산업제어설비 로그 정규화를 위한 AI 자동 매핑 및 휴먼인더루프(Human-in-the-Loop) 결합 모델에 관한 연구

오다희, 송현석, 박택근*

한전KDN 전력ICT연구원

{5_dahee_k, hyunseok.song.17, *reply_1997}@kdn.com

A Study on the AI-based Automatic Mapping Engine with Human-in-the-Loop for Normalization in Industrial Control Systems

Da-Hee Oh, Hyun-Seok Song, *Taek-Keun Park

KDN Electric Power ICT Research Institute

요 약

산업제어설비의 디지털 전환이 가속화됨에 따라 이기종 설비에서 발생하는 로그의 통합 관리 필요성이 증대되고 있다. 본 논문에서는 AI 기반 자동 매핑 엔진을 활용하여 로그 정규화 규칙을 생성하고, AI가 처리하지 못하는 예외 케이스를 사용자 GUI를 통해 보정하는 '휴먼인더루프(Human-in-the-Loop)' 결합 모델을 제안한다. 이를 통해 신규 설비 도입 시 정규화 규칙 생성 시간을 단축하고, 수동 보정 데이터를 피드백으로 활용하여 AI 모델의 정확도를 지속적으로 향상시킬 수 있는 선순환 구조를 구축하고자 한다.

Ⅰ. 서 론

산업 현장의 제어설비(PLC, DCS 등)는 제조사마다 로그 포맷이 상이하며, 설비 업데이트나 신규 장비 도입 시마다 정규화 규칙을 수동으로 생성해야 하는 번거로움이 있다. 기존의 정규표현식 기반 방식은 전문가의 개입이 필수적이며 유지보수 비용이 높다. 본 연구는 이러한 문제를 해결하기 위해 AI 학습 기반의 자동 매핑 엔진을 설계하고, 완전 자동화의 한계를 극복하기 위해 사용자 인터페이스(GUI)를 통한 보정 메커니즘을 결합한 하이브리드 정규화 방안을 제시한다.

비교 항목	기존 방식	제안 장식
규칙 생성 주체	보안 전문가(수동 작성)	AI + 전문가 보정
신규 설비 대응	신규 규칙 생성까지 공백	Few-shot으로 즉시대응
유지보수 비용	포맷 변경시 코드 수정	규칙DB 자산화 및 재학습
신뢰성 확보	휴먼 에러 가능성	인덱스 검증 및 알람

[표 1] 기존 방식과 제안 시스템의 비교

Ⅱ. 본론

2.1 T5 기반 Few-shot 학습을 통한 초기 매핑 가용성 확보

산업제어시스템 환경에서는 신규 설비 도입 시 참조할 수 있는 로그 데이터셋이 극히 제한적이다. 본 연구에서는 수만 건의 대규모 데이터 없이도 단 5~10건의 샘플 시퀀스만으로 매핑 규칙을 추론할 수 있는 T5 모델의 Few-shot learning 역량을 활용한다. T5 모델은 사전 학습된 언어적 문맥을 바탕으로, 낯선 로그 포맷에서도 IP, Timestamp, Status 등의 핵심 개체명을 식별해낸다. 이는 데이터 확보가 어려운 초기 설비 가동 단계에서도 최소한의 가용성을 확보하여 정규화 작업의 진입 장벽을 낮추는 핵심 기능이다.

