

# 금속 산업 용도별 전력 데이터 비율 기반 임계값 최적화를 통한 이상치 탐지

정민성<sup>1</sup>, 김준서<sup>1</sup>, 서지윤<sup>1</sup>, 이충호<sup>2</sup>, 허태욱<sup>2</sup>, \*이상금<sup>1</sup>

\*국립한밭대학교<sup>1</sup>, 한국전자통신연구원<sup>2</sup>

{jmss1101, star1209042, sillou0821}@gmail.com, {leech, htw398}@etri.re.kr,  
\*sangkeum@hanbat.ac.kr

## Outlier Detection by Threshold Optimization Based on Power Data Ratios of Metal Industry Applications

Minsung Jung<sup>1</sup>, Junseo Kim<sup>1</sup>, Jiyun Seo<sup>1</sup>, Chungho Lee<sup>2</sup>, Taewook Heo<sup>2</sup>,  
and \*Sangkeum Lee<sup>1</sup>

\*Hanbat National University<sup>1</sup>, Electronics and Telecommunications Research Institute<sup>2</sup>

### 요 약

본 연구는 그린 버튼 플랫폼의 금속 산업체 전력 데이터를 분석하고, 이를 산업용, 일반용, 심야, 농사용 4 가지 용도로 세분화하여 각 용도별 전력 데이터 비율을 파악한다. 그 결과 산업용이 98.2%로 대부분의 비율을 차지하며, 전력 사용량도 가장 많다는 것을 확인할 수 있다. 농사용을 제외한 용도별 전력 데이터에는 매일 일부 결측치와 이상치가 존재하며, 심야 전력 데이터는 22 시부터 8 시까지 수집된다. 이런 결측치와 이상치는 전력 데이터의 분포 및 비율 특성을 변화시켜 임계값 설정에 직접적인 영향을 미친다. 따라서 용도별 전력 데이터 비율의 불균형이 임계값에 미치는 영향을 최소화하기 위해 Z-Score 기반의 이상치 탐지 기법을 활용하여 가중 평균 기대 비율을 제안하고, 이를 기반으로 최적 임계값을 도출한다. 연구 결과를 통해 용도별 전력 데이터 비율에 따라 임계값을 조정함으로써 이상치 탐지의 효율성을 극대화하고, 전력 데이터의 무결성을 보장하여 산업체의 전력 데이터 품질 향상에 기여하고자 한다.

### I. 서 론

4 차 산업혁명으로 에너지 소비가 증가하면서, 이산화탄소 배출 증가로 인한 지구 온난화가 가속화되고 있다. 이로 인해 지속가능한 발전과 효율적인 에너지 사용은 현대 사회 산업체들이 달성해야 할 주요 목표로 떠오르고 있다. 현재 에너지 사용의 상당 부분은 전력이 차지하고 있고, 그 중에서도 산업부문의 비중이 커 산업체들은 전력 데이터 관리를 기반으로 한 체계적인 절약이 특히 중요하다[1]. 이러한 흐름 속에서 산업체들이 주목한 기술 중 하나가 그린 버튼 플랫폼이다. 이 플랫폼은 손쉽게 실시간 에너지 사용량을 확인할 수 있어 직관적인 에너지 사용량 모니터링 및 조정이 가능하고, 이는 에너지 절약을 통한 이산화탄소 감축을 실현할 수 있다[2]. 그린 버튼 플랫폼에 이런 산업체의 전력 데이터를 사용하면, 산업체를 효율적으로 관리할 수 있을 뿐만 아니라 안정적인 서비스 제공도 가능하다[3]. 그러나 산업체의 전력 데이터는 다양한 환경 요인과 비정상적인 운영으로 인해 이상치가 발생할 가능성이 높다. 이런 이상치는 전력 데이터의 신뢰성과 정확성을 저해할 수 있기 때문에, 신속하고 정확하게 탐지하는 것이 중요하다[4][5].

본 연구에서는 금속 산업체의 전력 데이터에서 비정상적인 전력 사용 패턴 등의 이상치를 조기에 탐지하기 위해 Z-Score 기반의 이상치 탐지 기법을 활용하여 용도별 전력 데이터에 따른 정상치를 파악하고, 정상치 비율인 기대 비율을 가중 평균 기대 비율과 일치하도록 조정해 최적의 임계값을 도출하는 방법을 제안한다. 이 방법은 데이터 분석 및 모델링의 품질을 향상시킴으로써, 이상치 탐지의 효율성과 정확성을 높이는 데 기여한다.

