

화공 산업 전력 데이터의 부하 패턴 특성 분석 및 계층적 클러스터링 기반 이상 탐지 프레임워크

정민성¹, 권우현¹, 강지수¹, 허태욱², 이충호², *이상금¹

*국립한밭대학교¹, 한국전자통신연구원²

{jmss1101, mfireon0520, k4ngj15u}@gmail.com, {htw398, leech}@etri.re.kr,

*sangkeum@hanbat.ac.kr

Characteristics Analysis of Chemical Industry Power Data and Hierarchical Clustering-based Anomaly Detection Framework

Minsung Jung¹, Woohyeon Kwon¹, Jisu Kang¹, Taewook Heo², Chungho Lee², *Sangkeum Lee¹

*Hanbat National University¹, Electronics and Telecommunications Research Institute²

요약

화공 산업의 전력 소비 패턴은 24 시간 가동되는 연속 공정부터 주문 수주 이후 진행되는 공정까지 설비 운용 방식에 따라 매우 이질적이고 복잡한 분포를 보인다. 이러한 특성을 고려하지 않은 기존의 단일 모델은 탐지 성능이 저하되는 한계가 있다. 본 연구에서는 화공 산업 단지에서 수집된 시계열 전력 데이터를 분석하여, 부하율(LF, Load Factor)과 저부하 지속 비율(BMR, Below Mean Ratio)을 기준으로 공정을 고효율 연속(Type A), 주간 단속(Type B), 간헐적 공정(Type C)으로 분류하고, 표준편차의 중앙값을 기준으로 Type A1 과 A2, Type B1 과 B2, Type C 로 추가로 분류하는 계층적 클러스터링 방식과 Multi-Head Transformer 를 결합한 2 단계 학습 기반의 이상 탐지 프레임워크를 구축하였다. 제안된 모델은 공유 인코더를 통해 시계열의 일반적인 문맥을 학습하고, 클러스터별 개별 헤드를 통해 미세 공정 특성을 학습한다. 실험 결과, 정형화된 패턴의 Type A1 은 F1-Score 0.905, A2 는 0.835 를 기록했다. 변동성이 큰 Type C에서는 0.871 의 높은 성능을 달성하였고, Type B1에서는 0.812, B2에서는 0.309 를 기록했다. 이는 향후 Type B2와 같은 불규칙 단속 공정에 대한 추가적인 성능 고도화가 요구됨을 의미함과 동시에 제안 방법론이 다양한 공정 패턴 속에서 미세한 이상 징후를 효과적으로 탐지함을 입증한다.

I. 서론

스마트 팩토리와 에너지 관리 시스템(FEMS, Factory Energy Management System)의 확산으로 산업 현장의 전력 데이터 수집이 용이해졌으나, 이를 효과적으로 분석하고 활용하는 데에는 여전히 어려움이 존재한다 [1]. 특히 화공 산업은 24 시간 가동되는 연속 공정, 주문 수주 이후 진행되는 공정 등 다양한 설비 운용 방식이 공존하여 전력 패턴이 매우 복잡하고 이질적이다 [2][3]. 본 연구는 부하율(LF, Load Factor)과 저부하 지속 비율(BMR, Below Mean Ratio)을 지표로 활용하여 전력 패턴을 3 가지 유형(Type A, B, C)으로 분류하고, 각 유형을 표준편차의 중앙값에 따라 Type A1, A2 와 Type B1, B2, Type C 로 분류하는 계층적 클러스터링 기반의 Multi-Head Transformer 이상 탐지 모델을 구축하여, 데이터의 특성에 최적화된 이상 탐지 방법론을 제안한다.

II. 데이터 및 부하 패턴 분석

2.1 데이터셋 개요

본 연구는 2020 년 1 월 1 일부터 2023 년 12 월 31 일까지 수집된 화공 산업 전력 데이터로 15 분 단위로 계측되어 하루 96 개의 시점을 가지는 시계열 데이터로 구성되어 있다.

