

Class Activation Map을 활용한 전력 설비 패턴의 주요원인 분석

장영준^o, 김지호*, 최영진*, 이홍철*

^o고려대학교 산업경영공학과,

*고려대학교 산업경영공학과

e-mail: {jihonav, youngjin1206, hclee}@korea.ac.kr*, yjang11@korea.ac.kr^o

Electrical equipment pattern analysis using Class Activation Map

Young-Jun Jang^o, Ji-Ho Kim*, Young-Jin Choi*, Hong-Chul lee*

^oDept. of Industrial Management Engineering, Korea University,

*Dept. of Industrial Management Engineering, Korea University

● 요약 ●

전력 생산의 효율을 높이고 지속적인 공정관리를 위해 전력 설비 데이터의 패턴을 분석하고 원인이 되는 주요 변수를 찾는 것이 중요하다. 따라서, 본 연구에서는 전력 설비 데이터의 패턴을 분석하기 위해 데이터를 군집화하고 연구 방법으로 Decision Tree, Random Forest와 ResNet을 이용하여 패턴을 분류하였다. Class Activation Map을 이용하여 설비데이터의 원인이 되는 주요 변수를 확인하였다. 본 연구를 통해 전력 설비 데이터의 분류 및 원인 분석이 가능한 통합적 솔루션을 제시하고자 한다.

키워드: Fault Detection Classification, Deep Learning, Machine Learning,
Class Activation Map, XAI

I. Introduction

최근 제조업에서는 설비 사용량을 분석하여 자원을 효율적으로 활용하기 위한 연구가 증가하고 있다.

특히, 공정 설비의 효율을 위해 다양한 센서에서 실시간으로 측정되는 시계열 데이터를 바탕으로 사용량의 패턴을 분석하거나 유사한 패턴끼리 군집화를 수행하여 정확한 분류를 수행하는 것이 중요하다. 따라서, 본 연구에서는 딥러닝을 이용한 전력 설비데이터의 분류 모델을 제안한다. 제안된 방법론으로 설비 전력 데이터를 이용하여 군집화를 통해 시계열 데이터를 분류하였다. Machine Learning, Deep Learning을 통해 분류 정확도를 확인하고 설비데이터의 원인이 되는 주요 변수를 확인함으로써 모델에 설명 가능성을 더해준다.

LSTM Autoencoder를 이용하여 신호 데이터의 특징 추출을 진행한 후, DBSCAN을 통하여 신호 데이터들을 군집화를 수행하였으며, Bidirectional LSTM을 통하여 시계열 분류 알고리즘의 성능을 측정하였다.

[2]는 시계열 데이터를 거리 기반으로 유사도를 측정하는 K-NN 알고리즘을 사용하였으며, 시계열 데이터를 문자 시퀀스로 변환하는 SAX-VSM, BOSSVS를 이용하여 시계열 분류 알고리즘의 성능을 측정하였다.

본 연구에서는 센서 데이터에서 추출된 시계열 데이터를 augmentation을 진행한 후 Machine Learning과 Deep Learning 방법론을 사용하여 분류 성능을 비교한 후 시계열 데이터의 설명 가능성 기능을 더해주고자 한다. 본 연구는 Fig. 1의 Architecture를 통해 수행하였다.

II. Preliminaries

1. Related works

공정 설비데이터의 효율적인 운영을 위해 시계열 데이터를 분류하고 분석하는 연구들은 지속적으로 진행되어왔다. [1]은

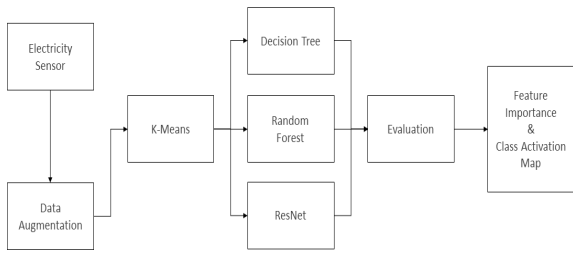


Fig. 1. Model Architecture

III. The Proposed Scheme

1. Data & Preprocessing

본 연구에서 사용된 데이터는 중소형 제조설비 모니터링 장치(MCD Unit: Measuring Communicating & Displaying Unit)에서 수집되었다. MCD Unit으로부터 수집된 데이터를 통하여 실시간 전력 사용 데이터를 계측하였다. 데이터 불균형 문제를 해결하기 위해서 기존에 노이즈를 부분별로 Scaling을 진행하였다. 스케일링은 데이터에 임의의 스칼라(scalar)값을 곱하여 데이터를 augmentation 한다. [3]

