
경영과학연구실 하계 워크샵

경영과학연구실 김윤석

외조기 공정에서의 상관관계를 고려한 조기 이상 감지 방법에 대한 연구

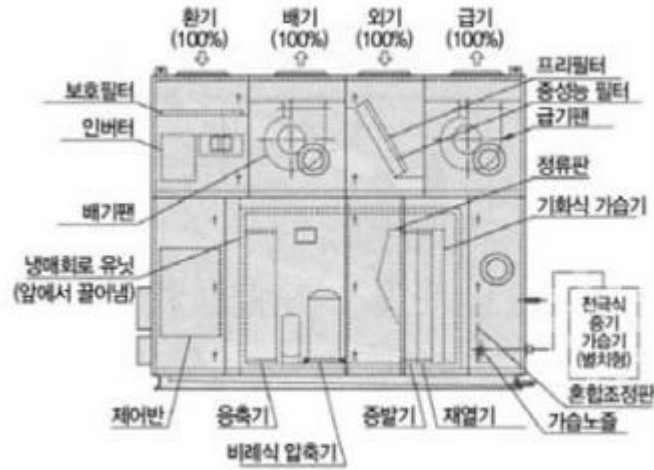
노승찬¹, 정용곤¹, 김주은¹, 백준걸¹
¹고려대학교 산업경영공학과

경영과학연구실 김윤석

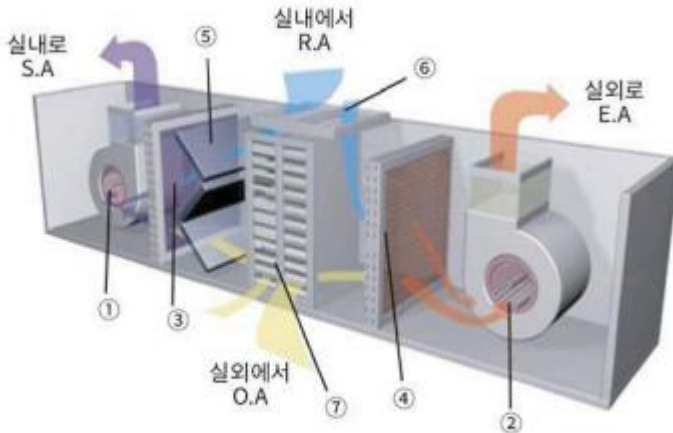
Background

❖ 외조기 공정(OA, Out Air Process)

- 외조기란?
 - 요구되는 온도, 습도 및 이슬점을 원하는 양으로 공급하는 것을 의미
- 외조기 공정: 공장 내에 요구되는 신선한 공기를 외부로부터 공급하는 공정
 - 실내의 공기를 100% 외부로 배출하고, 외부의 공기를 사용



[Figure 1] 외조기 유형 예시 1

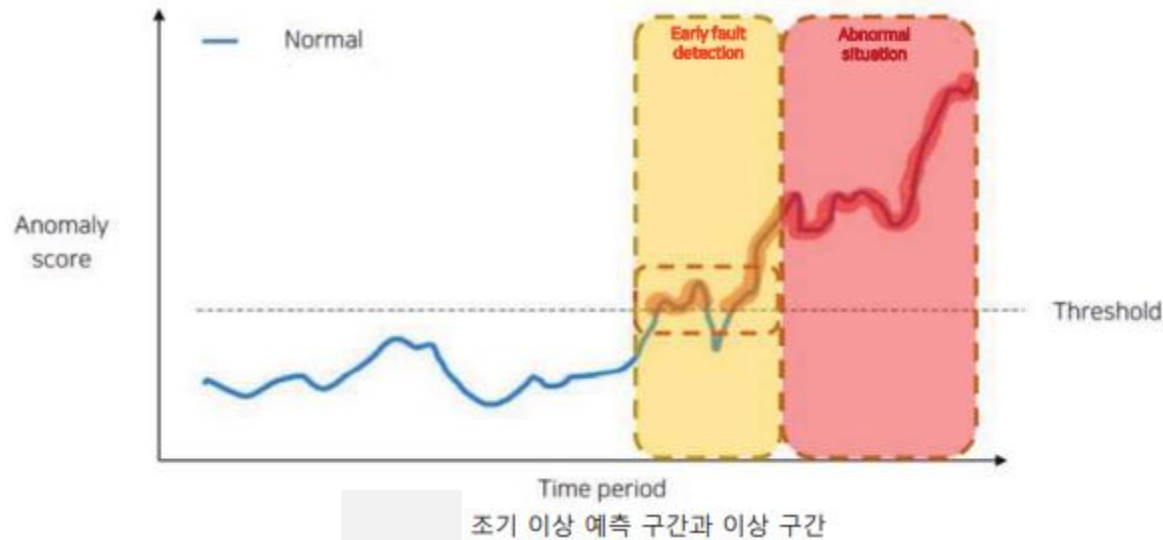


[Figure 2] 외조기 유형 예시 2

Problem statement

❖ Early Fault Detection

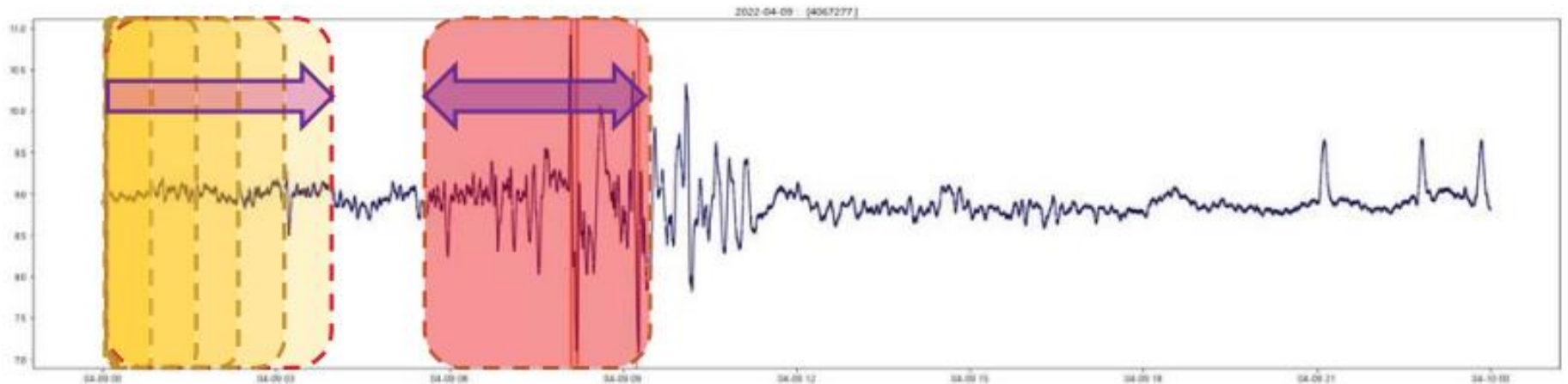
- 이상 상황의 조기 감지 필요
- 조기 감지 시 유지보수 비용 절감, 수명 주기 및 안전성 향상의 이점 존재
- 알람 발생 전 10분 ~ 60분 전 사이에 조기 이상 탐지를 목표로 함