2.2 부하 패턴 분류 기준 및 분석

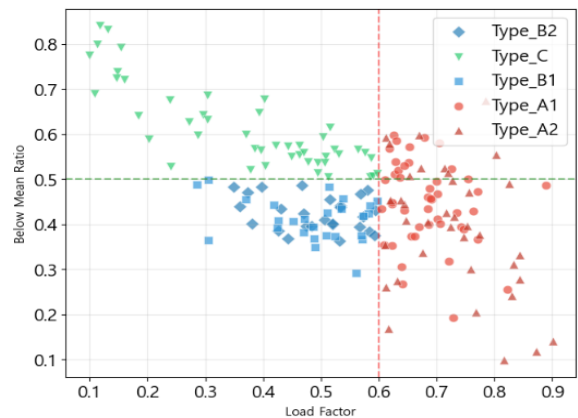


그림 1. Type A(빨간색), Type B(파란색), Type C(초록색) 분포

본 연구에서는 설비의 가동 효율과 지속성을 나타내는 두 가지 핵심 지표를 도출하여 패턴을 분류한다. LF 는 연간 최대 전력 대비 평균 전력의 비율로, 설비를 얼마나 꽉 채워 가동하는지를 나타낸다. BMR 은 전체 기간 중 연간 최대 전력의 20% 미만으로 가동된 시간의 비율로, 설비의 유휴 상태 빈도를 나타낸다. 이를 바탕으로 전체 미터기를 다음 3 가지 유형으로 정의하고 분석한다. 고효율 연속 공정인 Type A 는 LF 가 0.6 을 초과한다. Type C 는 간헐적/저활용 공정이며, LF 0.6 이하이고, BMR 이 0.5 이상이다. Type B 는 주간/단속 공정이며,

Type A 와 C 에 해당하지 않는 일반적인 공정이다. 주간 근무 패턴 혹은 변동이 명확하다.

III. 제안 방법론

3.1 계층적 클러스터링 (Hierarchical Clustering)

데이터를 물리적 특성과 통계적 패턴에 따라 2 단계로 세분화한다. 1 단계는 앞서 정의한 LF 와 BMR 을 기준으로 전체 데이터를 Type A, B, C 로 1 차 분할하고, Type 내부에서 클러스터 별 최적 임계값 설정을 위해 표준편차의 중앙값을 기준으로 Type A1, A2 와 Type B1, B2, Type C 로 2 차 클러스터링을 수행한다. Type C 는 데이터의 개수가 적어 2 차 클러스터링은 제외한다.

3.2 모델 아키텍처 (CNN & Multi-Head Transformer)

본 모델은 CNN 과 Transformer 를 결합하여 사용한다. 공유 특징 추출기(Shared Convolutional Encoder)의 Conv1D 레이어에서 시계열 데이터의 국소적 특징을 추출하고, 문맥 학습 레이어(Transformer Context Encoder)를 통해 추출된 특징 벡터의 장기 의존성 및 복잡한 시간적 상관관계를 학습한다.

이후, 재구성 헤드(Reconstruction Head)가 CNN 과 Transformer 를 통해 학습된 심층 표현을 입력받아 각 시계열 포인트의 정상 패턴을 재구성한다.

3.3 단계별 학습 (Two-Stage Training)

1 단계는 일반화 학습으로, 모든 데이터를 통합하여 공유 인코더와 공통 헤드를 학습시킨다. 이를 통해 모델은 시계열의 기본 문맥을 학습한다. 2 단계는 전문화 학습으로, 공유 인코더를 고정하거나 낮은 학습률을 적용한다. 그리고 각 클러스터별 데이터로 해당 개별 헤드를 미세 조정해 각 공정의 고유한 노이즈 분포와 미세 패턴을 학습하는 데 기여한다.

3.4 추론 및 이상 탐지

이상 탐지 임계값은 재구성 오차 분포의 MAD(Median Absolute Deviation)를 활용하여 설정한다. 재구성 오차(MSE)가 임계값을 초과할 경우 이상으로 판정한다.

IV. 실험 및 결과

4.1 실험 환경 및 이상 데이터 생성

제안한 방법론의 검증은 위해 실제 화공 전력 데이터에 Spike, Gradual, Cyclic 한 합성 이상(Synthetic Anomaly)을 주입하고 실험을 진행한다. 각유형의 이상 데이터는 1%씩, 전체 데이터의 3%를 차지한다.