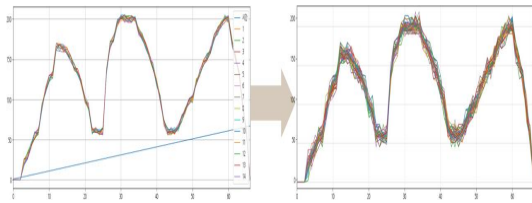


Fig. 2. Data augmentation

2. Methods

2.1 Decision Tree

Decision Tree는 데이터에 포함되어있는 패턴을 변수의 조합으로 나타내는 예측, 분류 모델을 나무의 형태로 만드는 Machine Learning 알고리즘이다. [4]

2.2 Random Forest

Random Forest는 Decision Tree 모델의 단점을 보완하기 위하여 고안된 Machine Learning 알고리즘이다. 기존에 Decision Tree 모델은 현재 가지고 있는 학습 데이터에 대한 정확도는 높으나 Overfitting의 단점이 있다. 단점을 보완하기 위해 Random Forest 여러 base 모델들의 예측을 다수결 법칙을 통하여 예측 정확성을 향상한다. [5]

2.3 ResNet

기존의 Convolutional Neural Network (CNN) 모델들은 깊은 Layer의 학습을 통해 학습 데이터에 존재하는 대표적인 특징을 추출함으로써 모델의 좋은 성능을 얻었다. 하지만, Layer가 깊어

짐에 따라 오차역전파(Backpropagation) 학습 시 기울기가 소실되는 단점이 있다. 문제를 해결하기 위해 ResNet을 이용한 분류 모델을 제시한다. ResNet은 네트워크의 깊이가 깊어짐에 따라 경사가 소실되거나 증폭하는 문제를 해결하고 VGG-19 모델의 구조를 기반으로, 컨볼루션 층을 추가로 깊게 쌓은 후 ResidualBlock의 shortcut을 추가한 것이 핵심이다. [6]

3. Experimental Result

3.1 Grid Search of Random Forest

Table 1은 Grid Search 이후에 Random Forest 모델의 최적의 하이퍼 파라미터를 기술한 것이다.

Table 1. Grid Search Hyperparameter

Hyperparameter	Value
Cross Validation	5
n_estimator	400
max_depth	4

3.2 Result of Classification Model

Data augmentation을 이용하여 Scaling을 진행하여 데이터의 수를 늘려 학습에 용이한 상태로 만들었다. 학습 데이터와 테스트 데이터는 8:2 비율로 나누어서 실험을 진행하였다. Table 2에서 Random Forest가 Decision Tree보다 더 좋은 성능을 보여주는 것을 확인하였다. 모델의 성능은 94% 이상의 정확도 얻었으며 ResNet의 분류 정확도와 모든 성능이 가장 높았다.

Table 2. Experiment Result

Model	Accuracy	Precision	Recall	F1-score
Decision Tree	94%	0.95	0.95	0.95
Random Forest	96%	0.98	0.94	0.96
ResNet	98%	0.98	0.98	0.98

3.3 Class Activation Map & Feature Importance

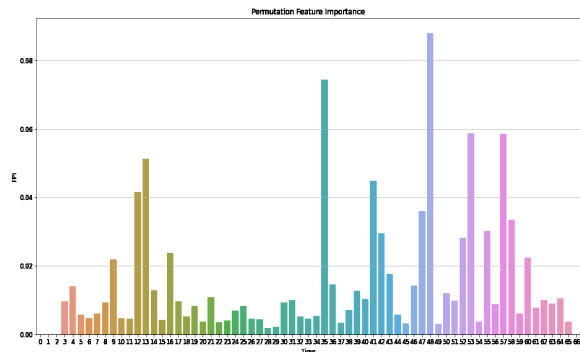


Fig. 3. RandomForest 변수 중요도

Fig. 3은 Random Forest 모델로부터 추출된 변수 중요도의 결과이다. 각각의 변수들이 서로 다른 중요도를 가지고 있는 것을 확인하였다. 본 논문의 시계열의 데이터를 분석해본 결과 48번째 주기 신호가 중요도가 가장 높은 변수로 확인하였다. Random Forest 모델은 예측력이 높지만, 다수 트리의 결과를 종합한 후 예측 결과를 확인하여 의사결정을 진행하기에 설명력을 잃는 단점을 지니고 있다.

