

기계학습기반 GHP 시스템 이상탐지

권혜영, 윤혜주, 전창재*
세종대학교

hy913149@sju.ac.kr, i3yyj566@gmail.com, *cchun@sejong.ac.kr

Machine Learning-Based Anomaly Detection for Gas Heat Pump System

Hyeyoung Gwon, Hyeju Yun, Chang-Jae Chun*
Sejong University

요약

건물에 공급되는 가스는 여름철 중앙제어식 냉방기 운용 및 겨울철 난방에 주로 활용된다. 이러한 Gas Heat Pump (GHP) 냉난방 시스템에 문제가 발생하게 되면 작게는 냉난방 중단부터 큰 사고로까지 이어질 수 있다. 따라서 건물 가스에너지 공급 시스템의 이상여부를 정확하게 사전에 탐지하는 것이 매우 중요하다. 본 논문에서는 여름 및 겨울철 가스 사용량 데이터를 바탕으로, 건물의 가스에너지 공급 시스템 이상탐지 모델을 제안하고 분석하였다. 시뮬레이션을 통해 LightGBM 알고리즘기반 이상탐지 모델이 사전 이상탐지 성능이 가장 좋은 것을 확인했다.

I. 서론

Gas Heat Pump (GHP) 냉난방 시스템은 가스 구동 엔진을 기반으로 한 냉난방 시스템이다. GHP 냉난방 시스템은 학교, 빌딩 등 주로 중앙집중식 냉난방이 필요한 곳에 설치되며 개인이 사용량에 따른 과금을 하지 않는다는 특징을 가진다. GHP 냉난방 시스템은 대형 설비인 만큼 규모가 크고 한번 이상상황이 발생하면 작게는 냉난방 중단부터 큰 사고로까지 이어질 수 있다. 따라서 이러한 대규모 에너지 시스템의 이상여부를 에너지 사용량 시계열 데이터를 활용하여 이상 사용 패턴을 정확하게 사전에 탐지하는 것이 매우 중요하다. 최근 연구들에서는 이러한 대규모 에너지 시스템의 이상상황에 대한 고려를 하고 있다. 연구 [1]에서는 실제 데이터를 활용하여 이상상태를 구별하는 통계적 방법을 제안하였다. 하지만, 에너지 시스템을 포함한 다양한 환경에서 수행되고 있는 많은 이상탐지관련 연구들에서는 이상상태를 탐지하는 모델을 개발하기에는 실제 이상상태 데이터가 충분하지 않은 경우가 많다. 특히 GHP 냉난방 시스템과 같은 대규모 에너지 시스템에서는 이상상황을 모의하기가 불가능에 가깝고, 이상상황이 발생하는 빈도 또한 매우 낮기 때문에 실제 이상상태 데이터를 수집하기가 매우 힘들다.

본 논문에서는 계절성이 있는 GHP 냉난방 시스템의 가스 소비량 데이터를 활용하여 이상탐지 모델을 구축했다. 특히, 확보하기 매우 어려운 이상 데이터는 [2]에서의 경우처럼 인공적으로 생성하였으며, 이상데이터의 크기와 시간에 따른 빈도의 정도에 따른 이상탐지 성능을 분석했다.

II. 본론

2.1 데이터셋

본 논문에서는 세종대학교 대양 AI 센터 옥상 가스 계량기에서 수집한 가스 소비량 데이터를 사용하였다. 데이터는 년, 월, 일, 시와 대양 AI 센터 내 4 층부터 12 층까지의 가스 사용량의 보정적산을 포함한다. 보정적산 열에서 이후 값과 이전 값의 차로 구성된 시간

당 가스 사용량 데이터를 얻었다. 24 개의 연속된 시간 당 Gas 사용량 데이터를 하나의 학습 및 테스트 데이터셋으로 사용하였다. 시간 당 Gas 사용량 데이터는 시계열 데이터로서, 길이 24 의 윈도우를 2 만큼씩 이동시키면서 733 개의 학습 및 테스트를 위한 데이터셋을 얻었다. 그 중 랜덤하게 선택한 정상데이터 366 개에 일정한 상수 (1.2, 1.3, 1.5, 2, 4)를 곱하여 이상데이터를 생성했다. 이후 본문에서 이상치 크기를 *1.2, *1.3, *1.5, *2, *4 와 같이 표현했다.