Key idea

❖ Anomaly detection using correlation

- Sliding window를 사용하여 상관계수를 계산
- 초기에는 시간을 누적해 가며 상관계수를 계산
- Window size만큼 누적이 되면 window를 이동해 가며 계산



❖ 데이터 설명

- 외조기의 개별 센서에서 데이터 수집
- 개별 센서에 대한 알람 시점 정보가 존재, 알람 발생 시점을 이상 상황으로 판단
- 외조기 시계열 데이터는 극히 적은 알람 상황이 기록된 불균형 데이터

[Table 1] 센서 데이터 형식

시간	설비 A				설비 B			
	sensor 1	sensor 2	sensor 3	...	sensor 4	sensor 5	sensor 6	...
2022-06-02 00:00:00	18,731 9	44,531 3	21,530 0		19,097 2	35,000 0	9,2477	
2022-06-02 00:00:03	18,645 1	44,534 9	21,560 0		19,560 2	35,000 0	9,2441	
2022-06-02 00:00:06	18,829 6	44,538 5	21,530 0		20,138 9	35,000 0	9,2513	
2022-06-02 00:00:09	19,024 9	44,534 9	21,560 0		19,820 6	35,000 0	9,2730	
2022-06-02 00:00:12	19,140 6	44,531 3	21,530 0		19,791 7	35,000 0	9,2839	
2022-06-02 00:00:15	19,050 2	44,527 6	21,560 0		19,820 6	35,000 0	9,2911	
...	

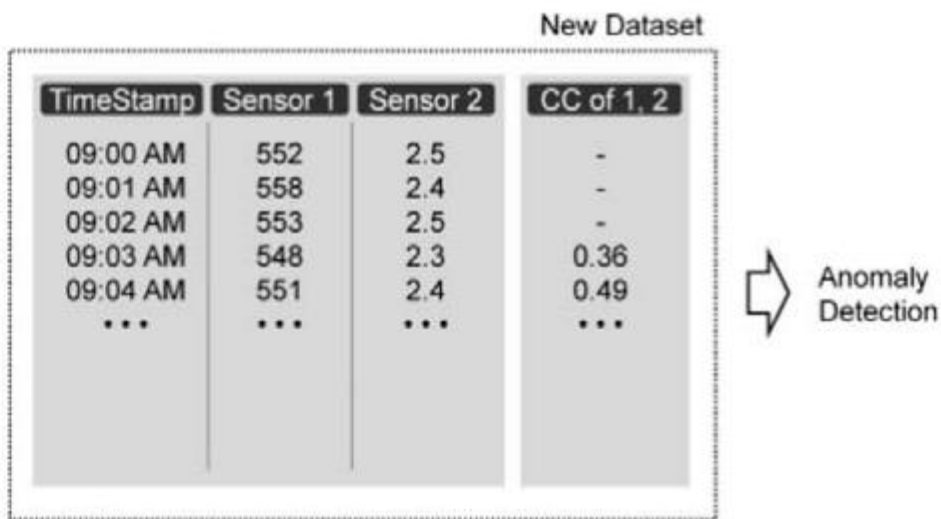
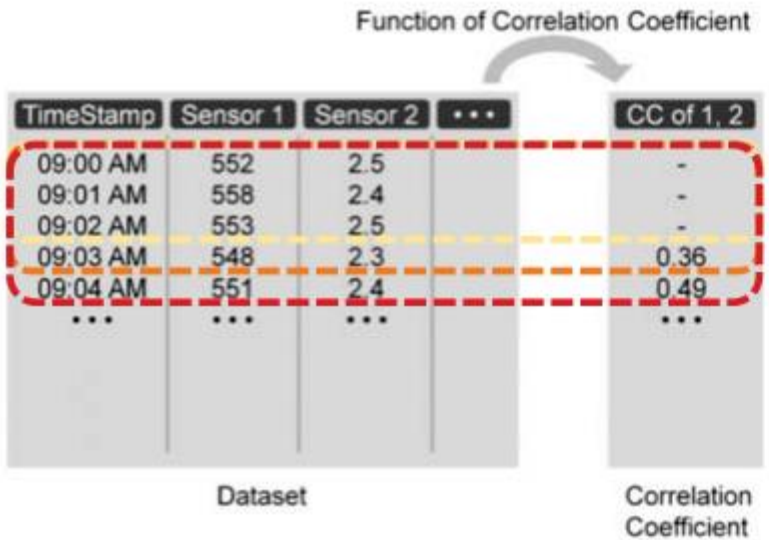
[Table 2] 알람 정보 데이터

센서	알람 시작 시간	알람 종료 시간
센서 13	2022-06-02 05:45:00	2022-06-02 09:02:00
센서 4	2022-06-03 03:56:00	2022-06-03 10:28:00
센서 61	2022-06-03 23:37:00	2022-06-04 00:25:00
센서 20	2022-06-03 23:46:00	2022-06-04 00:07:00
센서 96	2022-06-05 15:53:00	2022-06-05 16:18:00
센서 49	2022-06-09 12:09:00	2022-06-09 14:27:00
...

Add data

❖ Correlation coefficients 변수 추가

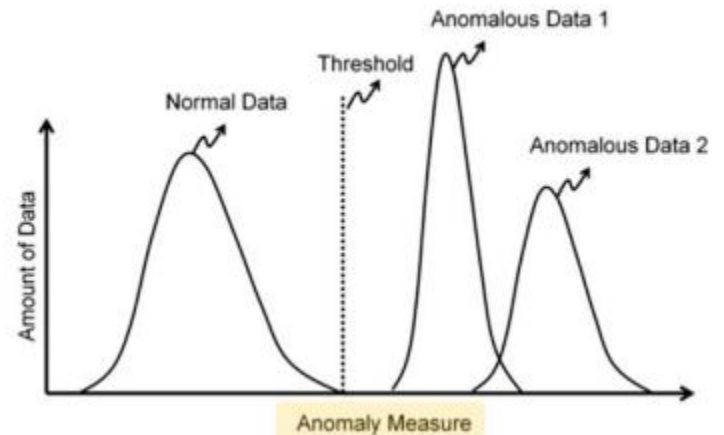
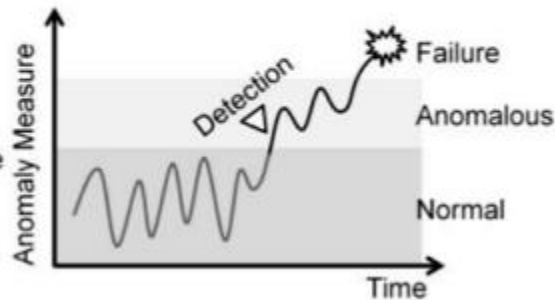
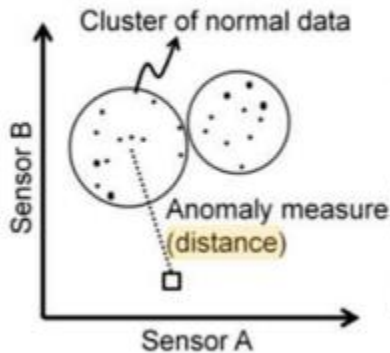
- 정상 상황에서 불규칙한 패턴을 갖고 있는 센서 데이터는 상관계수 변동이 심함
 - 규칙적인 패턴을 가지는 센서 데이터를 사용하여 상관계수 변수 생성
- 센서 데이터가 상관관계 변수보다 이상 상황에 대한 민감도가 클 수 있음
 - 기존의 센서 데이터를 제거하지 않고 상관계수를 dataset에 추가