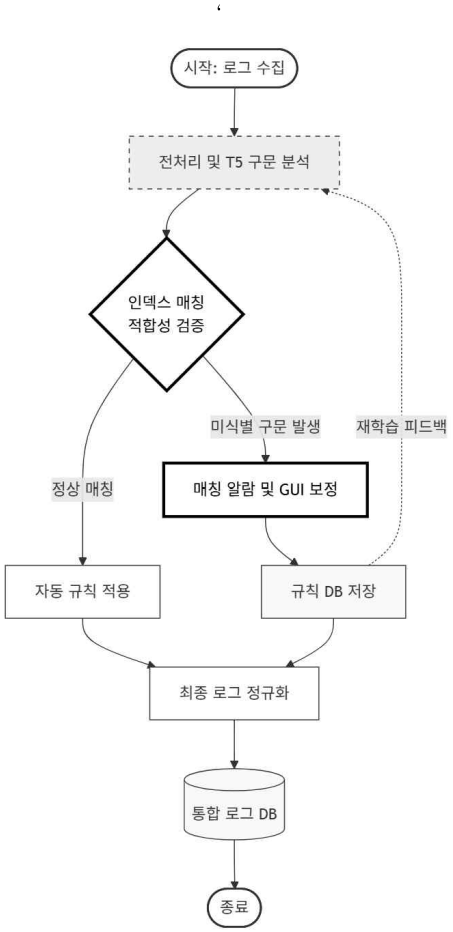
2.2 시맨틱 모호성 해소를 위한 전문가 참여형(HitL) 매핑 최적화

AI 모델이 필드를 식별하더라도, 보안 관제 관점에서는 데이터의 시맨틱 모호성이 치명적인 오판을 야기할 수 있다. 예를 들어, 단일 로그 내에 복수의 IP 주소가 존재할 때 AI는 이를 '출발지'와 '목적지'로 정확히 구분하지 못하고 단순 IP 필드로 통합 매핑할 위험이 있다. 본 논문에서 제안하는 GUI 인터페이스는 이러한 AI의 불확실성을 해결하는 최종 의사결정 계층 역할을 수행한다.

2.3 휴먼인더루프(Human-in-the-Loop) 피드백 메커니즘

사용자 GUI를 통해 교정된 매핑 정보는 정규화 규칙 데이터베이스에 즉시 자산화된다. 이 과정은 시스템의 운영 효율성을 극대화하는 두 가지 경로로 작동한다. 첫째, 규칙 기반의 즉각적 처리이다. 한 번 교정된 로그 포맷은 고유의 시그니처로 저장되어, 향후 동일한 포맷의 로그가 유입될 경우 AI 모델의 추론 단계를 거치지 않고 DB에 저장된 규칙에 따라 즉시 파싱된다. 이는 AI의 연산 자원을 절약함과 동시에 실시간 로그 처리에 있어 오차 없는 100%의 재현성을 보장한다. 둘째, AI 모델의 점진적 고도화이다. 전문가에 의해 검증된 정규화 데이터는 T5 모델의 파인튜닝 데이터셋으로 환류되어 모델의 가중치를 업데이트한다. 이를 통해 시스템은 아직 규칙 DB에 등록되지 않은 유사한 변종 로그가 발생하더라도, 전문가의 교정 패턴을 학습한 AI를 통해 이전보다 높은 정확도로 초기 매핑을 수행하게 된다. 결과적으로 본 메커니즘은 최소 1회의 전문가 검토, 규칙 DB 자동 등록, 동일 로그는 AI 개입 없이 즉시 통과의 선순환 구조를 형성하여 운영시간이 경과할수록 관리자의 개입 빈도를 지수적으로 낮추는 효과를 제공한다.

2.4 정규화 프로세스



[그림 1] T5 기반 로그 정규화 프로세스

본 논문에서 제안하는 로그 정규화 시스템은 T5 기반의 지능형 분석과 전문가의 보정 메커니즘이 결합된 하이브리드 구조를 가진다. 전체 프로세스는 크게 구문 분석, 인덱스 검증, 규칙 자산화의 3단계로 구성된다. 먼저, 이기종 설비로부터 수집된 비정형 로그 데이터는 전처리를 거쳐 T5 엔진에 입력된다. T5 엔진은 사전 학습된 문맥 이해 능력을 바탕으로 로그 내부의 구성 요소를 개체명 단위로 분리하고, 각 필드에 적합한 인덱스 구조를 생성한다. 이는 데이터가 부족한 초기 설비 도입 환경에서도 T5의 Few-shot 학습 능력을 통해 유연한 구문 분석을 가능케 한다.

검증 유형	판단 기준	위협 요소
구조적 필수 요소 누락	표준 스키마의 필수 인덱스 미식별	핵심 정보(시간, IP) 누락으로 분석 불가
데이터 타입 부정합	추출 필드와 스키마 정의 타입 불일치	데이터 왜곡
카디널리티/순서 위반	필드 중복 추출 및 논리적 순서 역전	로그 포맷 변조 가능성 및 파싱 신뢰도 저하

[표 2] 인덱스 매칭 부적합 판단 기준

생성된 인덱스 구조는 시스템의 표준 스키마와 비교되는 인덱스 매칭 적합성 검증단계를 거친다. T5 모델이 추출한 로그 엔티티 시퀀스는 사전에 정의된 표준 스키마 인덱스와 비교되며, [표 2] 정합성 규칙을 하나라도 위반할 경우 시스템은 매칭 실패로 간주하고 관리자에게 알람을 발생시킨다. 이 알람은 사용자 GUI로 전달되어 보안 전문가가 시각화된 인터페이스 상에서 미매칭 필드를 직접 교정하고 확정할 수 있는 환경을 제공한다.

