

LLM 기반 Adaptive Few-Shot Learning 을 통한 산업용 구동기 다중 고장 분류

송지혁¹, 유현창²

¹ 고려대학교 SW.AI 융합대학원 인공지능융합학과 석사과정

² 고려대학교 정보대학 컴퓨터학과 교수

jihy1325@gmail.com, yuho@korea.ac.kr

Multi-Fault Classification of Industrial Actuators Using LLM-based Adaptive Few-Shot

Jihyuk Song¹, Heonchang Yu²

¹Dept. of Applied Artificial Intelligence, College of SW.AI Convergence, Korea University

²Department of Computer Science & Engineering, College of Informatics, Korea University

요 약

본 연구에선 LLM 모델에 In-Context Learning 을 적용하여 구동기의 전류 데이터를 통해 이상 탐지 분류 시스템을 제안한다. 특히 Gemma-3-4B,12B 를 활용하여 정상, 축정렬 불량, 베어링 불량을 분리한다. 데이터 전처리로는 Hampel Filter 를 사용하여 전류에서 나오는 노이즈를 제거 하였고, 이 과정 후 Silding Window 중 특성을 가장 잘 표현한 $L = 21$ 값을 선택하여 In-Context Learning 에 적용하였다. 기존 Sliding Window 데이터를 넣기에는 토큰 수의 한계가 있어, Window 의 통계적 특성을 뽑아 적용하였다. In-Context Learning 의 방법인 Zero-Shot, One-Shot, Adaptive Few Shot 을 활용하여 정량적인 방법으로 비교 실험 하였다. 실험 결과 Adaptive Few Shot 을 적용 시 Zero Shot, One Shot 보다 성능이 유의미하게 개선되었다.

1. 서론

기술이 발전함에 따라 구동기의 활용 범위는 다양한 산업 전반으로 빠르게 확대되고 있다. 대표적인 예로 공장 자동화 설비, 로봇 팔, 엘리베이터, 에스컬레이터, 공조기 등 여러 분야에서 구동기의 사용량이 지속적으로 증가하고 있다[1]. 그러나 이러한 구동기의 고장 원인을 정확히 파악하는 것은 여전히 쉽지 않은 과제이다. 구동기에 문제가 발생할 경우 생산라인 중단이나 시설 운영 차질로 이어져 이용자의 불편과 더불어 상당한 산업적 비용 손실을 초래할 수 있다. 따라서 고장 원인을 신속하게 진단하고 적절히 조치하는 것이 점점 더 중요해지고 있으며, 이러한 맥락에서 효과적인 이상 탐지 기법의 필요성이 크게 부각되고 있다.

기존의 이상 탐지 연구에서는 CNN, GNN, LSTM 과 같은 딥러닝 기법과 전통적인 머신러닝 방식이 널리 활용되어 왔다. 그러나 이러한 접근법은 고장 유형마다 별도의 모델을 구축해야 하고, 정확한 고장 명칭에 따른 분류가 쉽지 않다는 한계를 지닌다. 대부분 정상과 비정상만을 구분하는 수준에 머물러 실제 적용에 제약이 크며, 많은 자원을 소모한다는 점에서

효율성도 떨어진다. 이에 따라 보다 효과적이고 실증적으로 활용 가능한 새로운 모델의 필요성이 제기되고 있다.

본 연구에서는 LLM(Large Language Model)을 기반으로 In-Context Learning(ICL)의 Adaptive Few-Shot 을 활용한 구동기 이상 탐지 기법을 제안한다. 실제 현장에서 수집한 구동기 데이터를 활용하여 제안 기법을 평가하였으며, 기존 방법과 달리 다양한 고장 유형과 정상 상태를 구분할 수 있음을 확인하였다.

2. 관련 연구

이상 탐지 연구는 전통적인 머신러닝 기법과 딥러닝 기법이 모두 활발히 연구가 이루어지고 있으며, 두 접근 방식은 발전 과정 속에서 상호 보완적으로 활용되고 있다. 초기 연구 단계에서는 SVM, Random Forest, Isolation Forest 등과 같은 전통적인 머신러닝 방법이 주로 사용되었다. 이들은 특징 추출 과정을 통해 비교적 단순한 구조로도 이상 여부를 판단할 수 있었으며, 해석 가능성이 높고 계산 자원이 제한된 환경에서도 효율적으로 적용될 수 있다는 장점이 있었다. 그러나 데이터의 규모와 복잡성이 점차 증가하고, 센

서와 IoT 기기에서 발생하는 비정형 데이터가 본격적으로 등장하면서 전통적 방법은 특징 설계와 일반화 능력에서 한계를 드러내기 시작하였다.

