

Tiny LLM 기반 Two-stage 미세 조정 전략을 활용한 점용접 공정 자율 제어 지시 생성 프레임워크에 관한 연구

김계은^{1, 2}, 허영순¹, 김거식¹, 신지수^{1, 2}, 박형준¹, 김진술^{2*}

한국전자통신연구원¹, *전남대학교²

kyeeun_kim@etri.re.kr, *jsworld@jnu.ac.kr

A Study on the Autonomous Control Instruction Generation Framework for Spot Welding Processes utilizing Tiny LLM-based Two-Stage Fine-Tuning

Kyeeun Kim, Yeongsoon Heo, Keosik Kim, Ji-soo Shin, Hyoungjun Park, Jin Sul Kim*

Electronics and Telecommunications Research Institute, *Chonnam National Univ.

요약

본 논문은 기존 규칙 기반 제어의 한계를 극복하고, 엣지 디바이스에 활용하기 적합한 Tiny LLM(TinyLlama-1.1B) 기반 점용접 공정 이상 탐지 및 자율 제어 지시 생성 프레임워크를 제안한다. 제한된 산업 데이터 환경에 최적화된 QLoRA 기반 2단계 미세 조정(Two-Stage Fine-Tuning) 전략을 도입하여 모델을 특화시킨다. 이 전략은 합성 데이터 사전 학습(Stage 1)을 통해 광범위한 이상 패턴 커버리지를 확보하고, 실데이터 정교화(Stage 2)를 통해 현장 적합성을 높이는 것을 목표로 한다. 본 연구는 LLM을 활용한 자율적이고 안전이 검증된 제어 지시 생성의 실현 가능한 경로를 제시하며, 향후 성능 최적화를 위한 연구의 기반을 제공하는데 의의가 있다.

I. 서론

자동차 차체 조립의 점용접 공정은 전류, 전압, 가압, 통전시간 등 4대 변수의 균형에 따라 품질이 결정된다. 기존 규칙 기반 제어는 다양한 이상 상황에 대한 적응적이고 정량적인 제어 지시에 한계가 있다.

본 연구는 저자원 LLM(TinyLlama-1.1B)를 활용하여 용접 공정 이상을 진단하고, 안전 제약 조건을 준수하는 정형화된 JSON 제어 지시를 생성하는 에이전트 프레임워크를 제안한다. 제안한 프레임워크는 QLoRA 기반 Two-stage 미세 조정 및 Policy Guard로 구성되며, 이는 LLM의 비결정적 출력이 산업 제어 환경에 맞는 안전한 범위에서 동작할 수 있도록 통합하는 기능을 포함한다.

II. 데이터셋 및 2단계 SFT 전략

2.1. 데이터 표준화 및 JSONL 구축

본 논문에서 사용한 데이터셋은 KAMP(Korea AI Manufacturing Platform)의 '용접기 AI 데이터셋(KAIST(울산과학기술원), (주)피엠플루션즈)'이며, 점용접기에서 수집된 로그 데이터로 구성된다.

본 데이터셋의 점용접 로그는 한글과 영어가 혼재되어 있고 단위 기재가 섞여있다. Raw data는 Robust reader로 인코딩(utf-8-sig/utf-8/cp949/ms949/euc-kr 등)한 후, 구분자를 자동으로 탐지하여 각 feature를 로드하였으며, 이후 헤더를 정규화하여 force/current/voltage/time 4열로 자동 정규화하였다.

특히, 0-1 스케일로 정규화된 데이터의 경우, 안전범위(I: 12-18 kA, V: 1.5-3.5 V, F: 1-12 bar, T: 30-120 ms)를 기준으로 실단위로 역스케일하여 모델의 입력과 출력이 현장 제어값과 일치하도록 처리하였다.

2.2. Two-Stage QLoRA 학습

본 프레임워크는 그림 1과 같이 LLM의 효율적인 도메인 특화를 위해 two-stage 전략으로 학습을 수행한다.

