

강화학습 적용을 위한 도시철도 전력사용량 데이터 전처리 연구

A study of preprocessing of Electricity Consumption Data for Urban Railway system for Reinforcement Learning

정호성*[†], 신승권*, 김형철*, 윤치명*

Hosung Jung*[†], Seungkwon Shin*, Hyungchul Kim*, Chimyeong Yun*

초 록 본 논문은 에너지 최적화 분야에 강화학습을 적용하기 위한 데이터 전처리 방안을 제시하였다. 실제 운영노선에서 측정된 2년간의 데이터를 바탕으로 정규화 과정을 통해 기준 전력소비 패턴을 제시하고 시계열 비교를 통해 기준모델을 벗어난 이상치 데이터를 제거하였다. 측정 샘플을 대상으로 대략 45.3%의 이상치 데이터를 제거하였다. 이와 같이 이상치가 제거된 데이터는 에너지 최적 활용을 위한 강화학습에 적합한 데이터로 활용 가능할 것이다.

주요어 : 강화학습, 전력사용량, 데이터 전처리, 정규화, 시계열 분석

1. 서 론

최근 들어 여러 산업분야에서 인공지능 기술의 적용 연구가 활발하게 진행되고 있다. 철도 분야에서도 에너지 최적 활용을 위해 실제 운영되고 있는 도시철도 전력사용량을 활용하여 강화학습을 적용하는 연구가 진행되고 있다 [1,2]. 이러한 강화학습 적용을 위해서는 학습에 적용되는 전력 데이터가 원하는 형태로 되어 있지 않은 경우에는 학습에 실패할 확률이 높아진다. 따라서 본 논문에서는 강화학습 적용을 위한 실제 운영되는 노선에서 측정된 전력 데이터에서 비정상적인 데이터를 제거하고 학습에 적합한 데이터로 가공하기 위한 전력사용량 데이터 전처리에 대한 방안을 제시하였다.

2. 본 론

2.1 이상치 제거 기법

운영노선에서 센서 고장, 통신 이상 등의 다양한 원인으로 인하여 그림 1과 같이 취득된 전력사용량 데이터에 이상치가 발생한다.

이러한 이상치 데이터는 하루 전체 데이터 또는 일정 시간에만 이상치가 발생할 수 있다.

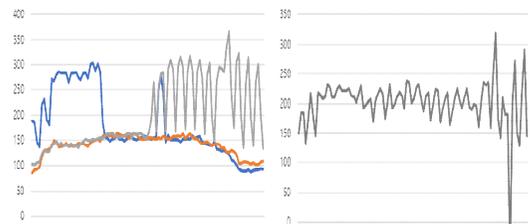


Fig. 1 Data samples including abnormal data

본 논문에서는 일일 기준으로 강화학습을 수행하기 위해 2년간의 전력 데이터 표준 분포를 제시하고 시계열로 시간에 따른 정규분포를 벗어난 경우에 해당 일에 대한 전력사용량 데이터를 제거하였다. 그림 2는 시계열 데이터에 대한 이상치 탐지 기법을 제시한 것이다. 다만, 실제 측정되는 시간대별 전력사용량이 정규화된 표준 분포로 될 수 없기 때문에 데이터 정규화(Normalization) 과정을 통해 데이터를 0~1의 값으로 나타내는 방법으로, 수요의 값보다는 패턴을 중심으로 판단하기 위하여 일간 수요를 정규화시켜 패턴화하였다. 식 (1)은 min-max 정규화 기법을 사용하였으며 매시간의 정규화 값을 구하는 과정이다[3].

† 교신저자: 한국철도기술연구원 전철전력연구팀
(hsjung@krri.re.kr)

* 한국철도기술연구원 전철전력연구팀

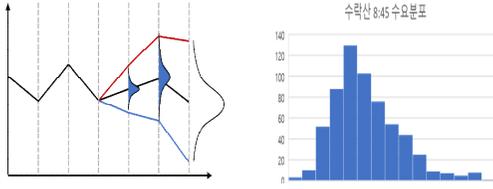


Fig. 2 Detecting outliers in time series power data

$$\overline{P_t^{Load}} = \frac{P_t^{Load} - P_{\min}^{Load}}{P_{\max}^{Load} - P_{\min}^{Load}} \quad (1)$$

$\overline{P_t^{Load}}$ 는 시간 t 에서의 정규화된 전력 값을 나타내며, P_{\max}^{Load} 및 P_{\min}^{Load} 은 해당일의 최대 및 최소 전력 값을 나타낸다. 그림 2는 이상치 데이터를 제거하고 정규화된 전력데이터를 통한 기준패턴을 제시한 것이다.

