

이상 탐지 기법을 활용한 IoT 센서 고장 진단에 관한 연구

성상하* · 최형림** · 박도명*** · † 김상진

*동아대학교 경영정보학과 박사과정생, **동아대학교 스마트물류연구센터 연구교수, ***동아대학교 스마트물류연구센터 책임연구원
† 동아대학교 경영정보학과 교수

요 약 : 고장 진단은 IoT 장비의 안전성과 효율성을 유지하는데 필요한 기술 중 하나이다. 따라서, 본 연구는 IoT 센서 데이터를 기반한 고장 진단 알고리즘을 개발하는데 목적이 있다. 본 연구는 알고리즘의 효율성을 개선하기 위해 기술통계량을 기반으로 데이터 차원을 축소하였으며, 이를 바탕으로 고장 진단 알고리즘의 정확도 및 연산시간을 개선하였다. 본 연구는 다양한 후보 알고리즘을 활용하여 고장진단을 수행하였으며, 정확도를 기반으로 가장 우수한 알고리즘을 선정하였다. 연구 결과, Isolation Forest 알고리즘이 가장 뛰어난 분류 결과를 나타내었다. 본 연구결과를 통해 IoT 센서의 안전성과 신뢰성을 향상시키는 데 도움을 줄 수 있다.

핵심용어 : 고장진단, IoT데이터, 차원축소, 머신러닝, 이상치탐지

1. 서론

연구 배경

- 최근 IoT 기술이 다양한 산업 분야에 활용되면서 IoT 센서의 중요성이 대두되고 있음
- IoT 센서는 실시간으로 데이터를 생성하고 전송하며, 이를 바탕으로 시설물의 상태 정보를 제공함
- 이러한 IoT 센서는 안정성과 신뢰성이 보장되어야 하며, 고장이 발생할 경우 신속한 대응과 예방 조치가 필요함
- 고장 진단 알고리즘을 개발함으로써, 고장 발생 시 빠른 대응이 가능해지고, 안정적인 운영을 지원함
- 따라서, 본 연구에서는 IoT 센서 데이터를 활용하여 시설물의 고장 유무를 진단할 수 있는 이상탐지 기반 알고리즘을 제안함

-1-

1. 서론

연구 필요성

- 고장 진단 알고리즘 개발을 통해 센서의 고장이 발생할 시, 빠르게 대응이 가능함
- 정확한 고장 진단을 통해 유지보수 비용 절감이 가능하며, 부품 교체나 장비 복구에 드는 비용을 최소화할 수 있음
- IoT 센서 데이터 분석을 통해 고장의 주요 원인에 대한 파악이 가능하며, 우선적인 예방정비를 수행할 수 있음
- 데이터의 이해를 통해 효과적인 전처리 방법 제시 및 저장 관리 효율성을 높일 수 있음
- 이를 위해 본 연구에서는 IoT 센서 데이터를 효과적으로 전처리할 수 있는 방법론을 제시하고, 다양한 후보 알고리즘 간 성능 비교를 통해 고장 진단에 가장 적합한 알고리즘을 제시함

-2-

2. 연구방법론

고장진단 연구방법론

- IoT 센서 데이터의 고장진단을 수행하기 위해 다음 3가지 알고리즘을 활용함
- 다음 알고리즘의 경우 다양한 종류의 데이터에 적용할 수 있으며, 데이터 분류 시 높은 성능을 나타내는 대표 알고리즘임
- 또한, 3가지 모형은 변수간의 관계를 해석하기 쉬우며, 이를 바탕으로 고장 진단 시 어떤 변수가 중요한지 이해할 수 있음