Anomaly detection using clustering

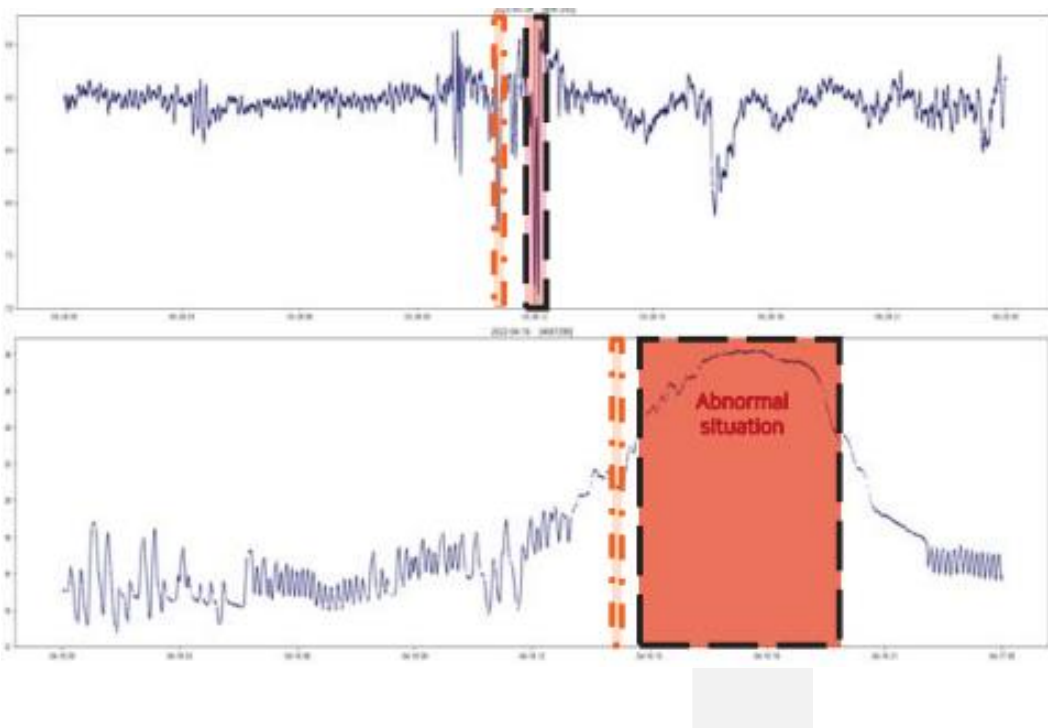
❖ K-means clustering based anomaly detection

- 정상 데이터로 만든 군집의 중심과 관측치의 거리가 이상 탐지의 지표
- 군집의 centroid 및 이상 판단 기준인 임계치 계산
- 군집의 centroid와 군집의 경계까지의 거리를 임계치로 설정



Experiment

- ❖ Test 기간에 대해서 조기 감지 여부 확인
 - 알림이 울리기 3시간 전부터를 test 기간으로 하여 검증
 - 두 설비의 dataset 모두 상관계수를 추가했을 때 보다 정확한 조기 이상 탐지



[Table 4] 상관계수 변수의 유무에 따른 결과 비교

설비	A설비	B설비
상관계수 추가 o	30분 전 알람	20분 전 알람
상관계수 추가 x	2시간 10분, 1시간 25분전, 50분, 15분전, 3분전	2시간 30분전, 1시간 50분전, 1시간 10분전, 43분 전 알람, 8분전

❖ 결론

- 상관관계 기반의 k-means clustering 방법론을 통한 early anomaly detection 제안
- 센서 데이터만 사용하는 경우 부정확하고 빈번한 조기 탐지
- 상관계수를 추가할 경우 보다 정확한 조기 탐지
- 정상 데이터가 규칙적이고 일정한 패턴을 가져야 좋은 성능을 보임

❖ 평가

- 규칙적이고 일정한 패턴을 갖는 시계열 데이터는 모든 이상 탐지기에서 좋은 성능을 기대할 수 있다고 생각
- Time stamp 간의 상관관계를 반영하는 방법이 이상 탐지에 유의미함을 보임

Limitation and Improvement

❖ Limitation

- 제안한 이상 탐지 방법은 복잡한 데이터에서 정상 데이터도 이상으로 판단할 수 있음
- 상관계수 추가로 인해 단순한 시계열 대비 시간 복잡도 증가가 클 것으로 예상

❖ Improvement

- Window size의 chart 또는 이미지에 대해 이상 탐지를 학습 하는 방법
- Window를 보고 이상을 판단한다면 이상 발생 전의 window를 보고 조기 이상 탐지가 가능할 것