2.2 모델 구성

이상탐지를 위한 전통적인 모델로는 Threshold 기반 방법을 고려했다. Threshold 기반 모델의 경우, 학습을 위한 데이터는 데이터 샘플의 평균값을 사용했다. 기계학습 모델의 경우, 학습을 위한 데이터는 데이터 샘플의 표준화된 Gas 사용량의 평균, 분산, 왜도, 첨도, 최대, 최고, 중앙값을 사용했다. 기계학습 모델로는, 먼저 Neural Network 기반 모델을 사용했으며, 입력층과 출력층, 그리고 7 개의 은닉층으로 이루어져 있다. 각 은닉층이 가지는 노드 수는 7, 50, 80, 120, 120, 80, 2 이다. 활성화함수는 Relu, 출력함수는 Softmax, 손실함수는 Categorical Crossentropy 를 사용했으며 최적화를 위해 Adam 알고리즘을 사용하였다. 학습 횟수는 150 epoch 이고 학습률은 0.0001 이다. 다음으로 Random Forest 기반 모델 LightGBM 기반 모델 Logistic Regression 기반 모델을 활용했으며, 각 모델을 학습하는데 필요한 최적의 파라미터를 grid search 방법을 사용하여 찾았다.

모든 모델의 훈련 데이터와 테스트 데이터의 비율은 8:2 로 설정하였으며, 모델 학습 및 테스트는 Python 3.8.13, Tensorflow 2.7.0 환경에서 수행하였다. Threshold 기반 모델, Neural Network 기반 모델, Random Forest 기반 모델, LightGBM 기반 모델, Logistic Regression 기반 모델을 활용하여 한 윈도우 이내에 다중으로 존재하는 (즉, 여러 빈도의) 이상 데이터와 샘플과, 정상 데이터 샘플의 4 배 이하 크기인 이상 데이터 샘플에 대한 이상 탐지 성능을 확인하였다

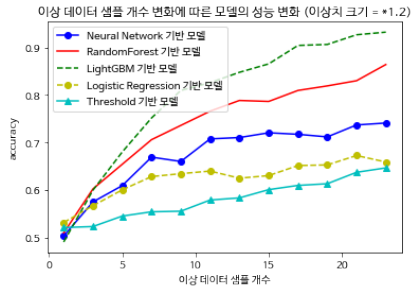


그림 1 이상 데이터 샘플 수 변화에 따른 모델의 성능 변화 (이상치 크기 = *1.2)

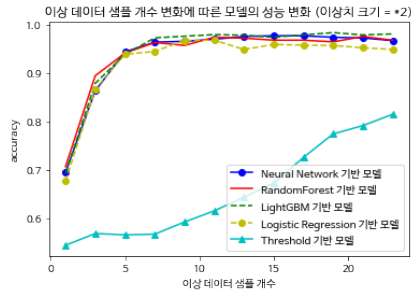


그림 2 이상 데이터 샘플 수 변화에 따른 모델의 성능 변화 (이상치 크기 = *2)

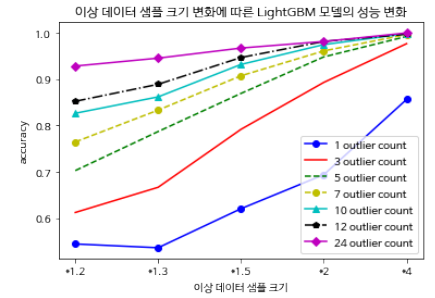


그림 3 이상 데이터 샘플 수 변화에 따른 LightGBM 모델의 성능 변화

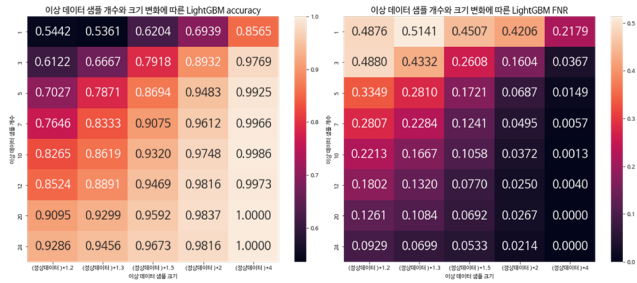


그림 4 이상 데이터 샘플 크기와 개수 변화에 따른 LightGBM 기반 모델의 정확도(좌) 및 FNR(우)

III. 실험 결과

3.1 모델 별 이상탐지 성능 확인

Threshold 기반 모델 및 4 가지의 기계학습기반 모델을 활용하여 이상치 개수 변화 (빈도)에 따른 모델의 성능을 이상치 크기가 *1.2 인 경우 (그림 1), *2 인 경우 (그림 2)를 비교하였다. 그림 3 에서는 이상치 크기변화에 따른 모델의 성능을 이상치 개수가 10 인 경우 모델 별 이상탐지 정확도를 비교했다. 그림 4 에서는 이상치 크기와 빈도에 따른 정확도 값과 FNR 값을 확인할 수 있다.