1. Stage-1(합성데이터 기반 사전학습): 합성된 SFT(train_aug)을 사용하여 2-3 epoch 동안 사전학습을 수행한다. 이는 LLM에게 용접 공정 데이터의 일반적인 분포와 이상 패턴에 대한 커버리지를 빠르게 학습 시키는 것을 목표로 한다.
2. Stage-2(실데이터 기반 미세 조정): 실데이터 SFT를 사용하여 1-2 epoch 동안 학습을 수행한다. 이때 학습률은 Stage-1 대비 0.5배로 모델을 파라미터를 정교하게 조정한다. 검증에는 실데이터 val만을 사용하여 모델이 현장 분포에 적합하도록 유도한다.

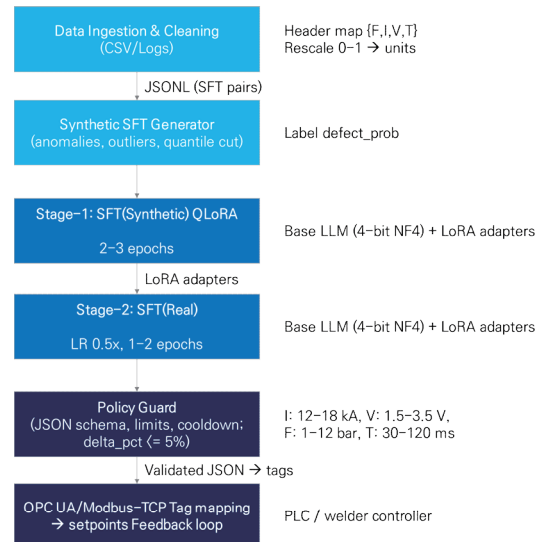


그림 1. 전체 파이프라인 블록도

표 1. 2단계 SFT 예시

단계(Stage)	모델	데이터	목표	Learning Rate
Stage 1 (Pre-training)	TinyLlama-1.1B	합성 데이터 (2,400 건, 이상 비율 30%)	폭 넓은 이상 패턴 커버리지 확보	2×10^{-4}
Stage 2 (Refinement)	Stage1 모델	실데이터 (Train 6,143 건)	실제 데이터의 경계 학습 및 JSON 출력 정교화	1×10^{-4} (0.5x)

표 2. 학습 하이퍼파라미터

항목	하이퍼파라미터
기본 모델	TinyLlama-1.1B
미세 튜닝	LoRA(r = 16, alpha = 32, 4-bit NF4)
배치	2, 유효 배치 16(grad-accum = 8)
하드웨어	NVIDIA RTX 3090(24GB VRAM)

III. 실험 및 결과

3.1. 실험 셋업 및 검증 환경

본 연구는 저자원 산업 환경 적용의 효율성을 입증하기 위해 Tiny LLM 계열인 TinyLlama-1.1B 모델 기반으로 QLoRA 미세 조정을 수행했다. 학습은 Two-Stage 전략에 따라 진행하였으며, 합성 데이터 사전 학습(Stage 1)을 통해 용접 공정의 일반적인 이상 커버리지를 확보하고, 실데이터 정교화(Stage 2)를 통해 현장 분포 적합성을 높이는 데 중점을 두었다. 최종 모델의 검증은 실데이터 기반의 val셋(총 1,535 건)을 사용하여 현장 적용 관점에서 평가하였다.

3.2. 정량적 평가 결과 및 성능 분석

Stage 2 학습 완료 후, 검증 셋에 대한 정량적 평가 결과는 LLM 기반 에이전트의 실질적인 성능과 안정성을 입증한다. 실험 결과 모델의 정확도는 0.7186 이었다.