이러한 한계를 극복하기 위해 하드웨어의 성능 발전과 함께 딥러닝 방법론들이 등장을 하며 이상탐지 분야에서도 주목하고 있다[2]. CNN 은 이미지나 공간적 패턴을 학습하는 데 강점을 보였고, LSTM-AE 와 같은 순환 신경망 기반 모델은 시계열 데이터의 장기 의존성을 포착하여 이상 탐지 성능을 크게 향상시켰다. 최근에는 그래프 구조 데이터의 확산과 함께 그래프 신경망(GNN)이 적용되어 복잡한 관계형 데이터에서의 이상 탐지 가능성을 넓히고 있으며, 이는 기존의 시계열-벡터 기반 접근이 다루기 어려운 영역까지 확장하는 중요한 흐름으로 평가된다.

3. 데이터 전처리

<표 1> 구동기 부하별 데이터

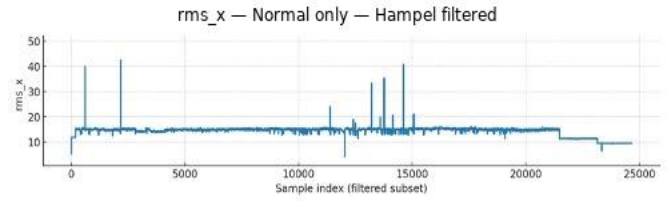
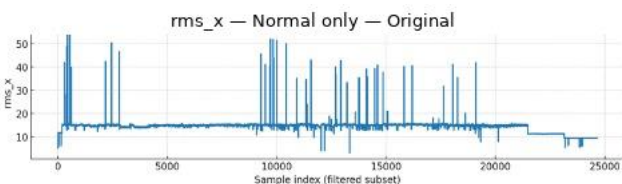
(단위: 건 이상)

부하별(KW)	구동량(전류/진동)					정상 데이터
	회전체 불평형	벨트 느슨함	축정렬 불량	베어링 고장	소계	
2.2	20,000	60,000	20,000	40,000	140,000	100,000
3.7	20,000	30,000	20,000	40,000	110,000	100,000
3.75	-	-	20,000	-	20,000	100,000
5.5	20,000	30,000	20,000	40,000	110,000	100,000
7.5	-	30,000	20,000	40,000	90,000	100,000
11	20,000	30,000	20,000	40,000	110,000	100,000
15	20,000	30,000	-	40,000	90,000	100,000
18.5	-	-	-	40,000	40,000	100,000
22	20,000	30,000	20,000	-	70,000	100,000
30	-	-	20,000	-	20,000	100,000
37	-	-	20,000	-	20,000	100,000
55	20,000	30,000	-	-	50,000	100,000
합계	140,000	270,000	180,000	280,000	870,000	1,200,000

본 연구에서 활용한 데이터는 AI Hub 에서 제공된 5.5KW~ 30KW 구동기 데이터[3] 중 안의 전류 데이터이다. 이 가운데 축정렬 불량, 벨트 느슨함, 정상 상태에 해당하는 데이터를 사용하여 실험을 진행하였다. <표 1>과 같이 고부하 구간(37KW 이상)은 고장 표본수가 적고 고장유형이 한정적이어서 데이터 편향 가능성이 커 제외 하였다. 또한, 저부하 구간(2.2KW~3.75KW)은 부하 프로파일로 인해 도메인 시프트(covariate shift)가 크므로 제외 하였다.

전류 데이터는 노이즈 발생하는 경우 정상과 이상 상태를 구분하기 힘들어 전처리 과정에서 노이즈를 적절히 제거하는 것이 중요하다. 또한, 노이즈를 효과적으로 처리할 수 있는 모델을 선택하는 것은 추론 과정에서 편향성을 줄이는 핵심적 방법이다.

3.1 Hampel Filter 기반 전처리

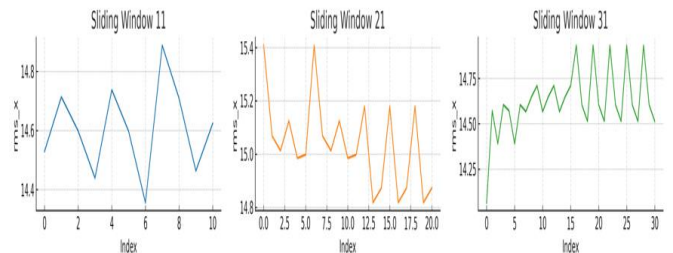


(그림 1) 정상 라벨의 RMS X 값의 원래 데이터와 Hampel Filter 적용 후 데이터.