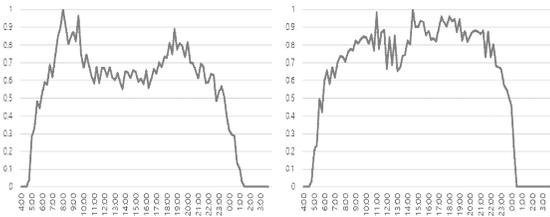


Fig. 3 Standard power data pattern(weekday, weekend)

그림 2와 같이 기준 전력사용량 패턴이 확정되면 식 (2)과 같이 기준치를 초과하는 데이터가 발생하는 경우 해당 일에 대한 데이터를 제거하는 방식으로 학습데이터 전처리를 수행하였다. 다만 전력사용량 패턴이 계절에 따라 상이하므로 실제 계측된 데이터를 이용하여 식 (3)과 같이 기준패턴을 지속적으로 업데이트를 수행하였다.

$$\left| \overline{P_t^{Load}} - \widehat{P_t^{Load}} \right| > \varepsilon \quad (2)$$

$$\widehat{P_t^{Load}} = \alpha \overline{P_t^{Load}} + (1 - \alpha) \widehat{P_t^{Load}} \quad (3)$$

α 는 0~1 사이의 값을 가지는 계수로 α 값이 클수록 기준패턴은 최근의 부하 패턴에 가까워진다.

2.2 이상치 제거 알고리즘 결과

표 1은 실제 운영노선에서 일일 전력사용량 데이터 수로부터 이상치 제거 알고리즘을 적

용하여 제거한 데이터와 학습에 활용가능한 최종 학습데이터 비율로 전체 데이터 중 대략 45% 데이터가 제거되었다. 또한 그림 4는 여름과 겨울의 전력사용 패턴에 따라 남은 일일 학습데이터 샘플로 여름에는 낮 시간대에 겨울에는 저녁/밤 시간대에 전력소비량이 증가함을 확인할 수 있다.

Table 1 Removed ratio

	Total data	Removed data	Remained data	Removed ratio
Number of data	880	399	481	45.3%

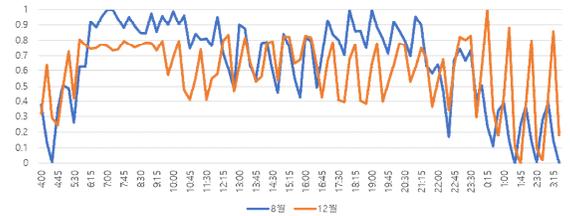


Fig. 4 Remained data samples(summer, winter)

3. 결론

본 논문은 실제 운영노선에서 측정된 2년간의 데이터를 바탕으로 강화학습을 위한 데이터 전처리를 위해 정규화 과정을 통해 기준 전력소비 패턴을 제시하고 시계열 비교를 통해 기준모델을 벗어난 이상치 데이터를 제거하였다. 측정 샘플을 대상으로 대략 45.3%의 이상치 데이터를 제거하여 강화학습에 활용할 수 있을 것이다.

참고문헌

- [1] 고락경(2020) 도시철도 시스템 피크요금 저감을 위한 강화학습 기반 에너지저장장치 제어 기법, 대한전기학회 하계학술대회, pp. 2,019-2,020
- [2] 운영란(2016), Machine Learning 기법을 이용한 BEMS 에너지 데이터 분석 사례, 설비공학논문집, 45(10), pp. 40-48
- [3] 신동하(2016), 하계 전력수요 예측을 위한 딥러닝 입력패턴에 관한 연구, 한국정보기술학회논문지, 14(11), pp. 127-134