다중 식별자를 이용한 Adversarial Autoencoder 기반 제조 공정 이상 탐지

이승희, 백준걸†
고려대학교 산업경영공학과
{9775lsh, jungeol}@korea.ac.kr

대한산업공학회
제 16회 석사논문 경진대회
2020. 11. 13

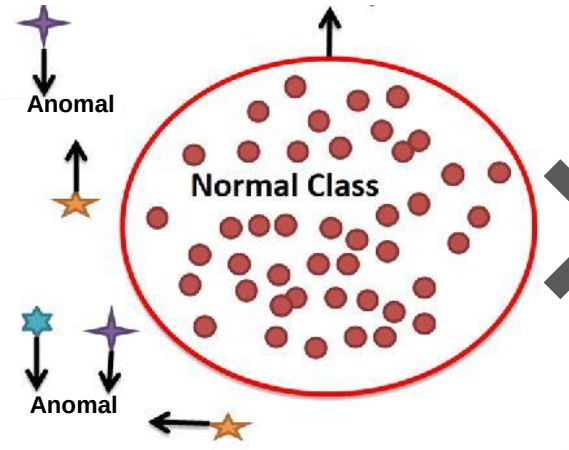
고려대학교
KOREA UNIVERSITY

SPS LAB

Contents

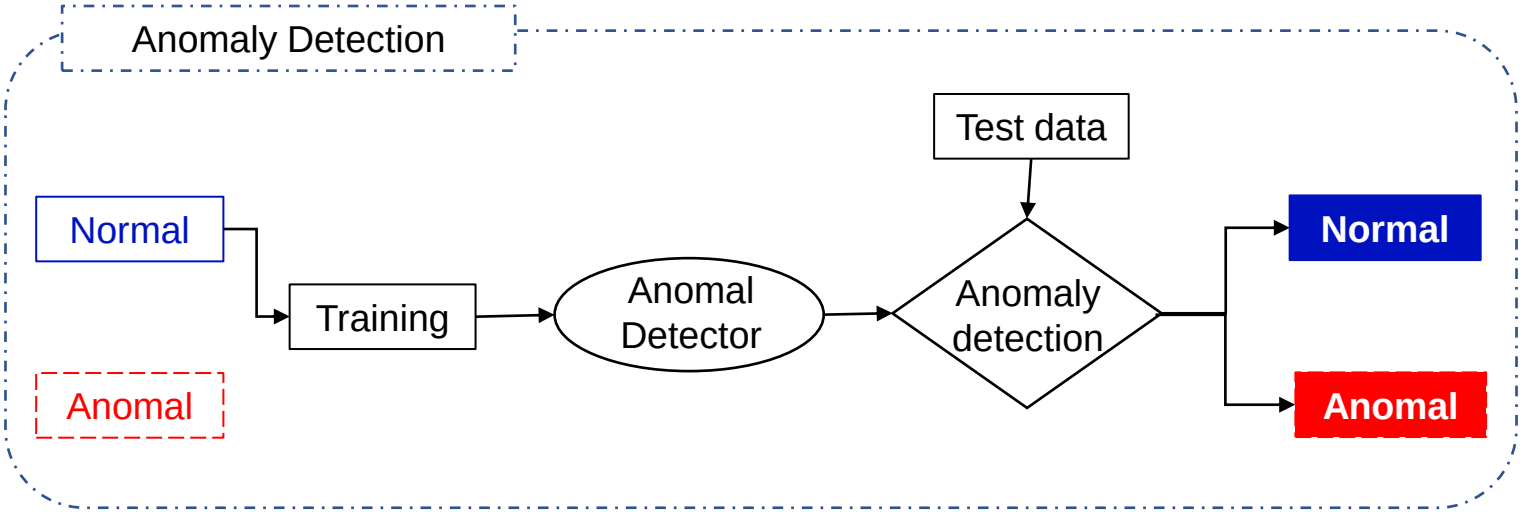
Anomaly Detection

- 공정이 진행되는 동안 정상 데이터와 이상 데이터가 생성됨
  - 센서의 이상 유무 판단 : 전문가들의 경험으로 알아내는 것이 대부분
  - 고도화된 공정 시스템 → **고도화된 이상 탐지 시스템** 필요