2.5 시스템의 특장점 및 기대효과

제안된 T5 기반 엔진은 기존의 규칙 기반 시스템과 달리, 특정 로그 포맷에 종속되지 않은 생성형 모델의 특징을 가진다. 이는 제조사가 다른 다양한 설비가 혼재된 산업 현장에서 정규화 규칙을 생성할 때, 개발자의 직접적인 코드 수정 없이도 모델의 추론만으로 대응할 수 있는 확장성을 제공한다.

AI 단독 시스템이 가질 수 있는 오탐의 위험성을 사용자 GUI 보정 단계로 보완함으로써 보안 관제의 핵심인 데이터 무결성을 확보한다. 특히 미학습된 신규 로그 포맷이 유입될 경우, 전문가의 즉각적인 교정을 통해 보안 공백을 최소화할 수 있다는 점에서 신뢰도 높은 보안 관제 환경을 제공한다.

사용자의 교정 데이터가 다시 모델의 재학습 데이터로 피드백되는 구조는 현장 운영자의 도메인 지식을 AI 모델에 지속적으로 내재화하는 효과를 가진다. 이는 시간이 흐를수록 시스템의 자동화율을 높여 운영자의 업무 부하를 실질적으로 경감시킬 것으로 기대된다.

Ⅲ. 결론

본 연구에서는 이기종 산업제어설비 로그의 효율적인 통합 관리를 위해 T5 모델과 전문가 피드백 루프가 결합된 지능형 정규화 아키텍처를 제안하였다. 제안된 시스템은 T5 모델의 Few-shot 학습 능력을 활용하여 데이터가 부족한 신규 설비 도입 초기에도 유연한 구문 분석을 수행하며, 특히 인덱스 매칭 적합성 검증 단계를 통해 AI의 추론 오류를 실무적으로 통제한다. 미식별 구문 발생 시 발송되는 매칭 요청 알람과 사용자 GUI 보정 메커니즘은 보안 데이터의 무결성을 보장하는 핵심 장치로 작용한다. 또한 확정된 규칙을 DB화하여 동일 로그에 대해 AI 개입 없이 즉시 처리하는 구조는 시스템의 운영 효율성을 극대화하고 보안 전문가의 업무 부하를 실질적으로 경감시켰다는 점에서 의의가 있다.

본 연구는 아키텍처의 설계와 논리적 타당성 검증에 집중하였으나 다음과 같은 한계점을 갖는다. 첫째, 실험에 사용된 로그 데이터가 실제 운영 환경의 대규모 트래픽을 완전히 대변하기에는 규모 면에서 제한적이다. 둘째, T5 모델의 언어적 추론 능력에 의존함에 따라 극도로 비정형화된 특수 제조사의 로그나 바이너리 형태의 데이터 처리에는 추가적인 전처리 로직이 요구된다는 점이 확인되었다.

향후에는 실제 로그 스트림 환경에서 본 모델을 적용하여 시스템 처리 지연 시간과 처리량을 정량적으로 검증할 계획이다. 사용자의 보정 행위를 정교하게 학습하기 위해 강화학습 기법을 도입하고, 다양한 산업 표준 프로토콜에 최적화된 모델 연구를 병행하여 범용 정규화 엔진으로 고도화하고자 한다.

참 고 문 헌

[1] C. Raffel et al., "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.

[2] X. Wu et al., "Human-in-the-loop machine learning: a state of the art survey," IEEE Transactions on Cybernetics, 2021.

[3] J. Zhu, S. He, J. Liu, and M. R. Lyu, "Tools and Benchmarks for Automated Log Parsing," 2019 IEEE/ACM 41st International Conference on Software Engineering, 2019, pp. 121-130, doi: 10.1109/ICSE-SEIP.2019.00021.

[4] S. He et al., "A Survey on Automated Log Analysis for Reliability Engineering," ACM Computing Surveys (CSUR), vol. 54, no. 6, pp. 1-37, 2021.