실제 AI 사고 사례에 적용함으로써, 이 체계의 해석 일관성과 실용성을 검토하였다. 전체 238건 중 157건에서 동일한 점수가 도출되었다는 결과는, 정렬 원칙과 점수 기준이 명확히 정의될 경우 고성능 LLM들이 높은 수준의 해석 일치성을 확보할 수 있음을 보여준다. 이는 정렬 기반 스코어링 체계가 해석자의 주관성을 최소화하고 재현 가능한 평가 도구로 기능할 수 있다는 가능성을 시사한다.

또한 해석 불일치가 발생한 81건 중 약 92%에 해당하는 75건이 조율 원칙을 통해 성공적으로 점수 재조정이 가능하였다는 점은, 단일 해석 체계로 발생할 수 있는 편향이나 누락을 다양한 해석 방식의 조율을 통해 상호 보완할 수 있음을 보여준다. 특히 본 연구에서 제안한 구조적 원칙 해석과 결과 중심 해석의 통합 조율 방식은, 정렬 원칙 기반 평가가 단순한 규범 검토를 넘어 실제 피해 수준까지 포괄적으로 반영할 수 있는 유연한 구조임을 입증한다.

이러한 조율 경험은 향후 정렬 위험 평가를 자동화할 때 고정된 기준(rule-based approach)과 확률적 모델(probabilistic LLM)의 혼합형 접근이 효과적으로 활용될 수 있음을 시사한다. 더불어, 사람이 직접 개입해야 했던 6건의 사례를 통해 현재 LLM 기반 평가의 한계 또한 확인할 수 있었으며, 이러한 한계 사례를 기반으로 판단 보완 알고리즘이나 해석 보조 인터페이스를 설계하는 후속 연구가 가능할 것이다.

정렬 원칙 기반 평가 체계는, 결과 중심의 단편적 사고 평가를 넘어서 GPAI의 설계 구조, 목표 설정, 작동 메커니즘 등 시스템 전반의 위험 징후를 사전에 포착할 수 있는 분석 도구로 발전할 수 있다. 본 연구는 그러한 가능성의 실증적 단초를 제공하며, 향후 정렬 기반 위험 프레임워크의 제도화 및 정책화 가능성을 제시한다.

ACKNOWLEDGMENT

This work was supported by the Korean MSIT (Ministry of Science and ICT) as Establishing the foundation of AI Trustworthiness(TTA).

참 고 문 헌

[1] J. Carlsmith, “Is Power-Seeking AI an Existential Risk?,” (https://arxiv.org/abs/2206.13353)

[2] S. Han, E. Kelly, S. Nikou, and E. O. Svee, “Aligning artificial intelligence with human values: reflections from a phenomenological perspective,” AI & Society, vol. 37, pp. 1383-1395, 2022, (https://doi.org/10.1007/s00146-021-01247-4)

[3] J. Ji, T. Qiu, B. Chen, B. Zhang, H. Lou, K. Wang, and Y. Duan, “AI Alignment: A Comprehensive Survey,” 2023, (https://ar5iv.org/abs/2310.19852)

[4] L. Aranda and K. Perset, “Name it to tame it: Defining AI incidents and hazards,” OECD AI Policy Observatory, (https://oecd.ai/en/wonk/defining-ai-incidents-and-hazards)

[5] Shin, Y. and Kang, S. “Systematic Analysis of Risk Factors in General Purpose Artificial Intelligence Centered on Alignment Principles,” ICFICE 2025, July 2025.

[6] OECD, “AI Incidents and Hazards Monitor,” 2025, (https://oecd.ai/en/incidents)

변화하는 디지털 전장: 인공지능 적용의 확대

차용준[1,2], 조한얼[1,3], 이상철[1], 장효석[1,2], 백승호[1,2], 강경일[4,+], 김찬수[1,2,+,*]

1 한국과학기술연구원 인공지능·정보·추론 (AI/R) 연구실

2 과학기술연합대학원대학교 AI-로봇

3 세종대학교

4 합동참모본부

* Correspondence should be addressed to eau@ust.ac.kr.

Evolving Digital Battlefields: Expanded Applications of AI

Yongjun Cha[1,2], Haneol Cho[1,3], Sangchul Lee[1], Hyoseok Jang[1,2], Seungho Baek[1,2], Kyung-il Kang[4,+], Chansoo Kim[1,2,+,*]

1 AI·Information·Reasoning (AI/R) Laboratory, Korea Institute of Science and Tech. (KIST)

2 The Department of AI-Robotics, University of Science and Technology (UST)

3 Sejong University

4 Joint Chiefs of Staff

+ Equally contributed.

요 약

러시아-우크라이나 전쟁과 이스라엘-이란 분쟁의 사례 연구를 통해 AI가 디지털 전장의 역학 관계와 군사 작전의 전략적 접근 방식에 어떤 영향을 미치고 있는지를 고찰한다. AI가 정보 환경을 재편하고 대중의 서사를 조작하는 데 활용되는 방식과 이와 관련된 윤리적, 우려, 특히 감시 증가 및 잠재적 민간이 피해 등의 도전 과제를 제시한다. 두 주요 사례에서 보듯, AI가 확대 적용되는 현대 디지털 전장에서 AI는 단순한 지원 도구를 넘어 전략적 게임 체인저로 부상했다. 미사일과 군수품뿐만 아니라 데이터, 알고리즘, 서사를 통한 전투가 점점 더 중요해지고 있음도 드러난다.

Ⅰ. 서 론

인공지능(AI)의 비약적인 발전은 현대 전장의 양상을 근본적으로 변화시키고 있다. 과거 전장에서의 정보 수집과 분석, 그리고 의사결정 과정은 인간 중심으로 이루어졌으며 방대한 데이터의 처리에는 한계가 따랐고, 복잡한 전술적 판단을 신속하게 내리기 어려웠다.

그러나 AI의 등장은 이러한 제약을 극복하게 하였고, 전장의 자동화와 디지털화를 가능하게 만들었다. 러시아 - 우크라이나 전쟁과 이스라엘 - 이란 분쟁은 AI가 실제 군사 작전에서 어떤 방식으로 활용되고 있는지를 잘 보여주는 대표적인 사례들이다. 이 둘은 AI가 단순한 지원 도구를 넘어 현대전에서 전략적 게임 체인저로 작용하고 있음을 보여준다 [1-3].

Ⅱ. 본론

1. AI와 러시아-우크라이나 전쟁

러시아 - 우크라이나 전쟁은 AI가 ISR(정보·감시·정찰), C2(지휘통제), 사이버전 전반에 걸쳐 어떻게 활용될 수 있는지를 보여주는 결정적 사례이다. 전통적 전쟁에서 정보 수집과 분석은 수작업 중심으로 이루어져 대량의 데이터를 신속히 처리하기 어려웠고, 지휘관의 판단 역시 시간 지연에 크게 영향을 받았다. 그러나 이 전쟁에서 우크라이나는 AI 기반 데이터 융합 및 분석 시스템을 적극적으로 활용함으로써 이러한 한계를 극복하였다.

우크라이나는 위성 이미지, 드론 촬영 영상, 그리고 소셜미디어와 같은 비전통적 데이터까지 통합하여 전장의 디지털 트윈을 구축하였다 [1, 3]. 이러한 체계는 실시간으로 적의 위치와 움직임을 추적할 수 있도록 하여, 기존의 전장 상황 인식보다 훨씬 신속하고 정밀한 목표 식별을 가능하게 했다. 더 나아가 AI 기반 정찰·공격 드론은 아군의 인명 손실을 최소화하면서도 정밀 타격과 감시 임무를 수행하며 전술적 우위를 제공했다 [3].

대표적으로, 우크라이나가 자체적으로 개발·운용 중인 'Delta 상황인식 시스템'은 클라우드 기반의 디지털 전장 관리 플랫폼으로, 상업위성 이미지, UAV(무인항공기) 영상, 군사 통신 자료, 공개 출처 정보(OSINT, Open Source Intelligence)까지 통합한다 [8]. 이러한 클라우드 인프라는 대규모 데이터를 실시간 저장·처리·공유하며, 위성·드론·통신 등 이종 센서 데이터를 결합하고 지휘부와 전선 간 저지연(low-latency) 협업을 지원한다 [16, 17, 22]. 이어서 AI는 이 자료를 전처리·분석해 적의 위치와 위협 수준을 시각화하고, 지휘관은 이를 바탕으로 최종 결정을 내린다 [2, 9].

무인기와 자폭형 드론에는 AI 기반 부분 자율 기능이 적용되어 통신 두절이나 GPS 교란 상황에서도 항법 유지, 회피 기동, 목표 재획득이 가능하다. 이를 통해 인명 손실을 줄이고 정밀 타격과 전술적 유연성을 확보했지만, 자율성은 아직 원격조종·자동화 단계에 머물고 교전 결정은 인간

이 맡고 있다 [9].

러시아는 Lancet 자폭형 드론(loitering munition)을 전쟁에서 실전 배치하였다. Lancet은 정찰과 공격 임무를 동시에 수행할 수 있으며, 우크라이나군의 포병 및 방공 자산을 타격하는 데 널리 활용되었다. 특히 Lancet은 영상 인식 및 표적 추적 알고리즘을 통해 목표를 탐지하고 타격한다 [11].

AI 기반 영상 인식·추적 알고리즘은 컴퓨터 비전과 딥러닝을 활용해 목표물을 실시간 탐지·추적한다. CNN과 객체 탐지 네트워크(YOLO, Faster R-CNN)는 군사 장비나 차량을 인식·분류하고 [12, 13], 추적 단계에서는 칼만 필터, DeepSORT, Siamese 네트워크 등이 적용되어 이동 목표를 안정적으로 추적한다 [14, 15]. 이러한 기술은 GPS 교란 상황에서도 표적 재식별과 회피 후 재공격을 가능케 하며, Lancet과 같은 드론 무기의 정밀성과 자율성을 뒷받침한다.

AI 기반 사이버 전도 향후 발발할 가능성이 있다. 러시아 - 우크라이나 전쟁에서 AI가 직접 활용된 사이버 공격이 수행되었다는 명확한 증거는 현재까지 확인되지 않았다. 실제로 보고된 사이버 작전은 주로 와이퍼 악성코드(HermeticWiper, CaddyWiper, Industroyer2 등), 분산 서비스 거부(DDoS), 피싱 캠페인 등이다 [31].

우크라이나의 사이버 방어는 전통적인 시그니처 기반 탐지를 넘어, 서방 보안 기업들이 제공한 AI/머신러닝 기반 보안 솔루션을 적극 활용한 것이 특징적이다. Microsoft는 Defender와 Sentinel에 탑재된 ML 기반 탐지 시스템을 통해 HermeticWiper, CaddyWiper 등 러시아의 파괴적 악성코드를 조기에 식별하고 차단했다 [32].

향후 사이버 영역에서도 AI는 중요한 역할을 수행할 수 있다. AI는 악성코드 탐지, 공격 패턴 분류, 위협 예측, 자동 대응 등에 활용되어 방어적 역량을 강화한다 [3]. 이러한 기술은 전장을 물리적 공간에 한정하지 않고, 사이버 공간을 병행하는 다차원적 전장으로 확장시켰다.

악성코드 탐지에는 머신러닝과 딥러닝 기법이 활용된다. SVM·랜덤 포레스트는 코드의 정적·행위 특징을 학습해 새로운 악성코드를 식별하고, CNN과 RNN/LSTM은 실행 파일이나 로그 데이터를 분석해 악성 행위를 탐지한다 [18]. 또한 이상 탐지와 클러스터링(k-means, DBSCAN) 기법은 정상·비정상 트래픽을 구분해 다단계 공격을 조기에 식별하는 데 기여한다 [19].

2. 이스라엘-이란 분쟁과 AI

이스라엘 - 이란 분쟁은 AI가 물리적 전장뿐만 아니라 정보전과 인지전(cognitive warfare)의 차원에서 얼마나 강력한 영향을 미치는지를 보여주는 중요한 사례이다. 과거의 정보전은 주로 전통적인 심리전 기법과 언론 조작에 의존했지만, 이번 분쟁에서는 생성형 AI가 핵심적인 역할을 하였다.

2024 - 2025년의 분쟁에서는 고도로 정교한 AI 생성 이미지와 영상이 대규모로 생산되었고, 파괴된 기반시설이나 격추된 항공기, 대규모 시위 장면과 같은 허위 시각자료가 실제처럼 묘사되었다 [4, 8].

이스라엘 - 이란 분쟁에서 군사 여론 분석에는 NLP가 활용되며, 최근 Transformer 기반 언어모델(BERT, GPT 등)이 등장해 분석의 정확성과 속도를 높였다 [10, 20, 21]. 이스라엘은 가자지구 작전에서 다국어 AI 챗봇으로 현지 여론을 실시간 모니터링 했으며, 동시에 생성형 AI는 합성 이미지·영상·음성을 생산·확산시켜 사실과 허구의 경계를 흐렸다.

이스라엘은 AI 기반 오디오 분석 도구를 사용하여 폭음이나 공습의 위치를 추적하였다. 안면인식 시스템은 부상으로 얼굴이 손상되거나 가려진 인물까지 식별할 수 있도록 개발되었다. 이러한 기술은 실제로 하마스 지휘관의 위치를 특정하는 데 활용되었으며, 이는 터널 복합체에 대한 공습으로 이어졌다 [7].

반면 이란은 가짜 뉴스 앵커를 활용한 딥페이크 방송으로 국제 정보 환경을 교란하였다 [6]. 딥페이크 기술의 핵심은 생성적 적대 신경망(GANs)으로, 가짜 데이터를 만드는 '생성자'와 이를 구별하는 '판별자'가 경쟁하며 학습해 실제와 거의 구분되지 않는 결과물을 만들어낸다 [23].

오토인코더(Autoencoders)는 입력 데이터를 압축·복원해 얼굴 교환에 활용되며, 원본 인물의 표정과 움직임을 유지한 채 다른 얼굴을 합성할 수 있다. 이러한 딥페이크 기술은 영화나 교육 분야에 활용되는 등 긍정적 가능성을 지니지만, 동시에 정치적 허위 영상, 가짜 뉴스, 사생활 침해 등 악용 위험도 크다 [24].

3. 두 분쟁을 통해 본 전쟁 패러다임의 변환

러시아 - 우크라이나 전쟁은 AI가 단순한 보조 기술에서 벗어나 물리적 전장의 효율성과 속도를 좌우하는 핵심 요소로 부상한 전환점이었다. 전쟁 이전에는 완전 자율 무기와 대규모 드론 스웜 운용이 실현될 것이라는 기대가 있었으나, 실제로는 항법·비행·목표 식별 수준에서만 자율성이 구현되었고 살상 여부는 여전히 인간이 통제하였다 [27].

또한 전역 단위 자동화를 가능하게 할 것이라 예상되었던 대규모 드론 스웜은 소규모 실험 수준에 머물렀으며, 네트워크 안정성과 명중률의 한계가 드러났다. 그 대신 Delta 상황인식 시스템을 통해 드론, 위성, 센서 데이터를 실시간으로 융합하여 수백여 개의 목표를 관리할 수 있었고, FPV·자살 드론과 같은 저비용 무기가 고가 장비를 무력화하는 사례가 확산되었다. 이처럼 전쟁 전 예상과 달리 실제로는 저비용·고효율 기술과 실시간 정보 융합 체계가 부상하면서, AI는 전장의 정보 우위와 작전 주도권을 결정하는 핵심 기술로 자리매김하게 되었다 [27, 28].

이스라엘 - 이란 분쟁에서는 AI의 역할이 물리적 전투를 넘어 정보전과 인지전의 중심으로 확장되었다. 전쟁 이전까지는 AI가 주로 정찰과 타격의 정밀성을 높이는 물리적 전투 지원 도구로 여겨졌으나, 실제로는 정보 환경과 여론을 통제하는 도구로 적극 활용되었다 [26, 29].

이스라엘은 가자지구 작전에서 다국어 AI 챗봇을 활용하여 현지 여론을 실시간 분석하였고, 동시에 생성형 AI는 합성 이미지와 영상, 음성을 대량 생산하여 소셜미디어를 통해 빠르게 확산시켰다. 이처럼 AI는 전쟁의 초점을 무기와 병력에서 여론과 인식으로 이동시켰으며, 전장의 범위를 물리적 충돌에서 정보와 서사 통제까지 확장시켰다. 이는 AI가 단순한 물리적 전투의 보조 기술을 넘어 국제 안보 전략의 핵심 변수로 부상했음을 보여준다 [30].

Ⅲ. 결론

러시아 - 우크라이나 전쟁과 이스라엘 - 이란 분쟁은 AI가 단순한 정찰·보조 기술을 넘어 전쟁의 성격 자체를 바꾸는 전략적 자산으로 부상했음을 잘 보여준다. 우크라이나 사례에서는 드론·위성 데이터를 실시간으로 통합하는 상황 인식 체계와 저비용 드론 전술이 전장의 물리적 양상을 변화시켰고, 이스라엘 - 이란 분쟁에서는 생성형 AI와 딥페이크 기술이 정보전과 심리전의 핵심 도구로 활용되면서 전쟁의 무게 중심이 물리적 충돌에서 여론과 인식의 장으로 확장되었다.

그러나 AI의 부상은 동시에 심각한 도전 과제도 드러냈다. 완전 자율 무기의 부재 속에서 'AI 제안 - 인간 결심' 구조가 유지되고 있으나, 잘못된 식별로 인한 민간인 피해, 허위 정보와 딥페이크 확산에 따른 사회적 신뢰 훼손, 자율 무기 사용의 법적 책임 공백 등은 여전히 해결되지 않은 문제들이다. 이는 AI 군사 활용이 기술의 문제가 아니라 국제 규범과 제도

적 장치 마련의 과제임을 보여준다. 앞으로의 국제 안보 질서는 AI의 발전과 이를 관리하는 규범적 틀에 의해 크게 좌우될 것이다.

ACKNOWLEDGMENT

This research was funded by the grant Nos. 2023-00262155; 2024-00339583; 2024-00460980; and 2025-02304717 (IITP) funded by the Korea government (the Ministry of Science and ICT).

참 고 문 헌

[1] International Centre for Defence and Security (ICDS). Russia's War in Ukraine: Artificial Intelligence in Defence of Ukraine, 2024.

[2] Tsur, M.; Morag, M. Navigating the AI frontier: Insights from the Ukraine conflict for NATO's governance role in military AI. 873 Defense & Security Analysis 2024.

[3] Ratti, P. From Ukraine to Gaza: Artificial intelligence in war. Critique internationale 2024, 243, 39 - 55.

[4] European Digital Media Observatory (EDMO). The First AI War: How the Iran-Israel Conflict Became a Battlefield for Generative 876 Misinformation, 2024.

[5] Nickel, D. AI slop spreads in Israel-Iran war. POLITICO 2025.

[6] Radware. Hybrid Warfare Unfolded: Cyberattacks, Hacktivism and Disinformation in the 2025 Israel-Iran War, 2025.

[7] Timesof Israel. Israel using AI to pinpoint Hamas leaders, find hostages in Gaza tunnels - report. The Times of Israel 2024.

[8] CSIS, "Does Ukraine Already Have a Functional CJADC2 Technology?" Center for Strategic and International Studies, February 21, 2024, https://www.csis.org/analysis/does-ukraine-already-have-functional-cjadc2-technology?

[9] NATO Allied Command Transformation, "The DELTA System and CWIX," NATO Allied Command Transformation, July 11, 2023, https://www.act.nato.int/article/delta-system-cwix/.

[10] Apurv Gupta, Suthikshn Channarayapatna Kumar, and Odelu Ojjela, "Sentiment Analysis for Defence Ecosystem and Armed Forces?" presented at the European Conference on Cyber Warfare and Security, June 2025, https://www.researchgate.net/publication/393048207_Sentiment_Analysis_for_Defence_Ecosystem_and_Armed_Forces.

[11] Nafuye, I. (2024). Weaponizing Artificial Intelligence in the Russia-Ukraine War.

[12] Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement. arXiv:1804.02767.

[13] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN. NeurIPS.

[14] Bewley, A., Ge, Z., Ott, L., Ramos, F., & Upcroft, B. (2016). Simple online and realtime tracking. IEEE ICIP.

[15] Bertinetto, L., Valmadre, J., Henriques, J. F., Vedaldi, A., & Torr, P. (2016). Fully-convolutional siamese networks for object tracking. ECCV.

[16] Mell, P., & Grance, T. (2011). The NIST Definition of Cloud Computing. NIST Special Publication 800-145.

[17] Zhang, Q., Cheng, L., & Boutaba, R. (2010). Cloud computing: state-of-the-art and research challenges. Journal of Internet Services and Applications, 1(1), 7 - 18.

[18] Saxe, J., & Berlin, K. (2015). Deep neural network based malware detection using two dimensional binary program features. MALWARE.

[19] Buczak, A. L., & Guven, E. (2016). A survey of data mining and machine learning methods for cyber security intrusion detection. IEEE Comm. Surveys & Tutorials.

[20] Goldberg, Y. (2017). Neural Network Methods for Natural Language Processing. Synthesis Lectures on Human Language Technologies.

[21] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL-HLT.

[22] Systematic. "The Silver Lining of Military Cloud Computing: Benefits and Solutions." Systematic, 27 July 2023, systematic.com/us/industries/defense/news-knowledge/blog/the-silver-lining-to-military-cloud-computing/.

[23] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative Adversarial Nets. In Advances in neural information processing systems (pp. 2672-2680).

[24] Verdoliva, L. (2020). Deepfake video detection methods: A survey. IEEE Journal of Selected Topics in Signal Processing, 14(5), 1083-1094.

[25] International Centre for Defence and Security. (2025). Russia's War in Ukraine: Artificial Intelligence in Defence of Ukraine. https://icds.ee/en/russias-war-in-ukraine-artificial-intelligence-in-defence-of-ukraine/

[26] European Digital Media Observatory. (2025, July 14). The First AI War? How the Iran-Israel Conflict Became a Battlefield for Generative Misinformation. https://edmo.eu/publications/the-first-ai-war-how-the-iran-israel-conflict-became-a-battlefield-for-generative-misinformation/.

[27] Bondar, Kateryna, and Samuel Bendett. "The Russia-Ukraine Drone War: Innovation on the Frontlines and Beyond." Center for Strategic and International Studies, 28 May 2025, www.csis.org/analysis/russia-ukraine-drone-war-innovation-frontlines-and-beyond.

[28] Hunder, Max. "Ukraine Rolls out Dozens of AI Systems to Help Its Drones Hit Targets." Reuters, 31 Oct. 2024, www.reuters.com/world/europe/ukraine-rolls-out-dozens-ai-systems-help-its-drones-hit-targets-2024-10-31/.

[29] The Times of Israel Staff. "Israel Using AI to Pinpoint Hamas Leaders, Find Hostages in Gaza Tunnels – Report." The Times of Israel, n.d., www.timesofisrael.com/israel-using-ai-to-pinpoint-hamas-leaders-find-hostages-in-gaza-tunnels-report/.

[30] European Digital Media Observatory. "The First AI War: How the Iran-Israel Conflict Became a Battlefield for Generative Misinformation." European Digital Media Observatory, 25 Apr. 2024, edmo.eu/publications/the-first-ai-war-how-the-iran-israel-conflict-became-a-battlefield-for-generative-misinformation/.

[31] Microsoft. (2024). Microsoft Digital Defense Report 2024. Retrieved from

https://www.microsoft.com/en-us/security/security-insider/threat-landscape/microsoft-digital-defense-report-2024

[32] Smith, B. (2022, June 22). Defending Ukraine: early lessons from the cyber war. Microsoft On the Issues. https://blogs.microsoft.com/on-the-issues/2022/06/22/defending-ukraine-early-lessons-from-the-cyber-war/

산불 발화점 추정을 위한 디지털 트윈 기반 감시 프레임워크

김동영, 김상원, 장인수, 김광주, 이경오
한국전자통신연구원

dongyoung.kim@etri.re.kr, eddiekim@etri.re.kr, jef1015@etri.re.kr, kwangju@etri.re.kr, longweek7@etri.re.kr

Digital Twin-Based Surveillance Framework for Wildfire Ignition Point Estimation

Dongyoung Kim, Sangwon Kim, In-su Jang, Kwang-Ju Kim, Kyoungoh Lee
Electronics and Telecommunications Research Institute (ETRI)

요 약

최근 산불은 전 세계적으로 발생 빈도와 피해 규모가 지속적으로 증가하고 있다. 초기 발견과 대응의 신속성과 정확성에 따라 피해 규모는 크게 달라진다. 따라서 조기에 산불 발화 지점을 정확히 추정하여 신속하게 전달하는 것이 중요하다. 하지만 기존 연구는 단순 영상 분석정도에 머물러 있어 실제 산불 발화 지점 추정하는 데 어려움이 있다. 본 연구는 산악 감시카메라 기반 산불 발화 지점 추정 알고리즘을 통해 발화점을 추정한다. 또한, 검출된 영상에서의 발화지점을 디지털 트윈 산악 지형 기반으로 실제 좌표로 산출하여 정확한 발화지점을 제공하는 통합 프레임워크를 제안한다.

1. 서 론

기후 변화와 건조한 기상 조건, 인위적 요인으로 인해 최근 국내외 산불 발생 빈도와 피해 규모가 급격히 증가하고 있다 [1]. 산불은 단순한 환경 재해를 넘어 대기 오염, 인명 피해, 지역 사회의 경제적 손실로 이어지며 국가적 차원에서도 심각한 위기 요인으로 대두되고 있다. 특히 산림 지형은 넓고 복잡하여 화재가 시작되는 지점을 신속하고 정확하게 파악하지 못할 경우, 초기 대응 지연으로 인해 피해가 기하급수적으로 확대될 수 있다.

이러한 이유로 산불 조기 감시와 발화 지점의 정확한 위치 추정은 산불 대응의 핵심 요소이다. 기존의 센서 기반 화재 감지 기술은 [2] 연기나 온도의 변화를 탐지하지만 범위와 환경이 제한적이고 산악 지형과 같은 광범위한 지역에서 활용하는 것은 어렵다. 딥러닝 모델을 활용한 영상 기반 산불 탐지 기술은 [3] 실시간성과 정확성으로 인해 주목을 받고 있으나, 대부분 영상에서의 화재의 경계 박스와 같은 형태의 단순히 발생 여부와 위치를 검출하는 수준에 머물러 있다. 즉, 실제 지형상의 발화 지점에 대한 정교한 좌표 혹은 지리적인 정밀한 위치까지 제공하지 못한다. 따라서 이러한 한계로 인해 영상 기반의 산불 감지 시스템에는 여전히 한계가 있다.

본 연구는 이러한 한계를 극복하기 위해 감시카메라 기반 발화점 추정과 디지털 트윈 지형 모델과의 매핑을 통한 실제 발화 지점 추정 프레임워크를 제안한다. 본 연구의 핵심 기여는 세 가지이다. 첫째, 합성 연기 데이터를 기반으로 학습된 객체 탐지 모델을 활용하여 2D 영상에서 객체탐지 모델을 통하여 연기를 검출하고 발화점을 추정한다. 둘째, 드론으로 촬영한 영상을 활용해 감시카메라 주변 지형의 디지털 트윈 산악 지형 모델을 구축한다. 셋째, 감시카메라 영상에서 추정된 발화점을 디지털 트윈 산악 지형 모델과 매칭하고, 이를 Unity 환경의 투사를 통하여 3D 좌표로 변환한다. 최종적으로, 제안하는 프레임워크는 단순 탐지를 넘어 발화 지점을 실제 좌표로 추정할 수 있어, 산불의 초기 진압 및 자원 배치 효율성을 크게 높일 수 있다.

2. 산불 발화점 추정 프레임워크

2.1 산불탐지 및 발화점 추정

우리는 산악 감시카메라 영상을 입력으로 하여 실시간 산불의 발화점 탐지를 수행한다. 모델 성능 향상을 위해, 우리는 Maya 시뮬레이터를 사용하여 가상의 산불 데이터셋을 구축하고, YOLOv4 [4] 객체탐지 모델을 사용하여 산불 데이터를 학습하고, 실제 산불 데이터로 검증한다. 산불은 하단에서 발생하여 상부로 확산되는 특성을 갖는다. 따라서 탐지 결과로부터 산불의 발화점을 추정하기위해 경계 검출 필터(Sobel)를 적용하여 연기 영역의 경계선을 추출한 뒤, 고강도의 픽셀 중 하단부에 위치한 픽셀 집합을 발화점 후보로 선정한다. 이후 이 후보 포인트들의 무게중심을 계산하여 최종 발화점으로 결정한다.

2.2 디지털 트윈 맵 생성

추정된 영상 좌표의 발화점으로부터 실제 위치 정보를 얻기 위해, 그림 1 의 (b)와 같이 실제 지형 정보를 3 차원 데이터와 같은 형태로 디지털화 한다. 드론을 이용해 감시카메라가 존재하는 산악 지역을 약 4km 씩 나누어

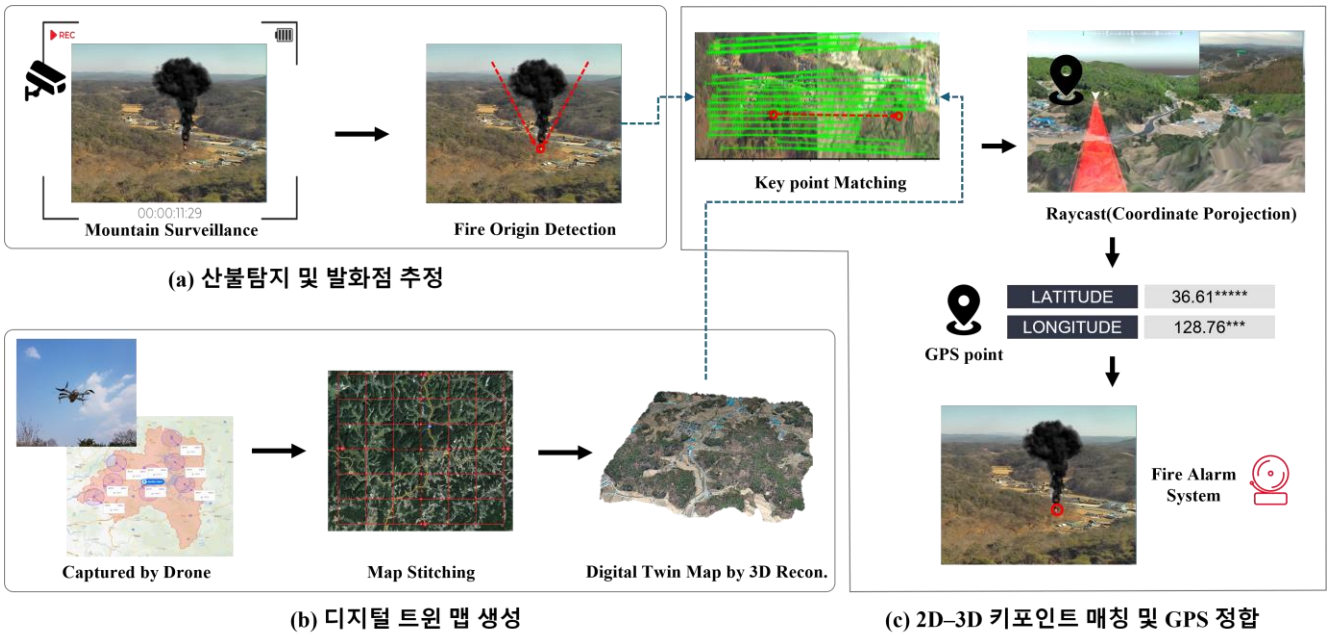


그림 1. 디지털 트윈기반 산불 발화점 추정 프레임워크

촬영한다. 그 뒤, Unity 기반의 3D 산악 데이터로 재구성하고, 랜드마크들의 실제 좌표를 이용해 3D 영상과 정합하여 실제 좌표를 반영한 디지털 트윈 지형 모델을 생성한다.

2.3 2D – 3D 키포인트 매칭 및 GPS 정합

감시카메라 영상의 발화점과 디지털화 된 산악 지형과 좌표 매핑을 통해 실제 물리적인 위치를 획득한다. 키포인트 매칭 모델을 사용하여, 감시카메라 영상과 디지털 트윈 뷰 간의 주요 특징점을 매칭하여 감시카메라 영상과 디지털 트윈 영상을 정렬한다. 그 뒤, 2D 좌표를 대응되는 3D 지형 좌표로 투영하여 해당 부분의 실제 GIS 좌표를 구하여 최종적으로 발화 지점의 위도·경도 좌표를 산출한다. 이를 통해 사용자는 그림 1 의 (c)와 같이 디지털 트윈 지형 모델상에서 발화 위치를 확인하고, 실제 GIS 좌표를 획득하여 산불 대응에 활용한다.

3. 결론

본 논문에서는 산불 조기 감시를 위해 감시카메라 영상을 활용한 디지털 트윈 기반 발화 지점 추정 프레임워크를 제안하였다. 제안된 방법은 연기 탐지를 통해 발화 후보 영역을 찾아내고, 컴퓨터 비전 기반 분석으로 발화 지점을 추정한 뒤, 디지털 트윈 영상과의 정합을 통해 실제 GPS 좌표로 변환했다. 이를 통해 단순 2D 기반 화재 여부 탐지에서 나아가 화재의 실제 위치 정보를 제공하여 초기 대응 효율성을 향상시킬 수 있었다. 이는 향후 산불 감시 체계에서 초기 진압 경로 최적화, 자원 배치 효율화, 피해 확산 방지에 활용할 수 있다.

하지만 본질적으로 감시카메라 영상의 한계로 인해, 지형적으로 산이나 건물에 의한 발화점이 가려지는 문제는 여전히 한계가 있다. 발생하는 오류나 실제 그 결과 랜드마크 정합시 발생하는 매핑된 GPS 좌표 오차와 같은 문제가 발생할 수 있다. 따라서 향후 연구에서는 3D 영상 기반 투사 과정에서 가려지는 문제를 최소화하여, 보다 정확한 발화 지점 추정과 랜드마크를 늘려 GPS 좌표 오차를 줄이고 우리의 프레임워크의 적용 범위 확장을 목표로 할 예정이다.

ACKNOWLEDGMENT

본 연구 논문은 한국전자통신연구원 연구운영지원사업의 일환으로 수행되었음. [25ZD1120, 대경권 지역산업 기반 ICT 융합기술 고도화 지원사업].

참 고 문 헌

[1] Cunningham, Calum X., Grant J. Williamson, and David MJS Bowman. "Increasing frequency and intensity of the most extreme wildfires on Earth." *Nature ecology & evolution* 8.8 (2024): 1420-1425.

[2] Gaur, Anshul, et al. "Fire sensing technologies: A review." *IEEE Sensors Journal* 19.9 (2019): 3191-3202.

[3] Sathishkumar, Veerappampalayam Easwaramoorthy, et al. "Forest fire and smoke detection using deep learning-based learning without forgetting." *Fire ecology* 19.1 (2023): 9.

[4] Bochkovskiy, Alexey, Chien-Yao Wang, and Hong-Yuan Mark Liao. "Yolov4: Optimal speed and accuracy of object detection." *arXiv preprint arXiv:2004.10934* (2020).

소수자 게임의 집단 효용 극대화: 강화학습의 보상 균형화 기반 신속 안정화 정책

백승호, 이상철, 조한얼, 김찬수*
과학기술연합대학원대학교, 세종대학교, *한국과학기술연구원
092236@kist.re.kr, eau@ust.ac.kr

Maximizing Collective Utility in the Minority Game: A Reward-Balancing Based Rapid Stabilization Policy for Reinforcement Learning

Seungho Baek, Haneol Cho, Chansoo Kim*
UST (University of Science and Technology), Sejong University,
*KIST (Korea Institute of Sci. and Technology)

요 약

본 연구는 경제학의 소수자 게임(minority game)의 대표적 사례인 El Farol Bar 문제를 강화학습(Q-learning) 기반으로 다룬다. 기존 접근은 집단적 참석자(attendance)가 수용량(capacity) 주변에서 불안정하게 진동하는 특성을 보였으며, 이는 사회적 낭비와 효율성 저하를 초래한다. 본 연구는 이러한 불안정성을 줄이고 집단적 효율성을 최대화하기 위해, 에이전트 특성을 반영한 보상 균형화(reward balancing) 방법을 제안한다. 제안 방법은 에이전트별 성향(초기 편향, 결정성, 관성)에 따라 유리/불리 유형을 구분하고, 이에 기반해 차등적인 보상 조정과 재분배를 수행한다. 시뮬레이션 결과, Q-learning에 보상 균형화 알고리즘을 더하여, attendance의 수렴 속도가 증가하고 자원 낭비(수용 인원 초과)가 감소함을 확인하였다. 이는 다중 에이전트 시스템에서 에이전트의 이질성을 고려한 정교한 보상 설계가 집단적 조정을 달성하는 데 매우 효과적임을 보인다.

Ⅰ. 서 론

엘 파롤 바(El Farol Bar) 문제는 소수자 게임(Minority Game)의 대표적 사례로, 제한된 자원을 두고 다수의 주체가 경쟁할 때 나타나는 집단적 의사결정 메커니즘을 설명한다. 경제학, 통계물리학, 복잡계 과학에서 집단 행동을 모사하는 전형적 모델로 다뤄져 왔으며, 개별 에이전트가 효용을 극대화하려는 과정이 오히려 집단적 비효율을 초래한다는 점을 보여준다. 이 문제를 분석하기 위해 Q-러닝과 같은 강화학습 알고리즘이 널리 활용되어 왔다. 에이전트들은 과거 경험으로부터 보상을 학습해 합리적 정책을 수립하지만, 전통적 접근에서는 총 참석자 수가 수용치(capacity)를 중심으로 크게 진동하며 예측 불가능한 변동성을 보인다. 이는 교통 체증, 전력망 과부하, 금융 시장 변동성처럼 현실 시스템의 비효율·사회적 비용과 직결된다.

그러나 기존 연구들은 대부분 모든 에이전트가 동일한 보상 구조와 학습 규칙을 따른다고 가정해, 현실의 다양한 성향(선호, 위험 감수, 학습 속도 등)을 충분히 반영하지 못한다. 이는 일부 에이전트가 국소 최적 전략에 고착되고, 전체 시스템의 불안정성이 완화되지 않는 한계로 작용한다.
본 연구는 표준 Q-러닝을 기준선으로 두고, 에이전트의 다양성을 반영한 보상 균형화 알고리즘을 제안한다. 개별 특성과 성과에 따라 보상을 동적으로 조정함으로써, 집단의 빠른 안정화와 사회적 비용 절감을 동시에 달성하는 가능성을 탐구한다.

Ⅱ. 본론

2.1 시뮬레이션 환경 및 행동(액션)

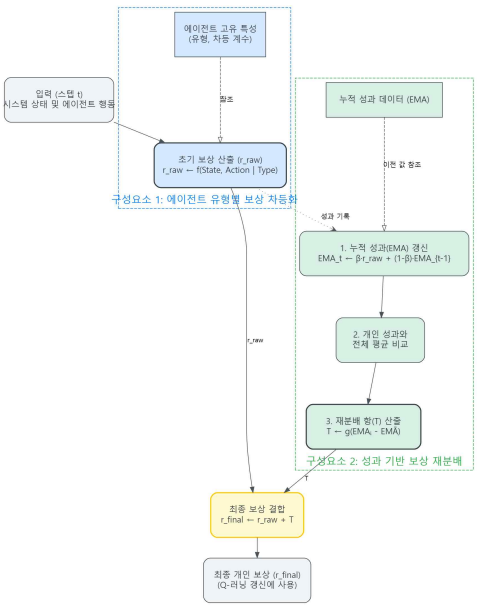


그림 1 보상 균형화 강화학습 알고리즘의 순서도

본 연구는 표준 보상 구조를 사용하는 기준 Q-러닝 모델과 보상 균형화가 적용된 Q-러닝 모델의 성능을 비교한다. 두 모델은 상태 공간, 행동 공간, Q-값 갱신 규칙 등 기본 프레임워크는 동일하지만 보상 설계에서 차이를 가진다. 기준 모델은 소수자 게임에 적용되는 전형적 Q-러닝 방식으

로, 에이전트는 최근 m 스텝의 혼잡도를 관찰해 과밀이 아닐 때 참석하거나 과밀일 때 불참하면 +1, 반대로 행동하면 -1의 보상을 받는다. 이 보상은 모든 에이전트에게 동일하게 주어진다.

이와 대조적으로, 본 논문이 제안하는 핵심 모델은 기준 Q-러닝 프레임워크 위에 '보상 균형화' 메커니즘을 통합한다. 그림 1에 나타난 바와 같이, 먼저 입력 단계에서 각 스텝의 시스템 상태와 에이전트 행동이 관찰된다. 이어서 구성요소 1: 유형 기반 보상 차등화에서 에이전트 고유 특성(행동 편향, 관성 등)을 참조하여 초기 보상 r_{raw} 이 산출된다. 이때 혼잡을 유발할 가능성이 큰 그룹에는 과밀 상황에서 더 큰 페널티를, 불리한 그룹에는 상대적으로 완화된 보상이 부여된다. 다음으로 구성요소 2: 성과 기반 보상 재분배에서는 각 에이전트의 누적 성과가 지수이동평균(EMA)으로 갱신되고, 이를 집단 평균과 비교해 성과 편차를 계산한다. 이 편차에 기반하여 재분배 항 T가 산출되며, 이는 성과가 과도하게 높은 에이전트의 보상 일부를 평균 이하 성과자에게 이전하는 역할을 한다. 마지막으로 두 신호가 결합되어 최종 보상 $r_{final} = r_{raw} + T$ 가 결정되고, 이는 Q-값 갱신에 사용된다. 이러한 방법을 통해 에이전트 간 이질성을 수용하면서도 집단 전체의 안정적 균형을 위한 보상 체계를 구현한다.

2.2 시뮬레이션 결과

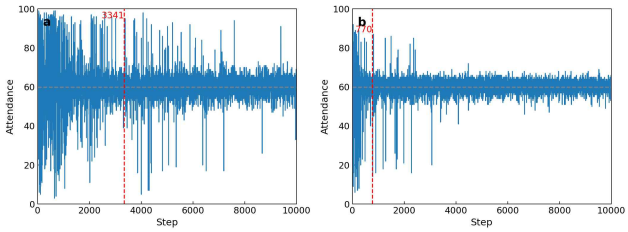


그림 2 Q-Learning 베이스와 보상 균형화 알고리즘이 추가된 모델의 비교

그림 2의 시계열 데이터는 두 모델의 동태적 성능 차이를 명확히 보여준다. 기준 Q-러닝 모델은 획일적인 보상 구조로 인해 에이전트들의 진동을 효과적으로 억제하지 못하였다. 그 결과, 시스템은 수용 인원을(60 명) 중심으로 큰 진폭의 진동을 지속하며 안정화되기까지 3341 스텝이라는 시간이 소요되었다.

반면, 개인화된 보상 균형화 모델은 시스템의 변동성을 현저히 감소시켜 수렴 시간을 770 스텝으로 단축했다. 이는 수렴에 필요한 시간을 77% 개선한 것으로, 제안된 메커니즘의 효율성을 입증한다. 이러한 결과는 유형별 보상 차등화와 성과 기반 재분배를 통해 에이전트들의 행동을 조율하고, 비효율적인 탐색 과정을 최소화하여 시스템을 안정화할 수 있었다. 결과적으로, 본 연구에서 제안한 보상 구조가 시스템의 불안정성을 효과적으로 제어하고 빠르게 수렴함을 실험적으로 확인하였다.

Ⅲ. 결론

본 연구는 엘 파롤 문제에서 표준 Q-러닝의 성능을 개선하기 위해 '보상 균형화 강화학습'을 제안했으며, 시스템의 수렴 시간을 77% 단축하는 효과를 확인했다. 이는 집단적 효율성을 높이는 데 보상 구조의 설계가 학습 알고리즘의 선택만큼 중요하다는 것을 보여준다. 따라서 본 연구가 제시한 '보상 균형화' 전략은, 교통, 에너지, 보건과 같은 소수자 게임을 포함하는 실제 복잡계 시스템에서 더욱 효율적인 전략을 수립할 가능성을 제공한다.

ACKNOWLEDGMENT

This research was funded by the grant Nos. 2023-00262155; 2024-00339583; 2024-00460980; and 2025-02304717 (IITP) funded by the Korea government (the Ministry of Science and ICT).

참 고 문 헌

[1] W. Brian Arthur, "Inductive Reasoning and Bounded Rationality: The El Farol Problem," American Economic Review: Papers and Proceedings, vol. 84, no. 2, pp. 406 - 411, 1994.

[2] Esteban Moro, "The Minority Game: An Introductory Guide," arXiv preprint, 2004.

[3] David Sherrington and Tobias Galla, "The minority game: effects of strategy correlations and timing of adaptation," arXiv preprint, 2003.

[4] Willemien Kets, "The minority game: An economics perspective," arXiv preprint, 2007.

[5] Radu Manuca, Yi Li, Rick Riolo, Robert Savit, "The Structure of Adaptive Competition in Minority Games,"

[6] Si-Ping Zhang, Jia-Qi Dong, Li Liu, Zi-Gang Huang, Liang Huang, Ying Cheng Lai, "Reinforcement learning meets minority game: Toward optimal resource allocation," Physical Review E, vol. 99, no. 3, Article 032302, 2019. DOI: 10.1103/PhysRevE.99.032302

[7] Guozhong Zheng, Weiran Cai, Guanxiao Qi, Jiqiang Zhang, Li Chen, "Optimal Coordination in Minority Game: A Solution from Reinforcement Learning," arXiv preprint, Dec. 2023.

[8] Duncan Whitehead, "The El Farol Bar Problem Revisited: Reinforcement Learning in a Potential Game," University of Edinburgh Technical Paper, 2008.

공공 AI 사업 상호호환성 확보를 위한 요구사항에 관한 연구

박눈솔, 최한경, 양진규, 김동희
한국지능정보사회진흥원
nspark@nia.or.kr, hotspot@nia.or.kr, jky@nia.or.kr, dhk@nia.or.kr

A Study on the Requirements for Ensuring Cross-Compatibility in Public AI Projects

NunSol Park, HanKyoung Choi, JinKyu Yang, DongHee Kim
NIA(National Information Society Agency).

요 약

본 논문은 공공 부문에서 인공지능(AI) 기반 시스템의 상호호환성(Cross-Compatibility)을 확보하기 위한 기준을 제시한다. 인공지능 기술의 발전과 상업적 모델의 다양화에 따라, 공공 부문에서는 AI 서비스 도입에 대한 수요가 급증하고 있다. 그러나 AI 시스템이 특정 클라우드(Cloud)나 AI 모델에 의존하여 개발될 경우 기술 변화에 대응하거나 시스템을 전환하고 확장하는 데에 제약이 발생할 수 있다. 이에 본 연구는 공공 AI 서비스에서 상호호환성 확보를 위한 설계 요구사항을 제시하여, 다양한 플랫폼 및 모델 간의 상호호환성을 확보할 수 있는 방법론을 제시한다. 이를 통해 공공기관은 특정 기술에 종속되지 않고, 다양한 플랫폼 및 모델을 유연하게 도입할 수 있으며, 장기적인 유지·보수 비용을 절감하고 기술적 자율성을 강화할 수 있다. 본 연구는 공공 AI 서비스의 지속 가능성 및 확장성 확보를 위한 기반을 마련하는 데 기여할 것이다.

Ⅰ. 서 론

최근 AI 기술의 발전과 상업적 모델의 다양화에 따라 공공 부문에서 AI 서비스를 도입하려는 수요가 급격히 증가하고 있다. 정부 및 공공기관은 행정 효율성 제고, 맞춤형 서비스 제공, 데이터 기반 정책 결정을 위해 AI 기술을 적극적으로 활용하고자 하지만, 실제 도입 과정에서는 여러 한계와 제약이 발생하고 있다. 특히 AI 시스템이 특정 클라우드나 모델에 종속적으로 설계될 경우, 기술 변화에 따른 시스템 전환, 확장 및 유지·보수 과정에서 높은 비용과 기술적 제약을 감수해야 하는 문제가 나타나고 있다[1]. AI 생태계가 빠르게 변화하는 환경에서 공공 부문 서비스의 안정성과 확장성을 보장하기 위해서는 AI 인프라, 클라우드 플랫폼, AI 모델 간 상호호환성(Cross-Compatibility) 확보가 필수적이다. 상호호환성은 서로 다른 AI 환경에서 서비스, 데이터, 모델이 원활하게 연계·활용될 수 있는 능력을 의미하며, 이를 통해 공공기관은 특정 기술에 종속되지 않고 다양한 공급업체 및 솔루션을 활용할 수 있다[2]. 나아가 상호호환성 확보는 장기적인 유지관리 비용을 절감하고, 공공 AI 서비스의 지속 가능성과 기술적 자율성을 강화하는 핵심 전략으로 작용한다.

그러나 현재 국내 공공 부문에서 추진 중인 다수의 AI 사업은 각기 다른 표준, 모델 형식, 플랫폼 환경을 기반으로 운영되고 있어, 모델 교체 및 기술 이전 과정에서 호환성 문제가 빈번하게 발생하고 있다[3]. 이러한 상황에서 공공기관이 다양한 AI 기술을 유연하게 도입하고 효과적으로 운영하기 위해서는 플랫폼 독립성과 개방형 표준을 기반으로 한 상호호환성 확보 전략이 필수적이다. 이에 본 논문에서는 공공 부문 AI 서비스에서 상호호환성을 확보하기 위한 설계 요구사항을 제시하고자 한다.

Ⅱ. 본론

1. 공공 AI 시스템의 도입 현황

최근 정부와 공공기관은 다양한 사회 문제 해결과 행정 효율성 제고를 목적으로 인공지능(AI) 기반 서비스의 도입을 적극 추진하고 있다. 민원 자동응답, 지능형 문서 분석, 예측 분석 시스템 등 다양한 활용 사례가 등장하고 있으며, 중앙정부는 물론 지방자치단체와 공공기관에서도 AI 실증사업이 확산되고 있다. 그러나 이러한 확산에도 불구하고, 많은 AI 프로젝트는 특정 플랫폼이나 개발사에 종속된 방식으로 구축되고 있어 기술 전환이나 확장 시 큰 제약이 발생하고 있다. 예컨대, AI 모델이 특정 딥러닝 프레임워크(PyTorch, TensorFlow 등)에 의존하거나, 자체 포맷 모델이 저장된 경우, 다른 환경에서의 재사용이나 전환이 매우 어렵다[4]. 이러한 문제를 해결하기 위해서는 AI 시스템 간 상호호환성 확보가 필수이다.

2. 상호호환성 개념 정의

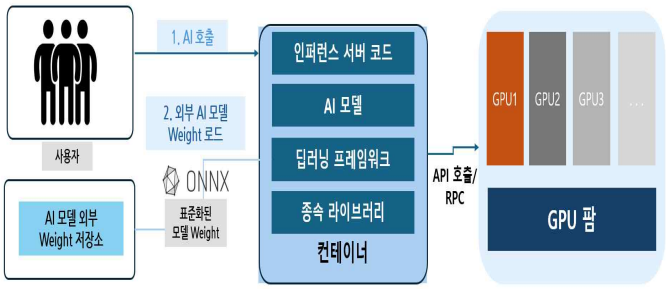
상호호환성(Cross-Compatability)은 AI 모델이 서로 다른 플랫폼, 시스템, 인프라 간에 기술적 제약 없이 기능, 데이터 등을 자유롭게 연계하고 이전·전환 할 수 있는 특성을 의미하며, 이를 통해 특정 기술에 종속되지 않는 유연한 AI 서비스 환경을 구현할 수 있다. 상호호환성 세부 구성 요소는 2가지로 나눌 수 있다. 첫 번째는 AI 인프라 변경에 따른 상호호환성이며 두 번째는 AI 모델 변경에 따른 상호호환성이다.

AI 인프라 변경에 따른 상호호환성은 클라우드, 서버, GPU 등 AI 인프라 환경의 변화에도 서비스 추가 개발 없이 AI 모델이 이식되어 동작하는 것을 의미한다. AI 모델 변경에 따른 상호호환성은 기존 시스템 구조나 API 포맷을 유지하면서, 다양한 AI 모델을 유연하게 교체하고 갱신하는 것을 의미한다.

3. 상호호환성 확보 방안

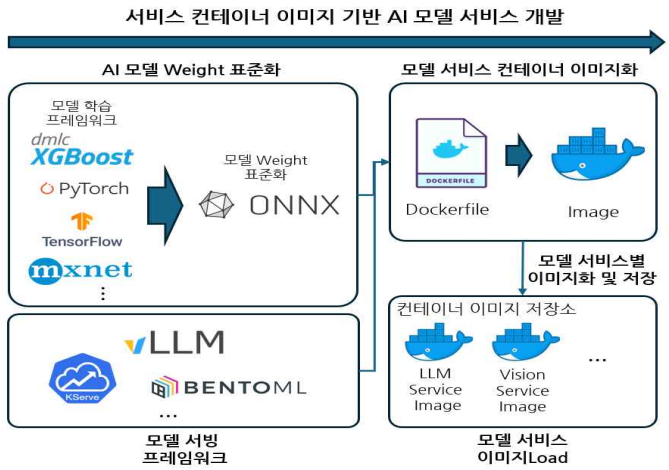
먼저 AI 인프라 변경에 따른 상호호환성 확보를 위한 요구사항은 다음과

같다. AI 모델은 컨테이너를 이용하여 저장 및 배포하는 것이다. 컨테이너 기반 환경(Docker)을 이용하여 AI 모델 실행에 필요한 코드, 라이브러리 등을 독립된 패키지로 묶어서 저장·배포하는 것이다. 그리고 이미지 기반으로 다양한 환경에 동일 방식으로 배포되도록 CI/CD 도구(GitLab, Jenkins 등)을 활용한 자동화 배포 파이프라인을 구성해야 한다.



[그림1] 컨테이너 기반 환경(Docker)을 이용한 AI 모델 구동

그리고 학습한 AI 모델은 표준 포맷 형식으로 저장 및 관리해야 한다. AI 모델이 하나의 프레임워크에서 다른 프레임워크로 전환되거나 동시에 사용될 수 있도록 AI 모델을 표준화된 방식으로 저장하는 것이다. AI 모델 유형에 따라 머신러닝(PMML), 딥러닝(ONNX), 언어모델(GGUF 등)로 구분하여 해당 포맷으로 저장하는 것을 권장한다.



[그림2] AI 모델 표준 포맷 적용을 통한 배포

다음은 AI 모델 변경(교체·갱신)에 따른 상호호환성 확보 방안이다. AI 모델을 학습할 때 데이터 재사용을 위해 학습 데이터 구조 및 포맷을 표준화해야 한다. AI 모델에 활용되는 학습 및 테스트용 데이터 구조(특성명, 형식, 단위 등)와 레이블 정의 기준이 일관되어야 동일한 조건으로 학습할 수 있기 때문이다.

포맷	적용 범위	설명
JSONL (JSON Lines)	NLP, LLM, 멀티모달까지 재활용 목표	- 각 row가 독립 JSON 객체 - 텍스트 등 복합 label 구조 표현 용이 - OpenAI, HuggingFace 등 표준으로 사용
CSV (+ 스키마 분리)	그 외 일반 fine-tuning	- 범용적 파일 형식 - 모든 모델이 ingest 가능 - 전처리 코드만 맞추면 이식 용이

[표1] 데이터 재사용을 위한 권장 포맷

AI 모델과 정보시스템 간의 인터페이스를 표준화해야 한다. AI 모델의 구조와 규모 측면에서의 빠른 발전 속도를 고려하여, 추후 해당 모델의 버전 갱신을 대비한 입출력 인터페이스 설계가 필요하다. 또한 이전 모델과 갱신된 모델과의 출력 호환성을 비교하는 기능의 구현 및 대규모 변경이 필요한 경우 API 버전의 별도 유지·관리를 해야 한다.

마지막으로는 프로세스 재사용을 위해 전주기 메타데이터를 작성·관리해야 한다. 데이터 가공→모델 학습 등 파이프라인 구성 요소의 제공(데이터 전처리 코드, 모델 구조, config의 모델 파라미터 명시)과 학습 환경(사용 프레임워크, 라이브러리 버전 등) 요소들을 메타데이터로 작성·관리하는 것이다. 동일한 형식의 데이터와 메타데이터 기반의 학습 환경을 구축함으로써, 모델 교체·재학습 시에도 일관된 품질 유지 및 안정성 유지할 수 있기 때문이다.

Ⅲ. 결론

본 논문은 공공 부문에서 인공지능(AI) 시스템을 도입함에 있어 발생하는 기술 종속성과 시스템 확장 한계 문제를 해결하기 위한 방안으로, 상호호환성(Cross-Compatibility) 확보의 중요성을 제시하고, 이를 위한 설계 원칙과 기술적 구현 방안을 체계적으로 제안하였다.

AI 기술이 빠르게 변화하고 플랫폼 및 모델 환경이 다양화되는 현시점에서, 공공기관은 특정 기술에 종속되지 않으면서도 다양한 서비스 요구에 유연하게 대응할 수 있는 인프라를 확보해야 한다. 상호호환성은 이러한 기술 유연성과 자율성을 보장하는 핵심 요소로서, 시스템 구성 요소 간의 연계성과 재사용성을 향상시킨다.

본 연구를 통해 제시한 상호호환성 확보를 위한 설계 원칙(모듈화, 표준화, 플랫폼 독립성 등)과 기술적 구현 방법(컨테이너화, 개방형 모델 포맷, 서빙 프레임워크 통합 등)은 실제 공공 AI 시스템에 적용 가능한 실질적 가이드라인을 제공한다. 이를 통해 공공 부문은 공급자 종속 문제를 해소하고, 유지관리 비용을 절감하며, 기술적 자립성을 강화할 수 있다.

향후 연구에서는 상호호환성 요구사항의 법제화 및 표준화, 도메인별 적용 사례의 축적, 그리고 AI 윤리 및 신뢰성 검증 기준과의 통합 방안이 함께 논의될 필요가 있다. 특히, 국가 차원의 AI 공공 사업 표준 제정과 연계하여, 실무 현장에서 바로 활용할 수 있는 구체적 프레임워크와 체크리스트 도출이 요구된다.

본 연구는 향후 공공 AI 서비스의 지속 가능성과 확장성을 확보하고, 책임 있는 AI 인프라 구축을 위한 기반 마련에 기여할 수 있을 것으로 기대된다.

참 고 문 헌

[1] Matthias Leese,"AI and Interoperability in Public Governance," Research Collection ETH Zurich, 2024.

[2] European Commission, “Artificial Intelligence for Interoperability in the European Public Sector,” EU Publications, 2024.

[3] OECD & UNESCO, “G7 Toolkit for Artificial Intelligence in the Public Sector,” 2024.

[4] Straub, J. et al., “Artificial Intelligence in Government: Concepts, Standards, and a Unified Framework,” arXiv, 2022.

스마트 플랜트 가상환경 구축을 위한 모바일 - 서버 연동형 AI 기반 메타정보 자동 생성 프레임워크

조윤호, 김세현, 최인하, 김승엽*

㈜올포랜드 사업부문 SD사업본부 SD3그룹

yhcho@all4land.com, sehkim@all4land.com, cso8892@all4land.com, *gisksy@all4land.com

Mobile - Server Integrated AI Framework for Automated Metadata Generation in Smart Plant Virtual Environment Construction

Cho Yun Ho, Kim Se Hyeon, Choi In Ha, Kim Seung Yub*

Software&Development Group 3, AllforLand Co., Ltd.

요 약

스마트 플랜트의 디지털 전환 중요성이 증대됨에 따라 현장 중심의 가상환경 구축 자동화에 대한 수요가 증가하고 있다. 그러나 기존 방식은 전문가의 수작업에 의존하여 초기 구축에 과도한 비용과 시간이 소요될 뿐만 아니라 현장의 세부 운영 특성과 변화를 충분히 반영하기 어렵다는 한계를 지닌다. 이를 해결하기 위해 생성형 AI를 활용한 2D 기반 3D 자동 생성 연구가 활발히 진행되고 있으나 품질과 정확도 향상을 위해서는 원본 이미지로부터 객체의 형상·속성 등의 메타정보를 추출하는 전처리 단계가 필수적이다. 본 연구에서는 이러한 메타정보 생성 과정을 모바일 디바이스 기반 통합 프레임워크를 설계하여 현장 작업자가 모바일 기기를 통해 이미지 입력부터 설비 객체 자동 분할, 이미지 전처리, 설비 태그 ID 인식까지 하나의 연속된 프로세스로 수행할 수 있는 방안을 제시하였다. 서버 기반 추론 방식을 통해 기능 통합과 실효성을 검증하였으며 향후 AI 모델 경량화와 모바일 기반 추론 최적화를 통해 온디바이스 AI 기반 시스템으로 확장하는 연구를 진행할 예정이다.

Ⅰ. 서 론

스마트 플랜트는 ICT 기반의 실시간 공정 데이터 수집과 분석을 통해 설비 상태를 예측하고 최적화하는 디지털 트윈 기술의 핵심 응용 분야이다. 최근 스마트 플랜트의 디지털 전환 중요성이 증대되면서, 실제 설비와 환경을 가상공간에 구현하는 가상환경 구축 자동화에 대한 수요가 빠르게 증가하고 있다. 그러나 기존 가상환경 구축 방식은 도면 수집, 3D 스캐닝, 전문가의 수작업 등 복잡한 절차를 포함하여, 초기 구축에 과도한 비용과 시간이 소요될 뿐만 아니라, 현장의 세부 운영 특성과 변화를 충분히 반영하기 어렵다는 한계를 지닌다.

이러한 한계를 극복하기 위해 생성형 AI를 활용한 2D 기반 3D 자동 생성 연구가 활발히 진행되고 있으나 고품질 3D 모델을 생성하기 위해서는 원본 이미지로부터 객체의 형상·속성 등의 메타정보를 정밀하게 추출하는 전처리 단계가 필수적이다.[1] 특히 현장 환경에서는 비전문가가 모바일 기기를 통해 직관적으로 작업할 수 있는 자동화 프로세스의 필요성이 크다.

본 연구에서는 이러한 요구를 반영하여 메타정보 생성 기능의 통합 프레임워크를 제안한다. 제안 시스템은 사용자 입력 포인트 기반 객체 분할(Segment Anything Model, SAM)[2]을 통한 이미지 내 설비 객체 추출, 인페인팅(inpainting)[3]을 활용한 누락·빈 영역 전처리, 기울기 자동 보정(rectification)[4], OCR(Optical Character Recognition)[5] 기반 태그 인식 등 주요 기술을 하나의 연속된 프로세스로 통합하였다. 이러한 구성은 현장 작업자가 복잡한 편집 도구나 전문 소프트웨어 없이도 설비 객체의 정제된 이미지와 메타데이터를 신속하고 효율적으로 확보하여 이를 3D 모델 자동 생성에 활용할 수 있게 한다.

Ⅱ. 모바일 디바이스 기반 메타정보 자동 생성 프레임워크

제안한 프레임워크는 React Native 기반 모바일 앱과 FastAPI 서버를 연동하여 구현하였다. 실험 단계에서는 서버 기반 추론 방식을 사용하여 기능 통합과 실효성을 검증하였다. 프레임워크의 전체 구성 및 처리 절차는 그림 1과 같이 이미지 업로드부터 객체 분할, 전처리, 태그 인식까지의 단계를 하나의 연속된 프로세스로 구성하였다.

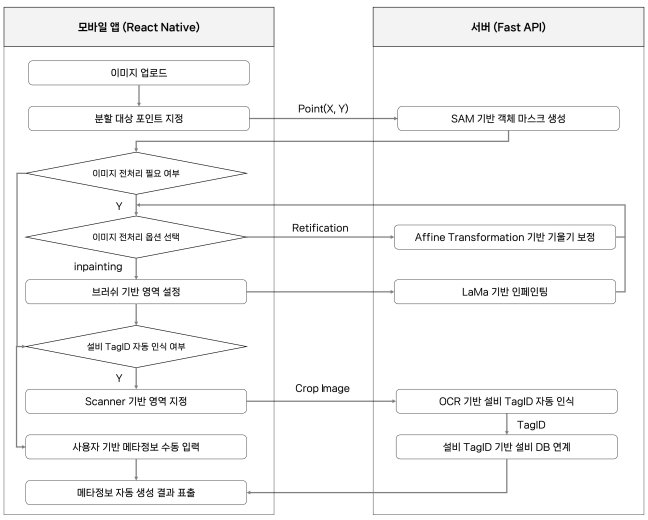
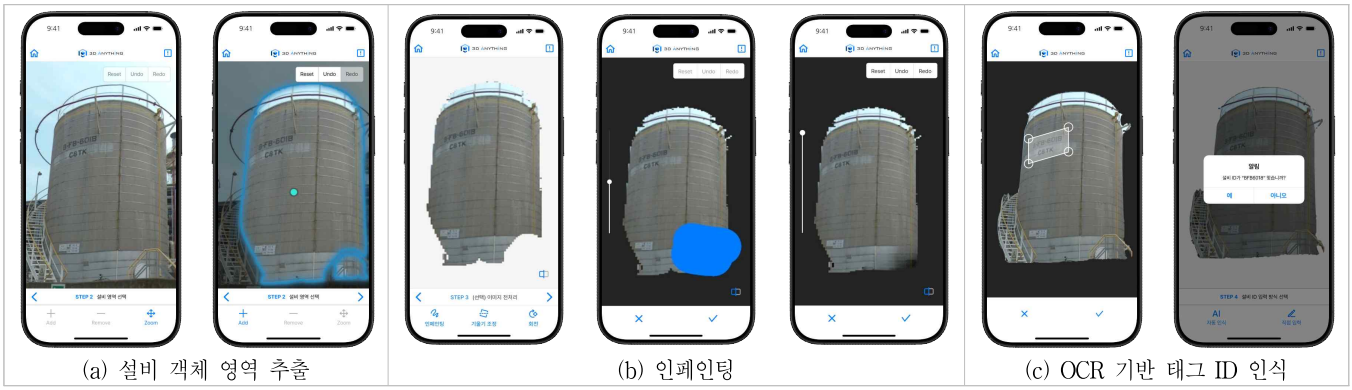


그림 1 모바일 디바이스 기반 메타데이터 자동 생성 프레임워크 흐름도



(a) 설비 객체 영역 추출 (b) 인페인팅 (c) OCR 기반 태그 ID 인식

그림 2 제안 프레임워크의 단계별 처리 결과

사용자는 앱에서 이미지를 업로드한 후 입력 기반 인터페이스로 분할할 설비 객체를 지정한다. 지정된 좌표는 서버로 전송되어 SAM을 통해 객체 마스크가 생성되고 LaMa 및 Rembg를 활용하여 배경이 제거된다. 이후 인페인팅 모듈을 이용해 객체 이미지 내의 누락되거나 빈 영역을 채워 시각적 완성도를 높인다. 설비가 기울어져 있는 경우에는 회전각을 분석해 자동으로 보정한다. 마지막으로 PaddleOCR로 설비 태그를 인식하며, 인식된 태그가 데이터베이스 항목과 일치하면 해당 설비의 메타데이터를 자동 조회·연계한다.

각 단계별 처리 결과는 그림 2에 제시하였다. (a) 객체 영역 추출 단계에서는 원본 이미지에서 관심 객체의 마스크를 생성하고, (b) 이미지 전처리 단계에서는 인페인팅을 통해 보정이 필요한 영역을 직접 지정하여 빈 영역을 자연스럽게 복원한다. (c) 설비 태그 인식 단계에서는 OCR로 인식된 태그 ID를 이미지에 표시하여 후속 3D 모델 생성 과정에서 활용 가능한 메타정보를 확보였다. 이를 검증하기 위해 본 연구에서는 Apple iPhone 16 Pro (RAM 8GB, iOS 18) 및 가상서버(NVIDIA GeForce RTX 3090 GPU) 환경에서 실험을 수행하였다.

실제 플랜트 설비 이미지를 대상으로 적용한 결과, 객체 분할·배경 제거·인페인팅·태그 인식이 모두 안정적으로 수행되었으며 설비의 기하학적 왜곡도 자동으로 보정되었다. 파이프라인의 각 단계별 처리 속도는 평균 약 1.5초, 최대 2초 이내로 측정되어 실시간 응용에 적합함을 확인하였다. 모든 과정이 짧은 시간 내에 완료되므로 사용자는 대기 시간에 대한 부담 없이 결과를 얻을 수 있으며, 비전문가도 단순히 몇 번의 클릭만으로 최종 결과를 획득 할 수 있어 실무 환경에서의 활용 가능성을 확인하였다

Ⅲ. 결론

본 연구에서는 현장 작업자가 모바일 기기를 활용하여 가상환경을 손쉽게 구축하기 위한 전처리 단계로 객체 분할, 배경 제거, 인페인팅, 태그 인식, 기하학적 왜곡 보정 기능을 통합한 모바일 - 서버 연동형 AI 프레임워크를 제안하였다.

배경이 포함된 실제 플랜트 제조공장 설비 이미지를 대상으로 실험한 결과 각 기능이 순차적으로 안정적으로 동작하였으며 직관적인 단계별 UI를 통해 비전문가도 짧은 시간 안에 결과를 획득할 수 있었다. 이를 통해 현장 중심의 디지털 트윈 구축 자동화 가능성을 확인하였다.

향후에는 SAM, LaMa, PaddleOCR 등 핵심 AI 모델의 경량화와 모바일 추론 최적화를 진행하여 서버 의존 없이 현장에서 즉시 처리 가능한 완전한 온디바이스 환경을 구현하는 것을 목표로 한다.

ACKNOWLEDGMENT

이 연구는 2024 년도 산업통상자원부 및 산업 기술 기획평가원(KEIT) 연구비 지원에 의한 연구임(2410001005, 플랜트 산업 가상 자율 제조를 위한 LOD 4 수준의 온디바이스 AI 기반 제조 현장 3차원 모델 자동 제작 기술개발 및 실증)

참 고 문 헌

[1] Lee, S. E., Chae, W. S., Yu, H.Y, Kim, N. W, and Kim, H. J. (2025). 플랜트 산업에서의 Image-to-3D 성능 향상을 위한 사전 이미지 처리 프레임워크. 한국컴퓨터 종합 학술대회 2025, 31-33.

[2] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.-Y., Dollár, P., & Girshick, R. (2023). Segment Anything. arXiv preprint arXiv:2304.02643. (https://doi.org/10.48550/arXiv.2304.02643)

[3] Suvorov, R., Logacheva, E., Mashikhin, A., Remizova, A., Ashukha, A., Silvestrov, A., Kong, N., Goka, H., Park, K., & Lempitsky, V. (2022). Resolution-robust Large Mask Inpainting with Fourier Convolutions. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) (pp. 2149 - 2159). https://github.com/advimman/lama

[4] OpenCV. Rotated Rectangle (cv.minAreaRect). https://docs.opencv.org/4.x/dd/d49/tutorial_py_contour_features.html, 2025.

[5] Cui, C., Sun, T., Lin, M., Gao, T., Zhang, Y., Liu, J., Wang, X., Zhang, Z., Zhou, C., Liu, H., Zhang, Y., Lv, W., Huang, K., Zhang, Y., Zhang, J., Zhang, J., Liu, Y., Yu, D., & Ma, Y. (2025). PaddleOCR 3.0 Technical Report. arXiv. (https://doi.org/10.48550/arXiv.2507.05595)

오류 누적 완화를 위한 의사결정 트랜스포머-흐름 매칭 통합 구조

김민준, 지창훈, 최요한, 눌란벡 키즈 아셀, 한연희†
한국기술교육대학교 컴퓨터공학과 미래융합전공

{june573166, koir5660, yoweif, aselbaekki, yhhan}@koreatech.ac.kr

A Flow Matching-Integrated Decision Transformer Architecture for Mitigating Compounding Errors

Min-Jun Kim, Chang-Hun Ji, Yo-Han Choi, Nurlanbek kyzy Asel, Youn-Hee Han†
Future Convergence Engineering, Dept. of Computer Science and Engineering, KOREATECH

요 약

강화학습 연구 분야의 한 갈래인 오프라인 강화학습(Offline Reinforcement Learning)은 오래전부터 연구되어온 일반적인 온라인 강화학습(Online Reinforcement Learning) 방법론이 실제 세계에서 직면하는 여러 실용적인 문제들을 해결한다. 본 연구는 오프라인 강화학습 알고리즘 중 하나인 의사결정 트랜스포머(Decision Transformer)와 흐름 매칭(Flow Matching)을 접목한 구조를 제안한다. 제안한 구조는 기존 의사결정 트랜스포머의 오류 누적(Compounding Error) 문제를 완화한다. 이를 검증하기 위해 간단한 환경에서 실험하여 순수 의사결정 트랜스포머보다 더 효과적으로 목표에 도달함을 확인하였다. 본 연구에서는 오류 누적 문제를 가진 알고리즘들의 개선 가능성을 제시하고, 오프라인 강화학습이 더 다양한 문제를 해결할 것으로 기대한다.

Ⅰ. 서 론

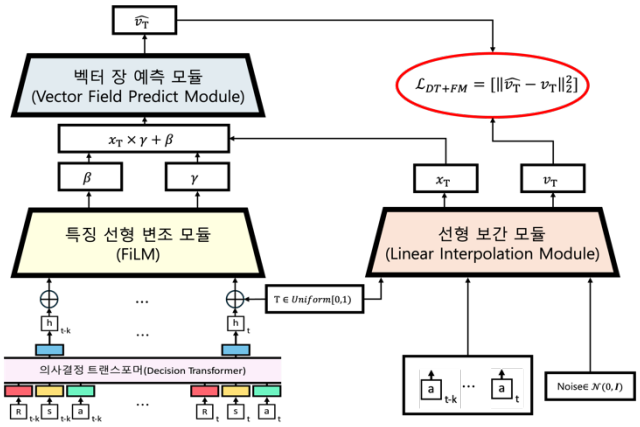
강화학습(Reinforcement Learning)은 에이전트가 환경과 상호작용하여 보상을 받으면서 누적 기대 보상을 최대화하도록 정책을 학습한다. 그런데 환경에 접근할 수 없거나 학습 데이터 수집 비용이 큰 경우, 오프라인 강화학습(Offline Reinforcement Learning)을 활용한다.

의사결정 트랜스포머(Decision Transformer, DT) [1]는 트랜스포머 구조를 기반으로 사전에 수집한 데이터를 활용해 학습하는 오프라인 강화학습 알고리즘이다. DT 모델은 시계열 데이터의 자기집중(Self-Attention) 기법을 통해 먼 과거의 행동과 현재 보상과의 관계를 이해할 수 있다. 하지만 DT 는 학습 과정에서 보지 못한 분포 밖(Out Of Distribution, OOD) 데이터 입력 이나 자기회귀적(Autoregressive) 추론 방식으로 인해 발생한 오류 누적(Compounding Error) 문제에 취약하다. 따라서 우리는 DT 의 오류 누적 문제를 완화할 수 있는 새로운 구조를 제안한다.

제안하는 구조는 DT에 흐름 매칭(Flow Matching, FM) [2]을 접목하여 초기 노이즈 데이터 분포에서 목표 데이터 분포로 향하는 벡터 장(Vector Field)를 학습한다. 오류가 발생하더라도, 학습된 벡터 장이 노이즈 분포를 목표 데이터 분포로 유도하기 때문에 안정적으로 수렴하며 결과적으로 오류 누적 문제가 완화된다. 이를 검증하기 위해 간단하고 제한적인 격자 (Grid) 환경에서 실험을 진행하였고, 오류 누적 문제를 완화함을 확인하였다.

Ⅱ. 의사결정 트랜스포머

DT 는 사전에 수집된 정적 데이터셋으로부터 정책을 학습한다. 데이터셋은 길이 k의 시계열 샘플로, 각 시점 t에서의 목표 누적 보상(Return-To-Go, RTG) R_t, 상태 s_t, 행동 a_t로 구성된다. 모델은 이러한 시계열 데이터셋



[그림 1] 의사결정 트랜스포머 + 흐름 매칭의 학습 구조

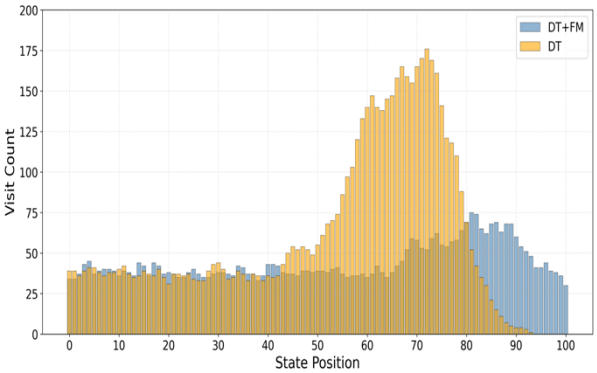
에 대해 주어진 RTG 를 만족하는 행동을 예측하도록 학습된다. 구체적으로, 모델이 예측한 행동 $\hat{a}_t$ 와 실제 데이터셋에 기록된 행동 a_t간의 차이를 최소화해 모델이 주어진 RTG 를 만족하는 행동을 정확히 예측하도록 학습한다.

