

2025년도 한국통신학회

## 동계종합학술발표회

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

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

저 자: 정민성, 김준서, 서지윤, 이충호(ETRI), 허태욱(ETRI), 이상금

발표자: 정민성

발표일: 2025.02.05

# 목차

1

서론

2

데이터 전처리

3

이상치 범위  
설정

4

임계값 최적화

5

결론

# 2025년도 한국통신학회 동계종합학술발표회

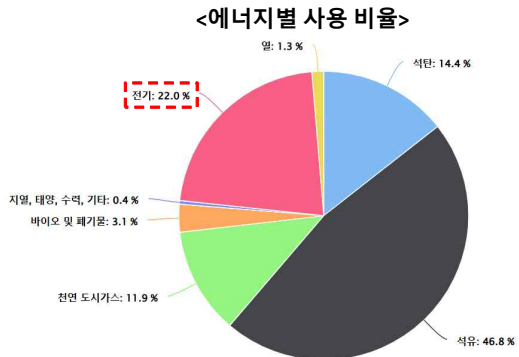
금속 산업 용도별 전력 데이터 비율 기반 임계값 최적화를 통한 이상치

탐지

## 서론

## 서론 (1/4)

### ➤ 전력 소비 관리의 필요성



출처 : 에너지통계월보 2024년 3월 자료

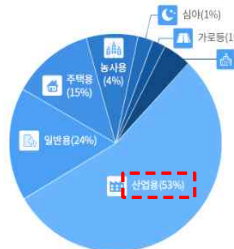
- 전력이 전체 사용 비율의 22%를 차지
- 전력 사용량은 계속 증가하는 추세
- 산업부문이 53%로 절반 이상을 차지

