인공지능 수어 인식 학습 모델의 성능 평가를 위한 사용자 집단 실증에 관한 연구

전기만

한국전자기술연구원

kmjeon@keti.re.kr

A Study on the User Group Validation for Performance Evaluation of AI Model Using Sign Language Recognition Learning

Jeon Ki Man

Korea Electronics Technology Institute

요 약

청각장애인들의 의사소통 수단인 수어는 국가별로 표준화 되어 있지 않을뿐더러 한국수어의 경우에도 사용지역이나 학습 형태, 신규 단어 및 문장의 생성 등으로 완벽한 표준화 실현 상태라 볼 수 없다. 장애인들도 일반인과 같은 최신 기술의 수혜를 받기 위해서는 기본적으로 의사소통이 필수이므로 수어 인식을 위한 인공지능 학습 모델 구현에 관한 연구는 다양하게 진행되고 있다. 본 논문은 수집된 수어영상 데이터를 기반으로 인공지능 학습을 통해 설계된 인식 모델의 실험적 성능평가가 아닌 실 사용자 대상으로 한 집단실증의 환경구성을 포함한 평가 과정과 실증 수행을 통한 평가의 분석결과를 설명한다.

Ⅰ. 서 론

정보통신 기술을 비롯한 IT융합 기술 발전으로 인해 사회 각 영역에서 인간의 편의 향상 개선과 안전한 삶을 위한 서비스가 제공되고 있으나, 사회적 약자들을 포함하여 정보기술 활용 적응도가 높지 않은 일반인의 경우 기술발전에 따른 수혜는 상대적으로 미흡할 수 있다. 청각장애인들의 경우 역시 수어가 의사소통의 매우 중요한 수단이지만 일상에서 지원되는 수어 서비스는 매우 제약적 이므로 인공지능 기술을 이용한 상호 의사소통 수단으로의 수어인식 모델이 활용될 경우 해당자의 편의성은 다소 상승할 것으로 기대된다. 딥러닝 기술의 발전으로 객체 인식, 동작 인식(Action Recognition), 얼굴 인식, 표정 인식(Face Expression Recognition), 자세 추정(Pose Estimation) 분야에서 큰 폭의 진전이 있었다. 그리고 각 분야의 연구 결과를 수어인식에 적용하려는 시도가 계속되고 있다[1]. 딥러닝을 수어 인식에 적용하기 위해서는 수어 동영상과 해당 영상에서 단어 혹은 수어소(手語素, chereme, 수어의 의미를 변별하는 가장 작은 단위로 음성언어의 음소에 대응)가 언제 등장하는지에 대한 주석이 필요하다. 하지만 주석 처리된 데이터 셋이 부족하여, 이를 극복하기 위해 weakly-supervised 방식 등을 제안하는 연구들이 있다[2].

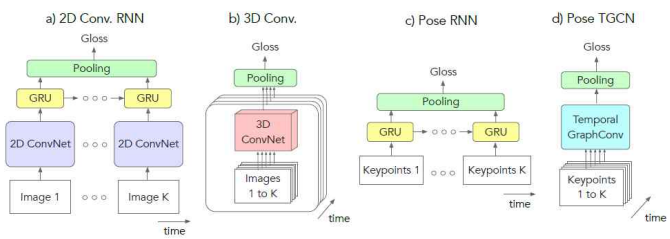

그림 1. WLASL 방법 비교에 관한 베이스라인 구조

해외 연구 사례에서도 대규모 단어 레벨의 수어 데이터 셋을 이용하여 여러 딥러닝 방법으로의 동작인식 성능평가를 시행하고 정확도 향상을 위한 제안이 확인되고 있으며, 그림 1은 전체적 시각 모습 방식과 2D 휴먼 포즈 방식의 비교에서 VGG backbone과 GRU 재학습에 의한 기준선을 제안하는 경우, 영상 포즈 추출을 통해 입력특징으로 사용 하는 경우 등의 단점을 보완하기 위한 temporal graph convolutional network(TGCN)을 제안하여 성능개선을 위한 시도하고 있는 내용을 나타낸다[3].

