

배전 전력설비 시계열 데이터의 비지도 학습 기반 분류 방법

박성철, 나호성*, 이성호, 정운중

한전KDN

sungchul_0126@kdn.com *sung_970529@kdn.com, letitbe_953387@kdn.com, dnswnd_0228@kdn.com

Unsupervised Learning - Based Classification of Time-Series Data from Distribution Assets

Park Sung Chul, Na Ho Sung*, Lee Sung Ho, Jung Un Jung
Kepco Kdn.

요약

본 논문은 배전설비에 부착된 방전 펄스 계측센서로부터 얻은 시간-진폭 형태의 시계열을 이용해 부분방전 여부를 자동 판별하는 방법을 제안한다. 실제 현장에서는 예방 정비에 주기적으로 이루어져 부분방전 사례 데이터는 적고 정상 데이터는 풍부하다는 점에 착안해, 정상 데이터만으로 LSTM 오토인코더(인코더-디코더)를 학습하는 비지도(One-Class) 방식을 사용한다. 운영 단계에서는 새로운 신호를 모델에 통과시켜 복원(재구성) 오차를 계산하고, 검증 데이터에서 정한 임계값(정밀도-재현율 균형 기준)을 적용해 정상/부분방전으로 판정한다. 이 방식은 라벨링 부담을 줄이고, 판정 절차를 문서화해 운영 표준화와 유지보수를 용이하게 한다. 모델은 현장 게이트웨이(IED) 또는 서버에 배치할 수 있어, 지연·전송비·보안 요구에 맞춰 유연하게 운용한다.

I. 서론

배전설비(변압기, 개폐기, 팬홀 등)에 설치된 방전 펄스 계측센서는 다양한 파형을 수집한다. 그러나 예방 정비로 인해 부분방전 발생 데이터는 희소하고, 정상 상태 데이터는 지속적으로 축적된다. 이 수집 현실을 고려하면, 여러 노이즈 유형을 모두 모아 라벨링하는 지도학습보다 정상 데이터로 기준선을 학습하고 그 기준에서의 이탈을 이상으로 판단하는 접근이 더 현실적이다. 이에 본 논문은 정상 데이터만으로 LSTM 오토인코더를 학습해 정상 파형의 전형성을 모델에 내재화하고, 운영 단계에서는 입력 신호의 복원 오차가 임계값을 초과하는지를 확인하여 부분방전 여부를 판정하는 절차를 제안한다. 임계값은 검증 데이터로 산정하며, 정밀도-재현율의 균형점을 우선 기준으로 하되 운영 정책(오탐/미탐 비용)에 따라 보수적으로 조정하고 정기적 재보정으로 분포 변화를 반영한다.

II. 본론

1) 관련 연구 및 기술배경

현장 데이터에는 정상·노이즈와 부분방전 신호가 함께 존재하지만, 부분방전 데이터는 드물고 정상 데이터는 상대적으로 풍부하다. 지도학습은 다양한 사례에 대한 정확한 라벨을 요구해 비용과 시간이 많이 든다. 본 연구는 이 한계를 줄이기 위해 정상 전용 학습(One-Class)을 채택하고, 정상 파형의 전형성을 직접 학습하도록 설계하였다.

2) 제안 방법 개요

본 연구는 정상 데이터만으로 LSTM 오토인코더(인코더-디코더)를 학습하고, 운영 단계에서는 입력 신호를 통과시켜 계산한 복원(재구성) 오차를 임계값과 비교해 정상/부분방전으로 판정하는 비지도(One-Class) 절차를 제안한다. 학습 단계에서 모델은 정상 파형을 최대한 정확히 복원하도록 훈련되어 정상 기준선이 형성되고, 운영 단계에서는 동일 절차·원도인을 적용해 산출된 복원 오차를 검증 데이터로 도출한 임계값(정밀도-재현율 균형점)에 따라 이분한다. 초기 무라벨 환경에서는 정상 데이터 상위 퍼센타일을 임계값으로 시작할 수 있으며, 데이터가 축적되면 균형점 기준으로 전환한다.

