

Deep Learning Applications

AM202216701
컴퓨터공학과
고지영

Indax

1. Introduction
2. Experimental method
3. Training and Algorithm application
4. results

1. 논문소개 Introduction

시계열 특징을 갖는 선박용 공기 압축기 전류 데이터의 이상탐지 알고리즘 적용실험

Anomaly detection algorithm application experiment of air compressor current data for ships with time series characteristics

본 연구의 목적은 선박의 대표적인 보조기계 중 공기압축기의

데이터 기반 고장 진단 시스템 구현에 필요한 기초적 연구를 수행하는 것이다. 본 연구에서는 실험실 규모로 구축된 공기압축 기를 이용해 운전 데이터를 확보하였으며, 이를 기반으로 시계열 데이터의 이상치 탐지 알고리즘의 일종인 오토인코더(Autoencoder)를 적용하여 이상치 탐지 모델을 구축하였다.

또한 공기압축기의 다양한 고장 유형을 구현한 고장 모사 실험을 통해 고장 데이터를 수집하였으며 이를 이용하여 이상 탐지 성능을 검증하였다.

The purpose of this study is to analyze the air compressor .In this study, operation data was obtained using an air compressor built on a laboratory scale, and based on this, an outlier detection model was built by applying an autoencoder.

이정형. (2021). 시계열 특징을 갖는 선박용 공기 압축기 전류 데이터의 이상 탐지 알고리즘 적용 실험. 해양환경안전학회지, 27(1), 127-134.

2. 실험 방법 Experimental method

실험 대상: 선박용 공기 압축기 (Air Compressor for Ships)

-> 다양한 기계요소가 구비되어있고 잦은 운전정지로 인해 고장 발생 가능성이 높음(**high possibility of failure**)
실험에 사용된 공기 압축기는 왕복동식으로서 압축기와 압축기를 구동하는 원동기(Motor), 압축 공기가 저장되는 저장탱크(storage tank), 자동 충전을 위한 제어기로 구성(controller for automatic charging)

정상 및 고장 데이터 셋 구축(data sets): 데이터는 단변량 시계열(Univariate time-series) 특징을 갖는 공기압축기의 전류 데이터. 공기압축기의 정상 운전과 10종의 고장모드에 대해 전류데이터를 취득

Acquisition of current data for normal operation of the air compressor and 10 types of failure modes

데이터 전처리(preprocessing): 정상 운전 데이터는 단일 특징값(244,274개)

->고장과 무관한 이상치가 수집되기 때문에 이들 데이터는 Z-score법을 이용해 제거

이후 전처리 과정은 먼저 원시 데이터(Raw data)는 1초 단위로 재샘플링(Resampling)한 뒤 최대-최소 정규화를 통해 값의 범위를 0 ~ 1이 되도록 하였다.