Ⅱ. 본론

본 논문에서는 기존 수행해 오던 수어 인식 모델 개발 과정 중 보유 데이터(Training, Test, Validation) 기반의 수어 인식에 관한 실험실 성능평가가 아닌, 실제 수어를 사용하는 청각장애인을 대상자로 수어 발화시 취득되는 신규 생성 영상 데이터를 통해 기존 학습 모델에 의해 추론된 단어나 문장을 출력하여 학습 모델에 대한 인식률을 평가하는 연구를 진행하였다.

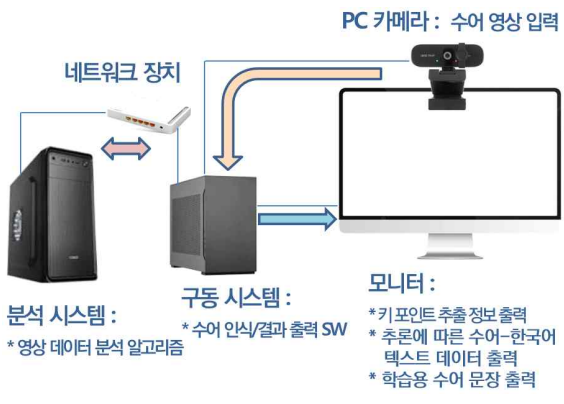

그림 2. 수어 인식 집단 실증 평가 환경

그림 2.와 같이 실험 환경은 PC 카메라를 통해 대상자의 발화 영상을 취득하여 구동시스템에 저장될 경우 분석 시스템의 영상 분석 알고리즘을 통해 수어 행위에 대한 인식 결과를 도출해 내게 되며 해당 결과를 텍스트 출력 SW를 통해 모니터 화면으로 출력하게 된다.

앞서 설명된 실험환경에서 사용자 집단실증 과정은 피 실험자 별로 카메라 앞에서 상반신과 양 손의 움직임이 확인되는 화각 내에서 측정되며, 피 실험자는 수어 인식 모델의 학습과정에 사용했던 데이터(숫자, 단어, 문장) 중에서 본인이 임의로 선택한 내용을 직접 발화하게 된다. 1단계로 선택한 숫자에 대해 최대 3회까지 발화 과정을 진행하며, 3회까지의 과정을 통해 의도한 의미가 출력되지 않을 경우 해당 인식 평가는 실패(Fail) 로 간주한다. 2단계와 3단계에서 역시 본인이 선택한 단어 및 문장에 대해 발화를 진행하고 3회 이전의 의도한 의미가 출력된 경우 합격(Pass), 상이한 의미가 출력된 경우 실패(Fail)로 판단하게 된다. 이와 동일한 실증 방법을 통해 총 30명의 피 실험자를 대상으로 수어인식 모델의 인식정확도 평가를 진행하여 출력된 정보와의 비교를 통해 인식률을 분석하는 방식으로 실험을 진행하였다.

그림 3. 청각장애인 수어 행위에 따른 학습모델의 인식 결과 비교 과정

그림 3은 일상에서 수어를 사용하는 청각장애인이 실시간 영상 취득을 위한 입력장치 정면에서 수어 발화를 진행하는 과정으로, 취득된 영상은 비교 데이터로 사용될 영상으로의 자동편집 과정을 거쳐 사람의 움직임(상반신)의 특징점을 생성하기 위한 키포인트 데이터를 추출하게 되고 분석 알고리즘을 통해 최종적으로 수어 인식에 대한 성공여부 결과를 출력해 주는 과정을 나타낸다. 본 과정을 통해 반복적 수어 발화에도 상이한 의미가 출력될 경우, 학습에 사용되었던 수어 영상을 선택적으로 재생하여 청각장애인 및 통역사가 사용하는 수어와의 유사도 여부 및 차이를 확인하여 해당 데이터에 대한 주석을 통해 향후 데이터 재가공 및 교체활용 등을 위한 분석 자료로 활용하게 된다.