### <전기 사용량 추이>

전기사용량 추이



2023년 종별 사용량 비중



출처:한국전력공사

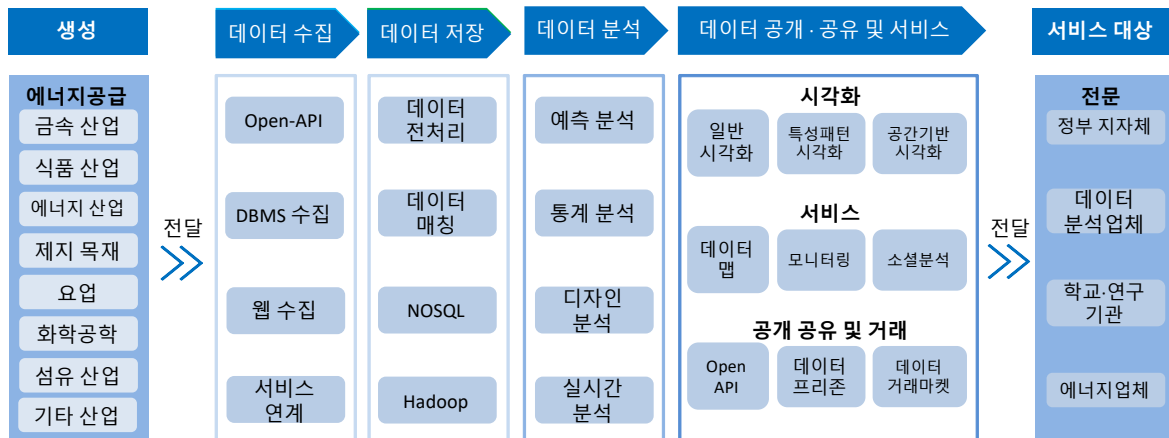
전력 사용량의 증가



산업체의 비용 증가,  
이산화탄소 배출 증가

## 서론 (2/4)

### ➤ 그린 버튼 플랫폼

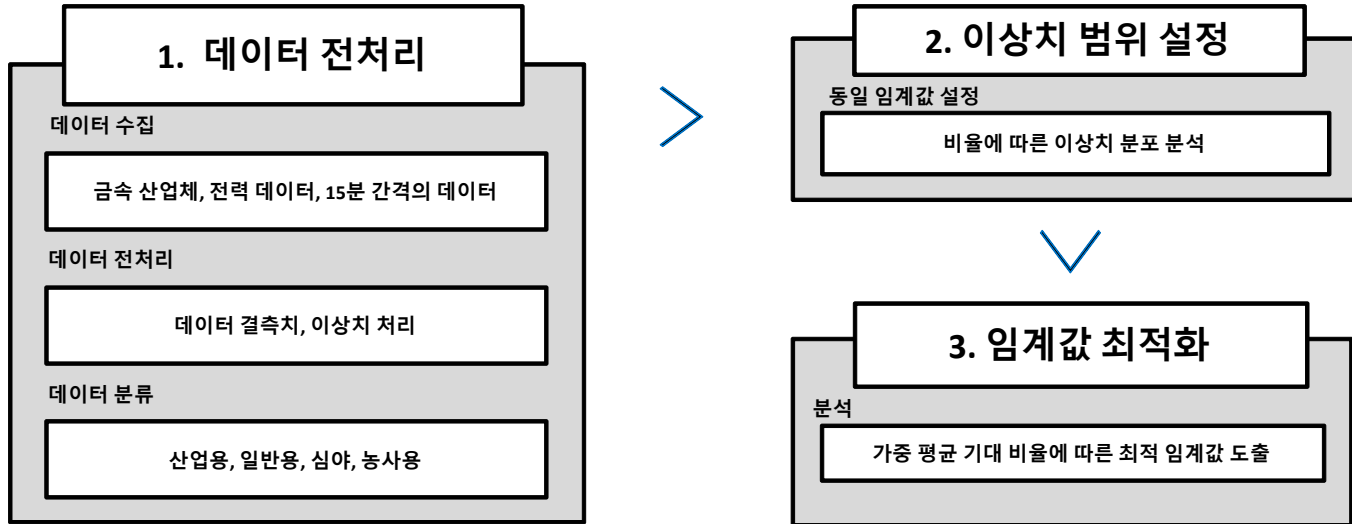


출처:한국에너지공단

- 표준화된 데이터를 기반으로 에너지 소비 정보를 효율적으로 활용할 수 있도록 지원
- 산업체의 실시간 전력 사용 데이터를 수집, 저장, 여러 요인에 따른 소비 패턴의 분석 과정을 거쳐 절약 방안 제시

## 서론 (3/4)

### ➤ 흐름도



## 서론 (4/4)

### ➤ 연구목표

- **Z-Score**를 기반으로 데이터의 분포 특성을 고려한 **이상치 탐지 기법을 개발**
- 데이터 품질 향상을 통해 산업체의 **전력 데이터를 효율적으로 관리**
- 산업체들의 **안정적인 서비스 제공 기반 마련**

2025년도 한국통신학회  
동계종합학술발표회

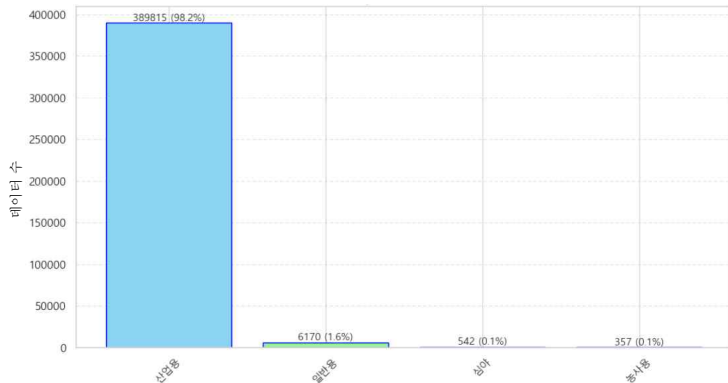
금속 산업 용도별 전력 데이터 비율 기반 임계값 최적화를 통한 이상치  
탐지