그러나 DT 는 모방 학습(Imitation Learning)의 학습 방식을 따라, 학습 과정에서 보지 못한 OOD 데이터 입력에 대해 오류가 발생하며, 더 나아가 자기회귀적 구조의 특성상 초기에 발생한 이러한 오류가 다시 입력으로 반영되면서 오류가 누적된다.

Ⅲ. 흐름 매칭

FM 은 데이터 분포의 변화를 확률 흐름(Probability Flow)으로 표현하고, 확률 흐름을 흐름 매칭 시간에 대해 미분하여 그 변화를 기술하는 벡터 장을 직접 회귀 방식으로 학습한다. 구체적으로, 흐름 매칭 시간 T에 대

† 한연희(Youn-Hee Han, yhhan@koreatech.ac.kr): 교신저자



[그림 2] 격자(Grid) 환경 실험 방문 수 비교 그래프

해 노이즈 분포 x_0 에서 목표 분포 x_1 로 향하게 하는 연속적인 확률 흐름 x_T가 정의된다. 이때 목표 벡터 장 v_T 는 이 확률 흐름을 T 에 대해 미분하여 얻어지며, 학습 과정에서는 v_T와 모델이 예측한 벡터 장 $\hat{v}_T$ 간의 차이를 최소화함으로써 노이즈 분포에서 목표 분포로 수렴하도록 유도한다. 따라서 학습 과정에서 보지 못한 OOD 데이터가 입력되어 오류가 발생하더라도, 학습된 벡터 장이 이를 목표 분포로 유도함으로써 오류가 보정되고 결과적으로 오류 누적 문제가 완화된다.

IV. 제안한 구조

그림 1 은 제안한 DT 와 FM 을 결합한 학습 구조이다. 먼저 DT 는 입력으로 윈도우 샘플 k만큼의 RTG, 상태, 대응하는 행동을 받아들인다. 이들을 임베딩하여 인과적 마스킹(Causal Masking)을 적용한 DT 를 통해 k 개의 특징(Hidden State)들을 출력한다. 이 특징들에 FM 을 적용하기 위해 임베딩한 흐름 매칭 시간 T 를 각각 더한다. 이어서 흐름 매칭 시간 T 가 더해진 특징들을 특징 선형 변조 모듈(FiLM) [3]에 통과시켜 β 와 γ 를 구한다.

한편, 선형 보간 모듈(Linear Interpolation Module)의 입력으로 초기 분포인 가우시안 노이즈, 목표 분포인 목표 행동, 흐름 매칭 시간 T 를 주어 확률 흐름 x_T 와 목표 벡터 장 v_T를 출력한다. 마지막으로 앞서 구한, β와 γ 를 사용해 확률 흐름 x_T 를 선형 변환한 뒤, 벡터 장 예측 모듈(Vector Field Predict Module)의 입력으로 넣어 예측 벡터 장 $\hat{v}_T$ 를 출력한다. 모든 모델의 학습은 아래의 손실 함수 $\mathcal{L}$ 을 최소화하는 방향으로 종단간(End-to-End) 진행된다.

$$\mathcal{L}_{DT+FM} = [\|\hat{v}_T - v_T\|_2^2] \quad (1)$$

학습 후 추론 시에는 목표 RTG R_0 와 초기 상태 s_0 를 DT 의 입력으로 주어 특징 h 를 출력한다. 이를 윈도우 샘플 k 만큼 생성한 $h_{0\sim k-1}$ 중 마지막 특징 h_{k-1} 와 노이즈를 디노이징 단계를 N번 거쳐 노이즈 없는 새로운 행동 a_{k-1}을 생성한다.

V. 실험

실험은 0 에서 시작해 100 에 도달하는 1 차원 격자 환경에서 진행되었으며, 에이전트는 매 스텝마다 왼쪽(-1) 또는 오른쪽(+1)로 이동할 수 있다. 에이전트가 계속 오른쪽 행동을 선택하여 100 에 방문하면 성공 처리하고 보상 1 을 받는다. 이외의 상태에는 보상이 주어지지 않는다. 100에 끝내 방문하지 못하고 200스텝 이상 지나면 타임 아웃하여 실패 처리한다. 학습에 사용한 정적 데이

<표 1> 격자(Grid) 환경 실험 결과 비교

Model	Success rate (%)	Most visited states (number of visits to that state)
DT	0.0	72 (176)
DT+FM	100.0	81 (75)

터셋은 에이전트가 항상 오른쪽으로 움직였을 때의 상태, 행동, RTG 이다.

추론 시에는 의도적으로 OOD 데이터 입력에 의한 오류 누적을 발생시키기 위해 새로운 환경을 설계하였다. 구체적으로 새로운 환경은 10% 확률로 추출한 행동과 무관하게 에이전트는 왼쪽(-1)으로 이동한다. 실험은 총 30 에피소드를 실행하였다. 표 1 은 이들의 성공률과 가장 많이 방문한 상태 및 해당 상태의 방문 횟수를 비교한 결과이다.

표 1 의 결과는 DT+FM 이 100%로 성공한 반면, DT 는 한 번도 성공하지 못한 것을 보여준다. 이는 DT 가 가진 OOD 데이터에 대한 오류 누적 문제를 보여주고, DT+FM 은 이를 효과적으로 해결했음을 나타낸다. 또한 가장 많이 방문한 상태를 비교하면, DT 는 72 상태를 176 번으로 가장 많이 방문한 반면, DT+FM 은 81 상태에서 75 번 방문하였다. 이는 DT 의 최대 방문 횟수가 DT+FM 의 2.3 배이고, 그 당시의 상태 또한, DT 가 더 빨리 나타나 오류 누적에 취약한 것을 보여준다.

그림 2 의 결과는 표 1 의 결과를 도식화한 것으로, 각 그래프의 100 상태와 피크 값을 통해 표 1 의 결과를 확인할 수 있다. 특히 주목할 점은 그래프가 상승하기 시작하는 지점으로, DT 는 43 상태부터 상승하기 시작해 오류 누적 현상이 나타나는 반면, DT+FM 은 64 상태부터 상승하여 오류 발생이 지연됨을 보여준다.

VI. 결론

본 연구에서는 DT 의 오류 누적 문제를 완화할 수 있는 FM 결합 구조를 제안하고, 격자 환경 실험 결과를 통해 제안한 구조가 오류 누적 문제를 완화함을 보였다. 더 나아가 본 연구는 오류 누적 문제를 가진 기존 알고리즘들의 개선 가능성을 제시하며, 오프라인 강화학습이 보다 다양한 실제 문제 상황에 적용될 수 있는 잠재력을 보여준다.

ACKNOWLEDGMENT

이 논문의 연구는 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행되었음

(No.NRF-2023R1A2C1003143&No.NRF-2018R1A6A1A03025526).

참 고 문 헌

[1] Chen, Lili, et al. "Decision transformer: Reinforcement learning via sequence modeling." Advances in neural information processing systems 34 (2021): 15084-15097.

[2] Lipman, Yaron, et al. "Flow matching for generative modeling." arXiv preprint arXiv:2210.02747 (2022).

[3] Perez, Ethan, et al. "Film: Visual reasoning with a general conditioning layer." Proceedings of the AAAI conference on artificial intelligence. Vol. 32. No. 1. 2018.

자산가격 모델링에서 효율적 트랜스포머 구조와 데이터 정규화에 대한 연구

김규진, 김동주*
고려대학교 SW·AI 융합대학원, *고려대학교 정보대학 뇌공학과
gklb@korea.ac.kr, *dongjookim@korea.ac.kr

Efficient Transformer Structures and Data Normalization on Asset Pricing Modeling

Kim Kyou Jin, Kim Dong Joo*
Graduate School of SW·AI Convergence, Korea University,
*Dept. of Brain and Cognitive Engineering, Korea University.

요 약

본 연구는 기존 자산가격(Asset Pricing) 연구에서 금융 데이터의 비선형성과 장기 의존성 문제를 해결하기 위해 다양한 Transformer 시계열 모델의 예측 성능을 분석하고, 금융 데이터 정규화가 미치는 영향을 연구하였다. 실험 결과, 인코더 구조와 Patch-Wise 방법을 사용하는 Transformer 모델이 다른 모델 대비 우수한 성능을 보였다. 이는 Transformer 모델이 전체 패턴보다 국소 패턴을 감지하는 것에 효과적임을 시사한다. 또한, 정규화된 데이터를 사용할 때보다 정규화되지 않은 데이터가 더 유효한 예측 성능을 기록하였다. 금융 데이터에서는 데이터의 비정규성이 중요한 정보가치를 지니며 정규화 과정이 되려 정보 가치를 제거할 수 있다. 본 연구는 자산가격 연구에서 Transformer 모델의 잠재력을 입증하고, 데이터 전처리에 대한 새로운 관점을 제시함으로써 향후 연구 방향에 기여할 것으로 기대된다.

Ⅰ. 서 론

금융 시장에서는 투자자의 효용 극대화를 위한 Asset Allocation 연구가 필수적 과제이며, 변동성(Volatility)과 자산가격(Asset Pricing), 2 변수가 주요 연구 대상이다[1]. 그러나 금융 데이터의 극심한 비정규성, 장기 의존성으로 인하여 기존 자산가격 모델 연구에서는 개별 종목의 재무 데이터 대신 팩터 노출도를 사용하는 접근법을 채택하여, 데이터의 비정규성을 제거하는 방법을 선택하였다[2]. 최근에는 시계열 처리에 특화된 Transformer 모델을 통해 금융 데이터의 장기 의존성 문제 해결 가능성을 제시하고 있다[3]. 본 연구는 자산가격 모델링 분야에서 다양한 시계열 Transformer 구조의 성능을 비교하고 효율적인 Transformer 구조를 제안한다. 둘째로 Transformer 구조 하에서는 금융 데이터 정규화에 대한 개념 전환이 필요함을 대해 논의한다.

Ⅱ. 자산가격 모델링에 Transformer Model 적용

Transformer 는 Self-Attention 메커니즘을 통해 데이터의 시점별 관계를 학습하며, 시계열의 장기적인 패턴을 포착하는 것에 우수하다. 본 연구에서는 Transformer 기반 시계열 모델들의 금융 시계열 예측 성능을 비교 분석한다. 예측 성능 비교 분석에 사용된 시계열 Transformer 모델은 Vanila Transformer, Informer, Autoformer, PatchTST, iTransformer 이다.

예측 목표는 종목별 52 주간 주가 데이터를 기준으로 향후 13 주 주가 데이터를 예측하는 것이다. 종목은 미국(Russell 3000)과 유럽(Stoxx 600), 일본(TOPIX 500), 한국(KRX 300)의 2001 년 12 월에서 2025 년 6 월까지 주간 데이터를 사용했다. 기존 자산가격 연구에서 주로 사용되던 정규화 데이터인 종목별 팩터 노출도는 BM, SMB, UMD, CMA, RMW, 6 가지 변수이다[4]. 대조군으로는 팩터 노출도 계산에 사용되는 비정규화 데이터 Value, Size, Momentum, Investment, Profitability, 5 가지 변수를 사용하였다.

Model	Structure	Methodology	Features
Vanilla Transformer [5]	Enc-Dec	Self-Attention	Temporal relationships
Informer [6]	Enc-Dec	ProbSparse, Point-Wise	Important temporal relationships
Autoformer [7]	Enc-Dec	Auto-correlation, Patch-Wise	Seasonality, local patterns
PatchTST[8]	Encoder	Patch-Wise	Local pattern
iTransformer [9]	Encoder	Variate-Wise	Inter-variable relationships

표 1. 연구에 사용된 Transformer 모델 방법론 비교

III. 모델 구조의 영향

Autoformer 와 같은 Patch-Wise 모델이 Vanila 모델 대비 성능이 개선되었으며, 패치 기반 모델 중에서는 인코더 구조가 가장 우수한 성과를 시현하였다. 결론적으로 Encoder 를 통한 계산 효율성 극대화가 중요하였으며, 전체 시계열 패턴보다는 Patch-Wise 구조를 통한 국소 패턴 포착이 금융 시계열 데이터 분석에 가장 효과적임을 관측하였다.

Data	Model	MAE	MSE	RMSE	MAPE
US	Vanila Transformer	1.46	2.98	1.73	3.53
	Informer	1.17	2.03	1.43	10.61
	Autoformer	0.58	0.56	0.75	74.27
	iTransformer	0.27	0.21	0.46	27.77
	Patch-TST	**0.24**	**0.18**	**0.43**	**61.75**
Europe	Vanila Transformer	1.58	3.33	1.82	1.25
	Informer	1.14	1.86	1.36	1.06
	Autoformer	0.57	0.57	0.75	2.21
	iTransformer	0.24	0.19	0.43	0.94
	Patch-TST	**0.22**	**0.18**	**0.43**	**1.22**
Japan	Vanila Transformer	1.17	1.96	1.40	0.82
	Informer	0.65	0.75	0.86	0.44
	Autoformer	1.13	1.98	1.41	1.31
	iTransformer	0.31	0.23	0.48	0.27
	Patch-TST	**0.25**	**0.16**	**0.40**	**0.23**
Korea	Vanila Transformer	1.53	3.25	1.80	3.89
	Informer	1.19	2.08	1.44	3.67
	Autoformer	1.11	1.85	1.36	46.30
	iTransformer	0.30	0.24	0.49	13.14
	Patch-TST	**0.26**	**0.21**	**0.46**	**4.77**

표 2. 국가별 Transformer 구조별 예측 성능

IV. 데이터 정규화의 영향

전통적인 자산가격 이론에서 팩터 노출도가 개별 종목의 고유 위험을 설명하는 효과적인 수단으로 간주되었던 것과는 달리, 원본 재무 데이터를 사용했을 때 성능이 더 우수하게 나타났다. 또한 정규화 데이터보다는 비정규화 데이터가 더 나은 예측 성능을 기록하였다. 즉, 금융 데이터의 비정규성(주가의 급등락과 같은 극단값) 자체가 중요한 정보가치를 지니고 있으며, 정규화를 통해 소실될 수 있음을 시사한다.

V. 결론

본 연구는 금융 데이터 분석에 있어서 Transformer 모델 중 패치 기반 인코더 모델들이 예측 성능을 크게 향상시켰음을 확인했다. 이는 시계열 예측에 있어 모든 시점의 관계를 학습하는 것보다, 국소적인 패턴(local patterns)을 포착하고 이를 기반으로 전체적인 예측을 수행하는 것이 더 효과적임을 시사한다. 또한 비정규화된 데이터의 예측 성능 상승 효과를 통해 금융데이터의 노이즈 자체가 정보 가치를 지니고 있으며, Transformer 모델과 같은 강력한 비선형 모델이 이를 해석하는 효과적인 도구가 될 수 있음을 확인하였다.

Data	MAE			MSE		
	Financial Data	Factor Exposure	Mixed	Financial Data	Factor Exposure	Mixed
US	0.239	0.244	**0.235**	0.183	0.184	**0.180**
Europe	**0.221**	0.224	0.222	**0.172**	0.183	0.173
Japan	**0.215**	0.254	0.259	**0.122**	0.157	0.163
Korea	0.272	0.265	**0.258**	0.214	0.209	**0.201**

표 3. 사용된 데이터별 예측 성능(PatchTST 기준)

Data	Not Normalized				Normalized			
	MAE	MSE	RMSE	MAPE	MAE	MSE	RMSE	MAPE
US	**0.24**	**0.18**	**0.43**	**0.70**	0.29	0.23	0.48	6.40
Europe	**0.22**	**0.17**	**0.42**	**0.24**	0.26	0.20	0.44	0.71
Japan	**0.22**	**0.12**	**0.35**	**0.21**	0.25	0.18	0.43	0.23
Korea	0.27	0.21	0.46	5.31	**0.25**	**0.15**	**0.39**	**0.22**

표 4. Standard Scaling 성능 개선 효과(PatchTST 기준)

ACKNOWLEDGMENT

본 논문은 고려대학교 정보대학 뇌공학과 김동주 교수님의 지원 하에 작성되었습니다.

참 고 문 헌

[1] Markowitz, Harry M. “Portfolio Selection: Efficient Diversification of Investments”. Yale University Press, 1959.

[2] Fama and French. “The Capital Asset Pricing Model: Theory and Evidence.” SSRN Electronic Journal, Jan. 2003

[3] Wang, Yuxuan, et al. “Deep Time Series Models: A Comprehensive Survey and Benchmark.” arXiv.org, 18 July 2024

[4] Fama, Eugene F., and Kenneth R. French. “A Five-factor Asset Pricing Model.” *Journal of Financial Economics*, vol. 116, no. 1, Oct. 2014, pp. 1- 22.

[5] Vaswani, Ashish, et al. “Attention Is All You Need.” *arXiv.org*, 12 June 2017

[6] Zhou, Haoyi, et al. “Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting.” *arXiv.org*, 14 Dec. 2020

[7] Wu, Haixu, et al. “Autoformer: Decomposition Transformers With Auto-Correlation for Long-Term Series Forecasting.” *arXiv.org*, 24 June 2021

[8] Nie, Yuqi, et al. “A Time Series Is Worth 64 Words: Long-term Forecasting With Transformers.” *arXiv.org*, 27 Nov. 2022

[9] Liu, Yong, et al. “iTransformer: Inverted Transformers Are Effective for Time Series Forecasting.” *arXiv.org*, 10 Oct. 2023

실시간 IPTV 방송스트림 기반 멀티모달 상품정보 추론 및 스코어링 시스템 개발

배주한, 허재호, 주현철, 나태영
SK 텔레콤

{juhan.bae, jaeho.hur, hc.joo, taeyoung.na}@sk.com

Multi-Modal Product Information Inference and Scoring System Based on IPTV Real-Time Broadcast Streams

Juhan Bae, Jaeho Hur, Hyunchul Joo, Taeyoung Na
SK Telecom

요 약

본 논문은 IPTV 실시간 방송 스트림을 분석하여 상품 정보를 추론하고 신뢰도 점수 기반으로 스코어링하는 멀티모달 시스템을 제안한다. 영상, 자막, 화면 내 텍스트를 통합 분석해 상품을 실시간 인식하고, 메타데이터를 구조화하여 전자상거래 플랫폼과 연동한다. 이를 통해 시청자는 방송 중 등장한 상품을 즉시 탐색·구매할 수 있으며, 본 연구는 스마트 TV 기반 실시간 커머스의 새로운 가능성을 제시한다.

Ⅰ. 서 론

스트리밍 미디어 소비가 증가하면서 방송 콘텐츠와 상업적 정보의 융합을 통한 새로운 소비자 경험이 주목받고 있다. IPTV 실시간 방송에서는 시청 중 상품 정보를 즉시 탐지하고 제공하는 기술 수요가 급증하며, 이는 인터랙티브 커머스(T-Commerce)로 발전하고 있다.

기존 광고 기반 커머스 모델은 사용자의 수동적 반응에 의존하거나 사전 정의된 상품 정보를 전달하는 한계를 가진다. Google YouTube 는 콘텐츠 제작자의 사전 등록이 필요하여 VoD 콘텐츠에 한정되고, Amazon Live 는 홈쇼핑 형태의 라이브 스트리밍 서비스를 제공하며, Facebook/Instagram Shopping 은 SNS 기반 라이브 방송 중 즉시 구매 기능을 구현하고 있다. 국내 서비스의 경우, SK Broadband 는 VoD 서비스 내 수작업으로 등록된 PPL 상품 정보 기반의 구매 QR 코드를 제공하고 있으나 실시간 방송을 지원하지 않는다. 네이버 쇼핑라이브는 스마트스토어 입점 업체 대상 라이브 커머스 툴을 지원하여 실시간 방송을 통한 상품 소개 및 즉시 구매 연계 서비스를 제공하고 있다.

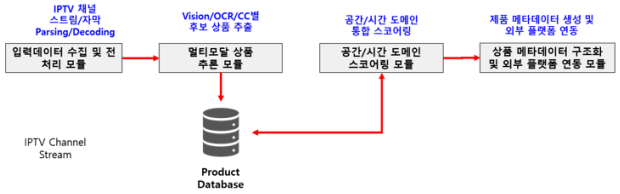
그러나 상기 서비스들은 사전 정의 의존성, 실시간 방송 적용 제한, 단일 모달리티 분석이라는 공통적인 한계를 지닌다. 기 구축된 상품 정보에 의존하여 새로운 상품의 실시간 반영이 어렵고, 대부분 VoD 콘텐츠 대상이며 실시간 연동이 불가능하다. 또한 영상, 자막, 화면 내 텍스트 등 다양한 소스의 통합 분석 기능이 제한적이다.

본 연구는 방송 중 등장하는 상품을 실시간 인식하여 즉시 추천하고 구매 연계하는 상호작용 기반 커머스 시스템을 구현하고자 한다. 관련 특허 기술로는 라이브 영상 스트림과 쇼핑 인터페이스 연동[1], 실시간 방송영상과 부가정보 동기화[2], IPTV 기반 인터랙티브 쇼핑[3] 등이 있으나, 멀티모달 데이터의 통합 분석과 실시간 상품 추론에는 한계가 있다.

Ⅱ. 본론

1. 제안하는 시스템 구조

본 논문에서 제안하는 시스템은 입력 데이터 수집 및 전처리 모듈, 멀티모달 상품 추론 모듈, 시간도메인 스코어링 모듈, 상품 메타데이터 구조화 및 외부 플랫폼 연동 모듈의 4 개 핵심 모듈로 구성된다.



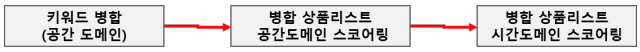
<그림 1. 전체 시스템 구조도>

입력 데이터 수집 및 전처리 모듈은 IPTV 방송 스트림으로부터 실시간 콘텐츠 데이터를 수집하고 멀티모달 분석을 위한 입력 형태로 전처리를 수행한다. 각 모달 데이터는 시간 동기화를 기준으로 병렬 수집되며, 통합 분석을 위해 멀티모달 상품 추론 모듈로 전달된다. 영상 프레임은 IPTV 스트림 파싱 및 디코딩 후 일정 주기로 키 프레임을 추출하며, 자막 정보는 IPTV 방송 스트림에서 제공되는 자막 데이터를 실시간으로 추출한다. 화면 내 텍스트는 방송 화면의 그래픽 문자, 오버레이 텍스트 자막 외 시각적 문자를 OCR(Optical Character Recognition) 기술을 활용하여 추출한다.

멀티모달 상품 추론 모듈은 전처리된 멀티모달 데이터를 입력으로 하여 방송 콘텐츠에 노출된 상품을 실시간 인식하고 상품 속성 정보를 추론한다. Vision 추론에서는 이미지 내 상품 탐지 및 상품 키워드, 관련 상품 크롭 영역을 추출하고, Caption 추론에서는 자막에서 텍스트 기반 상품 키워드를 추출한다. OCR 추론에서는 화면 내 텍스트 분석을 통한 텍스트 기반 상품 키워드를 추출한다. 최종 추론된 상품 정보는 스코어링 모듈로 전달된다.

2. 공간/시간 스코어링 알고리즘

시간공간 스코어링 모듈은 모달리티별 상품 추론 결과에 대해 통합 분석된 추천 상품 리스트 및 상품별 스코어를 산출한다. Vision, OCR, Caption 추론 결과인 상품 키워드를 대상으로 현재 시점 기준 최근에 집중적으로 여러 모달리티에서 반복해서 검출된 상품 키워드에 높은 스코어를 할당하며 아래와 같은 순서로 진행된다.



<그림 2. 시간공간 스코어링>

(1) 키워드 병합

공간도메인에서 상품 키워드 병합 과정은 다음과 같이 수행된다. 초기 모달리티별 기본 스코어($Score_{base}^{vision}$: $Score_{base}^{ocr}$: $Score_{base}^{cc}$)는 공간 도메인 상 각 모달리티별 중요도를 의미하며, 사용자 편의 및 콘텐츠 특징에 따라 변경 가능하며 시점 t 에서 모달리티별 추출 상품의 집합은 아래와 같이 정의한다.

$$ITEM_{method}(t) = \{item_{method}^{1}(t), item_{vision}^{2}(t), \dots, item_{method}^{max_method}(t)\}$$

시점 t 기준 공간도메인 중복상품 제거를 위해 병합 리스트($ITEM_{merge}(t)$)를 만들며 모달리티 간 상품 키워드들의 텍스트 인코딩된 값의 코사인 유사도(Cosine Similarity) 값이 임계치(Th) 이상일 경우 동일 상품 키워드라 간주한다. 또한, 다수 모달리티에서 추출된 상품에 높은 가중치를 주기 위하여 모달리티 카운트 ($modality_count_l(t)$)를 계산하며 아래와 같이 표현된다.

$$modality_count_l(t) = \sum_{\forall i} \Delta_{vision}^{l,i}(t) + \sum_{\forall j} \Delta_{ocr}^{l,j}(t) + \sum_{\forall k} \Delta_{cc}^{l,k}(t)$$

$$\Delta_{method}^{l,m}(t) = \begin{cases} 1 & ,if\ sim(E^T(item_{merge}^{l}), E^T(item_{method\in\{Vision,OCR,CC\}}^{m\in\{i,j,k\}})) > Th \\ 0 & ,otherrwise \end{cases}$$

l = Index of merged product list
i = Index of product list from vision
j = Index of product list from OCR
k = Index of product list from CC
E^T = Text Encoder

(2) 병합상품리스트 공간도메인 스코어링

$modality_score_l(t)$는 시점 t 기준으로 병합리스트의 상품 l이 다수의 모달리티에서 많이 검출되면 될수록 높아지는 값으로 모달리티별 기본점수($Score_{base}^{method}$)에 모달리티별 검출유무($\sum_{\forall i} \Delta_{method}^{l,i}(t)$)와 모달리기 검출 가중치 ($W_{spatial}^{modality_count_l(t)}$) 의 곱으로 표현된다.

$$modality_score_l(t) = max\left(Score_{base}^{vision} \times \sum_{\forall i} \Delta_{vision}^{l,i}(t), Score_{base}^{ocr} \times \sum_{\forall j} \Delta_{ocr}^{l,j}(t), Score_{base}^{cc} \times \sum_{\forall k} \Delta_{cc}^{l,k}(t) \right) \times W_{spatial}^{modality_count_l(t)}$$

$$W_{spatial}^{modality_count_l(t)} \begin{cases} W_{spatial}^{1} & if, modality_count_l(t) = 1 \\ W_{spatial}^{2} & if, modality_count_l(t) = 2 \\ W_{spatial}^{3} & if, modality_count_l(t) = 3 \end{cases}$$

(3) 병합상품리스트 시간도메인 스코어링

$W_{temporal}^{x}$는 시점(t)기준으로 이전 M-1 개의 프레임을 검색범위, 검색범위내 프레임 위치를 x라 가정할때 현재시점에 가까울수록 1 에 가까운 값을 가지며 아래와 같이 표현된다.

$$W_{temporal}^{x} = W_{temporal}^{base} + \left((1 - W_{temporal}^{base}) \times \frac{x}{M}\right)$$

따라서 t 시점 기준 병합상품 l 의 최종스코어 $(final_score_l(t))$는 아래와 같이 표현된다.

$$final_{score_l}(t = M) = \sum_{1 \le x \le M-1} \left(W_{temporal}^{x} \times modality_score_l(x) \times \sum_{\forall j} \Delta_{temporal}^{l,j}(x) \right) + W_{temporal}^{M} \times modality_score_l(M)$$

3. 상품 메타데이터 구조화 및 외부 플랫폼 연동

상품 메타데이터 구조화 및 전자상거래 플랫폼 연계 모듈은 스코어링 모듈에서 선출된 각 상품의 메타데이터를 정형화하고 외부 전자상거래 플랫폼과 연계하여 사용자가 실시간으로 구매 가능한 형태로 제공한다. 사용자 경험의 즉시성과 상호작용성을 극대화하여 실시간 커머스 환경을 구현한다.

Ⅲ. 결론

본 논문은 IPTV 실시간 방송 스트림 기반의 멀티모달 상품 정보 추론 및 스코어링 시스템을 제안하였다. 제안 시스템은 영상, 자막, 화면 내 텍스트의 통합 분석을 통해 방송 콘텐츠에 노출되는 상품을 실시간으로 인식하고, 신뢰도 점수 기반의 우선순위 추천정보를 제공하여 전자상거래 구매 시스템에 활용할 수 있도록 한다.

ACKNOWLEDGMENT

본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2024 년도 문화체육관광연구개발사업으로 수행되었음 (과제명: 전통예술 가무악의 융복합 공연제작 활성화를 위한 융복합 공연 기획/제작 플랫폼 기술 개발, 과제번호: RS-2024-00398536, 기여율: 100%)

참 고 문 헌

[1] "Live video streaming and shopping interface integration system," US Patent US10491958B2, 2019.

[2] "실시간 방송영상과 부가정보의 동기화 시스템," 한국특허 KR102123593B1, 2020.

[3] "IPTV 기반 인터랙티브 영상 쇼핑 기술," 한국특허 KR101124560B1, 2012.

문맥 인지 적응형 마스크 선택을 통한 일관된 참조 비디오 객체 분할

서준혁, 윤영우, 최원준, 이동규*
경북대학교
ssam2s@knu.ac.kr, dbsduddn@knu.ac.kr, sangju@knu.ac.kr, dglee@knu.ac.kr

Context-Aware Adaptive Mask Selection for Consistent Referring Video Object Segmentation

Jun-Hyeok Seo, Young-Woo Youn, Wonjun Choi, Dong-Gyu Lee*
Kyungpook National University

요 약

본 논문에서는 참조 비디오 객체 분할에서 프레임 간 시간적 일관성과 객체 수 정보에 기반한 문맥 인지 적응형 마스크 선택 기법을 제안한다. 기존의 참조 비디오 객체 분할 작업은 연속 프레임에서 시간적 일관성을 유지하는데 어려움을 겪는다. 이를 해결하기 위해 본 논문에서는 CLIP 기반 키프레임 선정, 마스크 일관성 점수 계산, 텍스트 기반 객체 수 판별, 그리고 시간적 마스크 보간으로 구성된 새로운 프레임워크를 제안한다. 제안하는 프레임워크는 벤치마크 데이터셋을 이용한 성능 비교에서 우수한 성능을 보임으로써 제안한 프레임워크가 참조 비디오 객체 분할 작업에 효과적임을 보여준다.

Ⅰ. 서 론

최근 딥러닝 기반 영상 이해 분야에서는 자연어로 지칭된 특정 객체를 영상 내에서 정확히 분할하는 참조 비디오 객체 분할(Referring Video Object Segmentation, RVOS)이 활발히 연구되고 있다[1, 2]. RVOS는 비디오 시퀀스와 함께 주어진 텍스트 설명을 기반으로, 해당 설명에 부합하는 객체의 픽셀 단위 마스크를 예측하는 문제로 정의된다. 그러나 기존 RVOS 연구들은 장면 내 시각적 폐색, 객체의 형태 변화, 배경 변화뿐 아니라, 프레임 간 시간적 일관성을 유지하는데 어려움을 겪는다. 특히, 프레임별로 독립적으로 예측된 마스크는 인접 프레임 간 경계 위치나 형태가 불안정해지는 문제가 발생하며, 이는 전체 비디오에서의 객체 추적 품질을 저하한다. 또한, 텍스트로 구성된 하나의 참조 표현이 복수 객체를 지칭하는 경우, 단일 객체 기반의 분할 방법은 정확도가 저하되는 한계를 가진다. 최근에는 대규모 비전-언어 모델(LVLM)과 사전학습된 범용 분할 모델을 결합하여 이러한 문제를 완화하려는 시도가 제안되었다. 예를 들어, Sa2VA[3]나 HyperSeg[4]와 같은 최신 프레임워크는 시각-언어 정렬을 통해 텍스트와 영상 간 의미적 대응을 강화하고, 범용 마스크 예측을 가능하게 한다. 그러나 이러한 방법들 역시 모델별 특성 차이로 인해 동일 장면에서도 결과 품질이 상이하며, 이를 효과적으로 결합하고 선택하는 방법은 충분히 연구되지 않았다.

이에 본 연구에서는 문맥 인지 적응형 마스크 선택 방법을 기반으로 한 RVOS 방법을 제안한다. 제안 기법은 다음과 같은 세 가지 핵심 요소로 구성된다. (1) CLIP[5] 유사도 기반 키프레임 선택과 양방향 마스크 비교를 통해 가장 일관성이 높은 마스크를 선택하는 일관성 인지형 마스크 선택, (2) 대규모 언어 모델을 활용해 참조 표현의 단·복수를 판별하고 객체 수에 따라 마스크를 조합하는 텍스트 기반 객체 수 판별, (3) 인접 프레임 정보를 활용하여 비어 있는 예측을 보완하는 시간적 마스크 보간이다.

본 논문에서는 대표적인 벤치마크 데이터셋에 대한 기존 연구들과의 객관적인 성능 비교를 진행하고, 제안하는 핵심 요소에 대한 절제 실험을 통해 제안하는 방법이 효과적임을 증명한다.

Ⅱ. 본론

본 논문에서는 문맥 인지 적응형 마스크 선택 방법 기반의 RVOS 프레임워크를 제안한다. 이는 주어진 영상과 참조 표현에 대하여, 시간적 일관성과 객체 수 정보를 고려하여 최적의 분할 마스크를 생성하는 것을 목표로 한다. 이를 위하여 전체 파이프라인은 (1) 일관성 인지형 마스크 선택, (2) 텍스트 기반 객체 수 판별, (3) 시간적 마스크 보간의 세 단계로 구성된다. 제안하는 프레임워크는 그림 1과 같이 구성되며, 비디오 시퀀스 $V=\{f_1, f_2, ..., f_T\}$와 참조 표현 E를 입력으로 받아, 최종 마스크 시퀀스 $\hat{M}=\{\hat{m}_1, \hat{m}_2, ..., \hat{m}_T\}$를 산출한다.

RVOS 문제에서 프레임별로 독립적인 마스크 예측은 인접 프레임 간 경계 불일치와 형태 변형을 초래할 수 있다. 본 연구에서는 이러한 문제를 완화하기 위하여 최첨단 모델들의 결과를 앙상블하여 시간적 안정성이 가장 높은 마스크를 선택한다. 우선, CLIP 기반 텍스트-이미지 유사도 함수를 정의하여 각 프레임 f_t에 대한 CLIP 임베딩과 주어진 참조 표현 E와의 코사인 유사도를 산출한다. 여기서 가장 높은 유사도를 갖는 프레임을 키프레임 f_k로 설정한다. 키프레임의 계산 과정은 아래 수식 (1)과 같다.

$$s_t = \frac{\phi_{img}(f_t) \cdot \phi_{text}(E)}{\| \phi_{img}(f_t) \| \, \| \phi_{text}(E) \|} \quad (1)$$

여기서 ϕ_{img}와 ϕ_{text}는 각각 CLIP의 이미지와 텍스트 인코더를 의미한다. 이후, 모델 A와 B가 각각 예측한 마스크의 픽셀 값이 1인 위치의 평균 좌표를 산출하여 중심점을 계산하고, 이전에 선택된 프레임과 현재 프레임의 중심점의 유클리드 거리를 측정한다. 이전 프레임의 중심점을 c_t라고 할 때, 중심점 거리는 아래 수식 (2)로 계산되며, 거리가 더 작은 마스크가 채택된다.

$$d_t = \| c_{t+1} - c_t \|_2 \quad (2)$$

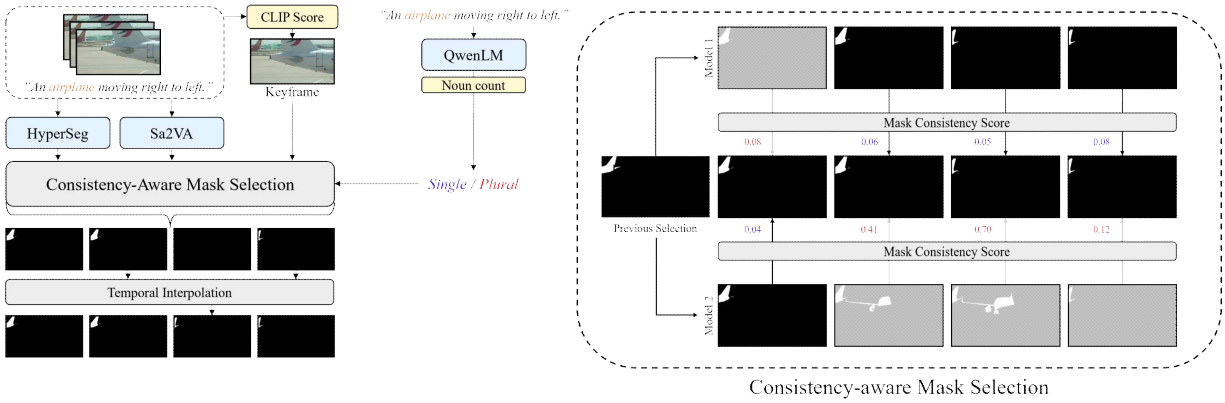


그림 1. 제안하는 프레임워크의 구조

참조 표현이 복수 객체를 지칭하는 경우, 단일 객체 분할 전략은 성능 저하를 초래한다. 이를 해결하기 위해 Qwen[6]을 이용하여 입력 텍스트의 명사 수를 판별하고, 단·복수 여부에 따라 마스크 선택 방식을 달리한다. 단수 객체의 경우 일관성 점수를 기반으로 한 단일 마스크를 선택하며, 복수 객체의 경우 두 모델이 예측한 마스크를 결합하여 다중 객체를 동시에 포함하도록 한다. 일부 프레임에서는 예측에 실패해 마스크가 비어 있는 경우, 시간적 불연속성이 발생한다. 이를 방지하기 위해, 해당 프레임의 이전 및 이후 프레임에서 예측된 마스크의 평균을 취하여 보간된 마스크를 생성한다. 이를 통해 연속적인 객체 추적과 시각적 일관성을 확보하도록 한다.

실험은 MeViS[7] 벤치마크 데이터셋을 사용하여 수행되었으며, 평가는 Jaccard Index(J), F-measure(F), 그리고 J&F 평균을 사용하였다. 실험에서는 Sa2VA-26B와 HyperSeg 두 모델을 앙상블하였다.

Methods	$J\&F$	J	F
MeViS	40.23	36.51	43.90
HyperSeg	58.95	55.32	62.59
Grounded-SAM2	59.26	54.59	63.92
Sa2VA-4B	55.10	50.74	59.46
Sa2VA-8B	59.28	54.95	63.62
Sa2VA-26B	61.38	56.53	66.23
Ours	**64.80**	**60.42**	**69.10**

표 1. 벤치마크 데이터셋에 대한 성능 결과 비교

실험 결과 본 논문에서 제안하는 방법이 기존 방법들보다 우수한 성능을 보이는 것을 확인할 수 있었다. 더 나아가, 본 논문에서 제안한 핵심 요소들의 기여를 분석하기 위해 절제 실험을 표 2와 같이 수행하였다. 그 결과, 모든 모듈을 결합했을 때 성능이 가장 높게 나타났으며, 특히 마스크 선택 기법과 객체 수 판별 모듈이 성능에 가장 큰 기여를 하는 것으로 확인되었다.

Mask Selection	Noun Counting	Interpolation	Keyframe	$J\&F$
✓	-	-	-	63.65
✓	✓	-	-	64.20
✓	✓	✓	-	64.56
✓	✓	✓	✓	**64.80**

표 2. 핵심 요소에 대한 절제 실험 결과 비교

Ⅲ. 결론

본 논문에서는 RVOS 문제에서 발생하는 시간적 불안정성과 객체 수 판별 문제를 해결하기 위해 문맥 인지 적응형 마스크 선택 기법을 제안하였다. 제안 방법은 CLIP 기반 키프레임 선정, 마스크 일관성 점수 계산, 텍스트 기반 객체 수 판별, 그리고 시간적 마스크 보간으로 구성되며, 서로 보완적으로 작용하여 높은 분할 정확도와 일관성을 확보하였다. MeViS 벤치마크 데이터셋에 대한 실험 결과, 제안 기법은 기존 최신 방법들보다 우수한 성능을 보였으며, 절제 연구를 통해 각 모듈의 독립적 기여를 검증하였다. 향후 연구에서는 본 방법을 다양한 대규모 비디오 분할 데이터셋과 실제 응용 환경에 확장 적용하고, 실시간 추론 환경에서의 효율성 향상 방안을 탐구할 예정이다. 또한, 현재는 두 개의 모델 출력을 비교·선택하는 방식이지만, 다수의 모델 출력을 동적으로 가중 결합하는 메타-앙상블 구조로 확장하는 연구도 기대할 수 있을 것으로 생각된다.

ACKNOWLEDGMENT

이 성과는 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원-대학ICT연구센터의 지원(IITP-2025-RS-2020-II201808, 50%)과 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. RS-2025-02214941, 50%).

참 고 문 헌

[1] Wu, Jiannan, et al. "Language as queries for referring video object segmentation." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.

[2] Wu, Dongming, et al. "Onlinerefer: A simple online baseline for referring video object segmentation." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.

[3] Yuan, Haobo, et al. "Sa2va: Marrying sam2 with llava for dense grounded understanding of images and videos." arXiv preprint arXiv:2501.04001 (2025).

[4] Wei, Cong, et al. "Hyperseg: Towards universal visual segmentation with large language model." arXiv preprint arXiv:2411.17606 (2024).

[5] Radford, Alec, et al. "Learning transferable visual models from natural language supervision." International conference on machine learning. PmLR, 2021.

[6] Bai, Jinze, et al. "Qwen technical report." arXiv preprint arXiv:2309.16609 (2023).

[7] Ding, Henghui, et al. "MeViS: A large-scale benchmark for video segmentation with motion expressions." Proceedings of the IEEE/CVF international conference on computer vision. 2023.

스포츠 중계 스트림에서 효과적인 중간 광고 구간 검출 기법

주현철, 허재호, 배주한, 나태영
에스케이텔레콤

{hc.joo, jaeho.hur, juhan.bae, taeyoung.na}@sk.com

An Effective Detection Method for Station Break in Sports Broadcast Streams

Hyunchul Joo, Jaeho Hur, Juhan Bae, Taeyoung Na
SK Telecom

요 약

본고에서는 스포츠 중계 스트림에서 효과적인 중간 광고 구간 검출 기법을 제안한다. 제안하는 기법은 중간 광고 구간에서 노출되는 스코어보드 특성을 활용하여 낮은 계산 복잡도로 대략적인 중간 광고 구간을 설정한다. 그리고, 경기 구간과 중간 광고 구간 간 전환 시 묵음이 발생하는 점을 활용하여 중간 광고 구간을 높은 정확도로 보정한다. 실험 결과에서 상용 환경에서의 제안한 기법의 성능을 보인다.

Ⅰ. 서 론

스포츠 경기들은 경기 중간에 휴식시간을 갖으며, 이를 이용하여 스포츠 경기를 중계하는 방송사들은 중간 광고를 송출한다. 인터넷 포털사이트 또는 통신사들은 방송국으로부터 중간 광고가 포함된 스포츠 중계 스트림을 수신하고, 자사의 서비스로서 활용한다. 인터넷 포털사이트 또는 통신사들은 수익 창출을 위한 자사가 수주한 광고 영상 혹은 자사 서비스 프로모션의 노출 기회로 활용하기 위한 영상으로 방송사의 중간 광고를 대체한다. 이를 위해 중간 광고가 송출되는 시간 구간을 정확하게 검출할 필요가 있으며, 다양한 연구가 진행되어 왔다 [1-2]. [1]에서는 일부 스포츠 경기는 야외에서 이루어지고, 관중들의 응원 등에 의해 경기 구간이 중간 광고 구간에 비해 오디오 신호에 높은 수치의 잡음이 포함되어 있다는 점에 착안하고 있다. 스포츠 중계 스트림에서 오디오 잡음 수치가 경기 구간의 오디오 잡음 수치보다 낮아지는지 여부를 판단하여 중간 광고 구간을 판단한다. 또한, 오디오 잡음 수치가 낮아지기 시작한 지점의 전/후 영상을 서로 비교하여 중간 광고 구간 인식의 정확도를 높이는 방안을 제시하고 있다. 그러나, 스포츠 경기 홍보 등의 중간 광고 소재는 경기 구간과 마찬가지로 관중들의 응원 등이 포함되어 오디오 잡음 수치가 경기 구간의 오디오 잡음 수치와 유사하므로, 이를 구별하기 어려우며, 중간 광고 시작 지점을 정확히 판단할 수 없다. [2]에서는 스코어보드 특성을 활용하여 중간 광고 구간을 검출한다. 통상적으로 중간 광고에서는 각 방송사 별 우상단의 정해진 위치에 (대결 팀 명, 득점 상황, 진행 중인 이닝/쿼터) 등의 정보를 포함하는 스코어보드를 노출한다. 이를 활용하여 우상단의 정해진 위치에 스코어보드가 인식될 경우, 이를 중간 광고 구간으로 판단한다. 그러나, 중간 광고 구간 내내 스코어보드가 노출되는 것은 아니므로, 이와 같은 방식은 정확한 중간 광고 구간을 식별할 수 없다.

Ⅱ. 제안하는 중간 광고 구간 검출 기법

이번 장에서는 제안하는 중간 광고 구간을 검출하는 기법을 다룬다. 제안하는 기법은 2 단계로 구성된다: (A) 코스 그레인드 (Coarse-grained) 중간 광고 구간 설정 단계 (B) 파인 그레인드 (Fine-grained) 중간 광고 구간 보정 단계. 코스 그레인드 (Coarse-grained) 중간 광고 구간 설정 단계는 중간 광고에서 노출되는 스코어보드 특성을 활용하여 낮은 계산 복잡도로 대략적인 중간 광고 구간을 설정한다. 파인 그레인드 (Fine-grained) 중간 광고 구간 보정 단계는 경기 구간과 중간 광고 구간 간 전환 시 묵음이 발생하는 점을 활용하여 앞선 단계에서 설정한 대략적인 중간 광고 구간을 정밀 보정한다.

(1) 코스 그레인드 중간 광고 구간 설정 단계

제안하는 코스 그레인드 중간 광고 구간 설정 단계에서는 낮은 계산 복잡도로 중간 광고에서의 스코어보드를 인식하여 대략적인 중간 광고 구간을 설정한다. 앞서 설명했듯이, 스포츠 중계 스트림의 중간 광고 내 일부 영상에서 스코어보드를 포함한다. 통상적으로 스코어보드는 중간 광고 영상의 우상단의 정해진 위치에 (대결 팀 명, 득점 상황, 진행 중인 이닝/쿼터) 등의 정보를 포함하는 스코어보드를 노출한다. 수신한 영상에서 광학 문자 인식을 통해 위의 정보를 포함하는 단어들이 검출될 경우, 중간 광고에 해당하는 영상이라고 손쉽게 판단할 수 있다. 그러나, 중간 광고 구간의 스코어보드 유/무를 판별하기 위해 수신한 영상의 전체 해상도 영역에 대해 광학 문자 인식을 수행하는 것과 매 영상 프레임마다 텍스트 탐지를 수행하는 것은 불필요하며, 높은 계산 복잡도를 요구한다. 제안하는 기법은 중간 광고에서 스코어보드가 존재할 만한 관심 영역에 대해서만 광학 문자 인식을 수행하고, 광학 문자 인식 절차 내 텍스트 탐지 단계와 텍스트 인식 단계가 동시 수행되지 않도록 분리하고, 텍스트 탐지 단계에서 기 검출된 바운딩 박스들을 재사용하여, 동영상 전체의 광학 문자 인식 과정에 필요한 계산량을 감소시킴으로써, 중간 광고 검출의 성능을 향상시킬 수 있다. 세부적인 절차는 하기와 같다.

절차 1: (바운딩 박스 저장소 = null, 대략적인 중간 광고 시작 지점 = null, 대략적인 중간 광고 종료 지점 = null)로 초기 변수 값을 설정한다.

절차 2: 스포츠 중계 스트림의 영상 입수 시, 해당 방송사의 스코어보드가 포함될 관심 영역을 크롭핑하고, 전처리 단계를 수행한다.

절차 3: 바운딩 박스 저장소 = null 일 경우, 절차 4 로 진행한다. 그렇지 않을 경우, 절차 5 로 진행한다.
절차 4: 관심 영역에 대해 텍스트 탐지 단계를 수행한다. 바운딩 박스들이 검출될 경우 바운딩 박스 저장소에 검출된 바운딩 박스들을 저장한다. 그리고, 절차 2 로 되돌아간다.
절차 5: 바운딩 박스 저장소에 기 저장된 바운딩 박스들에 대해 텍스트 인식 단계 및 후처리 단계를 수행한다.
절차 6: 절차 5 의 인식 결과 내 (대결 팀 명, 득점 상황, 진행 중인 이닝/쿼터) 등의 정보를 포함하고 있으면, 중간 광고 구간 내 스코어보드로 판단하고, 절차 7 로 진행한다. 그렇지 않을 경우, 바운딩 박스 저장소를 null 로 설정하고, 절차 8 로 진행한다.
절차 7: 대략적인 중간 광고 시작 지점이 null 일 경우, 이를 입수 영상의 시간 정보로 설정한다. 이 때, 영상의 시간 정보로는 PTS (Presentation Time Stamp)를 활용할 수 있다. 대략적인 중간 광고 시작 지점이 이미 설정되어 있다면, 입력 영상이 중간 광고 구간 중의 영상이라고 판단한다. 그리고, 절차 2 로 되돌아간다.
절차 8: 대략적인 중간 광고 시작 지점이 null 일 경우, 입수 영상이 중간 광고 구간 중의 영상이 아니라고 판단하여 절차 2 로 되돌아간다. 그렇지 않을 경우, 대략적인 중간 광고 종료 지점을 입수 영상의 시간 정보로 설정한다. 이 시점에서 대략적인 중간 광고 구간이 식별된다. 즉, 대략적인 중간 광고 시작 지점은 중간 광고에서 스코어보드가 노출되는 시점의 영상의 PTS 로 설정되며, 대략적인 중간 광고 종료 지점은 중간 광고에서 스코어보드가 소멸되는 시점의 영상의 PTS 로 설정된다. 추후, 다음 번 중간 광고 구간 재탐색을 위해 대략적인 중간 광고 시작 지점 및 대략적인 중간 광고 종료 지점을 null 로 초기화하고, 절차 2 로 되돌아간다.

(2) 파인 그레인드 중간 광고 구간 보정 단계

제안하는 파인 그레인드 중간 광고 구간 보정 단계에서는 경기 구간과 중간 광고 구간 간 전환 시 묵음이 발생하는 점을 활용하여 앞선 단계에서 설정한 대략적인 중간 광고 구간을 높은 정확도로 보정한다. 통상적으로 스포츠 중계 스트림에서 측정된 순간 음량 (Momentary Loudness)이 기 정의된 Loudness 임계값 (-70 LUFS) 보다 낮다면 묵음으로 판단한다. 스포츠 중계 스트림에서 경기 구간과 중간 광고 간 전환 시, 스위처의 입력 소스 변경으로 인해 묵음 구간이 발생한다. 또한, 경기 구간 내 사전 준비된 영상 (예를 들어, 하이라이트 영상, 선수 인터뷰 등)으로 전환 및 중간 광고 내 광고 소재 전환 시에도 묵음은 검출될 수 있으므로, 경기 구간과 중간 광고 구간 간 전환 시에 발생하는 묵음과 구별할 필요가 있다. 이를 위해 제안하는 기법은 앞선 단계에서 설정된 대략적인 중간 광고 구간 기준 전/후 시간적으로 가장 가까운 묵음 검출을 통해 중간 광고 구간을 보정하여 정확성을 높인다. 즉, 중간 광고 구간 내 스코어보드 노출/소멸 시점에서 가장 인접한 전/후 묵음 시점은 중간 광고 구간 시작/종료 지점 상에 발생하는 묵음이며, 경기 구간 내 발생되는 묵음들과는 시간 축에서 구별될 수 있다. 자세히 설명하자면, 앞선 단계에서 설정된 대략적인 중간 광고 시작 지점 이전의 묵음 구간 중 시간적으로 가장 인접한 묵음 구간을 중간 광고 구간의 시작 지점으로 재조정한다. 또한, 앞선 단계에서 설정된 대략적인 중간 광고 종료 지점 이후의 묵음 구간 중, 시간적으로 가장 인접한 묵음 구간을 중간 광고 구간의 종료 지점으로 재조정한다.

Ⅲ. 실험 결과

이번 장에서는 제안하는 기법을 이용한 시스템의 성능을 보인다. 실험 환경은 다음과 같이 구축되었다. 중간 광고 교체 서버의 UDP 소스 모듈은 IPTV 헤드엔드로부터 IP 멀티캐스트 방식으로 스포츠 중계 스트림을 수신하며, 제안하는 중간 광고 검출 모듈을 통해 중간 광고 구간을 식별하고, 이를 중간 광고 교체 모듈에 알린다. 그리고, 중간 광고 교체 모듈은 방송사의 중간 광고를 기 저장된 광고 소재로 대체하여 UDP 싱크 모듈을 통해 트랜스코더로 전달하며, 트랜스코더는 HLS (HTTP Live Streaming) 포맷으로 트랜스코딩하고, 이를 플레이어에게 서비스한다.

우선, (CPU: Intel(R) Xeon(R) Platinum 8260 CPU @ 2.30GHz & 2 GPU: A6000 & Memory: 188 Gbytes) 하드웨어 사양의 중간 광고 교체 서버에서 자원 사용량을 측정하였다. 대상 경기 종목은 야구 스포츠이며, 4 개 중계 채널을 동시 서비스하였다. 표 1 에서 볼 수 있듯이, 제안하는 중간 광고 검출 기법을 통해 광고 교체 서비스를 제공하더라도 매우 낮은 자원 사용량을 보여주고 있다. 이는 제안하는 기법이 광학 문자 인식 과정에 필요한 전반적인 계산량을 감소시킬 수 있음을 의미한다.

[표 1] 중간 광고 교체 서버에서 자원 사용량

CPU 사용률	3 ~ 4 %
메인 메모리 사용률	2.0 ~ 2.4 %
E2E 지연	30 ~ 100 msec
GPU 사용률	7 ~ 25 %
GPU 메모리 사용률	2 ~ 4 %

또한, 단순 스코어보드 특성 만을 활용하여 중간 광고 구간을 검출하는 [2]와의 검출 정확도를 비교하고자 한다. 제안하는 기법은 높은 정밀도로 중간 광고 구간을 식별할 수 있으므로, 방송국에서 제공된 중간 광고 소재들을 온전히 대체할 수 있었다. 반면, [2]에서는 중간 광고에서 첫 번째 대체 소재가 재생되기 전 및 마지막 대체 소재가 재생 된 후에 방송사에서 제공한 광고 소재가 각각 1.9 ~ 3.2 초 및 6.6 ~ 13.4 초 노출되었다.

Ⅳ. 결론

본고에서는 코스 그레인드 중간 광고 구간 설정 단계와 파인 그레인드 중간 광고 구간 보정 단계로 구성된 스포츠 중계 스트림에서 중간 광고 구간 검출 기법을 제안하였다. 제안하는 기법은 중간 광고 구간에서 노출되는 스코어보드 특성 및 경기 구간과 중간 광고 구간 간 전환 시 묵음이 발생하는 점을 활용하여 중간 광고 구간을 검출한다. 실험에서 제안하는 기법은 낮은 계산 복잡도로 높은 정밀도의 중간 광고 구간을 검출할 수 있음을 입증하였다.

ACKNOWLEDGMENT

본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2024 년도 문화체육관광연구개발사업으로 수행되었음 (과제명: 전통예술 가무악의 융복합 공연제작 활성화를 위한 융복합 공연 기획/제작 플랫폼 기술 개발, 과제번호: RS-2024-00398536, 기여율: 100%)

참 고 문 헌

[1] Pandora TV Co., “Method for changing advertisement in broadcasting”, Korea Patent 10-2013-0123037, filed Oct. 16 2013, and issued Dec. 4 2014.

[2] “SKTelecom A.TV Baseball”, 2022.

다축 평면 투영 기반 3D 객체 탐지를 활용한 ROI 보존형 포인트 클라우드 다운샘플링

임채윤, 박현희*
명지대학교 정보통신공학과

dlacodbs0627, hhpark*@mju.ac.kr

3D Object Detection based on Multi-Axis Plane Projection for ROI-Preserving Point Cloud Downsampling

Lim Chae Yun, Park Hyun Hee*
Dept. Information and Communication Engineering, Myongji University

요 약

본 논문은 3D 포인트 클라우드에서 시각적으로 중요한 영역을 보존하는 다운샘플링을 제안한다. 다축 평면 투영을 기반으로 3D region-of-interest(ROI)를 탐지하고 ROI 와 non-ROI 를 다른 품질로 다운샘플링한다. 동일 포인트 개수에서 기존 다운샘플링 기법 대비 중요 영역의 세부 구조를 더욱 효과적으로 보존한다. 실험을 통해 다운샘플링 결과의 시각적 품질뿐만 아니라, 탐지 작업을 위한 입력 데이터로도 활용이 가능하다는 것을 입증한다. ROI 보존형 다운샘플링 기법은 적응형 스트리밍에 활용할 수 있고, 핵심 정보를 중심으로 데이터를 전송하는 시맨틱 통신에도 기여할 수 있다.

I. 서 론

포인트 클라우드는 3 차원 공간에 정의된 점들의 집합이다. 객체와 공간을 입체감 있게 시각화할 수 있고, extended reality(XR, 확장현실)로의 연계 가능성 덕에 차세대 미디어 포맷으로 주목받고 있다. 하지만 데이터 크기가 방대하여 무선 통신 대역폭을 초과하는 경우가 발생한다[1]. 데이터 용량 축소 방법으로 포인트 수를 줄이는 다운샘플링 기법이 널리 사용되지만, 핵심 정보까지 함께 제거되는 한계점이 따른다. 특히 화상 통화와 컴퓨터 비전 등 시각적 품질이 중요한 분야에서는 사용자 경험 저하의 원인이 되기 쉽다. 따라서 핵심 정보 손실을 최소화하면서 포인트의 수를 줄이는 접근법이 요구된다.

본 연구는 다축 평면 투영 기반의 ROI 보존형 다운샘플링 기법을 제안한다. ROI 내의 포인트는 보존하고, non-ROI 내의 포인트는 더 많이 제거해 기존 다운샘플링 기법 대비 시각적으로 중요한 영역의 품질 저하를 최소화한다. 제안하는 기법은 다운샘플링을 활용한 데이터 크기 감소, 핵심 정보 보존이라는 두 목표를 함께 만족한다. 또한 낮은 품질에서도 ROI 정보가 크게 손실되지 않아 시각화 외에 이미지 탐지 작업에서도 안정성을 보인다.

(a) voxel grid downsampling (b) farthest point sampling
(c) ROI-preserving downsampling

그림 1. 기존 기법과 제안 기법의 시각화 비교

II. 본론

그림 1 은 기존 다운샘플링 기법과 제안하는 기법을 동일한 포인트 개수만큼 남긴 시각화 예시이다. 그림 1(a)과 그림 1(b)과 같은 기존 기법들은 전체 포인트를 균일하게 다운샘플링하지만, 얼굴의 상세 정보가 누락되는 문제를 보인다. 반면 제안 기법인 그림 1(c)은 덜 중요한 영역인 몸쪽의 포인트는 더 적은 수로 표현되지만, 중요한 영역인 얼굴의 상세 정보는 더욱 잘 보존한다.

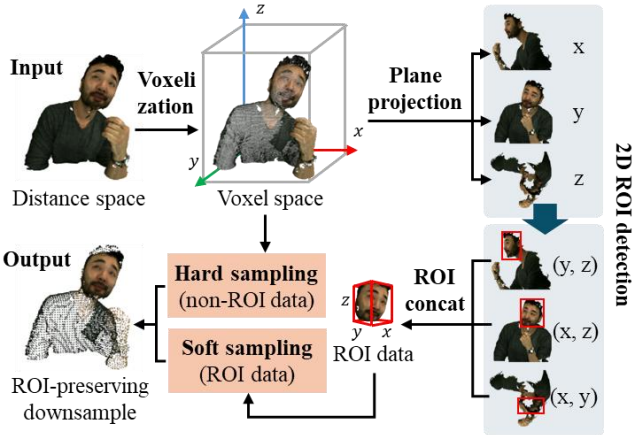


그림 2. ROI 보존형 다운샘플링 기법 전체 아키텍처

그림 2 에서의 원본 데이터는 프레임 당 약 150 만 개의 포인트로 인해 22.5MB 라는 매우 큰 용량을 갖는다. 해당 데이터를 미디어 콘텐츠에서 자주 사용하는 30 frames per second(FPS)로 수신하려면 5.4Gbps 의 전송률이 필요하다. 4K 평면 영상의 전송률[1]인 25Mbps 와 비교하면 매우 높은 수치이다. 이는 포인트 클라우드 전송에 데이터 용량 축소가 필수적이라는 것을 보여준다.

본 논문에서 포인트 클라우드의 3D ROI 를 얻어내는 방법은 크게 세 단계로 이루어진다. 먼저, 포인트 클라우드를 복셀화하여 균일한 밀도로 변환한다. 원본 포인트 클라우드 데이터는 객체의 특징에 따라 일부 영역에서 과밀집된다. 탐지, 분할 작업에서는 물체의 특징 정보로 활용할 수 있지만 시각화 측면에서는 포인트가 겹치거나 중복되는 문제를 보인다. 다음은 여러 축을 기준으로 평면 투영한다. 3D 포인트 클라우드 데이터를 2D 이미지로 변환하고, 객체 탐지와 같은 딥러닝 모델을 활용해 2D ROI 를 구한다. 마지막으로 여러 축의 2D ROI 좌표를 3D 좌표로 통합한다. 하나의 축에서만 ROI 가 감지되는 경우, 얻지 못한 좌표는 임의의 최댓값으로 대체한다. 얻어낸 3D ROI 좌표는 시각 정보 보존을 위해 사용한다.

III. 실험

품질 비교에 사용한 voxel grid downsampling 및 farthest point sampling 은 포인트 수를 균일하게 줄이는 기존 다운샘플링 기법이다. 여러 품질로 각 기법을 적용해 시각적 품질을 정성적으로 비교한다. 추가로, 제안하는 기법은 다운샘플링 이후에도 주요 정보를 보존할 수 있어 추후 탐지, 분할과 같은 작업에도 강건함을 보일 것으로 예상한다. 이를 실험으로 입증하고자 COCO 데이터셋으로 사전 학습된 YOLOv8 을 활용해 80 개 클래스 중 person 클래스의 탐지 결과를 정량적으로 측정한다.

데이터셋은 Microsoft Voxelized Upper Bodies - david 포인트 클라우드 영상을 활용한다. 해당 영상은 30FPS 의 주사율을 갖는다. 2D ROI 탐지 모델은 WIDAR FACE 얼굴 데이터셋으로 파인튜닝 된 YOLOv8 을 사용한다. 다운샘플링 품질의 경우 국내 통신 3 사 서비스 품질 평가에서의 5G 평균 다운로드 속도인 1,025.52 Mbps[2]를 실험의 최대 품질 기준으로 삼는다. 이 경우 30FPS 조건에 맞추기 위해 한 프레임에 수용할 수 있는 최대 포인트 수는 284,866 포인트이며, 이를 최대 품질(resolution) r_{max}라 정의한다. 수식 1 에서 r_k는 r_{max}에 대해 k 단계 강도의 다운샘플링을 적용한 포인트 수를 의미한다. k가 증가할수록 포인트 수는 지수적으로 감소하며, 소수점 이하는 내림 처리하여 정수로 표현한다.

$$r_k = \left\lfloor \frac{r_{max}}{2^k} \right\rfloor, \quad k = 0, 1, 2, 3, 4 \qquad (1)$$

Resolution	$r_0(r_{max})$	r_1	r_2	r_3	r_4
Voxel grid					
Farthest point Sampling					
ROI preserving (Ours)					
Point num	284,866	142,433	71,216	35,608	17,804
Bitrate @30fps (Mbps)	1,026	513	256	128	64

그림 3. 단계적 다운샘플링 시각화 및 전송률 품질 비교

그림 3 은 단계별 시각화 품질과 함께 30FPS 로 전송하는 데 필요한 전송률을 보여준다. 다운샘플링 기법의 하이퍼파라미터는 각 기법에 맞게 적절하게 설정한다. voxel grid downsampling 및 farthest point sampling 기법은 품질 단계가 낮아질수록 시각 정보가 전체적으로 소실되어 중요 영역의 세부 정보는 식별하기 어려워진다. 반면 제안하는 기법은 품질이 낮아지더라도 중요 영역에 대한 시각 정보를 유지한다.

표 1. YOLOv8 모델 기반 person 클래스 탐지 예측 결과

Type	Evaluation	Resolution					
		r_0	r_1	r_2	r_3	r_4	avg
Voxel grid	POD	1.0	1.0	**0.833**	0.097	0.222	0.631
	Confidence	0.903	0.864	0.649	0.322	0.398	0.627
Farthest point sampling	POD	1.0	0.963	0.458	0.051	0.153	0.525
	Confidence	0.892	**0.866**	**0.675**	0.363	0.434	0.646
ROI preserving (Ours)	POD	1.0	1.0	0.671	**0.727**	**0.736**	**0.827**
	Confidence	**0.905**	0.864	0.637	**0.543**	**0.517**	**0.693**

표 1 은 각 기법으로 다운샘플링된 결과 이미지를 탐지하여 평균 낸 결과이다. probability of detection(POD)은 person 객체가 성공적으로 감지된 확률이다. confidence 는 감지된 person 객체의 신뢰도 점수이다. 기존 기법들은 데이터 품질에 따라 POD 과 confidence 가 크게 불안정하다. 반면 본 논문에서 제안하는 기법은 전체적인 품질에서 탐지 작업을 안정적으로 수행한다.

Ⅳ. 결론

본 논문은 다축 평면 투영으로 얻은 3D ROI 를 활용해 중요 영역을 보존하는 포인트 클라우드 다운샘플링을 제안한다. 동일 포인트 개수에서 기존 기법 대비 시각적으로 중요한 영역을 효과적으로 보존한다. 또한 다양한 품질로 얻어낸 이미지를 활용해 탐지 작업을 수행했을 때도 POD 와 confidence 에서 안정적인 값을 보여준다.

ROI 보존형 다운샘플링은 별도 코덱 없이도 원본 데이터의 용량을 효과적으로 감축할 수 있다. 다양한 품질에서도 중요 정보를 효과적으로 보존하는 점에서 적응형 스트리밍 또는 시맨틱 통신에 활용할 수 있을 것으로 보인다. 다만 현재로서는 ROI 탐지 모델 성능에 따라 결과가 크게 변화하고, 시간 흐름 정보를 반영하지 않아 프레임 간 일관성이 낮을 수 있다는 한계점이 따른다.

향후 연구에서는 다중 클래스 탐지 또는 문맥 파악을 활용해 중요도 기반 다운샘플링으로 확장하고자 한다. 또한 사람 데이터 외에도 LiDAR 센서 기반 도시 데이터와 같이 대규모 포인트 클라우드에서의 다축 평면 투영 및 ROI 보존형 다운샘플링을 적용해 볼 계획이다.

ACKNOWLEDGMENT

이 논문은 산업통상자원부 및 산업기술평가원(KEIT)과 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. RS-2024-00469138, 국산 AI 반도체 기반 비전인식기술 소프트웨어개발도구 개발, No. 2022R1A2C2005705, 분산 머신러닝 기반 지능형 플라잉 기지국을 위한 AI-MAC 프로토콜)

참 고 문 헌

[1] Z. Liu et al., “Point Cloud Video Streaming: Challenges and Solutions,” in IEEE Network, vol. 35, no. 5, pp. 202-209, Sept. 2021, doi: 10.1109/MNET.101.2000364

[2] 과학기술정보통신부·한국지능정보사회진흥원. “2024 년 통신서비스 수신 가능 지역 점검 및 품질평가 결과 발표. 보도자료. Dec, 2024.

물체 분류를 위한 엣지 AI 기반 매니퓰레이터 동작 제어 연구

김민경, 정준호, 주우현, 윤주상*

동의대학교

kmksummer0902@gmail.com, jeong@junho.dev, joowh0207@gmail.com, *jsyoun@deu.ac.kr

A Study on Edge AI-Based Manipulator Motion Control for Object Classification

MinKyeong Kim, Junho Jeong, WooHyun Joo, Joosang Youn*

Dong-Eui Univ.

요 약

본 논문에서는 센서 데이터 처리 및 복잡한 AI 연산을 효율적으로 수행하여 실시간 로봇 작업을 가능하게 하는 물체 분류를 위한 엣지 AI 기반 매니퓰레이터 동작 제어 시스템을 제안한다. 시스템 모듈화를 위해 각 기능을 컨테이너로 구현하였으며, 시스템 지연 시간을 줄이기 위해 복잡한 딥러닝 모델 추론을 엣지 서버에서 처리하는 엣지 AI 구조를 도입하였다. 매니퓰레이터 제어 보드에서 센서 데이터를 수집하여 RTPS 토픽으로 엣지 서버에 전송하고, 엣지 서버에서 ResNet-50을 활용해 실시간 추론을 수행한다. 추론 결과를 바탕으로 엣지 서버의 스킬 저장소에서 적합한 작업 시퀀스를 선택 후 매니퓰레이터 제어 보드로 반환하여 매니퓰레이터를 제어한다. 실험 결과 본 시스템을 통해 매니퓰레이터가 객체 분류 결과에 따른 동작을 수행할 수 있음을 확인하였다.

Ⅰ. 서 론

최근 지능형 사물인터넷 및 로봇 분야의 발전으로 대용량 센서 처리와 복잡한 AI 연산을 빠르고 효율적으로 수행하는 기술이 요구되고 있다. 객체 분류 결과에 따라 로봇이 실시간으로 동작을 수행하기 위해서는 센서 데이터 처리와 인공지능 모델의 추론이 빠르게 처리되어야 한다. 하지만 로컬에서 모든 연산을 처리할 경우, 연산 자원의 한계로 인해 시스템 지연이 발생하여 실시간 로봇 작업의 안정성이 저하될 수 있다. 본 논문에서는 이러한 한계를 극복하기 위해 인공지능 모델 추론을 엣지 서버에 오프로딩하여 처리하는 물체 분류를 위한 엣지 AI[1] 기반 매니퓰레이터 동작 제어 시스템을 제안한다.

본 논문의 구성은 다음과 같다. 2장에서는 전체 시스템 구성을 설명하고, 3장에서는 로봇의 작업 수행 과정을 다룬다. 4장에서는 실험 과정 및 결과를 설명하며 5장을 결론으로 마친다.

Ⅱ. 시스템 구성

본 시스템은 그림 1에서 나타내는 것과 같이 엣지 AI 구조를 기반으로 매니퓰레이터 제어 보드, 엣지 서버, 카메라, 매니퓰레이터로 구성되어 있다. 매니퓰레이터 제어 보드는 카메라 데이터 수집 컨테이너와 매니퓰레이터 제어 컨테이너로 구성되며 엣지 서버는 객체 분류 컨테이너와 스킬 관리 컨테이너로 구성된다. 매니퓰레이터 제어 보드의 카메라 데이터 수집 컨테이너는 센서 데이터를 수집하여 RTPS 토픽을 엣지 서버의 객체 분류 컨테이너로 전달하고, 객체 분류 컨테이너는 객체 분류 모델을 이용해 객체 분류를 수행한다. 객체 분류 모델은 뛰어난 성능과 효율성을 갖춘 구조로 실시간 추론에 적합하며 다양한 고성능 모델들의 백본(backbone)으로 자주 활용되는 ResNet-50[2]로 선정했다. 스킬 관리 컨테이너는 객체 분류 모델의 추론 결과를 바탕으로 스킬 저장소에서 적합한 스킬을 선택해 매니퓰레이터 제어 보드의 매니퓰레이터 제어 컨테이너로 전송한다. 해당 스킬 정보를 바탕으로 매니퓰레이터 제어 컨테이너에서 목표 포즈를 계산 후 매니퓰레이터가 해당 포즈로 동작하도록 제어한다.

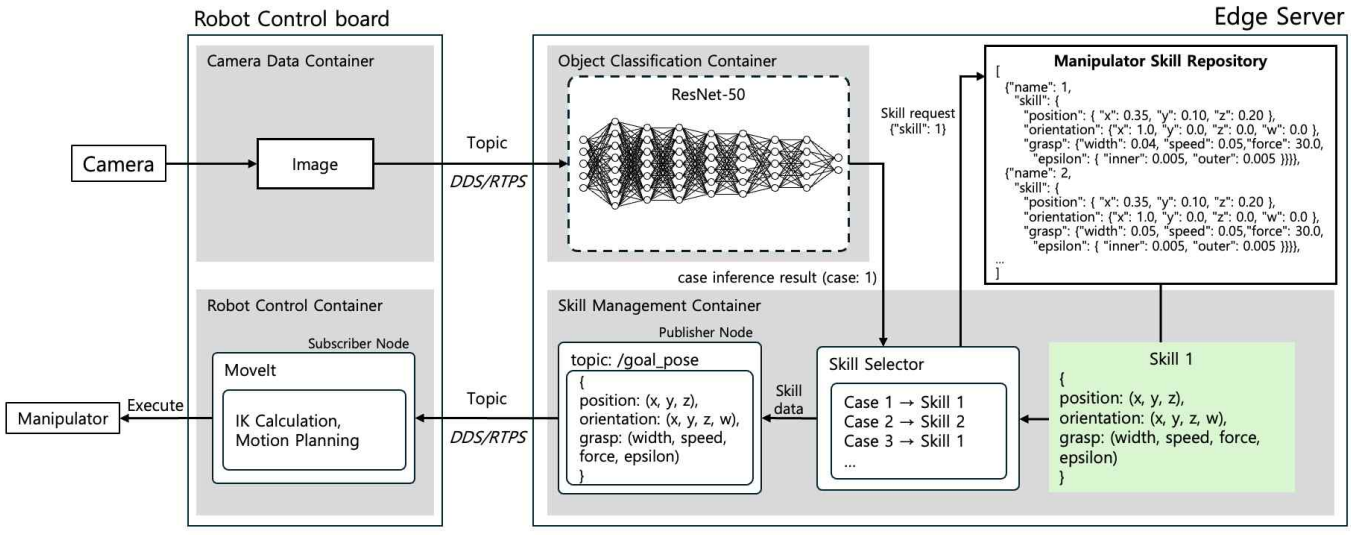


그림 1. 매니퓰레이터 제어 보드 및 엣지 서버 간 엣지 AI 기반 물체 분류 및 로봇 제어 시스템 설계도

Ⅲ. 로봇 작업 수행 과정

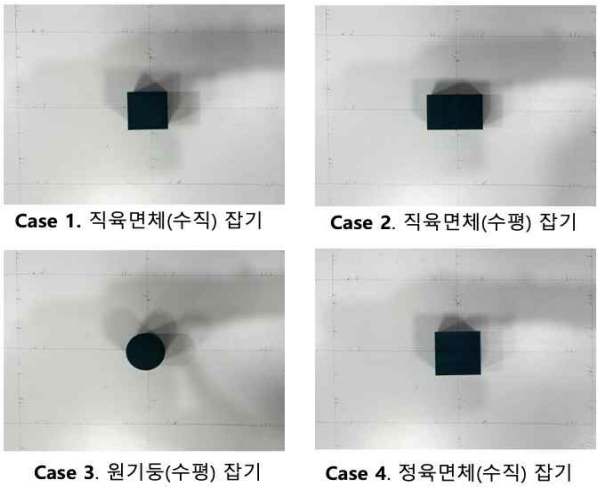


그림 2. 매니퓰레이터 작업 케이스

그림 2는 매니퓰레이터가 수행할 수 있는 4가지 작업 케이스를 나타내며, 각 케이스는 고정된 위치에 놓인 물체의 종류 및 배치 방향에 따라 정의된다. 스킬 저장소에는 매니퓰레이터의 엔드이펙터 위치와 방향, 그리퍼 개폐 정보가 JSON 형식으로 저장되어 있으며 객체 분류 모델의 작업 케이스 추론 결과에 따라 스킬 저장소에서 케이스에 적합한 스킬이 선택되어 매니퓰레이터 제어 컨테이너로 전송된다. 매니퓰레이터 제어 컨테이너의 MoveIt 프레임워크는 전송된 스킬 정보를 목표 포즈로 설정하고 역운동학을 통해 관절 각도를 계산한다. 이후 플래닝 알고리즘을 사용하여 충돌 회피를 고려한 최적 경로를 생성하고, 해당 경로대로 매니퓰레이터가 동작하며 그리퍼의 개폐도 함께 제어된다.

