

GAN 기반 건물 전력소비량 데이터 이상 탐지

황준화, 김보우, 서동준*
경북대학교

*corresponding author

{jhhwang1907, kbw5913, *dongjunsuh}@knu.ac.kr

GAN-based Anomaly Detection for Electricity Consumption Data

Junhwa Hwang, Bowoo Kim, Dongjun Suh*
Kyungpook National University

요약

이상 탐지는 데이터에서 비정상적이거나 예상치 못한 패턴을 식별하는 방법으로 금융, 의료, 사이버 보안 및 제조와 같은 다양한 분야에서 유용할 수 있다. 대규모 빅 데이터의 이상 데이터는 데이터 처리 및 분석의 정확성과 효율성 감소, 편향된 결과로 이어질 수 있다. 본 연구는 시계열 데이터인 건물의 전력소비량 데이터를 활용하여 GAN(Generative Adversarial Network) 기반 이상 탐지 모델을 적용하여 이상 데이터 탐지를 진행하였다. 이상 데이터 데이터 유형별로 이상 탐지 및 성능 평가를 진행하였다. 두 가지 이상 데이터 유형에서 높은 탐지 성능을 확인하였다.

I. 서론

이상 탐지 기술은 빅데이터 활용에 있어 중요한 부분이다. 방대한 크기의 빅데이터에서의 이상 데이터는 불가피하게 존재하게 된다. 이상 데이터를 처리하는 방법은 다양하지만 그 중 이상 탐지는 기존 데이터 분포에 포함되기 어려운 극단적이고 비정상적인 데이터를 식별하는 방법이다. 이상 탐지를 통해 데이터의 잠재적인 문제를 해결하고 신뢰성을 확보할 수 있다.

건물의 전력소비량의 이상 데이터는 다양한 원인을 통해 발생한다. 장비 오작동, 인적 오류, 극단적인 기상 조건과 계절 변화가 이상 데이터의 원인이 될 수 있다. 데이터의 이상 데이터에 대한 판단은 데이터 활용을 위해 중요하게 고려되어야 한다. 이상 데이터를 포함한 데이터는 데이터 처리 분석 과정에서 정확성과 효율성을 떨어트리며 편향된 결과를 가져올 수 있다. 적절한 이상 데이터 식별과 분류 및 처리는 데이터 활용과 분석의 결과에서 높은 품질을 얻을 수 있다.

전력 소비량 데이터는 시계열 데이터로 시간의 흐름속에서 연속적인 데이터이다. 시계열 데이터는 주기성과 경향을 가지고 있어 비정상적인 주기와 경향은 이상 데이터로 판단할 수 있다. 시계열 데이터에서의 이상 탐지 모델은 비정상적인 주기와 경향을 분석하여 이상 데이터로 판단하고 처리할 수 있다 [1]. 따라서, 본 연구는 건물의 전력소비량 이상 데이터 식별 및 처리를 위하여 GAN(Generative Adversarial Network) 기반의 이상 탐지 모델을 적용하여 성능 평가하였다.

2.1 GAN

본 연구는 GAN 모델을 기반으로 하여 이상 데이터를 식별하고 처리하였다. GAN 모델은 생성기와 판별기의 두가지 네트워크로 구성되어 있다 [2]. 생성기를 통해 원본과 유사한 합성 데이터를 생성하도록 훈련하며 판별기는 생성된 데이터와 원본 데이터를 구별하도록 훈련된다.

2.2 데이터 수집 및 전처리

본 연구에서 사용된 데이터는 시간 단위 전력 소비량 데이터(MW)로 PJM Interconnections LCC 를 통해 제공받았다 [3]. 제공 받은 2011 년 1 월부터 2018 년 8 월까지 데이터 중 2017 년 1 월에서 12 월까지의 데이터를 활용하였다. Standard Scaler 를 통해 평균이 0 이고 분산이 1 인 분포로 처리하였다.

다양한 이상 데이터 탐지를 위해 정상 데이터를 2 가지 종류의 이상데이터로 치환하였다. 첫번째는 단일 이상 데이터 포인트에 해당되는 값이다. 단일 이상 데이터 포인트는 일시적인 장비 오작동 또는 인적 오류, 정전 등으로 인한 이상 데이터 생성이 이에 해당된다. 두번째는 이상 데이터 구간에 해당된다. 10 시간을 기본 이상 데이터 구간으로 선정하였다. 장비 오작동과 긴 시간의 정전 등이 해당된다. 모든 이상 데이터는 기존 데이터 분포에 포함되기 어려운 비정상적인 값으로 랜덤 적용하였다. 포인트 및 구간 이상 데이터 모두 원본데이터의 4%로 생성하였다.