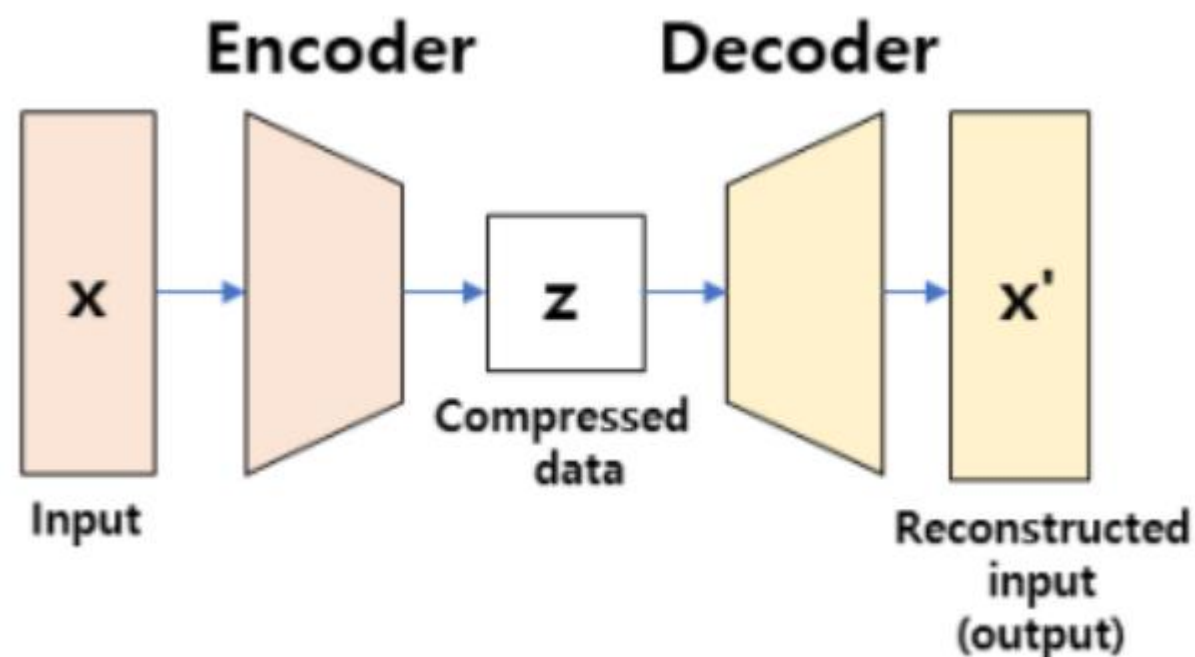
전처리 후 데이터 개수 -> 12,220개

3.학습 및 이상탐지 알고리즘 적용

Training and Algorithm application

오토인코더(Autoencoder)

신경망을 이용해 임의의 입력값과 같은 출력값으로 재생성하도록 하는 비지도 학습(Unsupervised learning) 기반의 알고리즘



-인코더와 디코더로 구성되어 있으며 입력층으로 입력된 다 차원 데이터가 은닉층으로 전달되면서 저차원으로 축소되며 이는 다시 출력층으로 전달되면서 입력데이터와 동일 차원의 데이터로 복원된다.

-은닉계층을 중심으로 인코더와 디코더는 대칭구조를 이루고 있으며, 입력계층의 노드 수와 출력계층의 노드 수는 항상 동일한 특징을 가지고 있다.

-구성하는 신경망 구조에 따라 적층형(Stacked), 변이형(Variational), 잡음제거형(Denoising) 등의 다양한 구조가 있다

-본 연구에서는 가장 단순한 형태의 **합성곱 오토인코더**(Convolutional autoencoder) 모델을 적용했다.

The input layer is reduced to a **lower dimension** as it is transmitted to the hidden layer, which is then transmitted to the output layer and restored to data of the same dimension as the input data.

Encoders and decoders are symmetrical.

3. 학습 및 이상탐지 알고리즘 적용

Training and Algorithm application

합성곱 오토인코더(Convolutional autoencoder)

:인코더(Encoder) 부에 해당하는 2개의 1차원 합성곱 층(Convolutional layer)과 디코더(Decoder) 부에 해당하는 2개의 합성곱 전치층(Convolution transpose)으로 구성되어 있으며 출력층을 제외한 모든 층에 활성화 함수는 Relu 함수 사용.

*Relu함수는 활성화 함수의 한종류로 Sigmoid와 tanh가 갖는 Gradient Vanishing 문제를 해결하기 위한 함수이다.x가 0보다 크면 기울기가 1인 직선, 0보다 작으면 함수 값이 0이 된다.sigmoid, tanh 함수보다 학습이 빠르고, 연산 비용이 적고, 구현이 매우 간단하다는 특징이 있다.