Ⅳ. 실험 및 결과

객체 분류 모델의 학습을 위해 데이터 증강 기법인 RandAugment[3]를 사용하여 원본 이미지 144장을 각 이미지당 30개의 증강을 적용해 총 4,320장의 데이터셋을 구성했다. 이 중 3,000장을 학습 데이터로 사용했으며 1,320장을 검증 데이터로 사용해 모델을 평가했다. 모델의 학습 결과 0.9803의 정확도를 달성했으며 표 1에서 각 클래스별 평가 결과를 확인할 수 있다. 또한 그림 3과 같이 객체 분류 결과를 바탕으로 매니퓰레이터가 물체의 위치와 방향에 맞는 동작을 수행함을 확인하였다.

표 1. 객체 분류 모델 평가 결과

클래스	정밀도 (Percision)	재현율 (Recall)	F1-Score	샘플 수 (Support)
case1	0.9649	0.9872	0.9795	390
case2	0.9636	0.9700	0.9668	300
case3	1.0000	1.0000	1.0000	390
case4	0.9956	0.9500	0.9723	240
Macro 평균	0.9810	0.9768	0.9787	1320
가중 평균	0.9806	0.9803	0.9803	1320

그림 3. 모델 추론 결과에 따른 매니퓰레이터 스킬 수행

Ⅴ. 결론

본 논문에서는 엣지 서버와 매니퓰레이터 제어 보드 간의 엣지 AI 구조를 사용한 매니퓰레이터 제어 시스템을 제안하였다. 센서 데이터 수집과 객체 분류, 매니퓰레이터 제어를 모듈화된 도커 컨테이너 및 RTPS 통신을 통해 효율적으로 연동함으로써 실시간성 및 확장성을 확보하였다. 또한 객체 분류 모델을 사용하여 작업 케이스를 분류하고, 매니퓰레이터의 스킬 정보를 MoveIt 프레임워크에 적용하여 경로를 계획함으로써 매니퓰레이터의 정밀한 작업 수행을 가능하게 하였다. 이를 통해 연산 자원의 한계를 극복하고 인공지능과 로봇 제어를 융합하여 로봇 응용 분야에서의 활용 가능성을 입증하였다.

다만, 현재 시스템은 매니퓰레이터가 4가지의 미리 정의된 작업 케이스만 수행하도록 설계되어 있어 적용 범위에 제약이 있다. 이를 보완하기 위해 작업 케이스를 20가지로 확장할 계획이며, 단순한 모형 블록뿐만 아니라 일상생활에서 접할 수 있는 물체를 학습 대상으로 포함시켜 시스템의 실제 응용 가능성을 높이고자 한다. 또한 로컬 AI와 엣지 AI 환경에서 시스템 지연 시간을 측정하여 제안 시스템의 성능을 정량적으로 분석할 계획이다. 이후 분할 지능 구조를 도입하여 연산 부하를 분산하고 데이터 전송 및 처리 지연을 최소화할 예정이며 이를 통해 효율성과 안정성을 향상시킬 수 있을 것으로 기대된다.

ACKNOWLEDGMENT

본 논문은 과학기술정보통신부 및 정보통신기획평가원의 복합지능자율행동체SW핵심기술개발사업(RS-2024-00346798)과 부산광역시 및 (재)부산테크노파크의 BB21plus 사업으로 지원된 연구결과임.

참 고 문 헌

[1] Hua, H., Li, Y., Wang, T., Dong, N., Li, W., & Cao, J., “Edge computing with artificial intelligence: A machine learning perspective,”in ACM Computing Surveys, 55(9), pp. 1-35, 2023.

[2] He, K., Zhang, X., Ren, S., and Sun, J., "Deep Residual Learning for Image Recognition," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-778, 2016.

[3] Ekin D. Cubuk, Barret Zoph, Jonathon Shlens, Quoc V. Le., “RandAugment:Practical automated data augmentation with a reduced search space,” in Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, pp. 702-703, 2020.

엣지 디바이스 인터워킹을 활용한 SDV 서비스 재구성 기법 연구

권태현, 성은산, 윤주상*
동의대학교
dhxk9513@gmail.com, ses0137112@naver.com, *jsyoun@deu.ac.kr

An Edge Device Interworking Method for SDV Service Reconfiguration

Taehyeon Kweon, Seong Eun San, Joosang Youn*
Dong-Eui Univ.

요 약

소프트웨어 정의 차량(Software-Defined Vehicle, SDV)은 서비스 지향 아키텍처(SOA)를 기반으로 하드웨어와 소프트웨어를 분리하며 고성능 연산 플랫폼 위에서 모듈화된 소프트웨어를 실행함으로써 다양한 기능을 재구성할 수 있다. 선행 연구에서 SDV 간 서비스 재구성을 지원하기 위한 네트워크 아키텍처를 제안하였으며 퍼블릭 및 프라이빗 마이크로서비스(Microservice, MS)를 구분하고 서비스 에이전트를 통한 탐색 및 데이터 흐름 제어를 통해 상호운용성과 확장성을 확보하였다. 그러나 최근 SDV에는 다중 센서 융합 기반 객체 인식과 주행 상황 예측 등과 같은 고성능 지능형 기능이 동시에 요구되며 이는 차량 내부 컴퓨팅 자원만으로는 한계가 존재한다. 이에 따라 본 논문에서는 이러한 한계를 해결하기 위해 SDV와 엣지 간 인터워킹을 통한 AI 서비스 재구성 기법을 제안한다. 각 에이전트는 외부의 에이전트와 토픽 정보를 공유하여 로컬 토픽을 글로벌 토픽으로 변환하고 데이터를 송수신내부 MS와 연결된 로컬 토픽을 글로벌 토픽으로 변환해 외부와 데이터를 교환하며 엣지 디바이스에 배치된 AI 마이크로서비스가 이를 처리한 후 결과를 다시 글로벌 토픽으로 배포한다. 이를 통해 SDV는 하드웨어 변경 없이도 엣지의 AI를 활용하여 연산량이 큰 지능형 기능을 유연하게 재구성할 수 있으며 네트워크 상태 변화나 자원 제약 환경에서도 안정적인 서비스 제공이 가능하다.

Ⅰ. 서 론

소프트웨어 정의 차량(Software-Defined Vehicle, SDV)은 서비스 지향 아키텍처(Service Oriented Architecture, SOA)를 기반으로 차량의 하드웨어와 소프트웨어를 분리하고 고성능 연산 플랫폼 위에서 모듈화된 소프트웨어를 실행함으로써 다양한 기능을 유연하게 재구성할 수 있다. 이로 인해 차량은 하드웨어의 변경없이 OTA(Over-The-Air) 업데이트를 통해 지속적으로 새로운 기능을 추가하거나 개선할 수 있으며 이는 자율주행과 지능형 모빌리티 구현의 핵심 기술로 주목받고 있다[1].

선행 연구에서는 SDV간 서비스 재구성을 지원하기 위한 네트워크 아키텍처를 제안하였다. 해당 아키텍처는 퍼블릭 서비스와 프라이빗 서비스를 구분하며 에이전트를 통해 서비스 탐색 및 데이터 흐름 제어를 수행함으로써 차량 내 외부의 다양한 마이크로서비스를 재구성 할 수 있는 구조를 제시하였다.

그러나 차량 내에서 요구되는 서비스는 다중 센서 융합 기반 객체 인식, 보행자 위험 행동 분석, 자율주행 상황 예측 등 복합적인 AI기반 기능으로 고도화되고 있으며 이로 인해 기존 서비스에 비해 많은 연산 자원을 필요로 한다. 이러한 경우 차량 내 HPC만으로는 연산 성능이 제한적이며 결국 더 고사양의 하드웨어를 도입하거나 업그레이드해야 하는 문제가 발생한다. 이는 제조 비용 증가와 유지보수 부담으로 이어져 SDV의 확장성을 저해할 수 있다[2].

이러한 한계를 극복하기 위해 본 논문에서는 기존의 에이전트를 엣지 디바이스의 연계하여 확장하는 방안을 제안한다. 이를 통해 SDV는 새로운 하드웨어 교체 없이도 엣지 자원을 활용하여 복잡한 지능형 서비스를 구성할 수 있으며 궁극적으로 하드웨어 의존도를 줄이면서 유연한 서비스 재구성이 가능하다.

Ⅱ. 관련 연구

기존 SDV 서비스 재구성은 차량 내 서비스로 이루어진다. 이에 따라 선행 연구에서 SDV간 서비스 재구성 기법을 제안했다. SDV 내 소프트웨어는 독립적인 기능 단위로 존재하며 이를 마이크로서비스(Microservice, MS)로 정의한다. MS를 외부 접근에 따라 프라이빗 MS와 퍼블릭 MS로 분리하고 MS의 유형에 따라 호스트 모드 네트워크, 브릿지 모드 네트워크 등 다르게 구성하였다. 또한 MS간 동적인 연결을 지원하기 위해 프록시 기능과 메시지 브로커 기능을 포함한 에이전트를 구성하였다. 이를 통해 퍼블릭 및 프라이빗 서비스간 통신을 중계하고 데이터 흐름을 효율적으로 설정할 수 있음을 보였다. 또한 서비스 레지스터를 구성하여 각 MS를 중앙에서 관리하고 등록 및 탐색 절차를 통해 SDV 간 상호 연결성을 확보하는 방안을 제시하였다[3]. 이러한 기법은 차량 내부의 컴퓨팅 자원을 통해 이루어지기 때문에 SDV가 점점 발전함에 따라 연산량이 큰 AI기반 기능이 동시에 요구되는 상황에서는 성능 저하 및 하드웨어적 한계가 존재한다.

Ⅲ. 본론

본 논문에서는 엣지와 인터워킹을 통한 SDV 서비스 재구성 기법을 제안한다. 기존에는 차량 내부에 AI MS를 구성했으며 이러한 서비스 수가 증가하는 경우 SDV 내 컴퓨팅으로는 한계가 존재한다. 이를 해결하기 위해 엣지 디바이스에 여러 AI MS를 구성하고 에이전트를 통해 SDV와 엣지 디바이스간 서비스 재구성을 진행한다. 선행 연구로 SDV 간 서비스 재구성 기법에 대해 제시하였으며 본 연구에서는 에이전트 기반 SDV와 엣지 디바이스 간 인터워킹 중점으로 연구하였다.

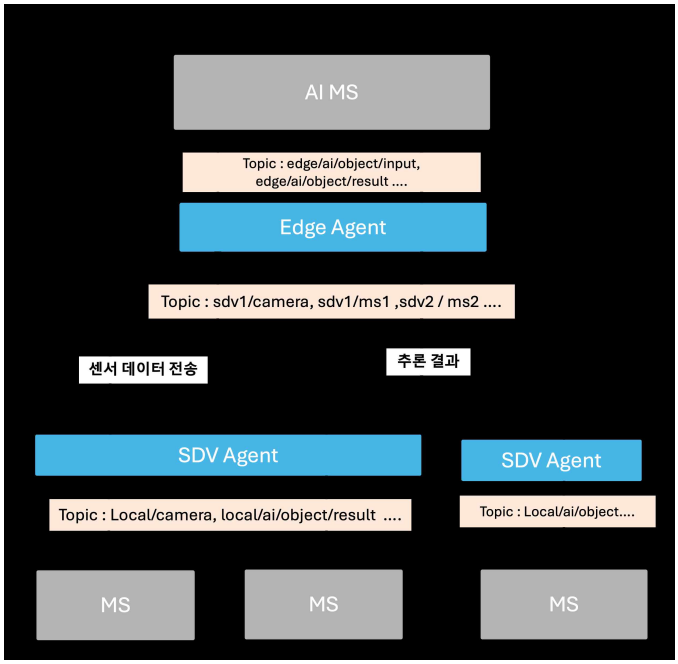


그림 1 SDV와 엣지 간 인터워킹 프레임워크

그림 1은 서비스 재구성 과정에서 SDV와 엣지 디바이스 간의 인터워킹 프레임워크를 도시한 것이다. 엣지 디바이스는 SDV와 동일한 구조를 가지며 각 에이전트는 내부 MS와 외부 간의 데이터 흐름을 중재하는 구조로 구성된다.

각 에이전트는 SDV 또는 엣지 내부의 MS와 연결되어 있으며 센서 데이터 및 처리 결과를 내부 MS와 통신은 로컬 토픽을 통해 이루어진다. 외부와의 데이터 교환이 필요한 경우 상대 에이전트와 토픽 정보를 공유하여 로컬 토픽을 글로벌 토픽으로 변환하고 데이터를 송수신한다. 이 과정에서 SDV 에이전트는 로컬 토픽을 통해 차량 내부의 카메라, LiDAR 등의 센서 데이터를 관리하는 MS로부터 전달받은 데이터를 글로벌 토픽을 통해 엣지 에이전트에 전달하고, 엣지 에이전트는 이를 로컬 토픽을 통해 AI MS에 전달하여 추론을 수행한다. 추론 결과는 다시 엣지 에이전트를 통해 글로벌 토픽으로 출판되며 이를 구독하고 있는 각 SDV 에이전트는 로컬 토픽을 통해 차량 내부의 관련 MS에 제공한다.

그 결과 모든 SDV는 엣지에 존재하는 AI MS가 산출한 결과를 공유하여 일관된 서비스를 실시간으로 제공할 수 있다. 이러한 구조는 각 에이전트가 메시지 브로커 역할을 수행하고 디스커버리 기능을 통해 필요한 MS를 탐색 및 연결함으로써 SDV와 엣지 간 고성능 기능을 가진 서비스 재구성을 동적으로 지원할 수 있다. 또한 엣지 디바이스에 복합적인 AI MS를 구성하여 연산을 수행함으로써 SDV의 제한된 컴퓨팅 자원만으로는 어려운 복잡한 서비스도 엣지와 협력하여 구현할 수 있다. 이는 SDV의 경량화를 가능하게 하며 보다 다양한 지능형 서비스를 확장하는 데에도 기여할 수 있다.

Ⅳ. 결론

본 논문에서는 SDV 간 여러 AI를 활용한 서비스 재구성을 함에 따라 연산 한계와 확장성 문제를 해결하기 위해 SDV와 엣지 간 인터워킹을 활용한 AI 서비스 재구성 기법을 제안하였다. 제안 기법은 SDV와 엣지 간 에이전트를 통해 로컬 및 글로벌 토픽을 변환 및 연계함으로써 엣지 AI MS의 추론 결과를 다수의 SDV에 실시간으로 분배할 수 있도록 하였다. 이를 통해 SDV는 하드웨어 변경없이 연산량이 큰 지능형 서비스 재구성할 수 있다. 향후 연구에서는 제안된 기법을 실제 SDV 테스트베드 환경에 구현하여 시나리오에 적용하여 성능을 정량적으로 검증할 계획이다.

ACKNOWLEDGMENT

본 논문은 과학기술정보통신부 및 정보통신기획평가원의 정보통신방송 표준개발지원사업(RS-2024-00398199)과 부산광역시 및 (재)부산테크노파크의 BB21plus 사업으로 지원된 연구임

참 고 문 헌

[1] S. Halder, A. Ghosal, and M. Conti, "Secure OTA Software Updates in Connected Vehicles: A Survey," arXiv preprint arXiv:1904.00685, 2019.

[2] X. Li, Y. Dang, and T. Chen, "Vehicular Edge Cloud Computing: Depressurize the Intelligent Vehicles Onboard Computational Power," arXiv preprint arXiv:1901.03761, 2019.

[3] T. Kweon, J. Park, and J. Youn, "A Study on Techniques for Service Reconfiguration Between SDVs," Manuscript submitted for publication, 2025.

교통약자의 보행 안전을 위한 경사도 기반 보행 취약지역 분석 및 최적 경로 탐색에 관한 연구

임윤정[1,2], 송규원[1,*], 박호건[2,*]

[1]차세대융합기술연구원, [2]성균관대학교

{yunjeong.im, gyuwon.song}@snu.ac.kr, hogunpark@skku.edu

A Study on the Terrain-aware Path Planning Using Digital Elevation Models to Enhance Safety and Accessibility

Yunjeong Im[1,2], Gyuwon Song[1,*], Hogun Park[2,*]

[1]Advanced Institute of Convergence Technology, Suwon, Republic of Korea,
[2]Sungkyunkwan University, Suwon, Republic of Korea

요 약

본 논문은 수치표고모델(Digital Elevation Model, 이하 DEM)을 통해 산출한 경사도와 클러스터링 기법을 활용하여 가파른 경사가 빈번한 지점을 보행 취약지로 도출하고, 경사도와 거리 가중치를 고려한 A* 알고리즘 기반 안전한 경로 탐색 방법을 제안한다. 이를 위하여 실제 인천시 보행 네트워크 데이터를 사용하여 도로 구간별 경사도를 산출하고, 경사도 5%를 초과하는 취약 구간을 식별한다. 취약 구간을 DBSCAN 기법을 적용함으로써, 연속적인 경사 구간을 취약 클러스터로 도출하였다. 분석 결과, 인천시 424,271개의 보행로 중 약 15%를 차지하는 63,641개가 경사도 5%를 초과하는 것으로 나타났다. 또한, 경사도 가중치를 고려한 A* 알고리즘은 급경사 구간을 식별하고 회피할 수 있어 교통약자가 접근 가능한 경로를 생성한다. 실험 결과, 제안한 방법은 최단 거리 기반 A* 대비 거리를 크게 늘리지 않고도 가파른 경사에 패널티를 주어 물리적인 장벽을 크게 줄여주는 것으로 나타났다. 본 연구를 토대로 교통약자의 이동에 있어 안전하고 편리한 경로를 제공해줄 수 있으며, 도시 보행 안전 정책 수립의 기초자료로 활용될 수 있음을 시사한다.

Ⅰ. 서 론

최근 도시 환경에서의 보행 이동성은 단순한 편의성을 넘어 교통약자의 사회적 포용성과 직결되는 중요한 주제로 부상하고 있다. 기존 내비게이션 서비스가 제공하는 최단 경로에는 종종 급경사나 계단과 같은 물리적 장애물이 포함되어 있어 유용성과 실용성이 제한된다. 또한 경로 탐색에 관한 연구는 주로 A* 알고리즘을 사용하여 경로의 유효성을 보장하지만, 대부분 2D grid 환경을 전제로 설계되어 실제 지형이나 경사도와 같은 요인을 충분히 반영하지 못한다. 따라서 도시 보행 네트워크 내에서 교통약자 친화적이고 현실적인 무장애 경로를 산출하기 위해서는 지형 특성을 고려한 경로 탐색 기법이 요구된다.

본 연구는 DEM 기반 지형의 고도 정보를 반영한 보행 네트워크를 구축하고, 경사도가 5%를 초과하는 지점이 밀집된 지역을 교통약자 보행 취약지역으로 정의하기로 한다. 가파른 경사 구간이 밀집된 지역을 도출하기 위해 DBSCAN(Density-based spatial clustering of applications with noise) 기법을 활용하였다. 또한, 경사도 가중치를 반영한 A* 경로 탐색 알고리즘을 개발하였다. 개발된 알고리즘은 경사 수준에 따른 패널티를 비용 함수에 반영하여 가파른 경사를 회피하고, 보다 원활한 이동 경로를 도출하는 것을 목표로 한다.

Ⅱ. 본론

본 연구에서 활용한 데이터셋은 인천광역시에서 제공한 보행자 네트워크 데이터와 1m급 DEM을 결합하여 구축하였다. 보행자 네트워크 데이터는 노드와 링크로 구성되며, 각 노드는 고유 식별자와 위·경도 좌표를 포함한다. 각 링크는 길이, 보행 시설 위치(예: 계단, 보도턱, 엘리베이터 등)을 속성으로 가진다. DEM은 xdem 라이브러리를 이용해 경사도 래스터로 변환하였으며, 이를 통해 각 링크의 평균 경사도를 계산하여 네트워크에 할당하였다. 인천시 전체 네트워크에서 무작위로 1,000개의 출발지와 도착지 쌍을 추출하여 경로 탐색 시나리오를 구성하였다.

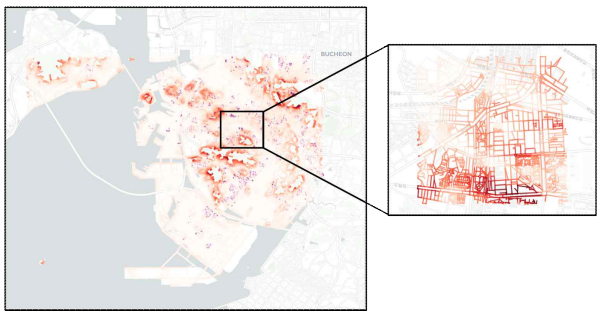


Figure 1. Illustration of the pedestrian network data incorporating DEM-derived slope information.

기존 A* 알고리즘은 유클리드 거리 기반 비용을 최소화하여 경로를 탐색하는 방식이다. A*는 OPEN 리스트에서 가장 작은 f(s) 값을 갖는 노드를 선택하여 확장해 나가며, 목표 노드에 도달할 때 최적 경로를 반환한다. 본 연구에서는 이를 확장하여 경사 정보를 반영한 A* 알고리즘을 개발하였다.

$$f(s) = g(s) + h(s) \quad (1)$$

$$g(s_j) = g(s_i) + d_{2D}(s_i, s_j) + \lambda \cdot |\Delta z| \quad (2)$$

$$\theta = \arctan\left(\frac{|\Delta z|}{d_{2D}(s_i, s_j)}\right) \quad (3)$$

여기서 $d_{2D}(s_i, s_j)$는 수평 거리, Δz는 고도 차, λ는 경사 페널티 계수를 의미한다. 또한, 경사각 θ가 미리 정의된 임계값(본 연구에서는 5%)를 초과할 경우, 해당 간선(edge)은 비가용 링크로 간주되어 탐색에서 배제된다.

Ⅲ. 실험 결과

본 연구에서는 예측된 경로가 접근성 제약을 얼마나 잘 준수하는지를 평가하기 위해 접근성 점수(Accessibility Score)를 사용하였다. 접근성 점수는 Jaccard 유사도 기반 지표로 정의된다. 여기서 P는 알고리즘이 탐색한 경로, P_{GT} 는 5%를 넘어가는 경사 구간을 제외한 접근 가능한 기준 경로를 의미한다. 점수가 높을수록 실제 환경의 지형 제약을 충실히 반영한 경로를 산출했음을 의미한다.

$$\text{Accessibility Score}(P, P_{GT}) = \frac{|P \cap P_{GT}|}{|P \cup P_{GT}|} \quad (4)$$

표 1은 제안한 알고리즘과 Standard A의 성능을 비교한 결과를 보여준다. 제안한 알고리즘은 총 1,000개의 탐색 시나리오에 대해 접근성 점수(Accessibility Score) 1.00을 달성하였다. 이는 모든 경로에서 5% 이상의 경사 구간을 완전히 회피했음을 의미한다. 반면 최단 거리 기반 Standard A*는 0.75 수준에 그쳐, 전체 경로의 약 25%가 여전히 비가용 구간을 포함하였다. 또한 평균 경사도는 제안 기법이 4.31%로, Standard A*의 5.45%보다 완만한 경로를 제공하였다. 이동 거리의 경우 제안 방법이 평균 551.4 m로 Standard A*의 539.4 m 대비 약 2.2% 증가에 불과하여, 접근성을 확보하면서도 효율적인 경로 탐색이 가능함을 입증하였다. 실험 결과, 경사도 기반 비용 함수와 접근성 평가 지표를 통해 도시 환경에서 실질적인 무장애 보행 경로 탐색이 가능함을 보여준다.

Table 1. Comparison of standard A* and slope-aware A* on accessibility, terrain slope, and path length.

Method	Acc. Score	Avg. Slope	Distance(m)
Standard A*	0.75	5.45	539.4
Ours	1.00	4.31	551.4

그림 2는 인천시 보행 네트워크에서 Standard A*와 제안한 알고리즘이 생성한 경로를 시각적으로 비교한 결과이다. 검은색 선은 Standard A*가 탐색한 경로를 나타내며, 파란색 선은 제안 방법이 탐색한 경로를 나타낸다. Standard A* 경로는 단순히 최단 거리를 우선시하기 때문에 경사가 5%를 초과하는 구간을 포함하고 있으며, 이는 휠체어 사용자와 같은 이동 제약이 있는 보행자에게 부적합하다. 반면 제안한 알고리즘의 경로는 다소 긴 거리임에도 불구하고 경사도가 낮고 계단을 회피하여, 실제 환경에서 무장애 경로 탐색이 가능함을 확인할 수 있다.

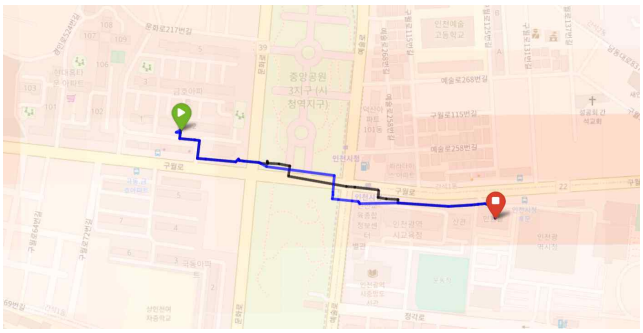


Figure 2. Comparison of paths generated by Standard A* (black) and slope-aware A* (blue).

Ⅳ. 결론

본 연구는 DEM 기반의 경사도를 인식한 A* 경로 탐색 알고리즘을 제안하고, 이를 실제 도시 보행자 네트워크에 적용하여 그 성능을 검증하였다. 제안 알고리즘은 Standard A*에 비해 접근성을 크게 향상시키면서도 이동 거리의 증가를 거의 유지하였으며, 결과적으로 이동 약자를 위한 현실적이고 포용적인 도시 보행 경로 탐색 가능성을 입증하였다.

향후 연구에서는 스트리트 뷰 및 항공 영상 등의 시각적 데이터를 활용하여 다양한 물리적 장애물을 반영하고, 비전-언어 모델을 활용한 장애물 인식 기능을 추가함으로써 보다 정교하고 실용적인 무장애 경로 탐색 체계를 구축할 예정이다. 이는 궁극적으로 도시 내 보행 약자의 이동권 보장과 자율주행 로봇의 안정적 운용을 동시에 실현하는 데 기여할 수 있다.

ACKNOWLEDGMENT

이 논문은 2024년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(No.RS-2024-00397607)과 한국연구재단의 지원을 받아 수행된 연구임 (No. RS-2024-00351159)

참 고 문 헌

[1] Davies R. W." The Data Encryption standard in perspective,"Computer Security and the Data Encryption Standard, pp. 129-132.

[2] Miles E. Smid, "From DES to AES," 2000, (http//www.nist.gov/aes).

[3] Shamir, A. "On the security of DES," Advances in Cryptology, Proc.Crypto '85, pp. 280-285, Aug. 1985.

[4] NIST, "Announcing the Advanced Encryption Standard(AES),"FIPS PUB ZZZ, 2001, (http://www.nist.gov/aes).

[5] Daemen, J., and Rijmen, V. "AES Proposal: Rijndael, Version2.," Submission to NIST, March 1999.

다중 LLM 교차 검증을 통한 선행 연구 검색 및 분석에서의 AI 환각 완화에 관한 연구

정한민
한국과학기술정보연구원, 과학기술연합대학원대학교
jhm@kisti.re.kr

Study on AI Hallucination Mitigation in Prior-Literature Search and Analysis Using Cross-validation across Multiple LLMs

Hanmin Jung
Korea Institute of Science and Technology Information, University of Science and Technology

요 약

생성형 AI의 환각 현상은 R&D 및 학술 영역에서 심각한 왜곡을 초래할 수 있어 그 탐지와 완화가 중요한 이슈이다. 본 연구는 다중 LLM의 교차 검증 구조를 기반으로, 선행 연구 검색 및 분석 과정에서 환각을 효과적으로 발견하고 제거하는 다단계 프롬프트 실험을 설계하였다. 실험 결과, ChatGPT-4o와 Grok-3 DeepSearch/Think 간의 교차 검증을 통해 환각이 발생한 선행 연구를 성공적으로 식별하였으며, 두 LLM이 완전히 동의한 결과는 환각 없이 신뢰할 수 있는 정보를 제공하는 데 기여함을 확인하였다.

Ⅰ. 서 론

AI 환각(Hallucination) 현상은 생성형 AI가 사실과 무관하거나 근거 없는 정보를 마치 사실처럼 만들어내는 현상으로, R&D 및 학술 영역에서 잘못된 정보가 포함된 요약, 문헌 생성, 또는 인용으로 인해 후속 연구 왜곡, 정책 결정 오류로 인해 실질적 피해가 발생할 수 있기 때문에 중요하게 다루어야 한다. LLM을 활용한 자동화된 문헌 분석이나 요약 도구가 빠르게 확산함에 따라, 연구자가 AI가 생성한 내용을 비판 없이 수용할 경우, 환각이 발생한 정보가 검증되지 않은 채 학술 생태계에 퍼지는 '2차 환각 전파'의 위험성도 증가할 수밖에 없다. 따라서 AI 환각을 탐지하고 완화하는 기술은 학술의 정확성과 연구 윤리를 보장하는 핵심 인프라로서 필요성이 크다. 특히, 참고문헌의 구조는 일정한 포맷(저자, 제목, 연도 등)을 따르기 때문에, 모델은 그럴듯한 형식을 자동으로 생성할 수 있지만, 실제 존재 여부에 대한 근거가 부족하므로 논문이나 책 제목을 그럴듯하게 만들어내는 경향이 있다[1-2].

환각을 줄이기 위한 연구가 활발하게 진행되고 있는데, [3]은 LLaMA-v2의 텍스트 생성 과정에서 발생하는 환각 현상을 시각적인 힌트를 통해 완화하는 방법을, [4]는 Nltk와 Rouge_scorer를 사용하여 QA 데이터의 성능을 정량적으로 분석하는 방식으로 결과의 정확도를 올리는 방법을 제안한다. 그렇지만, 이러한 방법들은 하나의 LLM에만 의존함으로써 학습 과정에서 원천적으로 완벽하지 않은 상황에 대처가 쉽지 않으며, 모델의 편향성을 해결할 방법이 없다. 이에 본 연구는 다중 LLM에 기반하여 교차 검증(Cross-validation)하는 방식으로 빈번하게 환각이 발생하는 선행 연구 검색 이슈를 해결하고자 한다.

Ⅱ. 본론

본 연구는 R&D 프로세스 내 '사전 단계' 중 '과거(선행) 연구 리뷰'를 효율적으로 수행하기 위한 목표 아래 다중 LLM의 교차 검증을 통해 환각을 최소화하는 방식으로 원하는 선행 연구를 정확하고 신속하게 검색과 분석하는 방안을 제시한다[5-6]. 실험은 "다중 LLM이 선행 연구 검색에 있어서 환각 현상을 줄여줄 수 있는가?" "다중 LLM이 선행 연구 내용에 대해 정확히 분석하는 데 도움을 줄 수 있는가?"라는 질문들에 대한 답을 얻기 위해 진행하였다. 본 실험을 위해, 4개의 프롬프트와 3개의 LLM(ChatGPT-4o, Grok-3 DeepSearch, Grok-3 Think)을 이용한다(표 1 참조). 첫 번째, 1번 프롬프트와 ChatGPT-4o를 이용하여 특정 조건에 맞는 선행 논문 10편을 검색하였다(표 2 참조). 두 번째, 2번 프롬프트와 Grok-3 DeepSearch를 이용하여 검색된 10편에 대해 교차 검증을 수행하여 환각이 발생했다고 의심되는 논문들을 제거하였다(표 2 참조). 세 번째, 3번 프롬프트와 ChatGPT-4o를 이용하여 검색된 10편에 대해 분석 과정을 수행하여 주요 기여점과 키워드 목록을 얻었다(표 3 참조). 네 번째, 4번 프롬프트와 Grok-3 Think를 이용하여 분석된 10편에 대해 교차 검증을 수행하여 환각 여부를 확인하였다.

표 1. 실험에 사용된 프롬프트와 LLM 목록

No.	Prompt	LLM
1	List 10 papers published in SCI(E) journals since 2023 that deal with the phenomenon of 'hallucinations in LLM', in order of citation index. List the author, paper title, source, year of publication, and citation index. If you disagree, please specify the reason along with the paper number.	ChatGPT-4o
2	The following papers are the top 10 papers published in SCI(E) journals since 2023 that deal with the 'hallucinations of LLM,' listed in order of citation index as suggested by ChatGPT-4o. Please review these papers to verify whether the sources actually exist, whether the sources are accurate, and whether the results align with the specified conditions (such as topic and publication year). Then, select only those papers that you fully agree with. + [Results of 1]	Grok-3 DeepSearch
3	Describe the main contributions of each of the 10 papers in one paragraph and select the five most important keywords for each.	ChatGPT-4o
4	The following is ChatGPT-4o's summary of the main contributions of 10 papers in one paragraph each, along with the five most important keywords for each. Please indicate which ones you fully agree with, partially agree with, or disagree with, and explain your reasons. + [Results of 3]	Grok-3 Think

표 2. 1번과 2번 프롬프트로부터 얻어진 LLM 결과

No.	Title	Publication	Year	Citation Index	Agreement	Reason for Disagreement
1	Survey of Hallucination in Natural Language Generation	ACM Computing Surveys	2023	~310	Fully Agreed	-
2	Siren's Song in the AI Ocean: A Survey on Hallucination in LLMs[1]	ArXiv	2023	~180	Partially Agreed	Not yet in journal

3	Detecting hallucinations in LLMs using semantic entropy	Nature	2024	~120	Fully Agreed	-
4	HILL: A Hallucination Identifier for Large Language Models	CHI	2024	~90	Fully Agreed	-
5	A Survey on LLM Hallucination via a Creativity Perspective	ArXiv	2024	~75	Partially Agreed	Not yet in journal
6	Hallucination is Inevitable: An Innate Limitation of LLMs	ArXiv	2024	~60	Partially Agreed	Not yet in journal
7	A Comprehensive Survey of Hallucination Mitigation Techniques in LLMs	Claimed IEEE/Elsevier	2024	~45	Disagreed	Not in journal, only on arXiv
8	Hallucination Detection and Mitigation: An Investigation	Claimed ACL/EACL	2024	~40	Partially Agreed	Not yet in conference or journal
9	On LLMs' Hallucination with Regard to Known Knowledge	NAACL	2024	~30	Fully Agreed	-
10	Quantifying Uncertainty in Answers and Enhancing Trustworthiness	ACL 2024	2024	~25	Fully Agreed	-

[1번 프롬프트로부터 얻어진 LLM 결과에 대한 Grok-3 DeepSearch 결과]

- Fully Agreed Papers (5): Papers 1, 3, 4, 9, and 10 meet all conditions, being published in high-impact venues since 2023 and focusing on LLM hallucinations.
- Partially Agreed Papers (4): Papers 2, 5, 6, and 8 are relevant and from 2023-2024, but they are currently only on arXiv, not yet in journals or top conferences.
- Disagreed Paper (1): Paper 7 does not meet the condition of being published in an SCI(E) journal, as it is only on arXiv despite the claim, leading to disagreement.

표 2에서 보듯이, Grok-3 DeepSearch에 의해 부분 동의(Partially Agreed)나 비동의(Disagreed) 된 논문 5편(2, 5, 6, 7, 8번) 중 2번 논문은 Grok-3 DeepSearch가 환각이 발생한 경우이며(빨간색), 7번 논문은 ChatGPT-4o가 환각이 발생한 경우이며(파란색), 5, 6, 8번 논문은 두 LLM 모두 환각이 발생한 경우(보라색)이다. 완전 동의(Fully Agreed) 된 논문 5편(1, 3, 4, 9, 10번)은 두 LLM 모두 환각이 발생하지 않은 경우이다. 표 3은 논문 10편을 대상으로 3번 프롬프트로부터 얻어진 결과 중, 두 LLM 모두 환각이 발생하지 않은 5편 논문의 주요 기여점과 키워드 목록을 보여준다. 4번 프롬프트를 이용하여 이 결과를 검증한 결과, Grok-3 Think 역시 모든 결과에 대해 완전 동의(Fully Agree) 의견을 제시하였다. 이는 문제 범위를 한정한 Retrieval-in-the-loop 구조가 사실 근거성을 높인다는 연구 결과를 뒷받침한다[7].

표 3. 3번 프롬프트로부터 얻어진 LLM 결과 중 일부

No.	Main Contributions	Keywords
1	Provides a foundational taxonomy of hallucination types (intrinsic vs. extrinsic), surveys evaluation metrics and mitigation methods across NLG tasks (summarization, QA, dialogue, MT, etc.), and highlights emerging challenges when scaling to transformer-based LLMs. Serves as a go-to reference for hallucination research.	intrinsic hallucination, extrinsic hallucination, evaluation metrics, mitigation strategies, faithfulness
3	Introduces semantic entropy as a robust signal for detecting hallucinations, demonstrating strong correlation between entropy spikes and factual inconsistency. Validated on GPT-series and open-source LLMs, it suggests real-time monitoring potential.	semantic entropy, detection, entropy spikes, factual inconsistency, real-time monitoring
4	Presents HILL, a user-centric tool that flags potential hallucinations via interface overlays, using model confidence and provenance signals. Empirical user studies show increased trust and reduced misinformation.	user interface, provenance logging, confidence score, human in loop, trust
9	Reveals that hallucinations occur even when knowledge is present in context—due to retrieval and attention misalignment. Proposes attention-aware retrieval refinement to improve grounding.	retrieval misalignment, attention bias, context grounding, knowledge retrieval, NAACL evaluation
10	Proposes probabilistic answer confidence scores applicable to any LLM, calibrates them on held-out factual queries, and shows using confidence thresholds significantly reduces hallucination rates.	uncertainty calibration, confidence scoring, probabilistic framework, calibration dataset, thresholding

Ⅲ. 결론

본 연구는 단일 LLM에 의존하는 기존 접근의 한계를 넘어, 다중 LLM을 활용한 교차 검증 방식이 특히 선행 연구 검색 및 분석에서 환각을 줄이는 데 효과적임을 확인하였다. 이처럼 다중 LLM 기반 교차 검증은 자동화된 문헌 조사에서의 정확성 확보뿐만 아니라, 후속 연구와 정책 결정의 신뢰도를 높이는 데 실질적인 기여를 할 수 있다. 향후 추가적인 LLM과 대용량 데이터 셋 기반 실험을 통해 좀 더 신뢰할 만한 환각 검색 및 분석 방식을 제안할 예정이다.

ACKNOWLEDGMENT

This work was supported by UST Young Scientist+ Research Program 2024 through the University of Science and Technology(No. 2024YS12).

참 고 문 헌

[1] Agrawal, A., Suzgun, M., Mackey, L., and Kalai, A. “Do Language Models Know When They’re Hallucinating References?,” In Proceedings of EACL 2024, pp. 912-928, Mar. 2024.

[2] Park, D. and Lee, H. “Literature Review of AI Hallucination Research Since the Advent of ChatGPT: Focusing on Papers from arXiv,” Journal of Information Policy 31(2), pp. 3-38, 2024.

[3] Noh, T., Yeo, H., and Han, K. “Using Deep Learning-Based Visual Hints to Mitigate Hallucinations in Large Language Model,” In Proceedings of KSC 2023, pp. 523-525, Dec. 2023.

[4] Bae, S. and Lee, D. “Exploring Hallucination Types in Question-Answering Generation and Limitation of Text Evaluation Metrics,” In Proceedings of KCC 2024, pp. 450-452, Jun. 2024.

[5] Jung, H., Jang, Y., Choi, G., Kim, H., and Park, J. “Scenario Design and Prototype Development to Support R&D Process,” In Proceedings of the KSCI 2023, pp. 131-132, Jan. 2023.

[6] Jung, H. “The Generation of ‘General Procedure’ and ‘Task-Oriented Procedure’ from LLM Reasoning Processes,” In Proceedings of the KSII 2025, Apr. 2025.

[7] Shuster, K., Poff, S., Chen, M., Kiela, D., and Weston, J. “Retrieval Augmentation Reduces Hallucination in Conversation,” In Proceedings of EMNLP 2021, pp. 3784-3803, Nov. 2021.

[1] Computational Linguistics, July 07 2025.

ZCD: LVLM 환각을 완화하는 Zero-Image Contrastive Decoding 기법

하지우, 소진현*
대구경북과학기술원

djiwoo20@dgist.ac.kr, *jinhyun@dgist.ac.kr

ZCD: Mitigating hallucinations in Large Vision-Language Models through Zero-image Contrastive Decoding

Jiwoo Ha, Jinhyun So*
DGIST

요 약

거대 시각-언어 모델에서 환각을 완화하는 문제는 모델의 신뢰도를 높이기 위해 반드시 해결되어야 할 과제이다. 본 논문에서는, 모든 픽셀 값이 0 으로만 채워진 영 이미지와의 대조적 디코딩을 활용한 환각 완화 기법인 ZCD(Zero-image Contrastive Decoding)를 제안한다. LLaVA-1.5 모델을 기반으로 한 실험 결과, ZCD 는 베이스라인 및 기존 완화 기법인 VCD 대비 POPE, MME, AMBER 벤치마크에서 더 높은 성능을 보였다. 이를 통해 ZCD 가 학습 없이 LVLM 의 객체 환각을 효과적으로 억제하는 디코딩 기법임을 입증하였다.

Ⅰ. 서 론

거대 시각-언어 모델(Large Vision-Language Model, LVLM)은 거대 언어 모델의 추론 능력을 활용하여 다양한 시각 과제를 처리할 수 있지만 여전히 환각(Hallucination) 문제에 직면해 있다. 환각은 주어진 시각 입력에 대해 LVLM 이 존재하지 않는 객체나 사실과 맞지 않는 관계·속성을 포함한 답변을 생성하는 현상을 뜻한다[5]. 이러한 현상은 의료 영상 분석, 자율주행 등 실제 응용에서 심각한 사고를 야기할 수 있다. 환각을 완화하기 위해 다양한 방법이 제안되었으며, 그중 VCD(Visual Contrastive Decoding) 기법은 학습 없이 환각을 완화하는 유용한 기법이다. VCD 는 원본 이미지 입력에 가우시안 노이즈를 더해 부정 로짓(negative logit), 즉 이미지와 관련 없이 언어 편향(language prior)에만 의존한 출력을 생성한다. 이후 이를 원본 로짓과 대조하여 모델이 언어 편향에만 의존하지 않고 시각적 정보를 더 반영한 출력을 생성할 수 있도록 한다[2]. 그러나 가우시안 노이즈가 추가된 이미지에도 여전히 원본 이미지의 정보의 일부가 남아있기 때문에, 완벽히 언어 편향에만 의존한 부정 로짓을 만들 수 없다[9]. 따라서 본 논문에서는 이러한 한계를 해결하기 위해 가우시안 노이즈를 더하는 대신, 원본 이미지에 대한 정보를 완전히 제거한 0 으로만 채워진 영 이미지(Zero-Image)와 대조적 디코딩을 하는 ZCD 기법을 제안한다.

Ⅱ. 배경 지식

θ로 모델링된 LVLM 은 시각 인코더를 통해 시각 입력 v를 처리하고, 텍스트 프롬프트 x와 결합하여 이에 대한 응답 y 를 생성한다. y 는 자기회귀적으로(auto-regressive) 생성되며, 각 시점 t 에서 토큰 y_t는 다음과 같이 확률적으로 선택된다.

$$y_t \sim p_\theta(y_t|v,x,y_{<t}) \propto \exp logit_\theta(y_t|v,x,y_{<t}) \quad (1)$$

여기서 $y_{<t}$ 는 $(t-1)$ 시점까지 생성된 토큰을 의미한다. $logit_\theta(\cdot)$ 은 후보 토큰에 대해 모델이 계산한 점수를 의미하며, 이후 $softmax$ 함수를 통해 확률 분포로 변환된다. 기존 연구에 따르면, 이러한 디코딩 과정에서 LVLM 은 시각 입력 v 보다 거대 언어 모델이 학습한 언어 편향에 의존하는 경향이 있다. 이러한 현상은 LVLM 의 객체 환각(object hallucination) 원인 중 하나로 지적되고 있다[2,8].

Ⅲ. 제안 기법

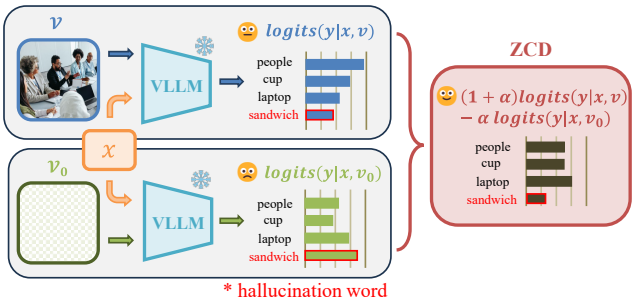


그림 1 ZCD 의 전체적인 진행 과정

본 논문에서 제안하는 ZCD(Zero-Image Contrastive Decoding)는 대조적 디코딩(Contrastive Decoding)과 적응형 개연성 제약(Adaptive Plausibility Constraint)의 두 단계로 구성된다. 그림 1 은 ZCD 의 전체 구조를 나타낸다.

가. 대조적 디코딩

ZCD 의 핵심 아이디어는 원본 이미지 v와 0 으로만 채워진 영 이미지 v_0를 각각 모델에 입력하여 두 개의 출력 분포를 생성하고, 그 차이를 활용해 시각 정보를 더 강조하는 것이다. 영 이미지는 모든 픽셀 값이 0 으로만 채워져 있어 추론에 필요한 시각 정보를 전혀 포함하지 않기 때문에 이에 기반한 출력 분포는 언어 편향에만 의존하게 된다. 이 두 분포의 차이를 이용해 VCD[2]의 식과 유사하게 새롭게 조정된 분포를 계산한다.

Setting	Method	Acc.↑	F1↑
Random	Regular	0.829	0.808
	VCD	0.848	0.834
	ZCD	**0.895**	**0.888**
Popular	Regular	0.811	0.792
	VCD	0.825	0.813
	ZCD	**0.873**	**0.868**
Adversarial	Regular	0.786	0.771
	VCD	0.799	0.791
	ZCD	**0.834**	**0.835**

표 1 POPE 결과. 가장 좋은 성능을 볼드체로 표기하였음.

Method	Acc.↑	Acc+↑	MME↑
Regular	0.68	0.46	114.20
VCD	0.71	0.50	121.44
ZCD	**0.77**	**0.59**	**136.31**

표 2 MME 결과. Accuracy+는 한 이미지에 대한 두 질문을 모두 맞춘 결과의 비율이며, MME score 는 Accuracy 와 Accuracy+를 더한 점수임.

Method	Generative task			
	CHAIR↓	Cover↑	Hal↓	Cog↓
Regular	11.9	49.6	48.8	4.4
VCD	9.8	51.2	43.3	4.7
ZCD	**8.0**	**53.3**	**40.6**	**4.3**

Method	Discriminative task				
	Acc.↑	Prec.↑	Recall↑	F1↑	AMBER↑
Regular	74.7	81.1	80.7	80.9	84.5
VCD	75.9	82.1	**81.3**	81.7	86.0
ZCD	**80.1**	**88.7**	80.2	**84.2**	**88.1**

표 3 AMBER 결과. CHAIR는 총 객체 중 환각 객체의 비율이며 Cover는 총 객체 중 정답 객체의 비율임. Hal은 환각 객체가 포함된 문장의 비율, Cog는 인간의 인지와 비슷한 환각 객체가 포함된 비율임. AMBER score는 CHAIR와 F1 점수를 종합적으로 고려한 지표임.

$$p_{zcd}(y|v, v_0, x) = softmax[(1+\alpha) logit_\theta(y|x, v) \\ -\alpha\, logit_\theta(y|x, v_0)] \quad (2)$$

이때, α는 두 출력 분포의 차이의 반영 정도를 결정하는 하이퍼파라미터(hyperparameter)로, $\alpha = 0$이면 일반적인 디코딩과 동일하게 작동한다.

나. 적응형 개연성 제약

영 이미지로부터 얻은 출력 분포는 왜곡된 시각 정보를 포함하고 있지만 기본적인 문법이나, 상식에 대한 정보 또한 여전히 포함하고 있다. 이러한 특성을 고려하지 않고 단순히 α를 통해 해당 출력 분포를 감쇠하는 것은 오히려 문법적으로 부자연스럽거나 의미가 부정확한 토큰을 선택하게 할 수 있다. 따라서 이러한 현상을 막기 위해 기존 연구[3]에서 제안된 방법과 유사하게 적응형 개연성 제약을 적용한다. 이 제약은 원본 시각 입력에 기반한 출력 분포에서 일정 수준 이상 신뢰할 수 있는 토큰 집합에서만 다음 토큰을 선택할 수 있도록 한다.

$$\mathcal{V}_{head}(y_{<t}) = \{y_t \in \mathcal{V}: \\ p_\theta(y_t|v, x, y_{<t}) \geq \beta \max_\omega p_\theta(\omega|v, x, y_{<t})\} \\ p_{zcd}(y_t|v, v_0, x) = 0, if\ y_t \notin \mathcal{V}_{head}(y_{<t}) \quad (3)$$

이때 $\mathcal{V}$는 LVLM 의 총 출력 어휘 집합이며, 토큰 후보 집합 $\mathcal{V}_{head}$는 원본 시각 입력에 기반한 출력 분포에서 확률이 $\beta \cdot$ *최대 확률* 이상인 토큰들로 정의된다. β는 필터링의 강도를 결정하는 하이퍼파라미터이다. β 값이 커질 수록 높은 확률인 토큰을 제외한 나머지 토큰을 다음 후보에서 제외하게 되므로 더 공격적인 필터링이 가능해진다. 두 기법에 기반한 ZCD 의 최종 디코딩은 다음과 같다.

$$y_t \sim softmax\{(1+\alpha) logit_\theta(y|x, v) \\ -\alpha\, logit_\theta(y|x, v_0)\} \\ subject\ to\ y_t \in \mathcal{V}_{head}(y_{<t}) \quad (4)$$

Ⅳ. 실 험

ZCD 기법의 성능을 검증하기 위해, 대표적인 LVLM 인 LLaVA1.5 모델[6]에서 실험을 적용하였으며, 다음 세 가지 설정을 비교했다.: (1) Baseline 인 기본 LLaVA1.5 (2) LLaVA1.5 + VCD (3) LLaVA1.5 + ZCD. 평가는 LVLM 의 환각 평가에 자주 사용되는 다음 벤치마크를 활용했다.: (1) 객체 존재 여부에 관한 POPE[4], (2) 14 개의 멀티모달 과제를 포함하는 MME[1], (3) 단답형, 생성형 벤치마크를 모두 포함하는 AMBER[7]. 디코딩 방식은 baseline 과 VCD, ZCD 방식 모두 샘플링 기반으로 진행하였다. 하이퍼파라미터 세팅은 VCD 의 경우 논문[2]의 세팅과 동일하게 $\alpha = 1.0, \beta = 0.1$ 로 설정하였으며, ZCD 의 경우 성능이 가장 높았던 $\alpha = 1.0, \beta = 0.4$ 로 설정했다. 모든 실험은 seed=55 에서 수행했다.

표 1 과 표 2 에서 확인할 수 있듯이, ZCD 는 POPE 와 MME 와 같은 discriminative task 에서 기존 baseline 과 VCD 기법보다 더 우수한 성능을 보였다. POPE 정확도에서 Baseline 대비 세팅 별로 6.6%p, 6.2%p, 4.8%p, VCD 대비 4.7%p, 4.8%p, 3.5%p 향상을 보였다. MME score 에선 baseline 대비 22.11 점, VCD 대비 14.87 점 향상하였다. 표 3 에서의 결과와 같이 AMBER 와 같은 생성형 벤치마크에서도 우수한 성능을 보여주었다. 생성형·단답형 과제 모두 반영한 AMBER score 에서 baseline 대비 3.6 점, VCD 대비 2.1 점 향상하였다.

이러한 결과는 영 이미지 기반으로 부정 로짓을 생성하는 ZCD 기법이 기존 기법인 VCD 보다 더 효과적으로 생성형·단답형 과제 모두 객체 환각을 제어할 수 있음을 보여준다.

Ⅴ. 결 론

본 논문에서는 거대 시각 언어 모델에서의 객체 환각을 완화하기 위해 이미지와 관련이 없는 영 이미지에 기반한 부정 로짓을 활용하는 ZCD 기법을 제안하였다. LLaVA1.5 기반 실험 결과 ZCD 는 베이스라인, VCD 기법대비 POPE, MME, AMBER 와 같은 다양한 벤치마크에서 일관적인 성능 향상을 보였다. 이는 ZCD 가 학습이나 추가적인 모델 없이도 LVLM 의 환각 완화에 강력하고 실용적인 방안이 될 수 있음을 보여준다. 향후 연구에서는 ZCD 를 다양한 디코딩 전략이나 환각 제어 기법과 결합하여 언어 편향을 더 효과적으로 제어하여 LVLM 의 성능을 더욱 강화할 수 있는 기법을 탐구하고자 한다.

ACKNOWLEDGMENT

이 논문은 2025 년도 정부(과학기술정보통신부)의 재원으로 2025 년 혁신거점 인공지능 데이터 융합과제 사업의 지원을 받아 수행된 연구임(S2201-24-1002)

This thesis was conducted with the support of the 2025 innovation base artificial intelligence data convergence project with the funding of the 2025 government (Ministry of Science and ICT) (S2201-24-1002)

참 고 문 헌

[1] C. Fu, P. Chen, Y. Shen, Y. Qin, M. Zhang, X. Lin, J. Yang, X. Zheng, K. Li, X. Sun et al., "Mme: A comprehensive evaluation benchmark for multimodal large language models,"arXiv:2306.13394, 2023

[2] Leng, Sicong, et al. "Mitigating object hallucinations in large vision-language models through visual contrastive decoding." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024.

[3] Li, Xiang Lisa, et al. "Contrastive Decoding: Open-ended Text Generation as Optimization." Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2023.

[4] Li, Yifan, et al. "Evaluating Object Hallucination in Large Vision-Language Models." *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*. 2023.

[5] Liu, Hanchao, et al. "A survey on hallucination in large vision-language models." *arXiv preprint arXiv:2402.00253* (2024).

[6] Liu, Haotian, et al. "Improved baselines with visual instruction tuning." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2024.

[7] Wang, Junyang, et al. "Amber: An llm-free multi-dimensional benchmark for mllms hallucination evaluation." *arXiv preprint arXiv:2311.07397* (2023).

[8] Yan, Hong, et al. "Overcoming language priors with self-contrastive learning for visual question answering." Multimedia Tools and Applications 82.11 (2023): 16343-16358.

[9] Zhao, Jianfei, et al. "Cross-image contrastive decoding: Precise, lossless suppression of language priors in large vision-language models." arXiv preprint arXiv:2505.10634 (2025).

LLM의 만진법 수치 처리 향상을 위한 전처리 방법론

김혜련, 목경서, 이호진, 이래형*
{helena1129, mok0726, leefine16, *rh}@kodata.co.kr

Preprocessing Methodology for Enhancing LLM Numeric Processing in Ten-Thousand System

Kim Hyeryeon, Mok Kyeongseo, Lee Hojin, Lee Raehyeong

요 약

대형 언어 모델(Large Language Model, LLM)은 자연어 이해에 뛰어난 성능을 보이는 반면, 재무 데이터의 수치 처리에서는 지속적인 한계가 관찰되고 있다. 본 연구는 만진법(萬進法) 체계가 LLM의 수치 처리에 미치는 부정적 영향을 완화하기 위해, 수치를 4자리씩 나누어 만(萬), 억(億) 등의 단위를 명시하는 전처리 방법론 Chunk4를 제안한다. 재무 데이터 기반 두 가지 태스크—수치 읽기(Spell-out)와 수치 변환(Conversion)에서 GPT-4o, Gemini2.5 Flash, Qwen2.5-72B를 대상으로 Raw, Comma, Chunk4 전처리 조건을 비교한 결과, Chunk4가 9–13자리의 '큰 수' 구간에서 정확도를 유의하게 개선하였다. 이는 만진법 체계와 LLM 간의 구조적 불일치로 인한 수치 처리 한계를 완화하고, 재무 데이터 분석의 신뢰성을 높이는 데 기여한다.

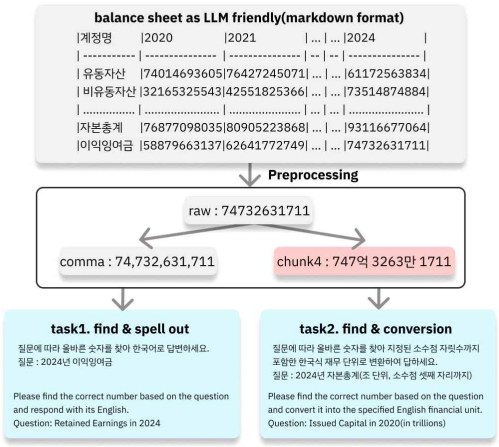
Ⅰ. 서 론

대형 언어 모델(Large Language Model, LLM)은 자연어 이해 및 생성 작업에서 뛰어난 성능을 보이지만, 수치 처리 영역에서는 지속적인 한계가 관찰되고 있다. 여기서 수치 처리란 숫자의 검색, 읽기(Spell-out), 변환 등의 작업을 포함한다[1]. 예를 들어 '123456'을 '십이만 삼천사백오십육'으로 읽거나, '856그램'을 '856000밀리그램'으로 변환하는 등의 작업을 지칭한다. 특히 재무 분야에서는 수치 처리 오류가 본래의 의미를 왜곡할 수 있어 정확성이 매우 중요하다.

LLM의 수치 처리 한계는 숫자의 크기가 커질수록 숫자 간 구분 능력이 떨어지는 경향성과 숫자의 토큰화 방향성에 기인한다. 첫째, LLM은 작은 숫자는 정밀하게 구분하지만, 큰 숫자로 갈수록 정밀한 구분 능력이 떨어진다[2]. 특히, 10^8-10^9구간에 정밀도가 떨어지는 현상이 관찰되었는데, 이는 재무 문서에서 자주 다루는 억(10^8)에서 조(10^{12}) 단위의 수치에서 오류가 발생할 가능성이 있음을 시사한다.

둘째, 숫자 토큰화의 방향이 수치 처리 성능에 큰 영향을 준다. GPT 계열의 기본 토크나이저는 긴 숫자를 왼쪽에서 오른쪽으로 세 자리씩 끊는 경향이 있어 자릿수에 대한 맥락을 잃어버릴 수 있다. 선행 연구[3]는 입력을 1,234,567처럼 표기해 오른쪽 기준으로 분절을 유도하면 숫자의 자릿수가 올바르게 정렬된 상태로 유지되어 산술 정확도가 크게 개선된다고 보고했다. 이는 모델을 재학습하지 않고도 입력 표현만 바꿔 성능을 높일 수 있음을 보여준다.

한편, 한자 문화권에서는 추가적인 도전 과제가 존재한다. 한자 문화권의 만진법(萬進法) 체계는 영어권의 천진법(千進法) 체계와 근본적으로 다른 명수법(命數法) 체계를 사용한다. 만진법은 수를 10,000 단위(조·억·만 등)로 구분하는 반면, 천진법은 1,000단위(billion · million · thousand 등)로 구분한다. 문제는 LLM이 주로 영어 기반 코퍼스로 학습되어, 만진법 체계에 대한 노출 빈도가 상대적으로 낮다는 점이다. 이에 따라 해당 체계와 관련된 숫자 표현이 충분히 학습되지 못해, 수치 해석에서 오류가 발생할 가능성이 높다. 실제로 중국어 연구에서는 이러한 성능 저하가 확인되었으나[1], 한국어 맥락에서 만진법 기반 수치 체계가 LLM에 미치는 영향을 실증적으로 검증한 사례는 여전히 부족하다.



[그림 1] 재무상태표를 Raw, Comma, Chunk4 형식으로 변환한 뒤, 수치 읽기와 수치 변환 태스크에 적용하는 과정

Ⅱ. 본론

Ⅱ.1. 제안 방법

본 연구에서 제안하는 Chunk4 방법론은 한국어의 만진법 수치 체계에 맞춰 설계된 전처리 방법론이다. 기존 연구[3]에서 효과가 입증된 3자리 콤마 구분법(예: '12,345,678')이 영어권의 천진법 체계를 반영한다면, Chunk4는 만진법의 4자리 단위 구분 방식에 최적화한 시도라는 점에서 의의가 있다. 구체적으로, 큰 숫자를 조, 억, 만 등의 단위로 분할하고 각 블록에 해당 단위명을 명시하는 방식이다. 예를 들어, '123456789'는 '1억2345만6789'로 변환된다.

Ⅱ.2. 실험 설계

본 연구는 Chunk4 방법론의 효과를 검증하기 위한 실험을 설계하였다. 연구에서 비교 대상으로 설정한 전처리 방법론은 (1) Raw: 원시 숫자 그대로 제시(예: 74732631711), (2) Comma: 선행 연구[3]에서 제시된 천 단위 쉼표 표기법(예: 74,732,631,711), (3) Chunk4: 숫자를 네 자리씩 분할하고 각 묶음에 해당하는 단위 명시(예: 747억3263만1711)의 세 가지이다.

전처리 방법론 간 차이를 검증하기 위해 [그림 1]과 같이 두 가지 태스크를 설계하여 실험을 진행하였다.

첫 번째, 수치 읽기(Spell-out) 태스크는 LLM이 재무상태표의 수치를 "칠백사십칠억 삼천이백육십삼만 천칠백십일"과 같이 자연어로만 이루어진 형태로 정확히 변환하는 능력을 평가한다.

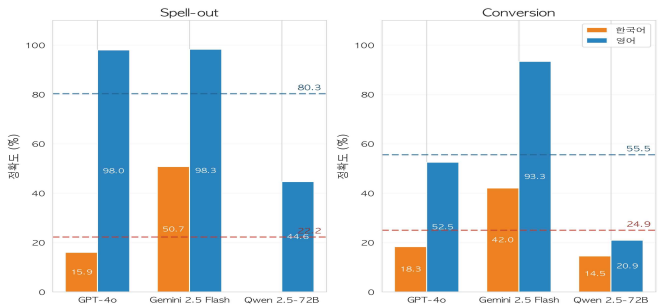
두 번째, 수치 변환(Conversion) 태스크는 LLM이 지정된 계정 항목의 수치를 정확히 식별하여 추출한 뒤, 주어진 수치 단위와 소수점 자릿수 규칙에 맞게 변환하는 능력을 평가한다. 예를 들어, "2024년 자본금을 억 단위로 소수점 둘째 자리까지 표시하시오."와 같은 요구에 LLM이 "931.16억"으로 정확히 추출하여 응답하는지를 평가한다. 이는 LLM이 만진법 체계를 이해하고, 금융권에서 통용되는 표현 방식 내에서 적절한 응답을 생성할 수 있는지를 종합적으로 평가한다.

데이터셋의 경우, 국내 상장사 115곳의 실제 재무상태표에서 추출한 숫자 데이터를 기반으로 한다. 이를 마크다운 형식으로 변환한 뒤, 프롬프트에 포함시켜 모델에 제공하였다. 수집된 재무 수치 데이터는 9-13자리 규모로, 10^8(억) 단위에서 10^{12}(조) 단위에 해당하는 범위에 분포하였다. 이 구간은 LLM이 수치를 정밀하게 구분하지 못하는 영역으로, 같은 간격 차이라도 작은 수치 구간 대비 구분 능력이 현저히 떨어지는 특성을 보인다.

모델의 경우, 상용 모델(GPT-4o, Gemini2.5 Flash)과 오픈소스 모델(Qwen2.5-72B)을 선정하였으며, 평가 기준은 금융권에서 통용되는 다양한 표현 방식을 모두 정답으로 인정하는 Pass Rate를 적용하여 실무 적용성을 제고하였다. 이를 통해 금융 문서 처리 과정에서 LLM이 직면하는 수치 인식의 한계를 실험적으로 검증할 수 있도록 구성하였다.

Ⅱ.3. 실험 결과 및 분석

먼저 한국어 환경과 영어 환경에서의 LLM의 수치 처리 성능 차이를 검증하기 위해 원시 숫자(Raw)에 대한 정확도를 분석하였다. 분석 결과, [그림 2]와 같이 성능에 유의한 차이가 있음을 확인하였다.



[그림 2] 언어별 정확도 비교

[그림 2]는 LLM의 수치 처리 정확도를 언어별로 비교한 결과를 보여준다. 수치 읽기 태스크에서는 영어 환경에서 평균 80.3%의 정확도를 기록한 반면, 한국어 환경에서는 22.2%에 그쳐 약 58%p의 성능 차이가 나타났다. 수치 변환 태스크에서도 언어 간 평균 약 30%p의 정확도 차이가 확인되었다. 이러한 경향은 모든 모델에서 일관되게 관찰되었으며, 이는 한국어 환경에서의 LLM의 수치 처리 성능이 상대적으로 낮아 한국어의 구조적 특성을 반영한 별도의 데이터 전처리 전략이 요구됨을 시사한다.

한국어 환경에서의 전처리 방법론 비교 실험에서는, 두 가지 태스크 모두에서 Chunk4 전처리 방법론을 적용했을 때의 평균 정확도가 높았다. [표 1]에서 보듯이, 수치 읽기 태스크에서는 전체 모델 평균 정확도가 Raw에서 22.22%, Comma 방법론에서 23.28%, Chunk4 방법론에서 45.70%로 나타나 Chunk4 방법론이 Raw 대비 23.48%p, Comma보다 22.42%p의 정확도 향상을 달성하였다. 수치 변환 태스크에서도 Chunk4 방법론이 평균 43.67%로 Raw 대비 18.74%p, Comma보다 13.82%p의 정확도 향상을 보여, 두 태스크 모두에서 일관된 성능 개선 효과를 확인하였다.

	Spell-out			Conversion		
	Raw	Comma	Chunk4	Raw	Comma	Chunk4
GPT-4o	15.94	12.75	50.15	18.26	30.43	57.68
Gemini2.5 Flash	50.72	57.10	77.68	42.03	48.41	45.51
Qwen2.5-72B	0.00	0.00	9.28	14.49	10.72	27.83
Average	**22.22**	**23.28**	**45.70**	**24.93**	**29.85**	**43.67**

(단위 : %)

[표 1] 한국어 환경에서의 전처리 방법론에 따른 태스크별 정확도 비교

모델별 세부 분석에서는 GPT-4o, Qwen2.5-72B 모델의 경우 두 태스크 모두에서 Chunk4 전처리 방법론이 가장 큰 개선 효과를 보였다. Comma 방법론은 Gemini2.5 Flash 수치 변환 태스크에는 가장 높은 개선 효과를 보였지만 Qwen2.5-72B에서 수치 읽기 태스크에서는 개선이 없었고, 수치 변환 태스크에서는 오히려 성능 하락(-3.77%p)을 보였는데 이는 천진법 기반의 구분이 한국어 수 체계 인식에 혼동을 유발할 수 있음을 시사한다.

Ⅲ. 결론

본 연구는 Chunk4 전처리 방법론을 적용한 입력 데이터 전처리 방식이 한국어 환경에서 효과적임을 입증하였다. 재무 데이터 기반 두 가지 태스크를 통해 Chunk4 전처리의 효용성을 검증했으며, 모든 구간에서 기존 Raw·Comma 방식 대비 일관된 성능 향상을 보였다. 이는 본 방법론이 한국어 수치 처리 문제 해결에 효과적임을 의미한다. 본 연구 결과는 한국어뿐만 아니라 중국어, 일본어 등 만진법 문화권 전반에 적용 가능할 것으로 기대된다. 특히 재무 분야에서 LLM 적용의 신뢰성과 실무적 안정성을 높이는 데 기여할 수 있다.

Ⅳ. 연구의 한계점 및 향후 연구 방향

본 연구는 LLM이 만진법 체계에서 겪는 수치 처리 오류를 완화하고, 안정성을 향상하기 위해 Chunk4 전처리 방법론을 제안하였으나, 적용 LLM 모델과 데이터 도메인이 제한적이라는 한계가 있다. 또한 실제로 금융 현장에 적용함에 있어 운영 적용 가능성이나 비용 - 효익 분석은 검토되지 못하였다. 향후 연구에서는 다양한 LLM 모델을 대상으로 한 검증, 과학·공공분야 등 타 도메인 확장, 실무적 효용성 평가를 통해 방법론의 일반화 가능성과 산업적 활용성을 강화할 필요가 있다.

ACKNOWLEDGMENT

본 연구는 한국평가데이터 주식회사의 지원을 받아 수행되었습니다.

참 고 문 헌

[1] A. Xu, M. Tan, L. Wang, M. Yang, and R. Xu, "NUMCoT: Numerals and Units of Measurement in Chain-of-Thought Reasoning using Large Language Models," 2024.

[2] H. V. AlquBoj, H. AlQuabeh, V. Bojkovic, T. Hiraoka, A. O. El-Shangiti, M. Nwadike, and K. Inui, "Number Representations in LLMs: A Computational Parallel to Human Perception," 2025.

[3] A. K. Singh and D. J. Strouse, "Tokenization counts: the impact of tokenization on arithmetic in frontier LLMs," 2024.

테스트 출력 예측을 위한 LLM 친화적 의사코드

김재진, 황승원
서울대학교

jaejin.kim@snu.ac.kr, seungwonh@snu.ac.kr

LLM-native Pseudocode for Test Output Prediction

Jaejin Kim, Seung-won Hwang
Interdisciplinary Program in Artificial Intelligence, Seoul National University

요 약

본 논문은 대규모 언어 모델(Large Language Models, LLMs)을 활용한 코드 생성(Code Generation) 과정에서, 테스트 출력 예측(Test Output Prediction) 성능을 향상시키기 위한 새로운 접근을 제안한다. 기존 연구들은 사람이 설계한 자연어 또는 의사코드(pseudocode) 형식을 활용하여 모델의 추론 과정을 유도했으나, 이러한 형식은 LLM 의 사전학습 분포와의 불일치로 인해 추론 성능을 저해할 수 있다. 이에 본 연구에서는 LLM 이 사전학습 중 익숙하게 접했을 가능성이 높은 표현 구조, 즉 LLM-native pseudocode 를 reasoning trace 로 활용하는 방법을 제안한다. 우리는 zero-shot 방식으로 LLM 에 자연스러운 pseudocode 형식을 탐색하고, 이를 few-shot 예시로 제시하는 간단한 프롬프트 구성만으로도 모델의 추론 안정성과 출력 정확도를 높일 수 있음을 실험적으로 확인하였다. LiveCodeBench 벤치마크를 기반으로 한 실험 결과, 제안된 방법은 기존 방식 대비 일관된 성능 향상을 보였으며, 특히 어려운 샘플에 대해서도 경쟁력 있는 결과를 달성하였다. 본 연구는 LLM-native 표현 방식이 코드 관련 추론 과제에서 효과적인 중간 표현이 될 수 있음을 시사하며, 향후 다양한 모델과 태스크로의 확장을 기대할 수 있다.

Ⅰ. 서 론

대규모 언어 모델(Large Language Models, LLMs)을 활용한 코드 생성(Code Generation) 연구가 활발히 진행되면서, 생성된 코드의 품질을 평가하거나 보완하기 위한 다양한 보조 과제들도 주목받고 있다. [4,5] 그 중 테스트케이스 생성(Testcase Generation)은 모델이 작성한 코드의 정확성과 일반화 능력을 검증하는 데 핵심적인 역할을 한다. [6]

테스트케이스 생성은 일반적으로 두 가지 하위 과제로 나눌 수 있다: 하나는 test input generation 이며, 다른 하나는 test output prediction 이다. 본 논문에서는 후자인 test output prediction 에 집중한다. 이 과제는 주어진 문제 설명(problem description)과 테스트 입력(test input)을 바탕으로, 그에 대응되는 출력(output)을 예측하는 태스크이다.

최근에는 이와 같은 추론 과정의 품질을 높이기 위해, Chain-of-Thought (CoT) [1] 방식을 활용하는 시도가 이루어지고 있다. CoT 는 모델이 최종 출력을 도출하기까지의 중간 추론 과정을 명시적으로 서술함으로써 정답률을 높이는 기법이다. 이 중간 reasoning chain 은 다양한 형식으로 표현될 수 있는데, 대표적으로 실제 코드(code) [2] 또는 의사코드(pseudocode) 형식이 사용된다. [3]

그러나 의사코드(pseudocode)는 실제 코드와 달리 명확하게 정해진 문법이나 구조가 없기 때문에, 모델이 중간에 생성한 의사코드의 표현 방식에 따라 성능이 민감하게 달라질 수 있다. 기존 연구에서는 사람이 설계한 형식을 few-shot 예시로 제공하거나, 일관된 스타일을 유도하기 위해 prompt 를 정교하게 구성하는 방식으로 이를 통제하려 했다. [3]

본 논문에서는 이러한 방식과는 다른 접근을 제안한다. 사람이 임의로 정의한 pseudocode 형식을 모델에게 강제하는 것보다, 사전학습(pretraining) 중 LLM 이 자주 접했을 법한 표현 방식, 즉 LLM-native 한 형식의 pseudocode 를 사용하는 것이 더욱 효과적이라는 것이다. 우리는 이러한 형식을 few-shot 예시로만 간단히 제시하더라도, 모델이 reasoning trace 를 더 자연스럽고 안정적으로 생성할 수 있음을 확인하였다.

즉, zero-shot 으로 LLM 에 친화적인 pseudocode 형식을 먼저 탐색한 뒤, 그 형식을 few-shot 예시 형태로 제시하는 것이, 전통적인 human-designed 형식에 기반한 prompting 보다 test output prediction 성능을 더 안정적으로 향상시킨다는 것이 본 연구의 핵심 주장이다.

Ⅱ. 과제 정의 및 접근 방식

Ⅱ-1. Test Output Prediction 과제 정의

Method	*Format*	*Pass@1*	*Easy-Pass@1*	*Medium-Pass@1*	*Hard-Pass@1*
Baseline	-	64.3	76.9	59.2	54.2
Constrained	Code	63.3	77.6	58.7	48.6
Constrained	T&E [3]	66.7	81.4	61.2	53.6
LPG	LLM-native	**67.6**	**81.6**	**62.8**	**54.2**

표 1. Test Output Prediction 과제에 대한 방법별 성능 비교
각 방법은 문제 설명과 테스트 입력을 바탕으로 정답 출력을 예측하는 과제를 수행하며, Pass@1, Easy-Pass@1, Medium-Pass@1, Hard-Pass@1 지표를 기준으로 평가되었다. Baseline 은 reasoning trace 없이 출력을 직접 예측하는 방식이며, Constrained 는 사람이 정의한 코드 또는 T&E(Trace-and-Execute) 형식의 pseudocode 를 활용하는 방식이다. LPG 는 본 논문에서 제안하는 LLM-native pseudocode 기반 접근법으로, 모든 난이도 구간에서 안정적이거나 우수한 성능을 보였다.

Test Output Prediction 은 문제 설명(problem description)과 특정 입력값(test input)이 주어졌을 때, 해당 입력에 대한 정답 출력을 직접 예측하는 과제이다. 이 과제는 실제 코드를 실행하지 않고도, 문제 해결 논리에 기반하여 정답 출력을 추론하는 것을 목표로 한다. 즉, 모델은 문제의 조건과 테스트 입력을 해석하고, 그에 따른 논리적 계산 및 흐름을 모델 내부에서 전개하여 출력값을 생성해야 한다.

본 과제는 기존의 코드 생성(Code Generation)이나 정답 선택(Answer Selection) 과제와는 달리, 출력값의 정확한 수치를 생성해야 하므로 reasoning 과정이 더 중요하게 작용한다. 이에 따라, 중간 reasoning trace 를 유도하는 Chain-of-Thought prompting 이 필수적으로 활용된다.

Ⅱ-2. LLM-native Pseudocode

기존 CoT 연구들은 중간 추론 과정을 자연어로 기술하거나, 사람이 설계한 pseudocode 형식을 삽입하는 방식으로 모델의 reasoning 을 유도하였다. 그러나 이러한 형식은 LLM 의 사전학습 분포와 괴리가 있는 경우, 오히려 모델의 혼란을 유발할 수 있다.

본 연구에서는 LLM 이 사전학습 중 익숙하게 접했을 가능성이 높은 표현 구조, 즉 LLM-native 한 pseudocode 형식을 사용하였다. 이를 위해 다양한 pseudocode 스타일을 탐색하고, 모델이 안정적으로 해석 및 생성할 수 있는 형식을 선택하였다. 이 형식은 특별한 규칙이나 키워드 없이, 자연스러운 파이썬 유사 문법과 논리적 구조만을 포함하며, 이를 few-shot 예시로 제공하였다.우리 접근은 다음의 특징을 가진다:

- 사람이 직접 설계한 형식이 아닌, 모델의 분포에 자연스럽게 호환되는 형식을 채택
- Zero-shot 으로 형식을 탐색한 뒤, 그 형식을 few-shot prompt 에 삽입하여 모델의 reasoning 을 유도
- 자연어 또는 실제 코드 대신 pseudocode 를 사용함으로써 추상화 수준을 조절하고 오류 가능성을 줄임

Ⅲ. 실험 및 결과

실험은 LiveCodeBench[7]의 test output prediction 벤치마크를 기반으로 수행되었으며, cutoff date 는 2023 년 5 월 1 일로 설정되었다. 생성 모델은 GPT-4-Turbo-2024-04-09 를 사용하였다.

