**빅데이터 기반 열 공급망의 이상탐지를 위한 딥러닝 알고리즘 성능 비교 분석**

**Performance Comparison of Deep Learning Algorithms for Detection of the big data based Heat Supply Chain**

오윤서

인공지능학과  
세종대학교  
서울특별시, 대한민국  
kidchammys@naver.com

|  |  |
| --- | --- |
|  | **요 약**  본 논문에서는 빅데이터 기반 열 공급망의 이상탐지를 위한 딥러닝 방법에 대해 연구하였다. 특히, 클래스 불균형 문제를 해결하기 위해 정상 데이터를 기반으로 인공 이상 데이터를 생성하는 방법을 제시하였으며 딥러닝(Deep Learning) 이상탐지 알고리즘을 사용해 성능 차이를 비교하였다. 비교 결과 1dCNN, LSTM, MLP 모델의 결과가 비슷한 정확도를 보이며 이상 탐지함을 확인하였다.  키워드: Anomaly detection, class imbalance, data augmentation, machine learning. |

|  |
| --- |
|  |

**1. 서론**

\* 교신저자

열 공급망은 중앙의 난방 기관에서 한 지역 내의 수용가에 온수나 증기를 보내는 열 공급 방식이다. 이러한 열 공급망은 보통 그 규모가 매우 크기 때문에, 한번 이상상황이 발생하면 그 피해에 따른 경제적 손실도 막대하다. 구체적으로, 열 공급망은 혹한의 날씨, 수격 작용(water hammer), 부품 노후화(밸브의 부식, 노후 배관) 등의 이유로 많은 고장이 발생한다. 실제로 2018년 백석역 인근에서 온수관 파열 사고가 발생하여 2,800가구의 난방과 온수가 끊기고, 많은 부상자와 사망자가 발생하였다. 청주에서도 비슷한 시기에 사고가 일어났으며, 두 사건의 원인은 노후화된 배관 관리 미흡이다. 최근 대한민국 수도권의 20년 넘은 노후 배관 비율이 26%를 넘는다[1]. 위와 같은 문제를 예방하기 위해서는 사전에 경고하는 것이 중요하다. 하지만 현재 빅데이터 기반 열 공급망의 고장을 사전에 감지하는 방법이 부족하다.

**2. 관련연구**

열 공급망과 규모가 유사한 에너지 공급시스템 중 하나인 전력 공급 시스템에서의 빅데이터 기반 이상탐지 연구는 여러 연구자에 의해 진행 중이다[2]. 전력 공급 시스템에서는 이러한 이상탐지 모델을 통해 과전압(over-voltages), 저전압(under-voltages)과 같은 이상 현상을 사전에 감지하여 사고가 발생하는 것을 방지한다[3].

반면에, 빅데이터 기반 열 공급 시스템에서의 이상탐지 관련 연구는 아직 활발하게 이루어지고 있지 않다. 열 공급 과정에서 사용되는 주요 핵심설비인 히트 펌프(Heat Pump)의 이상탐지 연구는 활발히 진행되었으나 [4,5], 전체적인 시스템 이상 상황에 대한 탐지 및 진단의 연구는 부족한 것이 사실이다.

**3. 제안방법**

본 논문에서는 열 공급망에서의 이상을 빅데이터 기반 딥러닝 알고리즘을 통해 사전에 감지하여 고장을 방지하는 것을 연구하였다. 일반적으로, 이상 탐지에 딥러닝 알고리즘에서 데이터 셋은 매우 중요하며, 가장 이상적인 데이터셋은 정상 데이터와 이상 데이터의 분포가 균일한 것이다. 그러나 실제 환경에서 이상상황의 발생빈도는 매우 적다. 특히 대규모 시스템에서 이상 상황 모의를 통한 데이터 수집 또한 제한적이다. 본 논문에서 고려하는 열 공급망 또한 이러한 제한점이 존재하여 수집된 데이터 셋은 정상 데이터에 비해 이상 데이터의 양이 절대적으로 부족하여 클래스 불균형 문제가 존재한다. 이에 본 논문에서는 정상 데이터를 기반으로 크기에 차이를 준 인공 이상 데이터를 생성하고, 이를 기반으로 빅데이터 기반 열 공급망 이상탐지 딥러닝 모델을 제안한다.

1dCNN 모델은 Input layer의 입력으로 시계열 데이터가 사용한다. Convolution layer는 여러 개의 convolution kernel로 구성되며, Pooling layer는 average pooling 방법을 채택한 후 Fully-connected layer에서 결과를 분류한다[6].

LSTM(Long Short-Term Memory) 모델은 기본적으로 RNN(Recurrent Neural Network)의 메모리를 확장하여 입력의 장기적인 종속성을 유지하고 학습할 수 있다. LSTM 모델은 입력에서 중요한 기능을 캡처하고 장기간에 걸쳐 이 정보를 보존한다. 따라서 LSTM 모델은 보전하거나 제거할 가치가 있는 정보를 학습한다[7].

MLP(Multi-Layer Perceptron)은 가중치를 연결하여 연결된 다수의 뉴런으로 구성된 feed-forward ANN이다. MLP 모델의 구조로 Input layer는 외부에서 입력 데이터를 수신한 다음 첫 번째 Hidden layer로 전달하고 최종적으로 Output layer에 도달할 때까지 전달된다[8].