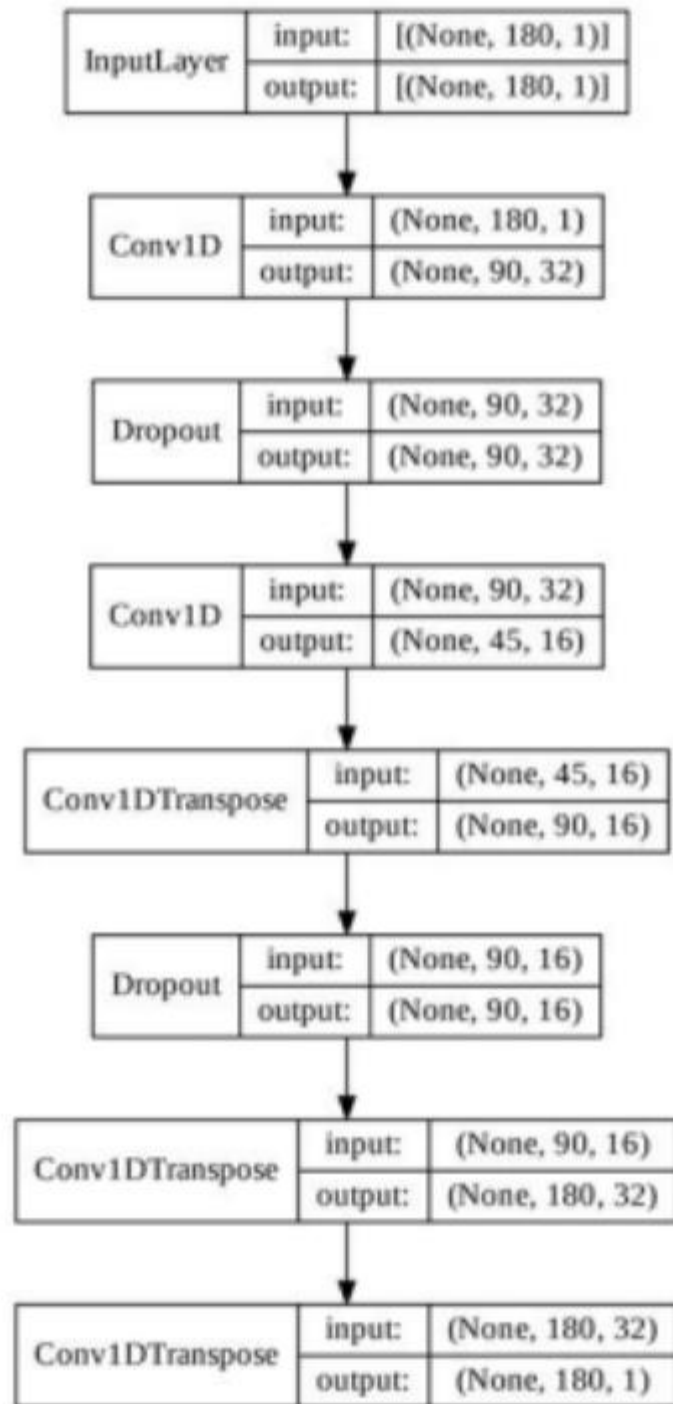


Fig. 2. Structure of a 1-D convolutional autoencoder.

Table 1. H/W and S/W specifications of workstation

Device/Library	Model/Version
CPU	Intel(R) Zeon W-2225 @4.6GHz
RAM	32G
GPU	NVIDIA Geforce RTX 2080 Ti 11GB (Cuda version: 10.1)
OS	Ubuntu 18.04
Python	3.7
Keras	2.1.5
Tensorflow	2.1.0

<연구 환경>

- 전체 데이터 셋(data sets) : ₩학습용 90% / 검증용 10%
- 학습계수(Learningrate): 0.001
- 배치크기(batch size): 128
- 5회연속 오차값개선이 없을경우 학습조기종료.
- Adam optimizer 사용
- 손실함수 : MSE

