

실시간 ESS 배터리 이상징후 탐지에서의 Spectrogram 기반 One-Class 모델 비교 연구

이연서, 김파란하늘, 현장훈*

국립한밭대학교

dustj0406@naver.com, light040901@gmail.com, jhhyeon@hanbat.ac.kr*

A Comparative Study of Spectrogram-Based One-Class Models for Real-Time Anomaly Detection in ESS Batteries

Yeonseo Lee, Paranhaneul Kim, Janghun Hyeon*

Hanbat National University

요약

최근 ESS 배터리 사용 증가에 따라 ESS 배터리 화재도 증가하고 있다. 특히, 리튬이온배터리로 인한 ESS 배터리 화재는 열폭주로 인해 발생하며, 열폭주로 인한 화재는 인접한 배터리로 빠르게 전파되어 화재 발생 전 배터리의 정보를 이용하여 사전에 예측해야 한다. 배터리의 실제 상태와 밀접한 관련이 있는 화재의 경우 정적인 임계값만 사용하여 화재를 예측하는 것은 노이즈에 민감하다. 따라서, 본 논문에서는 3가지의 One-Class Classification을 수행하여 본 연구에서 구축한 데이터에 대해 비교를 진행한다. 사용한 성능지표는 Precision, Recall, F1 Score, ROC-AUC, PR-AUC로 총 5가지이며 모든 성능지표에서 Local Outlier Factor의 성능이 가장 높았다. 이는 Spectrogram의 고차원 특징공간에서 정상 데이터가 형성하는 Manifold가 불규칙한 밀도 분포를 가지며, LOF가 이러한 Local Density Variation을 효과적으로 포착했기 때문이다.

I. 서론

최근 전력을 저장해두었다가 필요한 시기에 전력을 공급하는 시스템인 ESS (Energy Storage System) 배터리 사용량이 증가함에 따라 ESS 배터리 화재도 빈번히 증가하고 있다. ESS 배터리 화재는 리튬이온배터리로 인한 열폭주로 인해 발생하며 이는 주변 배터리로의 전파가 빨라 단시간 내에 더 큰 화재가 발생할 수 있다는 문제가 존재한다. 이러한 문제를 해결하기 위해서는 화재가 발생하기 전 배터리의 정보를 사용하여 화재 발생 여부를 조기에 예측하여야 한다.

간단한 방식의 기존 화재 예측 알고리즘은 특정상황에 대한 지표값이 사전에 정의된 임계값을 초과하면 화재라고 감지하기 때문에 노이즈가 존재할 때 높은 오경보율을 나타낼 수 있다.[1]

이러한 문제점을 해결하기 위해 본 논문에서는 Spectrogram으로 이미지화 후 OCC (One-Class Classification) Algorithm을 수행한다. 본 연구에서 구축한 데이터에 대해 3가지 OCC (One-Class Classification) Algorithm을 수행하여 비교를 진행한다.

이 다차원 배열로 들어오면 전처리 후 Spectrogram을 생성하기 위해 다차원 배열을 평탄화하여 STFT (Short-Time Fourier Transform)을 수행하여 Spectrogram으로 이미지화한다. 정상데이터의 특징만을 추출하기 위해 2차원 데이터의 지역적 특징을 추출할 수 있는 Convolutional Autoencoder(CAE)를 사용하여 학습을 진행한다. CAE에서 추출된 Latent Vector를 OCC Algorithm의 입력으로 사용하여 이상징후 탐지를 수행한다.

2) OCC (One-Class Classification) Algorithm

OCC Algorithm은 정상 데이터의 잠재 벡터만을 이용해 결정경계를 형성하며, 이 결정 경계를 통해 주어진 데이터의 이상을 판단하는 알고리즘이다.

새로운 데이터가 들어왔을 때, 정상 데이터로 학습된 영역 내에 포함되는지 여부를 기준으로 비정상 데이터를 판단한다. 이러한 판단 방식은 노이즈 민감도가 낮아 기존 알고리즘보다 안정적인 예측이 가능하다.

