

전기차 충전 데이터의 이상치 검출, 유형 분석 및 보정 방법

이예은, 김동주, *황의석

광주과학기술원

전기전자컴퓨터공학부

leeyeaun@gm.gist.ac.kr, {kimdj, *euisseokh}@gist.ac.kr

Outlier Detection, Feature Analysis and Compensation Method for Electric Vehicle Charging Data

Yeaun Lee, Dongju Kim, Euisseok Hwang*

School of Electrical Engineering and Computer Science,

Gwangju Institute of Science and Technology

요약

본 연구는 전기차 급속 충전 데이터의 특성을 고려한 이상치 유형 검출 및 보정 방법을 제시하였다. 전기차 충전 시간과 충전량의 관계성을 고려한 데이터셋 분석 과정에서 다양한 유형의 이상치가 발견되었으며, 이 중 일부는 적절한 조정을 통해 정상 데이터로 보정할 수 있는 것으로 확인되었다. 제안된 방법을 적용한 결과, 전체 데이터의 0.07%가 이상치로 검출되었고, 이 중 72.98%가 에러 유형 분석을 통해 성공적으로 보정되었다. 본 방법은 기존의 통계 기반 이상치 검출 알고리즘에 비해 최대 540.64% 더 많은 이상치를 검출해 내는 우수한 성능을 보여주었다.

I. 서론

전 세계적으로 전기차에 대한 관심이 빠르게 증가하고 있으며, 이에 따라 전기차 충전소 데이터의 중요성은 더욱 부각되고 있다. 2023년 전 세계 전기차 판매량은 약 350만 대로 전년 대비 약 35% 증가했으며, 공용 충전소는 약 380만 개로 40% 늘었다. 또한 전기차 및 관련 인프라 시장은 1조 달러를 초과하였다[1].

이처럼 전기차 및 충전소의 급속한 성장에 따라, 이에 관한 데이터가 중요한 자산이 되고 있다. 이는 충전소의 운영과 효율성 향상[2], 사용자의 충전 패턴 예측[3], 충전소 유형 분류[4], 인프라 개선 계획 수립[5] 등 다양한 목적으로 활용될 수 있다. 따라서 전기차 및 충전소 데이터의 분석과 활용은 앞으로도 더욱 중요해질 것으로 예상된다. 그러나 이를 위해선 데이터 내의 이상치를 정확히 탐지하고 처리하는 과정이 필수적이다.

본 연구는 이상치를 단순히 제거하는 것이 아니라, 이상치의 유형을 분석하고 이에 따른 보정 방법을 제안한다. 이를 통해 제한된 데이터를 최대한 활용하면서, 전기차 충전소 데이터의 정확도와 신뢰성을 향상할 수 있는 방법을 제시하고자 한다.

II. 본론

1. 실험 데이터

본 연구에서는 환경부 소속의 전기차 급속 충전소에서 2022년 한 해 동안 수집된 데이터를 대상으로 분석을 시행하였다. 총 6,721,672건의 충전데이터가 있으며, 각 충전 데이터는 충전소명과 충전기 ID, 충전 시작 일시와 충전 종료 일시, 충전 시간, 충전량으로 구성되어 있다. 전기차 충전방해금지법에 따라, 이 급속 충전기들은 40분의 충전 제한이 존재한다[6]. 이러한 규정들로 인해 데이터에서 충전 시간은 주로 40분에 정확하게 기록되는 경향이 있으며, 충전량 또한 55~80kW에 주로 분포한다.

2. 이상치 검출 및 유형 분류

본 연구에서는 이상치를 식별하는 방법으로, 먼저 데이터에서 충전 시작 일시와 충전 종료일시를 통해 차량이 충전소에 머물렀던 시간을 계산하였다. 이 계산된 시간은 데이터에 기록된 충전 시간과 비교 분석되어 이상치를 탐지하는 기준으로 활용되었으며, 충전 시간이 실제 차량이 충전소에 머문 시간과 상이한 경우를 이상치로 판단하였다.

또한, 전기차 충전 데이터 특성을 고려하여 충전 시간과 충전량을 함께 분석하여 이상치를 식별하였다. 급속충전기의 최대 출력이 200kW인 점을 고려하여, 이를 초과하는 출력 데이터를 이상치로 분류하였고, 충전기의 최소 출력은 전기차 배터리 잔량에 따라 변동될 수 있기 때문에 최대 출력만을 고려하였다.

위 방법으로 검출된 이상치 이들 중 일부는 적절한 조정을 통해 정상 데이터로 보정할 수 있는 것을 확인하였다. 이에 관련된 내용은 Table 1.에서 확인할 수 있다. E1, E2, E3, E4에 해당하는 유형들은 기록 오류 또는 기술적 문제로 발생한 것으로 추정되며, 충전 시간을 재계산함으로써 정상 데이터로의 변환을 시도할 수 있다. 반면, E5, E6, E7 유형은 충전 시간 정보의 누락으로 보정이 불가능한 것으로 판별되었다. 또한, 24시간을 초과하는 데이터는 현실적으로 존재할 가능성이 작고, 일반적으로 오기록으로 간주되어 제거하였다[6]. 일부 이상치는 단일 유형에만 국한되지 않고, 여러 유형이 겹쳐 나타나는 경우가 존재하였다.