# 데이터 전처리

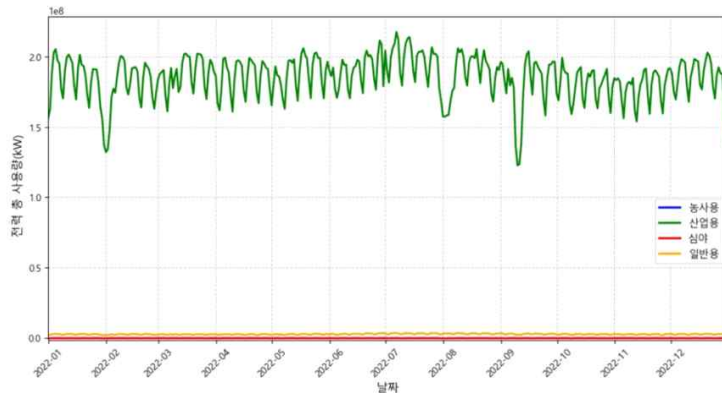


## 데이터 전처리 (1/2)

### ➤ 데이터셋 설명



<금속용 전력 데이터의 용도별 비율>



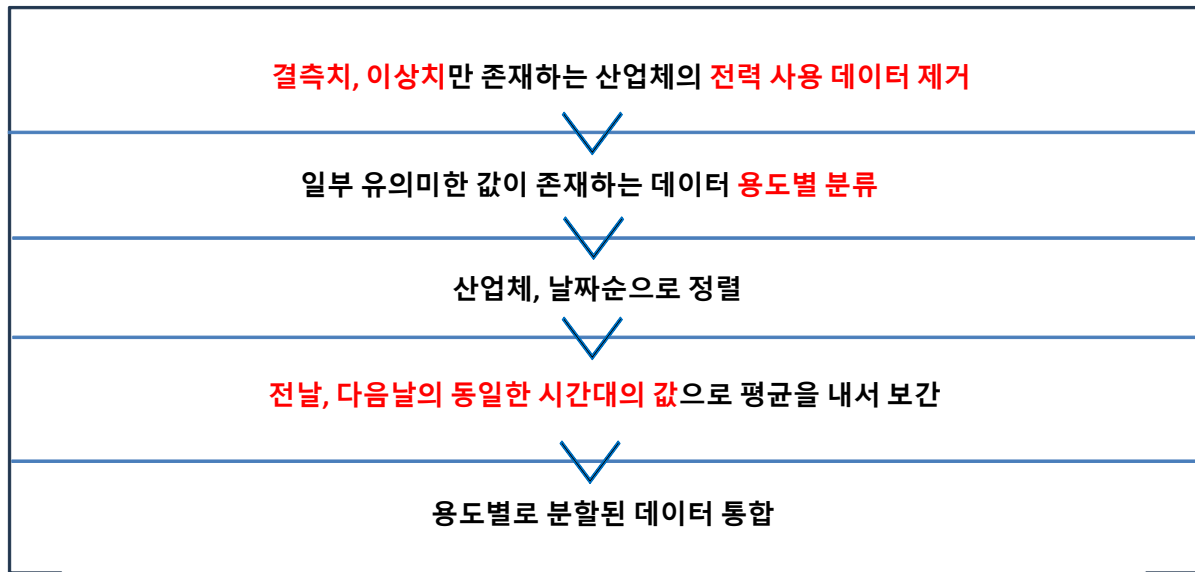
<금속산업체의 용도별, 일별 전력 총 사용량>

-용도별로 산업용, 일반용, 심야, 농사용으로 구분

-산업용의 데이터 수가 전체 데이터의 **98.2%**로 대부분을 차지

## 데이터 전처리 (2/2)

### ➤ 데이터 전처리



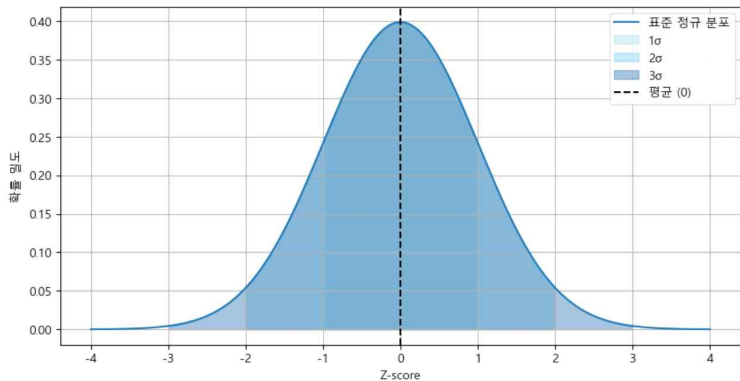