표 1. 피 실험자 발화에 따른 학습모델 인식률 평가

no.	숫자	인식 결과	단어	인식 결과	문장	인식 결과
1	20	Pass	금요일	Pass	너무 아파요	Pass
2	8	Pass	월요일	Pass	술 취한 사람이 방망이로 사람들을 때리고 있어요	Pass
3	7	Pass	경찰	Pass	강도가 들어온것 같아요	Pass
4	77	Pass	개	Pass	밖에 모르는 사람이 서성거려요.	Pass
5	95	Fail	엄마	Fail	윗집에 불이 났어요	Fail
6	60	Fail	구해주세요	Pass	도와주세요!	Fail
7	80	Fail	배고프다	Pass	집에 불이 났어요	Pass
8	26	Pass	공원	Pass	너무 아파요	Fail
9	2	Fail	집	Fail	빨리 와주세요	Pass
10	102	Fail	시청	Pass	아이를 공원에서 잃어버렸어요	Pass
11	24	Pass	고속도로	Pass	고속도로에서 자동차가 갑자기 멈췄어요	Pass
12	60	Fail	화장실	Pass	아기를 낳을것 같아요	Pass
13	15	Fail	가렵다	Fail	빨리 도와주세요	Pass
14	3	Fail	가스	Fail	배가 침몰하고 있어요	Fail
15	20	Fail	오빠	Fail	강아지가 뼈를 삼켰어요	Fail
16	100	Pass	가스	Pass	폭탄이 터졌어요	Pass
17	5	Pass	가렵다	Fail	폭탄이 터졌어요	Pass
18	47	Pass	창백하다	Pass	지하철 역 안에서 화재가 발생했어요	Pass
19	65	Pass	구해주세요	Fail	할머니가 갑자기 쓰러지셨어요	Pass
20	60	Pass	내년	Pass	도와주세요!	Pass
21	77	Pass	배고프다	Pass	너무 아파요	Pass
22	50	Fail	누나	Pass	아파서 못 참을것 같아요	Pass
23	77	Fail	아생한사람	Pass	도와주세요!	Fail
24	85	Pass	할머니	Pass	(사람)이 갑자기 쓰러졌어요	Pass
25	2	Fail	비상약	Pass	도와주세요!	Pass
26	44	Pass	가렵다	Fail	빨리 와주세요	Fail
27	95	Fail	금요일	Fail	집에 불이 났어요	Pass
28	66	Pass	개	Pass	이상한 사람이 저를 쫓아와요	Fail
29	15	Pass	경찰차	Pass	집이 흔들려요	Pass
30	82	Pass	경찰	Pass	아이를 공원에서 잃어버렸어요	Pass

피 실험자 10명씩의 3차 실증을 시행하였으며, 1차 실증에서 숫자, 단어, 문장 총 30항목 중 20개 항목이 Pass, 10개 항목이 Fail 로 확인되었고, 2차, 3차 실증 모두 총 30항목 중 각각 11개, 9개 항목에서 Fail 이 확인되어 최종 인식률은 약 66.67%로 판단 할 수 있었다.

Ⅲ. 결론

본 논문의 목적은 보유된 수어 영상 데이터의 학습을 통해 개발된 수어 인식 인공지능 모델의 인식률 향상과 기능적 안정화이다. 본 실험결과를 통해 개발 중인 수어 인식 모델의 인식 성공률을 약 66.67%로 확인했으나, 대체로 낮은 인식율의 원인은 밝기, 촬영 각도 등의 환경적 요인뿐 아니라, 발화자의 발화속도, 학습데이터와 다른 청각장애인 사용 언어, 그리고 다의어 등의 의미 차이로 확인되었다. 이러한 문제를 개선하기 위해 보유 데이터들의 재가공, 신규 학습 데이터의 적용을 계획하고 있으며 발화이후 추론에 따른 결과 출력까지의 시간을 단축하기 위한 지속적 연구를 진행할 예정이다.

ACKNOWLEDGMENT

본 연구는 2021년도 산업통상자원부 및 산업기술평가관리원(KEIT)연구비 지원에 의한 연구임(20014406)

참 고 문 헌

[1] Razieh Rastgoo, Kourosh Kiani, and Sergio Escalera, "Sign language recognition: A deep survey," Expert Systems With Application, 164:113794, 2021.

[2] 이원재, 이한규, “인공지능을 활용한 수어 인식과 아바타 수어 서비스”, 한국콘텐츠학회 종합학술대회 논문집, pp.23-24, 2022.

[3] Dongxu Li, Cristian Rodriguez Opazo, Xin Yu, Hongdong Li, “Word-level Deep Sign Language Recognition fromVideo: A New Large-scale Dataset and Methods Comparison”, IEEE/CVF (WACV),pp.1459-1469,March, 2020.