표 1 은 self-consistency 없이 수행한 다양한 방법들의 성능을 비교한 결과이다. Baseline 은 reasoning trace 없이 직접 출력을 예측하는 방식이며, Constrained 는 사람이 설계한 pseudocode 형식을 사용하였다. LPG 는 본 논문에서 제안한 LLM-native pseudocode 방식이다. 우리 방식은 전 구간에서 기존 방식과 유사하거나 더 나은 성능을 보였으며, 특히 Hard-Pass@1 에서도 Baseline 을 상회하는 결과를 나타냈다.

Ⅳ. 결론

본 논문은 test output prediction 과제에서 reasoning trace 로 pseudocode 를 활용할 때, LLM 에 친화적인 형식(LLM-native)을 사용하는 것이 성능 향상에 유리하다는 점을 실증적으로 보여주었다. 사람이 설계한 형식을 따르기보다는, 모델의 사전학습 분포에 기반한 표현을 사용하는 것이 reasoning 안정성과 결과 정확도 모두에 긍정적인 영향을 줄 수 있다.

향후 연구에서는 다양한 모델과 과제에 대해 LLM-native 형식의 일반성을 검증하고, 해당 형식을 자동으로 탐색하거나 생성하는 방법론으로 확장할 수 있을 것이다.

.

ACKNOWLEDGMENT

이 논문은 2025 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 [NO.RS-2021-II211343,인공지능대학원지원(서울대학교)]

참 고 문 헌

[1] Wei, Jason, et al. "Chain-of-thought prompting elicits reasoning in large language models." Advances in neural information processing systems 35 (2022): 24824-24837.

[2] Li, Chengshu, et al. "Chain of code: Reasoning with a language model-augmented code emulator." arXiv preprint arXiv:2312.04474 (2023).

[3] Chae, Hyungjoo, et al. "Language models as compilers: Simulating pseudocode execution improves algorithmic reasoning in language models." arXiv preprint arXiv:2404.02575 (2024).

[4] Li, Kefan, and Yuan Yuan. "Large language models as test case generators: Performance evaluation and enhancement." arXiv preprint arXiv:2404.13340 (2024).

[5] Fakhoury, Sarah, et al. "Llm-based test-driven interactive code generation: User study and empirical evaluation." IEEE Transactions on Software Engineering (2024).

[6] Wang, Wenhan, et al. "Testeval: Benchmarking large language models for test case generation." arXiv preprint arXiv:2406.04531 (2024).

[7] Jain, Naman, et al. "Livecodebench: Holistic and contamination free evaluation of large language models for code." arXiv preprint arXiv:2403.07974 (2024).

Multi-hop 질의응답을 위한 Five WS 기반 RAG Retrieval 기법

정일균[1]
[1]고려대학교 SW·AI 융합대학원
jik9210@korea.ac.kr

Five Ws-Enhanced Retrieval for Multi-hop Question Answering in RAG

Ilgyun Jeong[1]
[1]Korea University Graduate School of SW·AI Convergence

요 약

본 연구는 RAG(Retrieval-Augmented Generation)의 Multi-hop 질의응답 성능 향상을 위해 5W1H(Who, What, When, Where, Why, How) 태그 기반 검색 방식을 제안한다. 기존 임베딩 기반 검색은 복합적 문맥 연결성을 단일 벡터로 표현하는 데 한계가 있으며, 지식그래프 기반 방법은 연산 비용이 높아 실용성이 떨어진다. 이에 따라 GPT 를 활용해 질의와 문서 조각에서 5W1H 태그를 자동 생성하고, 문장 유사도와 태그 유사도를 결합한 검색 점수를 산출하였다.
HotpotQA 데이터셋을 활용한 실험 결과, 제안 방식은 Baseline 대비 F1-score 가 14.3% 향상되었고, Precision@2 와 Recall@2 또한 각각 10%와 14.3% 개선되었다. 특히 태그 가중치가 0.2~0.3 일 때 최적 성능을 보였으며, 초기 검색 정확도(P@1, MRR)는 유지되었다. 이는 5W1H 태그가 Multi-hop 추론에서 효과적인 보조 신호로 작용함을 보여준다.

Ⅰ. 서 론

대규모 언어모델(LLM)과 결합된 RAG(Retrieval-Augmented Generation) 기법은 외부 지식을 활용하여 환각을 줄이고 답변의 신뢰도를 높이는 방식으로 널리 사용되고 있다.[1] 그러나 실제 질의응답 환경에서는 단일 사실만을 요구하는 Single-hop 질의보다, 여러 문서와 단서를 종합해야하는 Multi-hop 질의응답이 빈번하게 등장한다. 기존의 임베딩 기반 검색 방식은 이러한 Multi-hop 상황에서 질의의 복합적인 문맥 연결성을 단일 벡터로 표현하는데 한계가 있어 성능 저하가 발생한다.
이를 해결하기 위해 엔티티 관계 추론이나 지식그래프 기반 검색이 제안되었으나, 이러한 접근은 높은 연산 비용과 복잡한 사전 처리 과정을 요구하는 실용적 한계를 지닌다.[2]
따라서 보다 단순하면서도 Multi-hop 질의에 효과적으로 대응할 수 있는 새로운 검색 방식을 제안한다.
본 연구에서는 이에 대한 대안으로 Five WS(Who, What, When, Where, Why, How) 기반 태그 정보를 활용한 Retrieval 방법을 제안한다. Five Ws 는 사실을 구조적으로 표현하는 기본적인 프레임워크로, 질의와 문서 조각을 의문사 단위에서 매칭할 수 있도록 한다. 예를 들어 "루즈벨트는 언제 사망했고 언제 대통령이 되었는가?"라는 Multi-hop 질의는 단순 임베딩 비교만으로는 정보 간의 연결성을 파악하기 어렵다. 하지만 [Who: 루즈벨트], [When: 사망 시점], [When: 대통령 취임 시점]과 같이 태그를 부여하면, 검색 시스템이 필요한 정보 조각 간의 관계를 더 명확히 인식할 수 있다.
따라서 본 연구의 목표는 Multi-hop 환경에서 RAG 의 Retrieval 성능을 향상시키기 위해 Five Ws 기반 태그 정보를 결합한 검색 방식을 설계하고, 실험을 통해 그 효과를 정량적으로 검증하는 데 있다.

Ⅱ. 본론

2.1 데이터셋

본 연구에서는 Multi-hop 질의응답 태스크를 평가하기 위해 HotpotQA 데이터셋을 사용하였다. HotpotQA 는 단일 문서에 근거하는 질문뿐만 아니라, 두 개 이상의 문서를 종합해야 정답을 도출할 수 있는 Multi-hop 질문을 포함하고 있어 본 연구의 목적에 적합하다.

1) 데이터 규모: 약 11 만 개의 질의응답 쌍(Train 90k, Dev 7.4k, Test 7.4k)

2) 질문 특성: 비교형 질문(예: "Which magazine was founded earlier, X or Y?")과 연결형 질문(예: "Who is the father of the author of …?")을 포함

3) 문서 컨텍스트: Wikipedia 문서를 기반으로 하며, Supporting Facts 가 제공되어 검색된 문서와 답변 간 연결성을 평가할 수 있음

2.2 태그 생성

검색 성능 향상을 위해 각 질의와 문서 조각에 대해 Five Ws(Who, What, When, Where, Why, How) 태그를 GPT-4.1 nano 를 통해 자동 생성하였다. 태그 생성 과정은 다음과 같다.

1) 프롬프트 기반 추출

LLM 에게 구조화된 프롬프트를 제공하여, 입력 텍스트에서 5W1H 요소를 JSON 형식으로 추출하도록 하였다.

표 1. 태그 생성 프롬프트 및 결과

프롬프트	아래 텍스트에서 5W1H 정보를 추출해주세요. JSON 형식으로만 응답하세요.
	텍스트: {question|context}
	{
	"who": [],
	"what": [],
	"when": [],
	"where": [],
	"why": [],
	"how": []
	}

결과	질문: *"Which American-born Sinclair won the Nobel Prize for Literature in 1930?"* 추출 결과: { "who": ["Sinclair", "American-born Sinclair"], "what": ["Nobel Prize for Literature", "won"], "when": ["1930"], "where": [], "why": [], "how": []}

2.3 방법론

태그 기반 검색 기법

본 연구에서는 각 질문과 문서 컨텍스트에서 Five Ws(Who, What, When, Where, Why, How) 태그를 추출하여 검색에 활용하였다. 검색 점수를 산출하기 위해 두 가지 임베딩 모델을 사용하여 비교하였다:

1) all-mpnet-base-v2 (Sentence-Transformers, 768 차원)

2) text-embedding-3-small (OpenAI, 1536 차원)

이들 모델은 질문과 문서 컨텍스트, 그리고 태그를 임베딩 벡터로 변환하는 데 사용되었으며, 변환된 벡터를 기반으로 코사인 유사도를 계산하였다.

검색 기법은 다음 두 가지 방식으로 구분된다.

1) Baseline: 질문과 문서 컨텍스트 전체를 임베딩하여 문장 유사도(SentenceSim)만으로 검색을 수행한다.

2) Proposed: Baseline 의 문장 유사도에 더해, 질문 태그와 문서 태그 간의 태그 유사도(TagSim)를 결합하여 최종 검색 점수를 산출한다.

계산식은 다음과 같다.

$$Score = (1-\lambda) \cdot SentenceSim + \lambda \cdot TagSim$$

여기서 λ는 태그 유사도의 가중치이며, 실험에서는 0.0, 0.2, 0.3, 0.4, 0.5 값을 적용하여 비교하였다.

Ⅲ. 실험결과

본 절에서는 제안한 5W1H 태그 기반 접근법의 효과를 검증하기 위해 수행한 실험 결과를 제시한다.

성능 평가는 아래의 4 개 지표를 사용하였다.

1) Precision@k (P@k): 상위 k 개 문서 중 정답 문서의 비율

2) Recall@k (R@k): 정답 문서가 상위 k 개에 포함되는 비율

3) F1 Score: Precision 과 Recall 의 조화 평균

4) Mean Reciprocal Rank (MRR): 정답 문서의 평균 역순위

3.1 Weight 변화에 따른 성능 분석

표 2 은 Weight 값을 0.0 에서 0.5 까지 변화시켰을 때의 성능 변화를 보여준다.

F1-score 는 Weight=0.2 및 0.3 에서 0.800 으로 최대치를 기록하였으며, Baseline(Weight=0.0)의 0.700 대비 14.3% 향상되었다.

Precision@2 와 Recall@2 또한 동일한 Weight 구간에서 0.800 을 기록하며 Baseline(0.700) 대비 각각 10% 및 14.3% 개선되었다.

반면, Weight 가 과도하게 증가한 경우(0.4 이상)에는 성능이 다시 하락하는 역 U 자 형태의 곡선이 관찰되었다. 이는 태그 가중치가 부족하거나 과도할 경우 모두 성능 저하로 이어짐을 시사한다.

Precision@1 과 MRR 은 Weight 값에 관계없이 각각 0.900, 0.950 으로 안정적으로 유지되었다. 이는 제안 기법이 초기 검색 결과의 품질에는 영향을 주지 않으면서, 다중 hop 검색 단계에서 성능을 집중적으로 개선함을 의미한다.

표 2. Weight 변화에 따른 성능 비교

Weight	F1	MRR	P@1	P@2	R@2
0.0	0.7	0.95	0.9	0.7	0.7
0.2	0.8	0.95	0.9	0.8	0.8
0.3	0.8	0.95	0.9	0.8	0.8
0.4	0.7	0.95	0.9	0.7	0.7
0.5	0.65	0.95	0.9	0.65	0.65

3.2 Baseline 대비 5W1H 태그 적용 효과

표 3 은 Baseline(Weight=0.0)과 5W1H 태그 적용 결과(Weight=0.2~0.3 구간)를 비교한 결과를 요약한 것이다.

Precision@1 은 두 방법 모두 0.900 으로 동일하여 초기 검색 정확도는 유지되었다.

Precision@2 는 0.700 에서 0.800 으로 10% 향상되었으며, Recall@2 와 F1-score 는 모두 0.700 에서 0.800 으로 증가하여 14.3%의 개선 효과를 보였다.

표 3. Baseline 대비 5W1H 태그 적용 성능 개선

Metric	Baseline	5W1H Tags	Improvement
P recision@1	0.900	0.900	0%
Precision@2	0.700	0.800	+ 10%

Recall@2	0.700	0.800	14.3%
F1-Score	0.700	0.800	14.3%

3.3 결과 해석

실험 결과, 5W1H 태그는 초기 검색 품질을 저해하지 않으면서 multi-hop 추론 성능을 개선하는 효과를 나타냈다. 특히 Weight=0.2~0.3 구간에서 최적의 성능 향상이 관찰되었으며, 이는 적절한 수준의 태그 정보가 문서 검색의 맥락적 일관성을 강화하는 데 기여함을 보여준다. 반면, 과도한 가중치는 불필요한 잡음을 유발하여 오히려 성능을 저하시켰다. 이러한 결과는 5W1H 태그가 단일 문서 검색보다 다중 hop 기반의 추론 문제에 특화된 보조 신호로 작동함을 시사한다.

Ⅳ. 결론

본 연구는 Multi-hop 질의응답에서 RAG 의 한계를 극복하기 위해 5W1H 기반 태그 정보를 활용한 검색 방식을 제안하였다. HotpotQA 실험 결과, 제안 기법은 Baseline 대비 F1-score 와 Recall@2 에서 최대 14.3% 향상을 보였으며, 초기 검색 정확도는 유지되었다. 이는 5W1H 태그가 Multi-hop 상황에서 문맥적 일관성을 강화하는 효과적 신호임을 입증한다. 다만, 태그 품질이 LLM 성능에 의존하고, 데이터셋이 제한적이라는 한계가 존재한다. 향후 연구에서는 다양한 도메인 검증과 동적 가중치 조정 방식을 탐구하여 제안 기법의 범용성과 실용성을 높일 예정이다.

참 고 문 헌

[1] P. Lewis, E. Perez, A. Piktus, et al., “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks,” Advances in Neural Information Processing Systems (NeurIPS), 2020.

[2] X. Zhu, Y. Xie, Y. Li, and W. Hu, “Knowledge Graph-Guided Retrieval Augmented Generation,” arXiv preprint arXiv:2502.06864, 2025.

LLM 기반 최적화 모델 자동 생성 파이프라인

강윤환, 조한얼*, 김찬수*, 이상철

1 한국과학기술연구원 인공지능·정보·추론 (AI/R) 연구실

LLM-based Optimization Modeling Automation Pipeline

Yoonhwan Kang, Haneol Cho*, Chansoo Kim*, Sangchul Lee

1 AI·Information·Reasoning (AI/R) Laboratory, Korea Institute of Science and Tech. (KIST)

요 약

본 연구는 전문가의 지식이 필수적인 수리 최적화 모델링과 코드화를 자동화하는 AI 기반의 지능형 파이프라인을 제안하고 그 성능을 검증한다. 여기에서는 대규모 언어 모델(LLM, Large Language Model)을 엔진으로 삼고 LangGraph를 프레임워크로 사용한다. 동 연구에서 개발한 파이프라인은 "생성-검증-피드백"의 순환적 자기 교정 구조를 코드 실행 시 발생하는 오류나 비가용성 문제를 스스로 해결할 수 있다. GPT-4o mini와 Llama-3 8B 모델을 대상으로 500개의 문제 데이터셋을 활용한 성능 평가를 통해, 약 87%의 높은 성공률을 얻어 실효성을 입증했다. 본 연구는 LLM이 단순 생성 도구를 넘어 능동적인 문제 해결 에이전트로 기능할 수 있는 가능성을 제시하고, 수리 최적화 기술의 새로운 접근법을 제안한다는 점에서 큰 의의가 있다.

Ⅰ. 서 론

최적화 기술은 생산 공정, 물류, 인력 배치 스케줄링 등 다양한 산업 분야에서 비용을 절감하고 효율성을 극대화하는 핵심 기술로 사용되어 왔다. 하지만 현실의 복잡한 문제를 수리적 모델로 정형화하고 이를 최적화 모델에 맞게 코드화 하는 과정은, 해당 분야의 전문 지식과 수리적 모델링 역량을 동시에 요구하는 매우 까다로운 작업이다. 이러한 이유로 여기에는 높은 시간 및 비용부담과 전문가 의존성으로 인해 보편적인 확대가 저해되어왔다.

최근 LLM 기술은 빠르게 발전하며, 자연어를 이해하고 복잡한 논리를 수학적 모델로 변환하는 능력을 보여주어 본 연구에서 적극적으로 활용하였다.

본 논문은 이러한 배경 하에, LLM을 핵심 엔진으로 사용하고 LangGraph 프레임 워크를 통해 전체 워크플로우를 구축함으로써, 수리 최적화 전과정을 자동화하는 새로운 파이프라인을 제시한다.

Ⅱ. 본론

최근 몇 년간, LLM은 문장과 언어를 기반으로 한 생성형 도구로 자리 잡았다. 전통적으로 전문가의 영역이던 최적화 모델링을 대규모 언어 모델(LLM)로 자동화하려는 시도가 최근 목격된다. LLM을 활용하여 사용자의 자연어 문제 기술을 바탕으로 수학적 모델을 자동 정형화하고[1, 2], 이를 특정 최적화 모델에서 실행 가능한 코드로 구현하는 접근법이 제시되었다 [3, 4]. 본 연구에서 제안하는 파이프라인은 LLM의 수리모델 및 코드 생성 능력을 기반으로, LangGraph를 프레임워크로 활용하여 순차적 모델 생성 및 피드백 루프의 자동화 과정을 구현한다.

이 파이프라인은 "문제 이해 → 모델 생성 → 코드 구현 → 검증 및 피드백→ 모델 재생성 → 조건만족시 종료" 이라는 순환적 구조를 통해, 기존의 단방향성 구조보다 높은 성능 및 정확성을 달성한다.

2.1. 파이프라인 구조 및 구성 요소

시스템은 각 기능이 유기적으로 연결된 모듈 형태로 구성된다. 전체 워크플로우는 LLM 에이전트 구조 설계에 적합한 LangGraph 프레임워크를 사용하여 상태를 공유하는 그래프 형태로 관리되며, 각 단계의 결과가 다음 단계의 입력으로 전달된다.

첫 단계로, 사용자의 문제 시나리오가 입력되면 RAG (Retrieval Augmented Generation) 기술을 통해 사전 구축된 최적화 문제 벡터 데이터베이스에서 가장 유사한 '문제-모델' 예제를 검색한다. 이렇게 검색된

컨텍스트는 LLM이 문제의 핵심을 명확히 파악하고 표준화된 형식의 수학 모델을 생성하도록 돕는다.

다음으로, 강화된 컨텍스트를 바탕으로 LLM은 목적 함수, 결정 변수, 제약 조건을 포함한 최적화 모델을 정형화하고, 이를 특정 최적화 솔버에서 실행 가능한 코드로 자동 변환한다.

마지막 단계로 파이프라인의 핵심인 실행 기반의 자동 검증 및 자기 교정 루프가 작동한다. 생성된 코드는 실제 실행 환경에서 구동되어 최적해 발견, 비가용성 확인, 오류 발생 등 모델의 상태를 실시간으로 확인한다. 만약 실행 결과가 비가용으로 판명되거나 오류가 발생하면, 시스템은 해당 실패 정보를 개선을 위한 피드백으로써 기존 최적화 모델과 함께 LLM에 다시 전달한다. LLM은 이 피드백을 바탕으로 제약 조건 완화 등의 수정을 거친 새로운 수학 모델을 생성하며, 이 모델은 다시 코드화 및 검증 단계를 거친다.

이러한 순환적 수정 과정은 LangGraph의 조건부 분기 통해 작동하며 시스템이 스스로의 실수를 통해 학습하여 유효한 해를 찾아 나가는 능동적인 문제 해결을 가능하게 한다.

2.2 실험 설계 및 결과

설계한 파이프라인의 실효성을 검증하기 위해 구체적인 실험환경을 설정하고, 정량적인 성능 평가를 수행하였다.

2.2.1 실험 환경 및 모델

- 베이스라인 모델: 기본적인 파이프라인의 성능을 확인하기 위해 OpenAI의 GPT-4o mini 모델 API를 베이스라인으로 사용
- 로컬 비교 모델: API 모델의 한계를 파악하고 파인 튜닝의 가능성을 탐색하기 위해, Llama-3 8B 모델을 로컬 환경에 구축하여 비교
- 프레임워크: 모델들은 LangGraph 기반의 워크플로우 내에서 동작하며, 비가용성 및 에러가 발생했을 때 이를 해결하기 위한 수정루프가 설정. 특히 비가용성 결과가 발생하면, 시스템은 ISS(Irreducible Inconsistent Subset) 분석을 통해 만족하지 못한 제약 조건을 파악하고, 이를 프롬프트에 추가하여 수학적 모델을 재생성

2.2.2. 데이터셋 및 평가 기준

- 데이터셋: 연구를 위해 총 500개의 "텍스트 시나리오와 정답(최적해)"으로 구성
- 평가 기준: 성공률을 핵심 지표로 한 성능지표로 선정. 생성된 코드가 도출한 최적 목적 함수 값이 데이터셋의 정답과 일치하는 경우를 성공으로 정의

2.2.3. 실험 결과 및 분석

데이터셋 테스트 결과, GPT-4o mini API를 사용한 베이스라인 모델은 약 87%의 성공률을 기록했다. Llama-3 8B 로컬 모델은 약 75%의 성공률을 보였다.

또한 파이프라인의 수정 루프가 필요한 여러 문제에서 시스템이 자율적으로 문제를 해결하는 것을 확인했다. 예를 들어 149번 문제의 경우, 최초 시도에서는 정답을 도출하지 못하였지만 두 번의 수정을 거친 후 정답을 도출하여 자기 수정 루프가 효과적으로 작동함을 확인하였다.

<표 1> 테스트셋 모델 별 실험 결과

	GPT-4o-mini w/ Feedback-loop	Llama-3 8B w/ Feedback-loop	GPT-4o-mini	Llama-3 8B
Success Rate	87%	75%	80%	71%
Pass Rate	99%	98%	81%	75%

Ⅲ. 결론

본 연구는 전문가 의존성이 높았던 수리 최적화 모델링 과정을 자동화하기 위해 LLM 기반 LangGraph 구조의 지능형 파이프라인을 제안하고 그 실효성을 검증했다. 제안된 파이프라인은 GPT-4o mini 모델기반으로 약 87%의 높은 성공률을 달성했으며, 특히 "생성-검증-피드백"의 순환적 자기 수정 루프 구조가 실제 문제 해결 과정에서 효과적으로 작동한다. 이는 우리의 연구에서 제안한 프레임워크가 단순히 정형화된 모델을 생성하는 단계를 넘어, 실행 결과를 바탕으로 스스로의 오류를 진단하고 수정하는 능동적인 문제 해결 주체로 기능할 수 있음을 보여준다.

ACKNOWLEDGMENT

This research was funded by the grant Nos. 2023-00262155; 2024-00339583; 2024-00460980; and 2025-02304717 (IITP) funded by the Korea government (the Ministry of Science and ICT).

참 고 문 헌

[1] A. Notz, D. T. Zhang, and N. P. Sudermann-Merx, "Unveiling the Potential of Large Language Models in Formulating Mathematical Optimization Problems," *arXiv preprint arXiv:2310.05210*, 2023.

[2] J. Hakizimana, A. Oustry, and L. Tible, "Optimus: Optimization Modeling Using MIP Solvers and LLM," *arXiv preprint arXiv:2402.13412*, 2024.

[3] G. Tack, C. N. Chou, M. D. SumAj, and K. N. Brown, "Autoformulation of mathematical optimization models using LLMs," *arXiv preprint arXiv:2309.09242*, 2023.

[4] K. Yang, W. H. Chen, J. H. Fu, S. Q. Liu, H. I. Ting, and C. J. Wang, "OR-LLM-Agent: Automating Modeling and Solving of Operations Research Optimization Problem with Reasoning Large Language Model," *arXiv preprint arXiv:2404.09825*, 2024.

SISEON: AIoT 인체공학 디스플레이 자동 제어 시스템

배제언, 차성현, 유형탁, 최인혁, 최혁진, 김형석
{jeeon98b, cjn112070, air2heo, inhyuck.choi0719, jjin6573, henna6967}@gmail.com

SISEON: AIoT-based Ergonomic Display Automatic Control System

Bae Je Eon, Cha Sung Hyun, Yu Hyeong Tak, Choi In Hyuk, Choi Hyeok Jin, Kim Hyoung Seok
Samsung Software·AI academy for Youth (SSAFY)

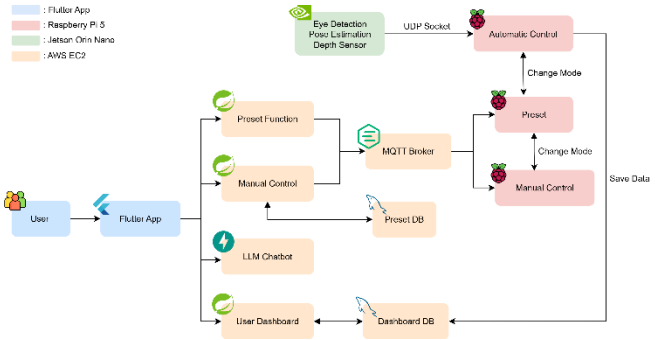
요 약

본 논문에서는 사용자의 눈 위치와 자세를 실시간으로 인식하여 모니터의 위치(X/Y/Z 축)를 자동으로 조정하는 인체공학 AIoT 디스플레이 제어 시스템(SISEON)을 제안한다. 제안 시스템은 YOLOv11 기반 AI 추론 모듈(Jetson Orin Nano), IoT 제어 모듈(Raspberry Pi 5), 4 관절 로봇 팔, Spring Boot API 서버 및 Flutter 모바일 앱으로 구성된다. 도메인 갭 해소를 위한 3단계 전이 학습을 적용한 시선 추적과 상체 중심의 경량화된 자세 추정 알고리즘을 통해 실시간성과 정확성을 확보했으며, 디스플레이 자동 조정, 프리셋, 자세 통계 등 통합 솔루션을 제공한다. 시제품 구현을 통해 시스템이 높은 정밀도로 사용자 작업 환경을 효과적으로 개선함을 입증했다. (코드 및 시현 영상: https://github.com/siseon-io/siseon)

Ⅰ. 서 론

원격 근무 및 스마트 오피스 환경이 보편화됨에 따라 모니터의 높이·거리·각도를 실시간으로 최적화하여 사용자의 자세를 보호하는 기술의 중요성이 날로 증가하고 있다[1]. 그러나 기존의 수동 모니터 암은 사용자의 자세 변화에 즉각적으로 대응하지 못해 인체공학적 최적 위치를 지속적으로 유지하기 어렵다.

이에 본 논문에서는 AIoT 기술을 활용해 사용자의 시선 및 자세를 실시간으로 인식하고 모니터 위치를 자동으로 조정하는 시스템 'SISEON'을 제안한다. 제안된 시스템은 NVIDIA Jetson Orin Nano[2] 기반의 실시간 AI 추론과 IoT 제어 기능을 결합하여 높은 사용자 편의성과 확장성을 동시에 확보한다.



[그림1: SISEON 시스템 아키텍처]

Ⅱ. 시스템 설계

SISEON은 AI 추론, IoT 제어, 백엔드, 모바일 앱으로 이어지는 4계층 파이프라인(그림 1)으로 구현된다.

(1). AI 추론 모듈(Jetson Orin Nano): YOLOv11[3] 경량 모델로 눈·상체 자세를 실시간(50ms/frame)에 추론해 좌표를 Raspberry Pi 5[4]에 전송한다.

(2). IoT 제어 모듈(Raspberry Pi 5): AI 모듈로부터 수신한 데이터와 Depth 센서 정보를 종합하여 4 관절 로봇 팔(Dynamixel+ U2D2)[5]을 ROS2로 구동하며 프리셋·수동 명령을 실시간 반영한다.

(3). Backend(Spring Boot, EC2): 사용자·디바이스·프리셋·자세 통계를 RDB에 관리하고, JWT 인증과 MQTT ACL로 전 구간 보안을 유지한다.

(4). Frontend(Flutter App): 모바일 UI에서 프리셋 설정, 수동 제어, 실시간 모니터링, 자세 통계 대시보드, 사용자 매뉴얼 챗봇을 제공해 사용자 편의성을 높인다.

Ⅲ 핵심 기술 구현

3.1 시선 추적모델: 3-Stage Transfer Learning

일반 이미지와 눈 클로즈업 이미지 간의 '도메인 갭' 문제를 해결하고자 점진적 3단계 전이 학습을 적용했다.

(1). 범용 특징 학습: 대규모 데이터셋[6]인 COCO로 YOLOv11 모델을 사전 학습하여 객체 검출에 필요한 기본적인 시각 특징 추출 능력을 확보한다.

(2). 도메인 특화 학습: 얼굴 관련 데이터셋[7] (iSEEDS)을 활용하여 도메인 간극을 1차적으로 해소한다.

(3). 눈4 영역 정밀 파인튜닝: 고해상도 눈 데이터셋[9] (Eye-D)을 사용하여 좌/우안, 및 홍채의 4개 영역을 정밀하게 조정함으로써 검출 정확도를 극대화한다.

3.2 자세 추정 모델: 상체 중심 키포인트 학습

디스플레이 제어에 필요한 상체 정보만 추출하도록, 기존 17~25개 키포인트를 10개(목·양어깨·팔꿈치·손목·가슴·등·머리)로 축소해 학습을 재설계했다. 이 10개의 키포인트는 사용자의 머리 높이와 상체 방향 벡터를 계산하는 데 충분한 정보를 제공해 인체공학적 모니터 위치 제어에 최적화되어 있다. 학습 파이프라인은 COCO 데이터셋으로 사전학습을 진행한 후, 상체 중심 데이터가 풍부한 FLIC[8] 데이터셋으로 미세 조정하여 최적화했다.

3.3 자세 평가 및 피드백 알고리즘

추정된 키포인트 (x, y, z)는 AI 모듈의 2D 좌표(x, y)와 Depth 센서의 거리(z)를 시간동기·외부보정으로 융합(RGB-D) 해 산출하였다. 평활화·좌표계 변환 후 목 굴곡/어깨 전인/상체 기울기 지표를 계산해 [10] 임계치로 판정하고, 지속 편차 시 모니터 ΔX/ΔY/ΔZ를 자동 보정하며 결과를 앱에 기록한다.

Ⅳ. 성능 평가

4.1 모델 성능

AI 모델의 성능 평가 결과는 [표1] 과 같다. 3단계 전이 학습을 적용한 시선 추적 모델은 2.5 px의 낮은 홍채 중심점 오차(MSE)를 달성했으며, 자세 추정 모델 또한 높은 정확도(AP@0.5=0.65)를 보였다.

4.2 시스템 주요 성능 지표

(1). 실시간성: 전체 응답 ≤ 500ms, Arm 이동 ≤ 1초
AI 추론 ≤50ms/frame, 자세 피드백 표시 ≤ ~1 s

(2). 보안: 전 구간 HTTPS (TLS 1.3 이상) 적용, JWT 기반 인증과 API Key 병행, 내부망 방화벽, MQTT 브로커 인증·토픽 ACL로 메시지 접근 제한

(3). 확장성/유지보수: 추상화된 제어 모듈, OTA, 테스트/배포 자동화, 대시보드 통계

모델	주요 역할	학습 데이터셋	성능 지표
시선 추적 (YOLOv11, 3단계 전이학습)	눈 위치/시선 추정	COCO, iSEEDS, Eye-D	홍채 중심점 오차 2.5 px (MSE)
자세 추정 (상체 중심)	10개 상체 키포인트 검출	COCO-POSE, FLIC	AP@0.5 = 0.65
자세 이상 탐지 (Rule-based)	잘못된 자세(거북목 등) 판별	의학 논문 및 산업 표준 기반	ISO 9241-5 임계치 기반 룰셋·표준 시나리오 정성 검증

[표1: 모델 성능 및 학습 데이터셋]

Ⅴ. 결론

본 논문에서는 3단계 전이 학습 기반 시선 추적과 경량화된 자세 추정 모델을 적용한 AIoT 인체공학 디스플레이 자동 제어 시스템 'SISEON'을 제안했다. 시제품 구현 및 평가를 통해, 제안 시스템이 높은 정확도와 실시간성으로 사용자의 작업 환경을 개선할 수 있음을 성공적으로 검증했다. 향후 연구에서는 사용자별 피드백을 반영한 개인화 모델 최적화와 다양한 환경 및 디바이스에서의 확장 가능성 검토를 통해, 보다 보편적이고 신뢰도 높은 인체공학 솔루션으로 발전시킬 계획이다.

참 고 문 헌

[1] ISO. ISO 9241-5:2024 Ergonomics of human-system interaction—Part 5: Workstation layout and postural requirements. International Organization for Standardization, 2024.

[2] NVIDIA. Jetson Orin Nano Super Developer Kit Datasheet. NVIDIA Corporation, 2024.

[3] Khanam, R., & Hussain, M. "Yolov11: An overview of the key architectural enhancements." arXiv preprint arXiv:2410.17725, 2024.

[4] Fingas, J. “Good News! The Faster Raspberry Pi 5 With a Custom Chip Just Dropped.” Lifewire, 2023.

[5] ROBOTIS. “OpenMANIPULATOR-X RM-X52-TNM Overview.” ROBOTIS e-Manual, 2025.

[6] Lin, T.-Y., et al. "Microsoft coco: Common objects in context." European conference on computer vision. Cham: Springer International Publishing, 2014.

[7] Roboflow Universe. “iSEEDS dataset: Emotion-Types-and-Eye-Location images and annotations in multiple formats for training computer vision models (version 8, May 19, 2023).” Roboflow Universe, 2023.

[8] Sapp, B. et al. “FLIC: A Dataset for Upper-Body Pose Estimation.” 2013.

[9] Roboflow Universe. “Eye Annotations - YOLO to VOC / Eye Detection KSO3D (Eye-D) dataset (version 4).” Roboflow Universe, 2024

[10] “Treatment of Chronic Neck Pain in Patients with Forward Head Posture.” PMC, 2025.

터널 내 딥러닝 기반 BLE RSSI 차량 측위

정기원, 장준영* 김민석**
충남대학교, *㈜와이파이브, **충남대학교

kwjung2571@o.cnu.ac.kr, *wkdwnsdud25@wifive.co.kr , **kms4105@cnu.ac.kr

Deep Learning-Based BLE RSSI Vehicle Localization in Tunnels

Jung Ki Won, Jang Jun Yeong ,Kim Min Seuk**
Chungnam National Univ., *WiFive Co.,Ltd. **Chungnam National Univ.

요 약

본 논문에서는 GPS (Global Positioning System) 음영 지역인 터널 내 정확한 차량 위치 추정을 위해, 저비용 BLE(Bluetooth Low Energy) 인프라와 딥러닝을 결합한 측위 시스템을 제안한다. 제안하는 시스템은 BLE RSSI (Received Signal Strength Indicator) 를 입력으로 MLP (Multi-Layer Perceptron) 모델을 학습시키고, 필터링을 적용하여 예측 안정성을 향상시켰다. 실험 결과, 평균 예측 오차 2.17 m 를 달성하였다. 본 연구를 통해 BLE 인프라와 딥러닝 및 필터링 결합이 GPS 음영 지역에서 신뢰성 높은 차량 측위 솔루션이 될 수 있음을 입증하였다.

Ⅰ. 서 론

자율주행 기술이 발전함에 따라, GPS 음영 지역에서 정밀 위치 정보 확보는 핵심 요구사항이 되었다. 본 연구는 저비용 BLE 인프라와 딥러닝 기술을 결합하는 새로운 접근법을 제안한다. BLE RSSI 를 fingerprinting 하여 위치를 추정하는 시스템 구축을 목표로 하며, 핵심 기여는 다음과 같다.

1. 통신 신호의 결측치 신호 간의 다변수 상관관계를 고려하는 MICE (Multivariate Imputation by Chained Equation) 보간법을 적용하여 입력 데이터의 신뢰도를 높였다.
2. 이미지나 텍스트와 같은 정형화된 데이터가 아닌, 통신 네트워크 데이터 (BLE RSSI) 를 딥러닝 모델의 입력으로 직접 활용하여 복잡한 비선형 관계를 학습(training) 하고 위치를 추정하는 시스템을 설계했다.
3. 딥러닝 모델의 출력에 포함된 outlier 를 완화하기 위해 후처리 필터링을 도입하여, 최종 결과의 안정성과 신뢰도를 크게 개선하였다.

본 연구에서는 실내에서 수 cm 수준의 높은 정밀도를 제공하는 UWB (Ultra-WideBand) 를 Ground Truth 로 활용하여[1] 불안정한 BLE 신호만으로도 높은 측위 정확도를 달성하고, 비용 효율성과 정확도를 모두 만족하는 터널 내 차량 측위 솔루션을 제시한다.

Ⅱ. 본론

본 연구에서 제안하는 측위 시스템은 데이터 전처리, 딥러닝 모델링, 후처리 필터링의 3 단계로 구성된다. 본 시스템은 "차량의 주행 상황과 관계없이, 특정 물리적 위치에서는 고유한 BLE RSSI fingerprint 가 형성될 것이다"라는 핵심 가정을 기반으로 한다. 이 가정의 타당성을 검증하기 위해 10 kph 등속 주행 데이터를 BLE RSSI 히트맵으로 시각화한 결과, 차량 이동에 따라 밝은 신호 강도가 대각선 패턴으로 순차적으로 나타나는 것을 확인했다[그림 1]. 이는 차량의 이동에 따른 물리적 거리 변화가 RSSI 신호 패턴에 명확히 반영됨을 의미한다.

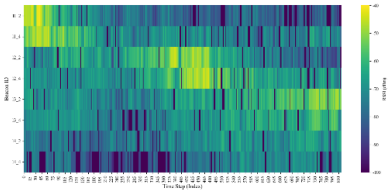


그림 1. 10kph 등속주행 BLE RSSI 히트맵

2.1 데이터 전처리

결측치 처리 단계에서는 다양한 주행 시나리오에 걸친 앵커 신호 간의 전역적 correlation 학습을 위해 모든 데이터를 통합 데이터셋으로 구축하고, 머신러닝 회귀 모델 기반의 MICE 기법을 사용했으며 이는 결측치가 있는 각 변수(앵커)를 정답(y)으로, 나머지 모든 관측된 변수들을 입력 특징(X)으로 삼는 예측 모델을 구축한다. 약 25,000 개의 훈련 데이터를 활용하여 각 앵커 신호 간의 복잡한 correlation 모델 (Imputer) 을 학습시켰다.

2.2 딥러닝 모델링 및 평가

본 연구의 측위 모델로는 MLP 을 채택하였다. 모델의 구조는 [입력층 - 은닉층 1(128) - 은닉층 2(64) - 은닉층 3(32) - 출력층(2)]으로 구성된다. 각 은닉층 다음에는 ReLU 활성화 함수를 사용하여 비선형성을 부여하였고, 배치 정규화와 Dropout (p=0.3) 을 순차적으로 적용하여 학습 과정을 안정시키고 모델의 과적합을 방지하였다.

모델의 성능은 식 (1) 과 같이 UWB Ground Truth $p_{true} = (X_{true}, Y_{true})$ 와 모델의 예측 좌표 $p_{pred} = (X_{pred}, Y_{pred})$ 간의 유클리드 거리 오차 $d(p_{true}, p_{pred})$ 를 기반으로 평가하였다.

$$d(p_{true}, p_{pred}) = \sqrt{(X_{true} - X_{pred})^2 + (Y_{true} - Y_{pred})^2} \quad \cdots \quad (1)$$

2.3 데이터 후처리

딥러닝 모델의 원시 예측 경로는 BLE RSSI 의 노이즈로 인한 급작스러운 위치 변동이나 역방향 이동 등 물리적으로 불가능한 현상을 포함할 수 있다. 이를 보정하는 후처리 필터링은 측위 시스템의 신뢰도와 안정성 향상에 필수적이다[2].

본 연구에서는 순차적 필터링 기법을 적용했다. 먼저 중앙값 필터를 적용하여 outlier 를 효과적으로 제거하고 그 다음으로 이동평균 필터를 통해 잔여 노이즈를 감쇠하고 평활화된 경로를 생성하였다.

시점 t 에서의 원본 예측 좌표를 $p_t = (X_t, Y_t)$ 라고 할 때, 중앙값 필터를 거친 좌표 p'_t 는 식 (2) 와 같다.

$$p'_t = \text{median}\{p_{t-k_1}, \ldots, p_t, \ldots, p_{t+k_1}\} \quad \cdots \quad (2)$$

k_1 은 중앙값 필터의 윈도우 절반 크기이며, 전체 윈도우 크기는 $N_1 = 2k_1 + 1$ 로 정의된다. 본 연구에서는 $N_1 = 10$ 으로 설정하였다. 다음으로 이동 평균 필터를 거친 최종 예측 좌표 p''_t 는 식 (3) 와 같이 계산된다.

$$p''_t = \frac{1}{N_2} \sum_{i=-k_2}^{k_2} p'_{t+i} \quad \cdots \quad (3)$$

여기서 k_2 는 이동평균 필터의 윈도우 절반 크기이며 $N_2 = 2k_2 + 1$ 은 전체 윈도우 크기를 의미한다. $N_2 = 15$ 로 설정하였다. 두 필터의 순차적 적용을 통해 outlier 제거와 경로 평활화를 달성하며, 작은 윈도우 크기 설정으로 실시간 적용이 가능하도록 하였다.

2.4 실험 환경 및 데이터 수집

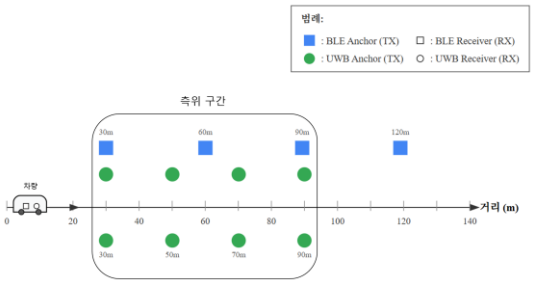


그림 2. 실험 환경

본 연구의 데이터 수집은 폭 9.1 m, 길이 475 m 의 실제 터널에서 수행되었다. 터널 한쪽 벽면을 따라 30 m 간격으로 설치된 4 개의 다방향 BLE 앵커(각 2 개 안테나) [3]로부터 RSSI 신호를 받아 총 8 차원 특징 벡터를 구성하였다. Ground Truth 는 UWB 앵커를 이용해 확보하였으며, 이는 모델이 예측해야 할 2 차원 (X,Y) 의 위치정보로 사용되었다. 7.5 × 60 m 의 측위 구간 내에서 10, 20, 30, 40 kph 정속 및 가·감속 등 다양한 시나리오로 차량을 주행하며 데이터를 수집하였고 이 과정을 통해 각 샘플이 8 차원의 RSSI 벡터와 2 차원의 좌표로 구성된, 총 25,000 개의 데이터셋을 구축하여 MLP 모델 학습에 사용하였다.

2.5 실험 결과

본 연구에서 제안한 측위 시스템의 성능을 검증하기 위해 딥러닝 모델 학습에 사용되지 않은 10 kph 등속 주행 데이터셋으로 제안 모델의 일반화 성능을 평가하였다. 실험 결과, 결측치 처리를 수행하지 않은 원본 데이터로 학습한 모델의 평균 예측 오차는 3.47 m 였다. 이는 상이한 주행 조건의 데이터가 혼재함에도 불구하고, 모델이 위치별 공통 신호 패턴을 성공적으로 학습했음을 시사하며, 핵심 가설의 타당성을 입증한다.

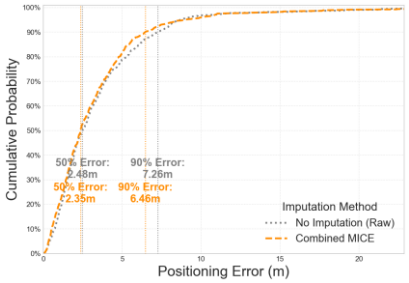


그림 3 MICE 보간법 적용 후 CDF

더 나아가, MICE 보간법을 적용한 데이터로 학습한 모델의 경우, 평균 예측 오차가 3.3 m 로 약 4.6 % 로 개선되었다. 전체 예측의 90 % 지점 오차는 원본 데이터 모델의 7.26 m 에서 6.46 m 로 약 11 % 감소하였다. [그림 3] 이후 예측 경로에 중앙값 필터와 이동평균 필터를 순차 적용한 결과 최종 평균 예측 오차는 2.17 m 로 크게 개선되었다. 중앙값 오차는 2.35 m 에서 1.8 m 로 23.4 % 감소하였고, 90 % 오차는 6.46 m 에서 4.31 m 로 33.3 % 감소하였다. [그림 4] 이는 제안된 필터링 기법이 시스템의 안정성과 신뢰도를 크게 향상시킨다는 것을 알 수 있다.

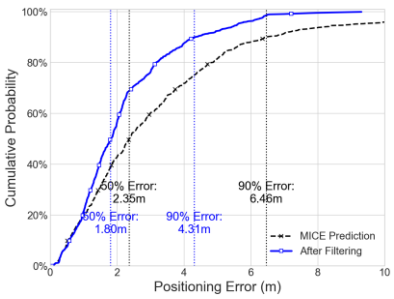


그림 4 필터링 후 CDF

Ⅲ. 결론

본 연구는 차량의 주행 조건과 무관하게, "특정 물리적 위치에서는 고유한 BLE RSSI fingerprint 가 형성된다"는 핵심 가설을 검증하는 것을 목표로 하였다. 실험 결과, 10-40 kph 의 다양한 속도 및 가감속 시나리오로 학습된 모델이 최종 평균 예측 오차 2.17 m 를 보임으로써 위 가설을 성공적으로 입증하였다. 특히 별도의 센서 퓨전 없이, BLE 신호만으로 안정적인 결과를 얻었다는 점에서 그 실용적 가치가 높다고 할 수 있다.

ACKNOWLEDGMENT

본 연구는 충남대학교 학술연구비에 의해 지원되었음

참 고 문 헌

[1] PS Farahsari, A Farahzadi, J Rezazadeh, A Bagheri, A survey on indoor positioning systems for IoT-based application, IEEE Internet of Things Journal, 2022

[2] Z Zhang, J Liu, L Wang, G Guo, X Zheng, X Gong, An enhanced smartphone indoor positioning scheme with outlier removal using machine learning, Remote Sensing, 2021 - mdpi

[3] 성태경, "실내위치 측정시스템 및 측정 방법". 대한민국 특허권 2616631, 18 12 2023.

SEA-Unet v2: 초음파 영상에서의 스펙클 노이즈 제거

이채령, 박지희, 전소연, 최동혁*
건양대학교 의공학부
chaeryounginji / jiheeinji / soyeoninji @gmail.com, *dhchoi@konyang.ac.kr

SEA-Unet v2: Speckle Noise Reduction in Ultrasound Images

Chaeryoung Lee, Jihee Park, Soyeon Jeon, Donghyuk Choi*
Biomedical Engineering, Konyang Univ.

요 약

본 논문은 초음파 영상에서 스펙클 노이즈를 효과적으로 제거하고 구조 정보를 보존하기 위해 Sobel 필터 기반 경계 정보 추출과 Attention U-Net 구조를 결합한 SEA-Unet을 제안한다. 기존 U-Net 및 Attention U-Net은 우수한 복원 성능에도 불구하고 파라미터 수와 연산량이 많아 실시간 적용에 한계가 있다. 이를 해결하기 위해 제안 모델은 Sobel 필터를 활용해 경계 정보를 사전 추출하고, 채널 폭 축소와 경량 업샘플링 구조를 통해 연산 효율성을 높였다. 제안 모델은 초음파 영상 데이터셋에서 기존 모델과 유사한 MSE, PSNR, SSIM을 유지하면서 파라미터 수를 약 75% 감소시켰고, ONNX 변환을 통해 모델 크기를 절반 이상 줄이며 처리 속도를 향상시켰다. 이러한 결과는 제안 모델이 경계 정보가 중요한 초음파 영상 복원에 적합하며, 휴대형 초음파 장비나 원격 진단 시스템과 같이 연산 자원이 제한된 환경에서도 실시간 적용이 가능함을 시사한다.

Ⅰ. 서론

초음파 검사는 방사선 노출이 없으며 비침습적이라는 장점으로 인해 다양한 임상 진단에 활용되고 있다. 그러나 초음파 영상에는 스펙클 노이즈가 빈번하게 발생하여 영상 해상도를 저하시키고, 그 결과 조직 경계와 세부 구조가 흐려져 정확한 진단을 저해한다. 이러한 문제를 해결하기 위해 기존의 필터 기반 기법과 다양한 딥러닝 모델이 제안되었으나, 일부 모델은 연산량이 많아 실시간 적용에 한계가 있는 것으로 보고되었다[1].

선행연구에서 본 연구팀은 Sobel 필터와 Attention U-Net 구조를 결합한 SEA-Unet 모델을 제안하였으며, 이를 통해 스펙클 노이즈 제거와 구조 정보 보존 성능을 검증하였다[2]. 그러나 해당 연구는 416장 규모의 소규모 데이터셋과 단일 커널 크기(3x3)의 Sobel 필터만을 사용하였기 때문에, 데이터 다양성과 필터 최적화 측면에서 한계가 존재하였다.

본 연구에서는 데이터셋 규모를 6,555장으로 확장하고, Sobel 필터의 커널 크기(3x3, 5x5, 7x7)와 계수 범위를 변화시켜 최적 구성을 탐색하였다. 또한 ONNX 변환을 적용하여 모델 크기를 크게 줄이면서도 성능 저하를 최소화하여, 임상 환경에서의 실시간 적용 가능성을 검증하였다. 이를 통해 제안 모델의 성능을 종합적으로 검증하고, 초음파 영상 품질 향상에 기여할 수 있는 가능성을 평가하였다.

Ⅱ. 본론

가. 데이터셋 구성

본 연구에서 사용된 초음파 영상 데이터는 Healcerion Co., Ltd.에서 제공받았다. 전체 데이터셋은 총 6,555장으로 구성되었으며, 이를 train 4,289장, valid 1,029장, test 1,237장으로 분할하였다. 정답 영상(ground truth)은 기존 SRI(Speckle Reducing Imaging) 필터를 적용한 결과 영상을 기준으로 구성하였다. 영상의 입력 크기는 기본적으로 256×256으로 설정하였으며, 추가로 원본 초음파 영상의 실제 크기에 가까운 512×512 크기에서도 학습과 평가를 수행하였다.

나. 제안 모델 구조(SEA-Unet)

제안하는 SEA-Unet은 초음파 영상에서 경계 정보를 정확하게 추출하고, 스펙클 노이즈 제거 과정에서 구조적 정보를 보존하도록 설계되었다. 먼저, 입력 영상에 대해 네 방향(vertical, horizontal, diagonal 1, 2)의 5×5 커널을 적용하여 edge map을 생성하였으며, 그 예시는 그림 1에 제시하였다.

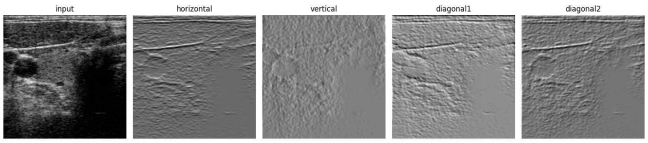


그림 1. 각각의 Filter를 거친 초음파 영상

생성된 네 개의 edge map은 원본 영상과 채널 차원(channel dimension)에서 결합된다. 이러한 결합 방식은 입력 데이터의 채널 수를 확장하여, 네트워크가 원본의 강도 정보와 각 방향의 경계 정보를 동시에 학습하도록 한다. 초기에는 3×3 Sobel 커널을 사용했으나, 커널 크기(3×3, 5×5, 7×7)와 계수 값을 변화시켜 실험한 결과 5×5 커널에서 가장 우수한 성능을 나타냈다. 이러한 경향은 입력 크기가 256×256인 경우뿐만 아니라 512×512 영상 학습에서도 일관되게 확인되었다.

이후 결합된 입력은 Attention U-Net에 전달되어 임상적으로 중요한 영역(조직 경계, 병변 등)을 강조하고 불필요한 배경 영역의 영향을 억제한다. 네트워크는 다중 스케일 특징 추출, skip connection, attention gate를 통해 구조 정보 손실을 최소화하며 스펙클 노이즈를 제거한다. 최종적으로 노이즈가 제거된 단일 채널 초음파 영상을 출력한다.

제안하는 SEA-Unet의 전체 아키텍처는 그림 2에 제시하였다.

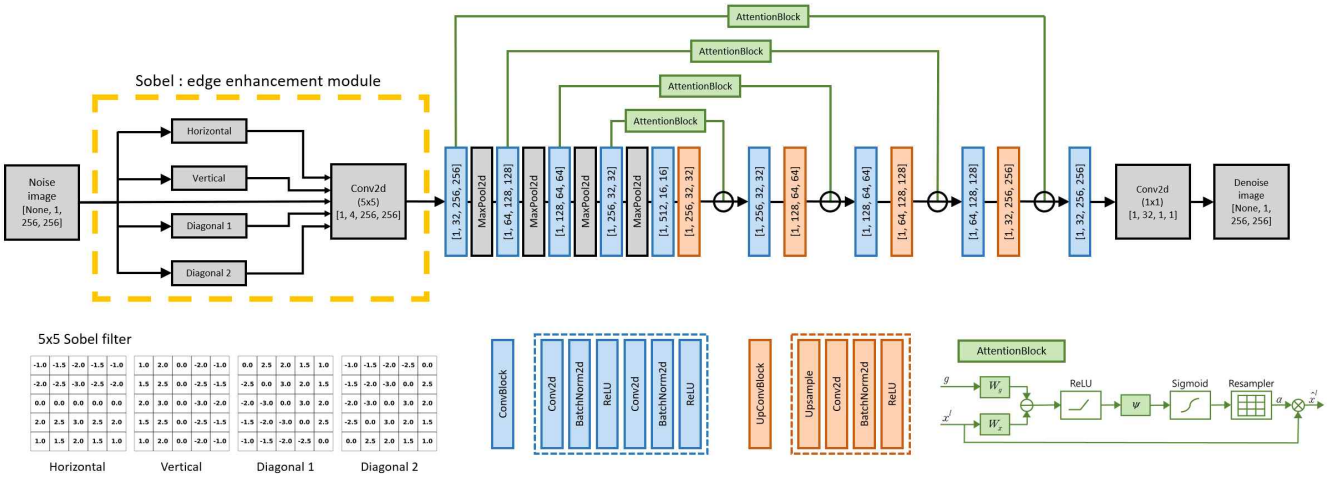


그림 2. SEA-Unet 전체 아키텍처

다. 손실함수 및 학습환경

손실함수는 EDCNN에서 제안한 Compound Loss 구성을 참고하였다[3]. Compound Loss는 MSE(Mean Squared Error)와 Multi-scale Perceptual Loss의 가중합으로 구성된다. MSE는 픽셀 단위의 정확도를 보장하고, Multi-scale Perceptual Loss는 구조적, 시각적 일관성을 유지하도록 한다. 두 손실을 결합함으로써 스펙클 노이즈를 효과적으로 억제하면서도 미세 구조를 보존할 수 있다고 판단하여 채택하였다.

학습은 NVIDIA GPU A100, keras 3.4.1, PyTorch 2.4.1+cu121 환경에서 수행하였다

라. 평가지표 및 모델 경량화

본 연구에서는 성능 평가를 위해 MSE, PSNR(Peak Signal-to-Noise Ratio), SSIM(Structural Similarity Index Measure)을 사용하였다[4]. 또한 모델 크기(MB), 이미지 당 처리 시간(ms), 파라미터 수(Params)를 측정하여 실시간 적용 가능성을 평가하였다.

모델 변환에는 다양한 딥러닝 프레임워크 간 호환성을 제공하는 개방형 형식인 ONNX(Open Neural Network Exchange) 포맷을 사용하였다.

마. 연구 결과

표 1의 비교 결과, 제안 모델(SEA-Unet)은 MSE 0.0012, PSNR 29.067 dB, SSIM 0.9132로 기존 Attention U-Net등과 동등한 복원 성능을 보였으며, 최근 제안된 Transformer 기반 모델 등과 비교해도 성능이 떨어지지 않았다. 파라미터 수는 872만 개로 Attention U-Net 대비 약 75% 감소했고, 처리 속도는 1.36 ms로 더 빨랐다. ONNX 변환 후에도 성능 저하를 최소화하면서 모델 크기를 33.40 MB에서 16.65 MB로 절반 이상 줄였으며, 처리 속도는 1.23 ms로 향상되는 것을 확인하였다.

표 1. 모델별 성능 지표 비교

	MSE	PSNR	SSIM	parmas	Size	time
U-net	0.0012	29.073	0.9174	31,030,593	100.68	1.72
Attention U-net	0.0012	29.314	0.9192	34,877,421	113.18	2.77
EDCNN	0.0013	28.728	0.9058	81,249	0.32	1.56
ViT + Unet	0.0024	26.120	0.8351	89,360,385	340.97	2.58
Swin + Unet	0.0043	23.635	0.7704	30,491,587	98.92	1.63
SEA-Unet	0.0012	29.067	0.9132	8,726,469	33.40	1.36
SEA-Unet ONNX	0.0012	29.070	0.9042	8,718,649	16.65	1.23

Ⅲ. 결론

본 연구에서는 Sobel 필터 기반 경계 정보 추출과 Attention U-Net 구조를 결합한 SEA-Unet을 제안하였다. 제안 모델은 기존 U-Net 및 Attention U-Net과 동등한 수준의 복원 성능을 유지하면서도 파라미터 수와 처리 속도를 크게 개선하였다. 또한 ONNX 변환을 통해 성능 저하를 최소화하며 모델 크기를 절반 이상 줄이고 처리 속도를 향상시켜, 임상 환경에서의 실시간 적용 가능성을 확인하였다. 본 연구의 성과는 경계 정보가 중요한 초음파 영상의 품질 향상뿐만 아니라, 휴대형 초음파 장비 및 원격 진단 시스템과 같이 연산 자원이 제한된 환경에서의 활용 가능성을 높일 것으로 기대된다.

ACKNOWLEDGMENT

"이 논문은 SW중심대학사업단 Lab-Corps 프로그램의 지원을 받아 작성되었습니다." (This paper was written with the support of the SW-Oriented University Project's Lab-Corps program.)

The authors acknowledge Healcerion Co., Ltd. (Republic of Korea) for providing the ultrasound dataset.

참 고 문 헌

[1] J. Lee, "EdgeSRIE: A hybrid deep learning framework for real-time speckle reduction and image enhancement on portable ultrasound systems," arXiv preprint arXiv:2403.02750, pp. 1-8, Mar. 2024.

[2] C.R. Lee, J.H. Park, S.Y. Jeon, and D.H. Choi,, "SEA-Unet: 초음파 영상에서의 스펙클 노이즈 제거," Proceedings of the 2024 KAICTS Fall Conference, pp. 60 - 61, 2024.

[3] Liang, Z., Zhang, H., and Xing, Y., "EDCNN: Edge enhancement-based Densely Connected Network with Compound Loss for Low-Dose CT Denoising," arXiv preprint, arXiv:2011.00139, 2020.

[4] Sara U., Akter M., and Uddin M.S., "Image Quality Assessment through FSIM, SSIM, MSE and PSNR—A Comparative Study," Journal of Computer and Communications, vol. 7, no. 3, pp. 8 - 18, 2019.

우수 관망 시뮬레이션을 활용한 맨홀 수위 예측 모델 연구

주우현[1], 성은산[1], 이태진[2], 윤주상[1]*

[1]동의대학교, [2](주)브로드웨이브

joowh0207@gmail.com, ses0137112@naver.com, LTJ@broadwave.co.kr, *jsyoun@deu.ac.kr

A Study on a Manhole Water Level Prediction Model Utilizing Stormwater Drainage Network Simulation

WooHyun Joo[1], Seong Eun San[1], Taejin Lee[2], Joosang Youn[1]*

Dong-Eui University[1], Broadwave[2]

요 약

최근 국지성 호우와 같은 강수 패턴의 증가로 도심 지역의 침수 피해 사례가 보고되고 있다. 침수 발생 시 CCTV나 강우량 데이터 분석을 통해 경고 및 조치하거나 AIoT 기반 통합 서비스로 대피를 돕는 방안이 제시되고 있다. 하지만 강한 우수 상황에서 침수가 발생하면 대피할 수 있는 시간이 충분하지 않다는 문제가 있다. 이를 해결하기 위해 본 논문은 침수 발생 이전에 수위를 예측하여 대피 시간을 확보할 수 있도록 맨홀의 수위를 예측하는 모델을 제안한다. 모델 학습을 위해 시뮬레이션을 통해 맨홀 수위 데이터를 확보하고, 이를 기반으로 RNN, LSTM, LSTM+Attention 모델을 학습하여 수위를 예측한다. 수위 예측 모델 평가를 위해 실제 맨홀의 수위 데이터를 사용하여 검증하였다. 실험 결과, LSTM 모델이 다른 모델에 비해 우수한 예측 성능을 보였다.

Ⅰ. 서 론

최근 국지성 호우와 같은 강수 패턴의 증가로 도심 지역의 침수 피해 사례가 보고되고 있다[1]. 침수 상황 발생 시 행정안전부에서는 인명피해를 최소화하기 위하여 물이 차오르거나 하수구가 역류하면 즉시 대피를 권고한다. 최근 연구에서는 센서, CCTV, 강우량, 실시간 분석 등을 활용하여 경고 및 대피를 지원하는 방법과 AIoT 기반 통합 서비스를 통해 침수 상황을 모니터링하고 대피를 지원하는 방안이 논의되고 있다[2][3]. 그러나 기존 방식의 경우 강한 우수가 발생하면 대피할 수 있는 시간을 충분하게 확보할 수 없다는 문제가 있다[4]. 이를 해결하기 위해 침수 발생 이전에 대피 시간을 확보하려면 맨홀의 수위를 사전에 예측할 방법을 모색할 필요가 있다.

이를 위해 본 논문은 '우수 관망 시뮬레이션을 활용한 맨홀 수위 예측 모델'을 제안한다. 시뮬레이션을 통해 맨홀 수위 데이터를 확보하고, 이를 기반으로 RNN, LSTM, LSTM+Attention 모델을 학습하여 수위를 예측한다. 또한, 실제 맨홀 수위 데이터를 사용하여 모델의 예측 성능을 비교한다.

논문의 구성은 2장에서 SWMM을 활용한 데이터 생성 과정에 대해 서술하고, 3장에서는 학습된 예측 모델들의 결과를 비교한다. 마지막으로 4장에서 결론을 맺는다.

Ⅱ. 우수 관망 시뮬레이션과 데이터 세트 구축

맨홀 수위 예측 모델을 학습하기 위해 서울시 광진구 중곡 1동 일부 맨홀에 레이저 거리 센서를 설치하여 맨홀의 수위 데이터를 수집하였다. 그러나 동일한 방식을 모든 맨홀에 적용하여 데이터를 확보하는 것은 비용과 운영 측면에서 한계가 있다.

실증지역의 맨홀의 수위 데이터를 확보하기 위해 선행 연구에서 SWMM(Storm Water Management Model)을 활용하여 환경을 구성하였다[5]. 구현한 SWMM 환경의 32개의 맨홀에서 10분 간격으로 맨홀 수위 데이터를 생성하여 총 6,312,480개의 시퀀스 데이터로 활용하며, SWMM으로 확보한 모든 맨홀 데이터만 학습 80%, 검증 10%, 테스트 10% 비율로 나누어 수위 예측 모델의 학습에 사용한다. 표 1은 구축한 데이터 세트의 구성을 보여준다. 모델의 평가를 위한 데이터 세트는 2025.06.18.부터 2025.06.20.까지 실제 맨홀에서 수집한 8,021개의 데이터를 사용하였다.

표 1. 데이터 세트 구성

학습(80%)	검증(10%)	테스트(10%)
5,049,984개	631,248개	631,248개

Ⅲ. 맨홀 수위 예측 모델

맨홀 수위 예측 모델은 시계열 데이터 학습에 적합한 RNN, LSTM, LSTM+Attention 알고리즘을 사용하여 학습을 진행했다. RNN과 LSTM은 셀 100개로 EarlyStopping을 적용하여 각각 52 epoch, 40 epoch만큼 학습을 진행했으며, LSTM+Attention은 셀 100개에 LSTM 레이어 뒤에 Attention 레이어를 추가하여 16 epoch만큼 학습을 진행했다. 각 모델은 출력 레이어에서 3개의 값을 출력하도록 하여 10, 20, 30분 후 수위를 출력하도록 하였다. 모델의 검증은 평균 제곱 오차(MSE), 평균 절대 오차(MAE), 결정계수(R^2)로 평가하였다.

1) RNN 모델

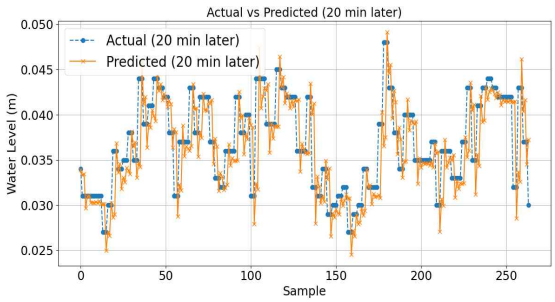


그림 1. RNN 모델의 20분 후 예측값과 실제 수위 값

그림 1은 RNN 모델에서 가장 좋은 결정계수 지폿값을 보인 20분 후 예측값과 실제 수위 값을 나타낸다. 수위 예측 모델의 예측값이 평균 절대 오차값 0.0101m만큼 실제 수위 값보다 낮게 예측하는 모습을 보이며, 최대 오차는 약 0.0122m로 다른 모델에 비해 가장 낮은 값을 보였다.

2) LSTM 모델

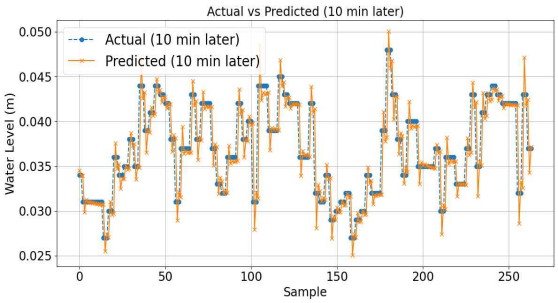


그림 2. LSTM 모델의 10분 후 예측값과 실제 수위 값

그림 2는 LSTM 모델의 10분 후 맨홀 수위 예측값과 실제 맨홀의 수위 값을 비교하였다. RNN 모델에 비해 오차가 약 1.51배 더 작지만, 수위는 평균 절대 오차값 0.0067m 낮게 예측하였다. 그리고 최대 오차는 약 0.0125m로 RNN 모델의 예측값보다 약 0.0024m 높은 값을 보였다.

3) LSTM+Attention 모델

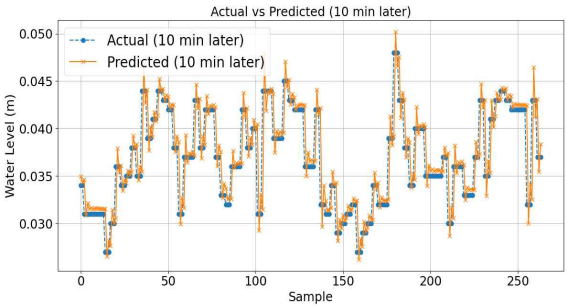


그림 3. LSTM+Attention 모델의 10분 후 예측값과 실제 수위 값

그림 3은 LSTM+Attention 모델의 10분 후 예측값과 실제 수위를 비교한 결과를 나타낸다. 해당 모델은 평균 절대 오차값이 0.01m 높게 예측하였으며, 최대 오차는 0.0125m로 LSTM 모델과 근사한 값을 보였다.

4) 모델별 평가

표 2는 RNN, LSTM, LSTM+Attention 알고리즘으로 학습한 모델의 10, 20, 30분 후 평가 결과를 도시하였다.

표 2. 모델별 평가 지표 결과 비교

모델(예측 시점)	MSE	MAE	R^2
RNN(10min 후)	0.0164m	0.0653m	-0.3477
RNN(20min 후)	0.0006m	0.0101m	0.9427
RNN(30min 후)	0.0164m	0.0654m	-0.3507
LSTM(10min 후)	0.0004m	0.0067m	0.9602
LSTM(20min 후)	0.0006m	0.0086m	0.9498
LSTM(30min 후)	0.0007m	0.0105m	0.9371
LSTM+Attention(10min 후)	0.0007m	0.0100m	0.9360
LSTM+Attention(20min 후)	0.0007m	0.0102m	0.9361
LSTM+Attention(30min 후)	0.0009m	0.0113m	0.9235

RNN으로 학습한 모델은 다른 모델에 비해 가장 낮은 평가 지폿값을 보였으며, 20분 후 예측 결과에서만 다른 시간대에 비해 높은 성능을 기록했다. LSTM 모델은 전반적으로 가장 우수한 성능을 보였지만, 예측 시간이 길어지면 결정계수 값이 평균 1.21%씩 낮은 값을 보였다. LSTM+Attention 모델은 LSTM 모델보다 모든 지폿값이 낮았지만, 예측 시간이 길어져도 결정계수 값의 하락 폭이 평균 0.67%로 다른 모델에 비해 작게 나타났다. Attention 레이어를 적용했을 때 LSTM 모델 대비 성능이 낮게 나타난 것은, SWMM으로 확보한 데이터의 단순한 특성으로 인해 LSTM+Attention 모델에서 과적합이 발생했기 때문으로 보인다.

모델의 성능을 비교한 결과, 수위 차이가 0.0125m로 작고 모든 성능 지표에서 우수한 결과를 보인 LSTM 모델이 수위 예측에 적합한 모델로 판단된다.

Ⅳ. 결론

본 논문에서는 SWMM을 활용하여 맨홀의 수위 데이터를 확보하고, 세 가지 모델(RNN, LSTM, LSTM+Attention)을 학습하여 수위 예측 성능을 비교 및 분석하였다.

평가 지표(평균 제곱 오차, 평균 절대 오차, 결정 계수)를 사용하여 비교한 결과, LSTM기반 모델이 가장 좋은 성능을 보였으며, 이를 통해 시뮬레이션 기반 수위 예측 모델의 활용 가능성을 보였다. 침수 발생 이전에 수위를 예측하여 대피 시간을 확보할 수 있는 기반을 마련하였다.

추후 연구에서는 실제 맨홀 수위 데이터를 확보하고, 학습 및 검증을 진행하여 예측 모델의 성능 변화를 분석하고자 한다.

ACKNOWLEDGMENT

본 연구는 중소벤처기업부의 기술개발사업[RS-2025-02317519]의 지원에 의한 연구임

참 고 문 헌

[1] Yoon, S. K., and Choi, H. S., “Improvement Measures for Ordinances and Legal Systems in Accordance with Enforcement of Urban Flood Prevention Act in Seoul,” The Seoul Institute, 2024.

[2] Choi, J., and Jang, Y., "A Study on Urban Inundation Situation Awareness and Risk Detection," Proceedings of KIIE/KORMS/KSS Spring Joint Conference, pp. 3389-3395, 2025.

[3] Song, Y. G., Shin, B., Jang, Y., and Kang, J. K., "AIoT-Based Urban Flood Prediction Model Utilizing LoRa Mesh Network," Proceedings of the 2025 KICS Spring Conference, vol. 29, no. 1, pp. 32-34, 2025.

[4] Kim, K. B., Kim, J. S., and Yoon, S. E., "A Numerical Modeling of Surcharged Manhole Flow with the Consideration of the Energy Loss Coefficient," Journal of the Korean Society of Civil Engineers, vol. 33, no. 2, pp. 521-528, 2013.

[5] Joo, W. H., Seong, E., Lee, T., Lee, G., and Youn, J., "Study on Water Level Inference Model in Urban Areas Using Urban Watershed Stormwater Drainage Simulation and LSTM Algorithm," Proceedings of KSCI Conference 2025, pp. 13-16, 2025.

차트 패턴 형성 전조의 비지도 유형화와 확률적 매핑을 통한 중장기 주가 방향 예측에 관한 연구

김진주[1], 정민찬[2], 김재윤*
순천향대학교
[1]wlswn@sch.ac.kr, [2]alscks3279@sch.ac.kr, *kimym38@sch.ac.kr

Forecasting Medium- to Long-Horizon Stock Direction via Unsupervised Typing and Probabilistic Mapping of Chart-Pattern Precursors

Kim Jinju[1], Jung Minchan[2], Kim Jaeyun*
Soonchunhyang Univ.

요 약

본 논문은 차트 패턴 완성 직전 전조를 연속형 시퀀스와 이산형 이벤트로 이중 표현하고 비지도 군집으로 유형화하여 전조 패턴과 차트 패턴의 경험적 확률 지도를 구축하여 거래일 예측에 투영하는 방법을 제안한다. 2000-01-01부터 2025-07-15 까지의 OHLCV를 70:30으로 분할하여 누수 없이 평가했다. 장기 지평(H)에서 정확도와 F1, Brier가 개선되었고 L≥20에서 spinning top 등 공통 전조 문법이 일관되게 관찰되었다. 본 연구는 차트 패턴을 사후 식별이 아닌 사전 예측 가능한 사건으로 다루며 실무 적용을 위한 해석 가능성을 함께 제시한다.

Ⅰ. 서 론

금융 시장 연구에서 캔들 차트 패턴 인식은 시장 상태를 진단과 매매 의사결정에 널리 활용되어 왔다.[1] 최근에는 패턴 탐지를 자동화하고 정량화하려는 시도가 증가하면서 캔들 차트와 지표 시퀀스를 입력으로 가격 수준이나 방향을 예측하는 연구도 활발히 진행되고 있다. 그러나 기존 연구들의 다수는 패턴이 형성된 뒤 그 성과를 평가하거나 개별 캔들 차트와 지표의 조합으로 가격 자체를 곧바로 예측하는 것에 초점을 둔다.[2]

차트 패턴은 패턴이 완성된 이후에 눈에 띄지만 실제로는 그 이전 단계에서 서서히 만들어진다. 변동성이 서서히 수축하고 가격의 리듬이 바뀌며 추세가 교체될 조짐이 차곡차곡 쌓인다.[3][4] 본 연구는 바로 그 완성 직전의 가격과 거래량, 흐름 등과 같은 이하 전조를 정량화하고 유형화하여 이후 어떤 패턴이 등장할지와 향후 H(30, 60, 90, 120, 150, 180, 200) 거래일에 대한 수익 방향을 확률적으로 예측하는 체계를 제안한다.

본 논문의 핵심 기여는 패턴 이전 전조 시퀀스의 정식화·대규모 정량화, 전조 조합의 빈도 - 성과 연결로 신호 식별, 패턴 발생 예측과 중장기 수익률 방향 예측까지의 단일 파이프라인 결합이며 이를 통해 차트 패턴을 사후 식별이 아닌 사전 예측이 가능한 사건으로 다루는 관점을 제시한다.

Ⅱ. 본론

1. 방법론

전조는 연속형과 이산형의 이중 표현으로 구성되며 시계열 시퀀스와 벡터인 연속형 표현에서는 종목·시대별 절대 규모 차이는 제거되고 형상과 리듬을 보존하기 위해 전조 구간의 OHLCV는 특성별 Min - Max 정규화로 0에서 1 사이의 스케일에 맞춘다. 이벤트열과 콤보인 이산형 표현에서는 같은 전조 구간에서 의미 있는 전조 이벤트를 시간순으로 기록하며 연속 쌍인 콤보를 집계하여 전조의 문법을 추출한다.

본 연구는 전체적으로 다음과 같이 진행된다. 슬라이딩 윈도우와 피봇 규칙으로 플래그, 변동성 수축, 더블탑, 더블바텀, 헤드 앤 숄더 등 고전 패턴을 자동 탐지하여 각 사례의 시작과 종료 시점을 확보한다. 탐지된 각각의 사례마다 t start 직전 L을 잘라서 연속형과 이산형 표현을 생성하고 전조 직후 예측 지평(H)별 미래 수익률도 계산하여 방향 라벨을 부여한다. 학습 구간의 전조 1D 시퀀스 L을 표준화한 뒤에 K 평균으로 군집화하며 K는 일반적으로 H 범위에서 실루엣 점수가 최대가 되도록 자동 선택한다. 전조 유형인 각 군집에 대해 이후 관측된 차트 패턴의 분포, 지평별 상승 하락 확률, 빈번한 전조 이벤트 콤보 Top N, 대표 시퀀스 등을 요약한다. 평가 구간의 매 거래일마다 각 종목의 최근 L의 전조 시퀀스를 만들고 학습 시 저장한 표준화 파라미터로 변환한 뒤에 학습 단계의 센트로이드 중 가장 가까운 것을 찾는다. 해당 (L, 전조 유형) 키로 확률 매핑을 조회하여 그날의 패턴과 방향 확률을 얻으며 L이 여러 개인 경우에는 다중 L 앙상블을 적용한다. 동일한 종목과 날짜에 대해 거래일 기준 지평을 적용하여 실제 수익 방향을 계산하며 구간 내 실제 발생한 패턴이 있으면 패턴으로 없으면 방향으로 적용하여 정답 패턴으로 병합한다. 예측 확률과 실제 방향을 비교하여 전체, 종목별에 대해 정확도, 정밀도, 재현율, F1, Brier, ROC AUC를 지평별로 산출하고 전조 콤보의 장기 성과, 전조와 패턴 분류 성능도 함께 정리하여 어떤 전조 콤보가 실증적으로 유의미한지 함께 도출한다.