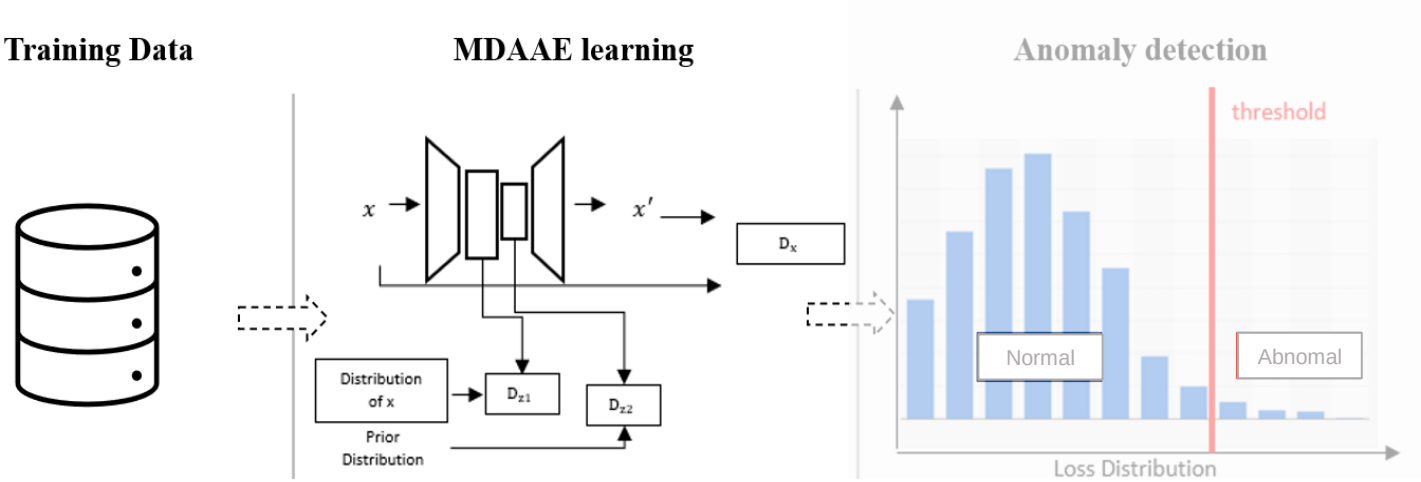
Anomaly Detection

- 이상 데이터
  - 품질 불량 유발, 수율(yield) 저하
  - 숫자가 적어 클래스 불균형 문제 발생 : 분류 성능이 떨어짐
  - **Anomaly detection**을 이용하여 문제를 해결하려고 함