3. 학습 및 이상탐지 알고리즘 적용 Training and Algorithm application

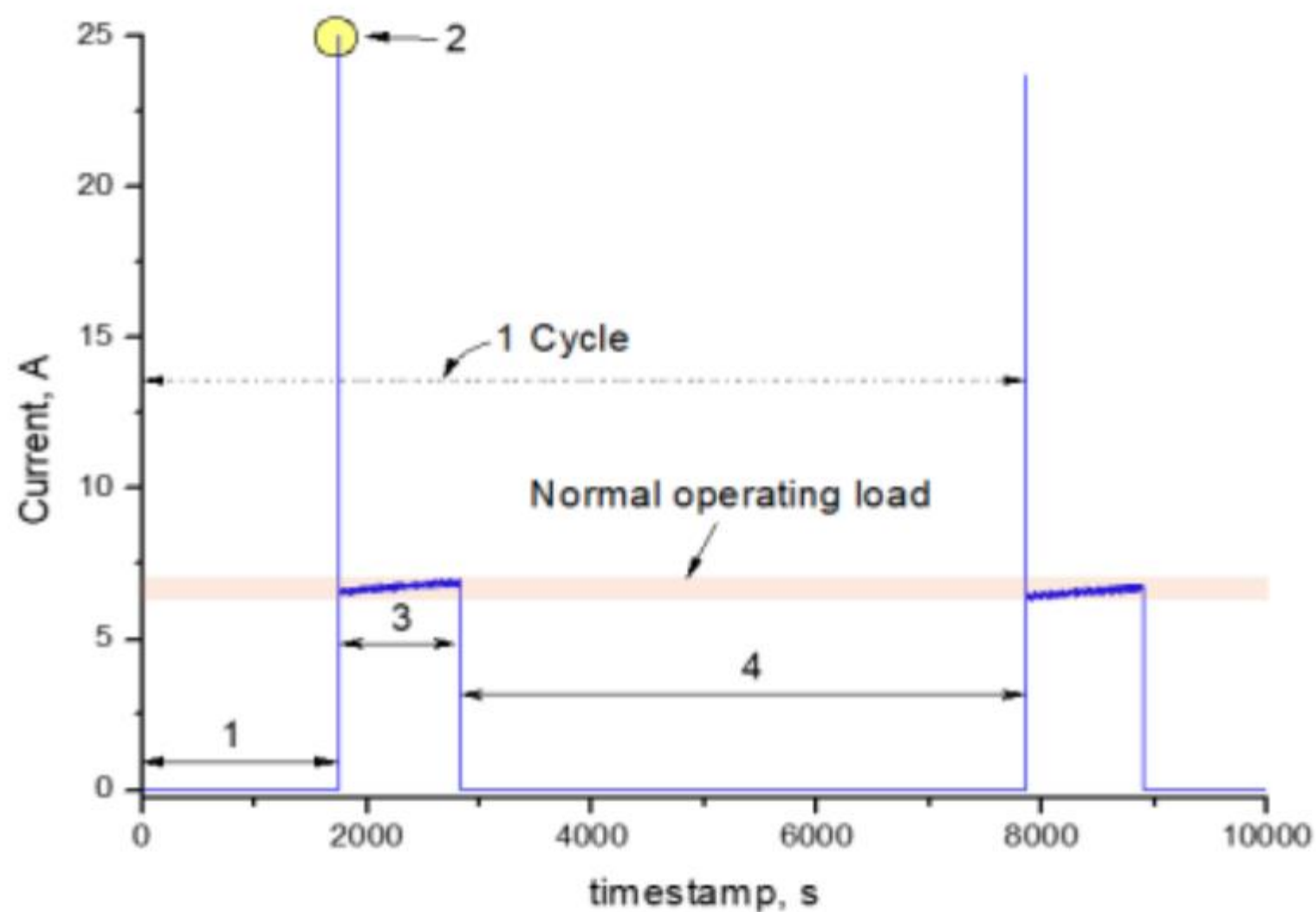


Fig. 4. Typical time-current curve of air compressor during normal operation.

공기 압축기의 정상 운전 중 1 사이클에 대한 시간에 따른 전류값의 변화 (기동-운전-자동정지에 이르는 기간)

① 공기 압축기 기동 전 대기 상태이며 0에 가까운 값

-> 압축 공기가 소모됨

Standby state before starting the air compressor, value close to zero

② 시점에서 압축기의 압력스위치의 작동에 의해 자동 기동

-> 기동부하 때문에 피크치가 관찰됨

Automatic start by the operation of the pressure switch of the compressor at this point

③ 정상 운전 시 약 7 A 정도 나타냄

-> 공기 압축기는 계속 운전되어 정해진 탱크 용량까지 보충

The air compressor continues to operate and replenishes up to the specified tank capacity.

④ 탱크의 압력이 압력스위치의 설정치 상한값에 이르면 압축기는 정지

-> 다음 기동까지 대기

When the tank pressure reaches the upper limit of the pressure switch, the compressor stops.