## 2025년도 한국통신학회 동계종합학술발표회

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

# 이상치 범위 분석

# 이상치 범위 분석 (1/3)

## ➤ Z-Score



<Z-Score 표준편차 그래프 >

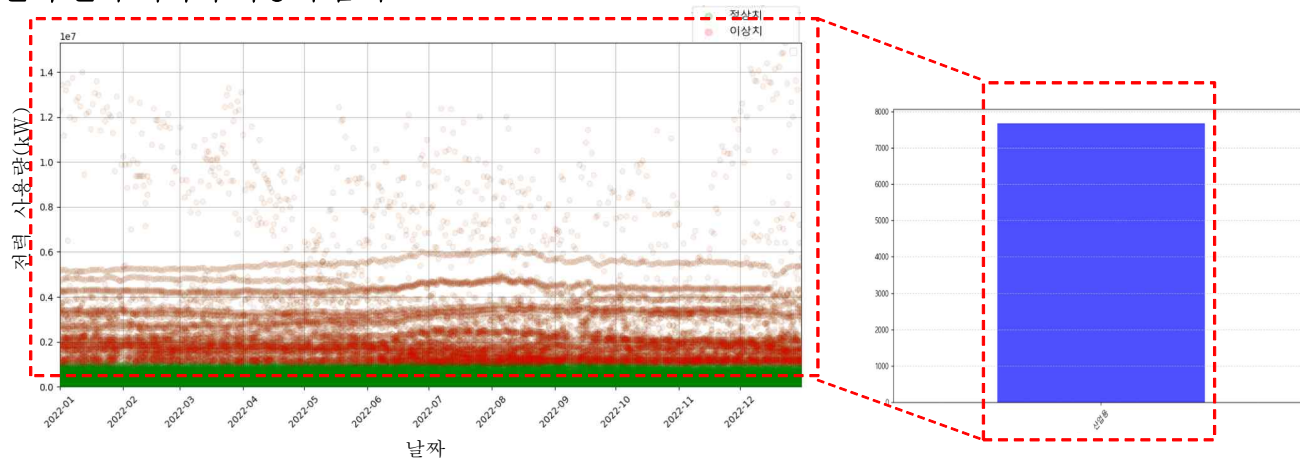
$$\text{If } |Z| = \left| \frac{\text{관측치} - \text{평균}}{\text{표준편차}} \right| > \text{임계값, } \text{이상치}$$