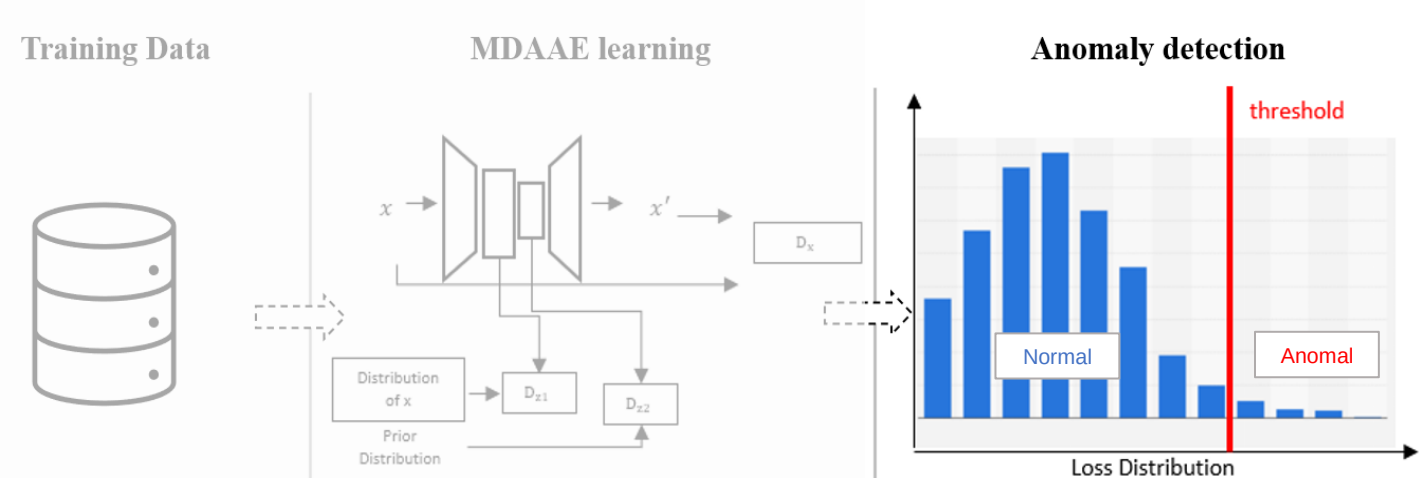
Overview

- AAE with Multiple Discriminators (MDAAE)
  - **AE 구조**(인코더+디코더)와 **3개의 식별자**로 구성
  - **정상 데이터**로만 알고리즘 **학습**
    - 충분하지 않은 이상 데이터 수
    - 이상 데이터의 종류를 정확히 모름



Overview

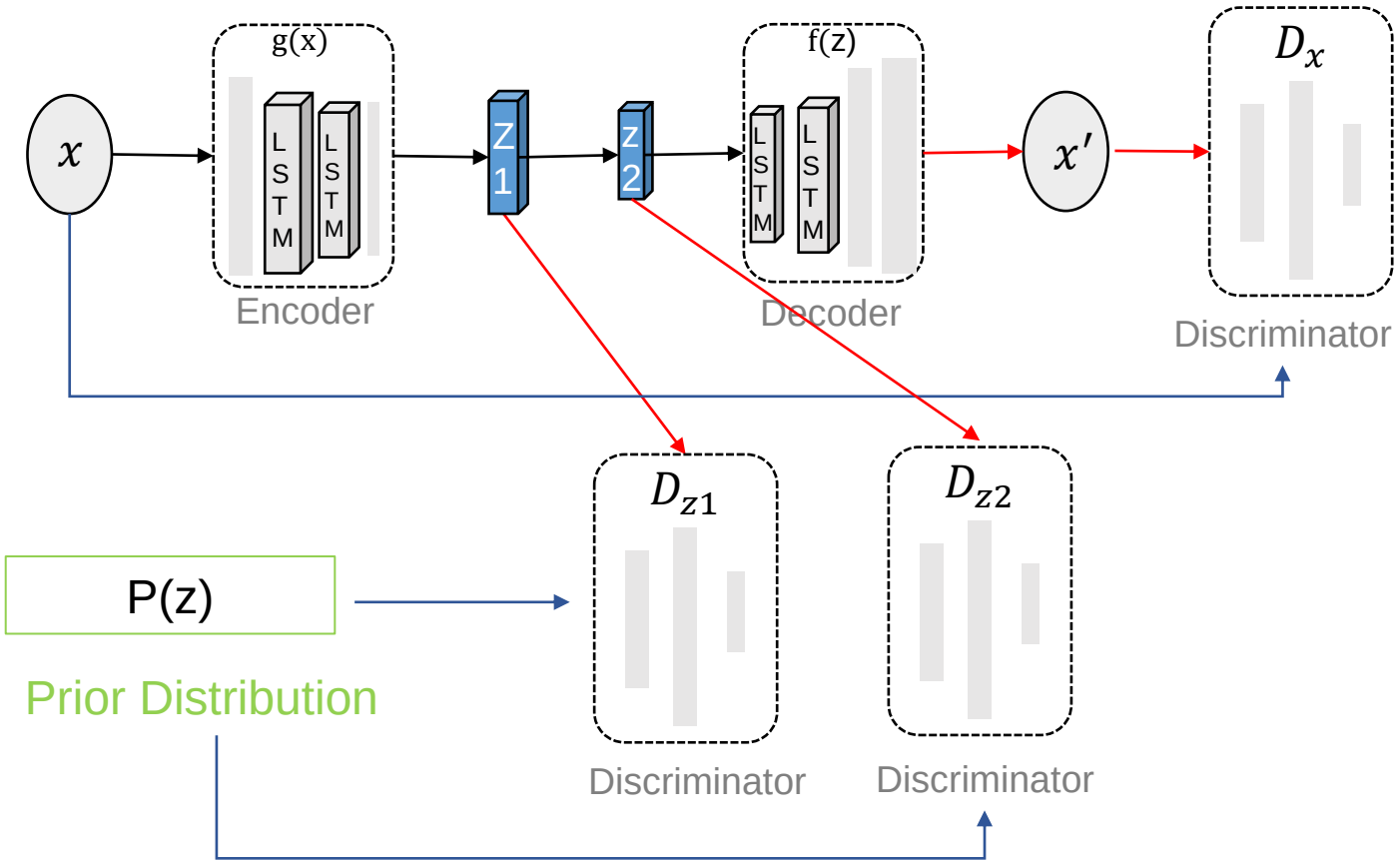
- Anomaly detection
  - **Loss 분포**를 활용하여 anomaly를 탐지하기 위한 **적절한 임곗값 설정**
  - 임곗값을 기준으로 이상 여부 판정
  - **Anomaly** : loss 값이 임곗값보다 높기 때문에 **정상 데이터와 특성이 다름**