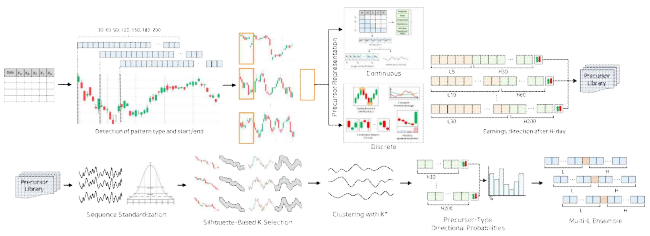


그림1. 전조 유형화 기반 중장기 방향 예측 전체 프레임워크

2. 실험 결과

연구 대상은 다수 종목·지수의 일간 OHLCV이다. 전체 기간은 2000-01-01부터 2025-07-15이고 학습과 테스트 기간은 70:30으로 분할 했다. 전조 창의 길이는 L[5, 10, 15, 20, 25, 30]이며 예측 지평은 H(30, 60, 90, 120, 150, 180, 200) 거래일로 설정했다. 표준화 파라미터 추정, 군집화, k 선택 및 모든 학습 단계는 학습 구간에서만 수행하여 테스트 구간에 대한 정보 누출을 원천 차단했다. 전조와 이중 표현에 대해서는 어떤 패턴의 시작일을 t start라 할 때 전조는 직전 L 구간 [t start − L, t start − 1]이며 미래 정보 누출을 막고 실제 예측에 필요한 최소 정보를 남기기 위해서 이 구간만을 사용했다. 테스트 시에는 학습된 표준화, 센트로이드, 확률 매핑만을 사용하여 진행하였다.

전조 단일 이벤트의 분포를 보면 창 길이 L이 길어질수록 특정 전조가 뚜렷이 부각 된다. L5에서는 PRE_TREND_UP(L-day Positive Slope)이 639건으로 우세하고 L10에서는 CDLENGULFING(Engulfing Pattern), L15에서는 CDLHIKKAKE(Hikkake Pattern)이 각각 최빈 전조로 나타났다. 반면 L≥20부터는 CDLSPINNINGTOP(Spinning Top)이 압도적으로 늘어나 L20, 25, 30에서 최다를 기록했다. 이는 창이 길수록 봉폭 축소와 결정 지연 같은 미세한 리듬 변화가 더 자주 포착된다는 것을 시사한다.

표1. 룩백(L)별 최빈 전조(Top1 precursor)와 발생 건수

L	precursor_pattern	occurrences
5	PRE_TREND_UP	639
10	CDLENGULFING	747
15	CDLHIKKAKE	1438
20	CDLSPINNINGTOP	2526
25	CDLSPINNINGTOP	3817
30	CDLSPINNINGTOP	5147

전조 패턴과 차트 패턴의 짝지음을 살펴보면 여러 패턴에서 CDLSPINNINGTOP(Spinning Top)이 전조로 가장 자주 동반된다. DNTL(Uptrend Line)에서는 1,002건, FLAGU(Bullish Flag) 2,620건, UPTL(Downtrend Line) 1,306건(으로 관찰됐다. FLAGD(Bearish Flag)에서는 CDLLONGLINE(Long Line Candle)이 최빈 전조로 2,005건에 달한다. HNSD(Head And Shoulders)와 HNSU(Inverse Head And Shoulders) 역시 전조의 평균 발생 횟수가 약 1.68~1.69회 수준에 모여서 단일 전조 이벤트가 한 전조 구간에서 소수 회 반복되는 경향을 보여준다. 패턴의 종류가 달라도 봉의 균형 붕괴 전의 머뭇거림(spinning top)과 긴 실체(long line) 같은 형상적 징후가 공통된 문법으로 기능한다는 것을 확인할 수 있다.

표 2. 차트 패턴별 최빈 전조와 발생 통계

chart	precursor	occurrences	avg count
DNTL	CDLSPINNINGTOP	1002	1.776
FLAGD	CDLLONGLINE	2005	1.777
FLAGU	CDLSPINNINGTOP	2620	1.719
HNSD	CDLSPINNINGTOP	2001	1.676
HNSU	CDLSPINNINGTOP	2157	1.688
UPTL	CDLSPINNINGTOP	1306	1.784

예측 성능은 지평이 길어질수록 점진적으로 개선되었다. 방향 예측 기준으로 정확도, 정밀도, F1은 H=30일에 0.512, 0.602, 0.581에서 출발하여 H=200일에는 0.728, 0.728, 0.843으로 상승했다. Brier 점수는 0.252에서 0.207로 감소하여 확률 예측의 평균제곱오차가 줄었다. ROC_AUC는 전 지평에서 0.50대 초반에 머무르는 것을 보여주며 값이 크지는 않지만 분류 지표가 개선되는 양상을 보였다.

표 3. 거래일 지평(H)별 방향 예측 성능

H	accuracy	precision	f1	brier	roc_auc
30	0.512	0.602	0.581	0.252	0.496
60	0.548	0.645	0.652	0.246	0.504
90	0.578	0.671	0.693	0.242	0.510
120	0.688	0.689	0.815	0.226	0.504
150	0.708	0.709	0.830	0.217	0.505
180	0.722	0.723	0.839	0.211	0.506
200	0.728	0.728	0.843	0.207	0.514

Ⅲ. 결론

본 연구는 패턴 완성 이후가 아닌 완성 직전의 전조 패턴을 분석 단위로 삼아서 전조를 연속형 시퀀스와 이산형 이벤트로 이중 표현하고 비지도 군집으로 유형을 추출한 뒤에 각 전형에 귀결되는 경험적 결과인 패턴과 방향 분포를 확률 지도로 고정하여 날마다 투영하는 방법을 제시했다.

룩백(L)이 커질수록 전조의 형상적 문법이 뚜렷해지며 이는 봉폭 축소 후 결정과 지연 후 추세 전환의 전야라는 공통 서사를 시사한다. 서로 다른 차트 패턴이라도 공통 전조가 반복적으로 동반되어 패턴 간에 공유되는 상위 문법이 존재함을 보여준다. ROC-AUC는 높지 않지만 캘리브레이션과 임계값 조정을 통해 운영 지표가 실무적으로 의미 있는 수준으로 개선될 수 있음을 확인했다.

향후 연구에서는 지평과 레짐별 확률 보정과 임계값 적응화로 운영 지표를 추가 개선, 전조 문법의 길이 확장 및 순서 민감도 모델링,동적 룩백 선택과 다중 L 가중 앙상블 학습을 추가하여 진행할 것이다. 또한 거래비용과 슬리피지, 체결 제약과 비용을 포함한 실거래 백테스트로 등으로 확장 시켜서 실전 성과 검증을 통하여 확장 시킬 예정이다.

ACKNOWLEDGMENT

“본 연구는 2025년 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학사업의 연구 결과로 수행되었음”(2021-0-01399)

참 고 문 헌

[1] Chen, Jun-Hao, and Yun-Cheng Tsai. "Dynamic deep convolutional candlestick learner." arXiv preprint arXiv:2201.08669 (2022).

[2] Zhu, Min, Yu Guo, and Yuping Song. "A parallel hybrid neural networks model for forecasting returns with candlestick technical trading strategy." Expert Systems with Applications 255 (2024): 124486.

[3] Shahbandari, Lida, Elahe Moradi, and Mohammad Manthouri. "Stock price prediction using multi-faceted information based on deep recurrent neural networks." arXiv preprint arXiv:2411.19766 (2024).

[4] Steinbacher, Matej. "Predicting stock price movement as an image classification problem." arXiv preprint arXiv:2303.01111 (2023).

사용자 피드백을 반영한 스마트 조명 목표 디밍 레벨 예측 모델

김서연[1], 이준희[2], 김대호[2], 성정식[2], 전윤호[3]
[1,3]국립한밭대학교, [2]한국전자통신연구원 자율형 IoT 연구실
[1]syeon@edu.hanbat.ac.kr, [2]{aloha, dhkim7256, jssung}@etri.re.kr, [3]yhjeon@hanbat.ac.kr

User Feedback based Predicton Model for Target Dimming Levels in Smart Light

Seoyeon Kim[1], Junhee Lee[2], Daeho Kim[2], Jung-Sik Sung[2], and Yunho Jeon[1]
[1]Hanbat National University, [2]Autonomous IoT Research Section ETRI

요 약

스마트 조명 시스템의 자동 제어 방식은 불필요한 에너지 소비를 줄이는 데 효과적이지만, 사용자 선호를 충분히 반영하지 못하는 한계가 있다. 본 논문에서는 사용자의 조명 사용 패턴과 피드백 데이터를 기반으로 조명의 목표 디밍 레벨을 자동 예측하는 시스템을 제안한다. 이를 위해 데이터 불균형 문제를 고려한 2단계 예측 모델 구조를 적용하였으며, 그 결과 사용자 선호를 반영한 보다 정밀한 목표 디밍 레벨 예측이 가능함을 확인하였다.

Ⅰ. 서론

스마트 조명 시스템은 에너지 절감과 사용자 편의성 향상을 동시에 달성하는 중요한 기술이다. 기존의 단순한 조명 제어 방식과 달리 스마트 조명 시스템은 조명 기기와 센서들이 유·무선 네트워크로 연결되어 있으며 사전에 정의된 시나리오에 의해 조명 디밍 레벨(밝기)을 자동 조절하거나 모션 감지 센서를 기반으로 점·소등 및 밝기 조절을 수행함으로써 사용자의 개입 없이도 효율적인 조명 환경을 제공할 수 있다[1]. 이러한 시스템은 불필요한 에너지 소비를 효과적으로 줄일 수 있다는 장점이 있지만 사용자의 피드백을 반영하지 못한다는 한계가 있다. 일부 사용자들은 자동 제어되는 조명 밝기에 만족하지 못해 수동으로 조명을 제어한다. 그러므로 사용자의 조명 사용 패턴과 선호도를 반영한 사용자 피드백 데이터를 기반으로 조명의 목표 디밍 레벨을 자동으로 예측하는 기능이 필요하다. 여기서 목표 디밍 레벨이란 특정 시점에서 조명이 도달해야 하는 목표 밝기값으로 현재 환경에서 수집된 센서 값, 사용자 선호도 등을 반영해 결정되어야 한다. 시스템은 환경 요인과 사용자 패턴을 학습하여 유사한 조건에서 요구되는 밝기 수준을 자동으로 예측하고, 이를 기반으로 목표 디밍 레벨을 향해 조명의 밝기가 자연스럽게 조정될 필요가 있다.

따라서 본 논문에서는 스마트 조명 시스템에서 사용자의 선호를 반영한 목표 디밍 레벨을 예측하는 방법을 제안한다. 이를 위해 데이터 불균형 문제를 고려한 2단계 예측 구조를 설계하고, 각 단계에 머신러닝 기법을 적용하였다. 1단계에서는 조명의 점등 여부를 예측하고, 2단계에서는 조명이 켜져 있는 경우에 한해 목표 디밍 레벨을 예측하도록 설계하였다. 이러한 순차적 예측 과정을 통해 데이터 불균형의 영향을 완화하면서도 보다 정확한 밝기 수준 예측이 가능하다. 본 연구의 궁극적 목표는 사용자가 수동으로 직접 밝기를 조절해야 하는 불편을 줄이는 동시에 에너지 절감과 사용자 만족도 향상에 기여하는 스마트 조명 시스템을 구현하는 데 있다. 두 과정이 순차적으로 이루어져 데이터 불균형의 영향을 줄이면서 비교적 정확한 밝기 예측이 가능하다. 이를 활용하여 사용자가 수동으로 밝기를 조절해야 하는 불편을 줄일 뿐만 아니라 에너지를 절감하며 사용자 만족도를 높이는 방향으로 시스템을 발전시키는 것을 목표로 한다.

Ⅱ. 주거 조명 데이터 전처리

1. 데이터 정보

스마트 조명 시스템의 사용자 맞춤형 목표 디밍 레벨을 예측하기 위해 아파트 거주 한 가구의 침실 조명 데이터를 사용하였다. [2]는 이 아파트 거주 가구들의 조명 에너지 소비 특성을 다양한 패턴으로 분석한 연구로, 본 논문에서 다루는 데이터의 특성과 해석을 이해하는데 중요한 참고가 된다. 본 연구에서 사용된 데이터는 2023년 8월 1일부터 2024년 12월 31일까지의 기간 동안 수집된 데이터로, 5분 간격으로 주기적으로 수집되는 리포트 데이터와 상태 변화가 발생하면 수집되는 상태 데이터들로 구성되어 있으며 총 약 14만 건의 데이터가 존재한다.

2. 데이터 전처리

시계열 데이터이기 때문에 시간의 흐름을 반영하고자 시간 정보를 분 단위로 표현하였다. 또한, 날씨 데이터와 사용자 조작 여부, 직전 디밍 레벨, 최근 20회 디밍 레벨의 평균 및 분산, 디밍 레벨의 지속시간, 변화량 등 시계열 관련 특징 변수(feature)를 추가하였다. 본 연구에서 사용된 모든 특징 변수는 데이터 누수(data leakage)를 방지하기 위해 과거 시점의 값으로만 구성하였으며, 이를 통해 시계열 데이터의 특성을 효과적으로 반영하도록 설계하였다. 실제 조명 시스템 환경에서는 설정이나 센서 특성으로 인해 결측값이 발생할 수 있다. 결측값을 임의로 채워줄 경우, 실제 조명 상태와 일치하지 않는 왜곡된 값이 생성될 수 있으며 채워준 값이 올바르다는 보장이 없다[3]. 따라서 결측값을 별도로 처리하지 않고, 결측값을 자연스럽게 학습 과정에서 처리할 수 있는 트리 기반 모델을 사용하였다. 모델이 미래 정보를 미리 학습하지 않으며 실제 환경과 유사한 조건에서 학습될 수 있도록 데이터를 전처리하였다.

3. 데이터 분할

총 14만 개의 데이터에서 약 8:1:1 비율로 train(학습):val(검증):test(확인) data를 구성하였다. 2024년 9월 1일 데이터를 기준으로 train과 val을 나누고 2024년 11월 14일을 기준으로 val와 test를 구분하였다. 이러한 분할 방식으로 시간 순서를 유지하였다.

Ⅲ. 조명 목표 디밍 레벨 산출 2단계 예측 모델

1. 2단계 예측 모델 설계 배경

조명 디밍 레벨을 살펴본 결과, 각 구간별 비율은 [표1]과 같다.

〔표 1〕 조명 디밍 레벨의 각 구간별 비율

조명 디밍 레벨 구간	비율(%)
0	91.07
1-10	0.53
11-30	0.82
31-60	1.32
61-80	1.21
81-100	5.04

조명 디밍 레벨의 구간이 0이 경우가 약 91%로 대부분을 차지한다. 이는 조명이 켜져 있는 경우가 12,781개, 조명이 꺼져 있는 경우가 130,384개로 매우 불균형하며 주기적으로 수집되는 리포트 데이터로 인해 on과 off 데이터의 수의 차이가 크게 나타나기 때문이다. 전체 데이터를 대상으로 목표 디밍 레벨을 예측할 경우, off 상태의 데이터가 예측 성능을 저해할 수 있다. 이를 해결하기 위해 예측 과정을 두 단계로 구분하였다. 1단계에서는 조명의 on/off 상태를 분류하고, 2단계에서는 on 상태에 한하여 목표 디밍 레벨을 회귀하도록 설계하였다.

2단계 구조에서 각각의 단계에 적용할 모델로는 Random Forest, XGBoost, LightGBM, Catboost를 사용하여 비교하였으며, 최종적으로 각 단계에서 성능이 가장 우수한 모델을 선택하여 사용하였다. 비교 과정에서 따로 파라미터를 설정하지 않고 각 모델의 기본 설정으로 학습하였으며, 결과의 재현성을 위해 랜덤 seed만 고정하여 진행하였다.

2. 조명 on/off 분류 모델

1단계에서는 전체 데이터를 활용하며 조명의 on/off 상태를 정확히 판단하는 것을 목표로 한다. 조명이 꺼져 있는 데이터가 많이 있기에 이 단계에서 정확하게 on 상태를 식별하는 것이 2단계 목표 디밍 레벨 예측의 성능에 매우 중요하다. 조명의 점·소등 예측 단계에서 모델 성능을 평가하기 위해 전체 정확도(Accuracy)와 on/off 클래스별 F1-score를 사용하였다.

〔표 2〕 on/off 분류 예측 성능

Model	Accuracy	F1_on	F1_off
RamdomForest	0.9999	0.9996	0.9999
XGBoost	0.9997	0.9992	0.9998
LightGBM	0.9901	0.9703	0.9941
Catboost	0.9999	0.9996	0.9999

on/off 분류 예측 성능은 모든 모델이 높은 정확도를 보였으나, 데이터가 불균형하므로 상대적으로 샘플 수가 적은 on 클래스의 F1-score를 주요 성능 지표로 삼았다. Random Forest 모델과 Catboost 모델이 on 클래스 F1-score 0.9996, off 클래스 F1-score 0.9999로 가장 우수한 성능을 보였다. on/off 예측 단계에 Random Forest 모델을 채택하였다.

3. 조명 목표 디밍 레벨 예측 모델

2단계에서는 1단계에서 on으로 분류된 데이터만을 대상으로 목표 디밍 레벨을 회귀 방식으로 예측한다. 즉, 조명이 실제로 켜져 있는 상황에서만 목표 디밍 레벨을 예측하도록 함으로써 off 데이터가 목표 디밍 레벨 예측에 영향을 주는 문제를 방지한다.

〔표 3〕 디밍 레벨 회귀 예측 성능

Model	RMSE_on
RamdomForest	2.969
XGBoost	1.906
LightGBM	2.121
Catboost	1.070

조명이 켜져 있을 때는 디밍 레벨 값이 다양해 정확하게 예측하는 것이 중요하므로 모델 성능 평가 지표는 조명이 켜져 있을때 RMSE를 사용하였다. CatBoost는 on 상태에서 rmse가 가장 낮아 조명 목표 디밍 레벨 예측 모델로 선택하였다.

4. 실험 및 평가

조명의 on/off 상태를 예측하는 1단계에서는 RandomForest를, 2단계에서는 목표 디밍 레벨을 예측하기 위해 CatBoost를 사용하였다. 또한, 최적의 하이퍼파라미터를 찾기 위해 베이지안 최적화 기반의 자동화 라이브러리인 Optuna를 활용하여 파라미터를 탐색하였다. 2단계로 나누어 진행했으므로 on/off 분류 정확도와 on 상태에서의 디밍 레벨 예측 정확도을 주요 성능 지표로 활용하였다.

〔표 4〕 test data 예측 성능

	단일모델	2단계 모델
on/off 분류 성능	-	0.9989
on 상태 rmse	3.916	1.525
전체 RMSE	1.729	1.235

표 4는 test 데이터를 제안 시스템에 적용한 예측 결과를 나타낸다. 단일모델과 본 논문에서 제안한 2단계 모델 방식에 대해 성능을 비교하였다. on/off 분류 정확도는 0.9989로 매우 높게 나타났으며, 2단계 모델은 단일 모델에 비해 RMSE를 낮추어 전반적인 예측 성능을 향상시켰다. 특히, on 상태에서의 RMSE가 3.916에서 1.525로 개선되어 목표 디밍 레벨 예측 정확도가 크게 높아졌고, 전체 RMSE 또한 1.729에서 1.235로 낮아져 제안한 2단계 구조의 효과성을 확인할 수 있었다.

Ⅳ. 결론

본 논문에서는 스마트 조명 시스템에서 사용자의 선호를 반영한 목표 디밍 레벨을 예측하는 시스템을 제안하였다. 제안된 시스템은 조명의 on/off 여부를 판단한 후 켜진 경우에만 목표 디밍 레벨을 예측하는 2단계 학습 구조를 통해 데이터 불균형 문제를 완화하고 기존 단일 단계 모델 대비 우수한 성능을 보였다. 다만, 한 세대 데이터만을 사용하므로 모델이 특정 사용자 패턴에 과적합될 가능성이 존재한다. 또한, 실시간으로 사용자 피드백이 발생하는 경우 기존 모델은 새로운 데이터를 즉시 반영할 수 없고 추가 학습이 필요하다는 한계가 존재한다. 향후 연구에서는 다양한 세대와 환경의 데이터를 포함하여 모델의 일반화 성능을 검증할 필요가 있으며, 시스템을 실시간으로 적응시키기 위해 강화학습을 적용하여 실시간으로 사용자 행동과 선호 변화에 따라 자동으로 목표 디밍 레벨을 조정할 수 있는 적응형 스마트 조명 시스템을 연구할 예정이다.

ACKNOWLEDGMENT

본 연구는 산업통상자원부(MOTIE)와 한국에너지기술평가원(KETEP)의 지원을 받아 수행한 연구 과제입니다. (No. RS-2024-00441420)

참 고 문 헌

[1] 성정식, 김대호, “스마트 조명 시스템의 통합 운영을 위한 요구사항, 구조 및 주소 체계,” 조명·전기설비 제38권, 제4호, pp29-34, 2024.

[2] A. Khan et al., “ Dynamic Temporal Analysis and Modeling of Residential Lighting Consumption for Energy Efficiency and Sustainability,” IEEE Access, Vol. 12, pp154365-154380, 2024.

[3] 이경재, 임현우, “건물 에너지 데이터 분석에서 결측치 처리방식에 따른 차원 축소 및 모델 예측 성능 비교,” 한국태양에너지학회 논문집, 제44권, 제1호, pp59-75, 2024.

생성형 AI 저작권 침해 대응을 위한 LangGraph 기반 LLM 에이전트 설계

정현우, 최윤서, 최동훈, 손경아*
아주대학교

jhw0714@ajou.ac.kr, chldbstj0512@ajou.ac.kr, dhc4003@ajou.ac.kr, *kasohn@ajou.ac.kr

LangGraph-based LLM Agent Design for Addressing Copyright Infringement by Generative AI

Hyun Woo Jeong, Yun Seo Choi, Dong Hun Choi, Kyung Ah Sohn*
Ajou Univ.

요 약

본 연구는 생성형 AI 기반 이미지의 저작권 침해 여부를 기술적으로 판별하고, 법률적 자문을 제공하는 LangGraph 기반 LLM 에이전트를 제안한다. 생성형 인공지능의 발전은 다양한 분야의 콘텐츠 제작에 혁신을 가져왔다. 그러나, 창작자의 동의 없이 수집된 데이터가 학습에 활용되거나, 원 저작물과 유사한 결과물이 무분별하게 활용되는 등 저작권 침해 가능성이 심화되고 있다. 이를 해결하기 위해 ViT 기반 합성 이미지 검사기와 저작권 판례 검색기를 구축하였고, LangGraph 를 이용하여 LLM 에이전트가 이를 활용하여 검색 증강 생성을 수행하도록 구현하였다. 실험 결과 이미지 검사기를 통하여 사용자가 의심되는 모방 이미지에 대하여 정량적인 판단 기준을 제공받을 수 있으며, 검색 증강 생성을 통하여 사용자가 문맥적 일관성과 법률적 연관성이 높은 응답을 받을 수 있음을 확인하였다. 본 시스템의 AI 생성 여부 판별 및 저작권 침해 분석 에이전트 시스템을 통해 생성형 AI 시대의 지식재산권 보호 체계 구축에 기여하고자 한다.

Ⅰ. 서 론

생성형 인공지능(Generative AI)의 발전은 음악, 미술 등 다양한 분야의 콘텐츠 제작 방식에 혁신을 가져왔다. 그러나 창작자의 동의 없이 수집된 데이터가 학습에 활용되거나, 원 저작물과 유사한 결과물이 무분별하게 활용되는 등 저작권 침해 가능성이 심화되고 있다[1,2].

본 연구는 생성형 AI 기반 이미지의 저작권 침해 여부를 기술적으로 판별하고, 법률적 자문을 제공하는 LangGraph 기반 LLM 에이전트를 제안한다. 제안하는 시스템을 통하여 AI 생성물 여부에 대한 판별을 거쳐 저작권 침해 분석을 통해 창작자의 권리 보호와 침해 대응을 지원함으로써, 생성형 AI 시대의 지식재산권 보호 체계 구축에 기여하고자 한다.

Ⅱ. 관련 연구

합성 이미지 탐지. Utkarsh Ojha 등[3]은 CILP 모델에서 추출한 사전 학습된 Frozen ViT 특징 추출기와 단순 선형 분류기만을 활용하여, 다양한 생성 모델에 대해 우수한 일반화 성능을 보이는 합성 이미지 탐지 모델을 제안하였다.

RAG 기반 법률 자문 시스템. Song-Yeon Kim 등[4]은 입력된 사실 관계를 바탕으로 관련 법령과 주장·증명책임을 도출하는 모델을 제안하였다. 해당 모델은 공개된 판례 데이터에 법학 전공자가 주석을 추가하여 데이터셋을 구축하고, 이를 RAG 기법과 결합하여 답변을 생성한다.

차별성. 본 연구는 ViT 기반 이미지 탐지 결과를 LLM 에이전트의 응답 생성에 적용하고, 저작권 판례 검색을 통한 법률 자문을 통합 제공함으로써 AI 생성물 저작권 침해 의심 사례에 실질적 대응을 지원한다는 점에서 차별성을 가진다.

Ⅲ. 시스템 설계

본 논문에서 제안하는 모델은 이미지 검사 모듈과 저작권 판례 검색 모듈로 구성되며, 이들은 LangGraph 기반의 에이전트 구조로 상호 연동된다. LangGraph 의 명시적인 워크플로우 로직 제어를 통하여 시스템의 신뢰성 및 설명 가능성을 향상시켰으며, 그래프 구조의 워크플로우로 다양한 이미지 도메인 및 AI 생성 여부를 고려한 다양한 전략을 사용할 수 있도록 설계하였다.

이미지 검사 모듈. Swin Transformer 구조를 기반으로 입력 이미지의 AI 생성 여부를 판별한다. 저작권 침해 분석을 위해 Multimodal 접근법을 적용하여 세 가지 영역에서 유사도를 계산하여 저작권 침해 가능성을 정량적으로 산출한다: (1) ImageNet 으로 사전 훈련된 ResNet50 을 사용하여 2,048 차원 특징 벡터를 추출하고, 이를 바탕으로 코사인 유사도를 계산하여 의미적 유사성을 계산한다. (2) Wang 등[5]이 제안한 SSIM 알고리즘을 통해 휘도, 대비, 구조 정보를 종합한 시각적 유사성을 정량화 한다. (3) RGB 각 채널별 256-bin 히스토그램을 계산하고, Pearson 상관계수를 산출하여 색상 분포의 일치도를 평가한다.

저작권 판례 검색 모듈. 한국저작권위원회에서 발간한 저작권 상담사례집[1]을 주요 데이터 소스로 수집하여 LLM 에이전트가 법률 자문 과정에서 판례 및 법률 조항을 활용하도록 하였다. 수집한 데이터를 GPT-4o-mini 모델을 활용하여 전처리하고 임베딩 벡터를 생성하였다.

[1]https://www.copyright.or.kr/business/law-case-counsel/index.do

*아주대학교 소프트웨어학과 교수, 교신저자, kasohn@ajou.ac.kr

맥락	2023 년 3 월, 172 회 서코에서 한 블루 아카이브 동인부스가 **AI 그림을 트레이싱하고 굿즈를 만들어 논란**이 되었다. 트레이싱이라고는 하나, 사람 손으로 새로이 그렸음에도 어색한 부분이 있어 걸렸다. 해당 굿즈들은 환불처리 진행중이다.
질문	제가 만든 블루아카이브 팬아트가 AI 트레이싱을 거쳐 굿즈로 판매되었습니다. 비록 일부 사람이 다시 그렸지만 여전히 제 원작을 기반으로 하고 있습니다. 이런 경우 저작권 침해를 주장하여 **판매 중단과 손해배상을 청구할 수 있을까요?**
참고 문서	Q: 21. 저작권 보호기간은 얼마나 되나요? A: 우리 저작권법은 특별한 규정이 있는 경우를 제외하고는 **저작자가 생존하는 동안과 사망한 후 70 년까지 저작권 보호기간을 인정**한다. 저작권은 저작물을 창작한 때부터 발생하지만 법에서 정한 보호기간이 경과하면 공유저작물(Public Domain)이 되어 누구나 자유로이 이용할 수 있다. 우리 저작권법은 특별한 규정이 있는 경우를 제외하고는 저작자가 생존하는 동안과 사망한 후 70 년까지 저작권 보호기간을 인정하고 있으며, 보호기간은 저작자가 사망한 다음 해부터 기산한다.
생성 결과	귀하의 블루아카이브 팬아트가 AI 트레이싱을 거쳐 굿즈로 판매된 경우, 저작권 침해를 주장할 수 있는 가능성이 있습니다. 저작권법에 따르면, **원작자의 저작물에 대한 권리는 저작자가 생존하는 동안과 사망 후 70 년까지 보호됩니다.** 귀하의 팬아트가 원작을 기반으로 하여 창작된 경우, 해당 팬아트의 저작권은 귀하에게 있으며, 이를 무단으로 사용한 경우 저작권 침해가 성립할 수 있습니다.

표 1. 저작권 침해 관련 사례 및 질의응답 생성 결과 예시

생성된 벡터는 FAISS (Facebook AI Similarity Search) 기반 벡터 데이터베이스(Vector DB)에 저장하여 고속 유사도 검색이 가능하도록 하였다. 검색 결과는 LLM 에 전달되어 법률 연관성을 고려한 응답을 생성한다.

Ⅳ. 실험 결과

프로토타이핑. 설계한 에이전트 시스템을 검증하기 위하여 프로토타입 모델을 개발하여 실험 및 검증을 진행하였다. 프로토타입 모델은 gpt-4o-mini 생성 모델을 기반으로 하며, streamlit GUI 를 이용하여 그림 1 과 같이 사용자 인터페이스를 구현하였다.

그림 1. 프로토타입 모델 인터페이스

저작권 상담 사례 검색. 저작권 침해와 관련된 10 개의 국내 및 해외 사례를 수집하였고, 사건 내용을 토대로 GPT-5 모델을 통해 법률 질문을 생성하였다. 표 1 과 같이, LLM 에이전트는 검색 문서를 적절하게 활용하여 최종 응답 생성에 활용하는 볼 수 있다.

	실제 이미지	AI 모방 이미지	같은 도메인의 다른 실제 이미지	다른 도메인의 실제 이미지
이미지				
의미적 유사성	-	0.8444	0.5232	0.6112
시각적 유사성	-	0.2082	0.1837	0.1348
색상분포 일치도	-	0.7023	0.4211	0.2884

표 2. 이미지 유사도 분석 결과

이미지 유사도 검사. 명화 및 애니메이션의 실제 이미지와 모방 이미지의 쌍으로 이루어진 총 20 장의 이미지를 준비하여 실제 이미지와 모방 이미지 간 이미지 유사도 검사 결과를 확인하였다. 표 2 를 보면, 실제 이미지를 모방한 이미지의 유사성이 실제 이미지와 관련 없는 다른 이미지보다 높았음을 확인하였다. 따라서 이미지 검사 모듈이 유사한 이미지를 판단할 수 있으며, 이를 통해 창작자가 의심되는 이미지에 대하여 정량적인 판단 기준을 제공받을 수 있음을 의미한다.

Ⅴ. 결론 및 향후 연구

본 연구에서는 생성형 AI 기반 이미지의 저작권 침해 여부를 기술적으로 판별하고, 법률적 자문을 제공하는 LangGraph 기반 LLM 에이전트 시스템을 제안하였다. 이미지 검사 모듈과 저작권 판례 검색 모듈을 통하여 정량적 판단 기준과 함께 문맥적 일관성과 법률적 연관성이 높은 답변을 얻을 수 있음을 확인하였다. 후속 연구로 캐릭터, 회화, 소설 등과 같은 다양한 유형의 컨텐츠 도메인별 침해 유형을 분류하고 대응 전략을 추천하는 알고리즘을 고도화 할 수 있을 것이다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 인공지능융합혁신인재양성사업 연구 결과로 수행되었음(IITP-2025-RS-2023-00255968)

참 고 문 헌

[1] 황지온. (2025). LLM 기반 생성 콘텐츠의 저작권 및 표절 관련 이슈 연구. 교양학연구, 12(1), 189-209. 10.55450/hilas.2025.12.1.189

[2] Li, O., Cai, J., Hao, Y., Jiang, X., Hu, Y., & Feng, F. (2025, July). Improving synthetic image detection towards generalization: An image transformation perspective. In Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 1 (pp. 2405-2414).

[3] Ojha, U., Li, Y., & Lee, Y. J. (2023). Towards universal fake image detectors that generalize across generative models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 24480-24489).

[4] 김송연, 김창현, 이상대, 박혜진, 김영민. (2024-12-18). 주장증명책임을 안내하는 RAG 기반 법률 LLM. 한국정보과학회 학술발표논문집, 전남.

[5] Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image Quality Assessment: From Error Visibility to Structural Similarity. IEEE Transactions on Image Processing, 13(4), 600-612. doi:10.1109/TIP.2003.81986

sLLM 양자화의 성능-효율성 트레이드오프에 대한 실증적 분석 연구

김강우, 함승재, 박상재, 성민규, 강재모*

경북대학교

{kkwo1013, hamsj99, psj1069, alsrb0351, *jmkang}@knu.ac.kr

sLLM Quantization : An Empirical Study of the Performance-Efficiency Trade-off

Kim Kang woo, Ham Seung Jae, Park Sang Jae, Sung Min Gyu Kang Jae-Mo*

Kyungpook National Univ.

요 약

최근 인공지능 분야는 LLM(Large Language Model)의 등장으로 다방면에 걸쳐 꾸준한 발전을 거듭하고 있다. LLM은 수천억 개에 달하는 파라미터를 기반으로 인간의 언어를 이해하고 생성하는 LLM의 능력은 다양한 산업 분야에서 새로운 가능성을 제시했다. 하지만 막대한 연산과 높은 비용, 느린 추론 속도라는 제약이 있다. 이러한 한계는 실시간 서비스나 On-device AI와 같이 빠른 응답 속도를 요구하는 환경에서 LLM의 도입을 가로막는 주요한 장벽이 되고 있다. 이러한 문제의 대안으로 적은 파라미터 수와 빠른 추론 속도를 기대할 수 있는 sLLM(small Large Language Model)이 강력한 대안으로 부상하고 있다. 본 논문은 sLLM의 효율성을 극대화하기 위한 양자화(Quantization) 알고리즘을 적용했을 때 발생하는 성능 저하와 그로 인해 얻을 수 있는 이점 사이의 상충 관계를 규명한다. 또한 향후 발전할 sLLM과 같은 언어 모델 연구를 위한 기반이 되는 자료를 제시한다.

Ⅰ. 서론

최근 인공지능 분야는 LLM의 놀라운 기술의 성장을 보여주고 있다. LLM은 인간의 언어를 이해하고 생성하는 능력을 바탕으로 AI 챗봇 상담, 소프트웨어 코드 생성, 콘텐츠 창작 등 산업 전반에 걸쳐 산업의 성장을 촉진하고 있다. 특히, Google, META, xAI와 같은 글로벌 플랫폼 기업들은 대규모 데이터 센터와 막대한 컴퓨팅 자원을 기반으로 하는 중앙 집중식 클라우드 모델을 통해 사용자 경험을 극대화하고 있다. 중앙 집중식 클라우드 방식을 기반으로 개인화된 AI 서비스를 제공하고 있으나, 막대한 자원과 높은 운용비용, 네트워크 지연 시간이라는 본질적인 한계가 존재한다. 사용자들은 서비스 이용을 위해 중앙 서버에 매번 요청을 보내고 받는다. 이러한 방식은 실시간 상호작용이 필수적인 응용 분야에서는 물리적인 제약으로 인해 불필요한 지연이 발생하며, 사용자 데이터를 중앙 서버로 전송해야 하는 과정에서 데이터 프라이버시와 보안에 대한 우려 또한 증대된다. 이러한 한계를 극복하기 위해서는 클라우드 의존성을 줄이고, E2E(End-to-End)에서 직접 AI 연산을 수행하는 On-device AI 패러다임의 필요성이 대두된다.

Ⅱ. 관련 연구

LLM : 최근 LLM의 파라미터 수가 증가하면서 막대한 자원을 소모하게 되자 LLM의 비용에 대한 중요성이 커지고 있다. Google의 Gemma[2]와 같이 QAT(Quantization Aware Training)[1]을 사용하여 메모리 사용량을 대폭 줄인 결과를 보여주거나 OpenAI의 gpt-oss 20b과 같은 모델과 같이 MoE(Mixture-of-Experts)[2] 아키텍처를 사용하여 추론 시 비용을 크게 절감한 LLM들이 등장하고 있다. 이는 고비용의 모델이 아닌 저비용 고성능의 모델이 필요하다는 것을 시사한다.

sLLM:최근 LLM의 발전과 함께 sLLM의 성장은 AI 기술의 개인 보급화를 앞당기는 핵심이다. Microsoft, Google, Meta 등은 거대 모델의 막대한 비용과 자원 한계를 극복하고자, 고도로 정제된 데이터와 최적화된 아키텍처를 통해 적은 파라미터와 비용으로도 놀라운 성능을 내는 sLLM 개발에 집중하고 있다. 이러한 연구는 중앙집중형 방식에서 개인의 단만 장비로 가져오는 On-device AI를 현실화하고 있다. sLLM의 발전은 인터넷 연결 없이도 빠르고 안전하게 AI를 사용하는 환경을 조성하며, 기술의 실용성과 접근성을 극대화하는 환경을 만들 수 있다.

양자화 : 기존 양자화는 FP4나 FP8과 같은 양자화는 성능의 감소를 감수하더라도 빠른 추론 속도와 낮은 비용으로 추론을 할 수 있도록 실용적 목표에 집중했다. 최근 양자화 기술은 GPTQ(Generative Pre-trained Transformer Quantization)[3]와 AWQ(Activation-aware Weight Quantization)[4]와 같은 방식들은 모델의 중요한 가중치를 보호하는 것과 같이 양자화 과정에서 성능 저하를 최소화하면서도 비용 효율성을 확보하여, 원본 모델의 성능을 거의 그대로 유지를 중심으로 발전하고 있다. 이는 모델의 성능과 비용의 균형이 중요한 것을 알 수 있다.

Ⅲ. 본론

본 논문에서는 최신 양자화 기법들을 적용한 sLLM의 성능을 변화를 측정한다. 이를 통해, 제한된 연산 자원을 가진 엣지 디바이스 환경에서도 충분한 가능성을 입증하고자 한다.

1.실험환경

본 논문에서 진행한 실험 환경은 표 1과 같다.

	Content
GPU	Nvidia RTX A5000 * 4
Framework	Pytorch, vLLM
models	EXAONE-3.5-2.4B-Instruct
	Gemma-3-1B-it
	Llama-3.2-1B

표 1. 실험 환경

	EXAONE-3.5-2.4B-Instruct			Gemma-3-1B-it			Llama-3.2-1B		
	vanilla	GPTQ	AWQ	vanilla	GPTQ	AWQ	vanilla	GPTQ	AWQ
MMLU	59.29	34.31	57.24	39.87	34.41	42.52	31.10	40.45	30.76
HellaSwag	54.04	42.88	53.01	42.45	40.88	42.54	48.34	41.86	36.91
Inference time	26.15 t/sec	8.72 t/sec	13.90 t/sec	16.34 t/sec	11.67 t/sec	13.03 t/sec	43.66 t/sec	34.29 t/sec	31.48 t/sec
Vram Usage	4.61GB	0.94GB	2.07GB	0.96GB	0.93GB	0.97GB	1.03GB	0.99GB	0.99GB

표 2. sLLm 정량적 성능 평가표

2.실험방법

표 2와 같이 성능을 평가하기 위해 57개 분야의 전문 지식을 평가하는 MMLU와 문맥 기반 상식 추론 능력을 평가하는 HellaSwag를 활용하여 기존 모델과 성능을 비교했다. 또한 양자화된 모델이 long-context를 처리하는 강건성을 검증하기 위해 LongBench 벤치마크를 활용하여 성능을 검증했다.

실험으로 활용한 sLLM 모델은 각기 다른 학습 방식을 통해 학습된 모델을 활용했다. EXAONE 3.5[5]는 사전학습 단계에서 학습 데이터 정제와 SFT(Supervised Fine-tuning)과 DPO(Direct Preference Optimization)을 통해 학습되었다. Gemma 3[6]는 QAT를 활용하여 비용을 감소시켰다. Llama 3.2[7]는 Pruning을 통해 성능은 높이고 비용은 감소시켰다. 이와 같이 현재 모델들은 각기 다른 학습 방식을 통해 성능을 높이거나 연산 비용을 낮추는 형식으로 학습되고 있다.

본 논문의 양자화는 autogptq, autoawq라이브러리를 사용했다. 양자화는 PTQ(Post-Training Quantization) 방식으로 GPTQ와 AWQ를 통해 이뤄졌다. GPTQ는 입력층부터 출력층까지 순차적으로 처리하는 Layer-wise Quantization을 사용하여 양자화를 했으며, wikitext2를 보정 데이터셋으로 사용했다. AWQ는 중요 가중치를 보존하는 Per-channel scaling방식을 사용하며, The Pile 데이터셋의 검증 데이터셋을 보정 데이터셋으로 이용했다.

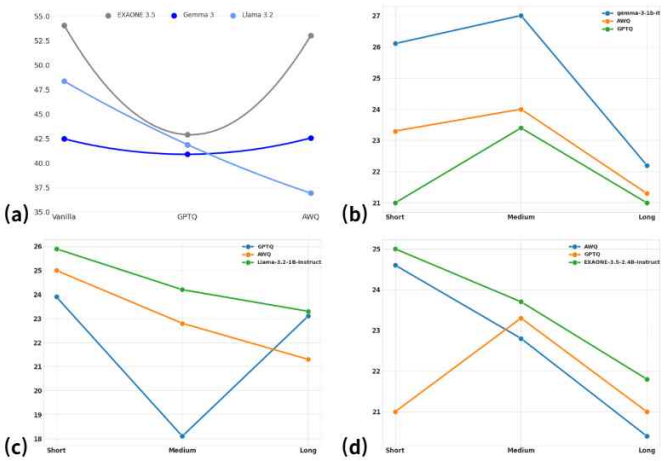


그림 1. (a) 모델별 HellaSwag 시각화, (b, c, d) 모델별 LongBench 문맥 길이에 따른 문맥 이해도 시각화

3.실험결과

그림 1의 (a)는 HellaSwag 벤치마크를 통해 모델의 양자화 방식에 따른 성능을 보여준다. 여기서 GPTQ는 양자화를 적용했을 때 모델의 종류와 관계없이 성능이 유사한 수준으로 수렴하는 경향이 있다. 이는 GPTQ의 성능이 모델 자체의 고유한 특성보다 양자화에 사용된 보정 데이터셋에 크게 의존할 수 있음을 시사한다. 즉, 특정 데이터 유형에서는 높은 성능을 보일 수 있으나, 범용적인 성능은 저하될 가능성이 있다. 반면, AWQ는 Llama모델을 제외한 원본(Vanilla) 모델과 거의 동등한 성능을 유지하는 것을 확인할 수 있다. 이는 AWQ가 추론 속도와 같은 효율성을 개선하면서도 모델의 성능을 잘 보존하는 경향이 있음을 의미한다. 한편, 그림 1(b, c, d)의 문맥 길이별 성능 그래프를 보면, 모든 양자화 적용 여부와 관계없이 긴 시퀀스 입력에 대해 유사한 패턴으로 성능이 감소하는 것을 확인할 수 있다. 이는 양자화 기술 자체가 모델의 장단기 기억 능력에 미치는 영향이 낮은 것을 알 수 있다.

표 2의 Vram 사용량을 보면 알 수 있듯 모델 크기에 따른 압축률 차이를 명확히 보여준다. 2.4B 파라미터의 EXAONE 모델은 높은 압축률을 달성했지만, 1B 규모의 다른 모델들은 그 효과가 미미했다. 이러한 결과는 현재 sLLM은 기본 파라미터가 작을수록 압축 잠재력 또한 낮아진다는 점을 암시한다.

Ⅳ. 결론

본 논문은 서로 다른 방식으로 학습된 sLLM에 최신 양자화 기법을 적용하여 그 성능을 비교 분석했다. 실험 결과로 모델의 학습 방식이 장단기 시퀀스 처리 능력에 미치는 영향은 미미한 것으로 나타났다. 결론적으로, sLLM에 양자화를 적용하면 추론 속도 향상이라는 이점은 분명히 존재하지만, 높은 수준의 모델 압축률을 기대하기는 어렵다는 한계를 확인했다.

이러한 한계점을 통해 특정 sLLM의 내부 가중치 분포와 활성화 패턴을 사전에 분석하여, 이에 최적화된 맞춤형 양자화 연구를 추후 진행할 계획이다.

ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원-대학ICT연구센터(ITRC)의 지원을 받아 수행된 연구임(IITP-2025-RS-2020-II201808)

참 고 문 헌

[1] Mengzhao Chen, Wenqi Shao, Peng Xu “EfficientQAT: Efficient Quantization-Aware Training for Large Language Models“

[2] William Fedus, Barret Zoph, Noam Shazeer “Switch Transformers: Scaling to Trillion Parameter Models with Simple and Efficient Sparsity“

[3]Elias Frantar, Saleh Ashkboos, Torsten Hoefler “GPTQ: Accurate Post-Training Quantization for Generative Pre-trained Transformers “

[4]Ji Lin, Jiaming Tang, Haotian Tang “AWQ: Activation-aware Weight Quantization for LLM Compression and Acceleration”

[5]LG AI Research “EXAONE 3.5: Series of Large Language Models for Real-world Use Cases”

[6]Gemma Team “Gemma 3 Technical Report”

[7]Llama Team “The Llama 3 Herd of Models”

A Multiple LLM RAG Platform Using Answer Templates and Matching-Based Integration

Athiruj Poositaporn[12], Jaehyun Kim[3], Hanmin Jung[12]
[1]University of Science and Technology, [2]Korea Institute of Science and Technology Information,
[3]Sungkyunkwan University
athiruj@kisti.re.kr, danyou7im@gmail.com, jhm@kisti.re.kr

정답 템플릿과 매칭 기반 통합을 이용한 다중 LLM RAG 플랫폼

Athiruj Poositaporn[12], 김재현[3], 정한민[12]
[1]과학기술연합대학원대학교, [2]한국과학기술정보연구원, [3]성균관대학교

Abstract

Retrieval-Augmented Generation (RAG) often depends on a single LLM, which can lead to hallucinations and limited domain-specific knowledge. This study proposes a multiple LLM RAG platform that uses an answer template with matching-based integration to combine outputs from different models. Our experiment showed that using meaning-based matching for word similarity metric produced better outputs than word-based matching. Future work will enable users to choose the answer template and the number and type of LLMs, extend evaluation with large-scale datasets and performance metrics such as F1, ROUGE, and BLEU, apply the approach to diverse query types, and incorporate algorithm-based result verification.

I. Introduction

RAG has emerged as a valuable framework for generating responses grounded in relevant knowledge [1–2]. However, most implementations rely on a single LLM, which could lead to limited domain knowledge and hallucinations [3]. Additionally, recent work by S. Chakraborty et al. (2025) has explored multiple LLMs in a RAG architecture, but it focuses on evaluation rather than methods for integrating outputs into a unified response [4]. To address this, we propose a multiple LLM RAG platform that applies an answer template and matching-based integration. In this study, we experiment on two integration methods: word-based matching (baseline) and meaning-based matching (proposed).

II. Multiple LLM RAG Platform

Figure 3 illustrates how information flows from the user prompt through to the final integrated results. The architecture is composed of four sequential modules:

1. LLM-based Question Answering: A user's query is combined with a 5W1H answer template and passed into the $LLM_{A/B}$ producing $Answer_{A/B}$ [5]. In this study, we use GPT-4o and Gemini 2.5 Flash for the LLM A and B. Each answer contains a set of words for each 5W1H component, as shown in Figure 1.

Q8: What did CERN do in 2012?

Answer$_A$ (GPT-4o):
Who = ["CERN", "Physicists", "ATLAS and CMS teams"]
What = ["Higgs boson", "Particle discovery", "Scientific breakthrough"]
When = ["2012", "July 4", "Early 21st century"]
Where = ["Large Hadron Collider", "Geneva", "Switzerland"]
Why = ["Standard Model validation", "Mass origin", "Fundamental physics"]
How = ["Proton collisions", "High-energy experiments", "Data analysis"]

Answer$_B$ (Gemini 2.5 Flash):
Who = ["CERN scientists", "ATLAS", "CMS"]
What = ["Higgs boson", "Discovery announcement"]
When = ["July 4, 2012"]
Where = ["Geneva", "Switzerland"]
Why = ["Standard Model", "Particle physics"]
How = ["Large Hadron Collider", "Particle collisions"]

Figure 1. An Example of The Answer$_{A/B}$

2. Classification-based Template Filling: $Answer_{A/B}$ are then transformed into six binary word-similarity matrices (one for each 5W1H component) using two methods: (i) word-based matching (baseline) by comparing exact word matches and (ii) meaning-based matching (proposed) by considering semantic similarities using Qwen3-Embedding-0.6B. A cosine of 0.7 is chosen as previously established as effective in semantic similarity evaluation [6]. The $Binary\ Matrices_{P/Q}$ are then classified to determine a relationship type on each 5W1H component using the following conditions:

$$K(M) = \begin{cases} Equal\ Sets & a = m \wedge b = n, \\ Disjoint\ Sets & c = 0 \wedge (m + n > 0), \\ Proper\ Subset & (a = m \wedge b < n) \vee (b = n \wedge a < m), \\ Overlapping\ Sets & otherwise \end{cases} \quad (1)$$

Where $M \in \{0,1\}^{m \times n}$ is the binary matrix of two word sets $p = \{w_1, w_2, \ldots, w_m\}$ *and* $q = \{w_1, w_2, \ldots, w_n\}$ from a given 5W1H component. m,n are sizes of the p *and* q. $c = \sum_{ij} M_{ij}$ is the total count of matches. $a = \sum_i 1\left(\sum_j M_{ij} > 0\right)$ is the number of elements in p matched at least one element in q, while the $b = \sum_j 1\left(\sum_i M_{ij} > 0\right)$ is the reverse. Thus, $k(M)$ classifies the relationship between p *and* q as one of four types: Equal Sets, Disjoint Sets, Proper Subset, and Overlapping Sets. Each relationship type is associated with an integration rule that specifies how the word sets p *and* q are merged across the 5W1H components, as shown in Figure 2. The outputs of this module are the $Fill\text{-}in\ Template_{P/Q}$.

Disjoint sets: Answer the question based on the intersection elements: $\{w_1, \ldots, w_n\}$. Also, give the answer with distinct aspects using the elements: $\{w_1, \ldots, w_n\}$ and another answer using the elements: $\{w_1, \ldots, w_n\}$.
Proper subset: Answer the question based the intersection elements: $\{w_1, \ldots, w_n\}$. Also, give the answer with distinct aspect on the relative complement elements: $\{w_1, \ldots, w_n\}$.
Equal sets: Answer the question based on the elements: $\{w_1, \ldots, w_n\}$.
Overlapping sets: Answer the question based on the intersection elements: $\{w_1, \ldots, w_n\}$. Also, give the answer with distinct aspects using the elements: $\{w_1, \ldots, w_n\}$ and another answer using the elements: $\{w_1, \ldots, w_n\}$.

Figure 2. Classification Relationship Types

3. Template-based Prompt Integration: The user query, the $Fill\text{-}in\ Template_{P/Q}$, and the system prompt are merged to form the $Integrated\ Prompt_{P/Q}$.
4. Integrated Result Generation: A LLMC (Sonnet-4) generate the Integrated ResultP/Q and return to the user interface.

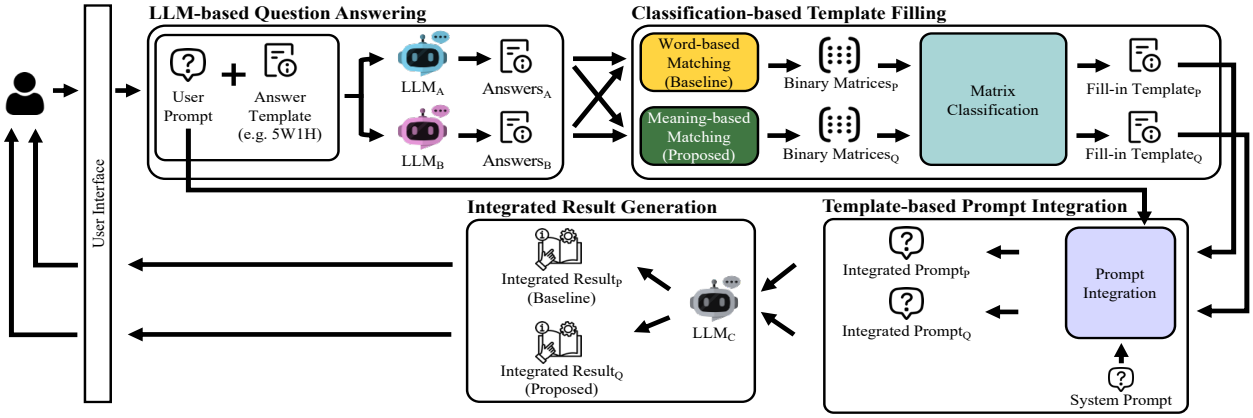


Figure 3. Multiple LLM RAG Platform Architecture

Table 1. Classification Results from Fill-in Template$_{P/Q}$ by Each Question and 5W1H Component

Question	Q1		Q2		Q3		Q4		Q5		Q6		Q7		Q8		Q9		Q10	
Method	WM	MM	WM	MM	WM	MM	WM	MM	WM	MM	WM	MM	WM	MM	WM	MM	WM	MM	WM	MM
Who	D	P	O	E	D	P	P	P	D	O	P	E	D	E	D	P	D	P	O	P
What	D	P	O	O	D	O	D	O	D	O	D	D	D	P	O	P	O	P	O	O
When	D	P	D	E	P	P	D	P	D	P	E	E	D	P	D	P	E	E	D	P
Where	P	E	D	O	D	P	D	D	O	E	D	P	P	P	P	P	D	E	D	P
Why	O	O	D	D	O	O	D	O	D	O	D	O	D	O	D	P	D	O	O	P
How	D	P	O	O	D	E	D	D	D	O	D	D	D	E	D	O	D	D	D	O

Word-based Matching (WM) | Meaning-based Matching (MM) | Equal Sets (E) | Proper Subset (P) | Overlapping Sets (O) | Disjoint Sets (D)

Table 2. A Generated Result Example of Word-based and Meaning-based Matching Methods

Q8: What did CERN do in 2012?
Word-based Matching: In 2012, **CERN, along with Physicists and the ATLAS and CMS teams, and separately CERN scientists with ATLAS and CMS,** achieved a major milestone with the Higgs boson—a Particle discovery regarded as a Scientific breakthrough and marked by a Discovery announcement. **The key moment occurred on July 4 in the Early 21st century, with July 4, 2012 as the exact date. The work took place in Geneva, Switzerland, specifically at the Large Hadron Collider. It mattered for Standard Model validation, explaining Mass origin, and advancing Fundamental physics, as well as addressing crucial questions in the Standard Model and Particle physics. The result was achieved through Proton collisions, High-energy experiments, and Data analysis, powered by the Large Hadron Collider enabling Particle collisions.**
Meaning-based Matching: In the Early 21st century, specifically on July 4, 2012, CERN, together with Physicists from the ATLAS and CMS teams — with a distinct contribution from CMS — announced a Scientific breakthrough: the Particle discovery of the Higgs boson. **This occurred in Geneva, Switzerland, at the Large Hadron Collider, where Proton collisions and other High-energy experiments were conducted. To achieve Standard Model validation and advance Fundamental physics, particularly clarifying Mass origin, the teams relied on extensive Data analysis of the resulting Particle collisions.**

III. Integrated Results Comparison between Word-based and Meaning-based Matching Methods

In the experiment, we asked 10 questions to the platform and the results of in Table 1 showed that word-based matching (WM) produced 41 disjoint sets, 2 equal sets, 6 proper subsets, and 11 overlapping sets. In contrast, meaning-based matching (MM) produced 6 disjoint sets, 11 equal sets, 25 proper subsets, and 18 overlapping sets. The dominance of disjoint sets in WM forced the model to responses long and wordy, while MM's higher rate of equal sets, proper subsets, and overlapping sets reduced redundancy and produced shorter, natural answers.

As an evident shown in Table 2. The WM output included repeated details, creating a wordy and less natural response. In contrast, the MM output was more concise, combining overlapping meanings into smoother text while still covering the same key information.

IV. Conclusion

This study proposed a multiple LLM RAG platform that combines multiple models using answer templates and matching-based integration. The experiments showed that meaning-based matching gave better results than word-based matching by producing shorter and clearer answers.

For future work, we will allow users to choose different answer templates and select how many and which LLMs to include. We also plan to conduct large-scale experiments on diverse datasets and evaluate the platform using performance metrics such as F1, ROUGE, and BLEU. Furthermore, we aim to apply the approach to a wider range of query types to test its adaptability across various scenarios. Finally, an algorithm-based verification step for the integrated result will be added, since the current process still requires manual checking.

Acknowledgment

This work was supported by UST Young Scientist+ Research Program 2024 through the University of Science and Technology. (No. 2024YS12)

References

[1] A. Poositaporn and H. Jung, "On RAG Framework Combined with Query Decomposition for Enhancing Client Engagement," *in Proceedings of KSII Spring Conference*, 2025.

[2] A. Poositaporn, H. Jung, and D. Lee, "A Study of a SQL-Generating RAG System for Improving Client engagement," *in Proceedings of The 17th International Conference on Future Information & Communication Engineering*, 2025.

[3] P. Feldman, J. R. Foulds, and S. Pan, "RAGged Edges: The Double-Edged Sword of Retrieval-Augmented Chatbots," *arXiv preprint arXiv:2403.01193*, 2024.

[4] S. Chakraborty et al., "A scalable framework for evaluating multiple language models through cross-domain generation and hallucination detection," *Scientific Reports*, 2025.

[5] Y. Cao et al., "5W1H extraction with large language models," *in Proceedings of Int. Joint Conf. Neural Netw. (IJCNN)*, pp. 1–8, 2024.

[6] T. J. Cann et al., "Using semantic similarity and text embedding to measure the social media echo of strategic communications," *arXiv preprint arXiv:2303.16694*, 2023.

전략적 의사결정에서 LLM의 편향과 유연성 탐색: 시나리오 기반 벤치마크 분석

최덕진
삼성전자
deokjin.choi@gmail.com

Exploring Bias and Flexibility in LLM Strategic Decision-Making: A Scenario-Based Benchmark Analysis

Deokjin Choi
Samsung Electronics

요 약

대규모 언어모델(Large Language Models, LLM)은 질의응답, 번역과 같은 전통적 과제에서는 우수한 성능을 보여주었지만, 실제와 유사한 경영 상황에서 전략적 선택을 어떻게 내리는지는 아직 충분히 규명되지 않았다. 그러나 이러한 맥락적 의사결정 능력은 LLM이 단순 도구를 넘어 의사결정 지원 시스템으로 확장될 수 있는지를 가늠하는 핵심 요소이다. 본 연구는 테슬라의 역사적 사례를 기반으로 여섯 개의 시나리오를 설계하고, 이를 익명(Generic)과 기업명(Specific) 조건으로 구분하여 제시하였다. 또한 각 시나리오에는 시장, 기술, 정책적 맥락을 추가하거나 변형하여 다양한 상황을 재현하였다. 이를 통해 모델이 선택하는 전략의 분포를 비교하고, 기업명 명시 여부와 맥락 변화가 전략적 판단에 어떤 영향을 주는지 분석하였다. 실험 결과, LLM은 전반적으로 기술 리더십(Technology Leadership) 전략을 과도하게 선호하는 경향을 보였으며, 기업명이 명시될 경우 전략 선택이 달라지는 브랜드 편향이 확인되었다. 또한 주어진 맥락이 긍정적 또는 부정적으로 변형될 때, 전략 분포가 뚜렷하게 이동하거나 분리되는 양상이 관찰되었다. 이러한 결과는 LLM이 단순히 사실 기반 답변을 제공하는 수준을 넘어, 맥락과 프레이밍에 민감하게 반응하는 전략적 추론 주체로 작동할 수 있음을 보여준다. 동시에 이러한 편향을 진단하는 과정은, 향후 LLM을 신뢰할 수 있는 의사결정 지원 도구로 발전시키는 데 중요한 단서를 제공한다.

Ⅰ. 서 론

LLM은 다양한 비즈니스 문제 해결에서 활용 가능성이 높지만, 실제 전략적 의사결정 맥락에서 어떤 편향과 추론 패턴을 보이는지는 충분히 밝혀지지 않았다. 본 연구는 이를 검증하기 위해 시나리오 기반 전략 벤치마크를 설계하였다.[1]

연구 방법론은 다음과 같다.

1.테슬라의 역사적 6개 시나리오(창업기, Roadster, Model S, Model X, Model 3, 에너지 인프라)를 기반으로 베이스 문제를 작성하였다.

2.각 시나리오는 익명(Generic)과 기업명 명시(Specific: Tesla) 두 가지 방식으로 제시하였다.

3.각 시나리오에는 시장·기술·정책적 맥락을 추가하거나 삭제하였으며, 특히 네 가지 변형 조건을 설계하였다.

-경쟁환경 정보 추가 (competitive_dynamics)

-부정적 정보 반영 (count_fact)

-긍정적 기회 강조 (opp_focus)

-수치 데이터 ±20% 변동 (randomized_numbers)

4.이를 통해 (a)특정 전략 선택 비율, (b)변형 시나리오 - 베이스 차이, (c)PCA 기반 시나리오 간 유사도를 도출하였다.[2]

Ⅱ. 본론

2.1 시나리오별 전략 선택 비율

본 연구에서 사용된 시나리오는 모두 문제 제시, 맥락 정보, 실행 옵션으로 구성된다. 구체적인 형식은 <그림 1>의 창업기(Founder Period) 시나리오 예시를 통해 확인할 수 있다.

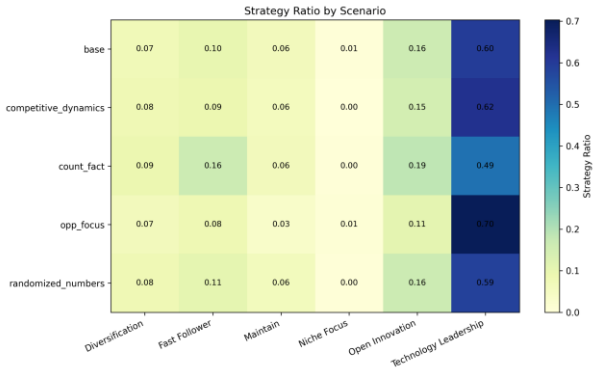
<그림 1. Base 시나리오 예시>

```
{
  "1_founder_period": {
    "problem_generic": "Problem: A new company must select an initial market entry strategy to prove its
    "problem_specific": "Problem: Tesla, a new company, must select an initial market entry strategy to
    "context_blocks": {
      "[Market] Global new car sales: ~60 million units/year.": "Global new car sales were roughly ~60 m
      "[Market] EV share: <0.02% of global new car sales.": "Battery electric vehicles accounted for les
      "[Technology] Li-ion cell cost: ~1,000 USD/kWh.": "Lithium-ion cell costs were around ~1,000 USD p
      "[Technology] Prototype performance: 0–60 mph in 3.6 s; ~300 km range.": "A high-performance proto
      "[Finance] Initial funding: ~USD 7.5M Series A (private).": "Initial external funding amounted to
      "[Market] Incumbent focus: Established automakers concentrated on hybrid and fuel cell technologie
    },
    "execution_options": {
      "A": { "name": "Build a portfolio of hybrid/fuel cell vehicles", "mapping": "Fast Follower" },
      "B": { "name": "Develop a high-performance EV sports car", "mapping": "Technology Leadership" },
      "C": { "name": "Co-develop an EV platform with a major automaker", "mapping": "Open Innovation" }
    }
  },
```


이 예시에서 문제는 “신생 기업(또는 Tesla)이 초기 시장 진입 전략을 선택해야 한다”는 것이다. 맥락으로는 당시 전기차 시장 점유율이 0.02% 미만, 배터리 단가가 약 1,000 USD/kWh 수준이라는 조건이 주어졌다. 이에 따라 제시된 실행 옵션은 (A) 하이브리드·연료전지 진입(Fast Follower), (B) 고성능 전기 스포츠카 개발(Technology Leadership), (C) 대형 완성차와 공동 플랫폼 개발(Open Innovation)이다.

이처럼 각 시나리오는 동일한 구조를 따르며, 이후 장에서는 이러한 설정에 따라 도출된 전략 선택 비율과 그 차이를 분석한다.

<그림 2. Strategy Ratio by Scenario Heatmap>



분석 결과, 모든 시나리오에서 Technology Leadership(TL) 전략이 가장 높은 비중을 차지했으며, Niche Focus와 Maintain은 지속적으로 낮았다. 그러나 케이스별 차이가 뚜렷했다.

-competitive_dynamics: TL이 base 대비 소폭 강화(0.60→0.62), 큰 변화 없음. Tesla framing이 들어가면 TL 선택이 고정되는 앵커링 효과.

-count_fact: TL이 크게 감소(0.60→0.49), 대신 Fast Follower(0.16)와 Open Innovation(0.19) 상승. 불리한 정보 삽입 시 공격적 전략 대신 보수적·협력적 전략으로 이동.

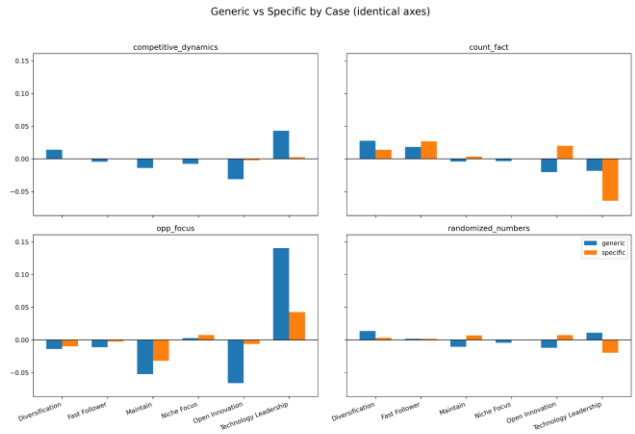
-opp_focus: TL이 크게 강화(0.70). 기회 강조 맥락에서 TL 쏠림이 극대화됨.

-randomized_numbers: base와 거의 동일. 수치 변동은 전략 선택에 유의미한 영향 없음.

종합하면, TL은 모든 조건에서 기본 전략이지만, 불리한 맥락에서는 약화되고 대안 전략으로 분산되며, 기회 강조에서는 TL 쏠림이 심화되는 패턴이 확인되었다.

2.2 변형 시나리오 - 베이스 차이

<그림 3. Delta from Base (Generic/Specific)>



변형 시나리오 – 베이스 차이를 Generic과 Specific 조건별로 비교한 결과는 다음과 같다.

-공통 패턴: TL에서 generic과 specific 간 차이가 가장 크며, Tesla 명시 시 TL이 약화되거나 상승폭이 줄어드는 경향이 나타났다.

-competitive_dynamics: generic에서는 TL이 +0.043 상승했으나 specific에서는 거의 0으로 수렴. 브랜드 안정화 효과.

-count_fact: generic은 TL −0.018, specific은 −0.064. Tesla 명시 시 불리한 정보에 과잉 반응하여 TL 억제, 대신 Fast Follower와 Open Innovation 강화.

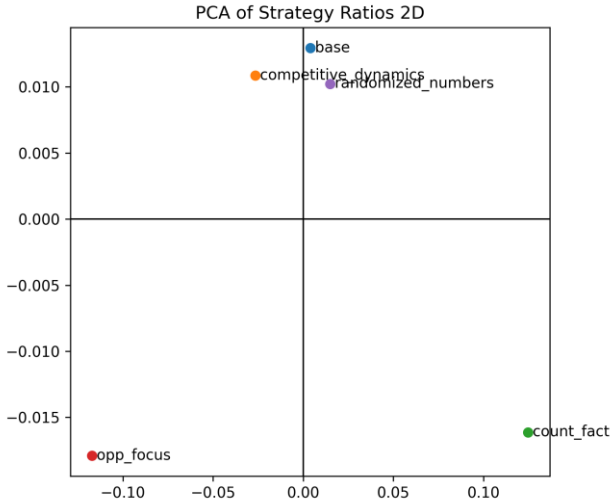
-opp_focus: generic은 TL +0.140의 강한 상승, specific은 +0.043으로 상승폭 축소. Tesla framing이 보수화를 유도.

-randomized_numbers: generic은 TL 소폭 양수(+0.011), specific은 음수(−0.019)로 반전. Tesla framing이 리스크 내러티브를 과잉 반영.

종합하면, 브랜드 효과는 단순 안정화만이 아니라, 불리한 정보나 불확실성이 추가될 때 공격적 전략(TL)을 억제하고 대안 전략으로 왜곡 이동시키는 효과를 동반함을 확인했다.

2.3 시나리오 간 유사도 (PCA 분석)

<그림 4. PCA of Strategy Ratios 2D>



시나리오별 전략 선택 비율 행렬에 PCA를 적용하여 2차원 공간에 투영하고, 이를 통해 시나리오 간 전략 분포의 상대적 유사성을 확인하였다.

분석 결과:

-base, competitive_dynamics, randomized_numbers는 서로 근접해 군집을 형성하였다.

-opp_focus와 count_fact는 멀리 떨어져 분리되어, 특정 맥락(긍정적 기회, 부정적 사실)이 전략 선택 패턴을 뚜렷하게 변화시킴을 확인할 수 있었다.

이는 LLM의 전략 선택 분포가 조건별로 명확히 구분되며, PCA가 이러한 차이를 효과적으로 드러내는 방법임을 보여준다.

Ⅲ. 결론

본 연구는 LLM의 전략적 추론 특성을 세 가지 지표로 분석했다. 전략 선택 비율은 TL이 기본 선택이며 맥락에 따라 강화 혹은 약화되는 양상을 보였고, 변형 시나리오 – 베이스 차이에서는 Tesla 명시 시 TL이 억제되며 불리한 맥락에서는 대안 전략으로 이동하는 패턴이 확인되었다. 또한 PCA 기반 유사도 분석에서는 특정 맥락(opp_focus, count_fact)에서 전략 분포가 뚜렷하게 분리되는 결과가 나타났다. 이 결과는 LLM이 단순한 의사결정 도구가 아니라 맥락과 브랜드 내러티브에 따라 전략적으로 편향된 선택을 할 수 있음을 보여준다.[3]

참 고 문 헌

[1] Zhang et al., "LLM as a Mastermind: A Survey of Strategic Reasoning with Large Language Models," arXiv, 2024.

[2] Jia et al., "Large Language Model Strategic Reasoning Evaluation through Behavioral Game Theory," arXiv, 2025.

[3] Lorè & Heydari, "Strategic behavior of large language models and the role of game structure versus contextual framing," Scientific Reports, 2024.

LoRA 기반 LLM 미세 조정 기법을 통합한 연구 논문 분류에 관한 연구

딥또 비스와스[1], 변태영[2], 길준민[3], 조재춘[4*]
[1,2]대구가톨릭대학교 컴퓨터소프트웨어학과, [3]제주대학교 컴퓨터공학과, [4]제주대학교 컴퓨터교육과
[1]dipto.biswas94@gmail.com, [2]tybyun@cu.ac.kr, [3]jmgil@jejunu.ac.kr, [4]jjo@jejunu.ac.kr

A Study on Research Paper Classification Using Fine-Tuning LLM Based on LoRA

Dipto Biswas[1], Tae-Young Byun[2], Joon-Min Gil[3], Jaechoon Jo[4*]
[1,2]Dept. of Computer Software Eng., Daegu Catholic Univ., [3]Dept. of Compuer Engineering, Jeju National Univ., [4]Dept. of Computer Education, Jeju Nation Univ.

요 약

대형 언어 모델(Large Language Models, LLM)은 텍스트 분류를 포함한 다양한 자연어 처리(Natural Language Processing, NLP) 작업에서 뛰어난 성능을 입증해 왔다. 그러나 이 모델을 특정 작업에 효과적으로 적용하려면 정교한 미세 조정(fine-tuning)과 고품질의 라벨링된 데이터셋이 필요하며, 이는 전문 학술 분야에서는 흔치 않다. 본 논문에서는 대형 언어 모델인 Meta AI의 LlaMA-3(3B)를 LoRA(Low-Rank Adaptation)을 활용하여 미세 조정함으로써, 정확하고 효율적인 연구 논문 분류(Research Papers Classification, RPC)를 가능하게 하는 접근 방식을 제안한다. LoRA는 사전학습된 모델을 소수의 파라미터만 업데이트함으로써 적은 비용으로도 높은 성능을 유지하면서 모델을 적응시킬 수 있게 한다. 본 논문에서는 LoRA 기반 미세 조정 기법을 RPC에 적용하고, 분류 시나리오에서 Top-n 기법을 활용하여 모델의 성능을 평가한다. 특히, 검색 질의의 맥락에 적합한 사용자의 맞춤형 추천을 제공하기 위해 의미론적 관계를 활용하는 Top-n 라벨링 기법을 제안한다. 실험 결과, LoRA를 적용한 LlaMA-3 모델이 기존의 기본 미세 조정 기법보다 높은 분류 성능을 보였으며, 이는 대규모 학술 연구 환경이나 자원이 제한된 시스템에서 제안된 분류 및 추천 방식이 적합한 방식임을 보여준다.

Ⅰ. 서 론

최근 연구 논문의 발행이 기하급수적으로 증가함에 따라, 연구자들이 관련 논문을 정확하게 식별하고 접근하는 것이 점점 더 어려워지고 있다[1]. 디지털 라이브러리 플랫폼이나 학술 저널 저장소는 연구 분야나 발행일 기준으로 콘텐츠를 관리하고 있으며, 기존의 전통적인 방식들은 학술 문헌의 복잡성과 최근의 기술적 특성을 효과적으로 반영하지 못한다. 기존의 인덱싱 및 필터링 시스템은 의미론적 이해가 부족하여 비효율적인 추천 결과를 초래하며, 정보 과부하 문제를 유발하고 있다. 이로 인해 연구자들은 방대한 양의 문헌을 수동으로 탐색해야 하며, 이는 많은 시간 소모를 유발하므로 비효율성을 초래한다. 이러한 정보 검색의 비효율성은 연구자의 생산성과 학술 연구의 전반적인 품질에도 부정적인 영향을 미친다. 특히, 융합 연구가 증가함에 따라 전통적인 분류 체계는 서로 다른 연구 주제 간의 미묘한 관계를 정확하게 반영하지 못하고 있으며, 여러 학문 분야에 걸친 주요 연구의 식별을 더욱 어렵게 만들고 있다. 연구자들은 핵심 논문을 놓치거나 관련성이 낮은 콘텐츠를 검토하느라 불필요한 시간을 소모할 수 있다. 이러한 문제를 해결하기 위해, 본 연구는 LLaMA-3 모델의 의미론적 이해 능력과 LoRA의 경량 적응성을 결합하여 이를 RPC 작업에 적용하고자 한다. 또한, 의미적 관계를 포착하고 검색 질의의 맥락에 맞는 정밀한 사용자의 중심 추천을 제공하기 위해 Top-n 라벨링 기법을 도입한다. 본 연구의 제안 기법을 통해 사용자의 선호도에 맞는 연구 논문을 분류하고 추천하는데 활용하고자 한다.