고장진단 방법론	주요 내용
Logistic Regression	<ul style="list-style-type: none"> Logistic Regression은 이진 분류의 가장 보편적으로 사용되는 선형 모델임 알고리즘의 가장 큰 장점은 해석이 쉽다는 점이며, 모델의 복잡도를 낮추고 예측 성능을 높일 수 있음 비선형 관계를 가진 데이터에 적용할 수 있도록, 선형 모델에 비선형 변환을 적용하여 학습 가능함 로지스틱 회귀는 데이터의 특성을 고려하여 모델의 성능을 높일 수 있으며, 간단한 구조로 해석할 수 있음 합성 데이터나 스칼라 데이터에 대한 적용이 쉬우며, 다양한 분야에서 활용되고 있음
Random Forest	<ul style="list-style-type: none"> Random Forest는 결정 트리의 앙상블 기법으로, 과적합을 줄이고 예측 성능을 높일 수 있음 각각의 결정 트리를 독립적으로 학습하고, 결과를 평균 내어 더 나은 결과를 얻음 다양한 변수 조합을 생성하여 학습을 통해, 모델의 성능을 높일 수 있으며, 다양한 데이터에 적용 가능함 Random Forest는 특히 고차원 데이터에 대한 학습을 위한 경우에서 강력한 성능을 보임 기능적으로 계산하여 데이터의 특성 정보를 제공할 수 있음
Isolation Forest	<ul style="list-style-type: none"> Isolation Forest는 이상치 탐지를 위한 특별한 알고리즘임 데이터를 랜덤으로 샘플링하여 이진 트리 구조로 학습하며, 이상치는 평균을 기준으로 잘림 이상치를 감지하기 위해 데이터 포인트의 노드 수를 구성하고, 이상치 탐지는 트리의 노드가 일정한 깊이에서 잘림 Isolation Forest는 높은 차원 데이터에 대한 학습을 위한 경우에서 강력한 성능을 보임 내부의 노드, 분할 기준에 대한 계산에서 가장 느린 등 다양한 분야에서 이상치 탐지에 활용됨

-3-

3. 데이터 분석

분석 프로세스

-4-

† skim10@dau.ac.kr 051-200-7484
* sangha@donga.ac.kr 051-200-8437
** hrchoi@dau.ac.kr 051-200-5611
*** home21cc@dau.ac.kr 051-200-5614

3. 데이터 분석

분석 대상 데이터

- 대상 데이터: 기계시설물 고장 예지 데이터
- 개요
 - ✓ 37MW LPAUH-Q2S 발전기 전류 데이터를 활용함
 - ✓ 본 데이터는 1초동안 수집된 전류 데이터로 총 2,000개의 값으로 구성되어 있음
 - ✓ 정상 데이터 1,208개와 고장 데이터 1,548개로 구성됨

-5-



3. 데이터 분석

데이터 전처리

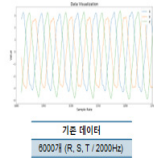


그림. 기술통계량 기반 자원 축소

기술통계량
Mean
Std.
Min
25%
Median
75%
Max

- R상, S상, T상으로 구성된 2,000개의 전류 데이터를 기술통계량으로 변환하여 총 21개 변수로 자원을 축소함
- 전류 데이터는 일정한 주기를 통해 반복적인 형태로 구성 되어 있으므로, 기술통계량을 통한 자원축소를 수행 하더라도 원본 데이터의 특성이 고유하게 반영됨
- 이에 따라 6,000개의 변수에서 21개의 변수로 자원이 축소 되었으며, 이는 전체 데이터의 0.35% 수준임

-6-



3. 데이터 분석

모델링

- 모델
 - 전처리 기법 적용 유무에 따른 성능을 측정하기 위해 6,000개의 변수 입력을 받는 모델과 21개의 변수 입력을 받는 모델을 각각 구성함
 - 실험간 편향 시드 값을 통일하여, 난수 발생으로 인한 요인을 통제함

성능지표

- Accuracy
 - Accuracy는 예측 값 중 올바르게 분류된 비율을 나타내는 성능지표로 모델의 전반적인 성능을 나타냄
 - 본 데이터의 경우 약 0.8:1로 비교적 균형적인 데이터이므로, 해당 지표를 활용함