Hampel filter 는 중앙값(median)과 중앙값 절대편차를 이용해 국소 윈도우에서의 급격한 스파이크, 이상치를 견고하게 제거하는 로버스트 통계 기반 필터이다. 각 시점 t 에 대해 길이 $2w+1$ 의 이동 창 $\{x_{t-w}, \dots, x_{t+w}\}$ 의 중앙값 m_t 와 $MAD_t = \text{median}\{|x_i - m_t| : i \in [t-w, t+w]\}$ 을 계산하고, 정규분포 가정 하의 스케일 보정계수 $c=1.4826$ 을 곱한 $\hat{\sigma}_t = c \cdot MAD_t$ 를 사용하여 $|x_t - m_t| > k \cdot \hat{\sigma}_t$ 이면 x_t 이상치로 판정한다. 이 방식은 평균, 표준편차 대신 중앙값, 중앙값 절대편차를 사용하므로 극단값의 영향에 둔감하여 노이즈에 강건하며, 결과적으로 학습 단계에서 스파이크나 이상치에 의한 편향 유입을 줄이고 모델이 국소적 잡음 패턴에 과도하게 적합하는 과적합을 완화하는 효과가 있다[4].

(그림 1)에서 보는 것과 같이 Hampel filter 를 적용하기 전에는 정상 데이터에서도 많은 노이즈들이 발생한 것을 확인 할 수 있다. 이는 데이터의 표준편차가 늘어나게 되어 과대 적합을 유발할 수 있다. 전류 시계열의 특성(잡음과 간헐적 스파이크)을 고려하여 Hampel filter 로 탐지된 이상치는 중앙값으로 대체하거나 결측치로 표기한 후 후속 보정 단계에서 처리하게 된다. 이 전략은 이상치가 섞인 값을 직접 학습에 사용하지 않게 하여 추정 편향을 더 효과적으로 줄여준다.

3.2 Sliding Window



(그림 2) Hampel 전처리 후 RMS_X 의 Sliding Window 별 (11,21,31) 비교.

해당 데이터를 단일 시점 값으로 적용이 신호의 형태적인 특성을 반영 할 수 없기 때문에 패턴을 보존하는 방법인 Sliding Window 로 단일 형태의 데이터를 여러 시계열 적인 형태로 바꾸는 것이 중요하다. 3.1 에서 적용된 Hampel filter 로 전처리한 시계열 데이터를 Sliding Window 를 $L \in \{11,21,31\}$ 로 나눠서 반영하여 Window 별 시각화 하여 비교 하였다. 시계열 데이

터의 특성인 연속성을 유지하기 위해 Sliding Window 적용 할 때 Overlay 를 50%로 적용하여 Window 를 구성하였다.

(그림 2)에서 보는 것과 같이 $L = 11$ 의 경우에는 고주파의 요동이 과도하게 포함되어 특징 표현이 불안정 하다. 이에 비해, $L = 31$ 의 경우에는 특징은 뚜렷하게 보이지만, 창이 커지게 되면서 미세 곡률이 줄어들게 되고, 세부 곡률이 평탄화 되어 유효한 데이터의 수도 줄어들게 된다. 반면, $L = 21$ 은 노이즈 억제와 세부 곡률 보존 사이의 균형이 가장 우수하고, 유효한 Window 가 유지 되는 모습을 (그림 2)에서 확인 할 수 있다.

4. In-Context Learning (ICL) 실험 방법

해당 실험을 위해 베이스 모델로 Google 의 Gemma-3 4B 와 12B 를 활용하였다. 해당 모델이 다른 모델들에 비해 토큰 수가 적음에도 불구하고 Chatbot Area Elo Score 에서 높은 점수를 받았다[5]. 또한, 해당 모델은 4B 부터 Prompt 에서 토큰을 128k 를 처리 할 수 있어, Fine Tuning 없이 ICL 을 적용할 때 더 많은 Shot 데이터를 적용할 수 있으며, 이를 통해 추론 정확도를 향상시킬 수 있다. 본 연구에서는 이러한 특성들을 고려해 Gemma 모델을 선택하고, ICL 적용 시 Shot 의 수에 따라 Prompt 를 달리 구성하였다.

Prompt 설계 시 모델 이해를 돕기 위해 정상과 고장 데이터를 각각 0, 1, 2(0=정상, 1=베어링 불량, 2=축정렬 불량)로 변환하였다. 원본 Window 데이터를 그대로 입력하면 토큰 소모가 커지고 Shot 예시 수에 제한이 생기므로, 각 Window 별로 mean, std, max 등 주요 통계치 5 개를 추출하여 입력 값으로 사용하였다.