Ⅱ. 시스템 모델 및 방법론

본 연구에서 제안하는 LLM 기반 RPC는 구조화된 파이프라인으로 시작되며, 다음과 같은 과정을 따른다. 먼저, 제목, 초록, 키워드(선택적으로 전체 텍스트나 메타데이터 포함) 등의 정보를 포함한 논문 데이터셋 D가 입력으로 주어진다. 각 논문 $p \in D$에 대해 텍스트 입력 $RPA = \{p_1, p_2, \ldots, p_n\}$는 연구 논문 초록의 정규화 및 전처리를 의미한다. 데이터처리 단계에서는 NLTK 모듈을 이용하여 원시 텍스트에서 문장 부호를 제거하고 불용어, 표제를 추출하며, 전체 단어를 명사형으로 변환하여 표현의 명확성, 일관성 및 일반화를 도모한다. 이를 통해 추상적 개념이나 행위를 명확한 개체로 표현할 수 있어 정보를 더욱 형식적으로 구조화할 수 있다. 전처리 함수 P는 다음과 같이 표현된다.

$$P(RPA) \rightarrow RPALSUP' \quad (1)$$

여기서 $RPALSUP'$는 전처리된 연구 논문 초록을 나타낸다.

그 다음 단계에서는 $RPALSUP'$를 LlaMA-3 모델에 맞는 SentencePiece 토크나이저로 토크화하고, 고정된 최대 토큰 길이 L (예: $L = 512$)로 자르거나 패딩 처리한다. 이로 인해 입력 시퀀스 X는 다음과 같이 정의된다.

$$X = \{x_1, x_2, \ldots, x_L\},\ x_i \in V \quad (2)$$

여기서 V는 어휘 공간을 의미한다. 입력 시퀀스 X는 임베딩되어 LlaMA-3 모델에 입력된다.

언어 모델로 사용하는 LlaMA-3는 다국어 및 다양한 도메인의 대규모 말뭉치에 대해 사전학습된 디코더 기반 트랜스포머 모델이다. 이 모델은 다층의 트랜스포머 블록으로 구성되어 있으며, 각 블록은 멀티-헤드 자기 주의(Multi-Head Self-Attention, MHSA)와 위치별 피드포워드 네트워크(Position-wise Feedforward Network, FFN)를 포함하고 있다. 입력 X에 대해 모델은 다음과 같이 문맥 임베딩 H를 생성한다.

$$H = Transformer(X) = LN(X + MHSA(X) + FFN(X)) \quad (3)$$

여기서 $MHSA$는 멀티-헤드 자기 주의, FFN은 위치별 피드포워드 네트워크, LN은 계층 정규화(Layer Normalization)를 의미한다.

LlaMA-3의 전체 파라미터를 조정하는 대신, 본 연구에서는 LoRA(Low-Rank Adaptation)[2]를 활용하여 어텐션(attention) 계층에 소규모 저랭크 행렬을 삽입해 선택적으로 미세 조정한다. 구체적으로, LoRA는 쿼리 W_Q와 값 W_V 행렬에 대해 훈련 가능한 랭크 분해 행렬 $A \in RLSUPd \times r$와 $B \in RLSUPr \times d$를 삽입한다. 업데이트된 가중치 행렬 $\widetilde{W}$는 다음과 같이 표현된다.

$$\widetilde{W} = W + \triangle W = W + \alpha \cdot AB \quad (4)$$

여기서 W는 고정된 사전학습 가중치, $\triangle W = AB$는 저랭크 적응 파라미터, α는 스케일링 계수이다.

연구 논문에 대한 입력 시퀀스를 받은 사전학습된 LlaMA 모델은 문맥 임베딩으로 인코딩한다. 이후 풀링 전략을 통해 최종 토큰의 임베딩

에서 전역 표현 $h \in RLSUPd$를 추출하고, 이를 LoRA가 적용된 선형 분류기에 입력하여 로짓(logits) $z \in RLSUPC$을 생성한다.

$$z = W_{cls}h + b + \alpha \cdot ABh \quad (5)$$

여기서 $W_{cls} \in RLSUPC \times d$, $b \in RLSUPC$는 분류기 층의 학습 가능한 파라미터이다. W_{cls}는 임베딩 h를 클래스 수 C에 해당하는 공간으로 투영하며, b는 각 클래스에 대한 편향 벡터이다.

로짓은 소프트맥스 함수로 변환되어 각 클래스에 대한 확률 분포를 형성한다.

$$\hat{y_i} = softmax(z_i) = \frac{eLSUPz_i}{\sum_{j=1}^{C} LSUPeLSUPz_j}, \ \forall_i \in \{1,2,...,C\} \quad (6)$$

여기서 z_i는 i-번째 클래스에 대한 로짓이며, $\hat{y_i}$는 예측 확률, C는 클래스의 총 개수이다. 클래스는 각각 Top-n 레이블을 기준으로 5와 10으로 설정하였다.

최종적으로, 예측된 레이블은 가장 높은 확률을 가진 클래스로 정의된다.

$$\hat{y} = \arg\max_i \hat{y}_i \quad (7)$$

본 연구에서는 단일 레이블 다중 클래스 분류에 적합한 범주형 교차 엔트로피(Categorical Cross-Entropy, CCE) 손실 함수를 사용하여 모델을 최적화한다. 한 샘플에 대한 손실은 다음과 같이 정의된다.

$$L_{CCE} = -\sum_{i=1}^{C} y_i \log(\hat{y}_i) \quad (8)$$

여기서 $y_i \in \{0,1\}$은 원-핫 인코딩된 정답 벡터이며, $\hat{y}_i$는 예측된 클래스 확률이다. 이 손실 함수는 상호 배타적인 클래스 예측을 유도하며, 단일 레이블 분류 작업에 효과적인 최적화를 가능하게 한다.

성능 평가를 위해서는 [3]의 성능 지표를 기반으로 소프트맥스 함수에서 가장 높은 점수를 갖는 클래스를 예측 결과로 선택하며, 정확도(accuracy), 정밀도(precision), 재현율(recall), F1-점수(F1-score) 등의 지표를 통해 라벨이 부착된 테스트 세트에서 모델의 성능을 평가한다.

Ⅲ. 실험 환경 및 결과

(1)데이터셋: 본 연구에서 활용한 데이터셋은 표 1과 같다. 표 1의 데이터셋은 "HuggingFace" 플랫폼의 API를 활용하여 수집되었다. 각 데이터셋은 논문 ID, 제목, 저자, 초록, 키워드, 출판 연도 및 추가적인 서지 정보 등으로 구성된 메타데이터 분야를 포함하고 있으며, 다양한 저널에 걸쳐 여러 학문 분야를 폭넓게 포괄하고 있다. 본 연구에서는 최신 연구 동향을 반영하기 위해 2023년에 출판된 논문들만을 선정하였다.

표 1. 본 연구의 실험에 활용한 데이터셋.

	arXiv	Elsevier
논문	35,000	32,072
문장	232,993	278,437
단어	1,217,714	1,496,324
학습 세트	28,000	25,657
테스트 세트	7,000	6,415
Top-n 라벨	5, 10	5, 10
평균 길이	278.65	370.72

(2) 모델 하이퍼파라미터: 이 설정에서는 "open_LlaMA_3b" 모델을 LoRA 구성 방식을 통해 전체 파라미터의 0.16%인 5,324,800개의 파라미터만 학습하도록 미세 조정하였다. 학습률은 5×10^{-5}, 학습 배치 크기는 8, 평가 배치 크기는 16이며, 총 5회의 에포크 동안 학습이 진행되었다. 과적합 방지를 위해 가중치 감쇠 $\lambda = 0.01$이 적용되었고, LoRA 관련 하이퍼파라미터로는 랭크 $r = 16$, 스케일링 계수 $\alpha = 16$, 드롭아웃 비율 0.5이다. 수치적 안정성을 위한 $\epsilon = 1\times10^{-6}$이 사용되며, 학습 상태는 10 스텝마다 기록된다. 이 구성은 전체 모델이 아닌 저차원 행렬에 집중해 효율적인 미세 조정을 가능하게 한다.

Ⅳ. 실험 결과

표 2와 그림 1은 LlaMA-3(LoRA)가 2가지 데이터셋(arXiv, Elsevier)의 Top-5 및 Top-10 레이블 설정에서 달성한 성능 향상을 보여준다. 특히, LlaMA-3(LoRA)는 정확도(ACC)와 F1-점수에서 기존 모델 대비 10% 이상의 성능 향상을 달성하였으며, 이는 최소한의 파라미터 조정만으로 모델을 최적화하는 LoRA 기반의 미세 조정과 논문 간 의미적 관계를 효과적으로 반영하는 Top-n 라벨링 방식 때문이다. 이러한 접근 방식은 개인화된 논문 분류(RPC) 작업에서 보다 정밀한 추천을 가능하게 한다. 특히, LlaMA-3(LoRA)는 Gemini-pro(S) 및 GPT-4 대비 12~15%의 성능 향상을 보여주며, 2가지 데이터셋에 대한 유연성과 도메인 특화된 학습에 대한 적응 능력을 보여준다.

표 2. Top-5 라벨 기반 제안 모델의 실험 결과

Model \ 데이터셋	arXiv		Elsevier	
	정확도	**F1-점수**	**정확도**	**F1-점수**
Gemini-pro(S)	0.4988	0.4780	0.4667	0.4546
GPT-4	0.5267	0.5059	0.4876	0.4820
LlaMA-2(8B)	0.5363	0.5298	0.4718	0.4258
LlaMA-3(LoRA)$^{Top-5}$	**0.6732**	**0.6529**	**0.7048**	**0.6786**

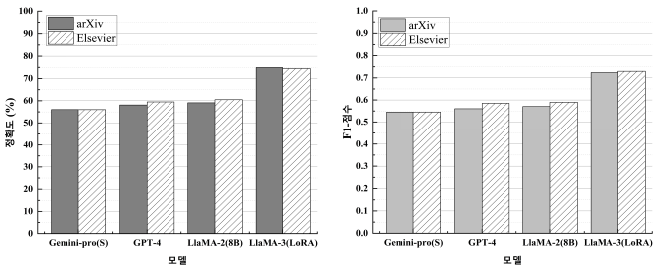


그림 1. Top-10 라벨 기반 제안 모델의 실험 결과

Ⅴ. 결론

본 연구에서는 LlaMA-3 모델을 LoRA로 미세 조정하여 연구 논문 분류(RPC)를 수행하는 접근 방식을 제시하였다. 제안한 모델은 사용자 맞춤 추천을 위한 분류 성능을 향상시키며, 실험 결과는 LoRA 기반 미세 조정이 Top-n 라벨링 방법과 결합될 때 분류 정확도를 크게 향상시킬 수 있음을 보여주었다. 제안 모델은 연구 논문 간의 의미적 관계를 효과적으로 포착하여 보다 정밀한 추천을 가능하게 해준다. 실험 결과는 LlaMA-3(LoRA)가 다양한 데이터셋에서 정확도 및 F1 점수 측면에서 기존 모델보다 높은 성능을 가짐을 보여준다. 또한, Top-n 라벨링의 도입은 보다 정확하고 사용자 중심적인 추천을 가능하게 하며, 학제 간 연구의 증가로 인한 복잡성을 해결하고 기존의 키워드 기반 분류 시스템의 한계를 극복하는데 기여할 수 있다. 한편, 모델의 적응성, 정확도, 개인화된 추천 성능을 향상시키기 위해 LoRA-FA(Feature Augmentation), 능동 학습, 멀티모달 데이터, 사용자 피드백을 통합하는 방안을 향후 연구로 수행할 예정이다.

ACKNOWLEDGMENT

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(RS-2025-16072365). 또한 2019년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 제주대학교 기초과학연구소 자율운영중점연구지원사업에서 수행된 기초연구사업임(RS-2019-NR040080).

참 고 문 헌

[1] Barbu, "Global trends in the scientific research of the health economics: a bibliometric analysis from 1975 to 2022," Health economics review, vol. 13, no. 1, p. 31, 2023.

[2] J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, "Lora: Low-rank adaptation of large language models," Proceedings of the International Conference on Learning Representations (ICLR), 2022.

[3] G. Naidu, T. Zuva, and E. M. Sibanda, "A review of evaluation

metrics in machine learning algorithms," Computer science on-line conference, pp. 15-25, Springer, 2023.

Time-LLM 프레임워크를 통한 시계열 예측 성능에 대한 LLM의 영향 분석

공무경, 임수환, 방대호, 박민수, 강재모*

경북대학교

{rhdanrud1000, tngks9317, dkrlwnstn1, operioncwal, *jmkang}@knu.ac.kr

Impact of LLM Backbones on Time Series Forecasting Performance: A Study Based on the Time-LLM Framework

Kong Mu Kyeong, Im Su Hwan, Bang Dae Ho, Park Min Soo, Kang Jae Mo*

Kyungpook National Univ.

요 약

시계열 예측은 금융, 에너지 등 다양한 분야에서 효율적 자원 분배와 위험 관리에 필수적이다. 그러나 ARIMA나 LSTM과 같은 기존 방법은 데이터 부족 상황에서 일반화 성능에 한계를 보인다. 이에 최근 방대한 사전학습을 통해 강력한 패턴 인식과 제로샷 능력을 지닌 대규모 언어 모델(LLM)을 활용하려는 연구가 주목받고 있다. 본 연구는 Time-LLM 프레임워크에서 서로 다른 LLM 백본을 교체하며 시계열 예측 성능에 미치는 영향을 분석하였다. 첫 번째 실험에서는 Qwen3 계열(4B, 8B, 14B)을 사용해 파라미터 수의 효과를 검증했고 두 번째 실험에서는 BBH 점수가 다른 Qwen3-8B와 Llama3-8B를 비교하여 추론 능력의 기여도를 평가하였다. ETTh1 데이터셋을 대상으로 동일한 학습 환경에서 MSE와 MAE로 성능을 측정한 결과, Qwen3-4B가 가장 낮은 오차를 기록하여 단순히 파라미터 규모가 성능 향상으로 이어지지 않음을 확인하였다. 반면 Qwen3-8B는 Llama3-8B보다 현저히 우수한 성능을 보여, 모델의 추론 능력(BBH)과 아키텍처 등 계열 고유의 특성이 성능에 중요한 영향을 미침을 시사한다.

Ⅰ. 서 론

시계열 예측은 금융, 에너지 등 다양한 산업에서 효율적인 자원 분배 및 위험 관리와 같은 합리적 의사결정에 필수적이다. 기존 ARIMA[1]나 LSTM[2]과 같은 방법론들은 특정 도메인에 특화되어야 하고 데이터가 부족할 경우 일반화 성능에 한계를 보여왔다. 이러한 문제를 해결할 대안으로 최근 방대한 데이터로 사전 학습되어 뛰어난 패턴 인식과 제로샷 학습 능력을 갖춘 대규모 언어 모델(LLM)을 활용한 연구가 활발히 진행되고 있다. LLM은 사전 학습된 지식을 바탕으로 데이터가 부족한 환경에서도 높은 일반화 성능을 기대할 수 있다는 점에서 기존 모델 대비 명확한 이점을 가진다.

그러나 LLM을 시계열 예측에 적용하는 데에는 근본적인 도전 과제, 즉 모달리티 간극이 존재한다. LLM은 이산적인 텍스트 토큰을 처리하도록 설계된 반면 시계열 데이터는 본질적으로 연속적인 숫자 값으로 이루어져 있기 때문이다. Time-LLM[3], Chronos[4], TimeCMA[5] 와 같은 최신 모델들은 이러한 간극을 해소하기 위해 연속적인 시계열을 LLM이 이해할 수 있도록 변환하는 다양한 '입력 표현 전략'을 제시한다. 특히 Time-LLM은 시계열 데이터를 LLM이 이해할 수 있도록 변환하는 Patch Reprogramming 과 도메인·태스크·통계 정보를 접두사로 제공하는 Prompt-as-Prefix(PaP) 를 핵심 아키텍처로 사용한다. 이를 통해 동결된 LLM 백본을 그대로 활용하면서도 시계열 예측이 가능하다.

본 논문은 Time-LLM 프레임워크에서 LLM 백본의 선택이 시계열 예측 성능에 중대한 영향을 미친다는 가설을 검증하고자 한다. 이러한 영향은 모델의 파라미터 규모를 넘어 특정 벤치마크로 측정된 모델의 고유 역량과도 상관관계가 있을 것으로 가정한다. LLM의 파라미터 규모와 추론 능력이 시계열 예측 정확도에 미치는 영향을 분석하여 향후 최적의 LLM 백본을 선택하는 데 기여하고자 한다.

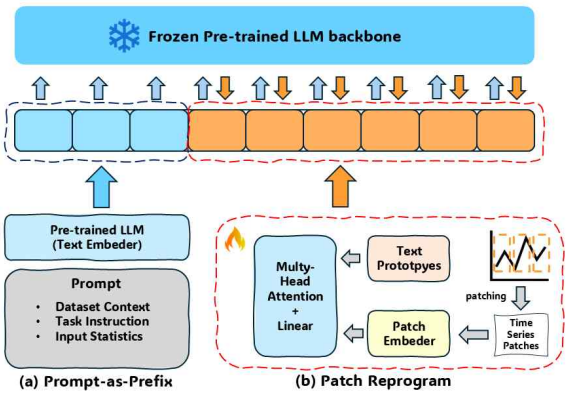


그림 1 Time-LLM 핵심 아키텍처

Ⅱ. 모델 선정 및 실험 설계

2.1. 모델 선정

본 연구는 LLM 백본 모델의 고유 역량이 시계열 예측 성능에 미치는 영향을 검증하기 위해 설계되었다. 이를 위해 모달리티 간극을 효과적으로 해소하면서도 LLM 자체는 수정하지 않는 Time-LLM 프레임워크를 실험 모델로 채택했다. Time-LLM은 LLM의 가중치를 frozen 상태로 유지하므로 백본 모델을 교체했을 때 나타나는 성능 변화를 LLM의 순수한 역량 차이로 귀속시켜 분석하기에 최적의 환경을 제공한다.

첫 번째 실험에서는 LLM의 일반화 능력이 시계열 예측 성능에 미치는 영향을 평가하기 위해 설계한다. 이를 위해 모델 선정 기준으로 파라미터 수를 채택하였다. 파라미터 규모는 모델이 학습 데이터에서 복잡한 패턴을 포착하고 새로운 상황에 적용할 수 있는 잠재력과 직접적으로 연관된다고 알려져 있다[6]. 따라서 동일한 아키텍처 내에서 파라미터 수가 다른

모델들을 백본으로 교체하여 규모 차이가 실제 예측 정확도에 어떠한 변화를 가져오는지를 분석하기 위하여 Qwen3[7] 계열의 4B, 8B, 14B 모델을 사용하였다.

두 번째 실험에서는 LLM의 추론 능력이 예측 정확도에 미치는 기여도를 검증하기 위해 설계되었다. 이를 위해 널리 사용되는 표준 벤치마크인 BIG-Bench Hard (BBH)[8] 점수를 모델 선정 기준으로 활용하였다. BBH는 논리, 상식, 알고리즘적 사고 등 단순한 패턴 학습을 넘어서는 복합적인 문제 해결 능력을 평가하는 지표로서 LLM의 범용적 추론 역량을 반영한다. 따라서 표 1에서 볼 수 있듯 파라미터 규모는 유사하게 유지하되 BBH 점수에서 큰 차이를 보이는 Qwen3-8B(78.40%)와 Llama3-8B[9](57.50%)를 비교 대상으로 선정하였다.

Model	Parameters (B)	BBH Score (%)
Qwen3-4B	4.02B	72.59
Qwen3-8B	8.19B	78.40
Qwen3-14B	14.8B	81.07
Llama3-8B	8.03B	57.70

표 1 실험에 사용된 백본 LLM의 파라미터 수와 BBH점수

2.2. 실험 설계

본 연구의 실험은 Time-LLM 프레임워크를 기반으로 하며 백본 LLM을 교체하는 방식으로 모델 간 성능을 비교하였다. 데이터셋은 ETT-small에 포함된 ETTh1 (Electricity Transformer Temperature, hourly) 시계열을 사용하였다. ETTh1은 전력 변압기의 시간 단위 온도 데이터를 포함한 다변량 시계열로 복잡한 계절성과 추세 패턴을 지니고 있어 시계열 예측 모델의 성능을 평가하는 표준 벤치마크로 널리 활용된다.

본 연구에서는 Time-LLM 프레임워크를 유지한 채 백본 LLM을 Qwen3-4B, Qwen3-8B, Qwen3-14B, Llama3-8B의 네 가지 모델로 교체하여 시계열 예측 성능을 비교 평가하였다. 모든 실험에서 LLM의 가중치는 동결(frozen) 상태로 유지하였으며 동일한 학습 환경 하에서 백본 교체로 인한 순수한 모델 역량 차이를 분석하였다.

성능 평가는 시계열 예측에서 표준적으로 사용되는 평균제곱오차(Mean Squared Error, MSE)와 평균절대오차(Mean Absolute Error, MAE)를 지표로 삼았다. MSE는 큰 예측 오차에 높은 페널티를 부여하여 모델이 이상치와 큰 편차에 얼마나 민감한지를 측정할 수 있고 MAE는 평균적인 예측 오차 크기를 직관적으로 보여준다. 두 지표를 함께 사용함으로써 모델의 정밀성과 안정성을 균형 있게 평가할 수 있도록 설계하였다.

Ⅲ. 실험 결과 및 결론

Model	MSE ↓	MAE ↓
Qwen3-4B	0.3756	0.4027
Qwen3-8B	0.3941	0.4133
Qwen3-14B	0.3821	0.4033
Llama3-8B	0.4731	0.4597

표 2 LLM 백본 모델별 시계열 예측 성능 비교 결과

3.1. 실험 결과

표 2는 Time-LLM 프레임워크에서 서로 다른 백본 LLM을 적용했을 때의 성능 비교 결과를 보여준다. 전체적으로 Qwen3 계열이 Llama3-8B보다 안정적으로 우수했으며, 특히 Qwen3-4B가 가장 낮은 MSE와 MAE를 기록하여 Qwen3-14B와 같이 파라미터 수가 상대적으로 큰 모델과 동등하거나 더 나은 성능을 달성하였다. 이는 모델 규모의 증가가 성능 향상으로 단조적으로 이어지지 않음을 의미한다. 반면, 동일한 8B 규모에서 Qwen3-8B는 Llama3-8B를 크게 상회하여 모델 계열과 추론 능력(BBH)이 성능 결정의 핵심 요인임을 확인할 수 있었다.

3.2. 결론

본 연구는 Time-LLM 프레임워크에서 LLM 백본 교체가 시계열 예측 성능에 미치는 영향을 분석했다. 실험 결과, 파라미터 수의 증가가 예측 정확도의 향상으로 직결되지 않는다는 점이 확인되었다. Qwen3-4B는 8B와 14B와 비교하여 가장 낮은 예측 오차를 기록하였고 이는 일반적으로 모델 규모가 커질수록 성능이 향상된다는 경향이 시계열 예측과 같은 특수한 태스크에서는 그대로 적용되지 않음을 보여준다. 다만 이는 실험에 사용된 ETTh1 데이터셋의 패턴이 비교적 단순하여 4B 모델의 용량으로도 충분히 분석 가능했기 때문일 수 도 있다. 즉, 이번 실험에서는 대규모 모델이 제공하는 고도의 패턴 인식 및 추론 능력이 요구되지 않았으며 단순한 모델 규모 확장이 항상 최적의 전략은 아님을 시사한다.

반면, 동일한 파라미터 규모를 가진 모델 간 비교에서는 뚜렷한 성능 차이가 관찰되었다. Qwen3-8B는 Llama3-8B 대비 현저히 우수한 성능을 보였으며, 이는 두 모델의 파라미터 수가 유사함에도 불구하고 BBH 점수로 측정되는 추론 능력과 계열적 특성의 차이가 실제 예측 정확도에 영향을 미친다는 점을 시사한다.

따라서 본 연구는 시계열 예측에서 LLM의 성능을 결정짓는 요인이 단순한 파라미터 규모가 아니라 모델 계열과 내재된 추론 능력임을 강조한다. 특히, ETTh1과 같은 단순한 데이터셋에서는 중간 규모 모델이 오히려 최적의 선택일 수 있음을 확인하였다. 그러나 파라미터 규모의 효과를 보다 정확하게 이해하기 위해서는 다양한 난이도와 복잡성을 가진 데이터셋을 활용한 추가적인 분석이 필요하다. 이러한 결과는 향후 LLM 기반 시계열 예측 시스템을 설계할 때, 모델 크기만을 기준으로 선택하기보다는 추론 능력, 계열적 특성, 데이터 특성을 종합적으로 고려한 전략적 접근이 필요함을 보여준다.

ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원-대학ICT연구센터(ITRC)의 지원을 받아 수행된 연구임(IITP-2025-RS-2020-II201808)

참 고 문 헌

[1] Box, George EP, et al. Time series analysis: forecasting and control. John Wiley & Sons, 2015.

[2] Graves, Alex. "Long short-term memory." Supervised sequence labelling with recurrent neural networks (2012): 37-45.

[3] Jin, Ming, et al. "Time-llm: Time series forecasting by reprogramming large language models." arXiv preprint arXiv:2310.01728 (2023).

[4] Ansari, Abdul Fatir, et al. "Chronos: Learning the language of time series." arXiv preprint arXiv:2403.07815 (2024).

[5] Liu, Chenxi, et al. "Timecma: Towards llm-empowered multivariate time series forecasting via cross-modality alignment." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 39. No. 18. 2025.

[6] Kaplan, Jared, et al. "Scaling laws for neural language models." arXiv preprint arXiv:2001.08361 (2020).

[7] Yang, An, et al. "Qwen3 technical report." arXiv preprint arXiv:2505.09388 (2025).

[8] Srivastava, Aarohi, et al. "Beyond the imitation game: Quantifying and extrapolating the capabilities of language models." Transactions on machine learning research (2023).

[9] Grattafiori, Aaron, et al. "The llama 3 herd of models." arXiv preprint arXiv:2407.21783 (2024).

Quantum-Hybrid Generative Flow Networks

Hosung Lee, Dasan Kim, Byeongoh Hwang, Myungjoo Kang
Seoul National Univ.

요 약

We propose a hybrid quantum simulation framework that integrates Generative Flow Networks (GFlowNets) with Trotterized quantum circuits for efficient and accurate simulation of the Ising model. By employing GFlowNets to sample intermediate quantum states between Trotter steps, our method significantly reduces circuit depth and mitigates cumulative noise, a major challenge in Noisy Intermediate-Scale Quantum (NISQ) devices. Experimental results on the FakeAlgiers simulator demonstrate that our approach achieves over four orders of magnitude improvement in mean squared error (MSE) compared to baseline simulations. This architecture enables long-time quantum dynamics with enhanced fidelity, offering a scalable strategy for realistic quantum simulations.

Ⅰ. 서 론

Quantum computation has emerged as a promising tool for simulating many-body quantum systems, particularly intractable models like the Ising spin chain. However, the limited fidelity and circuit depth constraints of NISQ-era devices severely degrade the quality of quantum simulations. In this context, techniques such as Trotter-Suzuki decomposition approximate time evolution but remain susceptible to noise as depth increases. To address these challenges, we propose a hybrid methodology wherein Generative Flow Networks are used to interpolate between discrete Trotterized time steps. GFlowNets construct intermediate quantum states by learning a reward-guided stochastic process, thus suppressing noise accumulation and improving accuracy.

Ⅱ. 본론

We target the simulation of the 1D transverse-field Ising model, governed by the Hamiltonian:

$$H = J\sum_{i=1}^{n-1}\sigma_i^z\sigma_{i+1}^z + h\sum_{i=1}^{n}\sigma^x$$

where σ^z and σ^x are Pauli matrices, J controls nearest-neighbor interactions, and h represents the transverse field strength.

To simulate time evolution, we apply Trotter-Suzuki decomposition, splitting the time-evolution operator:

$$e^{-iHt} \approx \left(e^{-iH_A\Delta t}e^{-iH_B\Delta t}\right)^{t/\Delta t}$$

with H_A and H_B representing the spin-spin and field terms respectively. This corresponds to alternating RZZ gates and RX gates on hardware.

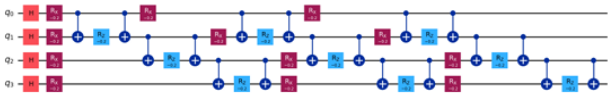


Fig. 1: Quantum Circuit Design for Ising Model Simulation. Example with 4 qubits.

To overcome depth-related noise, we implement a hybrid architecture using FakeAlgiers for coarse steps (Δt = 0.5), and GFlowNet to interpolate intermediate steps (Δt = 0.1). GFlowNet uses the reward:

$$R(s,t) = \alpha e^{-\beta E(s)} + (1-\alpha)\mathcal{L}_{\text{clo}}(s,t)$$

Here, E(s) is the classical Ising energy and L_colloc is a collocation loss to match QASM-simulated amplitudes.

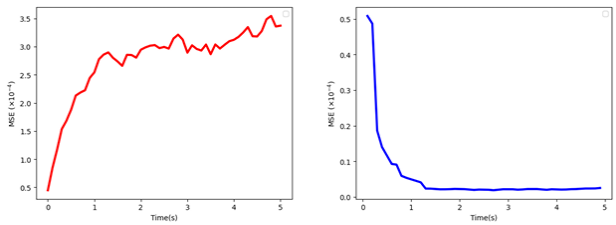


Fig. 2: Comparison of simulation fidelity. (Left) MSE between FakeAlgiers and QASM shows increasing error with time. (Right) GFlowNet-enhanced simulation maintains low MSE throughout evolution.

Experiment Summary:

- Setup: 8-qubit Ising model, J = h = 1.0, T = 5
- Baseline (FakeAlgiers-only): MSE ≈ 3.5×10^{-4}
- Hybrid (Ours): MSE ≈ 10^{-8}
- Optimal α = 0.5 balancing energy fidelity and data consistency.

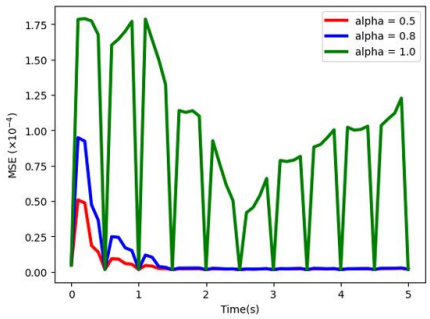


Fig. 3: MSE for different α values, showing optimal fidelity at α = 0.5.

Ⅲ. 결론

In this study, we presented a novel approach to enhancing the accuracy of Ising model simulations on quantum circuits by integrating Generative Flow Networks (GFlowNets). By using GFlowNet to sample intermediate states between Trotter steps, we effectively reduced the depth of quantum circuits and mitigated error accumulation over time. The results show that the GFlowNet-enhanced simulation maintained a lower Mean Squared Error (MSE) compared to simulations conducted solely on the FakeAlgiers simulator, demonstrating improved fidelity with respect to QASM outputs. This work highlights the potential of GFlowNets as a valuable tool for quantum simulations, especially in scenarios where noise and gate fidelity pose significant challenges due to increasing circuit depth. By optimizing intermediate states, GFlowNet provides a promising solution to manage error and maintain high simulation accuracy in long-time evolution studies. Future research could explore the application of this method to more complex quantum models or even experimental quantum devices. Additionally, refining the reward function or exploring adaptive time steps could further enhance the accuracy and efficiency of GFlowNet-integrated simulations.

ACKNOWLEDGMENT

IITP grant funded by the Korea government(MSIT) [NO.RS-2021-II211343, Artificial Intelligence Graduate School Program (Seoul National University)]

참 고 문 헌

[1] Liu, J., Qi, Y., Li, J., Gu, S.J.: Monte carlo studies of quantum systems with the sign problem. Journal of Statistical Mechanics: Theory and Experiment (2017)

[2] Cervera-Lierta, A.: Exact ising model simulation on a quantum computer. Quantum 2, 114 (2018)

[3] Kandala, A., Temme, K., C´orcoles, A.D., Mezzacapo, A., Chow, J.M., Gambetta, J.M.: Error mitigation extends the computational reach of a noisy quantum processor. Nature 567(7749), 491–495 (2019)

[4] Temme, K., Bravyi, S., Gambetta, J.M.: Error mitigation for short-depth quantum circuits. Physical Review Letters 119(18), 180509 (2017)

[5] Bengio, Y., Lahlou, S., Deleu, T., Hu, E.J., Tiwari, M., Bengio, E.: Gflownet foundations. Journal of Machine Learning Research 24, 1–76 (2023)

[6] Malkin, N., Shun, I., Zimmermann, H., Bengio, Y.: Gflownets and hierarchical variational inference for compositional distributions. arXiv preprint arXiv:2301.12594(2023)

[7] Shor, P.W.: Scheme for reducing decoherence in quantum computer memory. Physical Review A 52(4), 2493 (1995)

[8] Viola, L., Lloyd, S.: Dynamical suppression of decoherence in two-state quantum systems. Physical Review A 58(4), 2733 (1998)

온라인 콘텐츠의 부분 이미지 도용 탐지를 위한 특징점 기반의 효율적인 시스템

김다산[1]*, 황병오[1], 이호성[1], 강명주[2]

[1]서울대학교 공과대학 협동과정 인공지능전공, [2]서울대학교 자연과학대학 수리과학부
*dasan0409@snu.ac.kr

A Feature-based Efficient System for Detecting Partial Image Plagiarism of Online Contents

Dasan Kim[1]*, Byeoungoh Hwang [1], Hosung Lee[1] and Myungjoo Kang[2]
[1]Interdisciplinary Program in Artificial Intelligence, Seoul National University,
[2]Department of Mathematical Sciences, Seoul National University

요 약

본 논문은 현재 디지털 콘텐츠 시장에서 겪고 있는 불법 복제물을 탐지하기 위해 SIFT 와 FLANN 을 사용해서 웹툰이나 일러스트 등의 불법 복제물이 cropped 되어있거나 원본에 없던 watermark 가 붙어있어도 정확한 탐지가 가능하도록 모델을 만들었다. 웹툰과 일러스트 dataset 5000 장에 대해 실험을 한 결과, 제안한 모델이 부분적으로 잘려나가거나 watermark 가 붙은 이미지를 높은 정확도로 원본을 식별해내는 성능을 입증하였다.

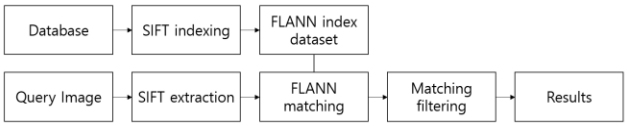
Ⅰ. 서 론

디지털 기술의 발전과 함께 웹툰, 전자책 등 디지털 콘텐츠 시장에서는 스크린 캡처나 원본 콘텐츠의 일부 영역만을 잘라내어(cropping) 해상도를 변경하거나 다른 사진 등을 추가하여 불법으로 유포하는 행위가 많아졌다. 이러한 부분 이미지 기반의 불법 복제물을 탐지하기 위한 기존의 접근법 중 하나인 딥러닝 기반의 이미지 비교 기술은 높은 정확도를 보일 수 있으나, 수만 장에 달하는 학습 데이터를 구축하는 과정이 필요하며, 고사양의 GPU 자원을 요구하고 추론 속도가 느려 실시간 대규모 탐지 시스템에 적용하기에는 한계가 존재한다.

본 연구에서는 이러한 한계를 극복하기 위해, 고전적이지만 강력한 컴퓨터 비전 기술인 ***SIFT***(Scale-Invariant Feature Transform)[1]와 고속 근사 근접 이웃 탐색 라이브러리인 ***FLANN***(Fast Library for Approximate Nearest Neighbors)[2]을 활용하여 효율적인 불법 복제물 탐지 시스템을 제안한다. 제안하는 시스템은 별도의 모델 학습 과정이 필요 없으며, 고사양의 하드웨어 없이도 오프라인&온라인 환경에서 빠르고 안정적으로 동작하는 것을 목표로 한다.

Ⅱ. 제안하는 시스템

본 연구에서 제안하는 부분 이미지 탐지 시스템은 크게 두 단계로 구성된다. 첫 번째는 원본 콘텐츠 데이터베이스의 모든 이미지에 대해 특징점 정보를 미리 추출하고 검색 가능한 구조로 만들어두는 오프라인 **'인덱싱 단계(Indexing Phase)'**이다. 두 번째는 의심되는 복제물 이미지가 입력되었을 때, 저장된 인덱스를 이용해 실시간으로 원본을 찾아내는 온라인 **'질의 및 매칭 단계(Query and Matching Phase)'**이다. 제안하는 시스템의 전체 구조는 [그림 1]과 같다.



[그림 1] SIFT 부분 이미지 탐지 시스템 pipeline

먼저, [그림 1]의 윗줄인 **인덱싱 단계**에서 모든 원본 이미지를 DB 에 넣고, SIFT 알고리즘을 사용해 이미지 내의 안정적인 특징점들과 각각에 해당하는 128 차원의 기술자 벡터를 추출한다. 그 다음 객체가 가진 핵심 속성(좌표, 크기, 각도 등)을 Tuple 형태로 변환한 후 FLANN 라이브러리를 이용해 k-d 트리 기반의 고속 검색 구조를 생성한다.

다음, [그림 1]의 아랫줄인 **질의 및 매칭 단계**에서 사용자가 탐지를 원하는 질의 이미지(예: 온라인 커뮤니티에 올라온 잘려진 이미지)를 입력으로 받는다. 질의 이미지에 대해서도 동일하게 SIFT 를 적용하여 키포인트와 기술자를 추출한다. 그 후, 사전에 생성된 FLANN 인덱스를 이용해, 질의 이미지의 각 기술자와 데이터베이스 전체의 기술자들 간의 k-Nearest Neighbors 탐색을 수행한다. 이때, 후속 필터링을 위해 k=2 로 설정하여 각 질의 기술자마다 가장 유사한 후보 2 개를 찾는다. 이제 찾은 매칭 후보 쌍들에 대해 Lowe 의 비율 테스트[3]를 적용한다. 즉, 가장 유사한 후보(1 순위)와의 거리와 두 번째로 유사한 후보(2 순위)와의 거리 비율을 계산하여, 이 비율이 설정된 임계값보다 낮을 경우에만 신뢰할 수 있는 '좋은 매칭(Good Match)'으로 판단하고 나머지는 폐기한다. 마지막으로 데이터베이스 내의 각 원본 이미지별로 '좋은 매칭'의 개수를 계산한다. 가장 많은 '좋은 매칭'을 기록한 원본 이미지를 질의 이미지의 출처로 최종 판정한다. 단, 신뢰도를 위해 최소 매칭점 개수 임계값을 두어, 이 기준을 넘지 못하면 '출처를 찾을 수 없음'으로 판단하여 오탐지를 방지한다.

Ⅲ. 실험 결과

데이터셋은 디지털 콘텐츠 시장에 초점을 맞춰 웹툰, 일러스트 합계 5000 장으로 만들었다. 한 번의 실험 당 총 100 번의 질의 이미지를 random 으로 만들어 물어보았고, 점수 체계는 억울한 피해자를 만드는 것보다는 원본을 못 찾는 것이 더 낫다고 생각해 100 점 만점으로 올바른 원본을 찾았을 경우 +1 점, 출처를 아예 못 찾았을 경우 0 점, 틀린 원본을 찾았을 경우 -0.5 점을 매겨 합산한 값을 최종 점수로 환산했다.

평균점수	0.7	0.75	0.8
10	97.3	98.7	98.1
15	98.5	**99.1**	98.4
20	96.6	97.3	95.2

[표 1] 10 번 테스트 평균점수. 가로축은 Lowe 의 비율 테스트의 임계값, 세로축은 최소 매칭점 개수 임계값.

[표 1]에서 보다시피 Lowe 의 비율 테스트의 임계값이 0.75, 최소 매칭점 개수 임계값이 15 개 일 때, 가장 좋은 performence 를 보였다.

Ⅳ. 결론

본 연구는 온라인 불법 복제 문제, 특히 원본의 일부를 잘라내어 유포하는 방식의 탐지 어려움에 주목하였다. 기존의 높은 자원을 요구하는 딥러닝 기반 기술의 한계를 극복하고자, 본 논문에서는 고전적이지만 성능이 검증된 컴퓨터 비전 기법인 SIFT 와 FLANN 을 활용하여 효율적이고 실용적인 불법 복제물 탐지 시스템을 제안하였다. 제안된 시스템은 별도의 학습 과정 없이 오프라인 인덱싱과 실시간 질의 단계로 구성되어, 하드웨어 제약이 있는 환경에서도 안정적으로 동작할 수 있도록 설계되었다.

자체 구축한 대규모 온라인 콘텐츠 이미지 데이터셋을 대상으로 한 실험을 통해, 제안한 시스템이 부분적으로 잘려나가거나 원본에 없던 watermark 가 붙은 이미지를 높은 정확도로 원본을 식별해내는 성능을 입증하였다. 특히, 매칭 파라미터에 대한 민감도 실험을 통해, 해결하고자 하는 문제 영역에서 최적의 성능을 발휘하는 시스템 구성 요소를 체계적으로 도출하였다. 또한, 오답의 위험성을 반영한 '신뢰도 인지 가중 점수' 체계를 통해 단순 정확도를 넘어 시스템의 실용적 가치를 정량적으로 평가하였다. 최종적으로, 인덱스 파일 캐싱과 FLANN 기반 고속 매칭을 통해 제안 시스템이 실제 환경에 적용될 수 있는 충분한 잠재력을 가지고 있음을 확인하였다.

ACKNOWLEDGMENT

IITP grant funded by the Korea government(MSIT) [NO.RS-2021-II211343, Artificial Intelligence Graduate School Program (Seoul National University)]

참 고 문 헌

[1] Lowe, D. G. "Distinctive Image Features from Scale-Invariant Keypoints," International Journal of Computer Vision, vol. 60, no. 2, pp. 91-110, 2004.

[2] Muja, M., and Lowe, D. G. "Scalable Nearest Neighbor Algorithms for High Dimensional Data," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, no. 11, pp. 2227-2240, 2014.

[3] Babenko, A., Slesarev, A., Chigorin, A., and Lempitsky, V. "Neural-Codes for Image Retrieval," Proceedings of the 13th European Conference on Computer Vision, pp. 584-599, 2014.

PPO 를 활용한 APF 알고리즘 인력/척력 계수의 환경 적응형 자동 최적화 모델

이종훈, 이세비, 유현석*
*㈜빅앤딥

jhlee@bigndeep.co.kr, leesebi@bigndeep.co.kr, *hsyoo@bigndeep.co.kr

Environmentally Adaptive Automatic Optimization Model of the Attractive/Repulsive Force Coefficient of the APF algorithm using PPO

Lee Jong Hun, Lee Se Bi, Yoo Hyun Suk*
*BIG & DEEP Co., Ltd.

요 약

최근 방위 산업 업계에서는 무인 이동 객체를 활용한 군사 작전 수립 전략에 딥러닝(Deep Learning)을 결합한 방산 인공지능이 주목받고 있다. 방산 인공지능 중에서도 무인 로봇의 경로 계획(Path Planning) 알고리즘은 아군 객체의 적진 우회, 군수물자 운반, 적군 추적 등 다양한 전시 상황에 응용될 수 있다. 본 연구에서는 지역 경로 계획 알고리즘 중 하나인 인공 퍼텐셜 필드(Artificial Potential Field, APF)의 고도화를 위해 강화학습 방법론을 활용한 APF 인력/척력 계수 결정 모델을 제시한다. 제안 방안은 국지 최소값 진입 현상을 인지하고 이탈하는 알고리즘을 포함하고 있으며, 시뮬레이터 기반 강화학습을 통해 제안된 모델이 APF 평가 지표에서 종합적으로 안정적인 성능을 보임을 확인하였다.

Ⅰ. 서론

로봇의 경로 계획은 전역 경로 계획(Global Path Planning)과 지역 경로 계획(Local Path Planning)으로 분류된다. 전역 경로 계획은 전체 지도를 기반으로 장거리 단위의 중간 지점을 생성함으로써, 최단 경로를 따르는 동시에 목표 지점까지의 안정적인 이동을 보장한다. 지역 경로 계획은 전역 경로 계획에서 생성한 중간 지점 사이의 이동 경로를 결정한다. 지역 경로 계획은 장애물 회피 능력과 실상황에서의 실시간 대응 계획 등 여러 경로 계획 요소의 통합적 고려가 요구된다.

지역 경로 계획 알고리즘인 APF(Artificial Potential Field)는 가상의 인력과 척력 포텐셜 필드를 사용하여 경로를 생성하는 알고리즘이다. APF 는 타 알고리즘 대비 낮은 계산 복잡도로 부드러운 경로 생성을 가능케 하나, 국지 최소값 (Local Minima)진입 현상을 극복하지 못하여 경로 생성에 실패할 가능성이 있다[1]. 이러한 국지 최소값 진입 현상은 알고리즘 수정[2] 또는 강화학습 적용 방법론[3]을 활용하여 개선될 수 있다. 본 연구에서는 강화학습 알고리즘 중 PPO(Proximal Policy Optimization)[4]를 접목한 환경 적응형 최적화 모델의 학습 방안과 이를 학습시키기 위한 시뮬레이터 구조를 제안한다.

Ⅱ. 환경 적응형 APF 인력/척력 계수 최적화 모델

경로 계획 에이전트는 이동 객체가 출발지에서 목적지로 향하게 하는 인력 계수 K_{att} 와 장애물로부터 이동 객체를 밀어내게 하는 척력 계수 K_{rep}를 결정하여 2 차원 평면에서의 APF 이동 경로를 생성한다. 에이전트는 높은 경로 생성 성공률, 최대 장애물 근접 거리, 최단 경로, 최소 Local Minima 진입 횟수를 도출하는 APF 인력/척력 계수 선택을 목적으로 학습한다.

2.1 인력 계수 기반 척력 계수 정의

본 연구에서는 인력 계수와 척력 계수가 특정 범위 내에서 반비례 관계를 갖는다는 특성에 주목하여 두 계수를 선형식에 근사하여 정의하였다. 인력 계수를 $K_{att} \in [K_{att_{min}}, K_{att_{max}}]$, 척력 계수를 $K_{rep} \in [K_{rep_{min}}, K_{rep_{max}}]$ 로 나타낼 때, K_{rep} 은 다음과 같이 정의할 수 있다.

$$K_{rep} = K_{rep_{max}} - \frac{K_{rep_{max}} - K_{rep_{min}}}{K_{att_{max}} - K_{att_{min}}} \times \left(K_{att} - K_{att_{min}}\right)$$

2.2 강화학습 기반 인력/척력 계수 자동 최적화 모델

지역 경로 계획을 수행하는 에이전트가 특정 시점에서 APF 경로 계획을 생성하기 위해서는 과거 상태가 아닌 현재 상태, 즉 현재 시뮬레이터의 결과와 장애물의 위치만을 필요로 한다. 따라서 해당 과정은 마르코프 특성을 만족하므로 강화학습 문제로 모델링할 수 있다.

(x^*, y^*) 좌표축으로 이루어진 평면에서 장애물 $O_{i\in[1,N]}$ 가 $\left(O_{i_{x^*}}, O_{i_{y^*}}\right)$ 에 위치해 있을 때, 에이전트가 관측하는 상태(State) S는 다음과 같다.

$$S = [O_{1_{x^*}}, O_{1_{y^*}}, \cdots O_{N_{x^*}}, O_{N_{y^*}}, K_{att}]$$

APF 에서는 장애물의 척력 세기에 따라 에이전트의 이동 경로가 달라지므로, 여러 장애물 간의 밀도와 분포는 에이전트의 학습에 있어 유의미한 입력으로 작용할 수 있다. 본 논문에서는 다수의 장애물과 에이전트의 상대적인 좌표를 일관된 상태로 표현하기 위해 기존 평면좌표계를 대체한 (x^*, y^*) 좌표계를 활용하였다. x^* 축은 기존 좌표계에서의 시작점과 도착점을 지나는 직선이며, 기존 시작점과 장애물들 간

거리를 비교하였을 때 가장 작은 Δx값을 갖는 장애물이 포함된 x^*축의 수직선을 y^*축으로 정의한다.

에이전트는 인식된 상태에 대해 실수 공간 $[K_{att_{min}}, K_{att_{max}}]$ 내에서 인력 계수 K_{att}를 추론하고, K_{att}와 K_{rep}에 따라 생성된 APF 에서 이동한 결과를 보상으로 얻는다. 경로 생성 여부 $P_{success}$ 에 따른 보상함수는 다음과 같다.

$$R = \begin{cases} r_{success} - \alpha \times \dfrac{D_{path} - D'}{D'} - \beta \times \dfrac{1}{{D_{obs}}^2} - \gamma \times N_{LM}, & if\ P_{success} \\ r_{fail}, & otherwise \end{cases}$$

경로 생성이 가능한 인력계수가 선택되었다면 에이전트는 기본 보상 $r_{success}$를 받게 되며, 해당 경로의 길이 D_{path}가 시작점과 도착점을 잇는 최소 거리 D'에 대해 차이가 적을수록 높은 보상을 얻는다. 에이전트는 가장 가까운 장애물과의 근접 거리인 D_{obs}가 작을수록 높은 페널티를 받으며, 국지 최소값 진입 횟수 N_{LM}이 적을수록 높은 보상을 얻을 수 있다.

2.3 국지 최소값(LM) 진입 감지 및 이탈 알고리즘

본 연구에서는 지역 경로 계획 에이전트의 국지 최소값 진입 현상을 완화하기 위해 LM 진입 감지 및 이탈 알고리즘[1]을 추가한 APF 시뮬레이터를 사용하였다. 에이전트가 최근 m 포인트 이내에 방문한 기록이 있는 위치에 다시 방문한다면, APF 시뮬레이터는 에이전트가 LM 상태에 진입한 것으로 판단한다. 이 때 시뮬레이터는 LM 발생 지점 주변의 포텐셜 값을 기존 최대값보다 높은 임의의 값으로 수정하여 에이전트로 하여금 해당 위치를 벗어나게 한다.

Ⅲ. 모의실험 실행 및 결과

강화학습 에이전트는 무작위로 분포된 N 개의 장애물 환경에서 학습한다. 본 모의실험에서는 $N \in [2,5]$ 개의 장애물이 무작위로 혼합되어 있는 2×10^4 개의 환경에서 학습을 진행하였다.

시뮬레이션에 사용된 인력 및 척력 계수는 이동 경로 길이가 약 $1\,km$, 장애물의 척력 적용 반경이 $100\,m$인 경우를 가정하여 각각 $K_{att} \in [0.01,\ 0.05]$, $K_{rep} \in [10^4,\ 5 \times 10^5]$ 로 설정하였다.

표 1 은 $N \in [2,\ 5]$ 혼합 환경에서의 비교군 대비 제안 방안의 성능 비교를 나타낸다. 성능 평가는 10^4 번의 테스트 환경에서 진행하였으며, 평균 경로 길이 D^*_{path}, 평균 장애물 근접거리 D^*_{obs}, 평균 LM 진입 횟수 N^*_{LM}, 평균 보상 R^*, 성공률 q 를 기준으로 학습 성과를 판단한다. $b_{i\in[1,B]}$ 는 인력 계수 범위를 B등분하여 만든 인력 및 척력 계수 조합 B개 중 i번째 조합을 나타내며, 본 실험에서는 $B = 5$로 설정하여 5개의 비교군에 대한

	b_1	b_2	b_3	b_4	b_5	제안방안
$D^*_{path}(m)$	1173	1125	1093	1057	1009	**1086**
$D^*_{obs}(m)$	97.40	85.24	75.21	63.65	41.02	**82.09**
N^*_{LM}	2.937	1.598	0.971	0.401	0.048	**0.358**
R^*	4.354	6.400	7.075	7.303	3.829	**8.004**
$q(\%)$	79.36	90.62	95.10	98.52	100.0	**99.55**

표 1. 혼합 환경에서의 비교군 대비 제안 방안 성능

제안방안의 성능을 평가하였다. 제안 방안으로 학습된 에이전트는 높은 경로 생성 성공률, 먼 장애물 근접 거리, 짧은 경로 길이, 최소 LM 진입 횟수를 동시에 갖는 인력 및 척력 계수를 선택함을 알 수 있다.

Ⅳ. 결론

본 연구에서는 5 개 이하의 장애물 배치 환경에서 APF 알고리즘의 인력 및 척력 계수를 적응적으로 결정하는 환경 적응형 APF 강화학습 모델을 제시하였다. 시뮬레이션 기반 강화학습 접근법을 통해 APF 시뮬레이터를 개발하였으며, 연속 공간에서의 PPO 알고리즘을 적용하였다. 행동 공간을 1 차원으로 축소하여 학습 난이도를 완화하고, 장애물 위치 정보를 전처리하여 분포 패턴을 정규화함으로써 유사한 구조적 특성을 갖는 환경을 동일하게 처리할 수 있도록 하였다. 결과적으로 다양한 장애물 분포 환경에서 장애물 배치에 강건하게 대응하며 효과적인 경로 계획을 생성할 수 있는 APF 에이전트를 강화학습을 통해 구현하였다.

본 연구의 결과를 바탕으로, 적군과 같은 위험요소가 존재하는 전장 환경, 고밀도 도시 환경, 안티드론 분야의 레이더 탐지 영역 분포 환경 등의 다양한 3 차원 실제 환경 하에서 실용성 있는 기능을 제공하는 강건한 강화학습 모델 연구를 할 계획이다. 또한 객관적인 성능 분석을 위하여 복잡한 실제 환경에서 D-APF(Dynamic Artificial Potential Field) 등 최신 알고리즘과 강화학습 에이전트 간의 상세한 비교 검증을 할 계획이다.

ACKNOWLEDGMENT

본 결과물은 방위사업청의 재원으로 국방기술진흥연구소의 방산혁신기업 100 프로젝트 기술개발 전용지원사업의 지원을 받아 연구되었음 (과제관리번호: R230106).

참 고 문 헌

[1] Park, Min Gyu, and Min Cheol Lee. "A new technique to escape local minimum in artificial potential field based path planning." KSME international journal 17 (2003): 1876-1885.

[2] Jayaweera, Herath Mpc, and Samer Hanoun. "A dynamic artificial potential field (D-APF) UAV path planning technique for following ground moving targets." IEEE access 8 (2020): 192760-192776.

[3] Shao, Mingzhi, et al. "Research on UAV Trajectory Planning Algorithm Based on Adaptive Potential Field." Drones 9.2 (2025): 79.

[4] Sakai, Atsushi, et al. "Pythonrobotics: a python code collection of robotics algorithms." arXiv preprint arXiv:1808.10703 (2018).

[5] Schulman, John, et al. "Proximal policy optimization algorithms." arXiv preprint arXiv:1707.06347 (2017).

Plug-and-Play Bi-CoT: 설명가능한 Multi-Hop QA 를 위한 자기인식 정방향 추론과 역방향 검증

강지헌 [1,*,©], 정수환 [2,*]
연세대학교 [1], 부경대학교 [2]

danieljiheon@yonsei.ac.kr[1], suhwan0823@gmail.com[2]

Plug-and-Play Bi-CoT: Self-Aware Forward Reasoning and Reverse Verification for Explainable Multi-Hop QA

Kang Ji Heon[1,*,©], Jeong Su Hwan[2,*]
Yonsei Univ.[1], Pukyong National Univ.[2]

요 약

본 논문은 Multi-Hop QA (MHQA)에서 추가적인 학습이 필요 없는 플러그인(Plugin) 모듈을 제안한다. 하위 질문에서 Self-Aware QA 를 적용해 단계별 Reasoning 품질을 높이고, 최종 단계에서 Reverse Verification(역방향 검증)을 통해 논리 경로와 증거 집합의 구조적 타당성을 확인하는 구조이다. 이는 기존 연구의 방향성을 강화하면서도 다른 도메인에서의 적용 가능성을 지닌다. HotpotQA 데이터셋에서 유의미한 성능 향상을 관찰할 수 있었고, 단순 출력만을 반환하는 기존 접근법과 다르게 각 단계별 질문·답변·근거를 명시적으로 제공함으로써 설명가능성을 크게 강화하였다.

Ⅰ. 서 론

Multi-Hop QA(MHQA)는 단일 문서나 사실로는 답할 수 없는 복합 질의에 대응하기 위해, 대규모 데이터 속에서 서로 다른 출처의 증거를 선택적으로 검색하고 연결하는 다단계 추론 과제이다. 최근 MHQA 연구들은 복합 질의를 해결하기 위해 질문을 하위 질문(Sub-question)으로 분해하여 Reasoning 품질을 높이는 전략을 취해왔다.[1]

그러나 이 방안은 두 가지 중요한 한계를 가진다. 첫째, 추론 경로의 불투명성이다. 대부분의 모델은 최종 답과 일부 근거 문서 전체를 반환하기 때문에, 사용자가 답변이 도출된 구체적 논리 구조를 확인하거나 오류 원인을 추적하기 어렵다. 둘째, Forward-only 구조에 따른 오류 전이다. 하위 질문에서 잘못된 답이 나오면 이후 단계에서 이를 바로잡을 기회가 거의 없어 최종 답변이 크게 왜곡될 수 있다.

본 연구는 이러한 한계를 해결하기 위해, 설명 가능성과 오류 완화를 핵심 목표로 하는, 어느 Retriever 과 QA 모델에도 결합해 사용할 수 있는 플러그인(Plugin) 모듈을 제안한다. 구체적으로 각 하위 질문의 답변과 근거를 명시적으로 기록하는 **Self-Aware QA** 단계를 통해 Reasoning 과정을 그대로 노출하고(설명 가능성 강화), Forward 추론 후 **Reverse Verification** 단계를 추가하여 논리 경로와 증거 집합을 다시 검증함으로써 오류 전이를 완화한다.

Ⅱ. 본론

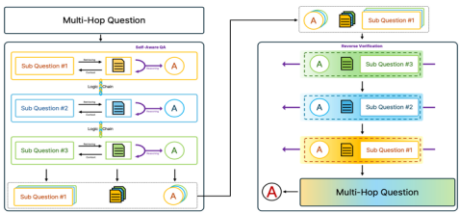


그림 1. 제안한 모듈의 파이프라인

본 모듈의 파이프라인은 하위 질문 분해로 시작해, 선택한 모듈의 Retrieval 과 1 차 QA, 그 결과를 바탕으로 한 Self-aware QA 로 순방향 Reasoning 강화, 그리고 Reverse Verification 을 통해 일관되고 신뢰성 있는 답변을 산출하는 네 단계로 구성된다.

(가) Sub-question 분해

원 복합 질문을 하위 질문(Sub-question)으로 분해하여 논리 과정을 구조화하는 단계이다. 이는 복합 질문을 단계별로 나누어 검색과 추론을 안정화한다.

하위 질문을 생성할 때 의존성(dependency)을 명확히 반영하고, 선행 답변이 필요한 경우에는 해당하는 질문에 치환을 적용한다. 예를 들어, “[Answer of SubQ1] is what kind of food?”로 질문을 생성하여 하위 질문간의 논리 구조를 종속성에 맞게 형성한다. 그리고 검색을 진행할 경우에는 실제 하위 질문 1 번의 답으로 치환하여 질문을 완성한다.

(나-0) Retrieval & QA *(Optional Choice)*

해당 부분은 기존의 MHQA 를 진행할 수 있는 어떤 Retriever 과 QA 모델도 사용 가능하다. 즉, 기존 MHQA 가 진행될 수 있도록 하는 기본 구조이며, 우리의 플러그인 모듈에 해당하지 않는다.

어떤 질문이 들어오면, 방대한 양의 데이터 속에서 관련된 정보를 가져오는 것을 Retriever 모델이 진행한다. 그리고 관련된 정보 속에서 질문의 대답을 찾는 것을 QA 모델이 진행하게 된다.

이 과정을 통해 추후 Reasoning 강화를 위해 필요한 기초 검색 데이터와 1 차 답변을 준비한다.

(나) Self-Aware QA (Forward CoT)

각각의 하위 질문에 대해 Retrieval 결과와 QA 답변을 기반으로, 모델이 스스로 답의 타당성을 점검하고 Reasoning 을 강화하고 경로를 명시적으로 기록할 수 있다.

* These authors contributed equally to this work.
© Corresponding author.

질문 속에서 필요한 조건을 추출하고, 그 조건을 가지고 온 데이터에 충족하는지 평가한다. 그 과정에서 구한 답변이 맞는지, 아니라면 새로운 답을 도출한다. 답을 찾는 과정에서 사용한 핵심적인 하나의 데이터, Evidence 를 얻게 된다.

(다) Reverse Verification (Reverse CoT)

Forward-only 구조는 초기 단계의 오류가 최종 답으로 전이된다. 이를 보완하기 위해 원질문에서 역방향으로 논리구조를 재구성하는 Reverse Verification 을 진행한다.

먼저 원질문과 하위 질문을 검토해 답 유형·비교 유형·구조적 문제를 판별한다. 여기서 불필요한 단계를 제거하고, 필수 정보만 남겨 역순으로 단순화한다. 역방향으로 검증하는 과정을 통해 하위 질문 분해에서 놓친 거시적 관점을 회복하고, 질문의 범위와 답 유형의 일치를 점검할 수 있다.

이후 검색 Evidence 를 다시 확인하면서 필요한 경우 단계를 병합하거나 롤백해 오류를 교정한다. 최종적으로 재구성된 논리 구조를 바탕으로 최종 답과 핵심 Evidence 를 함께 제시하여 Reasoning 의 일관성과 타당성을 높인다.

실험 결과

본 연구는 다량의 위키백과 문서가 들어있는 HotpotQA Full Wiki[2] 데이터셋에서 검증한다. 기본적인 MHQA 를 위한 Retriever 과 QA 모델은 각각 MDR[3]과 allenai/unifiedqa-t5-large[4]를 사용한다. MDR 의 성능 향상을 위해 각 하위 질문의 의미를 유지한 단일 Paraphrase 를 추가 생성하여 Retrieval 시 다양한 질의 표현을 확보하였다.

이를 Baseline 으로 잡고 우리의 플러그인 모듈을 추가하여 성능 향상을 비교한다. 플러그인 모듈은 GPT-4.1 과 GPT-4o 를 이용해 구성한다.

평가 지표는 정답과 완전히 일치한 비율인 EM(Exact Match)과 정답과 토큰 겹침 비율인 F1 을 사용한다.

1 번 실험

	EM (%)	F1 (%)
Our method	**42.77**(+18.61%p)	**55.40**(+22.25%p)
only Self-Aware QA	32.67(+8.51%p)	42.62(+9.47%p)
Baseline	24.16	33.15

표 1. Baseline 대비 성능 향상 지표

표 1 에서 보이듯, Baseline 대비 Self-Aware QA 만 적용해도 EM/F1 이 향상되었다. 그리고 Reverse Verification 까지 결합한 최종 방법은 성능이 모두 추가로 크게 향상되었다. 각 단계별 질문·답변·근거를 명시적으로 제공하여 논리 구성을 수정할 수 있게 되어, 오류 진단·교정이 된 덕분에 생긴 성능 향상으로 판단된다.

Forward CoT 만 적용한 경우, 과도한 하위 질문 분해나 불완전한 답변으로 인해 원 질문의 의도에서 벗어나는 경우가 있는데, 이를 Reverse CoT 가 잡아주기도 하였다. 이는 Reverse CoT 가 Forward CoT 의 초기 오류 전이 문제뿐 아니라, 불필요한 단계나 불완전 답변으로 인한 결론 누락까지 효과적으로 완화함을 보여준다.

한편, 전체 질의 중 약 65.3%만이 Reverse CoT 단계로 진입했으며, 나머지는 대부분 Retrieval 단계에서 필수 Evidence 를 확보하지 못해 역검증이 불가능했다. 이는 초기 데이터 검색의 중요성을 시사한다. 더 나은 성능의 Retriever 모델을 선택한다면, 성능 향상 효과 역시 커질 것으로 전망된다. 이에 따라 Reverse Verification 에서의 논리 구조 생성이 얼마나 정확한지를 비교하기 위한 실험을 진행하였다.

2 번 실험

	EM (%)	F1 (%)
DrKIT[5]	42.13	51.72
Transformer-XH-final[6]	51.60	64.07
Our method	**56.52**	**72.88**
AISO(SOTA)[7]	67.46	80.52

표 2. Reverse Verification(Reverse CoT)의 정확도와 기존 모델의 성능비교

기존 연구인 DrKIT, Transformer-XH 와 비교 시 본 방법은 경쟁력 있는 성능을 보였다. 그럼에도 SOTA 와 차이가 있는데, 이는 본 연구의 학습 없는 Inference-only 방식이 가지는 한계로도 해석 가능하다. 데이터셋에 최적화가 되지 않았기 때문이다. 하지만 Inference-only 방식이기 때문에 다른 모델과 다르게 플러그인 모듈로써 사용 가능하기도 하다. 이는 성능 좋은 모델에 결합할수록, 자세하고 좋은 논리 구조를 얻을 가능성을 시사한다.

Ⅲ. 결론

본 연구는 Multi-Hop QA 에서 제기된 두 가지 핵심 한계, 즉 추론 경로의 불투명성과 Forward-only 구조의 오류 전이를 해결하기 위해, Self-Aware QA 와 Reverse Verification 을 결합한 추가 학습 없는(Training-free) 양방향 CoT 파이프라인을 제안하였다.

HotpotQA 실험을 통해 단순 Retrieval+ QA 대비 유의미한 성능 향상을 입증했으며, 일부 Fine-tuning 된 MHQA 모델에 근접하거나 우세한 결과를 달성함으로써 학습 없이도 경쟁력 있는 수준의 성능을 보여주었다.

제안한 모듈은 각 단계별 질문·답변·근거를 명시적으로 기록하여 추론 경로를 그대로 드러내고, 역방향 검증으로 오류를 교정함으로써 설명 가능성과 신뢰성을 동시에 강화한다.

또한 플러그인 형태로 설계되어 기존 MHQA 모델은 물론 더 강력한 Retriever·QA 모델과 결합 시 더욱 정교하고 견고한 논리 구조를 제공할 수 있으며, 도메인 전이와 복합 추론이 필요한 다양한 응용 과제에도 확장될 수 있는 실용적 가치를 지닌다.

참 고 문 헌

[1] Trivedi, H. et al. *Interleaving Retrieval with Chain-of-Thought Reasoning*. ACL, 2023.

[2] Yang, Z. et al. *HotpotQA: A Dataset for Diverse, Explainable Multi-hop QA*. EMNLP, 2018.

[3] Xiong, W. et al. *Answering Complex Open-Domain Questions with Multi-Hop Dense Retrieval*. ICLR, 2021.

[4] Khashabi, D. et al. *UnifiedQA: Crossing Format Boundaries With a Single QA System*. arXiv:2005.00700, 2020.

[5] Dhingra, B. et al. *Differentiable Reasoning over a Virtual Knowledge Base*. ICLR, 2020.

[6] Zhao, C. et al. *Transformer-XH: Multi-Evidence Reasoning with eXtra Hop Attention*. ICLR, 2020.

[7] Zhu, Y. et al. *Adaptive Information Seeking for Open-Domain QA*. EMNLP, 2021.

제어 장벽 함수를 이용한 샘플링 기반 안전 필터 연구

박준영, 성현태, 안희진*
한국과학기술원

{junyoung766, hyeontae.sung, heejin.ahn}@kaist.ac.kr

Sampling-based Safety Filtering via Control Barrier Functions

Junyoung Park, Hyeontae Sung, Heejin Ahn*
Korea Advanced Institute of Science and Technology

요 약

본 논문에서는 제어 시스템의 안전을 보장하는 샘플링 기반 안전 필터 알고리즘을 제안하였다. 동역학을 알고 있는 이산 시간 시스템에서 현재 제어 입력을 기반으로 미래의 제어 입력 시퀀스를 샘플링하고, 각 샘플에 대해 롤아웃을 수행하여 안전성을 판단한다. 모든 롤아웃이 안전하지 않다고 판단되면, 안전 필터가 개입하여 주어진 제어 입력과의 차이가 작으면서도 안전한 제어 입력으로 대체한다. 안전성 판단에는 Control Barrier Function(CBF)을 활용한 제약 조건이 사용된다. 시뮬레이션을 통해 제시된 안전 필터가 로봇에 자율성을 부여하면서도 안전을 보장하는 것을 검증하였다.

Ⅰ. 서 론

자율주행과 같이 안전이 중요한 제어 시스템에서는 안전을 보장할 수 있는 제어 입력을 생성해야만 한다. 최근 연구되는 데이터 기반 제어기는 다양한 환경에서 높은 성능을 보이지만 안전을 보장하지 못하는 한계가 있다.[1] 따라서 이런 제어기의 안전을 보장하기 위한 장치가 필요하다. 기존의 성능 향상을 위한 제어기를 nominal controller 라고 할 때, 안전 필터는 nominal controller 의 제어 입력을 따르다가 필요한 순간에만 개입하여 안전을 보장하는 역할을 한다. 이러한 방법론은 안전한 상태를 유지하면서도 nominal controller 의 자율성을 최대한 보장한다.

본 연구에서 제안하는 샘플링 기반 안전 필터에서는 미래 제어 입력 시퀀스를 샘플링하여 현재 주어진 제어 입력의 안전성을 판단한다. 이와 같은 샘플링 기반 접근법은 기존의 최적화 기반 방식[2]과 비교했을 때 최적해를 구하기 어려운 복잡한 동역학 모델이나 비선형 시스템에서도 사용 가능성이 높다는 강점이 있다. 안전 필터는 필요한 경우 이전 샘플링 단계에서 저장한 안전 샘플들을 새로운 제어 입력에 활용한다.

논문에서는 샘플링 기반 안전 필터 알고리즘을 제시하고, 제시된 알고리즘이 제어 입력 제약이 있는 상황에서 로봇에 자율성을 부여하면서도 주어진 안전 조건을 만족하는 것을 시뮬레이션으로 확인하였다.

Ⅱ. 본론

본론의 2.1 절에서는 제어 입력의 안전성을 판단할 때 사용되는 CBF 의 개념에 대해 설명하였다. 2.2 절에는 본 논문에서 제안하는 샘플링 기반 알고리즘을 정리하고, 2.3 절에 실험을 통한 검증 결과를 나타냈다.

2.1 Discrete-Time Control Barrier Function[2]

다음과 같은 이산 시간 시스템을 생각하자.

$$x_{t+1} = f(x_t, u_t) \tag{1}$$

여기서 $x_t \in \mathcal{X}$ 는 시스템의 상태, $u_t \in \mathcal{U}$ 는 제어 입력을 의미한다. 안전 보장 제어를 위해 다음과 같은 안전 집합 S를 정의할 수 있다.

$$S \coloneqq \{x_t \in \mathcal{X} : B(x_t) \geq 0\} \tag{2}$$

이때 함수 $B: \mathcal{X} \to \mathbb{R}$가 모든 시스템 상태 $x_t \in \mathcal{X}$에 대해 다음 조건을 만족하면 discrete-time CBF(DCBF)라고 한다.

$$\exists u_t \in \mathcal{U}\ s.t.\ B\big(f(x_t, u_t)\big) \geq (1-\gamma)B(x_t),\ 0 < \gamma \leq 1 \tag{3}$$

DCBF B에 대해 부등식 (3)을 만족하는 제어 입력 u_t를 사용하면 안전 집합 S 내부의 상태를 유지할 수 있다.

2.2 샘플링 기반 안전 필터

본 논문에서 제안하는 샘플링 기반 안전 필터는 Algorithm 1 에 정리하였다. 알고리즘은 매 타임스텝마다 nominal controller 의 제어 입력 u_{des}를 사용하여 다음 스텝의 상태 x_{t+1}을 예측한다(line 1). 이후 T horizon 을 가지는 제어 입력 시퀀스 $U = [u_0, \dots, u_{T-1}]$를 샘플링하고, x_{t+1} 부터 롤아웃을 진행하여 N 개의 trajectory X 를 얻는다(line 3-5). CBF 제약 조건을 통해 trajectory 들의 안전성을 평가하는 $cost_{CBF}$를 계산한다(line 6).

$$cost_{CBF}(X) = \sum_{k=0}^{T-1} \max\{0, -B(X_{k+1}) + (1-\gamma)B(X_k)\} \tag{4}$$

여기서 X_k 는 trajectory X 의 k 번째 상태를 의미한다. 이때 안전한 제어 입력 시퀀스는 다음 타임스텝에 사용할 수 있도록 S_{t+1}^{safe}에 저장한다(line 7-8).

안전한 샘플이 하나라도 존재하면 안전 필터의 개입 없이 u_{des}를 사용한다(line 9-10). 안전한 샘플이 하나도 없는 경우에는 이전 타임스텝에서 저장한 안전한 제어 입력 시퀀스 집합 S_t^{safe} 에서 u_{des} 와 가장 가까운 샘플 U^{safe}를 찾는다(line 12). 이때 $cost_{deviation}$을 활용한다.

$$cost_{deviation}(U) = \|U_0 - u_{des}\|^2 \tag{5}$$

여기서 U_0는 시퀀스의 첫 번째 제어 입력을 의미한다.

이후 U^{safe} 의 첫 번째 제어 입력은 u_{safe}로 내보내고, 뒷부분은 다음 타임스텝에서 사용할 수 있도록 S_{t+1}^{safe} 로 넘겨준다(line 13).

Algorithm 1 Sampling-based Safety Filter

Given: System dynamics $f(\cdot)$, number of samples N, sampling horizon T
Input: Current state x_t, desired control input u_{des}, set of safe control sequences $S_t^{safe} \neq \emptyset$
Output: Safe control input u_{safe}, set of safe control sequences S_{t+1}^{safe}

$x_{t+1} \leftarrow f(x_t, u_{des})$
$S_{t+1}^{safe} \leftarrow [\,]$
for $i \leftarrow 0$ *to* $N-1$ *in parallel* **do**
  Sample $U^i \leftarrow [u_0^i, \dots, u_{T-1}^i]$
  $X^i \leftarrow rollout(x_{t+1}, f, U^i)$
  $cost_i \leftarrow cost_{CBF}(X^i)$
  if $cost_i == 0$ **then**
    $S_{t+1}^{safe}.append(U^i)$
if $S_{t+1}^{safe} \neq \emptyset$ **then**
  $u_{safe} \leftarrow u_{des}$
else
  $U^{safe} \leftarrow \arg\min_{U \in S_t^{safe}} (cost_{deviation}(U))$
  $u_{safe} \leftarrow U_0^{safe}$, $S_{t+1}^{safe} \leftarrow [U_{1:T}^{safe}]$
return u_{safe}, S_{t+1}^{safe}

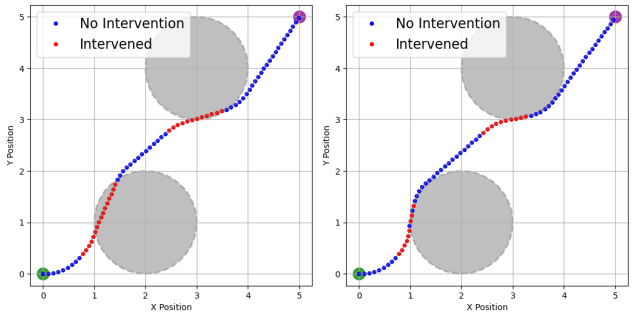
2.3 실험

시뮬레이션을 통해 알고리즘을 검증하였다. 시스템의 동역학으로는 Dubins car 모델을 사용하였다.

$$x_{t+1} = x_t + v\cos(\theta_t) \cdot \Delta t \quad (6)$$
$$y_{t+1} = y_t + v\sin(\theta_t) \cdot \Delta t \quad (7)$$
$$\theta_{t+1} = \theta_t + \omega_t \cdot \Delta t \quad (8)$$

상태 변수 x, y는 로봇의 위치 좌표, θ는 로봇의 heading angle 을 의미하고, v는 고정된 속력이다. 제어 입력은 각속도인 ω이고, 범위는 $\omega \in [-1.2, 1.2]\,rad/s$로 하였다. 타임스텝 간격 Δt는 $0.1s$로 설정하였다. 로봇의 목표는 $(0,0)$ 지점에서 출발하여 $(5,5)$에 도달하는 것이고, 안전 조건은 그림 1 의 회색 영역을 피하는 것으로 하였다.

안전 필터 적용에 필요한 CBF 로는 DeepReach[3]로 학습한 value function 을 사용하였다. 학습된 value function 은 미래 시점의 상태를 고려하므로, 장애물과의 거리만을 고려하여 안전성을 판단할 때에 비해 샘플링 horizon 이 짧을 때에도 더 나은 안전 보장 효과를 보인다. 그림 1 에서 CBF 학습 효과를 검증하였다. 필터가 개입하지 않은 시점은 파란색, 개입한 시점은 빨간색으로 궤적을 표시하였다.