### II. 본 론

#### 2.1 금속 산업체 데이터 설명

금속 산업체의 전력 데이터는 2022 년 1 월부터 12 월까지의 데이터가 15 분 단위로 기록되어 있으며, 용도별로는 산업용, 일반용, 심야, 농사용 총 4 가지로 구분된다. 산업용은 98.2%, 일반용은 1.6%, 심야와 농사용은 각각 0.1%의 비율을 차지하고 있으며, 그림 1을 통해 전력 사용량 역시 산업용이 가장 많다는 것을 알 수 있다. 농사용을 제외한 3 가지 용도별 전력 데이터에는 매일 일부 결측치와 이상치가 존재하고, 심야 전력 데이터는 특성상 22 시부터 8 시까지의 데이터가 수집된다.

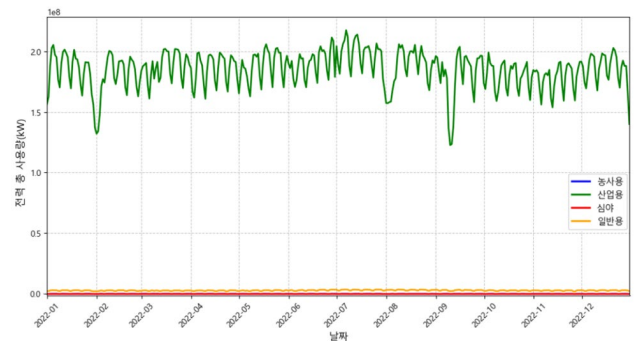


그림 1. 금속산업체의 용도별, 일별 전력 총 사용량

#### 2.2 데이터 전처리

전력 데이터 분석을 통해 결측치와 이상치만 가지고

있는 경우, 하루치 전력 데이터를 모두 제거한다. 결측치를 포함하고 있되 유의미한 값이 소량이나마 존재하는 날은 용도별로 분류하고, 전날과 다음날의 동일한 시간대의 값으로 평균을 내어 보간한다. 전력 데이터의 비교는 일별 총 전력 사용량의 평균값으로 진행한다.

### 2.3 Z-Score 를 이용한 이상치 범위 분석

금속 산업체의 전력 데이터는 시간대별, 용도별로 명확하게 분류되어 있어 이상치 탐지를 위해 통계적 접근법을 활용하기에 적합하다. 따라서 관측치의 편차를 Z-Score 를 활용해 전력 데이터를 분석한다. Z-Score 를 통한 이상치 탐지는 아래와 같이 정의한다:

$$\text{If } |Z| = \left| \frac{\text{관측치} - \text{평균}}{\text{표준편차}} \right| > \text{임계값, 이상치}$$

Z 의 절댓값이 특정 임계값을 초과하면 이상치로 간주한다. 이를 바탕으로 Z-Score 의 임계값(3)을 설정하고 이상치의 분포와 비율 차이를 비교했을 때, 전체 전력 데이터에서는 이상치가 모두 산업용에서 탐지되며, 용도별 전력 데이터로 각각 나눠서 본 경우 그림 2 와 같이 산업용(a)과 일반용(b)에서 빨간색 점으로 이상치가 탐지된다. 그 결과, 용도별 전력 데이터의 분포와 비율의 불균형이 임계값에 영향을 미치는 것으로 나타났다. 그 중 데이터 비율에 따른 임계값 조정의 필요성에 중점을 두고 분석을 진행한다.

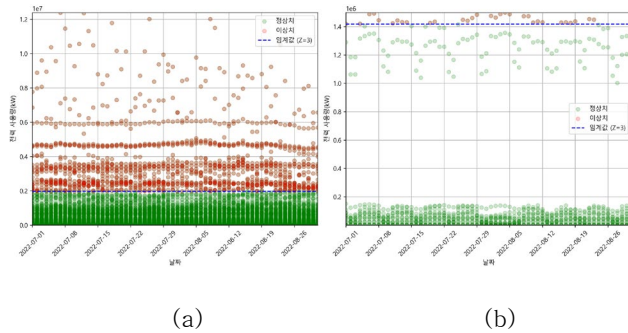


그림 2. Z-Score 의 임계값에 따른 산업용(a), 일반용(b)의 전력 사용량과 이상치

### 2.4 가중 평균 기대 비율을 통한 최적 임계값 도출

용도별 전력 데이터 비율의 불균형을 반영하기 위해 가중 평균 기대 비율을 제안한다. 가중 평균 기대 비율은 용도별 전력 데이터 비율을 기대 비율에 가중치로 반영하여, 보다 정확한 최적 임계값을 도출하기 위한 방법이다. 여기서 기대 비율은 Z-Score 에서 설정한 임계값에 따라 정상치가 차지하는 비율을 의미한다. 기대 비율과 가중 평균 기대 비율은 다음과 같이 정의한다:

$$\text{기대 비율} = 1 - \frac{\text{이상치 개수}}{\text{전체 데이터 개수}}$$

$$\begin{aligned} \text{가중 평균 기대 비율} &= \sum_{i=1}^n \left( \frac{k_i}{k_{\text{total}}} * k_i \text{의 기대비율} \right) \\ &= \sum_{i=1}^n \left( \frac{k_i}{k_{\text{total}}} * \left( 1 - \frac{k_i \text{의 이상치 개수}}{k_i} \right) \right) \end{aligned}$$