3. 학습 및 이상탐지 알고리즘 적용

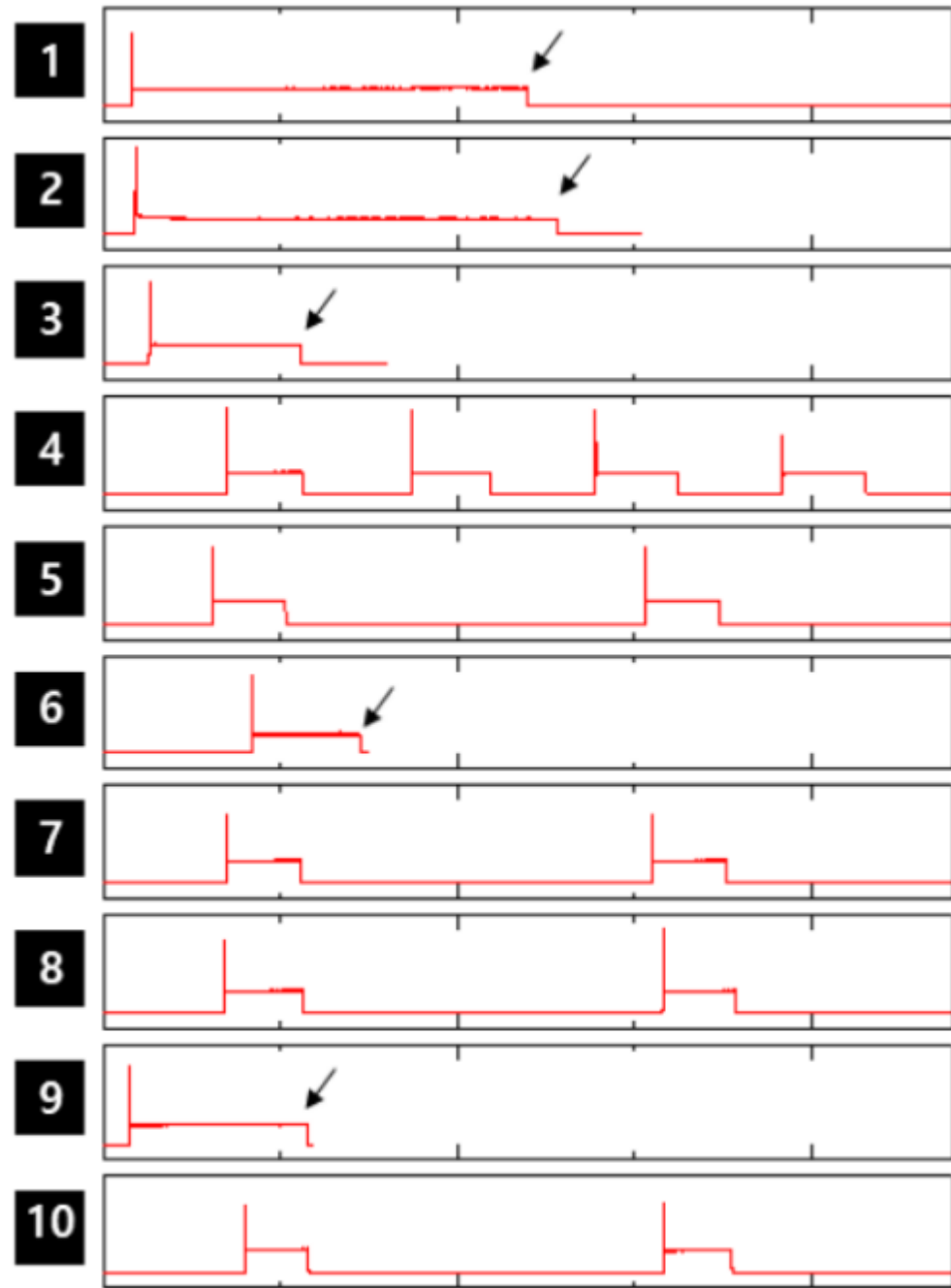


Fig. 5. Time-current curves showing temporal patterns with different fault modes of air compressor.

- 1) V-belt 이탈(V-belt departure)
- 2) 압축공기의 대량 누설(Massive leakage of compressed air)
- 3) 압축공기의 미량 누설(Very small leakage of compressed air)
- 4) 압축공기의 소량 누설(Small leakage of compressed air)
- 5) 모터 측 베어링의 손상(Damage to motor side bearings)
- 6) V벨트의 장력 부족(Insufficient tension of the V-belt)
- 7) V벨트의 장력 과다(Excessive tension in the V-belt)
- 8) 압축기 측 베어링의 손상(Damage to the bearing on the compressor side)
- 9) 압력스위치의 고장 (failure of the pressure switch)
- 10) 공기압축기의 윤활유 부족(Lack of lubricating oil in the air compressor)

10종의 고장 모드에 대한 고장 모사 실험으로부터 취득한 전류의 시계열 그래프

3.학습 및 이상탐지 알고리즘 적용

이상 기준치 설정 Abnormality threshold setting

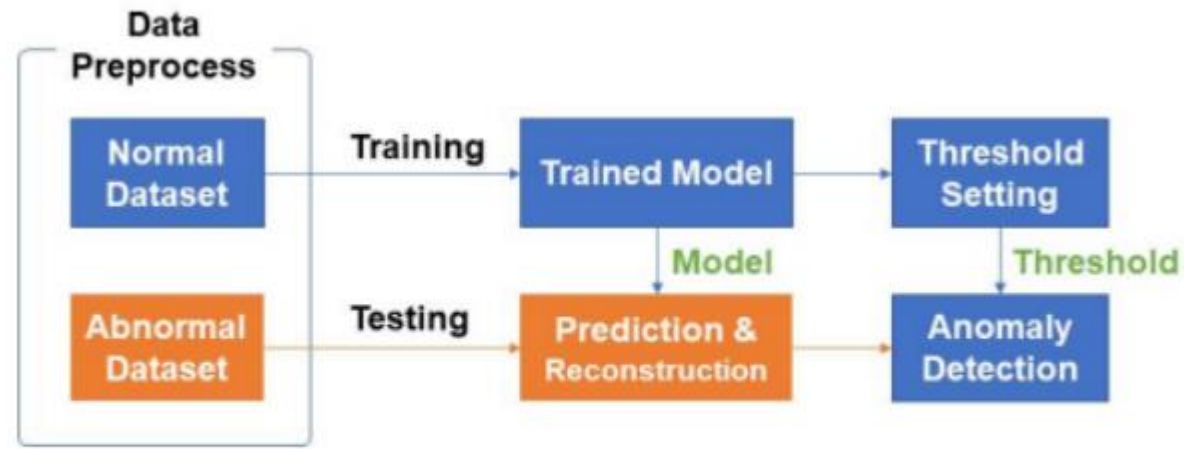


Fig. 6. Anomaly detection process using autoencoder.