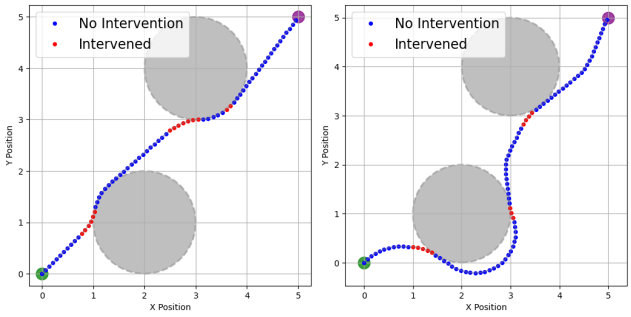


(a) (b)

그림 1. CBF 학습 효과 검증. 샘플링 horizon $\boldsymbol{T = 6}$일 때 (a)에는 장애물과의 거리만을 고려하고, (b)에는 학습된 CBF 를 사용하여 $cost_{CBF}$를 계산한 안전 필터의 적용 결과를 비교했다. 학습된 CBF 가 짧은 horizon 에서도 안전한 샘플을 잘 찾는 것을 확인하였다.

그림 2 에서는 서로 다른 nominal controller 에서 안전 필터의 효과를 검증하였다. 그림 2(a)에서는 최적화 기반 Model Predictive Control(MPC)을 통해 안전 조건과 관계없이 목적지 도달만을 고려하는 제어기를 nominal controller 로 사용하였다. 그림 2(b)에서는 실험자가 직접 조작하는 manual control 을 nominal controller 로 사용하였다.

그림 2(a)의 nominal controller 는 안전성을 고려하지 않는 제어기를 사용하기 때문에 원래는 회색 영역을 가로질러 가야 하지만, 안전 필터를 적용한 후에는 위험 상황마다 필터가 개입하여 로봇이 안전한 궤적을 따르게 된다. 그림 2(b)에서는 실험자가 임의로 제어 입력을 넣어 주는 상황에서도 안전 필터가 잘 동작하는 것을 보인다. 안전한 상황에서는 주어진 명령을 그대로 따르고, 위험 영역에 가까워지거나 의도적으로 진입하려고 할 때에는 안전 필터가 개입하여 안전한 상태가 유지된다.



(a) (b)

그림 2. Nominal controller 에 따른 실험 결과. (a)에는 목적지 도달만을 고려하는 최적화 기반 제어기, (b)에는 manual control 에 안전 필터를 적용한 결과를 나타냈다.

Ⅲ. 결론

본 논문에서는 샘플링을 통해 미래 궤적의 안전성을 판단하고 필요한 경우에만 제어 입력에 개입하여 안전을 보장하는 샘플링 기반 안전 필터 알고리즘을 제안하였다. 모델의 동역학을 알고 있는 상황에서 샘플링과 CBF 제약 조건을 활용하는 알고리즘을 고안하고 Dubins car 모델을 따르는 시뮬레이션 환경에서 검증하였다. 실험을 통해 알고리즘이 주어진 임의의 nominal controller 에 적절히 개입하여 안전을 유지하는 것을 확인하였다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학 ICT 연구센터사업의 연구결과로 수행되었음 (IITP-2025-RS-2023-00259991)

참 고 문 헌

[1] Brunke, Lukas, et al. "Safe Learning in Robotics: From Learning-Based Control to Safe Reinforcement Learning," Annual Review of Control, Robotics, and Autonomous Systems 5.1 (2022), pp. 411-444.
[2] Zeng, Jun, Bike Zhang, and Koushil Sreenath. "Safety-Critical Model Predictive Control with Discrete-Time Control Barrier Function," 2021 American Control Conference (ACC), pp. 3882-3889.
[3] Bansal, Somil, and Claire J. Tomlin. "DeepReach: A Deep Learning Approach to High-Dimensional Reachability," 2021 IEEE International Conference on Robotics and Automation (ICRA), pp. 1817-1824.

라이다 객체 검출을 위한 몬테카를로 드롭아웃 효과에 관한 연구

이종록, 박민철*
한국전자기술연구원, *한국전자기술연구원
jongrok@keti.re.kr, *mincheol.p@keti.re.kr

Rethinking Monte Carlo Dropout for LiDAR Object Detection

Jongrok Lee, Mincheol Park*
Korea Electronics Technology Institute, *Korea Electronics Technology Institute

요 약

본 연구는 라이다 객체 검출(LiDAR Object Detection) 분야에서 널리 사용되는 CenterPoint 모델을 기반으로, Monte Carlo Dropout(MC 드롭아웃) 기법의 새로운 활용 가능성을 탐구한다. MC 드롭아웃은 완전연결(FC) 계층에서 과적합 방지를 위해 널리 사용되었으나, 이미지 기반 비전 모델에서는 거의 활용되지 않는 기법이다. 그러나 본 연구에서는 MC 드롭아웃이 라이다 기반 3D 객체 검출에서 Spatial Reasoning 능력을 향상시킬 수 있음을 실험적으로 확인하였다. 구체적으로, 백본 네트워크의 특징맵 단계에 MC 드롭아웃을 적용하여 학습과 추론 과정 모두에서 불확실성을 반영하도록 하였으며, 그 결과 보행자(Pedestrian) 클래스에서 평균 정밀도(AP)가 2~3포인트 향상되는 성과를 얻었다. 제안 방법은 구현이 간단하여 CenterPoint뿐 아니라 다양한 LiDAR 객체 검출 네트워크에도 손쉽게 적용 가능하다. 본 연구는 드롭아웃의 활용을 재고찰함으로써, 3D 객체 검출의 성능 향상을 달성할 수 있는 실용적인 접근법을 제시한다.

Ⅰ. 서 론

LiDAR 기반 3D 객체 검출은 자율주행과 로봇 분야에서 핵심 기술로, BEV(Bird's-Eye View) 표현을 활용하는 CNN 백본 구조가 널리 사용되고 있다. 그러나 포인트 클라우드의 희소성(sparsity)으로 인해 BEV 백본이 확보할 수 있는 공간적 특징(Spatial Feature)의 밀도가 제한되며, 이는 특히 보행자나 자전거와 같이 작은 객체의 인식 성능 저하로 이어질 수 있다. 기존의 CNN 연구에서는 과적합 방지를 위해 드롭아웃(Dropout) 기법이 활용되었으나, 개별 픽셀 단위로 무작위 제거를 수행하는 Monte Carlo Dropout(MC 드롭아웃)은 공간적 특징을 훼손하는 부작용이 있어 주로 채널 단위로 제거하는 Dropout2D나 Spatial Dropout이 사용되어 왔다.[1] 하지만 LiDAR 도메인에서는 제한적인 Spatial Feature를 오히려 다양한 형태로 학습하게 함으로써 공간 추론(Spatial Reasoning) 능력을 강화할 가능성이 있다. 본 연구는 이러한 가설을 검증하기 위해, 학습 시 BEV 백본의 특징맵 단계에 MC 드롭아웃을 적용하여 제한적인 공간 특징의 활용을 다양화함으로써 객체 인식 성능을 향상시키는 접근을 제안한다. 또한 최근 제안된 BoS(Bounding Box Stability)[2] 연구에서 착안하여, MC 드롭아웃을 통한 특징 변동성이 모델의 일반화에 긍정적인 영향을 줄 수 있음을 고려하였으며, 간단한 드롭아웃 레이어 추가만으로도 보행자(Pedestrian) 클래스에서 AP가 약 2~3포인트 향상되는 결과를 확인하였다.

Ⅱ. 본론

본 연구에서는 제안한 Monte Carlo Dropout(MC 드롭아웃) 기법의 효과를 검증하기 위해 KITTI 데이터셋[3]을 활용하여 3D 객체 검출 실험을 수행하였다. 검출 모델로는 라이다 기반 객체 탐지 모델 중, BEV 구조를 사용하는 CenterPoint-PointPillar[4] 구조를 선택하였으며, 입력 포인트 클라우드를 BEV 형태로 변환하기 위해 두 가지 다른 보셀 크기(voxel size)를 설정하였다. 구체적으로, 0.16m와 0.32m 두 가지 보셀 크기를 적용하여 BEV 상 Spatial Feature 해상도 차이에 따른 드롭아웃 성능 변화를 비교하였다. 제안 기법은 백본 네트워크의 특징맵 단계 1과 2에 MC 드롭아웃 레이어를 삽입하는 방식으로 구현하였다. 학습 단계에서는 일반적인 드롭아웃과 동일하게 무작위 feature를 제거하며, 추론 단계에서도 드롭아웃을 유지하여 불확실성을 반영하는 Monte Carlo 추론을 수행하였다. 이를 통해 희소성이 큰 BEV 특징맵에서 다양한 공간적 표현을 학습할 수 있도록 하였다.

실험 결과, Car 클래스에서는 성능 차이가 거의 나타나지 않았으며, Cyclist 클래스의 경우 소폭 성능 저하가 관찰되었다. 반면, BEV상 크기가 작아 가장 제한적인 Spatial Feature를 가지는 Pedestrian 클래스에서 두 보셀 크기(0.16, 0.32) 모두 평균 정밀도(AP)가 약 2~3 포인트 향상되는 개선 효과가 확인되었다. 이는 포인트 클라우드의 희소성으로 인해 BEV 백본이 충분한 공간 정보를 확보하기 어려운 상황에서, MC 드롭아웃이 오히려 다양한 공간적 표현 학습을 유도하여 Pedestrian 검출의 Spatial Reasoning 능력을 강화한 결과로 해석할 수 있다.

표 1. (Moderate) MC 드롭아웃에 따른 모델 별 성능 변화 (AP)
Table 1. Performance Variation by Model with MC Dropout (AP)

	Car	Pedestrian	Cyclist
0.32	63.41	30.29	45.67
0.32_MD	62.58	**33.19**	45.89
0.16	72.24	40.18	59.85
0.16_MD	72.57	**42.33**	58.49

표 2. (Hard) MC 드롭아웃에 따른 모델 별 성능 변화 (AP)
Table 2. Performance Variation by Model with MC Dropout (AP)

	Car	Pedestrian	Cyclist
0.32	58.99	28.21	42.92
0.32_MD	58.34	**30.49**	42.56
0.16	66.54	37.95	55.83
0.16_MD	**69.53**	**39.39**	54.42

Ⅲ. 결론

본 연구에서는 LiDAR 기반 3D 객체 검출에서 Monte Carlo Dropout(MC 드롭아웃)의 가능성을 재고찰하였다. 기존 CNN 이미지 도메인에서는 공간적 특징(Spatial Feature)을 훼손하는 문제로 인해 거의 사용되지 않던 MC 드롭아웃을, 포인트 클라우드의 희소성으로 인해 Spatial Feature가 부족한 BEV 백본에 적용함으로써 오히려 Spatial Reasoning 능력을 강화할 수 있음을 실험적으로 확인하였다.

이는 제안 방식이 단순히 드롭아웃 레이어를 추가하는 것만으로 구현 가능하다는 점에서, 다양한 LiDAR 객체 검출 네트워크에도 손쉽게 확장될 수 있는 장점이 있다. 다만 본 연구에서는 특정 백본 구조와 제한된 데이터셋에 국한하여 검증하였으므로, 향후에는 PV-RCNN, BEVDet 등 다양한 3D 검출 모델로의 확장 실험과 Waymo, nuScenes와 같은 대규모 데이터셋에 대한 검증이 필요하다.

ACKNOWLEDGMENT

※ 본 연구는 산업통상자원부의 제조 기반 생산 시스템 사업(No. 00442974, 대규모 토석 운반 자동화를 위한 덤프트럭용 자율작업 및 운영 시스템 개발)의 지원을 통해 연구가 수행되었음

※ 본 연구는 과학기술정보통신부가 주최하고 정보통신기획평가원 및 한국정보통신기술협회가 주관하는 「2025년도 인공지능 챔피언(AI Champion) 대회」의 지원을 받아 수행되었음.

참 고 문 헌

[1] Tompson, J., Goroshin, R., Jain, A., LeCun, Y., & Bregler, C. (2015). Efficient object localization using convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 648-656).

[2] Yang, Y., Wang, W., Chen, Z., Dai, J., & Zheng, L. (2024). Bounding box stability against feature dropout reflects detector generalization across environments. arXiv preprint arXiv:2403.13803.

[3] Geiger, A., Lenz, P., Stiller, C., & Urtasun, R. (2013). Vision meets robotics: The kitti dataset. The international journal of robotics research, 32(11), 1231-1237.

[4] Yin, T., Zhou, X., & Krahenbuhl, P. (2021). Center-based 3d object detection and tracking. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 11784-11793).

실시간 복합음향 잡음 저감을 위한 WaveNet-VNNs 기반 ANC 시스템 설계 및 구현

이상민*, 최승렬**, 김태리***, 이예지****, 신애리*****, 김우현*****, 유길상*****

영남대학교*, 인천대학교**, 경북대학교***, 상명대학교****, 고려대학교*****
cross01wmin@naver.com, csy1736@naver.com, telle2409@knu.ac.kr, thisprevision@gmail.com,
shinairi0129@gmail.com, kimwoo320@naver.com, ksyoo@korea.ac.kr

Design and Implementation of a WaveNet-VNNs-Based ANC System for Real-Time Composite Acoustic Noise Reduction

Sangmin Lee*, Seungryeol Choi**, Taeri Kim***,Yaeji Lee****, Aeri Shin*****, Woohyun Kim*****, Gilsang Yoo*****

Yeungnam University*, Incheon National University**, Kyungpook National University***,
Sangmyung University****, Korea University*****

요 약

소음은 스트레스 유발 등 정신적 피해뿐 아니라 장기간 노출 시 심혈관 질환, 청력 손상과 같은 신체적 문제까지 초래할 수 있다. 이러한 소음을 저감하기 위한 능동형 소음 제어(ANC)는 소음의 반대 위상 신호를 생성하여 소음을 상쇄하는 방식으로 주목받고 있다. 그러나 기존 ANC 연구는 주로 시간 지연 및 저감 성능에 초점을 두고 있어 복합 환경에서의 적용에는 한계가 있다. 본 연구에서는 복합음향 중 원하는 소리만을 선택적으로 저감할 수 있도록 WaveNet-VNNs기반 인과적 확장합성곱을 적용하여 신호의 장기 시간 의존성을 학습하고, Volterra Neural Networks를 통해 고차 비선형 특성을 모델링함으로써 역위상 신호를 생성하는 새로운 ANC 시스템을 제안하였다. 실험결과, 음성 분리 및 분류, 소음 저감 과정을 경량화해 계산량을 최소화하면서도 실시간 처리 성능을 확보하였으며, dBA -37.61과 NMSE -35.28의 우수한 저감 성능을 확인하였다. 본 시스템은 실시간 맞춤형 소음 저감 솔루션의 구현을 위한 기술적 기반을 마련하고, ANC 기반 소음 제어 연구의 적용성·범용성·실용성을 높이는 데 기여할 것으로 기대된다.

Ⅰ. 서 론

소음은 현대 사회에서 삶의 질을 저하시키는 주요 환경 요인 중 하나이다. 환경부 보고에 따르면 소음은 스트레스 증가, 수면 방해, 심혈관계 질환 등 인체 건강에 부정적인 영향을 미치는 것으로 알려져 있으며, 이러한 문제를 해결하기 위해 다양한 소음 저감 연구가 진행되어 왔다. 기존 연구에서는 물리적 차폐를 기반으로 하는 수동적 소음 제어 방식이 주류를 이루었으나 인공 지능 및 공학 기술의 발전은 능동형 소음 제어(ANC, Active Noise Control)의 적용 가능성을 확대시키며 물리적 차폐를 보완하거나 대체할 수 있는 새로운 가능성을 열어 주고 있다.

기존 ANC 관련 연구들은 주로 비선형 특성의 반영, 시간 지연(time latency) 문제의 개선, 그리고 전체 소음 저감 성능 향상에 초점을 맞추어 발전해 왔다. 하지만 이러한 시스템들은 모든 입력 음향을 일괄적으로 저감한다는 한계가 있어, 실환경 적용 시 필수 정보가 포함된 신호(예: 음성 안내, 경고음 등)까지 함께 감소시키는 문제가 존재한다. 만약 특정 소리만을 선별적으로 저감하는 방식이 구현된다면, 일상생활뿐 아니라 산업 현장과 같은 복합 환경에서도 필요한 소리는 유지하면서 유해한 소음만을 효과적으로 줄일 수 있어 소음으로 인한 신체적 피해를 획기적으로 감소시킬 수 있다. 따라서 본 연구에서는 복합 음향 환경에서 특정 소리만을 선택적으로 저감할 수 있는 실시간 능동형 소음 제어(ANC) 시스템을 설계하고 구현하였다. 이를 위해 WaveNet-Volterra Neural Networks(WaveNet-VNNs)를 기반으로 한 음성 분리 및 분류 모듈을 설계하고, 해당 모듈을 ANC 프레임워크에 통합함으로써 필수 음향은 유지하면서 원치 않는 소음만을 저감할 수 있는 선택적 소음 제어 기법을 제안하였다. 본 연구는 핵심 기술은 다음과 같다. 첫째, 기존 ANC 기술의 한계로 지적되던 전체 음향 일괄 저감 문제를 해결할 수 있는 선택적 소음 제어 구조를 제시하였다. 둘째, 복합 환경에서도 실시간성이 유지될 수 있도록 모델 경량화와 연산 최적화를 수행하였다. 셋째, 제안된 시스템은 실제 환경에서의 테스트를 통해 실효성을 검증함으로써 차세대 맞춤형 소음 제어 솔루션의 기술적 기반을 제공한다.

Ⅱ. 본론

제안한 선택적 소음 저감 시스템의 전체 구조는 그림 1과 같다. 먼저 입력된 음향 신호는 음성 분리 모듈을 통해 여러 구성 요소로 분리되며, 이후 분류 단계에서 저감 대상 음향(A)과 비저감 대상 음향(B)으로 구분된다. 분류 과정에서 A로 판별된 신호에 대해서만 능동형 소음 제어를 적용함으로써 해당 신호에 대한 반대 위상 신호를 생성하고 제거한다. 이는 단일 채널 입력 환경에서도 소음을 독립적으로 분리하여 실시간으로 역위상 신호를 생성함으로써 효율적인 선택적 소음 저감을 가능하게 한다.

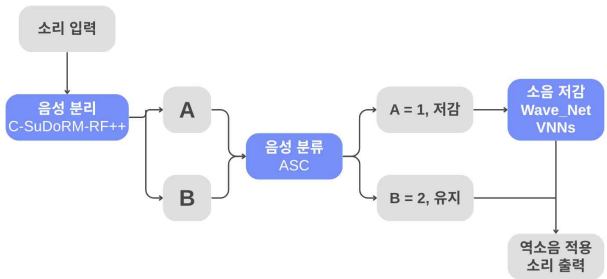


그림 1. 제안한 복합음향 선택 저감용 ANC 모델의 전체 구성도

(1) 데이터 수집 및 전처리

소음 저감 모델 학습은 기존 연구와의 정량적 비교를 위해 타 논문과 동일한 데이터 및 전처리 조건을 적용하였다. 소음 저감 학습에는 DEMAND 및 MN-SNSD 데이터를 벤치마킹 데이터로 활용하였으며, 분리 및 분류 모델 학습에는 서울교통공사 1-8호선 안내방송 음원과 도시 환경 소리(철도

및 항공기 소음)를 수집하였다. 모든 음원은 .wav, 16 kHz, mono 형식으로 통일하였으며, 30초 단위로 랜덤 분할하여 방송(spk1) 및 소음(spk2) 클래스로 저장한 후, 데이터 불균형 문제를 보정하기 위해 인덱스 기반 랜덤 복제를 적용하였다. 두 음원의 혼합은 더 짧은 신호 길이에 맞추어 진행하였으며, 목표 SDR 5 dB가 되도록 두 신호의 에너지 비율을 계산하여 조정하였다. 최종 데이터셋은 train:val:test = 8:1:1의 비율로 분할하였고, 분할 전 원천 데이터의 규모는 철도 49시간 30분, 항공기 24시간 55분이다.

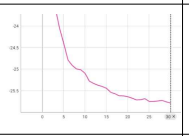
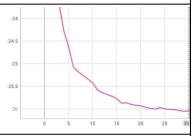
(2) 소음 분리 및 분류

소음 분리 과정에서는 C-SuDoRM-RF++[1]의 인과적 변형 구조를 기반으로 하여 단일 채널 혼합 음향으로부터 안내 방송과 소음을 실시간으로 분리된다. 제안한 분리 모델은 Encoder-U-ConvBlock-Mask Estimator-Decoder 구조로 구성되며, 모든 합성곱 연산을 인과적 1D convolution으로 대체함으로써 스트리밍 기반 실시간 처리가 가능하도록 설계하였다. 분리된 신호는 경량 1D CNN 기반 Audio Segment Classifier(ASC)를 통해 방송과 소음으로 분류되며, 본 분류기는 전역 평균 풀링과 sigmoid 기반 확률 산출 방식에 따라 각 프레임의 음향을 판별한다. ASC는 소음으로 판별된 구간만을 이후의 소음 저감 단계로 전달하는 게이트 역할을 수행함으로써 전체 시스템의 연산 효율이 향상되도록 하였다.

(3) 소음 저감

소음 저감 단계에서는 WaveNet-VNNs[2] 구조를 기반으로 인과적 확장 합성곱을 적용하여 신호의 장기 시간 의존성을 학습하고, Volterra Neural Networks를 통해 고차 비선형 특성을 모델링함으로써 역위상 신호를 생성하였다. 본 모델의 성능과 실시간성 간의 균형을 확인하기 위해 기존 모델의 설정과 비교하며 블록 수(num_stacks과 dilations) 구성에 따른 연산량과 저감 성능을 비교한 결과는 표 1과 같다.

표 1. 블록 수 변화에 따른 dBA/NMSE 성능 및 추론 속도 비교

블록 수	기존 모델	num_stacks = 3 dilations = 8	num_stacks = 2 dilations = 9
train_loss	-		
dBA(dB)	-35.30	-37.61	-37.81
NMSE(dB)	-43.97	-35.28	-41.31
추론속도(ms/sec)	10.60	10.30	9.50

num_stacks=3, dilations=8 설정은 dBA(A-weighted decibel) 기준에서 가장 높은 소음 저감 성능을 나타냈으나, 연산량 증가로 인해 추론 속도가 가장 느리게 측정되었다. 반면, num_stacks=2, dilations=9 설정은 NMSE(Normalized Mean Squared Error)가 개선되었으나 dBA 성능은 다소 저하되었으며, 연산 효율상 속도 개선 역시 제한적으로 나타났다. 그림 2의 결과는 원음 d(n), 역소음 u(n), 그리고 감쇠된 신호 e(n)를 나타내고 있으며, 이는 입력 신호로부터 소음이 효과적으로 감쇠되고 있는 것으로 나타났다.

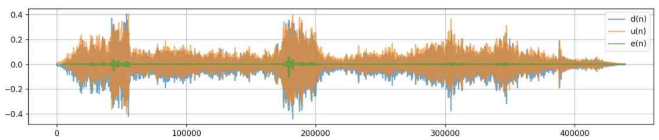


그림 2.제안한 저감 모델의 실행 결과

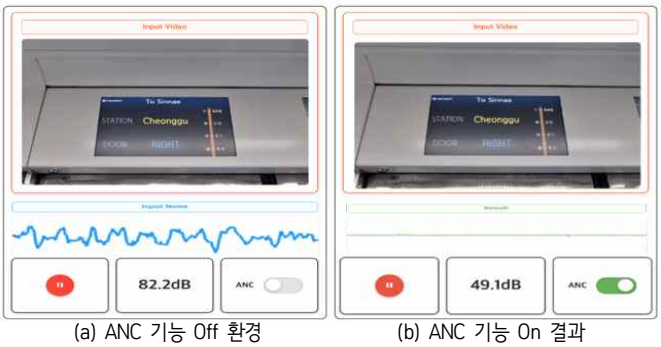
(4) 공동 학습

공동 학습((Joint Training)에서는 분리, 분류, 저감 단계로 구성된 다중 작업(Multi-task) 구조를 하나의 통합 프레임워크로 학습시키기 위해 각 단계별 손실 항목을 최적의 가중치로 조합한 Total Loss 함수를 설계하였으며, 공동 학습에서의 Total Loss 구성 요소 및 가중치 값은 표 2와 같다. 공동 학습을 통해 분리 모듈과 ANC 모듈 간의 과도한 경쟁을 억제하면서 전체 파이프라인의 품질을 균형 있게 향상시키고, 통합 손실 최적화를 기반으로 실제 추론 단계에서 18.9 ms/sec의 지연 시간을 달성하여 실시간 처리 성능[3]을 충족한다.

표 2. 공동 학습에서의 Total Loss 구성 요소 및 가중치

구성 요소	목적	가중치
Final_quality	전체 파이프라인 사용자 체감 품질 향상	0.12
Separation	음성 분리 단계 성능 반영	0.45
ANC_total	소음 저감 단계 성능 반영	0.28
Classifier	분류 인식 단계 성능 반영	0.15

(5) 모델 구현 및 실제 환경 적용



(a) ANC 기능 Off 환경 (b) ANC 기능 On 결과

그림 3. 모델 구현 및 실제 환경 적용 결과

모델의 성능은 서울 지하철 6호선 환경에서 수집한 실제 데이터를 대상으로 검증하였다. 안내 방송음과 배경 소음을 동시에 포함하는 입력 신호를 사용하여 분리 및 저감 성능을 실험한 결과는 그림 3과 같다. 현장 실험 결과, 평균 30 dB 내외의 소음 저감 효과가 확인되었다. 특히, 철도 주행 소음이 효과적으로 억제되면서 안내방송이 또렷하게 전달됨을 확인하였다.

Ⅲ. 결론

본 연구에서는 분리-분류-저감 파이프라인을 기반으로 실시간 환경에서 특정 소음을 선택적으로 처리하는 새로운 ANC 시스템을 제안하였다. 모듈 간 오버헤드를 최소화하고 스트리밍 처리를 전제로 설계함으로써 연산 효율성과 지연 관리 측면에서 유의한 이점을 확보하였다. 실험 결과, 비선형 및 복합 소음 환경에서도 dBA와 NMSE 지표의 개선을 안정적으로 달성하였으며, 교통, 모빌리티, 웨어러블 등 실시간 처리가 필수적인 응용 분야에 즉시 적용 가능한 수준의 일관된 성능을 확인하였다. 향후 연구에서는 임베디드 환경 최적화, 데이터 다양화, 자기 적응 학습(self-adaptive learning) 기법을 통해 시스템의 적용 범위를 확장하고 성능향상을 진행할 계획이다.

참 고 문 헌

[1] Tzinis, E., Wang, Z.-Q., & Smaragdis, P. (2020). SuDoRM-RF: Efficient and compact convolutional neural networks for audio source separation. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 29, 1683-1696.

[2] van den Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., ... & Kavukcuoglu, K. (2016). WaveNet: A generative model for raw audio. arXiv preprint arXiv:1609.03499.

[3] 3GPP. (2024). Universal Mobile Telecommunications System (UMTS); LTE; Enhanced Voice Services (EVS) codec; ANSI C code (fixed-point) (ETSI TS 126 442 V18.0.0, May 2024). European Telecommunications Standards Institute (ETSI).

통신 음영 지역의 에너지 효율 향상을 위한 강화학습 기반 지능형 반사 표면 제어 기법 연구

김지형, 김정현*
한국전자통신연구원, *세종대학교

savant21@etri.re.kr

A Study on a RL-based RIS Control Scheme for Energy Efficiency Enhancement in Coverage Hole Environments

Kim Jihyung, Kim Junghyun*
ETRI, *Sejong Univ.

요 약

본 논문은 통신 음영 지역의 에너지 효율(EE) 최대화를 위해, 지능형 반사 표면(RIS)의 실질적인 운용 비용(전력/지연)을 보상 함수에 직접 통합하는 심층 강화학습(DRL) 기반 제어 기법을 제안한다. 제안 기법은 시뮬레이션을 통해 기존 방식 대비 낮은 SNR(신호 대 잡음비) 환경에서의 우수한 EE 성능과 개선된 복잡도-성능 효율성을 입증하였다.

Ⅰ. 서 론

차세대 무선 통신 시스템의 핵심 기술 중 하나인 지능형 반사 표면(RIS)은 거의 수동적인 반사 소자를 통해 무선 채널 환경을 프로그래밍함으로써, 통신 성능을 유의미하게 향상시킬 수 있는 잠재력을 가진다 [1]. 특히, 셀 경계나 실내 음영 지역과 같이 채널 이득이 낮은 환경에서 RIS 는 비용 효율적으로 커버리지를 확장하고 신호 품질을 보장하는 핵심 조력자로 주목받고 있다.

그러나 RIS 의 실용적인 구현을 위해서는 소자의 위상 전환과 제어부 구동에 소모되는 전력 등 현실적인 운용 비용을 반드시 고려해야 한다 [2]. 신호 대 잡음비가 극도로 낮은 환경에서는 RIS 를 활성화하여 얻는 빔포밍 이득이 이러한 운용 비용을 상쇄하지 못할 수 있으며, 이 경우 RIS 를 비활성화하는 것이 시스템 전체의 EE 를 오히려 향상시키는 최적의 선택이 된다. 따라서, 시시각각 변하는 채널 환경에 맞춰 RIS 의 성능 이득과 운용 비용 간의 복잡한 상충 관계를 실시간으로 최적화하는 동적 제어 정책이 필수적이다.

본 논문에서는 이러한 다중 목표 최적화 문제를 해결하기 위해 심층 강화학습(DRL)에 기반한 접근법을 제안한다. DRL 은 불완전한 채널 상태 정보 하에서도 동적인 자원 할당 정책을 효과적으로 학습할 수 있는 강력한 프레임워크이다. 구체적으로, 본 논문은 RIS 제어에 따른 전력 및 지연 비용을 명시적인 페널티 항으로 포함하는 새로운 보상 함수를 설계하고, 이를 기반으로 EE 를 최대화하는 TD3 에이전트를 제안한다. 제안하는 에이전트는 채널 품질에 따라 RIS 를 선별적으로 활성화하는 정책을 학습하여, 열악한 통신 환경에서 시스템의 에너지 효율을 최적화한다.

Ⅱ. 시스템 모델 및 전력-지연 인지 TD3 (PL-TD3)

본 논문에서는 RIS 가 적용된 일반적인 무선 통신 환경을 고려한다. 대표적인 시스템 모델로서, 하나의 기지국과 RIS 의 도움을 받아 통신하는 단일 사용자가 존재하는 단일 셀 하향링크 환경을 가정한다. 여기서 사용자는 셀 경계나 장애물로 인해 채널 이득이 낮은 통신 음영 지역에 위치하는 상황을 상정한다. 따라서 본 논문의 목표는, RIS 운용 비용을 포함한 시스템의 총 EE 를 최대화하는 최적의 RIS 제어 정책을 찾는 것이다. 이를 위해 EE 는 식 (1)과 같이 정의된다.

$$EE = \frac{R(\mathbf{W}, \Phi)}{P_{tx} + P_{ctrl}(\Phi) + P_{circ}(\Phi)} \quad (1)$$

여기서 $R(\mathbf{W}, \Phi)$는 기지국 프리코더 $\mathbf{W}$ 와 RIS 위상 행렬 Φ에 의해 결정되는 채널 용량 기반의 데이터 전송률이다. P_{tx}는 기지국의 송신 전력이며, 본 논문의 핵심 고려사항인 $P_{ctrl}(\Phi)$와 $P_{circ}(\Phi)$는 각각 RIS 의 위상 구성 및 제어 신호 처리에 소모되는 제어 전력과 소자 및 제어부 구동에 필요한 회로 전력을 의미한다.

이러한 동적 환경에서 EE 를 최적화하기 위해, 본 논문은 PL-TD3 (Power and Latency Aware TD3) 프레임워크를 제안한다. PL-TD3 의 목표는 최적의 정책 $\pi(s_t)$를 학습하는 것이다. 여기서 최적의 정책이란, 각 상태에서 장기적으로 기대되는 누적 보상의 합을 최대화하는 일련의 행동을 결정하는 규칙, 즉 상태를 행동으로 매핑하는 함수를 의미한다. 에이전트의 상태 s_t는 이전 시간대의 채널 품질 정보(CQI), 데이터 전송률, 그리고 선택했던 RIS 구성을 포함하여 현재 채널 환경을 포괄적으로 인지할 수 있도록 구성된다. 이를 바탕으로 에이전트는 RIS 의 동적 활성화를 위한 게이트 $g_t \in \{0, 1\}$와 코드북 $\{\Phi_m\}$ 내에서 최적의 위상 행렬을 선택하기 위한 인덱스 m_t 의 조합으로 정의되는 행동 a_t 을 결정한다. 제안 기법의 핵심인 보상 r_t는 에이전트가

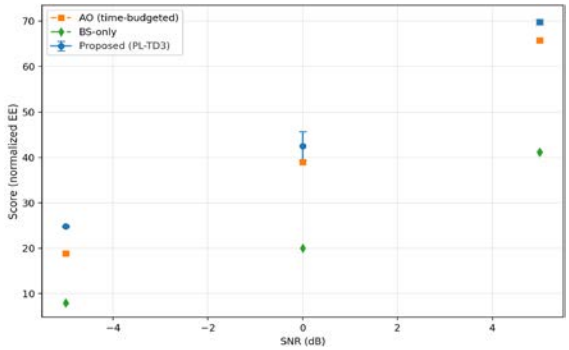


그림 1. SNR 에 따른 에너지 효율(EE)

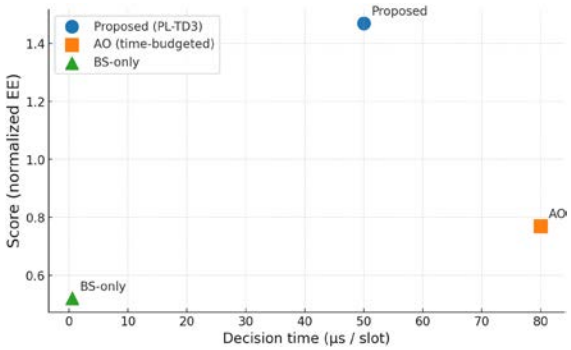


그림 2. 결정 시간과 에너지 효율(EE)

EE 와 운용 비용 간의 상충 관계를 학습하도록 아래 식 (2)와 같이 설계된다.

$$r_t = \alpha EE_t - \lambda P_{ctrl}(\Phi_t) - \lambda_c P_{circ}(\Phi_t) - \mu \Delta t_t \quad (2)$$

(2)에서 첫 번째 항은 시스템의 주 목표인 에너지 효율이며, 이후의 항들은 각각 제어 전력, 회로 전력, 그리고 정책 결정 지연 시간에 대한 페널티를 나타낸다. 제안하는 PL-TD3 에이전트의 학습 과정은 매 시간 슬롯마다 상태 s_t 를 관측하고, 정책 $\pi(s_t)$에 따라 행동 a_t 를 선택하여 시스템에 적용한다. 그 결과로 관측된 성능 지표들을 바탕으로 (2)에 따라 보상 r_t 를 계산하고, 상태 천이 정보 (s_t, a_t, r_t, s_{t+1})를 리플레이 버퍼에 저장하여 액터-크리틱 신경망을 TD3 알고리즘에 따라 업데이트한다. 이러한 학습 절차는 5G/6G 이동통신 시스템의 제어 프레임워크 내에서 적용될 수 있으며, 에이전트의 상태 관측은 기지국이 기존 신호를 기반으로 채널을 측정하고 사용자로부터 CQI 를 피드백받는 과정에 해당하며, 행동 적용은 결정된 RIS 구성을 기지국이 하향링크 제어 채널 등을 통해 RIS 제어기에 전달하는 과정으로 실행될 수 있다.

Ⅲ. 시뮬레이션 및 결과 분석

제안하는 기법의 성능을 정량적으로 검증하기 위해, 통신 음영 지역 시나리오를 대상으로 시뮬레이션을 수행하였다. 이번 시뮬레이션의 핵심은 제안하는 에이전트가 RIS 의 채널 이득과 운용 비용 간의 상충 관계를 효과적으로 학습하는지를 확인하는 것이다. 성능 지표로는 시스템의 총 EE 와 정책 결정 시간을 분석하였으며, 기존 기법들과의 비교를 통해 제안 기법의 효용성을 확인하였다. 성능 비교를 위해 전통적인 반복 최적화 방식인 교대 최적화(AO)와 RIS 를 사용하지 않는 BS-only 를 비교 대상으로 설정하였다 [3]. 공정한 비교를 위해 AO 는 제안 기법과 유사한 시간 예산 내에서 수행되도록 하였으며, BS-only 는 RIS 적용에 따른 순수 성능 향상을 확인하기 위한 기준선으로 사용된다. 시뮬레이션 결과는 그림 1 과 2 에 나타내었다. 그림 1 에서 제안하는 PL-TD3 은, 특히 신호 대 잡음비가 낮은 환경에서 비교 대상들보다 우수한 EE 를 달성함을 확인하였다. 이는 에이전트가 RIS 운용 비용을 보상 함수를 통해 인지하고, 채널 이득이 불충분할 경우 불필요한 활성화를 회피하도록 학습했기 때문이다. 비용을 인지하지 못하는 AO 는 RIS 를 기계적으로 활성화하여 오히려 EE 가 저하되는 것과 대조적인 결과이다. 또한 그림 2 는 제안 기법이 AO 대비 훨씬 짧은 결정 시간 내에 더 높은 EE 를 달성함을 보여준다. 이는 단일 순방향 연산만으로 동작하는 DRL 추론 방식이 반복적인 최적화가 필요한 AO 에 비해 실시간성에 매우 유리하기 때문이다. 결론적으로, 제안 기법은 더 나은 성능을 더 효율적으로 달성하며 실제 통신 환경에서 높은 적용 가능성을 보여준다.

Ⅳ. 결론

본 논문은 통신 음영 지역의 에너지 효율 최적화를 위해, RIS 의 실질적인 운용 비용을 고려하는 심층 강화학습 기반의 PL-TD3 제어 기법을 제안하였다. 시뮬레이션 결과, 제안하는 기법은 기존 교대 최적화 방식 대비 낮은 SNR 영역에서 우수한 에너지 효율 성능을 달성함을 확인하였다. 특히 RIS 를 선별적으로 비활성화하는 정책을 학습하여 불필요한 전력 소모를 효과적으로 방지하였다. 또한, 제안 기법은 AO 방식보다 훨씬 낮은 결정 시간 내에 동작하여 복잡도-성능 측면에서도 우수한 효율을 보여주었다. 이러한 결과는 제안 기법이 성능과 실시간성을 동시에 만족함을 보여주고 있다. 따라서 제안하는 PL-TD3 기법은 동적 제어가 필수적인 차세대 무선 통신 시스템의 에너지 효율을 향상시키는 효과적인 해결책이 될 수 있을 것으로 기대된다.

ACKNOWLEDGMENT

본 논문은 2025 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No. RS-2022-II220436, 지능형 재구성 표면 중계기 표준기술개발)

참 고 문 헌

[1] M. Di Renzo et al., "Smart Radio Environments Empowered by Reconfigurable Intelligent Surfaces: How It Works, State of Research, and The Road Ahead," IEEE Journal on Selected Areas in Communications, vol. 38, no. 11, pp. 2450-2525, Nov. 2020.

[2] C. Huang et al., "Reconfigurable Intelligent Surface Assisted Communications: A Power Consumption Model," in Proc. IEEE Global Communications Conference (GLOBECOM), pp. 1-6, Dec. 2020.

[3] Q. Wu and R. Zhang, "Intelligent Reflecting Surface Enhanced Wireless Network via Joint Active and Passive Beamforming," IEEE Transactions on Wireless Communications, vol. 18, no. 11, pp. 5394- 5409, Nov. 2019.

DDQN을 활용한 NR-V2X 모드2의 동적 리소스 예약 조절 기법

김성현*, 김채현*, 양영준*, 김동균*
*경북대학교
*sunghyunkim@knu.ac.kr, *hyeon917@knu.ac.kr, *youngj719@knu.ac.kr, *dongkyun@knu.ac.kr

A DDQN-based Control Method for Dynamic Resource Reservation in NR-V2X Mode 2

*Kim Sung Hyun, *Kim Chae Hyeon, *Yang Young Joon, *Kim Dong Kyun
*Kyungpook National University

요 약

본 논문은 3GPP Release 16 표준에 정의된 NR-V2X의 차량 대 차량(V2V) 통신 환경을 다룬다. V2V 통신 모드 2에서는 반-영구 스케줄링(SPS) 방식을 통해 자원을 예약하며, 이때 RC(Reserved Counter)와 SRRI(Sidelink Reserved Resource Interval) 파라미터가 사용된다. 하지만, 차량의 주행 환경 및 트래픽 패턴 변화로 인해 데이터 크기가 동적으로 변할 경우, 사전에 고정된 RC 값으로 인해 리소스 낭비 및 활용률 저하 문제가 발생한다. 본 논문에서는 이러한 문제를 해결하기 위해, 강화학습 기법 중 하나인 DDQN(Double Deep Q-Network)을 활용하여 주변 환경 변화에 따라 동적으로 리소스 예약을 조절하는 이론적 접근법을 제안한다.

Ⅰ. 서 론

5G 네트워크 기반의 NR-V2X 기술 표준화를 위해 3GPP는 Release 16 문서를 통해 V2V 통신 방식을 정의 하였다[1]. V2V 데이터 통신을 위해 모드2의 리소스(Sidelink) 예약 기법을 Release 16에서 소개하였다. 이 중 모드 2는 기지국(eNodeB)의 개입을 최소화하며 차량이 자율적으로 통신 자원을 선택하는 방식으로, SPS(Semi-Persistent Scheduling) 기법을 사용한다. 차량은 데이터 전송이 필요할 때 RC(Reserved Counter)와 SRRI(Sidelink Reserved Resource Interval)값을 기반으로 미래의 특정 시점까지 자원을 미리 예약한다[2].

하지만, 이러한 정적 자원 예약 방식은 실시간 트래픽 수요 변화에 유연하게 대응하기 어렵다는 한계가 있다. 예를 들어, 데이터 전송이 조기에 완료되거나 전송할 데이터의 크기가 급격히 줄어들었음에도, 초기에 설정된 RC 값만큼 자원이 불필요하게 예약되어 다른 차량이 사용하지 못하는 비효율이 발생한다. 이러한 문제는 전체 네트워크의 리소스 활용률을 저하시키는 주된 원인이 된다.

본 논문에서는 이러한 문제를 해결하기 위해 강화학습을 적용하고자 한다. 특히, DDQN(Double Deep Q-Network) 에이전트를 통해 주변 통신 환경을 인지하고, 최적의 RC 값을 실시간으로 결정하여 자원 낭비를 최소화하고 활용률을 극대화하는 기법을 제안하며, 2장에서는 제안기법을 서술하고 3장 결론으로 마무리 한다.

Ⅱ. 본론

2.1 문제정의

NR-V2X 모드 2에서 차량은 패킷이 발생하면 eNodeB로부터 할당받은 자원 풀 내에서 특정 자원을 선택하고, RC 값을 설정하여 SRRI 간격으로 해당 자원을 예약한다. 이 SPS 방식은 데이터 충돌과 혼잡을 줄이는 데 효과적이다.

하지만 차량 밀도나 주변 환경 변화로 인해 통신 트래픽의 특성은 계속해서 변한다. 예를 들어, 안전 메시지의 생성 주기가 바뀌거나 전송 데이터의 크기가 감소하면, 기존에 예약된 자원이 과도하게 되어 다른 차량의 통신 기회를 박탈하고 채널 낭비를 유발한다. 본 논문에서는 '정적 RC 값'과 '동적 트래픽 수요' 사이의 불일치 문제를 해결하는 것을 목표로 한다.

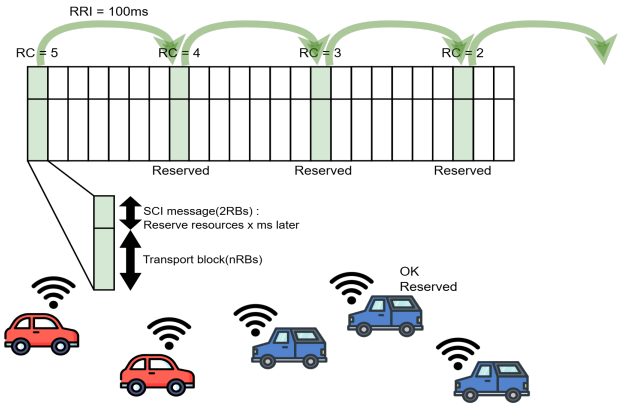


그림 1 NR-V2X 모드2 RC와 SRRI System

2.2 제안 기법: DDQN 기반 동적 RC 조절

본 논문에서는 DDQN을 이용하여 RC 값을 최적화하는 기법을 제안한다. DDQN은 기존 DQN(Deep Q-Network)이 갖는 Q-가치 과대평가(Overestimation) 문제를 완화하여 학습 안정성을 높인 알고리즘이다[3]. 제안하는 강화학습 모델의 주요 파라미터는 다음과 같이 정의한다.

● State (상태): 에이전트(차량)가 최적의 행동을 결정하기 위해 관찰하는 정보의 집합. 현재 통신 환경을 정확히 나타낼 수 있는 파라미터들로 구성된다.

· S = {RSSI (Received Signal Strength Indicator), 현재 PDR (Packet Delivery Ratio), 현재 설정된 RC 값, 잔여 데이터 버퍼 크기}

● Action (행동): 에이전트가 특정 상태(State)에서 취할 수 있는 정책적 결정. 본 문제에서는 RC 값을 동적으로 조절하는 것이 목표이다.

· A = {RC 값 1 감소, RC 값 유지, RC 값 1 증가} 와 같이 정의하거나,

특정 범위 내의 RC 값 중 하나를 직접 선택하도록 설계할 수 있다.

● Reward (보상): 에이전트가 취한 행동에 대한 성능을 평가하는 지표. 보상 함수는 리소스 활용률을 높이고 패킷 전송 성공률을 보장하는 방향으로 설계되어야 한다.

· R = (w1 * PDR) - (w2 * Wasted_Resources)

여기서 w1과 w2는 가중치이며, PDR은 높을수록 긍정적 보상을, 예약되었지만 사용되지 않은 자원의 양을 의미하는 Wasted_Resources는 클수록 부정적 보상(패널티)을 제공한다.

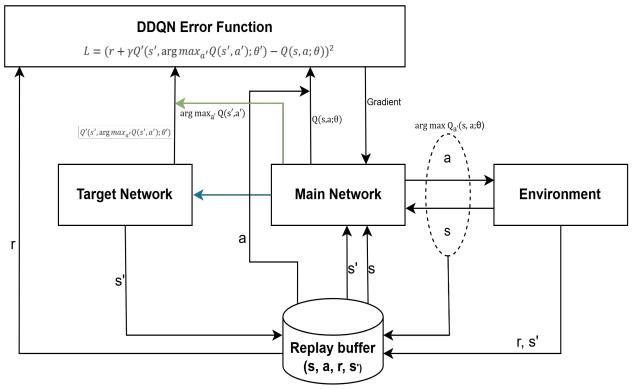


그림 2 DDQN

DDQN 에이전트는 '상태'를 관찰하고 '행동'을 수행한 뒤, 그 결과로 '보상'을 받으며 학습을 진행한다. 이 과정을 반복하여, Main Network와 Target Network 간의 오차(Loss)를 최소화하는 최적의 정책(Policy), 즉 각 상황에 맞는 최적의 RC 값을 찾는 규칙을 학습하게 된다. 이를 통해 데이터 크기 변화에 유연하게 대응하고 리소스 활용률을 극대화할 수 있다.

2.3 실험 및 결과

본 제안 기법의 실험은 NS-3를 이용하여 PDR(Packet Delivery Ratio)와 리소스 활용률에 대해 기존 고정 RC와 비교 분석 하였으며 실험 환경은 아래의 표1과 같다.

파라미터(Parameter)	값(Value)
차량 수	10 ~ 100(10대씩 증가)
시뮬레이션 시간	300s
주파수 대역	5.9GHz
대역폭	10MHz
SRRI	100ms
전송 전력	23dBm
패킷 크기	200 ~ 500 Bytes

표 1 NS-3 시뮬레이션 파라미터

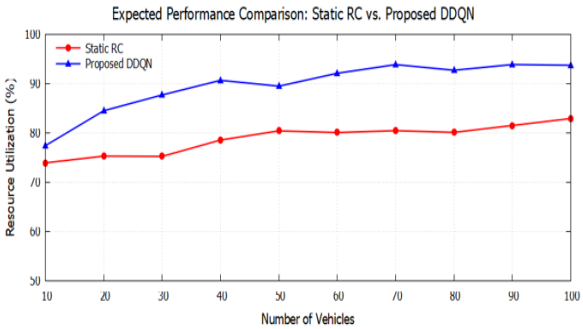


그림 3 리소스 활용률

그림3은 제안 기법이 차량 수가 증가해도, 제안하는 DDQN 기법은 85% 이상의 높은 리소스 활용률을 꾸준히 유지한다. 반면, 고정된 RC를 사용하는 기존 방식은 약 80% 수준에서 정체된다. 이는 DDQN 에이전트가 변화하는 트래픽 수요에 맞춰 불필요한 자원 예약을 효과적으로 줄여 자원 낭비를 최소화했음을 의미한다.

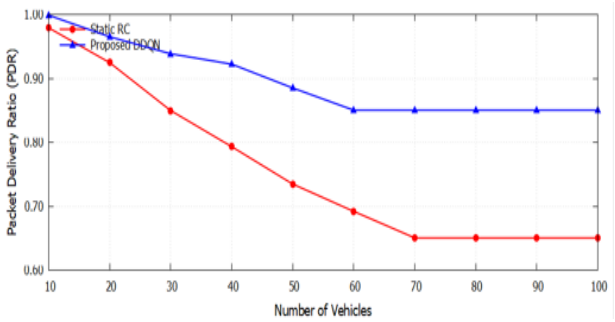


그림 4 PDR(Packet Delivery ratio)

그림4는 차량 수가 늘어날수록 두 기법 간의 PDR 성능 차이는 더욱 두드러진다. 기존 방식의 PDR은 65%까지 급격히 하락하는 반면, DDQN 기법은 85% 수준을 안정적으로 방어한다. 이는 DDQN이 유연한 자원 관리를 통해 패킷 충돌을 줄이고 통신 안정성을 크게 향상시켰음을 보여준다.

Ⅲ. 결론

본 논문에서는 NR-V2X V2V 통신의 모드 2 환경에서 고정된 RC와 SRRI 값으로 인해 발생하는 리소스 낭비 및 활용률 저하 문제를 지적했다. 이에 대한 해결책으로, 강화학습 기반의 DDQN 알고리즘을 적용하여 주변 환경과 트래픽 수요에 맞춰 RC 값을 동적으로 조절하는 기법을 제안했다. 상태, 행동, 보상을 명확히 정의함으로써, 제안하는 모델은 통신 환경 변화에 실시간으로 적응하여 자원 효율성을 높일 수 있다.

추후 연구로는 제안된 기법을 NS-3와 같은 네트워크 시뮬레이터에 구현하여 성능을 검증하고, 실제 차량 통신 환경과 유사한 시나리오에서 PDR, 지연 시간, 리소스 활용률 등의 지표를 종합적으로 평가할 계획이다.

ACKNOWLEDGMENT

This work was supported in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education under Grant NRF-RS-2018-NR031059 and this work was supported by the Commercialization Promotion Agency for R&D Outcomes(COMPA) grant funded by the Korean Government(Ministery of Science and ICT). (RS-2023-00304695)

참 고 문 헌

[1] Technical Specification Group Radio Access Network, “Study onNR Vehicle-to-Everything (V2X)” 3GPP, Tech. Rep. TR 38.885,Mar. 2019.

[2] S. Jung, H.-R. Cheon, and J.-H. Kim, “Reducing consecutivecollisions in sensing based semi persistent scheduling forcellularV2X,”in Proc. IEEE Vehicular Technology Conference,Honolulu, Hawaii, USA, September 2019, pp. 1-5.

[3] Hado van Hasselt, “Double Q-learning”, Advaces in Neural information Processing Systems 23, 201

이기종 가속기 기반 엣지 AI 시스템의 시간-공간 하이브리드 스케줄러 설계

김한결, *안재훈, 김영환
한국전자기술연구원
taylor4359@keti.re.kr, *corehun@keti.re.kr, yhkim93@keti.re.kr

Design of Time-Space Hybrid Scheduler for Edge AI Systems based on Heterogeneous Accelerators

Hangyeol Kim, Jaehoon An*, Younghwan Kim
Korea Electronics Technology Institute

요 약

엣지 환경에서는 제한된 하드웨어 자원 위에서 학습과 추론을 동시에 수행해야 하며, 이로 인한 자원 경쟁과 간섭이 추론 지연(SLA) 악화 및 전체 활용도 저하로 이어진다. 본 논문은 GPU와 NPU로 구성된 이기종 가속기 기반 엣지 시스템을 대상으로, **시간 분할(Time-Slicing)**과 **공간 분할(Spatial Partitioning)**을 결합한 하이브리드 스케줄러 아키텍처를 설계한다. 제안 스케줄러는 (1) 워크로드 관측(요청 도착률, 대기열 길이, p95 지연 등)과 모델 특성(지연 민감도, 메모리/연산 요구)을 실시간 수집하고, (2) 공간 분할 단계에서 GPU MIG 인스턴스 크기/개수와 NPU 오프로딩을 통해 격리를 보장하며, (3) 각 파티션 내부에서는 우선순위 선점형 시간 분할로 추론에는 짧은 타임슬라이스와 높은 우선순위를, 학습에는 긴 타임슬라이스와 백그라운드 실행을 부여한다. 또한 파티션 재구성의 오버헤드를 줄이기 위해 부하 추세 기반의 완만한 파티션 적응 정책을 적용하고, 모델 크기·정확도 요구에 따라 GPU/NPU 선택을 동적으로 병행한다.

Ⅰ. 서 론

엣지 컴퓨팅 환경에서 제한된 자원으로 AI 모델의 동시 학습(Training)과 추론(Inference)을 수행해야 하는 요구가 증가하고 있다. 그러나 하나의 장치에서 학습과 추론 작업을 병행하면 자원 경쟁으로 인해 추론 지연이 커지고 학습 속도가 느려지는 문제가 발생한다. 기존에는 학습 작업과 추론 서비스를 분리된 장치나 별도 단계로 운영하는 경우가 많았지만, 이러한 격리 운용은 GPU 유휴 시간 증가 등 자원 비효율과 새로운 데이터에 대한 모델 적응 지연을 초래한다 [1]. 예를 들어 추론 전용으로 GPU를 독점하면 지연은 낮아지지만, 전체 추론 처리량이 낮고, 반대로 학습 작업에 GPU를 전적으로 할당하면 추론 요청을 처리할 수 없어 서비스 공백이 생긴다. 따라서 엣지 단말의 이기종 가속기(예: GPU와 NPU)를 효율적으로 활용하여 동시에 학습과 추론을 수행하면서도 추론 지연은 최소화하고 자원 활용도를 높이는 스케줄링 기법이 필요하다.

본 논문에서는 시간분할(Time-Slicing)과 공간 분할(Spatial Partitioning) 기법을 병행한 하이브리드 스케줄러 아키텍처를 설계한다. 시간 분할은 하나의 가속기를 여러 작업이 시간상으로 공유하도록 하는 것으로 높은 활용도를 가능케 하나, 작업 간 간섭으로 지연 변동이 커질 수 있다. [2]. 공간 분할은 하드웨어 자원을 물리적으로 분할하여 작업별로 할당함으로써 격리와 예측 가능성을 제공하지만, 자원 단편화로 인한 비효율이 우려된다 [3]. NVIDIA Ampere GPU의 MIG(Multi-Instance GPU) 기술은 하나의 GPU를 하드웨어적으로 여러 인스턴스로 나누어 각 인스턴스에 전용 메모리와 연산 코어를 할당함으로써 워크로드 간 간섭을 줄이고 지연을 낮추는 대표적 공간 분할 기법이다 [4]. 한편 MIG로 분할된 각 인스턴스 내부에서도 다중 작업을 시분할로 공유할 수 있어 사용자 수용에 따라 유연하게 처리할 수 있다. [5]. 이러한 두 접근법을 결합한 스케줄링이 혼합된 작업 환경에서 성능과 효율의 균형을 달성할 것으로 기대된다.

본 연구에서 제안하는 스케줄러는 엣지 AI 디바이스의 GPU와 NPU 자원을 통합 관리하며, 작업 유형과 실시간 부하를 모니터링하여 공간적 자원 할당과 시간 슬라이스 조정을 동적으로 수행한다. 이를 통해 지연 민감한 추론 작업에는 전용 자원과 짧은 시간 조각을 우선 할당하고, 연산 집약적인 학습 작업은 잔여 자원을 활용하여 백그라운드로 진행한다. 이하에서 시스템 설계 내용을 자세히 설명하고 아키텍처 개요와 워크로드 유형별 자원 분할 정책에 대해서 설명한다.

Ⅱ. 본론

1. 관련 연구

엣지 AI 시스템의 이기종 가속기 스케줄링에 관한 최근 연구들이 활발히 보고되고 있다. Corun 프레임워크는 하나의 GPU에서 다수의 추론 요청과 지속적 모델 재학습을 동시 처리하는 기법으로, 사전 오프라인 프로파일링을 통해 메모리 초과나 지연 급증 없이 최적의 동시 실행 개수를 찾아내는 스케줄링 방법을 제안하였다. 그 결과 Corun은 단일 GPU로 여러 작업을 병행하여 추론 처리량을 크게 높이면서도 지연 증가는 최소화하였으며, 별도 GPU를 각각 사용할 때 대비 비용 효율적임을 보였다.

LeMix 시스템은 대규모 언어 모델(LLM)의 서빙과 훈련 작업을 멀티 GPU 환경에서 동일 노드에 혼재시키는 방안을 제시하였다. LeMix는 오프라인 프로파일러와 실행 예측 기법, 계층적 자원 할당 및 런타임 메모리-인지 스케줄러를 통합하여 작업들의 상호 간섭을 완화하면서 동적으로 자원을 할당한다. 이를 통해 사용률이 낮은 시기에 노드 수를 줄이고, 부하가 높을 때도 추론 응답성을 크게 저하시키지 않으면서 GPU 자원 활용

도를 22% 이상 향상 시켰다고 보고 하였다.

2. 시스템 아키텍처 설계

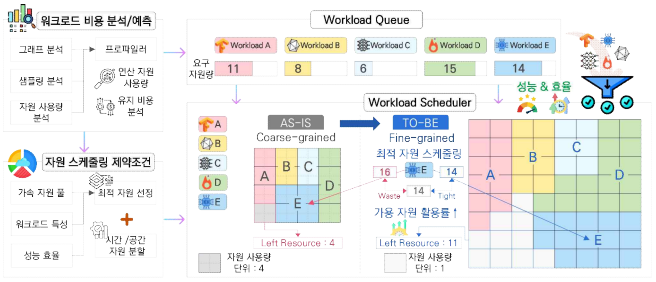


그림 1. 이종 AI 가속기 환경 최적 자원 배치 스케줄러 설계

본 논문에서 제안하는 하이브리드 시간 - 공간 자원 분할 스케줄러의 구조. 입력 대기열에서 추론 요청과 학습 작업이 도착하면, 모니터링 모듈이 작업 유형과 성능 지표(요청 빈도, 마감 시간 등)를 추적한다. 스케줄링 정책 모듈은 이를 바탕으로 자원 할당 방식을 결정하는데, 한편으로는 공간 분할 관점에서 GPU를 여러 파티션으로 나누거나(NVIDIA MIG 등의 기술 활용) 작업을 NPU로 오프로딩할지를 결정하고, 다른 한편으로는 시간 분할 관점에서 각 작업의 실행 우선순위와 타임슬라이스 길이를 조정한다. 최종적으로 하드웨어 분할 제어기가 이러한 결정 사항을 하드웨어 레벨에서 적용하여, GPU의 MIG 인스턴스를 구성하거나 작업을 해당 자원으로 디스패치(dispatch)한다.

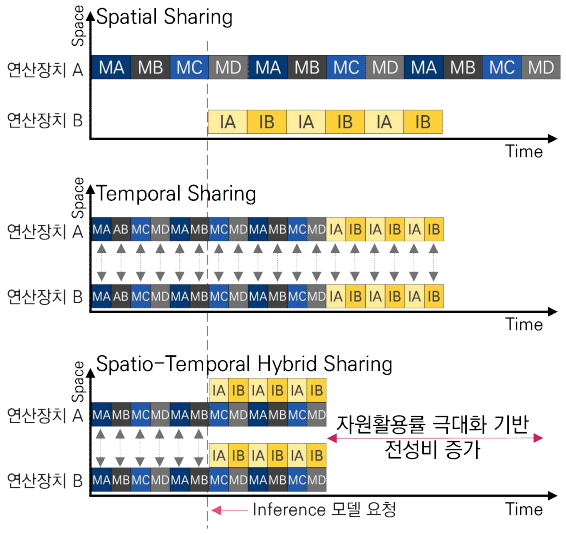


그림 2. 시간/공간 분할 스케줄러 설계

스케줄러의 공간 분할 구성은 현재 워크로드 구성에 따라 동적으로 변경된다. 예를 들어 지연에 민감한 추론 요청이 지속적으로 발생하면 GPU 자원의 일정 부분을 소형 MIG 인스턴스로 확보하거나 NPU 전용으로 할당하여 해당 인스턴스에서는 오직 추론 작업만 실행하도록 격리한다. 반대로 추론 부하가 적을 때는 더 큰 비중의 GPU 자원을 학습 작업에 할당한다. 이러한 적응형 GPU 파티셔닝을 통해 불필요한 리소스 재할당(overhead)을 줄이면서도 다양한 작업에 대응할 수 있다. 단, MIG 파티션 크기를 조정하는 작업은 잦은 변경 시 오버헤드가 발생할 수 있으므로, 일정 기간 모니터링한 부하 추세에 기반하여 완만하게 조정된다. NPU 역시 전용 메모리 및 연산 엔진을 가지고 있으므로, 지원하는 모델의 추론 작업은 가능하면 NPU로 보내 GPU 부하를 줄인다. 다만 NPU는 온보드 메모리 용량이 작아 대형 모델을 수용하기 어렵고 복잡한 DNN 실행 시 정확도 저하를 초래할 수 있다는 한계가 있다. 그러므로 모델 크기나 정확도 요구사항에 따라 GPU와 NPU를 선택적 병행 활용하여 성능과 정확도의 균형을 잡는다.

한편 시간 분할 스케줄링 측면에서는, 각 자원 파티션 내에서 다중 작업이 존재할 경우 선점형(priority preemptive) 스케줄링을 적용한다. 특히 실시간 추론 작업에는 높은 우선순위를 부여하여 실행 중인 학습 커널보다 우선 배치하고, 필요한 경우 학습 작업을 일시 중단하여라도 즉각 추론을 처리한다. 또한 추론 작업에는 짧은 타임슬라이스를 적용해 응답 지연을 낮추고, 학습 작업에는 비교적 긴 타임슬라이스를 부여하여 문맥 전환 overhead를 줄이면서 연산 효율을 높인다. 이러한 QoS-인식 시간조율을 통해 추론 요청의 응답시간 SLA를 보장하면서도 남는 시간에는 학습을 진행하여 자원을 최대한 활용한다. 스케줄러는 모니터링을 통해 추론 대기열 길이나 마감 시간 준수 여부를 계속 점검하며, 상황에 따라 학습 작업의 실행 슬라이스를 더 잘게 쪼개거나 일시 정지하는 등 동적 조정을 수행한다.

Ⅲ. 결론

본 논문에서는 GPU와 NPU로 구성된 엣지 AI 장치에서 동시 학습 및 추론을 효율적으로 지원하기 위한 하이브리드 자원 스케줄러 설계를 제안하였다. 제안된 스케줄러는 작업 특성 모니터링을 통해 공간적 자원 분할(MIG 기반 GPU 파티셔닝, NPU 할당)과 시간적 분할(우선순위 및 슬라이스 조정)을 동적으로 병행함으로써, 제한된 자원 환경에서도 추론 서비스의 지연을 보장하면서 남은 성능으로 학습 작업을 진행할 수 있다. 향후에는 실제 프로토타입 구현을 통해 제안 기법의 유효성 검증을 진행할 예정이다. 또한 FPGA, DSP 등 더 다양한 종류의 가속기에 대한 지원이나, 워크로드 예측을 통한 선제적 스케줄링 최적화 기법, 그리고 분산 엣지 환경에서의 자원 조율 등으로 연구를 확장할 계획이다.

ACKNOWLEDGMENT

This work was supported by the Institute of Information & Communications Technology Planning & Evaluation(IITP) grant funded by the Korea government(MSIT) (No.RS-2025-25441574, Development of an On-Device Integrated Edge AI Server System)

참 고 문 헌

[1] NVIDIA, “NVIDIA Multi-Instance GPU User Guide (RN-08625-v2.0),” 2024.

[2] T. Wang, S. Li, B. Li, Y. Dai, A. Li, G. Yuan, Y. Ding, and Y. Zhang, “Improving GPU Multi-Tenancy Through Dynamic Multi-Instance GPU Reconfiguration,” arXiv:2407.13126, 2024..

[3] Y. Li, Z. Li, Y. Zhu, and C. Liu, “LeMix: Unified Scheduling for LLM Training and Inference on Multi-GPU Systems,” RTSS 2025.

[4] Y.-M. Tang, W.-F. Sun, H.-T. Ting, M.-H. Chen, et al., “PCIe Bandwidth-Aware Scheduling for Multi-Instance GPUs,” in Proc. HPCASIA ’25, ACM, 2025.

[5] S. Zhang, A. Xu, Q. Chen, H. Zhao, W. Cui, Z. Wang, Y. Li, L. Xiao, and M. Guo, “Efficient Performance-Aware GPU Sharing with Compatibility and Isolation through Kernel Space Interception,” in USENIX ATC 2025, 2025.

EEG 신호를 활용한 그룹 대조 학습 기반 신경적응형 CBT 반응 생성

이진권[1 2 *], 김효빈[1 *], 황세미[2 3*], 김지윤[2 *], 송명주[2 *], 오하영[1 †]
[1] 성균관대학교, [2] 모두의 연구소, [3] 연세대학교 원주연세의료원
rnjs8316@skku.edu, hyobinkim@gmail.com, semi.hwang11@gmail.com, jjyoooon.work@gmail.com,
myungjoo.song12@gmail.com, hyoh79@skku.edu

EEG-Based Group Contrastive Learning Framework for Neuroadaptive CBT Response Generation

Jinkwon Lee[1 2 *], Hyobin Kim[1 *], Semi Hwang[2 *], Jiyoon Kim[2 *], Myungjoo Song[2 *], Hayoung Oh[1 †]
[1] Sungkyunkwan University, [2] Modulabs, [3] Yonsei University Wonju Health System

요 약

본 연구는 저채널(4채널) 소비자용 EEG를 활용해 실시간 뇌파-텍스트 변환으로 치료 전략에 부합하는 CBT 응답을 생성하는 프레임워크를 제안한다. 핵심은 CBT 전략별 대화 맥락 클러스터링과 그룹 대조학습을 통한 EEG-텍스트 정렬, 그리고 경량 추론으로 1초 미만 응답 지연을 달성하는 것이다. 28명으로 구성된 EEG-CBT 데이터셋에서 제안 모델은 BLEU-4, ROUGE-L, BERTScore 및 전문가 치료 적절성 평가에서 기존 방법 대비 개선을 보였고, 지연시간은 1초 미만으로 단축되었다.

Ⅰ. 서 론

우울증과 불안은 21세기 가장 시급한 공중 보건 문제 중 하나로, 전 세계 수억 명의 사람들에게 영향을 미치며 의료 시스템에 큰 부담을 주고 있다[1]. 인지행동치료(CBT)는 부적응적인 사고를 인지 재구성 및 공감적 반영을 통해 교정하는 표준 치료법으로 인정받고 있다. 디지털 CBT 도구와 챗봇은 치료 접근성을 높이고 사회적 낙인을 줄이는 데 기여했지만, 실시간 반응이 느리고 개인화 수준이 낮다는 한계를 가진다. 이러한 시스템들은 주로 텍스트 입력과 정해진 템플릿 기반의 답변에 의존하여, 사용자의 순간적인 감정 변화나 개별적인 요구를 포착하지 못한다.

이러한 한계를 극복하기 위해, 본 연구는 언어적 표현 이전에 인지-정서 상태를 실시간으로 객관적으로 측정할 수 있는 뇌파(EEG) 기술을 활용한다. 기존 EEG-텍스트 변환 연구들이 기술적 가능성을 보여주었으나, 대부분 고밀도 채널 EEG 장비를 요구하며 CBT라는 특정 치료 영역에 최적화되지 않았다[2].

본 논문은 CBT에 특화된 EEG-텍스트 변환 프레임워크를 제안하며, 다음과 같은 세 가지 핵심 혁신을 포함한다. (1) EEG 특징과 CBT 전략(인지 재구성, 공감적 반영 등)을 정렬하는 클러스터 기반 그룹 대조 학습, (2) 저채널 소비자용 EEG 장치에 최적화된 표현 증류 기법, (3) 상호작용형 치료를 위한 실시간 추론 능력이다. 이 접근법은 신경생리학적 통찰과 치료 원칙을 결합하여, 확장 가능하고 개인화된 정신 건강 치료를 가능하게 한다.

Ⅱ. 본론

EEG-CBT 모델 아키텍처

제안하는 EEG-CBT 모델은 EEG 인코딩, 이산 코덱 생성, CBT 맥락 클러스터링, 그룹 대조 학습, 그리고 LLM 디코더의 5단계 순차 처리 과정으로 구성된다. 각 단계는 CBT 도메인의 특성을 반영하고 저채널 EEG 환경에서의 최적화를 고려하여 설계되었다.

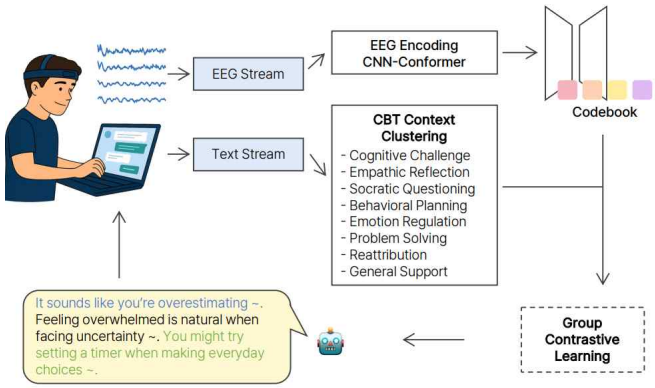


그림1 본 프레임워크는 EEG 스트림과 텍스트 스트림을 동시에 처리하여 신경적응형 CBT 반응을 생성한다. EEG 신호는 CNN-Conformer 기반 인코더를 통해 시간·공간적 패턴을 추출한 뒤, 벡터 양자화를 거쳐 코드북 표현으로 변환된다. 텍스트 입력은 Sentence-BERT 임베딩을 기반으로 CBT 전략(인지적 도전, 공감적 반영, 소크라테스식 질문, 행동 계획, 정서 조절 등)별 맥락 클러스터링 과정을 거친다. 이후 그룹 대조 학습을 통해 EEG 표현과 CBT 전략 클러스터 간 정렬을 수행하며, 디코더는 이를 바탕으로 치료적으로 적절하고 자연스러운 CBT 반응을 실시간으로 생성한다.

가. EEG 인코딩 및 이산 코덱 생성

입력된 EEG 신호는 CNN-Conformer 아키텍처 기반의 인코딩 모듈을 통해 시간적, 공간적 패턴을 효과적으로 포착한다. 추출된 연속적인 EEG 표현은 VQ-VAE(Vector Quantized Variational Autoencoders)를 사용하여 언어 모델과 호환 가능한 이산적인 코덱으로 변환된다[2].

나. CBT 맥락 클러스터링

본 연구의 핵심 혁신 중 하나인 CBT 맥락 클러스터링은 치료 전략에 따라 대화 내용을 의미적으로 그룹화하는 과정이다. Sentence-BERT 임베딩을 사용하여 각 CBT 대화 발화를 벡터로 변환한 후, K-평균 클러스터링 알고리즘을 적용하여 '인지적 도전', '공감적 반영', '소크라테스식 질문', '행동 계획' 등 8개의 주요 CBT 기법으로 군집화한다[3,4].

다. 그룹 대조 학습 메커니즘

그룹 대조 학습은 기존의 쌍(pair) 기반 대조 학습을 확장하여, 치료 전략 클러스터 간의 구조적 관계를 학습한다. '강한 긍정 쌍(동일 세션, 동일 클러스터)', '약한 긍정 쌍(다른 세션, 동일 클러스터)', '부정 쌍(치료적으로 대조적인 클러스터)' 등 계층적 쌍 구조를 정의하고, 가중치가 적용된 InfoNCE 손실 함수를 사용하여 클러스터 간의 치료적 유사도에 기반한 학습을 수행한다. 이를 통해 모델은 단순한 의미적 유사성을 넘어 치료적 맥락에 맞는 미묘한 차이를 학습하게 된다.

라. LLM 디코더

사전 훈련된 BART 모델 기반의 디코더는 이산화된 EEG 코덱을 최종적인 CBT 반응 텍스트로 변환한다. 대규모 CBT 대화 코퍼스를 사용한 도메인 적응 훈련을 통해, 디코더는 치료적으로 일관되고 언어적으로 자연스러운 문장을 생성하도록 최적화된다.

Ⅲ. 실험 및 결과

28명의 참가자로부터 수집된 약 21시간 분량의 EEG 신호 및 CBT 대화 데이터로 구성된 자체 데이터셋을 사용하여 모델 성능을 평가했다. 텍스트 생성 품질(BLEU-4, ROUGE-L), CBT 적절성(전문가 평점), 실시간 성능(응답 지연 시간)을 종합적으로 측정했다.

실험 결과, 제안 모델은 대표적인 EEG-텍스트 변환 모델인 DeWave 및 일반적인 대조 학습을 적용한 SimCLR-EEG 모델과 비교했을 때 모든 평가지표에서 월등한 성능을 보였다. 특히 BLEU-4 점수는 0.63으로 DeWave(0.28) 대비 68% 이상 높았으며, 응답 지연 시간은 890ms로 실시간 상호작용 기준인 1초 이내를 만족시켰다.

또한, 구조화된 EEG-텍스트 쌍의 중요성을 검증하기 위해 EEG-텍스트 쌍을 무작위로 섞거나, 클러스터 레이블을 뒤섞는 등의 조건과 비교 분석한 결과, 본 연구의 구조화된 접근법이 언어적 일관성과 치료적 적절성 모두에서 월등한 성능을 유지함을 확인했다. 이는 의미 있는 EEG-텍스트 정렬이 치료 효과에 결정적임을 시사한다.

Ⅳ. 결론

본 연구는 텍스트 기반 디지털 치료의 한계를 극복하는 최초의 실시간 EEG 적응형 CBT 시스템을 제안했다. 주요 기여는 (1) 뇌-행동 분석을 위한 CBT 특화 EEG 데이터셋 구축, (2) 치료 전략의 의미론적 모델링을 위한 클러스터 기반 대조 학습 개발, (3) 실용적인 배포를 위한 저채널 EEG 환경에서의 성능 최적화이다. 초기 실험 결과는 기존 시스템 대비 향상된 치료 결과와 사용자 참여를 보여주었다.

다만, 28명의 참가자 데이터는 개념 증명 수준이며, 장기적인 효과 검증과 다양한 인구 집단에 대한 일반화는 과제로 남아있다. 향후 연구는 심박변이도(HRV), 안면 및 음성 분석 등 다중 모드 통합, 대규모 무작위 임상 시험을 통한 임상적 검증, 그리고 연합 학습과 동형 암호화를 결합한 고급 개인정보 보호 프레임워크 개발에 초점을 맞춰야 한다. 또한, 급성 정신 건강 위기 상황을 위한 실시간 위기 탐지 및 개입 시스템 개발 역시 중요한 다음 단계가 될 것이다.

ACKNOWLEDGMENT

이 연구는 과학기술정보통신부가 주관하고 정보통신기획평가원(IITP)이 관리하는 메타버스융합대학원 지원사업 (IITP-2025-RS-2023-00254129)과 글로벌 석학 초청사업 (RS-2024-00459638), 문화체육관광부가 주관하고 한국콘텐츠진흥원이 지원하는 문화체육관광 연구개발사업(스포츠·관광 R&D 프로그램) — 「청소년 정신건강(심리·행동 조절) 관리를 위한 게임기반 디지털치료제 기술 개발」 (RS-2024-00344893), 정보통신기획평가원(IITP)이 지원하는 AI 스타 펠로우십 지원사업(성균관대학교) (RS-2025-25442569) 및 체화형 비전 AI 멀티에이전트 기반 차세대 인간지향 범용 인공지능(AGI) 기술 개발 사업 (RS-2025-25443884), 그리고 성균관대학교 퓁리스 AI 성남연구센터(K-하이테크 플랫폼)의 지원을 받아 수행되었습니다.

참 고 문 헌

[1] Liu, C.; and Miao, W. 2022. The role of employee psychological stress assessment in reducing human resource turnover in enterprises. Frontiers in Psychology, 13.