-관측치의 오차를 표준편차로 측정한 값

-Z-Score의 절댓값이 **특정 임계값을 초과하면 이상치**로 간주

## 이상치 범위 분석 (2/3)

### ➤ 전체 전력 데이터 이상치 탐지



<전체 전력 데이터의 이상치 분포>

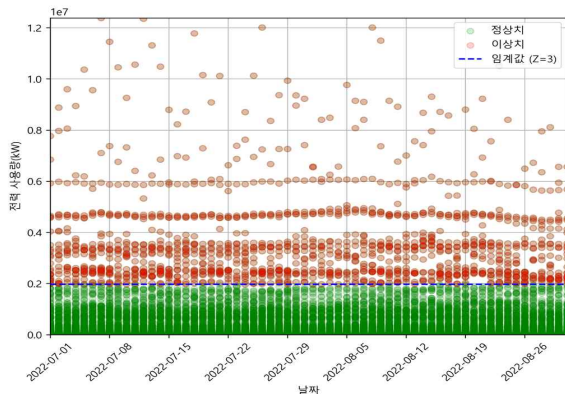
<탐지된 이상치의  
용도>

- 전체 전력 데이터에서 Z-Score의 임계값을 3으로 설정하여 진행

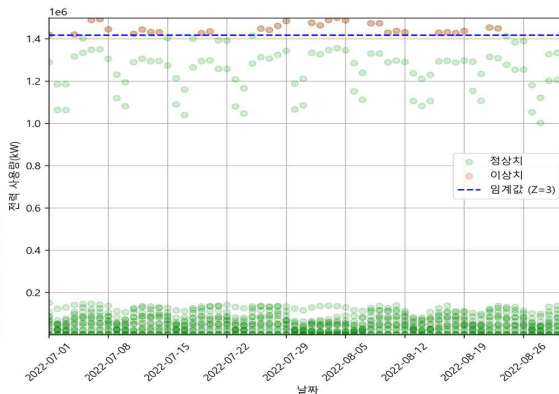
- 이상치를 각 용도별로 살펴봤을 때 모두 산업용에서 탐지

## 이상치 범위 분석 (3/3)

### ➤ 용도별 전력 데이터 이상치 탐지



<산업용 전력 데이터의 이상치 분포>



<일반용 전력 데이터의 이상치 분포>

— 용도별 분류 후 각각 임계값을 3으로 설정했을 때 산업용, 일반용에서만 이상치가 탐지됨

➤ 데이터의 분포와 비율의 불균형이 이상치 탐지를 최대화 하는 최적 임계값 설정에 영향을 미침

## 2025년도 한국통신학회 동계종합학술발표회

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

# 임계값 최적화

## 임계값 최적화 (1/3)

### ▶ 기대 비율과 가중 평균 기대 비율

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

$n$  = 용도 수,  $k_1, k_2, k_3, k_4$  = 일반용, 산업용, 심야, 농사용

- 용도별 전력 데이터의 비율이 최적 임계값에 미치는 영향을 줄이기 위해 **가중 평균 기대 비율**을 제안
- **용도별 전력 데이터 비율**을 기대 비율에 **가중치로 반영**해 더 정확한 최적 임계값 도출

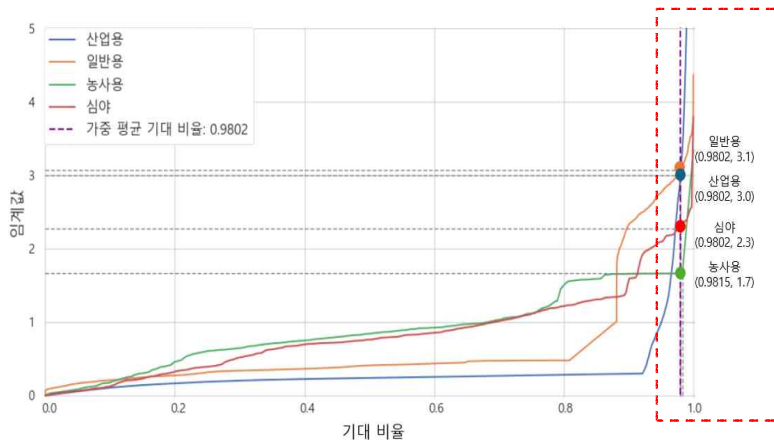
$$\text{기대 비율} = 1 - \frac{\text{이상치 개수}}{\text{전체 데이터 개수}}$$

