

DBA K-Means 군집화 알고리즘 기반 화공산업 전력 사용량 이상치 탐지

장현석¹, 박범도¹, 강한성¹, 박선아¹, 이충호², 허태욱², *이상근¹

*국립한밭대학교¹, 한국전자통신연구원²

{seokchu123, pbeomdo, alfm917, ssunaa0322}@gmail.com, {leech, htw398}@etri.re.kr,

sangkeum@hanbat.ac.kr

Outlier Detection in Power Consumption of the Chemical Industry Based on the DBA K-Means Clustering Algorithm

Hyeonseok Jang¹, Beomdo Park¹, Hanseong Kang¹, Suna Park¹, Chungho Lee²,

Taewook Heo², and *Sangkeum Lee¹

*Hanbat National University¹, Electronics and Telecommunications Research Institute²

요약

본 논문은 매년 악화하는 기후 문제를 해결하고자 화공 산업체의 전력 소비 데이터를 분석하고, 이상치를 탐지함으로써 탄소 배출 절감 및 지속 가능한 에너지 관리 구현에 기여할 수 있는 실질적인 플랫폼 구현을 목표로 한다. 계절별로 전력 소비량의 평균값과 시간대별 이용 패턴의 차이를 보여 이를 바탕으로 데이터셋을 구성한다. 전처리 과정에서는 결측치를 제외했으며, 평균 0, 분산 1로 표준화한다. 전처리된 데이터는 시계열 데이터의 비선형적 특성에 적합한 군집화 알고리즘인 DBA(Dynamic time warping and Barycenter averaging) K-Means를 적용하여, 실루엣 점수를 반영한 최적의 군집 수로 분류한다. 이상치 탐지 과정에서 군집의 중심점과 데이터 샘플 간의 거리가 먼 상위 3%를 이상치로 간주한다. 산업 현장에서 전력 소비량을 최적화할 수 있는 그린버튼 플랫폼의 신뢰성 향상에 기여할 수 있을 것으로 기대한다.

I. 서론

기후 변화의 심각성이 날로 증가함에 따라, 전 세계는 탄소중립 달성을 최우선 과제로 삼고 있다[1]. 산업 공정에서 발생하는 온실가스 배출은 탄소 농도를 증가시키는 주요 원인 중 하나로, 이를 줄이기 위한 체계적인 에너지 관리와 배출 저감 기술의 도입이 필수적이다. 에너지 관리 플랫폼인 그린버튼은 이러한 문제를 해결하기 위한 핵심 도구로 주목받고 있다[2]. 산업 현장에서 그린버튼의 활용은 전력 소비 효율화를 통해 온실가스 배출을 줄이고, 에너지 비용을 절감하는 데 중요한 역할을 할 수 있다. 그러나 산업 공정에서 생성되는 방대한 양의 전력 소비 데이터는 종종 이상치를 포함하고 있으며, 이는 에너지 관리의 신뢰성과 효율성을 저하할 수 있다. 이를 극복하기 위해 본 연구에서는 산업 전력 소비 데이터를 분석하고 DBA K-Means 군집화 알고리즘 기반 이상치 탐지 방법론을 제안한다. 해당 알고리즘은 비선형적 시간 변형을 허용하는 DTW(Dynamic Time Warping)와 최적의 매핑 경로를 활용해 중심점을 계산하여 시계열 데이터의 군집화를 안정적으로 수행한다. 이러한 접근은 산업체의 전력 소비 데이터 분석과 이상치 탐지의 정확도를 높이고, 탄소중립 실현을 위한 에너지 관리 기술 발전에 이바지할 것으로 기대된다.

II. 본론

2-1. 이상치 탐지를 위한 데이터셋 재구성

화공 산업체의 전력 사용량을 분석하기 위해 3년간 (2020년 1월 ~2022년 12월) 15분 단위의 데이터를 활용하였다. 그림 1은 시간에 따른 전력 사용량의 24시간 패턴을 계절별로 나타낸 결과이다. 실선과 점선은 전력 사용량과 3시간 단위의 전력 사용량의 평균값을 의미한다. 이상치 식별을

위해 특정 시간 시퀀스의 패턴이 명확하게 구분되는 여름과 겨울의 12시 ~ 15시 사이의 데이터를 사용한다. 결측치는 이상치 결과를 왜곡할 가능성이 있으므로 전처리 과정에서 제거한다. 패턴 분석에 집중하기 위해 평균 0, 분산 1로 표준화한 데이터를 사용한다.

