

자동 인코더(autoencoder)를 이용한 비지도 학습 기반 베어링 이상 탐지 연구

이주하¹, 배훈기², 최명현³, 이길종⁴¹국립부경대학교 기계조선공조공학전공 학부생²국립경국대학교 기계설계공학과 학부생³서울사이버대학교 전기전자공학과 학부생⁴HD 현대중공업 성능평가연구실 책임연구원

Jooha999@naver.com, aass6192@naver.com, toto4264@naver.com, autoship88@gmail.com

Unsupervised Bearing Anomaly Detection Using an Autoencoder

Joo-Ha Lee¹, Hoon-Gi Bae², Myeong-Hyeon Choi³, Kil-Jong Lee⁴¹Major. of Mechanical, Shipbuilding & Air conditioning Engineering, Pu-Kyong National University²Dept. of Machine Design Engineering, Gyeon-Guk National University³Dept. of Electrical & Electronic Engineering, Seoul Cyber University⁴Dept. of Performance Evaluation Research, HD Hyundai Heavy Industries

요 약

본 연구는 공개 베어링 진동 데이터셋을 이용해 정상 데이터만으로 학습하는 자동 인코더(autoencoder) 기반 이상 탐지 파이프라인을 구현하고, 파일 조건(회전수 · 고장레벨)에 따른 재구성 오차 분포의 차이를 확인하였다. 데이터는 UPM Bearing Database의 구성을 따르며(F0-F4, 200/350/500 rpm, 40 kHz, 30 s, 3 축 가속도) 파일명 체계와 측정 배치 정보를 기준으로 전처리하였다[1]. 모델은 F0(정상)만으로 학습하고 F1-F4에 대해 재구성 오차를 이용해 이상 여부를 판별하였다. 결과적으로 F0 대비 F1-F4에서 오차의 중앙값 · 상위 분위가 증가하는 경향을 관찰하였고, 단순 임계치 방식의 한계를 확인하여 정상 오차의 상위 분위수(예: 99% 분위)를 기준으로 한 데이터 기반 임계치 설정 방식을 제안한다. 구현은 로컬 PC 환경에서 완료되었으며, 실시간 이식은 추후 과제로 남긴다.

1. 서론

선박 설비의 예지보전은 고장 데이터의 희소성 때문에 지도학습 적용이 어렵다. 이에 비지도 학습을 이용해 정상 패턴에서의 이탈을 탐지하는 접근이 대안으로 주목받고 있다[3]. 자동 인코더(autoencoder, AE)는 입력을 스스로 복원하도록 학습되는 신경망으로 정상 패턴에서의 재구성 오차를 이상 지표로 활용할 수 있다[2]. 본 연구는 실제 운용 가능성을 염두에 두고 간결한 단일 파이프라인을 구현해, 공개 베어링 데이터셋에 적용 · 검증한 과정을 보고한다.

2. 데이터셋 및 전처리

본 연구는 UPM Bearing Database (F0- F4, 200/350/500 rpm, 40 kHz, 30 s, 3 축 가속도)를 사용하였다. 분석은 결함 영향이 민감한 Rod_1을 기본 채널로 하고, 필요 시 3 축 평균을 비교용으로 사용하였다.

전처리는 (1) 파일명으로부터 회전수 · 결함레벨 · 반복 메타데이터 파싱, (2) 단일 채널 구성(Rod_1 기본), (3) 고정 길이 윈도우 분할 및 각 윈도우 z-정규화, (4) F0만 학습/검증, F1- F4 전 구간 테스트 순으로 수행하였다.

3. 방법

3.1 자동 인코더(autoencoder) 구조

입력은 고정 길이 시계열(1D)이다. AE는 인코더 - 병목 - 디코더 구조로 설계하고, 활성화 함수는 ReLU/선형(출력층) 조합을 사용하였다. 재구성 손실(평균제곱오차, MSE)을 최소화하도록 학습하였다. 병목 차원은 충분히 좁게 설정해 정상 패턴의 특징만 압축하도록 유도하였다[2,5].

3.2 이상도 계산 및 임계값

테스트 구간의 이상도는 재구성 오차 통계량(예: 평균 MSE)으로 계산한다. 단순 고정 임계치의 민감도/특이도 트레이드오프 문제를 완화하기 위해, 정상(F0) 학습/검증셋의 오차 분포 $p(\epsilon)$ 에서 상위 q -분위수(예: 95-99%)를 임계값으로 채택하는 데이터 기반 규칙을 제안한다. 이 방식은 운용 환경이나 회전수에 따른 분포 이동에도 비교적 견고하다[3].

3.3 구현 세부

Python(TensorFlow/Keras, NumPy, Pandas, Matplotlib) 기반으로 파이프라인을 구현하였다. 저장된 모델 로딩 시 발생하는 Keras 직렬화 이슈는 `load_model(..., compile=False)`로 해결해 재현성을 확보하였다.