오토인코더를 활용한 펌프 데이터의 비지도 이상 탐지

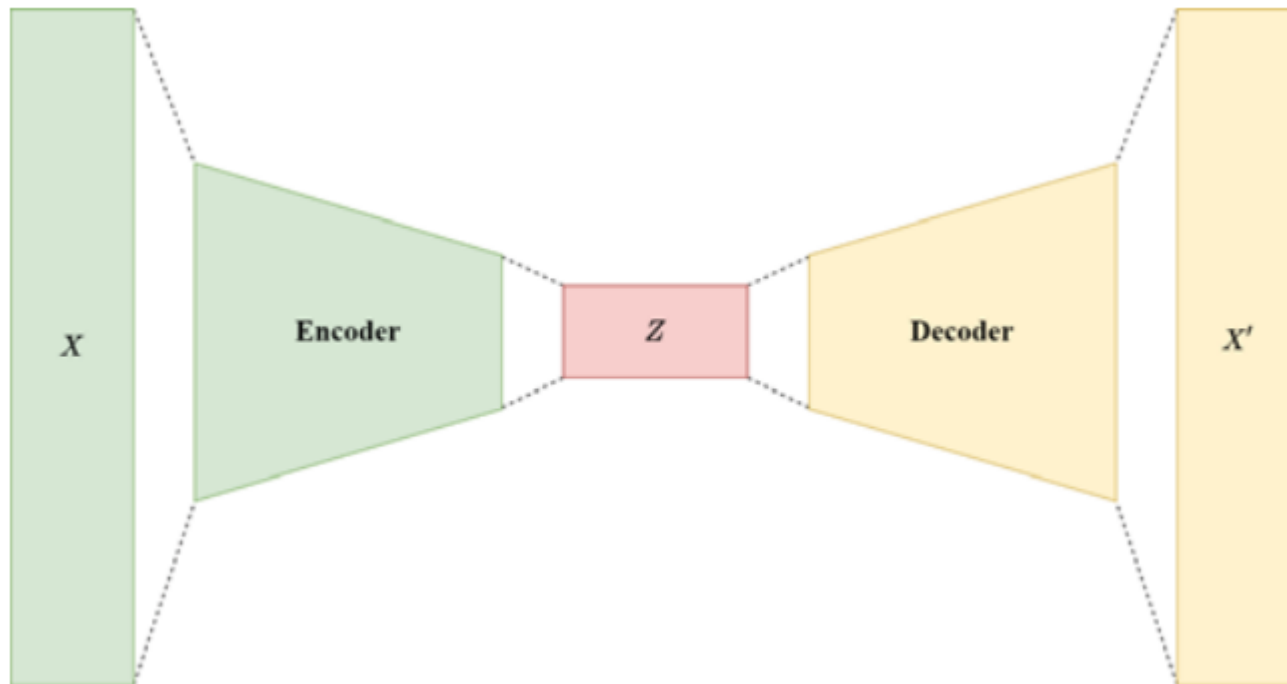
이유진¹, 송승환¹, 황우영¹, 백준걸¹
¹고려대학교 산업경영공학과

경영과학연구실 김윤석

Background

❖ AutoEncoder-based Methods

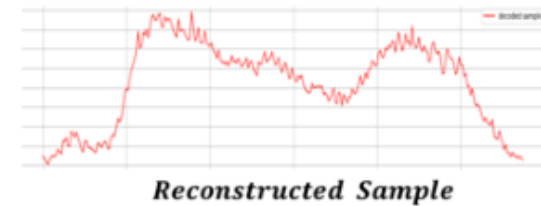
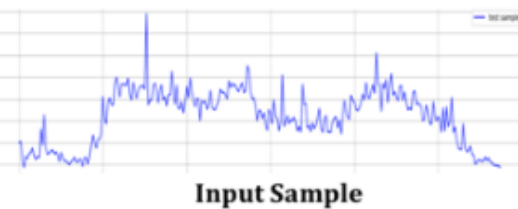
- 정상 데이터를 학습에 사용하여 모델이 정상 패턴만 학습함
- 데이터에 라벨을 필요로 하지 않아 비지도 학습으로 실제 응용에 유리함



Problem statement

❖ Restruction performance

- 실제 시계열 데이터는 복잡한 패턴으로 구성
- 대칭적 AutoEncoder는 재구성 성능을 보장하기 어려움
- 재구성 성능을 향상 시켜 정확한 이상 탐지기가 필요

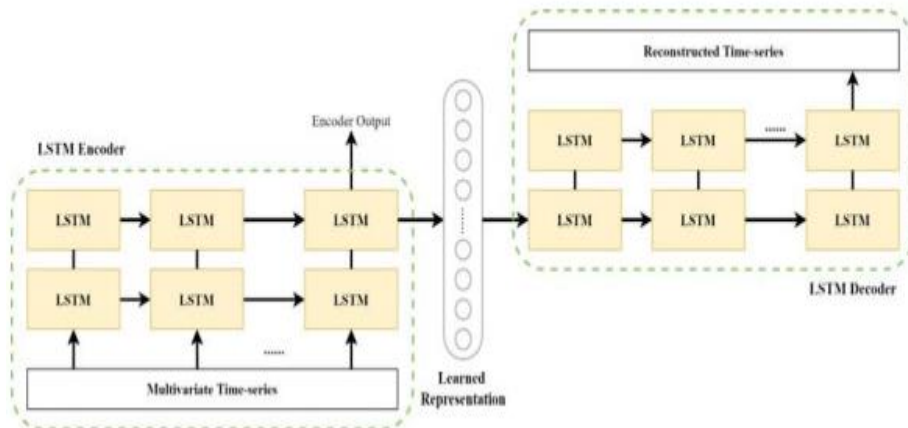


Key idea

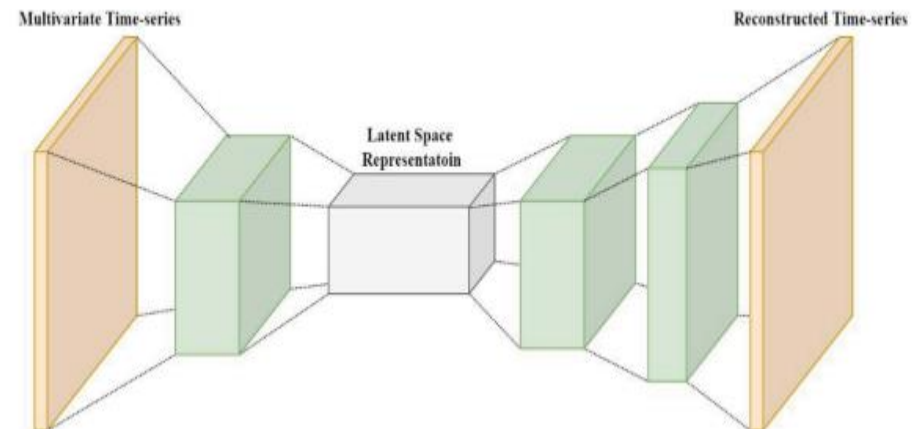
❖ LSTM AutoEncoder & Convolutional AutoEncoder

- LSTM은 RNN의 장기 의존성 문제를 해결하여 시계열 데이터에 많이 사용
- Convolutional 필터는 시계열 데이터를 generic non-linear transformation 함
- 다양한 Convolutional 필터는 시계열의 다양한 discriminative features를 학습 함

LSTM AutoEncoder

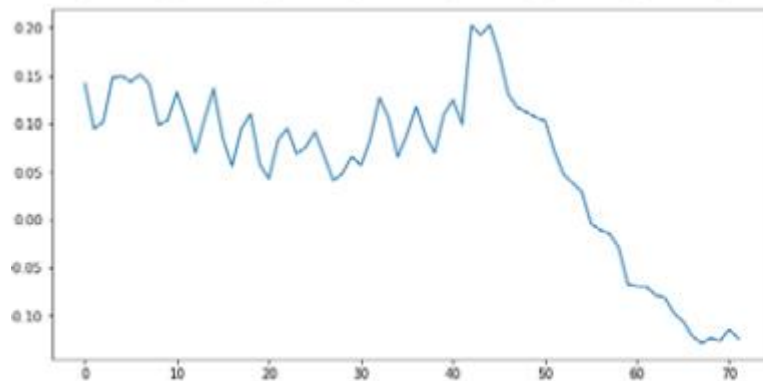
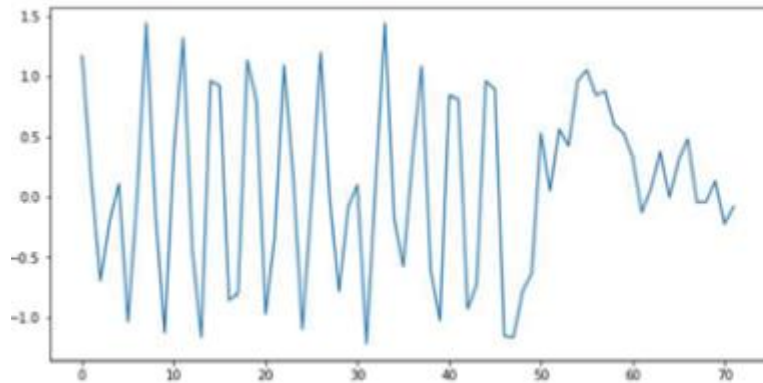