**4. 실험방법**

**4.1. 데이터셋**

본 논문에서는 청주 지역에 대한 한국지역난방공사 열 공급량 정보(2014-2018년)데이터를 사용하였다. 데이터는 날짜, 요일, 시간, 공급량으로 이루어져 있다.

**4.2. 인공 이상 데이터 생성**

이상상태의 정도에 편차를 두어 성능을 비교하기 위해 원래의 데이터 값에서 50% 값의 증감으로 이상 데이터를 생성하였다. 인공 이상 데이터의 위치는 랜덤으로 선택되며, 열 공급량의 편차가 음의 값을 초래하면 0으로 대체하였다.

**4.3. 실험환경**

본 논문에서는 한국지역난방공사 열 공급량 데이터를 주기는 1일, 간격은 12시간으로 설정한 정규화 데이터셋 7,304건 (정상 3,936 건, 이상 3,368건)을 학습하였으며, 7:3 비율로 구분하여 각각 Training set과 Test set으로 활용하였다. 학습 횟수는 100 epoch이고 최적화는 Adam 알고리즘을 사용하였다. 모델 학습 및 테스트는 Python 3.8, Tensorflow 2.4.0 환경에서 수행하였다. 이상탐지에 활용된 모델의 기본 구조는 표 1. 과 같고, Input layer에서의 5개의 노드는 정규화 데이터셋의 속성으로 각각 평균, 최소값, 최대값, 분산, 중앙값으로 설정하였다. GPU는 NVIDIA Geforce RTX 2080 Ti을 사용하였다.

표 1. 모델 기본 구조

|  |  |
| --- | --- |
| Layer | 세부 내용 |
| Input layer | 5 nodes |
| Output layer | 2 nodes, softmax activation |

**5. 실험결과**

제안한 이상탐지 모델을 활용하여 각각의 딥러닝 알고리즘의 성능 비교 결과, 표 2. 에서와 같이 이상데이터를 동일한 편차로 생성했을 경우 이상탐지의 정확도가 비슷함을 확인하였다.

표 1. 알고리즘 정확도 비교

|  |  |
| --- | --- |
| 알고리즙 | 정확도 |
| 1dCNN | 78.695(%) |
| LSTM | 80.018(%) |
| MLP | 79.125(%) |

**6. 결론**

본 논문에서는 이상 감지를 위해 여러가지 딥러닝 알고리즘과 인공 이상 데이터 생성 방법을 제안하였다. 특히 정상 데이터를 기반으로 크기에 차이를 준 인공 이상 데이터를 생성하여 이를 기반으로 빅데이터 기반 열 공급망 이상탐지 모델을 제안했다. 실험을 통해 3가지 딥러닝 알고리즘의 성능을 비교 평가하였으며, 비슷한 정확도를 가지며 이상을 탐지함을 확인하였다.

향후의 연구에서는 이상데이터 생성과 기존 모델의 구조를 보완하여 더 높은 검출 성능을 갖도록 연구할 것이다.

**Acknowledgement**

이 논문은 산업통산자원부 "열공급망 수용가 빅데이터 기반 에너지관리 기술개발 및 실증" 사업의 연구비지원(20212020900150)에 의해 수행되었습니다.

이 논문은 2022년도 정부 (과학기술 정보통신부)의 재원으로 국가과학기술연구회의 지원을 받아 수행된 한국전기연구원 기본사업임 (No. 22A01019).

**참 고 문 헌**

[1] 전준서, “정적 구조 해석을 이용한 복합하중 및 경계조건에 따른 열수송관의 안전성 평가,” *The Society of Convergence Knowledge Transactions*, vol. 9, no. 2, pp. 1-10, Apr. 2021.

[2] P. Lipcák, M. Macak and B. Rossi, “Big Data Platform for Smart Grids Power Consumption Anomaly Detection,” in *Proc. FedCSIS*, Leipzig, Germany, Sep. 2019, pp. 771-780.

[3] L. Wen, K. Zhou, S. Yang and L. Li, “Compression of smart meter big data: A survey,” *Renewable and Sustainable Energy Reviews*, vol. 91, pp. 59-69, Aug. 2018.

[4] Z. Sun, H. Jin, J. Gu, Y. Huang, X. Wang and X. Shen, “Gradual fault early stage diagnosis for air source heat pump system using deep learning techniques,” *International Journal of Refrigeration*, vol. 107, pp. 63-72, Nov. 2019.

[5] J.L. Casteleiro-Roca, H. Quintián, J.L. Calvo-Rolle, E. Corchado, M.C. Meizoso-López and A. Piñón-Pazos, “An intelligent fault detection system for a heat pump installation based on a geothermal heat exchanger,” *Journal of Applied Logic*, vol. 17, pp. 36-47, Sep. 2016.

[6] S. Huang, J. Tang, J. Dai and Y. Wang, “Signal Status Recognition Based on 1DCNN and Its Feature Extraction Mechanism Analysis,” *Sensors*, vol.19, no. 9, pp.1-19, April. 2019.

[7] S. Narmini, N. Tavakoli and A. Namin, “The Performance of LSTM and BiLSTM in Forecasting Time Series,” *IEEE international conference on big data(big data)*, pp.3285-3292, 2019

[8] I. R. Widiasari, L. E. Nugroho and Widyawan, "Deep learning multilayer perceptron (MLP) for flood prediction model using wireless sensor network-based hydrology time series data mining", *2017 International Conference on Innovative and Creative Information Technology (ICITech)*, pp. 1-5, Nov. 2017.