본 연구에서는 OCC Algorithm은 GMM, IsoForest, LOF 총 3가지를 사용하였다.

2-1) GMM (Gaussian Mixture Model)

데이터가 여러 개의 가우시안 분포가 섞여서 만들어졌다고 가정하는 확률 모델로, 밀도 추정으로 이상의 정도를 정량화하는 데 사용된다. EM(Expectation Maximization) Algorithm으로 학습을 진행하여 MLE 관점에서 파라미터를 추정한다. Likelihood가 임계값 이하인 신호는 결함 신호로 판단한다.

2-2) IsoForest (Isolation Forest)

이상 샘플이 데이터의 최소한 영역에 위치할 가능성이 높다는 가정을 기반으로 하며 이상 샘플들은 상대적으로 적은 수의 분할만으로도 고립되어 해당 노드에 도달하기까지 짧은 경로 길이를 가진다.

II. 본론

1) 전체 알고리즘

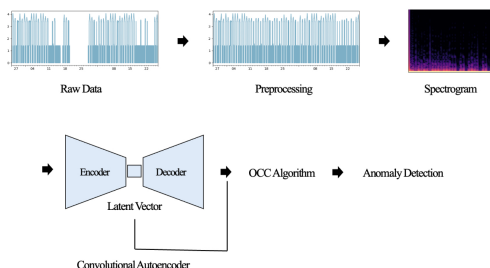


그림 1. 전체 알고리즘 개요

본 연구에서 제안한 알고리즘은 [그림 1]과 같다. ESS 배터리의 전압값

$$score = 2^{(-1/b \sum_{i=1}^b h_i(z_{\neq w})/c(n))} \quad (1)$$

이상점수는 (1)과 같이 정의되며 높은 이상점수를 갖게 되면 결함신호로 판단한다.

2-3) LOF (Local Outlier Factor)

데이터의 지역 밀도 정보를 고려하여 새로운 latent feature vector의 이웃 간 평균 거리를 통해 지역 밀도를 정의한다. 새로운 데이터가 정의된 지역 밀도로부터 얼마나 떨어졌는가로 이상점수를 계산하고 높은 이상점수를 갖게 되면 결함신호로 판단한다.

III. 연구과정

1) 데이터 구성

ESS 배터리는 Rack, Module, Cell로 구성되며 연구에서 사용한 데이터는 가장 작은 부분은 Cell에서 1분 단위로 측정된 전압값이다. 배터리는 총 화재가 나지 않은 데이터(정상)와 화재가 난 데이터(비정상)로 구분된다.

	Battery	Rack * Module	Cell Spectrogram
Train	4	696	22968
Test	7	216	17396

표 1. 데이터셋 구성

[표 1]은 본 연구에서 사용한 Spectrogram의 개수를 나타낸다. 학습을 진행할 때에는 정상 배터리 데이터 4개, 테스트를 진행할 때는 정상 배터리 데이터와 비정상 배터리 데이터를 합쳐 총 7개를 사용했다.

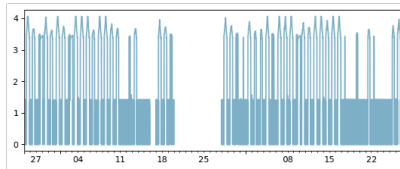


그림 2. 결측된 데이터 시각화

각 데이터의 측정날짜는 다양하게 존재하며, 하드웨어의 결함이나 네트워크 문제와 같은 불안정한 센싱 환경에 노출되는 경우 값이 전달되지 않아 [그림 2]와 같이 결측되는 날짜가 존재한다.[2] 0으로 측정된 값들도 위와 같은 이유로 간주하여 결측치로 처리하였다.

2) 전처리

위에서 서술한 결측치의 길이가 하루 단위일 경우 결측된 날짜의 바로 이전 날짜의 값을 복사하여 보간을 진행하였고, 결측된 날짜가 2일 이상일 경우 결측된 날짜의 길이의 이전 날짜부터 결측이 시작된 바로 전 날짜까지의 값을 복사하여 보간을 진행하였다.