[2] Duan, Y.; Zhou, J.; Wang, Z.; Wang, Y.; and Lin, C. 2023. DeWave: Discrete EEG Waves Encoding for Brain Dynamics to Text Translation. arXiv (Cornell University).

[3] Fenn, K.; and Byrne, M. 2013. The key principles of cognitive behavioural therapy. InnovAiT Education and inspiration for general practice, 6: 579-585.

[4] Cully, J. A.; and Teten, A. L. 2008. A therapist's guide to brief cognitive behavioral therapy.

[5] Attridge, M., Morfitt, R. C., Roseborough, D., & Jones, E. R. (2020). *Internet-Based Cognitive-Behavioral Therapy for College Students With Anxiety, Depression, Social Anxiety, or Insomnia*. JMIR Formative Research, 4.

[6] Baradari, D., Kosmyna, N., Petrov, O., Kaplun, R., & Maes, P. (2025). *NeuroChat: A Neuroadaptive AI Chatbot for Customizing Learning Experiences*.

MobileNetV2 기반의 딥러닝 모델을 활용한 넙치 다중 질병 분류 연구

이채령, 이명철*

건양대학교 의공학부, *한국전자통신연구원

chaeryounginji@gmail.com, *mclee@etri.re.kr

Multi-Disease Classification of Flatfish Using a MobileNetV2-Based Deep Learning Model

Chaeryoung Lee, Myungcheol Lee*

Konyang Univ., *Electronics and Telecommunications Research Institute

요 약

본 논문은 국내 양식 산업에 심각한 경제적 손실을 초래하는 넙치 질병을 신속하고 정확하게 진단하기 위하여 딥러닝 기술을 활용한다. 기존 연구에서 사용된 VGG16과 같은 모델은 과도한 연산량과 상대적으로 낮은 정확도로 인해 모바일 환경에서의 실시간 진단에 한계가 있다. 이에 따라 MobileNetV2 기반의 경량 딥러닝 모델을 제안한다. 넙치 질병 이미지 데이터셋에 MobileNetV2를 미세 조정(fine-tuning)하여 다중 질병 분류 성능을 최적화하였다. 또한, 기존 VGG16을 경량화한 TFLite 모델과 비교 분석을 통해 MobileNetV2 모델의 성능과 모바일 적용 효율성을 검증하였다. 제안 모델은 분류 성능 측면에서 기존 모델 대비 약 10% 증가한 87.53%의 F1 score를 달성했으며, 모델 크기 약 4.8배 감소 및 추론 속도 6.56ms로 크게 향상하였다. 이는 넙치 질병 진단에 있어 정확성과 실시간성을 동시에 확보하여, 양식 산업의 효율적인 질병 관리에 기여할 수 있다.

Ⅰ. 서 론

국내 양식 산업에서 넙치는 가장 중요한 양식어종 중 하나로, 국내 생산량에서 높은 비중을 차지하며 산업 전반에서 핵심적인 위치를 확보하고 있다[1]. 그러나 넙치 양식에서는 다양한 감염성 질병으로 인한 대규모 폐사가 발생해 양식 산업의 경제적 손실로 이어질 수 있으며[2], 이러한 질병은 다중 감염 형태로 빈번하게 발생하여 신속하고 정확한 다중 질병 진단이 중요하다.

기존 연구에서는 VGG16 기반의 딥러닝 모델을 활용하여 넙치 질병 이미지를 분류하였으나[3], VGG16은 구조적으로 연산량과 모델 크기가 커서 모바일 환경에서의 실시간 적용에는 적합하지 않다. 이에 본 연구에서는 MobileNetV2 기반 다중 질병 분류 모델을 설계하고, TensorFlow Lite(TFLite)로 변환하여 모바일 환경에서도 적용할 수 있는 모델을 제안하며, 기존 VGG16 기반 모델과의 성능을 비교하여 경량 모델의 현장 적용 가능성을 검증하였다.

Ⅱ. 본 론

가. 데이터셋 구성

본 연구는 국내 양식 넙치의 주요 질병 8종을 대상으로 경량화된 딥러닝 기반 다중 질병 분류 모델을 설계하였다. 데이터는 AI-Hub에서 제공하는 넙치 질병 이미지로, 그림 1과 같이 바이러스성출혈성패혈증, 림포시스티스병, 여윔병, 스쿠티카병, 연쇄구균증, 비브리오병, 에드워드병, 기타로 총 8개 질병 클래스로 구성된다. 이때 비브리오병은 5개의 세부 질병(h, a, ang, s, pd)을 하나로 통합하여 단일 클래스로 정의하였다.

원본 이미지는 약 6,000×4,000 크기로 제공되었으며, 학습을 위해 224×224로 리사이즈하였다. 전체 데이터는 총 31,482장으로 train, valid, test 데이터를 8:1:1 비율로 나누어 사용하였으며, 데이터 불균형 문제를 완화하기 위해 소수 클래스는 오버샘플링과 데이터 증강을 적용하고, 일부 클래스에 대해서만 언더샘플링하였다.

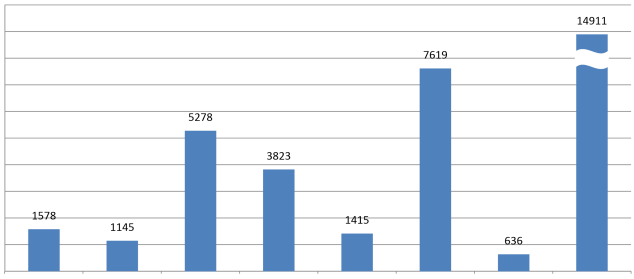


그림 1. 넙치 질병 클래스별 이미지 분포

나. 모델 구성

VGG16 기반 모델[4]과 MobileNetV2 기반 모델[5]을 TFLite 형식으로 사용하여 모바일 환경에서의 적용 가능성을 검증하였다.

먼저 사전에 학습된 VGG16 기반 모델은 Conv Block 1-5로 구성된 특징 추출부와 Flatten 이후 fully connected layers(1024-512-7)로 이어지는 분류기로 이루어진다.

반면 MobileNetV2 기반 모델은 그림 2와 같이 특징 추출부의 block 14~16만 학습하고, 나머지 블록은 동결(freeze)하였다. 이는 초기 블록의 사전 학습 가중치를 유지하여 기본적인 특징 표현을 보존하면서, 상위 블록만을 미세 조정하여 질병 분류 성능을 향상하고 과적합을 방지하기 위함이다. Global Average Pooling을 통해 특성을 벡터화한 뒤, fully connected layers(1024-512-256-8)를 거쳐 최종 출력을 도출하였으며, 각 층에는 Dropout과 L2 정규화를 적용하여 경량화된 구조에서도 안정적인 학습과 과적합 방지를 도모하였다.

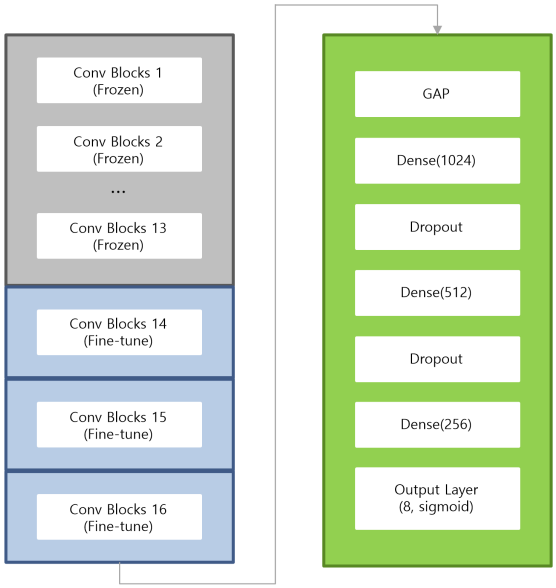


그림 2. MobileNetV2 기반 넙치 질병 분류모델 구조

다. 손실함수 및 평가지표

본 연구의 넙치 질병 분류 문제는 한 이미지에 여러 질병이 동시에 나타날 수 있는 다중 레이블 분류 문제이다. 이에 따라 출력층에는 클래스별 확률을 독립적으로 산출하기 위해 시그모이드(sigmoid) 활성화 함수를 적용하였다. 학습 단계에서 손실함수는 Asymmetric Loss(ASL)를 사용하였다. ASL은 희소 클래스의 학습 신호를 강화하고 hard negative의 영향을 완화하여 초기 수렴을 촉진하는 손실함수로, 본 연구에서는 구현 시 프레임워크 기본 설정에 따라 클래스별 손실을 평균하여 적용하였다.

$$ASL_{(y,p)} = -\sum_{c=1}^{C}\left[y_c(1-p_c)^{\gamma^+}\log(p_c) + (1-y_c)p_c^{\gamma^-}\log(1-p_c)\right]$$

여기서 y_c는 c번째 클래스의 실제 레이블이며, p_c는 예측 확률, γ^+와 γ^-는 각각 양성 및 음성 표본에 대한 조정 파라미터를 의미한다.

검증 단계에서는 F1 score를 최대화하기 위하여 validation 데이터에 기반한 클래스별 임계값(threshold) 탐색을 수행하였다. 추가적으로 Top-K 방식을 적용하여 각 입력의 예측 확률 벡터에서 상위 K개 클래스를 우선적으로 선택함으로써, 확률 분포가 여러 클래스에 분산될 때 발생할 수 있는 과도한 양성 판정을 완화하고 보다 안정적인 평가를 가능하게 하였다.

모델 성능 평가는 F1-score, Precision, Recall, Specificity, AUPRC를 활용하였다[6]. 이때 F1-score는 모든 클래스에 대한 참·거짓 예측을 표본 단위로 합산하는 micro 방식을 사용하였다. 또한, 모델의 실용성을 평가하기 위해 모델 크기(MB)와 이미지 한 장당 처리 시간(ms)을 측정하였다.

라. 실험 결과

본 비교는 TFLite 변환 모델을 기준으로 수행하였다. VGG16은 과거 학습된 모델로 기타 질병 클래스가 포함되지 않아, 본 데이터셋에서도 이를 제외한 7개 클래스(입력 112×112)로 평가하였고, MobileNetV2는 기타 질병을 포함한 8개 클래스(입력 224×224)로 학습·평가하였다. 두 모델 모두 비교의 일관성을 확보하기 위하여 임계값 0.8을 적용하였으며, 사전 검토에서도 해당 설정이 가장 안정적이고 우수한 성능을 보였다. 최종 결과는 표 1에 제시하였다. VGG16 TFLite의 F1-score는 77.49%였던 반면, MobileNetV2 TFLite는 87.53%로 VGG16 대비 약 10% 향상되었으며, 모델 크기도 76.16MB에서 15.98MB로 약 4.8배 감소하였다. 또한 TFLite 모델 기준 CPU 추론시간도 크게 단축되었다. 이러한 결과는 MobileNetV2가 정확도와 효율성 모두에서 우수하여 모바일 환경에서의 실시간 진단에 적합함을 보여준다.

모든 학습은 NVIDIA A40 GPU 4장(각 48GB 메모리)을 탑재한 서버에서 TensorFlow 2.13.1 기반, Python 3.8.10 환경에서 수행하였으며, CPU는 Intel Xeon Gold 6448Y(64코어, 128스레드)를 사용하였다.

표1. 실험 결과 비교

	VGG16	VGG16 TFLite	MobileNetV2	MobileNetV2 TFLite
F1-score	75.75	77.49	87.54	87.53
Precision	81.13	83.74	86.73	86.71
Recall	71.04	72.10	88.36	88.32
Specificity	95.28	95.90	97.29	97.18
AUPRC	70.55	72.77	94.82	94.80
Model Size	228.98	76.16	19.45	15.98
Time(GPU)	1.72	-	2.13	-
Time(CPU)	54.92	103.23	12.60	6.56

Ⅲ. 결 론

본 연구에서는 넙치 질병 8종을 분류할 수 있는 MobileNetV2 기반 경량 모델을 제안하고, TFLite 변환을 통해 모바일 환경에서의 실시간 적용 가능성을 검증하였다. 제안 모델은 기존 VGG16 대비 모델 크기와 연산량을 크게 줄이면서도 F1-score와 AUPRC 등 정확도 지표가 향상되었으며, 추론 시간 역시 6.56ms로 단축되었다. 향후 실제 양식장 환경에서 모바일 앱과 연계한 실시간 검증을 통해 제안 모델의 실효성을 확인하고 국내 양식 산업의 질병 관리 효율성 향상에 기여할 수 있을 것으로 기대된다.

ACKNOWLEDGMENT

· 이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2021-0-00136, 다양한 산업 분야 활용성 증대를 위한 대규모/대용량 블록체인 데이터 고확장성 분산 저장 기술 개발).

· 이 연구는 과학기술정보통신부의 재원으로 한국지능정보사회진흥원의 지원을 받아 구축된 "넙치 질병 데이터"를 활용하여 수행된 연구입니다. 본 연구에 활용된 데이터는 AI 허브(aihub.or.kr)에서 다운로드 받으실 수 있습니다.

참 고 문 헌

[1] 한국통계정보서비스(KOSIS), “시도·시군구별 양식방법별 어종별 양식현황(입식, 생산, 사육, 먹이[사료] 투입량),” 자료갱신일 2025-05-30.

[2] Shim, J.D., Hwang, S.D., Jang, S.Y., Kim, T.W., and Jeong, J.M., “Monitoring of the mortalities in olive flounder (Paralichthys olivaceus) farms of Korea,” Journal of Fish Pathology, vol. 32, no. 1, pp. 29-35, 2019.

[3] 한국지능정보사회진흥원(NIA), AI-Hub, “넙치 질병 이미지 데이터셋,” 2020.

[4] Simonyan, K., and Zisserman, A., “Very Deep Convolutional Networks for Large-Scale Image Recognition,” ICLR, 2015.

[5] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.C., “MobileNetV2: Inverted Residuals and Linear Bottlenecks,” CVPR, pp. 4510-4520, 2018.

[6] Zhang, M.L., and Zhou, Z.H., “A Review on Multi-Label Learning Algorithms,” IEEE TKDE, vol. 26, no. 8, pp. 1819-1837, 2014.

최근 정보통신기술 One-Step Image-to-Image Translation 의 원격주행용 Diffusion Model 기반 Semantic Compression 적용에 관한 연구

윤성빈, 김예준, 강민우, 서효운
성균관대학교

emsekr@g.skku.edu, kyj8130@g.skku.edu, kmw8123@g.skku.edu, hyowoonseo@skku.edu

A Study on the Application of Diffusion Model-Based Semantic Compression For One-Step Image-to-Image Translation in Remote Driving Systems

Yun Sung Bin, Kim Ye Jun, Kang Min Woo, Seo Hyowoon
Sungkyunkwan Univ.

요 약

본 논문은 의미 압축 프레임워크인 GESCO 모델이 야간 주행 환경에서 겪는 성능 저하 문제를 해결하고자 도메인 변환 모듈을 결합한 복원 체계를 제안한다. 제안 방식은 야간 이미지를 주간으로 변환하여 의미 압축을 수행하여 mIoU 와 LPIPS 지표를 각각 평균 707.01%와 12.88% 향상시켰다. 이를 통해 저조도 환경에서도 의미 보존과 지각 품질을 개선하고 통신 시스템의 견고성을 확보하였다.

Ⅰ. 서 론

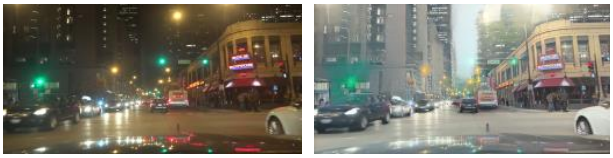
의미 기반 통신은 원본 데이터의 모든 세부 정보를 그대로 전송하는 대신, 수신 단말의 의사결정에 필요한 의미적 요소를 추출 및 압축하여 전송함으로써 대역폭 사용량을 획기적으로 절감할 수 있다.[1] Generative Semantic Communication(GESCO) 모델은 확산 모델 기반 의미 압축 프레임워크이다. 송신단이 One-Hot Segmentation Map 을 압축·전송하고 수신단이 FDS 블록과 SPADE 기반 조건부 확산 모델을 통해 의미를 보존하며 이미지를 재생성하는 방식으로, 낮은 SNR 환경에서도 높은 의미 복원 성능을 유지한다.[2] 그러나 해당 모델은 주간 주행 환경에서 안정적인 성능을 보였으나, 야간 주행과 같이 조명 조건이 극도로 변화하는 상황에서는 복원 성능이 급격히 저하되는 한계가 존재한다.

이러한 한계를 극복하기 위해, 본 연구는 GESCO 모델에 One-Step Image-to-Image Translation 기법을 결합하여 야간 주행 데이터를 주간 데이터로 변환한 후, 이를 GESCO 송신단에 입력해 의미 압축·복원을 수행하는 방식을 제안한다. 이를 위해 CycleGAN-Turbo 구조를 기반으로 한 One-Step 변환 프레임워크를 채택하였다. 이 접근법은 다단계 확산 기반 이미지 변환 모델보다 연산 복잡도가 낮으면서도, 원본 영상의 장면 구조를 유지한 채 조명·색상 특성을 효과적으로 변환할 수 있는 장점이 있다.[3]

결과적으로 제안 시스템은 원격 자율주행 환경에서 다양한 조명 조건에도 안정적인 의미 복원 성능을 보장하며, 특히 야간 주행 상황에서도 주간 대비 유사한 mIoU, LPIPS 성능 확보를 목표로 한다. [4,5]

Ⅱ. 본론

본 논문에서는 67 epoch 를 학습한 GESCO 모델에 야간 주행 데이터로 의미 압축·복원을 수행했을 때와 동일한 모델에 야간 주행 데이터를 One-Step Image-to-Image 로 주간 도메인 주행 데이터처럼 변환 후 의미 압축·복원했을 때의 결과를 비교한다. 각 데이터에 대한 결과 품질 평가는 생성 이미지에서 주요 객체의 보존 정도를 평가하는 mIoU, 생성 이미지와 원본 이미지 간의 시각적 유사도를 정량화하기 위한 지표인 LPIPS 를 사용하였다.



<그림 1. One-Step Image-to-Image Translation 을 이용해 야간 주행 데이터(좌)를 주간 주행 데이터(우)로 변환한 모습>

<그림 2. 원본 야간 데이터셋 GESCO 출력 결과(좌)와 주간으로 변환한 데이터셋 출력 결과(우)>

<그림 1.>처럼 야간 주행 데이터를 주간 주행 데이터처럼 변환해 색, 명암 대비 측면에서 보다 개선된 데이터를 <그림 2.>와 같이 GESCO 모델의 입력으로 하여 의미 압축·복원 시 차·도로, 건물 등 의미상 주요 객체들의 보존 정도가 향상되었다.

SNR[dB]	mIoU ↑ [%]		mIoU 향상 비율[%]
	야간 이미지	주간 변환 이미지	
1	0.12	0.32	166.71
5	0.13	0.25	92.31
10	0.11	1.67	1418.18
15	0.11	2.28	1972.72
20	0.44	2.36	436.36
30	0.44	2.71	515.91
100	0.64	2.86	346.88

<표 1. 동일 학습 GESCO 모델에서 출력한 야간 이미지와 주간 변환 이미지의 mIoU 평가 결과 비교>

생성 이미지 16 개에 대한 SNR 별 mIoU 값을 비교한 결과, <표 1.>에서 알 수 있듯 야간 주행 데이터의 주간 주행 데이터 형식으로 변환 후 평균 707.01%로 매우 높은 수준의 mIoU 향상을 보였다.

SNR[dB]	LPIPS ↓ 평균값 [%]		LPIPS 향상 비율[%]
	야간 이미지	주간 변환 이미지	
1	0.7892	0.7164	9.22
5	0.7863	0.6878	12.53
10	0.7829	0.6457	17.52
15	0.7463	0.6246	16.31
20	0.6979	0.6097	12.60
30	0.6687	0.6008	10.15
100	0.6758	0.5959	11.82

<표 2. 동일 학습 GESCO 모델에서 출력한 야간 이미지와 주간 변환 이미지의 LPIPS 평가 결과 비교>

생성 이미지 16 개에 대한 SNR 별 LPIPS 평균값을 비교한 결과, <표 2.>에서 알 수 있듯 야간 주행 데이터의 주간 주행 데이터 형식으로 변환 후 평균 12.88%의 준수한 LPIPS 향상 수준을 보였다.

mIoU 및 LPIPS 모두 향상된 것을 보이며, 이는 도메인을 변환함으로써 조명, 색상 개선을 통해 객체 의미 보존 및 시각적 유사도를 향상시킨 결과이다.

Ⅲ. 결론

본 논문에서는 확산 모델 기반 의미 압축 모델에 CycleGAN-Turbo 기반 도메인 변환 모듈을 결합하여, 원격 자율주행 환경에서 야간 주행 데이터 전송 시 성능 저하 문제의 해결 방향을 제안하였다.

제안 방식은 모든 SNR 구간에서 mIoU, LPIPS 가 유의미하게 향상되었다. mIoU 는 평균 707.01%, LPIPS 는 평균 12.88%의 향상을 보여 야간 도메인에서의 성능 개선을 입증하였다.

그러나 본 연구는 몇 가지 한계점이 존재한다. 첫째, 절대적인 mIoU 값이 낮다는 점이다. GESCO 선행 논문에서 보고된 90% 수준의 mIoU 에 비해 본 연구의 절대 수치는 최대 2~3% 수준에 불과하다. 이는 사용한 Segmentation 모델, 야간 데이터셋 품질, 변환 후 색채/조명 재현과 학습 환경의 한계 등에 기인할 수 있다. 둘째, 도메인 변환 모듈의 일반화 관점에서, 즉 CycleGAN-Turbo 가 기존 학습 데이터 분포와 크게 다른 야간 환경(여름과 겨울의 차이 등)에서는 변환 품질이 저하될 수 있다는 한계점도 존재한다. 셋째, 실시간 처리 성능 부족 또한 한계가 될 수 있다. 변환과 복원 모두 GPU 연산을 요구하며, 고해상도(2048×1024) 환경에서 실시간성 확보가 어렵다는 점은 본 모델의 실시간 원격 주행 시스템에 즉각적인 도입을 불가하게 만든다.

향후 연구에서는 다양한 야간 환경 조건을 포함한 학습 데이터 확충, 도메인 변환과 의미 복원기의 공동 최적화를 통한 품질 개선, 그리고 경량화 및 최적화 기반의 실시간 처리 가능성 검토를 진행할 예정이다. 이를 통해 제안 방법은 원격 자율주행 뿐만 아니라 저조도 영상 통신, 스마트 시티 감시 등 다양한 차세대 통신·비전 융합 응용에 폭넓게 적용 가능할 것으로 기대된다.

ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(RS-2025-00555907).

참 고 문 헌

[1] X. Luo, H.-H. Chen, and Q. Guo, "Semantic communications: Overview, open issues, and future research directions," IEEE Wireless Communications, vol. 29, no. 1, pp. 210-219, Feb. 2022.

[2] E. Grassucci, S. Barbarossa, and D. Comminiello, "Generative semantic communication: Diffusion models beyond bit recovery," arXiv:2306.04321, Jun. 2023. (https://arxiv.org/abs/2306.04321)

[3] G. Parmar, T. Park, S. Narasimhan, and J.-Y. Zhu, "One-step image translation with text-to-image models," arXiv:2403.12036, Mar. 2024. (https://arxiv.org/abs/2403.12036)

[4] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The Cityscapes dataset for semantic urban scene understanding," Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 3213-3223, 2016.

[5] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 586-595, 2018.

어안렌즈 카메라 캘리브레이션 전역 품질 지표에 관한 연구

김한구, 강신재, 차대웅, 한동석*
경북대학교 IT 대학 전자공학부
khangu0729@knu.ac.kr, kangsj129@knu.ac.kr, dwcha@auto-it.co.kr, dshan@knu.ac.kr*

A Study on the Global Quality Metric of Fisheye Camera Calibration

Han Gu Kim, Shin Jae Kang, Dae Woong Cha, Dong Seog Han*
School of Electronics Engineering, Kyungpook National Univ.
Automative Information Technology Inc.

요 약

본 논문은 어안렌즈(fisheye) 카메라 캘리브레이션 과정에서 낮은 재투영 오차(reprojection error)에도 불구하고 이미지에서 전역적으로 왜곡이 존재할 수 있는 함정을 분석한다. 이러한 문제는 캘리브레이션을 위한 데이터가 이미지 중앙에 편중될 때 주로 발생하며, 기존 캘리브레이션 품질 지표로 널리 사용되는 재투영 오차만으로는 실패 여부를 정확히 판단하기 어렵다. 이를 해결하기 위해, 본 논문에서는 보정된 이미지의 전역 품질을 정량화하는 새로운 지표 GDQI(global distortion quality index)를 제안한다. 이를 통해 어안렌즈 카메라 캘리브레이션의 품질 평가에서 재투영 오차의 한계를 보완하고, 전역적 관점의 보정 품질 평가를 위한 새로운 접근을 제시함으로써 SLAM(simultaneous localization and mapping) 등의 비전 알고리즘에서 은폐된 왜곡 오류로 인한 치명적 영향을 사전에 진단하고 방지할 수 있을 것으로 기대한다.

Ⅰ. 서 론

어안렌즈 카메라는 180° 이상의 매우 넓은 화각을 제공하여 자율주행, AR, VR 등 다양한 비전 분야에서 많이 활용된다. 그러나 광각으로 인한 강한 비선형 왜곡 때문에, 정확한 캘리브레이션을 통한 보정이 필수적이다. 일반적으로 카메라 캘리브레이션의 품질 평가는 재투영 오차(reprojection error, RPE)로 이루어지며, 이는 캘리브레이션 데이터에 대한 적합도를 의미한다. 하지만, 이 지표만으로는 보정의 전역적 성공 여부를 담보하지 못한다. 캘리브레이션 데이터가 영상 전체에 고르게 분포될 경우, RPE가 전역 품질을 잘 반영하지만, 데이터 분포에 편향이 있으면 해당 영역에서는 오차가 잘 보정되더라도 관측되지 않은 영역에서는 왜곡이 크게 남을 수 있다. RPE는 관측된 지점에 대해서만 계산되므로, 보드 외곽이나 이미지 가장자리의 왜곡 상황은 제대로 평가되지 않는다. 이러한 이유로 보정 결과에 대해 잘못된 판단을 할 수 있다.

이에 대해, 본 연구의 목표는 재투영 오차만으로 놓칠 수 있는 보정 실패를 전역 관점에서 진단하는 것으로, 세계의 직선 구조물이 보정 후에도 곧게 유지되고, 화면 전역에 걸쳐 균일한 배율을 갖는지를 정량 평가하는 전역 보정 품질 지표를 개발한다. 이는 기존의 RPE 지표를 보완하여, 어안렌즈 카메라와 같이 왜곡이 큰 광각 카메라의 캘리브레이션 품질을 더욱 엄격하고 신뢰성 있게 평가할 수 있도록 한다.

Ⅱ. 본 론

1. GDQI(Global Distortion Quality Index) 개념

GDQI는 보정 영상의 전역적 왜곡 잔존 여부를 수치화하기 위한 지표로서, 스케일 균일성 오차(scale uniformity error, SUE), 직선 보존 오차(line straightness error, LSE)를 종합한다. 두 값이 모두 낮을수록 이미지 전역에 걸쳐 올바른 보정이 이루어졌음을 의미하며, 반대로 이들 중 하나라도 높으면, 특정 국소 영역에 잔류 왜곡이나 비일관적 축척 변화가 존재함을 의미한다.

1.1. 스케일 균일성 오차

이론상 완벽한 보정에서는 이미지 전역에서 등일한 배율을 가진다. 즉, 왜곡되지 않은 이상적인 카메라에서 서로 다른 두 영역에서 동일한 크기의 물체가 투영될 때 비슷한 크기로 나타나야 한다. 그러나 캘리브레이션이 잘못된다면 보정된 영상에서 위치에 따라 서로 다른 배율로 보정이 이루어져, 어떤 부분은 확대되어 보이고 다른 부분은 축소되어 보이는 현상이 발생한다. 이를 정량화하기 위해, 본 연구에서는 자코비안(jacobian) 행렬과 보정에 따른 국소 면적 변화율을 표현한 행렬식을 사용했으며, 스케일 균일성 오차(SUE)는 보정 영상의 전체 화소들에 대한 값들의 분산으로 정의한다. SUE 값이 클수록 보정 배율이 위치에 따라 심하게 달라짐을 의미하며, 이는 전역적인 왜곡 보정 불균형을 나타낸다.

1.2. 직선 보존 오차

이론상 완벽한 보정에서는 직선 구조물이 보정된 2D 영상에서도 직선으로 표현되어야 한다. 올바르게 보정되었다면, 어안렌즈 렌즈의 왜곡으로 굽어졌던 직선들은 다시 곧게 펴져야 한다는 것이다. 본 연구에서는 보정 결과 영상에서 LSD(line segment detector) 알고리즘을 통해 직선 성분을 추출한 후, 각 직선상의 화소들이 해당 직선 모델에서 벗어나는 평균 거리를 계산한다. 이 값을 직선 보존 오차(LSE)로 정의하며, 단위는 픽셀이다. 여러 직선을 검출한 경우, 전체 LSE는 검출된 모든 직선에 대한 평균 거리로 산출한다. LSE 값이 클수록 보정 후에도 영상 내 일부 직선이 눈에 띄게 굽어있다는 뜻이며, 이는 해당 국소 영역에 왜곡이 잔존하고 있음을 의미한다.

1.3 GDQI 지표 활용

최종적으로 GDQI는 LSE와 SUE를 동시에 고려하기 위해 두 값을 단순 합산하거나 다차원 벡터로 취급할 수 있다. 두 오류가 모두 임계치 이하로 낮을 때만 보정 성공으로 판단하는 게이트 함수로 활용하거나, 혹은

가중합한 형태로 사용할 수 있다.

본 연구에서는 이해를 돕기 위해 LSE와 SUE 값을 개별적으로 관찰하고 논의하며, 절대적인 GDQI 스코어 산출은 추후 과제로 남겨둔다. 중요한 것은, GDQI의 도입으로 재투영 오차 하나만으로 간과될 수 있던 보정 품질을 다양한 관점에서 평가할 수 있게 되었다는 점이다. LSE와 SUE는 각각 기하학적 정확도(직선성)와 광학적 일관성(배율 균일도) 측면에서 보정 결과를 감지하므로, 두 지표를 함께 사용하면 전역 왜곡에 대한 높은 민감도를 얻을 수 있다.

2. GDQI을 적용한 캘리브레이션 지표 비교 실험

실험을 통해 본 논문에서 제안하는 지표 GDQI가 실제로 재투영 오차로는 드러나지 않는 보정 실패를 잘 식별해 내는지를 확인하였다. 데이터셋 A는 ChArUco 보드 패턴을 이미지 중앙부 위주로 촬영하여 보정 실패를 의도적으로 유도한 경우이고, 데이터셋 B는 ChArUco 보드를 중앙부터 가장자리까지 고르게 분포하도록 촬영한 경우로, 두 데이터셋을 각각 OpenCV 어안렌즈 모델로 캘리브레이션 한 후, 얻어진 카메라 파라미터로 동일한 테스트 이미지를 Balance = 0.5 동일 조건 아래에서 보정하였고, 캘리브레이션 결과의 RPE, 보정 이미지의 SUE, LSE를 각각 계산하여 비교하였다.

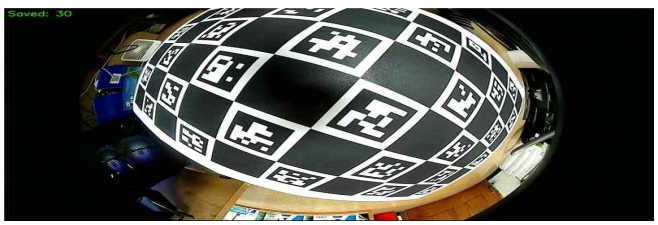
[표 1]에 따르면, 데이터셋 A와 B 모두 RPE는 0.5 픽셀 내외로, 수치상 두 경우 모두 보정이 잘 된 것처럼 나타난다. 특히, 실패 사례 A는 RPE가 0.365 픽셀로 B의 0.472 픽셀보다도 낮은 값을 보여주어, RPE만 봐서는 A의 보정이 더 좋아 보인다. 그러나 A의 경우에서 SUE(var) = 1.199, SUE(mean) = 10.709, LSE = 16.124 픽셀이고, B의 경우에서 SUE(var) = 0.546, SUE(mean) = 2.631, LSE = 12.282 픽셀이다. 이는 데이터셋 A 캘리브레이션 파라미터로 보정한 영상이 데이터셋 B 캘리브레이션 파라미터로 보정한 영상에 비해, 영상 속 위치별로 픽셀 크기가 심하게 불균일함을, 영상 속 직선들이 더 많이 휘어져 있음을 나타낸다.

원본 영상인 [그림 1]과 보정된 [그림 2], [그림 3]을 비교하면, GDQI 지표가 지적한 품질 차이를 명확히 확인할 수 있다. [그림 2]는 영상의 가장자리 부분이 좌측이 비정상적으로 확대되는 스케일 불균일 현상 및 원본의 많은 부분이 화면 밖으로 밀려 나갔음을, [그림 3]에 비해 ChArUco 보드의 격자선들이 미세하게 더 휘어져 있음을 확인할 수 있다. [그림 3]은 상대적으로 스케일이 균일하고 격자선들이 곧게 유지되어, [그림 2]에 비해 원본 정보를 더욱 보존하며 안정적인 결과를 보여준다.

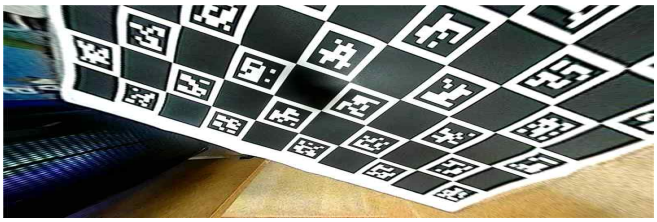
이처럼 GDQI 지표는 두 사례 간 보정 품질 차이를 극명하게 드러낸다. 재투영 오차만으로는 구별되지 않던 차이가, SUE와 LSE를 통해서 명확히 수치화되었다. 이는 전역 왜곡 관점에서 GDQI가 RPE에서 확인할 수 없는 캘리브레이션의 성공과 실패를 가려내는 데 유용함을 뒷받침한다.

[표 1] 데이터셋에 따른 RPE, SUE, LSE 비교표

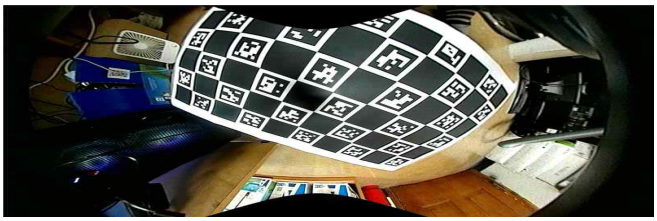
데이터셋	RPE (pixel)	SUE (var)	SUE (mean)	LSE (pixel)
A	0.365	1.199	10.709	16.124
B	0.472	0.546	2.631	12.282



[그림 1] 보정 전 원본 영상



[그림 2] 데이터셋 A 캘리브레이션으로 얻은 파라미터로 보정한 영상



[그림 3] 데이터셋 B 캘리브레이션으로 얻은 파라미터로 보정한 영상

Ⅲ. 결 론

본 논문에서는 어안렌즈 카메라 캘리브레이션 평가에서 널리 사용되는 재투영 오차(RPE) 지표의 함정을 지적하고, 이를 보완하기 위한 전역 왜곡 품질 지표(GDQI)를 제안하였다. 제한된 영역의 점 오차만을 보는 RPE와 달리, GDQI는 직선 유지 정도(LSE)와 배율 균일도(SUE)를 통해 보정 결과의 전역적 왜곡 수준을 정량화하였으며, 이를 통해 숨겨진 보정 오류를 조기에 발견할 수 있다.

향후 연구로, 본 논문에서 제시한 SUE와 LSE를 통합하여 단일 GDQI 점수를 산출하는 표준화 방안을 마련하고자 한다. 또한, 제안한 접근을 다양한 렌즈 유형(광각, 어안, 파노라마 등)에 적용하여 각 경우에 최적화된 품질 평가 방법을 탐색할 계획이다.

궁극적으로, 본 연구는 카메라 보정 품질을 평가하고 향상하는데 새로운 통찰을 제공하며, 왜곡 보정이 필요한 모든 비전 응용 분야에서 신뢰성 향상에 이바지할 것으로 기대한다.

ACKNOWLEDGMENT

이 성과는 정부(과학기술정보통신부)의 재원으로 과학기술사업화진흥원의 지원을 받아 수행된 연구임(RS-2025-25444562).

참 고 문 헌

[1] Rameau, F., Park, J., Bailo, O., and Kweon, I. S., “MC-Calib: A generic and robust calibration toolbox for multi-camera systems,” Computer Vision and Image Understanding, vol. 217, Art. no. 103353, Jan. 2022.

[2] Von Gioi, R. G., Jakubowicz, J., Morel, J. M., and Randall, G., “LSD: A fast line segment detector with a false detection control,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 4, pp. 722 - 732, Apr. 2010.

[3] Garrido-Jurado, S., Muñoz-Salinas, R., Madrid-Cuevas, F. J., and Marín-Jiménez, M. J., “Automatic generation and detection of highly reliable fiducial markers under occlusion,” Pattern Recognition, vol. 47, no. 6, pp. 2280 - 2292, Jun. 2014.

Wi-Fi CSI 데이터를 활용한 재실 인원 감지

오태린, 김범석, 정기원, 김민석*
충남대학교, *충남대학교

5234065@o.cnu.ac.kr, seeksick1009@o.cnu.ac.kr, kwjung2571@o.cnu.ac.kr, *kms4105@cnu.ac.kr

Occupancy Estimation Using Wi-Fi CSI Data

Oh Tae Rin, Kim Beom Seok, Jung Ki Won and Kim Min Seuk*
Chungnam National Univ, *Chungnam National Univ

요 약

본 연구는 송수신기가 고정된 device-free 환경에서 Wi-Fi 신호의 (Channel State Information) CSI 의 진폭을 분석하여 실내 인원을 감지하는 시스템을 제안한다. CSI 진폭의 동적 변화를 보다 정밀하게 구분하기 위해 진폭 값을 기반으로 슬라이딩 윈도우 기법을 적용하였으며, 각 샘플에 대해 진폭의 분산과 진폭으로부터 계산한 곡률 값을 기반으로 한 평균과 표준편차 특징을 추출했다. 추출된 특징을 다양한 머신러닝 및 딥러닝에 적용하여 성능을 검증했으며, 제안하는 시스템은 실제 환경에서 96 % 의 높은 정확도를 보인다. 이는 실내 인원 변화에 따른 CSI 신호의 변동성을 토대로 효과적으로 재실 인원 수를 구분할 수 있음을 시사한다.

Ⅰ. 서 론

최근 (Internet of Things) IoT 기술의 발전으로 Wi-Fi 기반의 비접촉 센싱 기술에 대한 관심이 높아지고 있다. Wi-Fi 는 대부분의 실내 환경에 이미 구축되어 있어 별도의 센서 장비 없이 구현이 가능하며, 영상이 아닌 전파 신호를 기반으로 하므로 사생활 침해가 없다는 장점을 가진다. [1]

Wi-Fi 통신에서 얻을 수 있는 정보에는 단일 신호 세기 값인 (Received Signal Strength) RSS 와, 고차원 정보인 (Channel State Information) CSI 가 있다. RSS 는 신호 세기를 하나의 스칼라 값으로 나타내기 때문에 환경 변화에 따른 정밀한 감지에는 한계가 있다. 반면, CSI 는 (Orthogonal Frequency Division Multiplexing) OFDM 기술을 통해 각 부반송파의 진폭과 위상 정보를 포함하는 복소수 형태로 제공되며, 미세한 움직임이나 인원 변화로 인한 다중 경로의 변화까지 정밀하게 분석이 가능하다. 이러한 특성을 활용하여, 본 연구에서는 Wi-Fi 기반의 비접촉 센싱에 정밀한 환경 감지를 하는 CSI 정보를 활용하여 실내 인원 감지 시스템을 설계하였다.

본 연구의 주요 기여는 다음과 같다.

1. device-free 방식을 적용하여 Wi-Fi 송수신기가 고정되어 있는 상황에서 사람의 움직임에 의한 CSI 특징을 효과적으로 추출하였다.
2. CSI 진폭 값에서 분산과 곡률을 추출하고 슬라이딩 윈도우를 적용해 기계학습 및 딥러닝의 입력으로 사용하는 기법을 적용하여 인원 변화에 민감한 신호 변동을 효과적으로 판별했다.
3. 여러 기계학습 및 딥러닝 기법을 적용하여 최고 96 % 의 재실 인원 감지 성능을 얻어 그 실효성을 입증하였다.

II. 시스템 모델

2.1 시스템 구조

그림 1 은 제안하는 인원 감지 프레임워크를 나타낸다. CSI 데이터를 수집한 후에 진폭을 추출하여 슬라이딩 윈도우 기법을 적용한다. 이후 분산과 곡률 특징을 추출하여 머신러닝과 딥러닝 모델에 적용한다.

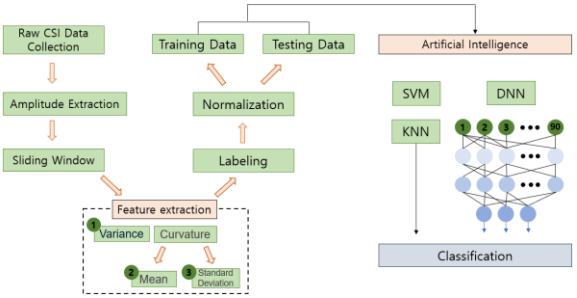


그림 1 인원 감지 프레임워크

2.2 특징 추출

CSI 데이터로부터 패킷-부반송파 행렬 형태의 진폭 값을 추출했다.[2] 시계열 특징을 활용하고자 슬라이딩 윈도우를 적용했다.[3] 또한 홉 크기를 최대한 작게 설정하여 윈도우 경계에서의 패킷 손실을 최소화했다.

진폭의 분산은 정적/동적 상태를 구분하는 지표가 된다. 부재 시에는 진폭이 안정되어 분산 값이 낮게 나타났고, 재실 시에는 사람의 움직임으로 인해 진폭 변동성이 커져 분산 값이 뚜렷하게 증가했다.

또한 우리는 곡률 특성을 도입했다. 곡률은 인원 수가 증가하여 다중 경로가 발생할수록 함께 높아지는 특징이 있다. 따라서 각 윈도우에서 계산된 곡률의 평균과 표준편차를 사용하여 인원 수 추정의 정확도를 높였다. CSI 진폭 y 에 대한 곡률을 구하는 식은 (1) 과 같다.

$$\kappa = \frac{|y''|}{(1+(y')^2)^{3/2}} \quad (1)$$

2.3 기계학습 및 딥러닝

본 연구에서 사용한 데이터셋은 3 개의 클래스(0 명, 1 명, 2 명)로 구성된다. 데이터는 클래스별 3 개씩, 총 9 개의 파일로부터 수집되었으며, 각 파일에서 약 980 개의 윈도우 샘플을 추출하여 전체 데이터셋은 총 8,825 개의 샘플로 구성했다. 모든 샘플 입력은 모델 학습 전에 Min-Max 정규화를 거쳤다.

특징 벡터 분류 성능을 비교 분석하기 위해 머신러닝 모델인 (Support Vector Machine) SVM, (K-Nearest Neighbor) KNN 와 딥러닝 모델인 (Deep Neural Network) DNN 을 사용했다. SVM 은 다중 클래스 분류를 위해 (One-vs-One) OVO 과 (Radial Basis Funcion) RBF 커널을 사용했다. KNN 은 최근접 이웃의 수인 하이퍼파라미터 K 를 35 로 설정했다. DNN 모델은 입력층, 두 개의 은닉층 그리고 출력층으로 구성된다.

III. 실험 결과

3.1 실험 환경 설정

그림 2 와 같이 가로 6m, 세로가 9m 크기의 실내 강의실 환경에서 송신기와 수신기를 각각 1 대씩 책상 위에 배치하였으며, 두 장치 간 거리는 4m 이다. 각 장치는 Intel 5300 Wi-Fi (Network Interface Card) NIC 에 3 개의 안테나를 장착하였다. 패킷당 30 개의 서브캐리어에 대한 CSI 데이터를 수집하였다. 재실의 경우 송신기와 수신기 사이를 시계 방향으로 이동하는 상황에서 데이터를 수집하였다. 이후, 패킷을 윈도우 크기 W(10, 20, 30)로 샘플링하여 W×30 형태의 배열을 구성하고, 진폭의 분산, 곡률의 평균, 표준편차 특징 총 90 개의 값을 추출하여 머신러닝 및 딥러닝 모델의 입력으로 사용하였다. 홉 크기는 5 로 설정하였다.

그림 2

3.2 모델 별 곡률의 영향

그림 3 에서 진폭의 분산(Variance)만을 이용한 경우 SVM, KNN, DNN 의 정확도는 각각 64 %, 68 %, 66 % 로 나타났다. 곡률을 전처리 입력으로 추가한 결과, SVM 과 KNN 의 정확도는 각각 96 %, 92 % 으로 32 %, 24 % 향상되었으며, DNN 은 96 % 로 30 % 향상되었다.

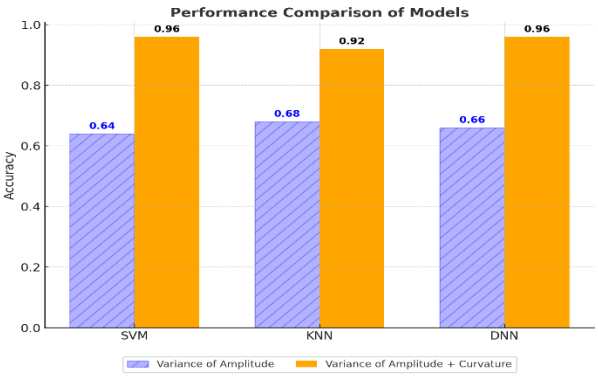


그림 3

3.3 모델 별 윈도우 크기의 영향

윈도우 크기를 W=10, 20, 30 으로 변화시키며 성능을 비교하였다. 그림 4 와 같이 W=10 에서 SVM, KNN, DNN 의 정확도는 각각 88 %, 83 %, 84 % 였다.

W=30 으로 변경 시, SVM, KNN, DNN 의 정확도는 각각 8 %, 9 %, 12% 향상되어 96 %, 92 %, 96 % 를 기록하였다. 이러한 결과는 샘플링 윈도우 크기가 모델 성능에 영향을 미친다는 것을 보여준다.

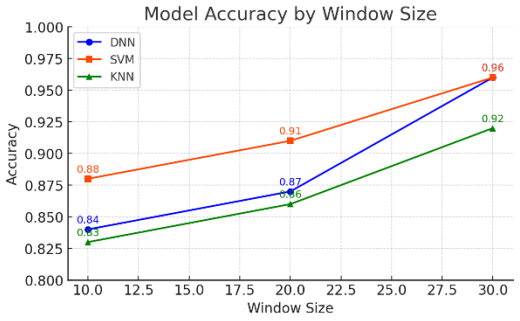


그림 4

최종적으로 그림 5 의 confusion matrix 와 같이 DNN 과 윈도우 크기 W = 30, 곡률 정보를 입력으로 사용했을 때 96 % 의 재실 인원 감지 성능을 확보하였다.

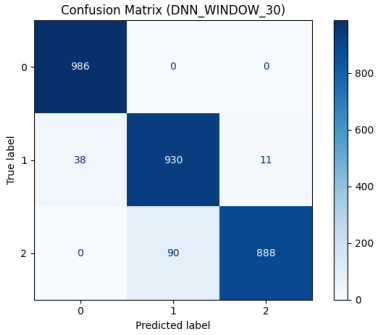


그림 5

IV. 결론

본 논문에서는 Wi-Fi CSI 데이터를 활용하여 재실 인원 감지를 수행하는 기계학습 및 딥러닝 기법을 제안하였다. CSI 진폭의 분산은 실내 환경의 정적/동적 상태를 명확하게 구분하는 기준을 제공했다. 또한 CSI 진폭의 곡률(Curvature) 특징을 전처리 입력으로 추가하고 슬라이딩 윈도우 기반의 데이터 구성 방식을 적용하여 재실 인원에 따른 분류 성능을 향상시켰다. 제안하는 기법은 DNN 모델에서 최대 96 % 의 정확도를 달성하여 효과적인 재실 인원 감지 기능을 수행함을 입증하였다.

ACKNOWLEDGMENT

본 연구는 충남대학교 학술연구비에 의해 지원되었음

참 고 문 헌

[1] I. Sobron, J. Del Ser, I. Eizmendi, and M. Vélez, "Device-free people counting in IoT environments: New insights, results, and open challenges," IEEE Internet Things J., 2018.

[2] Z. Liu, R. Yuan, Y. Yuan, Y. Yang and X. Guan, "A Sensor-Free Crowd Counting Framework for Indoor Environments Based on Channel State Information," IEEE Sens. J., 2022.

[3] O. Oshiga, et al., "Human detection for crowd count estimation using CSI of WiFi signals," in Proc. 15th Int. Conf. Electron., Comput. Comput. (ICECCO), Keffi, Nigeria, Dec. 2019, pp. 1-6.

CSI 기반 실내 위치 추정을 위한 구역 선택형 전문가 혼합(RSMoE)

한승우, 문하은, 이서인, 김민석*
충남대학교, *충남대학교

hsw2266@g.cnu.ac.kr, cherishuu02@o.cnu.ac.kr, sslynn22@o.cnu.ac.kr, *kms4105@cnu.ac.kr

Region-Selective Mixture-of-Experts for CSI-based Indoor Localization

Han Seung Woo, Moon Ha Eun, Lee Seo In, and Kim Min Seuk*
Chungnam National Univ., * Chungnam National Univ.

요 약

Wi-Fi 기반 채널 상태 정보(Channel State Information, CSI)를 활용한 실내 위치 추정은 NLoS(Non-Line of Sight) 로 인한 왜곡과 구역별 전파 특성이 크게 달라지는 문제를 동시에 안고 있다. 본 연구는 이 문제를 구역 전문가 라우팅 문제로 재정의하고, 이를 해결하기 위한 구역 선택형 전문가 혼합 모델(RSMoE)을 제안한다. 제안 모델은 먼저 Gating 네트워크로 입력 CSI 가 속한 구역을 예측하고, 이후 해당 구역의 단일 전문가 CNN(Convolutional Neural Network) 모델만이 활성화되는 단일 전문가 라우팅을 사용한다. 기존 VSDL(Viewpoint-Selective Deep Learning) 모델의 평균 위치 오차(4.01 m) 대비 3.05 m(≈24 % 개선)의 평균 위치 오차를 달성하며, 불필요한 연산을 줄이면서도 CSI 특성을 구역별로 분할 학습하여 전문성을 높였다.

Ⅰ. 서 론

Wi-Fi 기반 채널 상태 정보(Channel State Information, CSI)를 활용한 실내 위치 추정은 고정밀 예측이 가능하다는 장점으로 주목받고 있지만, 실제 환경에서는 벽, 장애물, 건물 구조 등으로 인한 Multipath·NLoS(Non-Line of Sight) 현상으로 높은 변동성과 잡음이 커지는 문제가 있다 [1].

이 문제를 해결하기 위해, 우리는 선행 연구에서 LoS 환경에 있는 뷰를 선별하고 뷰 가중치를 부여하는 VSDL(Viewpoint-Selective Deep Learning) 모델을 제안한 바 있다 [2]. VSDL 은 각 AP(Access Point) 그룹의 정보를 뷰 단위로 재가중하여 단일 회귀 모델의 입력으로 사용하였다.

그러나 이러한 구조에는 두 가지 본질적인 한계가 존재한다. 첫째, 단일 회귀 모델은 모든 구역에서 수집된 정보를 하나의 입력으로 함께 처리한다. 이로 인해 특정 구역의 고유한 전파 특성에 맞춘 전문화된 학습이 어렵고, 모든 구역의 정보가 반영된 평균적인 예측을 수행하게 된다. 둘째, 위치 추정을 위해 모든 AP 그룹에 대한 연산이 필요하므로 계산 효율이 저하된다. 가중치가 낮은 구역이라도 연산 과정에서 제외되지 않아, 전체 처리 속도 및 성능이 저하된다.

이러한 한계를 극복하고자, 본 연구에서는 환경을 K 개의 구역(Region)으로 분할하고, 각 구역에 특화된 전문가 모델을 할당하여 위치를 추정하는 구역 선택형 전문가 혼합(Region-Selective Mixture of Experts, RSMoE) 아키텍처를 제안한다 [3]. 이 방식은 구역별 모델 전문성을 극대화하고, 불필요한 연산을 줄여 계산 효율을 향상한다. 실제 실험에서 제안 모델이 단일 전역 회귀 모델 대비 의미 있는 성능 향상을 보였으며, 복잡한 실내 환경일수록 구역 기반 분할 학습 방식의 효과가 우수함을 확인하였다.

Ⅱ. 본론

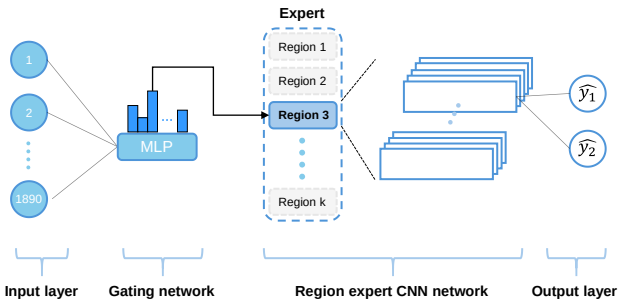


그림 1. RSMoE 모델 아키텍처

2.1. 제안 모델 아키텍처

본 논문에서 제안하는 구역 선택형 전문가 혼합 모델(RSMoE)은 분할 정복(Divide and Conquer) 전략에 기반하여, 복잡한 위치 추정 문제를 다수의 단순한 하위 문제로 나누어 해결한다. 이 아키텍처는 서론에서 언급된 단일 회귀 모델의 한계를 극복하기 위해, 그림 1 과 같이 Gating 네트워크와 구역별 전문가 CNN(Convolutional Neural Network) 네트워크라는 두 가지 핵심 요소로 구성된다. 전체 모델의 작동 원리는 수식 (1) 로 요약할 수 있다.

$$\hat{k} = \underset{k}{argmax}\big(G(X)\big), \qquad \hat{y} = E_{\hat{k}}\left(\varphi\big(X, \hat{K}\big)\right) \quad \cdots \quad (1)$$

가. Gating 네트워크

Gating 네트워크 G 는 전체 1,890 차원의 CSI 입력 벡터 X 를 받아 해당 데이터가 속한 구역을 선별하는 라우터 역할을 수행한다. 이 네트워크는 MLP(Multi-Layer Perceptron) 기반 구조로 K 개의 구역에 대한 확률 분포를 예측한 뒤 argmax 연산을 통해 가장 높은

값을 갖는 단일 구역 $\hat{k}$ 를 선택한다. 예측된 확률은 실제 구역 레이블과의 Cross-Entropy 손실을 최소화하도록 학습된다. 이후, 선택된 구역의 전문가만 활성화하는 단일 전문가 라우팅 방식을 사용하여, 계산 효율성을 높이는 동시에 각 구역의 학습을 독립적으로 분리한다.

나. 구역별 전문가 CNN 네트워크

수식 (1) 에서 Gating 네트워크는 입력 X 에 가장 적합한 구역 $\hat{k}$ 를 예측하고, 해당 전문가 네트워크 $E_{\hat{k}}$ 만이 활성화된다. 이때 변환 함수 $\varphi(X,\hat{K})$ 는 전체 입력 X 에서 $\hat{k}$ 구역에 해당하는 입력만을 동적으로 추출하고 2 차원 이미지 형태로 변환하여, $E_{\hat{K}}$ 에 전달하는 역할을 수행한다. 이러한 구조는 단일 회귀 모델이 모든 구역의 특성을 함께 학습해야 했던 한계를 극복하고, 각 전문가가 담당 구역의 고유한 전파 특성에 최적화되도록 한다.

실내 환경에서는 LoS/NLoS 조건과 Multipath 구성의 차이로 인해 구역별 전파 특성이 뚜렷하게 달라진다. 이러한 차이는 동일한 CSI 패턴이라도 구역마다 다른 해석을 요구하므로, 구역별로 독립적으로 학습한 전문가 CNN 은 해당 영역의 고유한 시간(packet sequence)–주파수(subcarrier)의 2 차원 패턴을 보다 정밀하게 학습할 수 있다.

최종 출력인 $(\widehat{y_1},\widehat{y_2})$ 위치 좌표는 평균 제곱 오차(Mean Squared Error, MSE) 손실을 최소화하는 방향으로 학습된다. 이러한 Divide-and-Conquer 학습 절차는 각 모듈이 고유 기능에 집중하도록 하여 계산 효율을 높이고, NLoS 환경에서 발생하는 미세한 신호 변화를 효과적으로 포착함으로써, 단일 회귀 모델 대비 위치 추정 정확도와 강건성을 크게 향상한다.

2.2. 실험 환경 및 결과

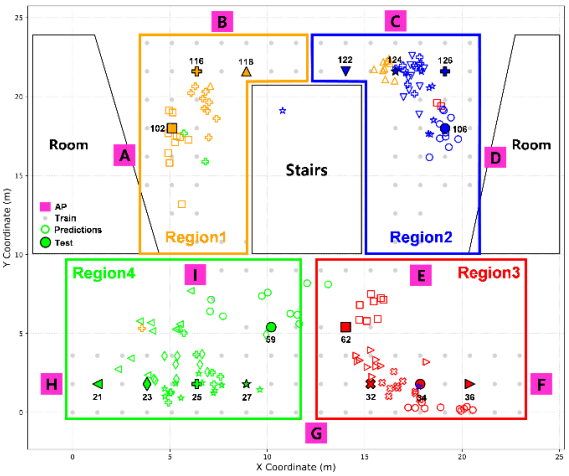


그림 2. 실험 토폴로지 및 위치 추정 결과(충남대학교 도서관)

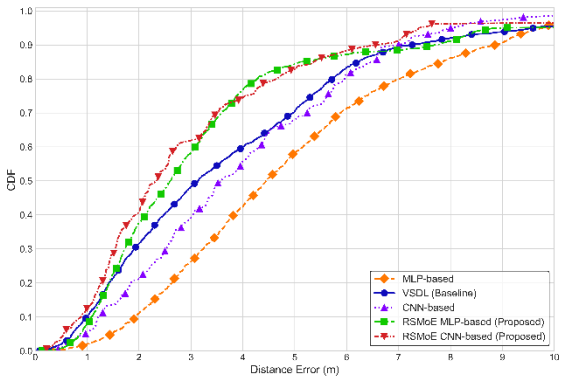


그림 3. Localization Distance Error CDF

본 연구의 실험 결과는 그림 2 와 그림 3 에 요약되어 있다. 실험은 유동 인구가 많은 충남대학교 도서관 실내 환경에서, 141 개 위치에서 수집된 1,890 차원의 CSI 데이터셋을 사용하였다. 이 중 125 개 위치는 학습에, 16 개 위치는 평가에 활용하였다.

성능 비교 결과, MLP 기반 모델은 4.65 m, CNN 기반 모델은 3.99 m 의 평균 위치 오차 성능을 보였다. 제안하는 RSMoE 구조는 MLP 기반 모델에서 3.56 m, CNN 기반 모델에서 3.05 m 의 평균 오차를 달성하여, 기존 VSDL 모델(4.01 m) 대비 약 24 % 의 성능 개선을 보였다. 이러한 성능 향상은 구역 선택 Gating 네트워크와 구역별 전파 특성에 최적화된 전문가 CNN 네트워크의 구조를 통해, 실제 환경에서 발생하는 미세한 신호 변화와 잡음에 효과적으로 대응한 결과로 분석된다.

그림 2 는 제안된 모델의 우수성을 명확히 입증한다. Gating 네트워크가 약 94 % 의 높은 정확도로 실제 구역을 선별하여 최적의 전문가 모델을 할당함으로써, 최종 예측값은 복잡한 구조 및 NLoS 환경에서도 해당 구역 내에 안정적으로 집중된다.

또한 그림 3 의 누적 분포 함수(CDF) 분석 결과, 오차 범위 내에서 RSMoE 의 누적 비율이 단일 회귀 모델 대비 월등히 높게 나타나, 제안 방식이 전반적인 위치 추정 정확도와 안정성을 모두 개선했음을 입증한다.

Ⅲ. 결론

본 논문은 Multipath·NLoS 의 영향으로 구역별 특성이 급격히 달라지는 실내 환경에서 Wi-Fi CSI 기반 위치 추정의 한계를 극복하고자, 구역 선택형 전문가 혼합(RSMoE) 아키텍처를 제안하였다. 제안 모델은 Gating 네트워크가 선택한 구역에 따라 해당 구역에 특화된 전문가 CNN 이 위치를 추정하는 구조로, 실제 데이터셋 실험에서 평균 위치 오차 3.05 m 를 기록해 단일 전역 회귀 모델 대비 약 24 % 의 정확도 향상을 달성하였다. 이는 복잡한 실내 공간을 구역별 하위 문제로 분할해 학습하는 접근이 실질적인 성능 개선으로 이어짐을 입증한다.

향후 연구에서는 군집화 기반 자동 구역 분할과 소프트 라우팅을 도입하여, 다양한 실내 환경과 규모에서의 확장성을 확보하는 동시에 구역 설정과 모델 관리에 필요한 시간적 비용을 줄여 더 실용적인 방향으로 발전시킬 것이다.

ACKNOWLEDGMENT

본 연구는 충남대학교 학술연구비에 의해 지원되었음.

참 고 문 헌

[1] A. Tzur, O. Amrani, and A. Wool, "Direction finding of rogue Wi-Fi access points using an off-the-shelf MIMO-OFDM receiver," *Physical Communication*, vol. 17, pp. 149–164, 2015.

[2] 한승우, 문하은, 이서인, 강민서, and 김민석, "Viewpoint-Selective Deep Learning 을 활용한 NLoS 환경에서의 WiFi CSI 기반 실내 측위," in 한국통신학회 학술대회논문집, 2025, pp. 1045–1046.

[3] W. Fedus, B. Zoph, and N. Shazeer, "Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity," J. Mach. Learn. Res., vol. 23, pp. 1–39, 2022.

영상광용적맥파 기반 혈압 추정: 관심 영역 지정 여부에 따른 딥러닝 성능 평가

김연준 [1,2], 김진원 [2,3], 이성한 [2], 정인철 [2,3]*
[1] 한림대학교 소프트웨어학부
[2] 한림대학교 뇌혈관질환선도연구센터
[3] 한림대학교 인공지능융합학과
yeonjunkim70@gmail.com, *incheol.jeong@hallym.ac.kr

Impact of Region-of-Interest on Deep Learning Performance in Imaging Photoplethysmography-based Blood Pressure Estimation

Yeonjun Kim[1,2], Jinwon Kim[2,3], Sunghan Lee[2], In cheol Jeong[2,3]*
[1]Department of Software Engineering, Hallym University
[2]Cardiovascular Disease Research Center, Hallym University
[3]Department of Artificial Intelligence Convergence, Hallym University

요 약

본 연구는 영상 기반 딥러닝을 이용한 혈압 추정에서 관심 영역(region of interest, ROI) 설정 여부가 모델 성능에 미치는 영향을 검증하였다. 자체 수집한 데이터로 합성곱 신경망 (convolutional neural network, CNN)과 양방향 게이트 순환 신경망 (bidirectional gated recurrent unit, BiGRU) 을 합성한 모델을 학습하였다. ROI 기반 입력 구조와 전체 영상 입력 구조를 비교한 결과, 전체 영상 입력 모델이 수축기 혈압 2.93 mmHg, 이완기 혈압 4.90 mmHg 의 평균 절대 오차 (mean absolute error, MAE)를 기록하며 가장 우수한 성능을 보였다. 또한 학습, 검증, 시험 데이터셋 간 성능 편차가 낮아 안정적인 학습 특성이 확인되었다. 이러한 결과는 ROI 지정 과정 없이도 정확하고 편리한 비접촉식 연속 혈압 모니터링이 가능함을 시사하며, 임상 및 일상 환경에서의 활용 가능성을 제고한다.

Ⅰ. 서 론

가. 연구 배경

혈압은 임상 현장에서 환자의 건강 상태를 평가하는 핵심적인 생리학적 지표이다 [1]. 기존의 측정 방식은 침습적 방법과 비침습적 방법으로 구분된다. 침습적 방법은 높은 정확도를 제공하나 일상적 적용이 어렵고, 비침습적 방법 중 커프 (cuff) 기반 측정은 압박감과 불편감으로 인해 연속적 측정에 한계가 있다 [2, 3]. 심전도 (electrocardiography, ECG)와 광용적맥파 (photoplethysmography, PPG)를 활용한 맥박 전달 시간 기반 혈압 추정은 연속 측정에 유리하지만, 신체에 전극 및 광센서를 부착해야 하므로 사용자 편의성이 떨어진다. 이러한 제약을 극복하기 위해 최근에는 영상광용적맥파 (imaging PPG, iPPG)를 활용한 비접촉 혈압 추정 기술이 활발히 연구되고 있다 [4]. iPPG 는 카메라 영상에서 피부 표면의 미세한 색 변화를 분석하여 맥파를 추출하고, 이를 통해 혈압을 간접적으로 산출하는 방식이다. 기존 연구들은 일반적으로 피부 영역을 관심영역 (region of interest, ROI)으로 지정하여 신호를 추출하고, 이를 딥러닝 모델에 입력하는 절차를 거친다 [5, 6]. 그러나 ROI 는 수작업 지정 또는 별도의 검출 알고리즘이 필요해 실제 적용에서 제약이 따른다. 본 연구는 ROI 를 지정하지 않고 전체 영상으로부터 직접 신호를 학습하는 딥러닝 기반 혈압 추정 모델을 제안하고, 이를 ROI 기반 모델과 비교하여 성능을 검증함으로써 보다 간편하고 실용적인 접근법을 제시하고자 한다.

Ⅱ. 본론

나. 데이터 수집 및 전처리

혈압 추정을 위한 데이터를 수집하기 위해 그림 1 과 같이 PPG 와 영상을 동시에 획득하는 실험을 구성하였다. 실험은 심폐 장애 및 움직임에 제한이 없는 19 세 이상의 건강한 성인 남녀 30 명을 대상으로 진행되었으며, 한림대학교 생명윤리위원회 (HIRB-2024-014)의 승인을 받아 수행되었다. 혈압 추정을 위한 데이터는 그림 1-(a)와 같이 PPG 와 영상 신호를 동시에 획득하는 실험을 통해 수집하였다. 영상데이터는 EX-F1 카메라 (Casio Computer Co., Ltd., Japan)를 이용해 해상도 384×512, 초당 30 프레임으로 얼굴과 왼손바닥을 촬영하여 수집하였다. 왼손 검지에는 반사형 PPG 센서 (BN-PPGED, BIOPAC Systems Inc., USA)를 착용하여 1,000 Hz 로 신호를 기록하였으며, 오른팔에는 혈압계를 부착하여 데이터 수집 시작 시점, 5 분, 종료 시점에 각각 측정하였다. 세 가지 신호는 동기화 과정을 거쳐 동시에 확보되었다. 실험은 안정 시와 경미한 운동 단계로 구성되었다 (그림 1-(b)). 안정 단계에서 참가자는 침대에 누운 상태로 10 분간 신호를 측정하였으며, 이후 2 분간 스텝밀을 이용한 운동을 수행하였다. 전체 30 명 중 동기화에 실패한 18 명을 제외하고, 최종적으로 12 명의 데이터가 분석에 활용되었다. 획득된 혈압 값은 200 초 단위로 구간화 하여 레이블로 할당하였으며, 영상은 10 초 단위로 분할 후 PPG 및 혈압 값과 동기화하였다. 관심 영역 설정 여부에 따른 성능 비교를 위해 아래와 같이 네 가지 입력 구조를 구성하였다:

1 **손 ROI:** 왼손바닥의 중앙에서 30×30 영역 추출
2 **얼굴 ROI:** 코와 눈을 제외한 왼쪽 볼에서 30×30 영역 추출
3 **다중 ROI:** 손 ROI 와 얼굴 ROI 를 함께 사용
4 **전체 영상:** 전체 프레임을 다운 샘플링 (128×128)

다. 모델 학습

그림 1-(c)는 제안하는 모델의 구조를 나타낸다. 입력 데이터는 두 개의 CNN 블록 (3×3 커널, 채널 수 16 및 32)을 거쳐 공간적 특징을 추출하였다. 이후 BiGRU 블록에서 50 차원의 시계열 특징을 추출하였으며, 양방향 구조를 통해 최종 100 차원 벡터로 변환하였다. 입력 구조 1 (손 ROI), 2 (얼굴 ROI), 4 (전체 영상)은 동일한 모델 구조를 적용하였다. 입력 구조 3 (다중 ROI)은 손 ROI 와 얼굴 ROI 를 각각 독립적으로 CNN-BiGRU 모듈에 입력한 후, 각각의 100 차원 특징 벡터를 병합하여

200 차원 벡터를 형성하였다. 이후 선형 계층을 통해 차원을 축소하여, 수축기 혈압 (systolic blood pressure, SBP)과 이완기 혈압 (diastolic blood pressure, DBP)을 mmHg 단위로 추정하도록 설계하였다. 모델 학습은 평균 절대 오차 (mean absolute error, MAE)를 손실 함수로, Adam 최적화 알고리즘 (학습률 0.001)을 사용하였다. 배치 크기는 16, 학습 에포크는 200 으로 설정하였다. 데이터셋은 피험자 단위로 분리하여 3-폴드 교차 검증을 수행하였으며, 12 명의 데이터 중 학습 8 명, 검증 2 명, 시험 2 명으로 구분하여 모델을 평가하였다.

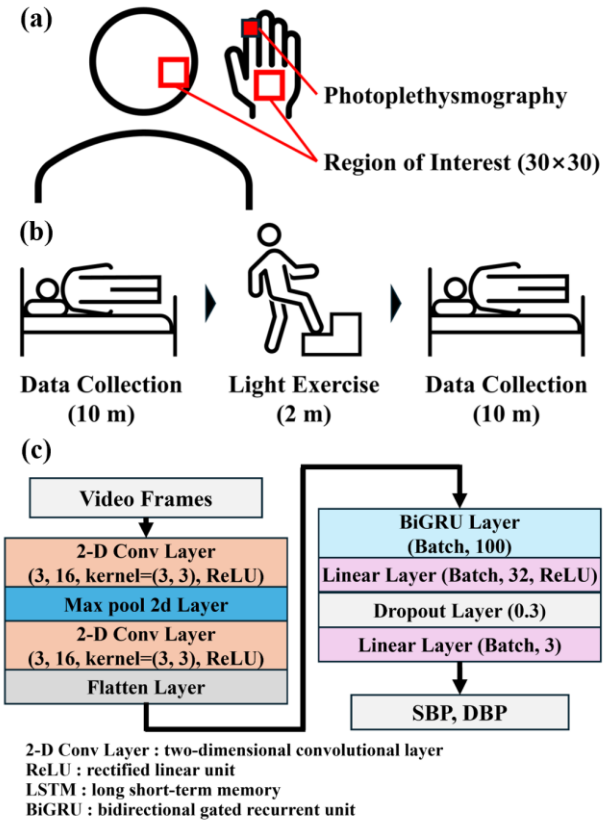


그림 1 실험 절차 개요도. (a) 실험 참가자 촬영 장면 및 ROI 예시; (b) 실험 과정 개요도; (c) 모델 구조도

라. 결과

표 1 은 입력 구조 별 시험 데이터셋 성능을 평균 절대 오차 (mean absolute error, MAE)로 평가한 결과이다. Hand ROI, Face ROI, Multi ROI 입력 모델은 각각 SBP 8.97, 8.96, 8.97 mmHg, DBP 8.55, 8.45, 8.45 mmHg 로 유사한 성능을 보였다. 반면 전체 영상 입력 모델은 SBP 2.93 mmHg, DBP 4.90 mmHg 로 가장 우수한 성능을 기록하였다.

Input data	SBP (MAE±STD)	DBP (MAE±STD)
Hand ROI	8.97 ± 00.32	8.55 ± 0.11
Face ROI	8.96 ± 06.63	8.45 ± 6.55
Multi ROI	8.97 ± 11.14	8.45 ± 6.54
All Frame	**2.93 ± 02.05**	**4.90 ± 4.35**

SBP: systolic blood pressure, DBP: diastolic blood pressure, MAE: mean absolute error, STD: standard deviation, ROI: region of interest

표 1 입력 구조에 따른 모델의 성능

표 2 는 전체 영상 입력 모델의 학습, 검증, 시험 데이터셋별 성능을 나타낸다. SBP MAE 는 각각 2.62, 2.89, 2.93 mmHg, DBP MAE 는 4.94, 4.86, 4.90 mmHg 로, 세 데이터셋 간 성능 편차가 작았다.

Data set	SBP (MAE±STD)	DBP (MAE±STD)
Train	2.62±2.12	4.94±4.38
Validation	2.89±1.98	4.86±4.33
Test	2.93±2.05	4.90±4.35

표 2 전체 영상 입력 모델의 데이터 세트 별 성능

Ⅲ. 논의

표 1 의 결과는 ROI 기반 모델보다 전체 영상 입력 모델이 더 효과적으로 학습되었음을 보여준다. ROI 방식은 수작업 지정 또는 별도 검출 알고리즘이 필요해 실제 적용에서 제약이 크다. 반면 전체 영상 입력 방식은 추가 절차 없이도 높은 정확도를 달성하여, 사용성과 성능 모두에서 우수함을 입증하였다. 이는 비접촉식 혈압 추정 기술의 실제 적용 가능성을 뒷받침한다.

다만 본 연구는 제한된 데이터셋에 기반한다는 한계를 가진다. 학습·검증·시험 데이터가 총 12 명으로 각각 8 명, 2 명, 2 명으로 구성되어 결과의 일반화 가능성이 충분히 확보되지 않았을 수 있다. 향후 연구에서는 본 연구 데이터 전수를 포함한 확장된 데이터와 공개 데이터셋을 활용하여 다양한 조건에서 성능을 검증할 필요가 있다.

Ⅳ. 결론

본 연구는 영상 기반 혈압 추정에서 ROI 설정 여부가 딥러닝 모델 성능에 미치는 영향을 평가하였다. 직접 수집한 30 명의 데이터 중 12 명을 활용하여 CNN-BiGRU 기반 모델을 학습한 결과, ROI 입력 모델들은 유사한 성능을 보였으나 전체 영상 입력 모델은 SBP 2.93 mmHg, DBP 4.90 mmHg 의 MAE 로 가장 우수한 성능을 기록하였다. 또한 학습, 검증, 시험 데이터셋 간 성능 차이가 작아 안정적인 학습을 확인할 수 있었다. 데이터 규모가 제한적이라는 한계는 존재하지만, 향후 확장된 데이터셋으로 검증할 경우 ROI 지정 과정을 생략한 영상 기반 혈압 추정이 정확성과 신뢰성을 동시에 확보할 수 있을 것으로 기대된다.

ACKNOWLEDGMENT

이 논문은 정부 (과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (NR070859)

참 고 문 헌

[1] S. Magder, “The meaning of blood pressure,” Critical Care, vol. 22, p. 257, 2018. doi: 10.1186/s13054-018-2171-1.

[2] A. S. Meidert and B. Saugel, “Techniques for non-invasive monitoring of arterial blood pressure,” Frontiers in Medicine, vol. 4, p. 231, 2018. doi: 10.3389/fmed.2017.00231.

[3] A. Charmoy, G. Würzner, et al., “Reactive rise in blood pressure upon cuff inflation: cuff inflation at the arm causes a greater rise in pressure than at the wrist in hypertensive patients,” Blood Pressure Monitoring, vol. 12 (5), p. 275, 2007. doi: 10.1097/MBP.0b013e3282c9ac9a.

[4] R. H. Goudarzi, S. Mousavi, and M. Charmi, “A Novel Method for Blood Pressure Estimation Using an Imaging Photoplethysmography (iPPG) Signal,” Sensors, vol. 22, no. 18, p. 7037, 2022. doi: 10.3390/s22187037.

[5] R. Meng and K. Li, “A blood pressure prediction method based on imaging photoplethysmography in combination with machine learning,” Biomedical Signal Processing and Control, vol. 62, p. 102328, 2020. doi: 10.1016/j.bspc.2020.102328.

[6] H. Cheng, J. Xiong, J. Chen, et al., “Deep learning-based non-contact iPPG signal blood pressure measurement research,” Sensors, vol. 23, no. 12, p. 5528, 2023. doi: 10.3390/s23125528.

www.koreaai.org

제6회 한국 인공지능 학술대회

The 6th Korea Artificial Intelligence Conference

KICS
한 국 통 신 학 회