Convolutional AutoEncoder

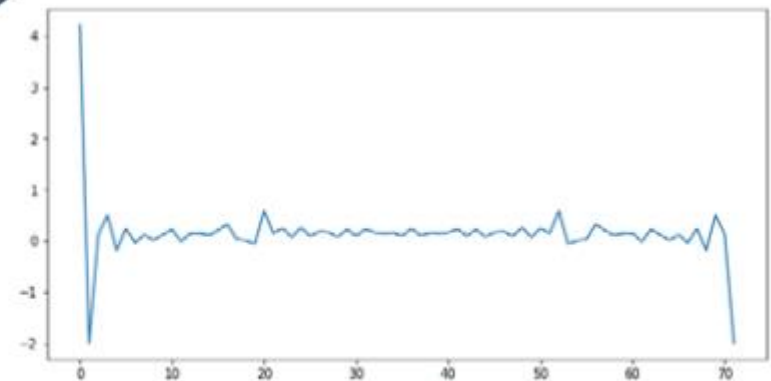
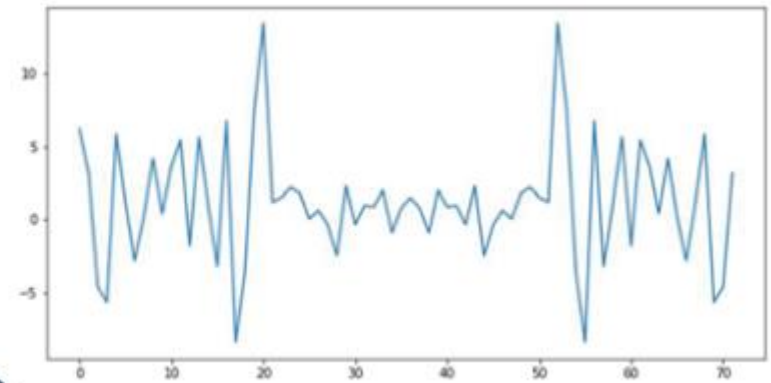


❖ Fast Fourier Transform on Pump Dataset

- Fast Fourier Transform (FFT)은 신호를 개별 스펙트럼 구성 요소로 변환하여 신호에 대한 주파수 정보를 제공함



[FFT 미적용]

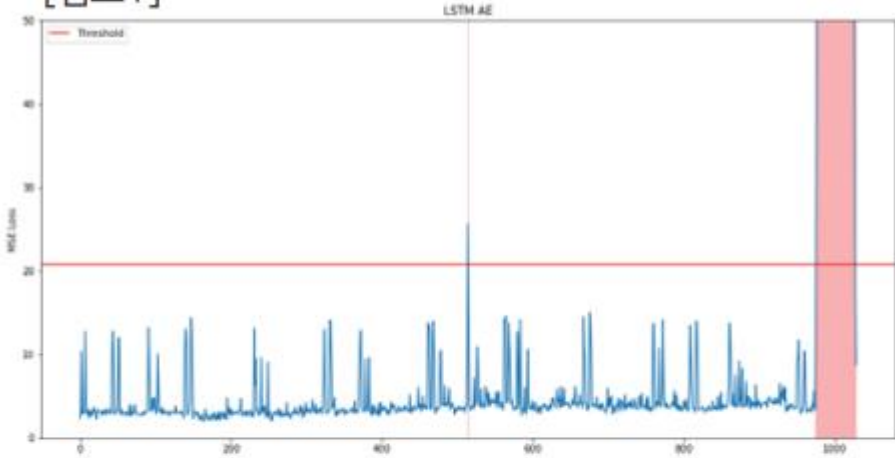


[FFT 적용]

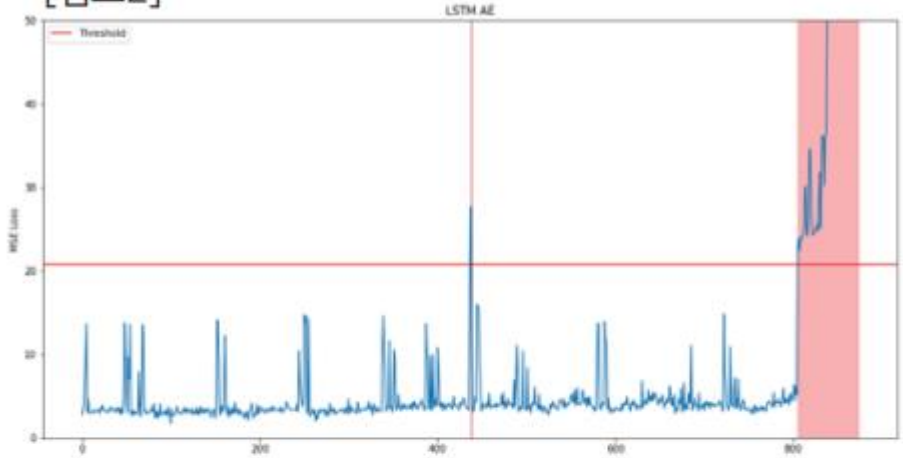
Result of LSTM AutoEncoder

❖ LSTM AutoEncoder

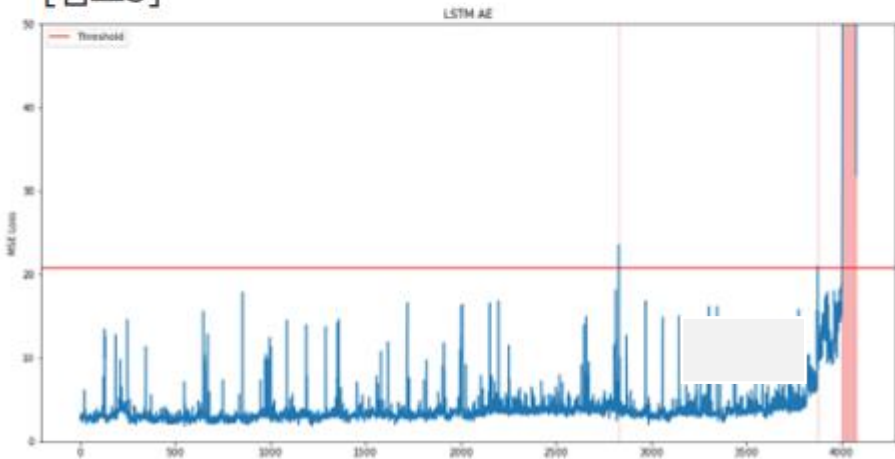
[펌프1]