- **기대비율**은 Z-Score에서 설정한 **임계값에 따라 정상치가 차지하는 비율**을 의미



## 임계값 최적화 (2/3)

### ➤ 용도별 최적 임계값



<가중 평균 기대 비율과 용도별 전력 데이터의 최적 임계값>

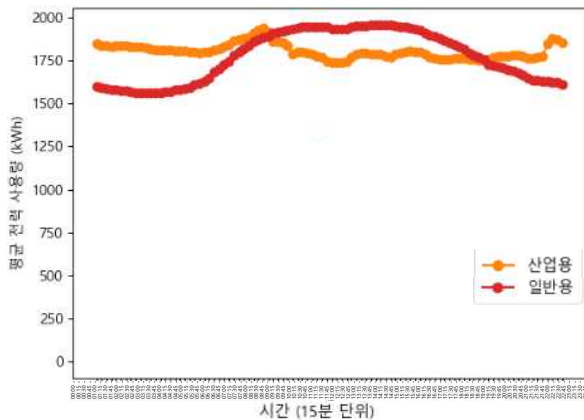
- 임계값 변화에 따른 기대비율과 가중평균 기대비율이  
최소한의 오차를 보이는 값을 **최적 임계값**으로 선정

- 일반용 3.1, 산업용 3.0, 심야 2.3, 농사용 1.7

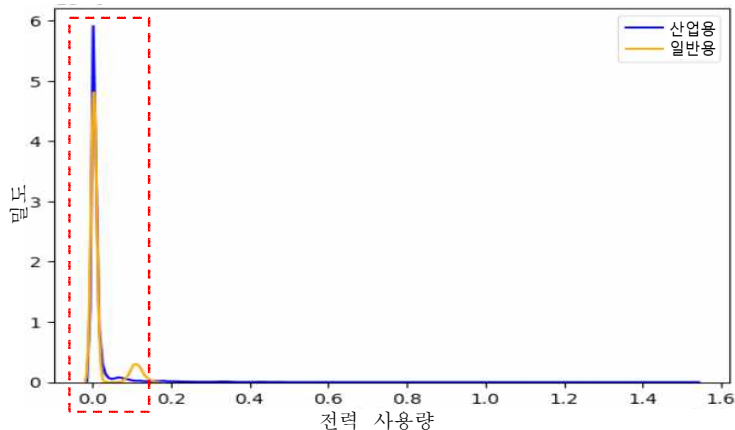
> 데이터의 **비율이 낮을수록 최적 임계값도 낮은 경향**

## 임계값 최적화 (3/3)

### 예외



<용도별 평균 전력 사용량 그래프(15분 단위)>



<산업용과 일반용 전력 데이터의 커널 밀도 추정 그래프>

- 일반용과 산업용의 전력의 **평균이 유사**
- 일반용이 산업용보다 **밀도가 낮게** 나타남
- **일반용의 IQR값이 75883, 산업용은 68908으로 보다 높음**

> 데이터의 **평균이 유사**하고 **값이 더 고르게 분포**되어 있어  
최적 임계값이 **더 높게** 나타남

## 2025년도 한국통신학회 동계종합학술발표회

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

### 결론 및 향후 연구

## 결론

1. 금속 산업 전력 데이터에 통계적 접근법을 활용해 **임계값 최적화**를 목표로 연구  
진행
2. **Z-Score**를 이용해 용도별 전력 데이터의 분포와 비율에 따른 **이상치 분포 분석**
3. 용도별 전력 데이터의 임계값에 영향을 미치는 요인으로 **데이터의 비율 선정**
4. 기대 비율에 가중치를 부여하는 가중 평균 기대 비율 **수식 도출**
5. 가중 평균 기대 비율 활용을 통해 **최적 임계값 확인 및 분석**

## 향후 연구 및 기대효과

금속 이외 산업의 전력 데이터를 대상으로 적용 가능성 검증



전력 데이터의 신뢰성과 정확성 강화



지속 가능한 에너지 관리와 효율적인 산업체 운영

# 2025년도 한국통신학회 동계종합학술발표회

감사합니다

저자: 정민성, 김준서, 서지윤, 이충호(ETRI), 허태욱(ETRI), 이상금

발표자: 정민성

발표일: 2025.02.05