- 오토인코더 기반의 이상탐지는 재구성 오차(Reconstruction error)의 계산과 임계치(Threshold) 설정에 의해 이상 여부를 판정
- 재구성 오차는 복원 오차라 고도 하며 입력과 출력 사이의 오차이며 오토인코더의 학습은 이 재구성오차가 최소화되도록 진행
- 재구성 오차는 이상치 점수(Anomaly score)로 작동하게 되며 임계치와의 비교를 통해 이상여부를 결정.

*재구성오차 :원본 데이터와 복원 데이터를 n차원의 좌표 공간안의 점으로 간주하고, 두 점의 거리를 입출력 데이터 차이의 지표로 사용하는 방식

->즉, 정상 데이터로 학습시킨 모델은 정상 데이터에 대해서는 낮은 재구성 오차를 나타내지만, 이상 데이터에 대해서는 높은 재구성 오차를 나타낸다. 정상 데이터를 대상으로 모델을 학습시키고 재구성 오차에 대한 임계치를 결정하게 되며 이를기준으로 입력 데이터의 이상/정상 여부를 판정

That is, a model trained with normal data shows a low reconstruction error for normal data, but a high reconstruction error for abnormal data. The model is trained on normal data and the threshold for the reconstruction error is determined. Based on this, it is determined whether the input data is abnormal or normal

*Reconstruction error: A method in which the original data and the restored data are regarded as points in the n-dimensional coordinate space, and the distance between the two points is used as an index of the difference between the input and output data.

3.학습 및 이상탐지 알고리즘 적용

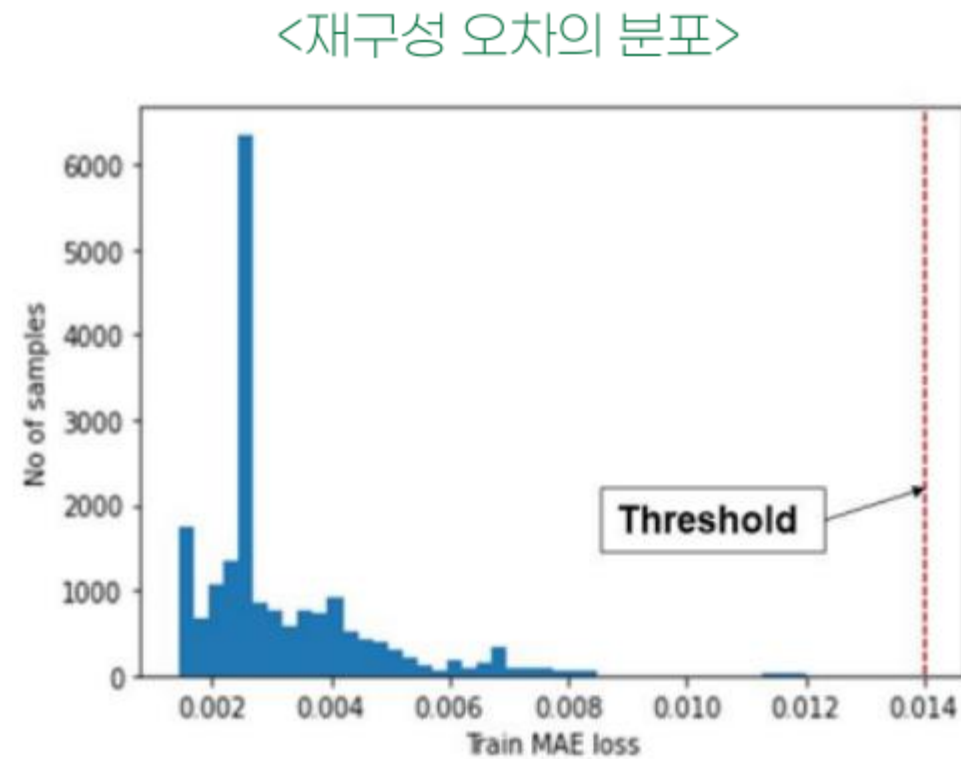


Fig. 7. Histogram of reconstruction error on normal training dataset.