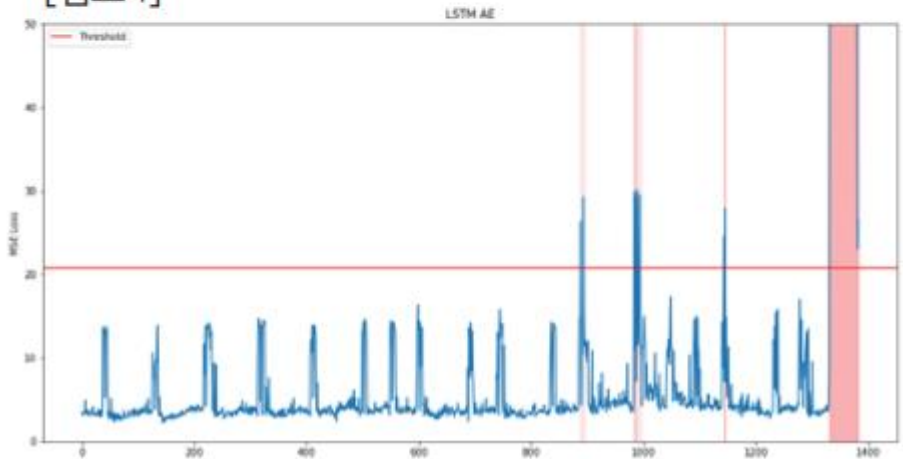
[펌프2]



[펌프3]

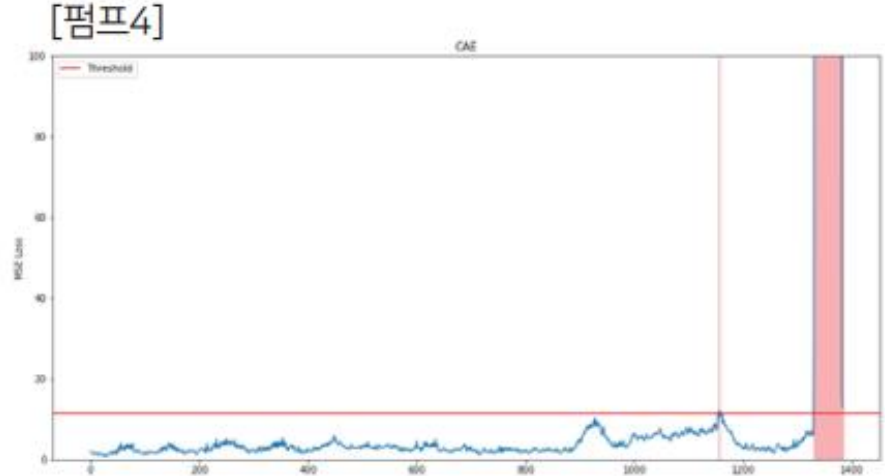
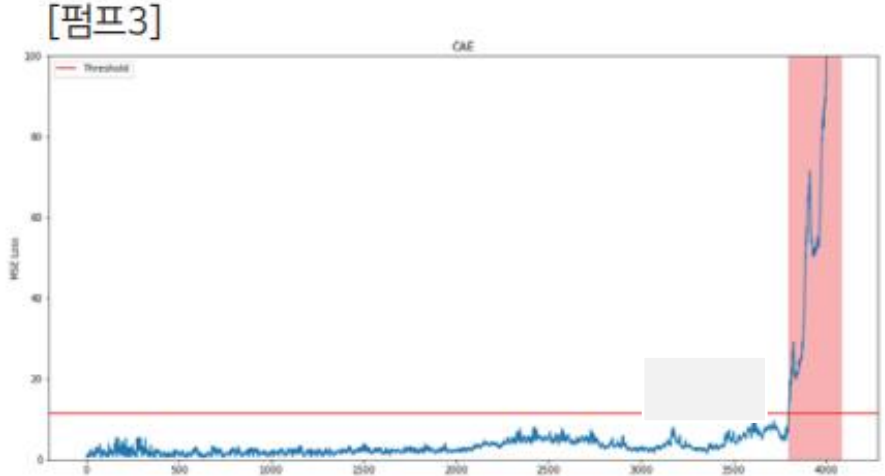
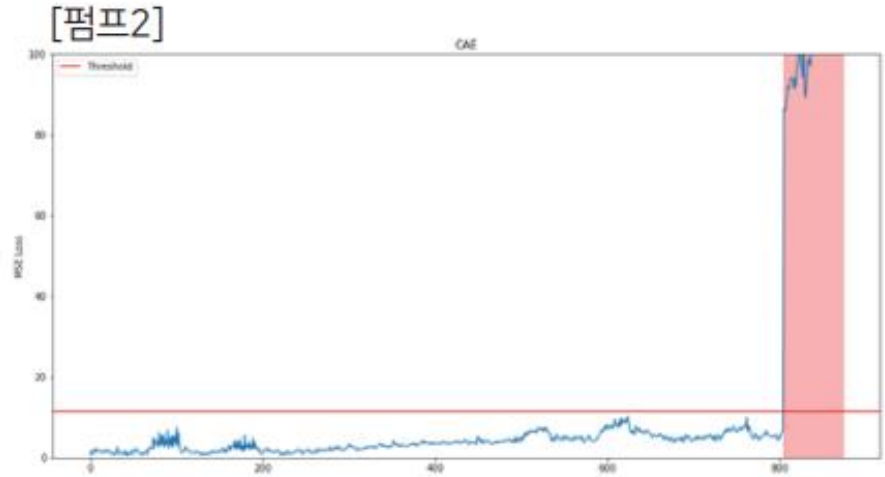
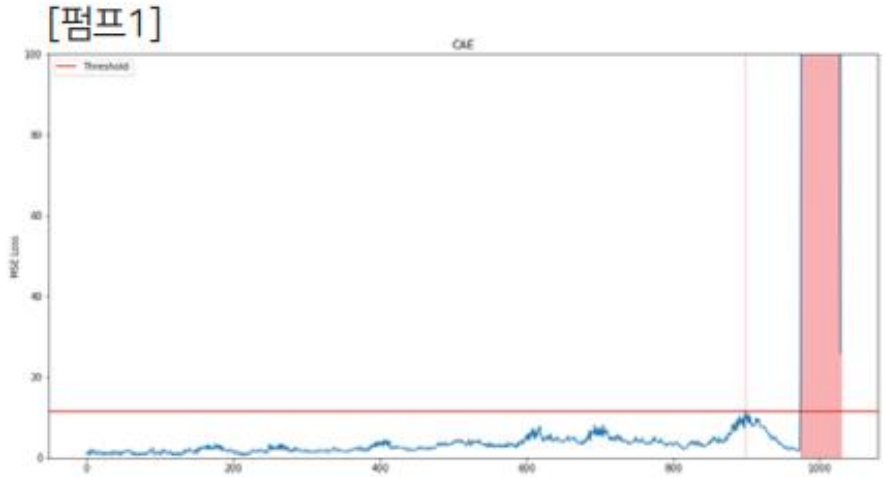


[펌프4]