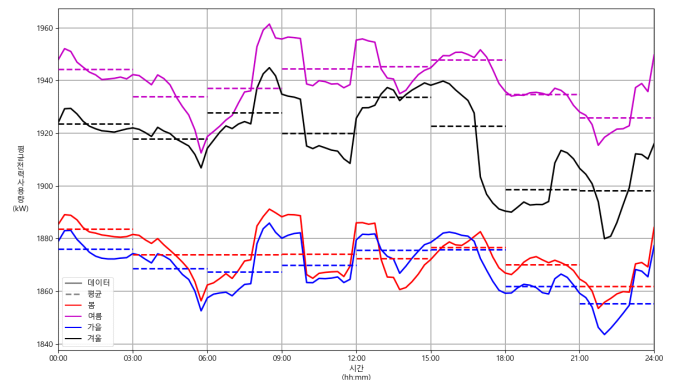


그림 1. 화공 산업체의 계절별 평균 전력 사용량

2-2. DBA K-means 군집화 이상치 탐지

산업 공정에서 생성되는 전력 소비 데이터는 복잡한 패턴과 불규칙적인 이상치를 포함한다. 이 데이터를 효과적으로 분석하고 이상치를 탐지하기 위해 실루엣 점수에 따라 군집 개수를 선정한 DBA K-Means 군집화 알고리즘을 사용한다. DBA는 DTW거리에 따라 최적 매핑 경로를 활용해 중심값을 산출한다. DTW는 시간 축을 비선형적으로 정렬하여 시퀀스의 길이와 관계없이 유사성을 측정한다. 거리는 다음과 같이 정의된다[3]:

$$D(A_i, B_j) = \delta(a_i, b_j) + \min \left\{ \begin{array}{l} D(A_{i-1}, B_{j-1}) \\ D(A_i, B_{j-1}) \\ D(A_{i-1}, B_j) \end{array} \right\}$$

A_i 는 시퀀스 A 의 처음 i 개 요소로 구성된 부분 시퀀스이다. 즉, $A_i = \langle a_1, a_2, \dots, a_i \rangle$ 로 정의되며, B_j 도 A_i 와 동일한 방식이다. $D(A_i, B_j)$ 는 누적 거리, $\delta(a_i, b_j)$ 는 두 시점 간의 거리를 나타낸다. 이러한 거리를 기반으로 최적의 매핑 경로를 계산해 데이터를 정렬한다. Barycenter averaging은 중심값을 정렬된 값의 평균으로 계산하고, 값이 수렴할 때까지 반복하는 과정이다. 기존의 거리 계산 방식과 달리, DBA는 데이터의 복잡한 패턴에 대해서도 강건한 군집화 성능을 보인다.

이상치 식별을 위해 테스트 데이터를 학습 데이터와 동일한 방식으로 전처리 후 모델에 적용한다. 군집의 중심과 각 데이터 샘플 간 거리를 DTW로 계산하고 거리가 먼 상위 3%의 데이터 샘플을 이상치로 간주한다. 탐지 결과를 시각화 하기 위해 이상치를 색상의 차이로 구분하였다.