시뮬레이션을 통해 이상치 크기가 정상 데이터의 1.2 배인 데이터에서 LightGBM 의 정확도가 다른 모델에 비해서 높았으며 이는 LightGBM 모델이 이상치 크기가 작을 때, 즉 초기 이상상황에서 가장 탐지 성능이 높다는 것을 알 수 있다.

3.2 계절성 있는 데이터 기반 모델 별 성능 비교

GHP 냉난방 시스템과 같이 계절성이 있는 데이터의 경우, 여름과 겨울의 Gas 사용 양상이 달라 이상 탐지 성능에 영향을 미칠 수 있다. 이전 실험에서 가장 좋은 성능을 보인 LightGBM 기반 모델을 사용하여 다음의 네 가지 Case 에 대해 이상탐지 정확도를 확인했다. 이상 데이터 샘플 개수는 4 개의 모델 모두 한 윈도우 내 10 개로 동일하다.

Case I: 여름데이터로 모델을 학습하고, 여름데이터로 테스트

Case II: 여름 및 겨울 데이터로 모델을 학습하고 여름데이터로 테스트

Case III: 겨울데이터로 모델을 학습하고, 겨울데이터로 테스트

Case IV: 여름 및 겨울 데이터로 모델을 학습하고 겨울데이터로 테스트

표 1 의 결과 에서와 같이, 여름과 겨울을 동시에 고려한 이상탐지 모델의 성능보다 여름과 겨울 각각

이상탐지 모델을 구축한 경우의 성능이 압도적으로 좋은 것을 확인했으며, 이는 가스라는 에너지원은 동일하지만 냉방과 난방 에너지 소비 패턴의 큰 차이로 인한 결과로 분석된다.

이상치크기	Case I	Case II	Case III	Case IV
*1.2	0.829932	0.539455	0.792857	0.52251
*1.3	0.863946	0.568149	0.792857	0.500682
*1.5	0.911565	0.629842	0.871429	0.507503
*2	0.986395	0.66858	0.95	0.473397
*4	1	0.888092	0.992857	0.480218

표 1 계절 통합 LightGBM 기반 모델과 계절별 LightGBM 기반 모델의 이상 데이터 샘플의 크기에 따른 성능비교

IV. 결론

본 논문에서는 인공적으로 이상 데이터를 생성하였고, 이를 활용하여 GHP 시스템의 이상상태를 탐지하는 모델을 제안했다. 또한 GHP 냉난방 시스템과 같이 계절성이 있는 데이터의 이상탐지의 경우, 여름과 겨울 모델을 각각 구성하는 것이 더 좋은 성능을 보임을 확인하였다.

ACKNOWLEDGMENT

이 논문은 2022 년도 정부(중소벤처기업부)의 재원으로 중소기업기술정보진흥원-제조데이터 공동활용 플랫폼 기술개발(인공지능비전 기반, 데이터 증강 및 결함 탐지 플랫폼 개발)사업의 지원을 받아 수행된 연구임(RS-2022-00156456). 이 논문은 2022 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2022-0-00106, 복합도메인의 에너지 수요 빅데이터를 활용한 설명 가능한 인공지능 기반 진단 및 분석 프레임워크 개발).

참 고 문 헌

- [1] X. Li, C. P. Bowers, and T. Schnier, "Classification of energy consumption in buildings with outlier detection," *IEEE Trans. Ind. Electron.*, vol. 57, no. 11, pp. 3639–3644, Nov. 2010.
- [2] M. De Nadai and M. van Someren, "Short-term anomaly detection in gas consumption through ARIMA and artificial neural network forecast," *IEEE workshop on Environmental, energy and structural monitoring systems*, pp. 250–255, 2015.