Result of Convolutional AutoEncoder

❖ Convolutional AutoEncoder

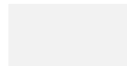


❖ Conclusion

- 실제 데이터에 AutoEncoder 기반의 다양한 이상 탐지 방법론 적용
- Decoder를 깊게 쌓은 비대칭 AutoEncoder를 통해 데이터의 재구성 성능을 높임
- 비대칭 AutoEncoder를 통해 이상 탐지 성능 향상

❖ Evaluation

- 재구성 성능 증가를 확인할 수 있는 실험 결과 부재
- 이상 탐지 성능 향상에 대해서도 확인이 어려움

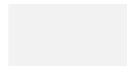


❖ Limitation

- Fast Fourier Transform의 적용할 이유가 없어 보임
- 이상 탐지에 재구성 오차의 필요성을 모르겠음

❖ Improvement

- FFT와 임계치를 이용하여 진동수를 이용한 이상 조기 탐지 방법 제안
- 이상 발생 전 일정 기간의 주파수에 대해 분석



주기성을 갖는 단변량 시계열 데이터를 위한 이상탐지 방법론

윤기오¹, 전여진¹, 이예슬¹, 백준걸¹
¹고려대학교 산업경영공학과

경영과학연구실 김윤석

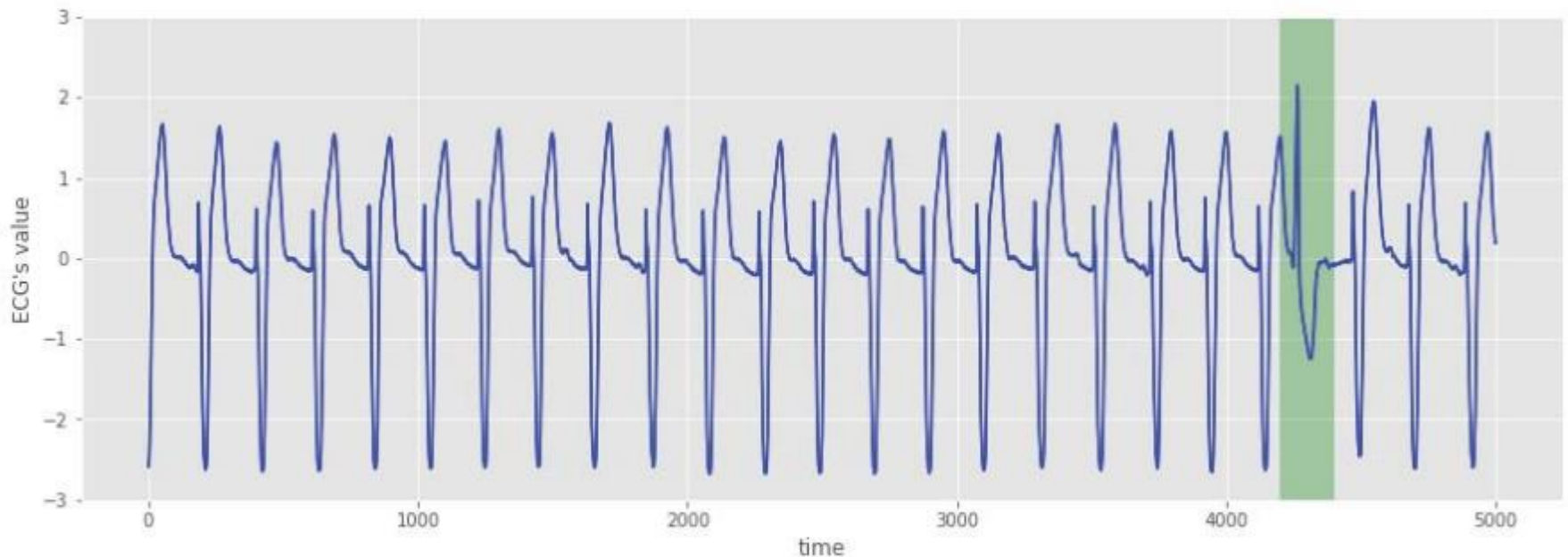
❖ Lack of abnormal data

- 많은 딥러닝을 이용한 이상탐지에서 이상 데이터는 극히 일부
- 균형이 무너진 데이터로 학습 했을 때, 모델의 accuracy가 높더라도 이상 탐지 성능은 기대하기 어려움

Problem statement

❖ 이상탐지기 성능 고도화

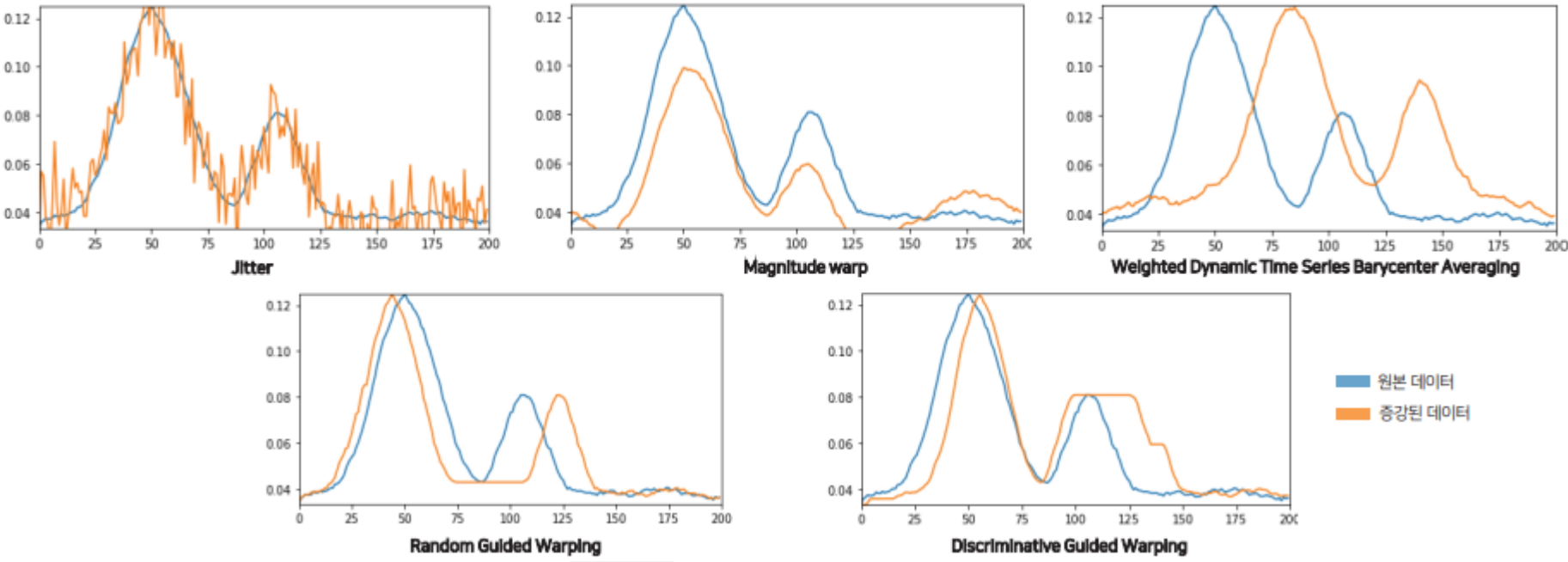
- 시계열 이상 탐지에서 비지도 학습의 성능을 기대하기는 어려움
- 지도 학습을 위해 이상과 정상을 분류하는 과정 필요
- 지도 학습을 하기 위해 데이터 불균형 문제를 해결할 필요가 있음