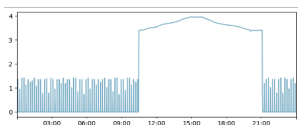


그림 3-1. 정상적인 흐름의 데이터

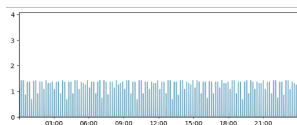


그림 3-2. 비정상적인 흐름의 데이터

위의 방법대로 보간을 진행하게 되면 전체 시계열의 패턴을 훼손하지 않고 보간을 진행할 수 있다. 하지만 [그림 3-2]와 같이 이전 날짜의 측정값이 원래 시계열 데이터의 흐름과 동일하지 않다면 다른 정보훼손이 발생

할 수 있다. 따라서, 방전종지전압인 3.3V 이하 값이 일정 비율 이상이면 결측치로 간주하고 위의 방법으로 보간을 진행한다.

3) 실험 결과

	Isolation Forest	Mixture of Gaussian	Local Outlier Factor
Precision	0.6561	0.6906	0.7914
Recall	0.4832	0.6937	0.9982
F1 Score	0.5566	0.6921	0.8828
ROC-AUC	0.2836	0.3024	0.8024
PR-AUC	0.6221	0.6574	0.9200

표 2. Test 데이터 성능

테스트 데이터의 성능을 평가하기 위해 사용한 성능지표는 Precision, Recall, F1 Score, AUC-ROC, PR-ROC 총 5가지이다. [표 2]를 보면 모든 성능지표에서 LOF가 가장 좋다. 특히, LOF의 Recall 값이 0.9982로, 조기에 화재를 예측해야 하는 본 연구의 목적에도 부합하는 알고리즘이라는 것을 알 수 있다.

4) 결과

모든 성능지표에서 좋은 성능을 보인 LOF의 경우 GMM과 다르게 데이터의 분포에 대한 가정이 없어 단순히 데이터들이 모여있는 지역적인 밀도만을 보기 때문에 정상 데이터들이 모여있는 영역 자체를 정상 영역으로 인식한다. 또한, 비정상 데이터에서 발생하는 가끔의 정상 흐름을 절대적인 밀도가 아닌 상대적인 밀도의 차이를 보기 때문에 모든 성능지표에서 좋은 성능을 보인다.

V. 결과

본 논문에서는 임계값만을 사용하여 화재를 예측하는 기존 알고리즘의 문제점을 해결하기 위해 원래 데이터를 Spectrogram으로 이미지화한 후 CAE로 학습시킨다. 학습된 Latent Vector를 3가지 OCC Algorithm의 입력으로 넣어 성능을 비교하였다.

LOF의 성능이 모든 성능지표에서 높게 나타났고 특히 Recall이 0.9982인 것은 화재를 조기에 예측해야 하는 본 연구의 목적을 달성했다고 볼 수 있다. LOF의 성능이 높았던 이유로는 데이터 분포 가정 없음, 상대적인 밀도 차이를 보는 것 두가지이다.

향후 연구에서는 머신러닝이 아닌 딥러닝을 사용한 OCC Algorithm을 본 연구에서 사용한 데이터에 적용할 예정이다.

ACKNOWLEDGMENT

본 연구는 2025년도 정부(교육부)의 재원으로 한국연구재단의 지원(No. RS-2025-25432454)과 2025년도 산업통상 자원부 및 산업기술기획평가원(KEIT) 연구비 지원(과제번호: RS-2024-00432506)에 의한 연구임.

참고문헌

- [1] Y. Heo, S. Seo, W. Shim, and J. Kang, "A Study on the Real-time Fire Classification Algorithm Based on Stacked LSTM," J. Korean Soc. Hazard Mitig., vol. 21, no. 3, pp. 93-104, Jun. 2021. doi: 10.9798/KOSHAM.2021.21.3.93
- [2] 정재익, 구태연, and 박완기, "협업필터링 오토인코더 기반 센서 데이터 결측치 처리," in 한국통신학회 학술대회논문집, 2022, pp. 80-81.