$n = \text{용도 수}$ ,  $k_1, k_2, k_3, k_4 = \text{일반용, 산업용, 심야, 농사용}$   
위의 식을 이용해 가중 평균 기대 비율과 기대 비율 간

의 오차가 최소가 되는 임계값을 최적값으로 선정한다. 임계값의 범위(0~5)를 설정하고, 계산한 가중 평균 기대 비율(0.9802)을 이용하여 얻은 결과를 그림 3 을 통해 볼 수 있다. 그림 3 을 보면, 최적 임계값은 일반용(3.1), 산업용(3.0), 심야(2.3), 농사용(1.7) 순으로 용도별 전력 데이터 비율에 비례해 높게 나타나는 경향을 보인다. 이때, 일반용(3.1)과 산업용(3.0)의 경우는 데이터의 분포가 고를수록 최적 임계값이 더 높아지는 예외가 있음을 시사한다. 이는 용도별 전력 데이터 비율 불균형이 임계값에 미치는 영향을 최소화하여 이상치 탐지의 효과를 높이는 데 기여한다.

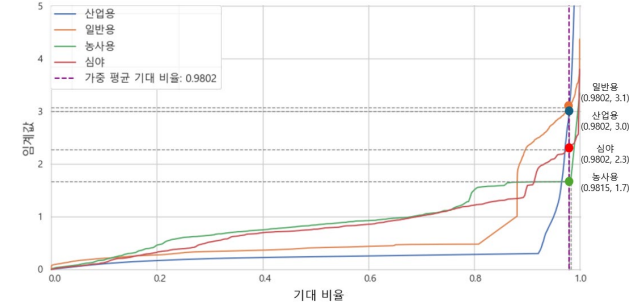


그림 3. 가중 평균 기대 비율과 기대 비율의 오차가 최소가 되는 용도별 전력 데이터의 최적 임계값

## III. 결론

본 연구는 금속 산업체의 전력 데이터를 분석하고, Z-Score 기반의 이상치 탐지 기법을 활용하여 이상치를 효율적으로 탐지하였다. 이때 용도별 전력 데이터 비율의 불균형이 미치는 영향을 최소화하기 위해 가중 평균 기대 비율을 이용한 최적 임계값 도출 방법을 제시하였다.

연구 결과, 최적 임계값은 일반용(3.1), 산업용(3.0), 심야(2.3), 농사용(1.7) 순으로 용도별 전력 데이터 비율에 비례해 높게 나타나는 경향을 보인다. 이때, 일반용(3.1)과 산업용(3.0)의 경우는 데이터의 분포가 고를수록 최적 임계값이 더 높아지는 예외가 있음을 시사한다. 이러한 임계값 조정은 용도별 전력 데이터 비율의 불균형을 최소화하여 이상치 탐지의 정확성과 신뢰성을 높이는 데 기여할 수 있다.

향후 연구에서는 금속 산업체뿐만 아니라 다른 산업체의 전력 데이터를 대상으로 본 기법의 적용 가능성을 검증할 예정이며, 이를 통해 전력 데이터의 신뢰성과 정확성을 강화하고, 나아가 지속 가능한 에너지 관리와 효율적인 산업체 운영에 기여할 수 있을 것으로 기대된다.

## ACKNOWLEDGMENT

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

## 참 고 문 헌

- [1] 에너지경제연구원. (2024). 「2024.12. 에너지통계월보」
- [2] 최경호. (2019). 4 차 산업혁명 핵심기술을 활용한 기후변화 대응 및 관련 법제 연구 - 빅데이터를 중심으로 -. 강원법학, 58, 779-815. 10.18215/kwir.2019.58..779
- [3] 이상금, 도윤미, 이충호, 허태욱. (2022-02-09). GRU 오토인코더를 이용한 스마트미터링 데이터 이상치 탐지. 한국통신학회 학술대회논문집, 강원.
- [4] 정병희, 박선아, 김주령, 이충호, 이상금. (2024-01-31). 산업 전기데이터 무결성을 위한 LSTM 네트워크를 이용한 결측치 처리 정보보호 플랫폼. 한국통신학회 학술대회논문집, 강원.

- [5] 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.