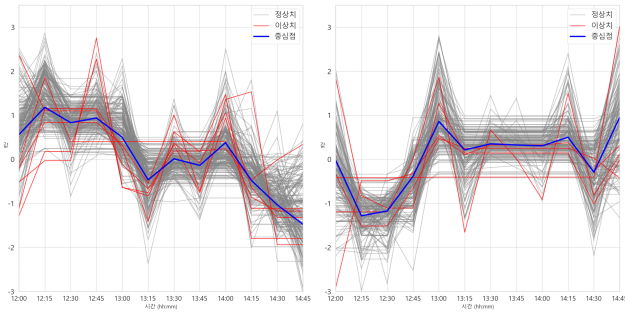


그림 2. 여름철 전력 소비 데이터의 이상치 탐지 결과

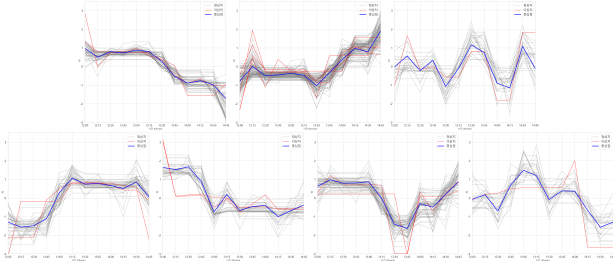


그림 3. 겨울철 전력 소비 데이터의 이상치 탐지 결과

그림 2, 3과 같이 여름철은 상승, 하강 두 가지 패턴에 따른 2개의 군집으로 분류되었지만, 겨울철은 상승, 유지, 하강 세 가지 패턴에 따른 7개의 군집이 형성된다. 이러한 차이는 계절적 특성으로 인한 화학 공업의 공정 운영 방식에서 기인한 결과로 해석된다. 계절에 따라 군집 개수가 다른 현상을 관찰하고 계절적 특성이라는 인과적인 의미를 알아보기 위해서 귀추적 추론을 한다.

2-3. 군집화 결과의 귀추적 추론

계절별 군집 차이의 원인을 알아보기 위해 사용한 데이터를 산업 분야 별로 분석한 결과, 석유화학이 큰 비중을 차지했다. 석유화학 공정에서 초기 작업은 증류탑과 가열로를 사용해 원료를 성분별로 분리하는 상압증류 공정이다[4]. 각 성분을 증류하기 위한 적정 온도가 존재하며, 이를 유지하기 위한 전력 소모는 기온의 영향을 받는다. 2022년 대한석유협회 통계에 따르면, 겨울철 보일러에 사용하는 'B-C유'와 같은 연료가 여름철 대비 수요가 높다[5]. 이에 따라 겨울철 난방으로 인한 전력 소비는 여름철 에

컨 사용으로 인한 전력 소비량보다 상대적으로 적다. 따라서 겨울철에는 난방보다 생산과 산업활동에 사용되는 전력 소비 패턴이 더욱 두드러지게 나타나게 되며, 여름철보다 겨울철에 복잡한 패턴이 발생하는 원인을 귀추적 추론으로 분석했다.

III. 결 론

화학 산업체의 전력 사용량을 분석한 결과, 여름과 겨울의 12시~15시 데이터가 명확한 패턴 차이를 보였다. 패턴이 달라서 발생하는 비선형적인 특징의 데이터를 효과적으로 군집화하는 데 적합한 DBA K-Means 군집화 알고리즘을 활용한다. 분석 결과, 여름철에는 2개의 군집, 겨울철에는 7개의 군집이 형성되었다. 이를 귀추적으로 추론한 원인으로, 겨울철은 난방보다 생산 활동에 의한 전력 소비가 두드러져 여름철보다 소비 패턴이 더 복잡하게 나타났다. 또한, 군집 중심점과 데이터 샘플 간의 거리가 먼 상위 3%의 데이터 샘플을 이상치로 간주하여 시각화하였고, 이를 통해 모델의 신뢰성을 확인하였다. 향후 연구에서는 최적의 파라미터 조합을 도출하여 모델 성능을 개선하고, 그린버튼 플랫폼의 신뢰성을 강화하여 에너지 소비 효율을 극대화하고자 한다.

ACKNOWLEDGMENT

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

참 고 문 헌

- [1] 김두수, "생물자원 이용을 통한 탄소중립 실현 관련 국제적 동향과 우리나라의 법정정책 대응방안," 국제경제법연구, vol. 21, no. 3, pp. 199-228, 2023. (10.46271/KJIEL.2023.11.21.3.199).
- [2] 정범희, 박선아, 김주령, 이충호, and 이상금, "산업 전기데이터 무결성을 위한 LSTM 네트워크를 이용한 결측치 처리 정보보호 플랫폼," 한국통신학회 학술대회논문집, pp. 104-106, 강원, 2024-01-31.
- [3] Petitjean, François & Ketterlin, Alain & Gancarski, Pierre. (2011). A global averaging method for dynamic time warping, with applications to clustering. Pattern Recognition. 44. 678-. 10.1016/j.patcog.2010.09.013.
- [4] 대한석유협회, "국내 석유산업 현황 및 지속가능 발전방안," 2024년 12월 11일.
- [5] 대한석유협회, "월별·석유제품별 국내소비통계,"(2022).