3) 데이터와 전처리

입력 데이터는 시간-진폭(mV) 형태의 펄스 시계열이다. 모델 입력의 일관성을 위해 파형을 고정 길이 원도우로 분할해 투입한다. 채널·장비 간 스케일 차이를 줄이는 정규화, DC 오프셋 제거와 과도치 완화 같은 기본 보정, 원도 길이·중첩률 등 원도잉 규칙은 운영 환경(센서 사양, 잡음 환경, 목표 민감도)에 맞춰 설정하고 문서화한다.

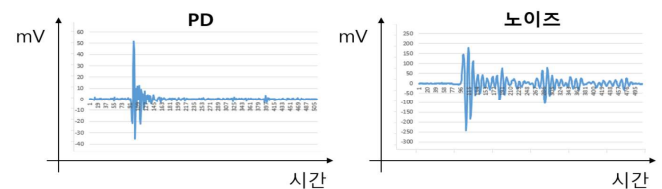


그림 1 부분방전, 노이즈 펄스

4) 모델 구조

모델은 인코더-디코더로 구성된 LSTM 오토인코더이다. 인코더는 입력 시계열에서 핵심 정보를 추려 잠재표현(압축된 특징)으로 요약하고, 디코더는 이를 바탕으로 입력을 복원한다. LSTM을 사용함으로써 시간적 연속성과 상승·하강 구간, 반복성 등 정상 파형의 시간 맥락을 안정적으로 학습할 수 있다.

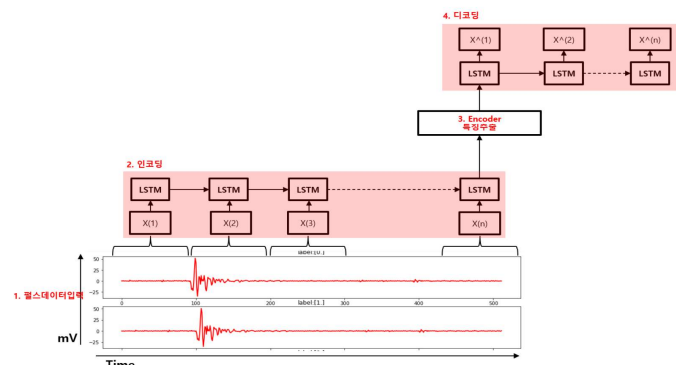


그림 2 인공지능 신경망 구성도

5) 학습 - 판정 절차

학습 단계에서는 정상 데이터만을 사용해 모델이 정상 과형을 최대한 정확히 복원하도록 훈련한다. 판정 단계에서는 운영 중 유입되는 입력을 동일 전처리·원도잉으로 모델에 통과시켜 복원 오차를 계산하고, 임계값과 비교해 정상/부분방전으로 구분한다.

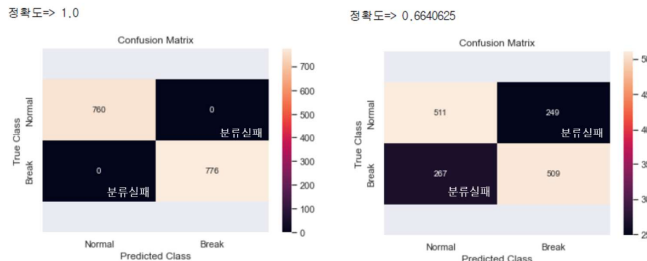


그림 3 임계값 기반 결과 혼동행렬 표기

6) 임계값 설정

임계값은 임의로 정하지 않고 검증 데이터로 산정한다. 기본 원칙은 정밀도와 재현율의 균형점을 선택하는 것이다. 무라벨 초기에는 정상 검증 데이터의 상위 퍼센타일을 임계값으로 설정해 보수적으로 시작할 수 있으며, 이후 일부 라벨이 확보되면 균형점 기준으로 전환한다. 설비 교체·계절 변화 등으로 분포가 달라질 수 있어 월/분기 단위 재보정을 권장한다.