Proposed Method

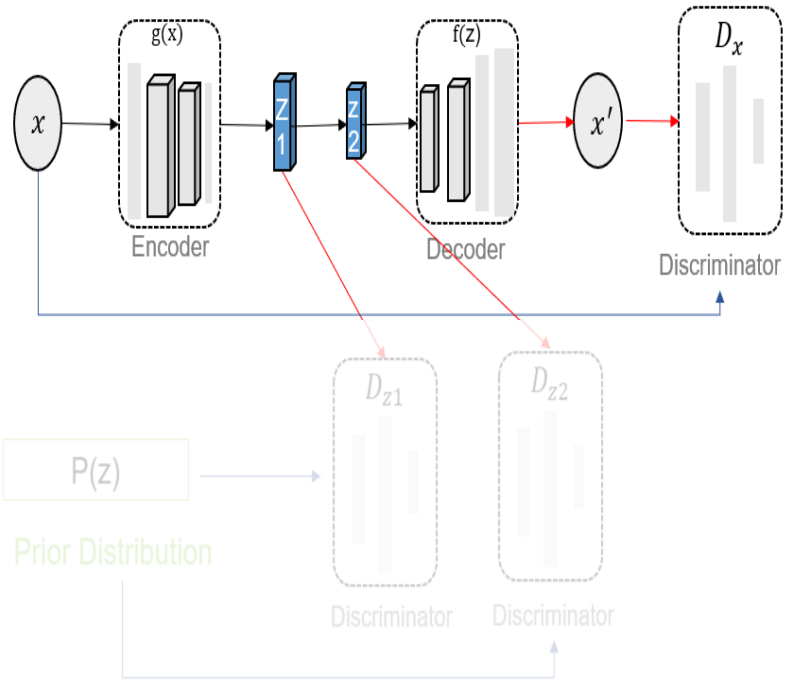
MDAAE

- MDAAE : Adversarial Autoencoder with Multiple Discriminators



MDAAE

- MDAAE : Adversarial Autoencoder with Multiple Discriminators

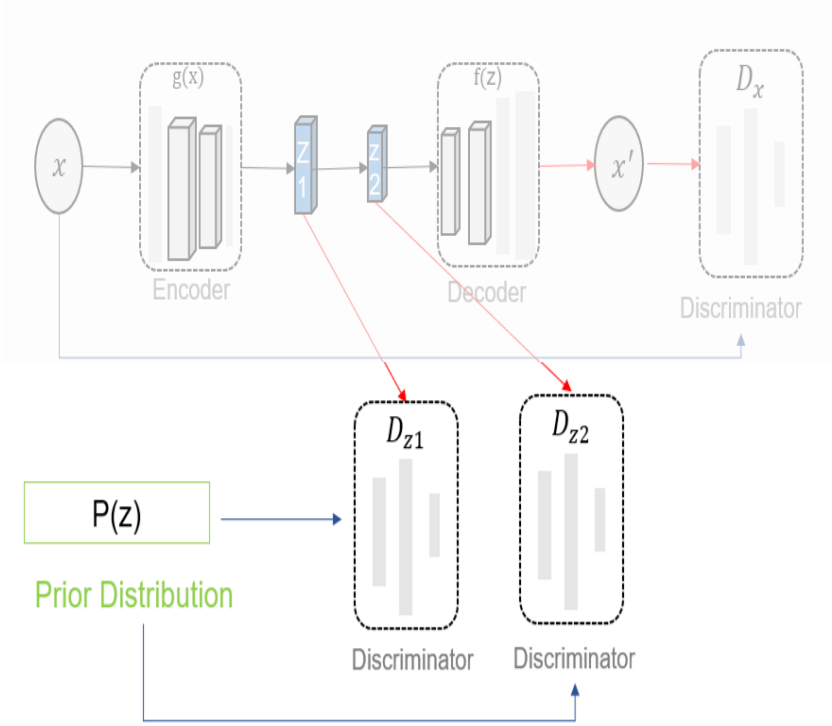


- Encoder
  - 고차원에서 저차원 매핑(mapping)
  - 노이즈 제거 작업
  - 입력 데이터 특질 추출
- Decoder
  - 저차원에서 원 데이터의 차원인 고차원으로 복원
  - 노이즈가 제거된 데이터 복원
- Discriminator_x
  - 경쟁 학습
    →입력 데이터 x와 출력 데이터 x'
  - 입력 데이터와 출력 데이터를 똑같게 복원해내는 역할

Proposed Method

MDAAE

- MDAAE : Adversarial Autoencoder with Multiple Discriminators



- Discriminator_z1
  - 경쟁학습
    → 저차원 $z1$과 사전 분포 $p(z)$
  - 입력 데이터의 분포 $p(z)$와 저차원 $z1$이 같아지도록 학습
- Discriminator_z2
  - 경쟁학습
    → 저차원 $z2$과 사전 분포 $p(z)$
  - 입력 데이터의 분포 $p(z)$와 저차원 $z2$이 같아지도록 학습

MDAAE objective

- 식별자 D_{z1} : $Max\ L_{adv-d_{z1}}(x, g, D_{z1})$