-> 임계치는 오차의 최댓값인 0.014로 설정

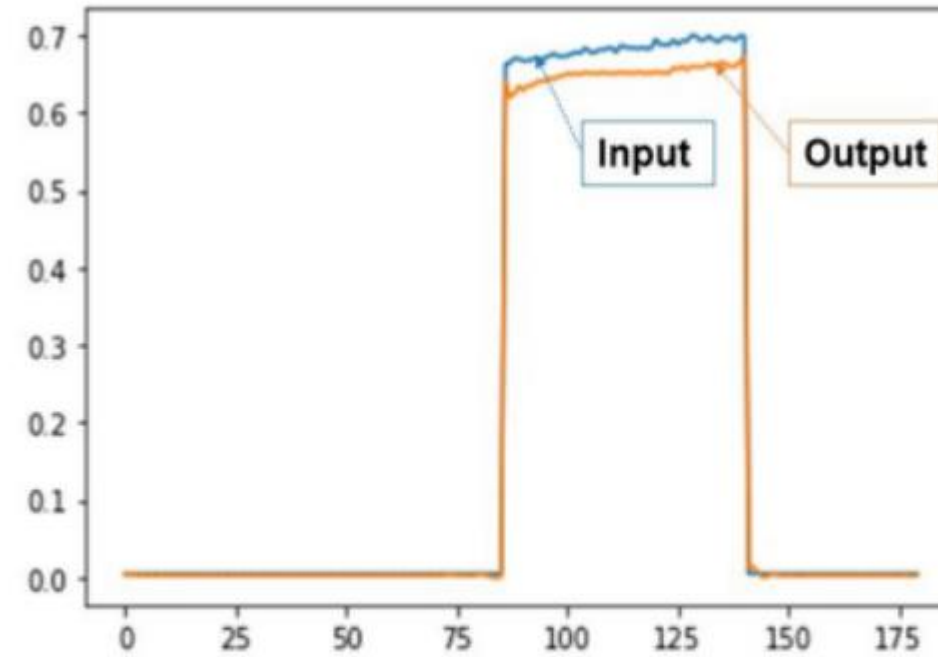


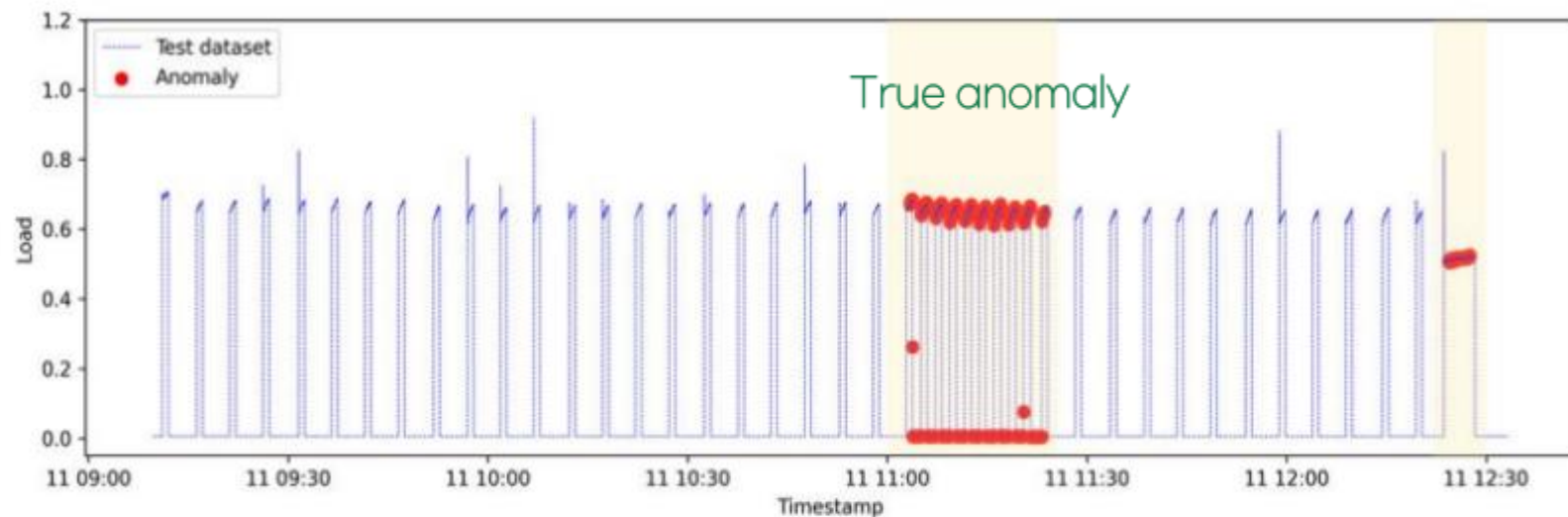
Fig. 8. Input sequential data vs reconstructed input (output) of first sample data.