II. 본론

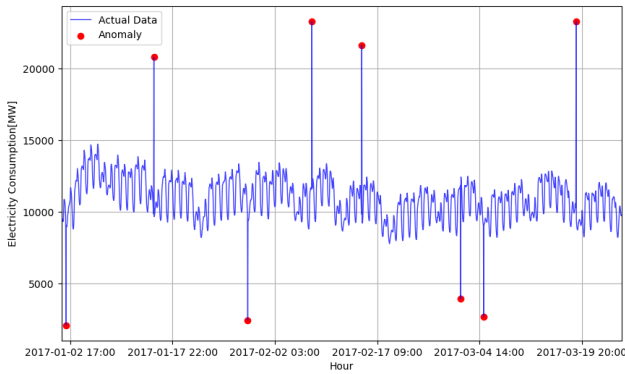


Figure 1. Graph of Point Anomaly Detection

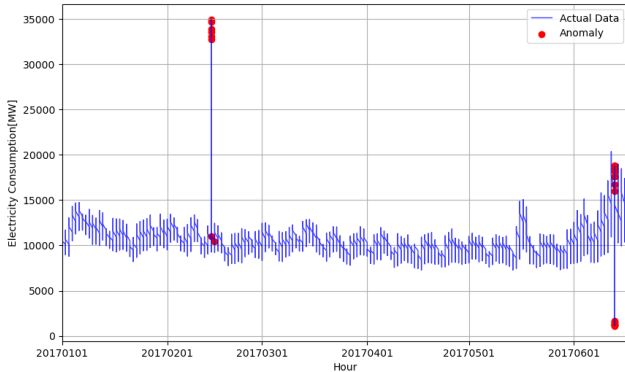


Figure 2. Graph of Subsequence Anomaly Detection

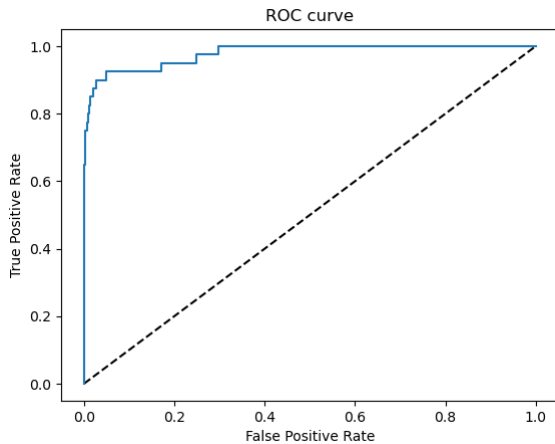


Figure 2. ROC Curve of Point Anomaly Detection

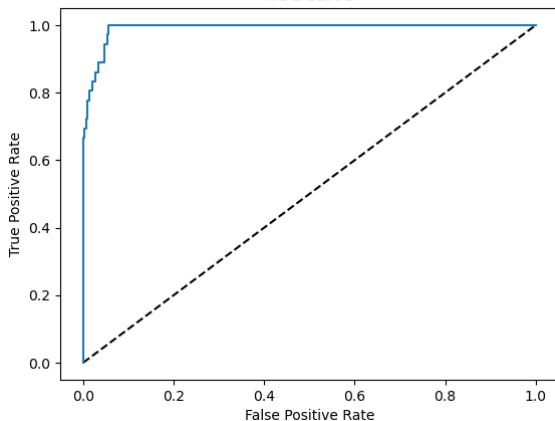


Figure 4. ROC Curve of Subsequence Anomaly Detection

Table 1. Evaluation of Anomaly Detection

AUC	Point Anomaly	Subsequence Anomaly
GAN	0.9910	0.9781

2.3 실험

전력소비량 데이터의 2 가지 종류 이상 데이터 데이터 탐지를 위하여 GAN 모델을 사용하였다. GAN 모델의 이상 데이터 탐지 성능 평가를 위하여 ROC Curve 를 사용하였다. Figure 1, 2 는 포인트 및 구간 이상 데이터 탐지 결과 그래프를 나타낸다. 시계열 데이터의 분포에 포함되기 어려운 비정상적인 데이터를 탐지한 것을 확인할 수 있다. Figure 2, 3 그리고 Table 1 은 GAN 모델의 탐지 성능을 나타내었다. 두 가지의 이상 데이터 모두 AUC 값 0.95 이상으로 높은 탐지 성능을 얻었다.

III. 결론

본 연구에서는 전력 소비량 데이터의 이상 데이터 탐지를 진행하였다. 2017 년 PJM 시간 단위 전력 소비량 데이터를 활용하였으며 다양한 이상 데이터 탐지를 위하여 두 가지 종류의 이상 데이터 데이터를 생성하였으며 GAN 모델 기반의 이상 탐지 모델을 활용하여 이상 데이터를 탐지하였다. 포인트 및 구간 이상 데이터의 두가지 형태에서 모두 높은 탐지 성능을 얻었으나 구간 이상 데이터 탐지에서 조금 낮은 성능을 확인할 수 있다.

ACKNOWLEDGMENT

본 연구는 정부(과학기술정보통신부, 교육부)의 재원으로 한국연구재단의 지원을 받아 수행되었음 (과제번호: 2021R1A5A8033165, 2021R1I1A3049503).

참 고 문 헌

- [1] 김혜연, 김형석, 강필성. (2021). WaDGAN-AD 를 이용한 전력 소비 패턴의 비지도 학습 기반 이상 탐지. 대한산업공학회지, 47(5), 421-432.
- [2] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio. Generative adversarial nets Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2, MIT Press (2014), pp. 2672-2680
- [3] "Daily energy market offer data", Jun. 2018, [online] Available: <http://www.pjm.com/markets-and-operations/energy/real-time/historical-bid-data/unit-bid.aspx>. (<http://www.nist.gov/aes>).