-7-



4. 실험 결과

성능 평가 결과

- 본 연구에서 제시한 후보 모델을 통해 고장진단을 수행한 결과, 자원 축소를 수행한 Isolation Forest의 고장진단 성능이 가장 우수하게 나타남
- Random Forest와 분류 성능은 동일하게 나타났지만, 학습 시간에서 크게 개선되었음
- 또한, 모든 알고리즘이 데이터 자원축소를 수행할 경우 성능이 개선되었으며, 특히, Logistic Regression의 경우 성능이 크게 개선되었음
- 이에 따라, 기술통계량 기반 자원축소의 효과성을 입증함

분류기법	Accuracy	Training Time	분류기법	Accuracy	Training Time
Logistic Regression	49.55%	0.96 sec	Logistic Regression	76.45%	0.03 sec
Random Forest	91.30%	91.30 sec	Random Forest	99.82%	0.23 sec
Isolation Forest	92.21%	1.24 sec	Isolation Forest	99.82%	0.15 sec

< 기본 데이터 >

< 자원축소 데이터 >

-8-



5. 결론

결론

- 본 연구는 IoT 전류 데이터를 기술통계량 기반으로 자원축소하고, 이를 기반으로 고장진단을 수행할 수 있는 알고리즘을 제시하였음
- 본 연구를 수행한 결과, 기술통계량을 활용한 데이터 자원 축소 기법은 연산 효율성과 분류 성능 개선에 크게 도움을 줄 수 있음을 확인하였으며, 이를 바탕으로 향후 수집된 복잡한 데이터를 효율적으로 다룰 수 있음
- 또한, Isolation Forest의 뛰어난 고장진단 성능은 IoT 센서의 안전성과 신뢰성을 향상시키는 데 기여할 수 있음
- 특히, Isolation Forest는 고장 레이블이 없는 센서의 전류 데이터에서도 이상치 탐지를 수행할 수 있으며, 이를 통해 실제 환경에서 새로운 데이터를 신속하게 분석하고 고장진단을 수행할 수 있음
- 추후 연구에서는 추가 데이터를 활용한 모델 성능 검증 및 실제 시나리오 검증을 수행할 필요가 있음

-9-



6. 참고 문헌

참고 문헌

- >> K.H. Sun, H.Huh, Tama, B.A, S.Y Lee, J.H. Jung, and S. Lee (2020), Vision-Based Fault Diagnostics Using Explainable Deep Learning with Class Activation Map. IEEE Access 8:129169-129179.
- >> DataRPM, P. (2017), Anomaly Detection & Prediction Decoded: 6 Industries. Copious Challenges. Extraordinary Impact. Technical Report.
- >> 장현준, 김경주, 김상훈 (2022), 기계 시설물 고장 여부 및 고장 범주 분류 기법 비교, 한국통신학회 학술대회 논문집, pp.942-943
- >> 서재홍, 박준성, 유준우, 박희준 (2021), LSTM-VAE를 활용한 기계 시설물 장치의 이상 탐지 시스템, 품질경영학회지, 49권 4호, pp. 581-594.
- >> Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008). Isolation forest. In Proceedings of the 2008 Eighth IEEE International Conference on Data Mining, pp. 413-422.
- >> Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.
- >> Cox, D. R. (1958). The regression analysis of binary sequences. Journal of the Royal Statistical Society: Series B (Methodological), 20(2), 215-242.
- >> AIHub (2020), 기계 시설물 고장 예지 센서, <https://aihub.or.kr/aihubdata/data/view.do?curlMenu=115&topMenu=100&aihubDataSe=exam&dataSetSn=238>

사 사

이 논문은 2023년도 해양수산부 재원으로 해양수산과학기술진흥원의 지원을 받아 수행된 연구임 (20210636, 스마트항로표지 현장시설 고도화)

This research was supported by Korea Institute of Marine Science & Technology Promotion(KIMST) funded by the Ministry of Oceans and Fisheries(20210636).