  - Latent space z1 distribution & Prior distribution
  - $E\left[\log\left(D_{z1}(p(z)\right)\right] + E\left[\log\left(1 - D_{z1}\left(g(x)\right)\right)\right]$
- 식별자 D_{z2} : $Max\ L_{adv-d_{z2}}(x, g, D_{z2})$
  - Latent space z2 distribution & Prior distribution
  - $E\left[\log\left(D_{z2}(p(z)\right)\right] + E\left[\log\left(1 - D_{z2}\left(g(x)\right)\right)\right]$
- 식별자 D_x : $Max\ L_{adv-d_x}(x, D_x, f)$
  - Decoded data & Real data
  - $E[\log(D_x(x))] + E[\log(1 - D_x(f(p(z))))]$
- AE : $Min\ L_{error}(x, g, f)$
  - Reconstruction error
  - $-E_z[\log(p\left(f(g(x)|x\right)\right)]$

Experimental design

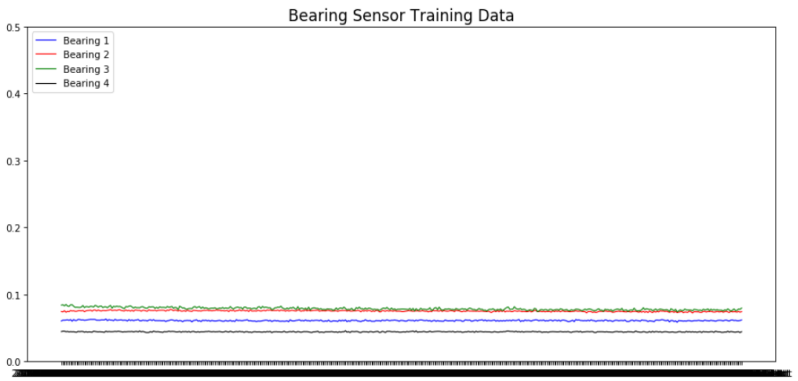
- Hypothesis
  - Loss가 낮을수록 정상 데이터의 특징을 잘 학습한 모델임
  - 정상 데이터의 특징을 잘 학습한 모델은 anomaly detection 성능도 좋을 것임