4.1 Zero Shot

Zero Shot (K=0)
RMS 통계 분류 (0=정상,1=베어링 불량,2=축정렬 불량). 예시는 제공하지 않는다. 한 숫자만 출력.
--- 질문 ---
[D]: <질문 윈도우의 15개 통계>
[I]:

(그림 3) Zero Shot Prompt.

Zero Shot 의 경우에는 어떠한 예시를 주지 않고 LLM 이 직접 값을 보고 추론하여 분류를 하는 것이다. (그림 3)에서 보는 것과 같이 어떠한 예시를 주지 않고 한 분류를 하게 만드는 것이다. 이것은 LLM 이 추론하고 그냥 찍는 것과 다름이 없어 일관성 있는 결과를 보여줄 수 없다.

4.2 One Shot

One Shot (K=1)
RMS 통계 분류 (0=정상,1=베어링 불량,2=축정렬 불량).
--- 예시 ---
[D]: <정상 샷#1 통계>
[I]: 0
[D]: <베어링 샷#1 통계>
[I]: 1
[D]: <축정렬불량 샷#1 통계>
[I]: 2
--- 질문 ---
[D]: <질문 윈도우의 15개 통계>
[I]:

(그림 4) One Shot Prompt.

One Shot 은 예시를 한 개를 Prompt 에 넣어 LLM 이 예시를 통해 분류를 추론하는 것이다. (그림 4)에서 보는 것과 같이 각 통계치 RMS_X, Y, Z Window 통계치 각 5 개씩 들어가 15D 를 만들고 각 분류별 한개씩의 예시가 들어간다. 이를 통해 LLM 이 추론하여 분류를 하도록 했다.

4.3 Adaptive Few Shot

Adaptive Few Shot (K=3)
RMS 통계 분류 (0=정상,1=베어링 불량,2=축정렬 불량).
--- 예시 ---
[D]: <정상 샷#1 통계>
[I]: 0
[D]: <베어링 샷#1 통계>
[I]: 1
[D]: <축정렬불량 샷#1 통계>
[I]: 2
..... 각 통계치 3개씩 선택
[D]: <축정렬불량 샷#3 통계>
[I]: 2
--- 질문 ---
[D]: <질문 윈도우의 15개 통계>
[I]:

(그림 5) Adaptive Few Shot Prompt.

Few Shot 의 대표 값을 뽑기 위해서는 Adaptive Few Shot(이하 ‘AFS’)로 이용하여 Prompt 를 작성 하였다. 먼저 학습 집합에서 클래스 균형 지원 풀을 만들고, 토큰과 탐색 비용을 줄이기 위해서 클래스당 최대 4000 개의 무작위로 서브 샘플링을 하였다[6]. 이후 풀 통계 (μ_p, σ_p)로 15 차원 통계를 Z-점수 표준화 하여 $\hat{f}$ 로 변환한 후, 제공된 윈도우의 통계의 아래 유클리드 거리로 근접도를 계산한다.

$$d(\hat{f}_i, \hat{f}(q)) = |\hat{f}_i - \hat{f}(q)|_2 = \sqrt{\sum_{j=1}^d (\hat{f}_{ij} - \hat{f}_{qj})^2}$$

$$S_c(q) = \arg \min_{S \subset \mathcal{P}_c, |S|=K} \sum_{(\hat{f}_i, y_i) \in S} |\hat{f}_i - \hat{f}(q)|_2^2$$

각 클래스 $c \in \{0,1,2\}$ 에서 거리 기준 K -최근접 이웃을 선택하여 $S_c(q)$ 를 얻고, 이들을 합쳐 총 $3K$ 개의 예시를 Prompt 에 사용한다. 계산 효율을 위해 제공거리로 동일 순위를 사용한다.

Prompt 는 “지시문 $\rightarrow$ 예시($K=3$) $\rightarrow$ 질문”의 고정 템플릿을 따르며, $0 \rightarrow 1 \rightarrow 2$ 클래스 순으로 교차(interleave) 배치해 위치 편향을 줄였다. 추론 Gemma-3-12B-IT 와 Gemma-3-4B-IT 에 빠른 속도와 prompt 의 토큰 수를 초과하지 않게 $K=3$ 를 기본으로 설정하였다. 또한, 로짓 마스킹을 적용하여 출력 토큰을 $\{0,1,2\}$ 로 제한하고, greedy 로 한 자리(max new tokens=1)만 생성하여 최종 라벨로 채택한다. 이런 AFS 구성은 질의 적응형으로 클래스 균형 예시를 제공해 토큰 효율성과 안정성을 동시에 확보한다.