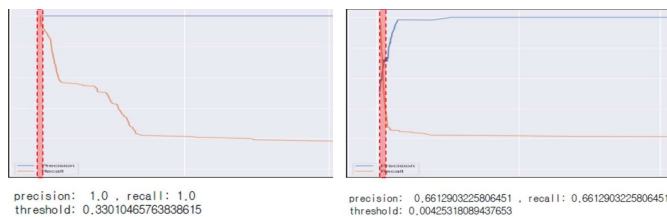


그림 4 정밀도 재현율 기반 임계값 설정

7) 시스템 구성과 배치

배치는 두 가지가 가능하다. (i) 현장 게이트웨이(IED)에 경량화 모델을 탑재해 지연과 전송량을 줄이는 방식, (ii) 서버 배치로 중앙에서 일괄 판정·모니터링·모델 교체를 관리하는 방식이다. 선택 기준은 지연 허용치, 전송비용, 보안·규제, 모델 크기이며, 변경 시 사유와 절차를 기록해 일관성을 유지한다.

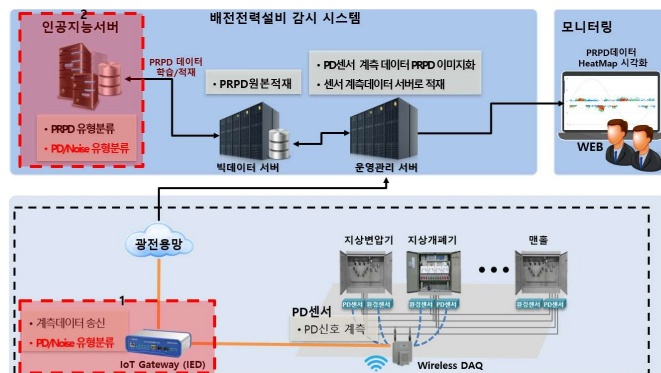


그림 5 전력설비 감시 시스템 구성도

8) 기대 효과와 한계·보완

본 방법은 정상 데이터만으로 학습하므로 라벨링 부담이 작고, 수작업 특징 정의 없이 과형 자체를 학습해 판정 기준의 일관성을 높일 수 있다. 또한 임계값 산정·재보정 절차를 문서화함으로써 운영과 감사 대응이 용이하다. 반면 성능은 정상 데이터의 대표성·품질과 전처리 안정성에 민감하며, 환경 변화로 분포 이동이 생기면 성능 저하가 있을 수 있다. 이를 보완하기 위해 임계값의 정기적 재보정, 설비/센서군별 차등 임계값 적용, 반복 발생 빈도·세기를 반영한 이벤트 단위 판단과의 결합을 권장한다.

III. 결론

본 논문은 정상 데이터만으로 LSTM 오토인코더를 학습하고, 운영 단계에서는 복원 오차의 임계값으로 정상/부분방전을 판정하는 방법을 제시하였다. 이 방식은 다양한 노이즈를 별도로 모델링하지 않아도 되어 라벨링 부담을 낮추고, 판정 절차를 표준화해 운영과 유지보수를 용이하게 한다. 임계값은 정밀도-재현율의 균형점을 우선 기준으로 하되, 현장 정책에 맞춰 보수적으로 조정하고 정기적으로 재보정해 분포 변화를 반영한다. 모델은 IED 또는 서버에 배치할 수 있어, 현장의 지연·전송비·보안 요구에 맞춰 유연하게 운용할 수 있다.

참고 문헌

- [1] Barrios, S., Comech, M. P., Gilbert, I., Orue, I., "Partial Discharge Classification Using Deep Learning Methods—Survey of Recent Progress," *Energies*, 12(13):2485, 2019. (doi:10.3390/en12132485) MDPI
- [2] Mantach, S., et al., "Deep Learning in High Voltage Engineering: A Literature Review," *Energies*, 15(14):5005, 2022. MDPI
- [3] Govindarajan, S., et al., "A review on partial discharge diagnosis in cables," *Measurement*, 216:112882, 2023. CoLab