4.2 이상 탐지 성능 분석

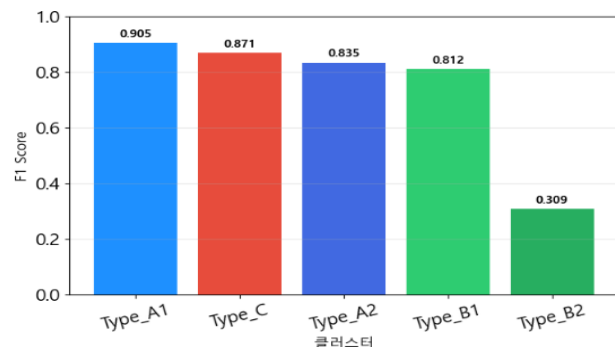


그림 2. 각 클러스터별 미터기 이상 탐지 결과 F1-Score 비교

실험 결과, 제안된 계층적 클러스터링 기반 프레임워크는 세부 공정 특성에 따라 차별화된 탐지 성능을 보였다. 그림 2 는 각 서브 클러스터별 F1-Score 를 나타낸다.

우선 Type A 의 서브 클러스터인 Type A1 의 미터기는 F1-Score 0.905 를 기록하였다. A2 는 0.835, Type C 는 0.871 를 기록했다.

Type B 에서는 서브 클러스터 간 극명한 성능 차이가 발생하였다. Type B1 은 0.812 로 안정적인 성능을 보인 반면, Type B2 는 0.309 로 매우 저조한 성능을 기록하였다. 분석 결과, LF 와 F1-Score 는 양의 상관관계를 보여, LF 가 높을수록 탐지 성능이 향상됨을 확인했다. 반면, BMR 과 표준편차는 음의 상관관계를 나타내어, 전력 사용 패턴이 불규칙할수록 탐지 성능이 저하됨을 보였다. Type B2 는 일별 전력 패턴의 변동성이 Type B1 대비 28% 높게 나타났으며 표준편차 또한 4.4 배 큰 것으로 확인되었다. 이는 Type B2 와 같이 높은 분산을 보이는 클러스터에서 정상적인 변동과 이상치 간의 경계가 모호해져 탐지 성능이 저하됨을 나타내며, 이러한 고분산 패턴에 대한 추가적인 모델 고도화가 요구됨을 시사한다.

V. 결론

본 논문은 화공 산업 전력 데이터의 분석을 통해 데이터의 이질성을 정량적으로 규명하고, 이를 반영한 계층적 이상 탐지 프레임워크를 제안하였다. 또한, 이를 기반으로 한 Multi-Head Transformer 는 단일 모델의 한계를 극복하고 다양한 공정 패턴에 유연하게 대응할 수 있음을 입증하였다. 본 연구 결과는 향후 스마트 팩토리의 에너지 효율화 및 예지 보전 시스템 구축에 있어, 데이터 분석에 기반한 맞춤형 모델링의 중요성을 시사한다.

ACKNOWLEDGMENT

본 연구는 산업통상자원부(MOTIE)와 한국에너지기술 평가원 (KETEP)의 지원을 받아 수행한 연구 과제입니다. (No. RS2023-00237018)

참 고 문 헌

- [1] S. Lee, S. H. Nengroo, H. Jin, Y. Doh, C. Lee, T. Heo, and D. Har, Anomaly detection of smart metering system for power management with battery storage system/electric vehicle, *ETRI Journal* 45 (2023), 650–665. <https://doi.org/10.4218/etrij.2022-0135>
- [2] Kwon, W.; Jung, M.; Park, J.; Lee, S. FFT-Guided Multi-Window USAD with DTW- Isolation Forest for Reliable Anomaly Detection in Industrial Power Time-Series. *Energies* 2025, 18, 6584. <https://doi.org/10.3390/en18246584>
- [3] Jung M, Jang H, Kwon W, Seo J, Park S, Park B, Park J, Yu D, Lee S. A Comparative Study of Customized Algorithms for Anomaly Detection in Industry-Specific Power Data. *Energies*. 2025; 18(14):3720. <https://doi.org/10.3390/en18143720>