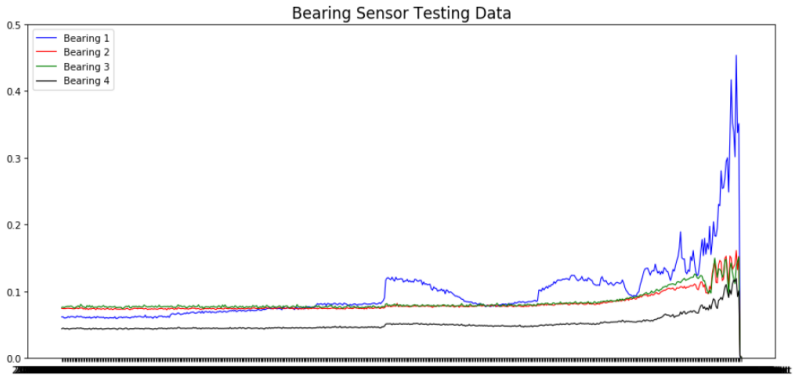
- 1 Step : 모델의 **복원력** 평가
  - Class 정보 없는 데이터 사용 – NASA Bearing dataset
  - Loss of distribution을 이용하여 복원력 측정

- 2 Step : 모델의 **anomaly detection 성능** 평가
  - Class 정보 존재하는 데이터 사용 – Wafer dataset
  - 임곗값을 이용하여 anomaly detection 성능 측정

Step 1 : 모델의 복원력 평가

- Data description
  - NASA Bearing Sensor Data (https://www.kaggle.com/rkuo2000/nasa-bearing-sensor-data)
  - 4개 베어링의 기계적 열화 센서 판독 값
  - 베어링 당 983 길이의 데이터로 구성
    - 983 x 4개 시계열 데이터
    - Train set : 정상 445 x 4개 (45%)
    - Test set : 정상+비정상 : 538 x 4개 (55%)



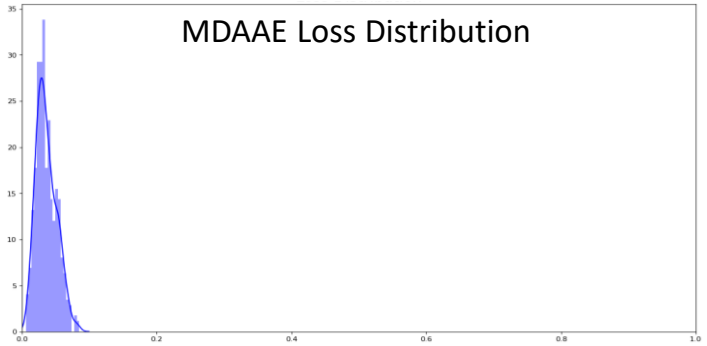


Step 1 : 모델의 복원력 평가

- 성능 평가 및 모델 비교
  - 성능 평가 지표로 **MAE**(Mean absolute error)를 사용
    - MSE는 오차를 제곱하기 때문에 오차가 커지면 커질 수록 손실이 제곱으로 커짐

      → MSE에 비해 MAE는 이상치 영향을 적게 받음
    - **이상치에 강건함**
  - MDAAE와 모델들을 비교
    - MDAAE가 정상 데이터의 특징을 좀 더 정확히 학습할 수 있음

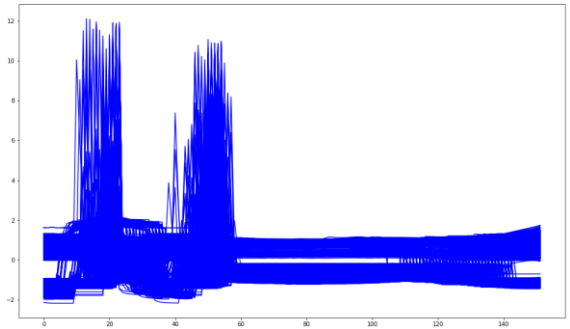
Model	Q2 of Loss
Autoencoder	0.0912
Adversarial Autoencoder	0.0513
GPND	0.0485
MDAAE	**0.0332**