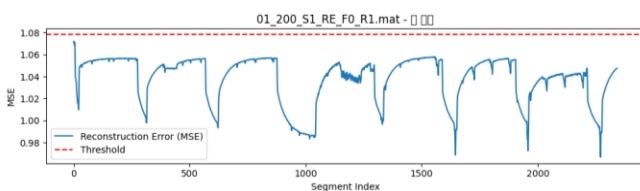
4. 실험 및 결과

4.1 설정

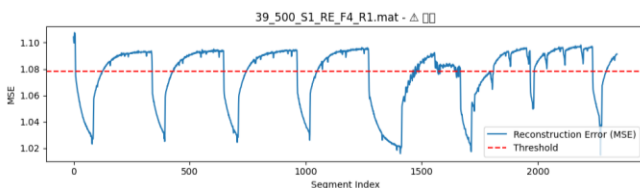
학습은 F0 전 구간을 회전수별로 균형 샘플링하여 수행하였고, 테스트는 F1-F4(200/350/500 rpm)에 대하여 Rod_1 중심으로 적용하였다. 평가는 윈도우 단위 재구성 오차 분포(중앙값 · IQR · 상위 분위)와 이상도 시계열로 진행하였다.

4.2 관찰 결과

정상(F0) 대비 결함(F1-F4)에서 오차 분포의 중앙값과 상위 분위가 전반적으로 상승하였으며, 레벨이 높을수록 증가폭이 커지는 경향이 확인되었다. 다만 일부 회전수 · 파일 조합에서는 F0 상위구간과 F1 하위구간의 중첩이 나타나 고정 임계치의 한계가 드러났다. 이에 F0 오차 분포 상위 분위(예: 99%)를 임계로 적용한 결과, F3-F4는 위험 구간이 명확히 표기되고 F1-F2는 단계적 증가 양상이 시각화되어 실무적 활용성이 높음을 확인하였다.



(그림 1) 정상 데이터(F0)의 재구성 오차 시계열.



(그림 2) 결함 데이터(F4)의 재구성 오차 시계열.

5. 논의

정상 데이터 기반 학습과 재구성 오차만으로도 결함 레벨 증가에 따른 경향을 확인하였다. 다만 회전수 · 개별 파일 조건에서 분포 중첩이 발생하며, 이는 (1) 윈도우 길이 · 중첩률, (2) 채널 결합 방식(Rod_1 단독 vs 3 축 결합), (3) 병목 크기/정규화 전략 등의 영향으로 해석된다.

고정 임계치의 한계는 데이터 기반 임계값(분위수, 극단값 이론), 적응형 기준(회전수별 분리 학습/판정), 혹은 군집/밀도 추정 결합(예: IF, kNN, KDE)로 보완 가능하다[3,4]. 또한 주파수영역 특징(예: 스펙트럼 에너지 대역비)을 AE 입력에 포함하는 방법은 회전수 차이로 인한 분포 편차를 줄이는 데 유효할 것으로 기대된다.

6. 결론 및 향후 과제

본 연구는 공개 베어링 진동 데이터셋을 이용하여 정상 데이터만으로 학습하는 AE 기반 이상 탐지 파이프라인을 구현 · 적용하고, 결함 레벨 증가에 따른 재구성 오차의 상승 경향을 확인하였다. 또한 파일 · 회전수 조건에 따라 중첩 구간이 존재함을 관찰하였으며, 정상 오차 분포 상위 분위 기반 임계값이 실무 적용 가능성이 높음을 확인하였다. 향후 과제는 (1) 회전수 조건별 도메인 정규화와 스펙트럼 특징 결합, (2) IF 등 비선형망 기법과의 앙상블, (3) 정량 지표(AUC, F1, AU-PR) 도입, (4) 엣지(임베디드) 환경 이식 및 실시간 스트리밍 대응이다.

사 사

본 논문은 해양수산부 스마트 해운물류 융합인재 및 기업지원(스마트해운물류 x ICT 멘토링)을 통해 수행한 ICT 멘토링 프로젝트 결과물입니다.

참고 문헌

- [1] Soto-Ocampo, C. R., Mera, J. M., Cano-Moreno, J. D., Garcia-Bernardo, J. L., Bearing Database: Induced Failure Test Data on Rolling Elements of a Spherical Roller Bearing (FAG 22205E1KC3), Geneva, Zenodo, 2020.
- [2] Hinton, G. E., Salakhutdinov, R. R., "Reducing the dimensionality of data with neural networks," Science, 313, 5786, 504-507, 2006.
- [3] Chandola, V., Banerjee, A., Kumar, V., "Anomaly detection: A survey," ACM Computing Surveys, 41, 3, 1-58, 2009.
- [4] Liu, F. T., Ting, K. M., Zhou, Z.-H., "Isolation Forest," IEEE International Conference on Data Mining (ICDM), Pisa, 2008, 413-422.
- [5] Goodfellow, I., Bengio, Y., Courville, A., Deep Learning, Cambridge (MA), MIT Press, 2016.