3. 이상치 보정 결과

Fig. 1.은 원본 데이터의 충전 시간과 충전량의 데이터 분포를 나타내며, Fig. 2.는 본 연구에서 제안된 방법론을 적용하여 이상치를 제거한 후의 충전 시간과 충전량의 데이터 분포를 나타낸 것이다. 이 두 그래프 모두에서 법률에 따라 정해진 충전 시간 제한인 40분을 선으로 명확히 표시하여, 이를 기준으로 데이터를 쉽게 식별할 수 있도록 하였다. 충전 시간의 데이터에 오류가 발견된

Table 1. 이상치 유형과 설명, 보정 여부

이상치 유형	특징	개수
보정 가능		
E1	시간과 분의 단위가 각각 분과 초로 바뀌어 잘못 기록된 경우	2,977
E2	특정 충전시간을 초과하면 시간이 기록되지 않고, 분과 초만 기록된 경우	457
E3	충전 시작 일시, 충전 종료 일시, 충전량이 존재하나 충전시간이 0으로 기록된 경우	197
E4	충전 종료 일시와 충전 시작 일시가 뒤바뀐 경우	7
보정 불가		
E5	충전량은 존재하나, 충전 종료 일시가 충전 시작 일시와 동일하게 기록된 경우	1,184
E6	충전량이 충전시간에 비해 과도하게 높은 경우	90
E7	충전시간이 24시간을 넘는 경우	41

경우, 차량이 머문 시간을 계산하여 충전 시간을 대체하였기 때문에 Fig. 2.에선 40분을 초과하는 관측치의 수가 증가한 것을 볼 수 있다. 결과적으로, 총 6,721,672개의 데이터 중에서 1,274개의 이상치를 제거하고, 3,441개의 데이터를 성공적으로 보정함으로써 최종적으로 6,720,399개의 데이터를 확보하였다.

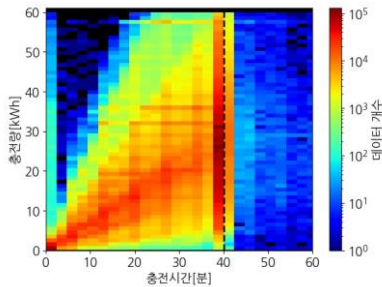


Fig. 1 원본 데이터 충전 시간과 충전량 분포.

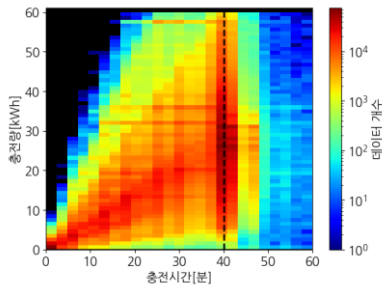


Fig. 2 이상치 보정 후 충전 시간과 충전량 분포.

4. 다른 방법과의 비교

이상치 제거에 널리 사용되는 두 가지 방법인 Interquartile Range (IQR) 방법과 Z-score 방법을 동일한 데이터에 적용하여 비교 분석하였다. 충전량과 충전 시간, 차량이 머문 시간 세 항목에 대하여 해당 방법을 적용하였다. Z-score 방법의 경우 항목별로 Z-점수가 3 이상일 때 이상치로 간주하였다. 두 방법 모두 3가지 데이터 중 2가지 이상의 데이터가 이상치로 분류될 경우 해당 데이터를 이상치로 식별하는 방식으로 진행되었다.

IQR 방법은 총 2,905개의 이상치를 식별하였고, Z-score 방법은 736개의 이상치를 검출하였다. 반면, 본 연구에서 제안된 방법은 4,715개의 이상치를 검출해내었고, 특히 충전시간이 5분 미만임에도 불구하고 충전량이 과도하게 높은 데이터를 이상치로 분류하는 데 효과적이었다. 이는 IQR 방법보다 약 62.31%, Z-score 방법보다 약 540.63% 더 많은 이상치를 식별한 것으로 볼 수 있다.

III. 결론

본 논문에서는 환경부 소속 전기차 충전소의 2022년 데이터를 분석하여, 다양한 유형의 이상치를 검출하고 이를 보정하는 방법을 제시하였다. 제시된 방법으로 검출 및 보정을 진행한 결과, 총 6,721,672건의 데이터 중 4,715건의 이상치 (전체 데이터의 0.07%)를 검출하였다. 이 중 전기차 충전 데이터의 특성을 고려하여, 3,441건의 데이터를 보정할 수 있었다.

이러한 정제된 데이터셋은 향후 전기차 충전 인프라의 최적화에 중요한 기초 자료로 활용될 수 있다. 특히, 충전소의 위치 최적화, 충전 시간의 효율성 개선, 그리고 충전 요금 정책의 조정과 같은 분야에서 응용될 수 있다. 또한, 정제된 데이터를 활용하여 시간대별, 지역별 충전 패턴을 분석함으로써, 충전 인프라의 수요 예측 모델을 개발하는 것도 가능하다.

ACKNOWLEDGMENT

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2024-2021-0-01835) supervised by the IITP (Institute of Information & Communications Technology Planning & Evaluation).

참고 문헌

- [1] IEA, "Global EV Outlook 2024," IEA, 2023. (<https://www.iea.org/reports/global-ev-outlook-2024>) (accessed May 07, 2024).
- [2] Yang et al. "A data-driven approach for optimizing the EV charging stations network." IEEE Access 8 (2020): 118572-118592.
- [3] Ye et al. "Short-term Forecasting for Utilization Rates of Electric Vehicle Charging Stations." 2021 IEEE International Smart Cities Conference (ISC2). IEEE, 2021.
- [4] Huang et al. "EV charging load profile identification and seasonal difference analysis via charging sessions data of charging stations." Energy 288 (2024): 129771.
- [5] Hafeez et al. "Utilization of EV charging station in demand side management using deep learning method." IEEE Access 11 (2023): 8747-8760.
- [6] 조재환, "'10일째 충전중'...환경부, 전기차 충전 정보 혼란," August 25, 2022. (<https://www.econovill.com/news/articleView.html?idxno=587199>) (accessed May 07, 2024).