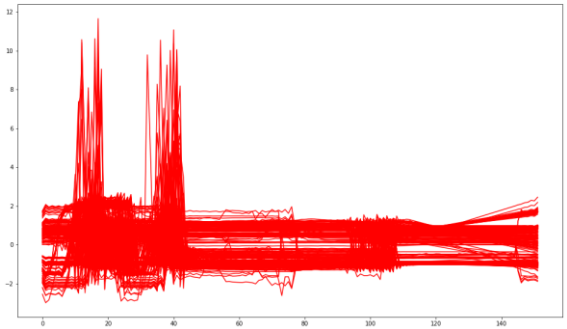
Experiments

Step 2 : 모델의 anmaly detection 성능 평가

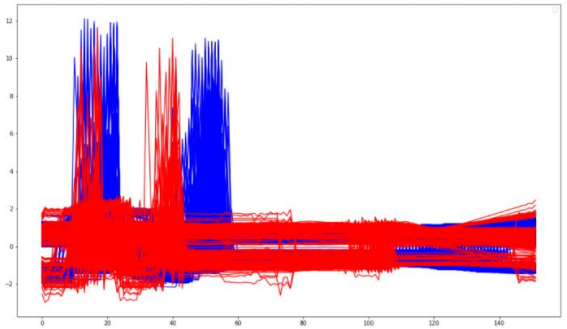
- Data description
  - Wafer Sensor Data (http://timeseriesclassification.com/description.php?Dataset=Wafer)
    - 반도체 제조용 실리콘 웨이퍼 처리 중 다양한 센서에서 기록된 인라인 프로세스 제어 측정 값
    - This dataset was formatted by R. Olszewski as part of his thesis Generalized feature extraction for structural pattern recognition **in time-series data** at Carnegie Mellon University, 2001.
    - One sensor during the processing of one wafer by one tool
  - Large class imbalance : 정상 6,402개(89.4%), 비정상 762개(10.6%)
  - Train set : 정상 5,499개
  - Test set : 1,665개 (정상 : 903개 / 비정상 : 762개)



<Normal of Train set>



<Abnormal of Train set>



<Normal + Abnormal>

Experiments

Step 2 : 모델의 anmaly detection 성능 평가

- 모델 비교
  - **Threshold** : 모델 손실 **분포의 95분위수** 사용
    - 정상 클래스의 데이터를 이상 데이터로 탐지하는 오류를 최소화 하기 위함
    - 임곗값을 넘으면 이상 데이터로 판정
  - 성능 평가 지표로 분류 성능을 사용
  - MDAAE와 모델들을 비교
    - MDAAE의 이상 분류 성능이 제일 우수함
    - 제조 생산 환경에서는 정상 데이터보다 이상 데이터 분류가 매우 중요함

Classification accuracy

	AE	AAE	GPND	MDAAE
Normal	94.4% (853/903)	**96.0% (867/903)**	95.9% (866/903)	95.3% (861/903)
Abnormal	99.2% (756/762)	99.0% (755/792)	98.6% (751/762)	**99.7% (760/762)**
Acc	97.2% (1,609/1,665)	**97.4% (1,622/1,665)**	97.1% (1,617/1,665)	97.3% (1,621/1,665)

Summary and Future study

- Summary
  - MDAAE를 통해 **시계열 데이터의 이상 여부**를 판정하는 방법 제시
  - 엔지니어가 경험적으로 확인했던 이상 판정 방식에 **객관성**을 더함
  - 자동 분류를 통하여 **편의성**을 제공함

- Future study
  - 임곗값 설정에 대한 통계적 검증
  - 입력 데이터 x와 복원 데이터 x'이 똑같아지도록 경쟁 학습시키는 식별자 D_x 의 구조 보완
    - 시간적 특성이 잘 드러나도록 LSTM layer로 변경 등
  - Bearing dataset에 임의로 클래스를 지정하여 분류 성능 측정

참고문헌 및 사사

- 참고문헌

[1] Bonggyun Goh, Jun-Geol Baek, (2019), Anomaly Detection with Variational Autoencoders to Prevent System Malfunctions, *Journal of the Korean Institute of Industrial Engineers* 45(2), 2019.4, 138-145(8 pages)

[2] Pidhorskyi, S., Almohsen, R., and Doretto, G, (2018), Generative probabilistic novelty detection with adversarial autoencoders, *In Advances in neural information processing systems* (pp. 6822-6833).

- Acknowledgement

이 연구는 2019년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원으로 수행된 연구(NRF-2019R1A2C2005949)이며, 2019년도 정부(산업통상자원부)의 재원으로 한국산업기술진흥원의 지원을 받아 수행된 연구입니다(P0008691, 2019년 산업전문인력역량강화사업).

Thank You

Q & A

고려대학교 산업경영공학과
Smart Production Systems Lab.
석사과정 이승희