표 3. 학습 결과 예시

Classification Report	Precision	Recall	F1-Score	Support
Class 0 (Normal)	0.8802	0.7050	0.7829	1105
Class 1 (Anomaly)	0.4985	0.7535	0.6000	430

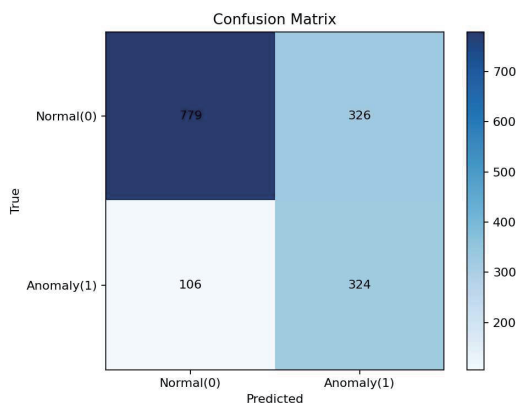


그림 2. Confusion matrix

학습 결과는 다음과 같이 해석할 수 있다.

- 정상 탐지 성능(Class 0): 모델은 정상 샘플(normal) 1,105건 중 779건을 정상으로 예측(TP)했으며, 정밀도(Precision)는 0.8802, 재현율은 0.7050 이었다.
- 이상 탐지 성능(Class 1): 이상 샘플(anomaly) 430건 중 324건을 이상으로 탐지(TN)하여 정밀도(Precision)는 0.4985, 재현율(Recall) 0.7535 이었다.
- 미세 조정된 모델은 이상 탐지 성능에 대한 재현율을 극대화하는 방향으로 분류 경계가 설정되도록 학습된 것으로 생각되며, 이는 이상(anomaly)을 정상(normal)으로 오판(FP)하는 사례를 줄이는데 중점을 둔 모델로 해석되었다.

IV. 결론

본 연구는 점용접 공정 데이터에 저자원 LLM(TinyLlama-1.1B)과 QLoRA 기반 Two-Stage 미세 조정 전략을 적용하여, 기존 규칙 기반 제어의 한계를 극복하는 자율 제어 지시 생성 에이전트 프레임워크를 제안하고 성능을 검증하였다.

Two-Stage SFT 학습을 통해 LLM은 요청된 JSON 스키마를 산업 제어 환경에 맞도록 안정적으로 출력을 생성하였으며, 이는 Policy Guard와 연동하여 설비 제어 페루프에 통합함으로써 구조적 타당성을 입증하였다. 또한, 경량 모델을 사용했음에도 불구하고 ROC-AUC 0.7440을 달성했으며, 특히 산업 안전에서 중요한 이상 클래스(Class 1)의 재현율을 0.7535로 달성함으로써 실제 현장에 적용될 수 있는 가능성을 확인하였다.

향후 연구는 LLM의 탐지 성능을 획기적으로 개선하는 데 집중될 것이다. 특히, 부정확한 규칙 기반 라벨링으로 인해 발생할 수 있는 모델 편향 문제를 해결하기 위해, 고품질 전문가 라벨링의 보강과 손실 함수에서의 클래스 가중치 적용이 필수적이다. 본 프레임워크는 이러한 성능 최적화 연구의 기반이 될 것이며, TinyLLM의 잠재력을 활용하여 실제 산업 현장에서의 엣지 디바이스에서 자율 제어 효율성을 극대화할 수 있을 것으로 기대한다.

ACKNOWLEDGMENT

본 연구는 한국전자통신연구원 연구운영지원사업의 일환으로 수행되었음. [25ZK1100, 호남권 지역산업 기반 ICT 융합기술 고도화 지원사업] 중소벤처기업부, Korea AI Manufacturing Platform(KAMP), 용접기 AI 데이터셋, KAIST(울산과학기술원, ㈜이피엠텔루트즈), 2020.12.14., www.kamp-ai.kr

참 고 문 헌

- [1] Yi Jiang, Lu Liu, Gang Feng, "TFully distributed adaptive control for output consensus of uncertain discrete-time linear multi-agent systems," Automatica, Vol. 162, 2024. 111531.