따라서, Random Forest 모델의 가지고 있는 설명력을 잃게 되는 단점을 보완하기 위하여 본 연구에서는 ResNet 모델을 이용하여 시계열 분류 성능을 측정된 후, Class Activation Map을 적용하여 시계열 데이터에서 어떠한 주기가 결정적인 영향을 미치는지 확인할 수 있도록 시각화를 진행하였다.

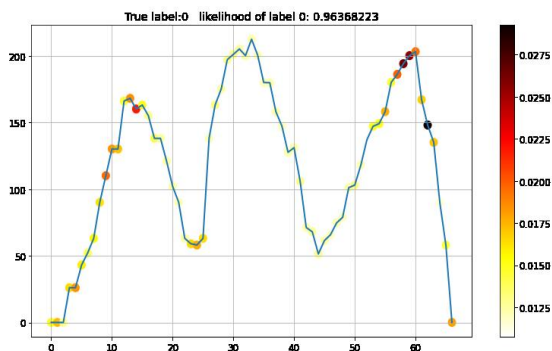


Fig. 4. ResNet 모델의 Class Activation Map

Fig 4는 ResNet 모델을 적용하여 Class Activation Map 시각화 시킨 결과이다. 색의 명암에 따라 변수 중요도를 확인할 수 있다. 어두운 빨간색을 지니고 있을수록 높은 변수 중요도인 것을 의미하며 연한 노란색을 지니고 있을수록 낮은 변수 중요도를 가지고 있는 것을 의미한다. ResNet 모델에서는 57번째 시계열 주기 신호가 가장 높은 변수 중요도인 것을 확인하였다.

IV. Conclusions

본 연구에서는 MCD Unit의 전력 설비 센서에서 수집된 소량의 시계열 데이터를 augmentation 후 머신러닝, 딥러닝을 모델을 통하여 시계열 데이터를 분류한 후에 해석 가능성을 더하였다. 적은 데이터의 패턴 분류는 일반적인 시계열 데이터 분류에 비해 많은 제약점을 가지고 있다. 본 연구는 다양한 방법을 제시하고 성능을 비교하기보다 라벨링이 되어있지 않은 짧은 시계열 데이터를 분류하기 위한 하나의 프로세스를 수립하는데 목적을 두었다. 추후 연구로는 다양한 노이즈를 적용하여 데이터 augmentation과 군집화 및 기존 시계열 유사도를 구하는 기법인 Dynamic Time Warping(동적 시간 위핑)을 적용하여 다양한 시도를 해보고자 한다. 추가로 전문가의 의견이 포함되어 있는 라벨링 된 시계열 데이터셋에 추가적인 딥러닝 모델을 구축하여 시계열 분류 모델의 성능을 고도화해보고자 한다. 향후, 다른 설비 시계열 데이터로 모델을 구축한 후, 전이 학습을 수행하는 방법을 진행할 예정이다.

ACKNOWLEDGMENT

본 연구는 4단계 두뇌한국21에 의해 지원되었습니다.

REFERENCES

- [1] Sang-Hyun Kwon, Min-Jeong An, Hong-Chul Lee. "Fault Detection and Classification of Process Cycle Signals Using Density-based Clustering and Deep Learning." *Journal of the Korean Institute of Industrial Engineers* 44.6 (2018): 475-482.
- [2] Jiyeon Park, Dongho Seo, Byungchan Choi, Haewoon Nam, 「Comparison and Analysis of Time Series Data Classification Algorithm Performance」, Korea Institute Of Communication Sciences, Proceedings of Symposium of the Korean Institute of communications and Information Sciences, 2021.2, 112
- [3] Um, Terry T., et al. "Data augmentation of wearable sensor data for parkinson's disease monitoring using convolutional neural networks." *Proceedings of the 19th ACM International Conference on Multimodal Interaction*. 2017.19th ACM International Conference on Multimodal Interaction (pp. 216-220).
- [4] Safavian, S. Rasoul, and David Landgrebe. "A survey of decision tree classifier methodology." *IEEE transactions on systems, man, and cybernetics* 21.3 (1991): 660-674.
- [5] Breiman, Leo. "Random forests." *Machine learning* 45.1 (2001): 5-32.
- [6] He, Kaiming, et al. "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.