입력(청색)에 대해 복원된 출력(오렌지색)이 형상적으로는 근접하게 복원되었으나 약간의 차이를 나타냄을 알 수 있다. 오토인코더의 출력이 완벽하게 입력을 복원하였다고 해서 반드시 이상탐지 성능이 좋은 것은 아니다

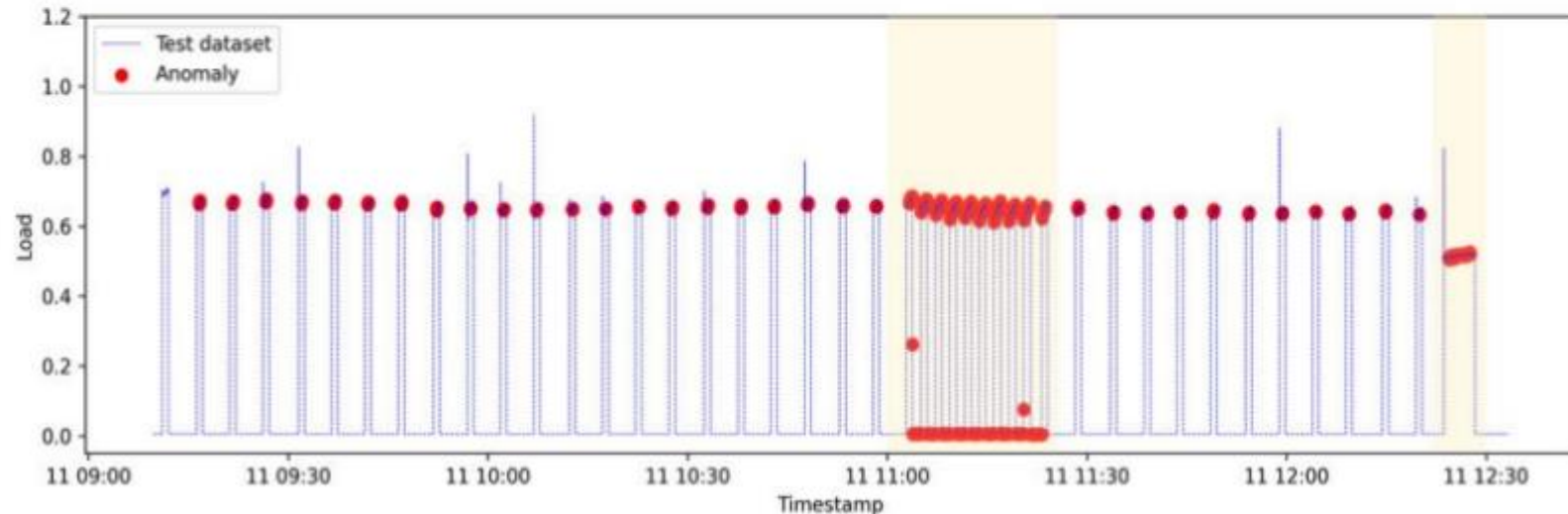
It can be seen that the restored output (orange) for the input (blue) is geometrically closely restored, but shows a slight difference. Even if the output of the autoencoder perfectly restores the input, it does not necessarily mean that the anomaly detection performance is good.

3. 학습 및 이상탐지 알고리즘 적용

이상탐지 가능성 검토 Review of anomaly detection potential



(a) 입력시퀀스의 길이가 180초(A)



(b) 입력시퀀스의 길이가 240초(B)

음영으로 표시된 부분 - > 실제 고장이 발생한 부분(True anomaly)

빨간 색 점 -> 이상탐지모델에 의해 이상이 탐지된 부분

모델 A는 2개의 실제 고장에 대해 이상탐지가 잘 되었다.
모델 B는 실제 고장에 대한 이상 탐지뿐만 아니라 정상 운전 데이터에 대해서도 이상이라 거짓 탐지한 부정오류(False negative) 문제를 보여주고 있다.

한계: 비지도학습 기반의 이상탐지 모델 특성 상 정답 레이블이 없으므로 Accuracy, Recall, Precision, F1-score 등을 통한 성능 평가를 적용하기 어려움

2개의 오토인코더를 이용한 모델 구축

Limitations: Due to the nature of the anomaly detection model based on unsupervised learning, it is difficult to apply performance evaluation through Accuracy, Recall, Precision, F1-score, etc. because there is no correct label.

4. 결론

오토인코더를 이용한 시계열 데이터의 이상탐지 모델은 입력 시퀀스의 길이와 초모수 조정에 따라 이상 탐지 성능이 상이하다는 결론을 얻었다. 다음연구에서는 이상탐지기능의 최적화를 통한 성능향상 방법을 찾는 노력이 더 필요할 것이다.

The anomaly detection model of time series data using an autoencoder obtained the theory that the anomaly detection performance is different depending on the length of the input sequence and the adjustment of the hyperparameters. In the next study, more efforts will be needed to find a way to improve performance through optimization of the anomaly detection function.

감사합니다