5. 실험 결과

<표 2> 각 모델 별 정확도와 F1-score 결과

	Accuracy		Macro F1-Score	
	Gemma-3-12B	Gemma-3-4B	Gemma-3-12B	Gemma-3-4B
Zero-Shot	27.89	30.99	20.93	16.62
One-Shot	87.42	39.63	87.48	35.06
ADAPT-3 Shot	91.68	64.4	91.75	63.04

LLM 에 ICL 을 적용하여 구동기의 고장 및 정상 분류를 실행하였으며, 평가는 Accuracy 와 Macro F1 Score 로 진행하였다. <표 2>에서 보는 것과 같이 12B 의 경우, Zero Shot 의 Accuracy 는 약 27%, F1 Score 는 약 21%로 낮은 성능을 보인다. 그러나 One Shot 에서는 87.42%, 87.48%로 성능이 개선 되었고, Adaptive-3 Shot 에서는 One shot 보다 성능이 소폭 상승하여 91%를 기록하였다. 이와 유사하게 4B 모델은 Zero-Shot 에서 Accuracy 30%, F1 Score 16%로 나타났으나, One shot 에서는 39.63%, 35.06% 성능이 개선 되었고, Adaptive-3 Shot 을 적용 하였을 때 64.4%, 63.04% 성능 개선을 보였다.

특히 4B 에서 Zero-shot 에서는 Accuracy 와 F1 Score 간 큰 성능 차이보였다. 이는 클래스의 불균형을 의미한다. 반면, One Shot 과 Few Shot 에서는 두 지표 간 차이가 거의 없어 균형 잡힌 분류 성능을 확인할 수 있었다. 이 실험을 통해 12B 모델의 Adaptive-3 Shot 을 적용한 결과로 F1-score 91.75%로 가장 높은 성능을 보였다.

6. 결론 및 시사점

본 연구는 Hampel filter 전처리와 50% Overlay 방식

의 Sliding Window 길이 21을 적용하여 ICL 탐지 기법을 제안했다. <표 2>에서 보는 것과 같이 Adaptive Few-Shot 은 제안했던 두 모델에서 가장 우수한 성능을 보였다. 특히 12B 모델에서는 Adaptive 3-Shot 에서 가장 우수한 성능을 보였고, 4B 성능은 One shot 대비 Adaptive 3-Shot 이 큰 폭의 성능 향상을 보여 잠재적 활용 가능성을 입증하였다. 이를 통해 소형 모델인 4B 또한 실제 응용 환경에서 충분히 발전 가능성이 있음을 확인할 수 있었다.

향후 연구를 통해 5 가지 고장 분류 모델을 확장 적용하고, 전류 데이터뿐 아니라 진동 데이터를 포함하여 보다 정밀하고 다양한 분류를 수행하는 것이 중요하다. 또한, 4B 모델에 예시 Shot 을 성능 향상을 검증하고, CoT(Chain of Thought)기법을 활용하여 12B 모델 수준의 Accuracy 와 Macro F1 Score 성능을 개선이 이루어진다면, 온 디바이스 환경에서도 실질적인 적용이 가능할 것으로 기대된다. 나아가 소형 모델에서도 우수한 성능이 검증된다면, 실제 산업 현장에서의 적용을 통해 구동기 고장 분류 과정에서 수리 비용과 시간을 크게 절감할 수 있을 것이다.

사사

본 연구에 적용된 데이터(‘기계시설물 고장 예지센서’)는 AI 허브(aihub.or.kr)에서 다운로드 받아 활용하였습니다.

참고문헌

- [1] 한국 전기 연구원 한필환 책임연구원 "기업과 환경을 모두 살리는 ‘슈퍼 프리미엄’급 산업용 전동기 개발”, 2023
- [2] Chalapathy, R., Chawla, S., “Deep learning for anomaly detection: A survey.”, arXiv:1901.03407v2
- [3] 대전 광역시, “인공지능 데이터 구축·활용 가이드 라인-기계시설물 진동/전류 데이터-”, 2021
- [4] 임지수, 이종욱, 신진호, “Hampel Filter 와 Cross-Dimensional Self-Attention 기반 태양광 발전량 이상 검출 데이터 탐색 및 결측치 보정”, 2022년도 대한전기학회 산하 스마트그리드연구회 추계학술대회 논문집, 2022
- [5] google Deep Mind, “Gemma 3 Technical Report”, arXiv:2503.19786v1, 2025
- [6] Lei Tang, Jinghui Qin, Wenxuan Ye, Hao Tan1, Zhijing Yang, “Adaptive Few-shot Prompting for Machine Translation with Pre-trained Language Models”, arXiv:2501, 2025