Data Augmentation

❖ Augmentation

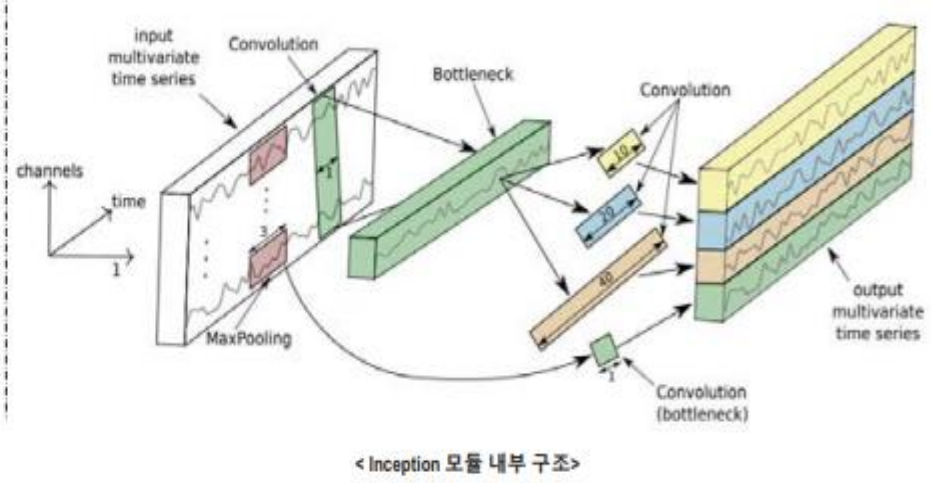
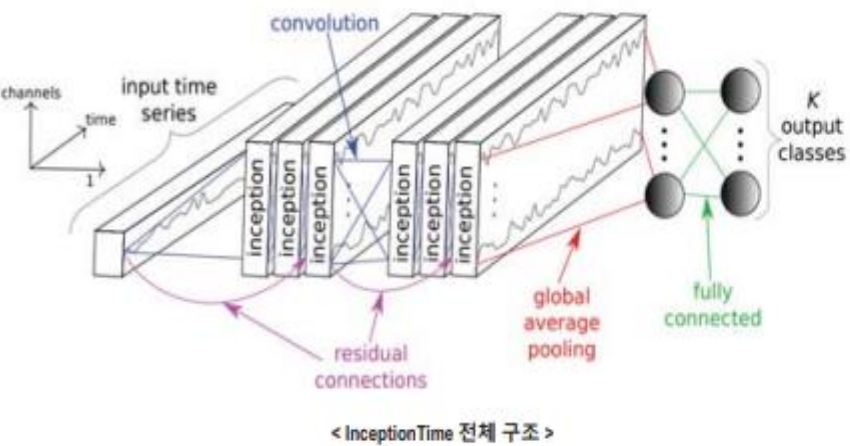
- 데이터 불균형을 해결하기 위해 Augmentation 기법 사용
- Time series data augmentation 기법 적용



Classification

❖ InceptionTime

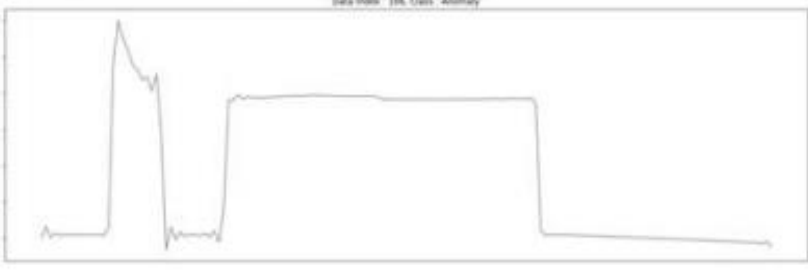
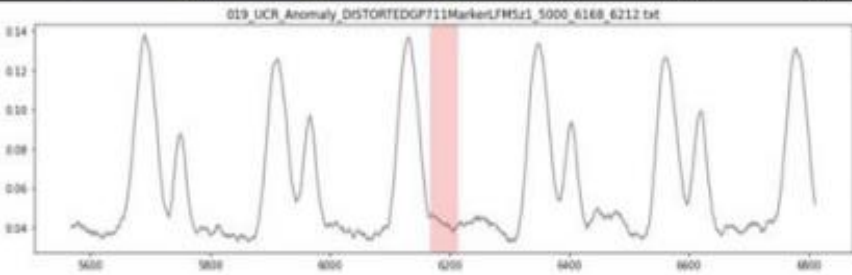
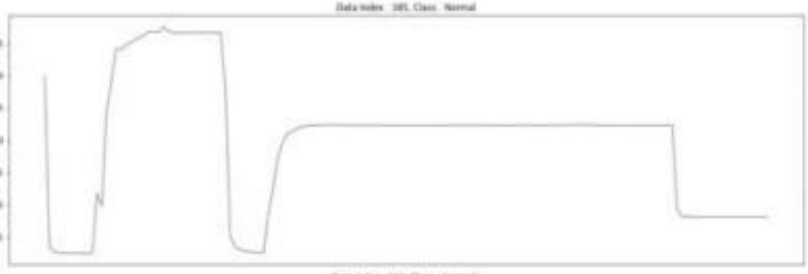
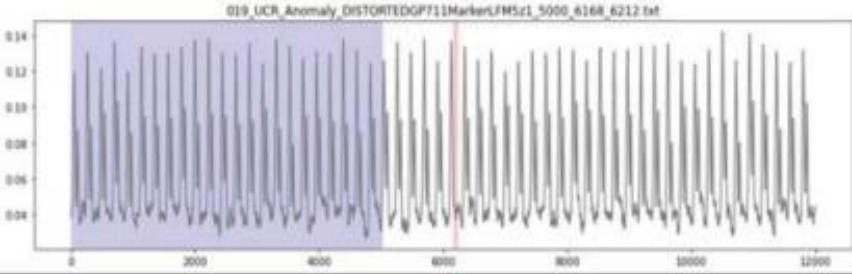
- Inception-v4에 영감을 받아 만들어진 1D-CNN 시계열 분류 알고리즘
- CNN의 앙상블로 시계열 데이터의 local 및 global 패턴 파악
- Inception network 5개의 앙상블로 시계열 분류 문제에서 높은 성능을 보임



Datasets

❖ UCR Time Series Anomaly Detection 2021, Wafer Time Series

- 두 데이터셋 모두 불특정 시점에 비정상 신호가 발생
 - 이는 실제 반도체 공정에서 발생하는 데이터의 형태와 유사
- Wafer time series: 웨이퍼 제작 공정 시계열 데이터로 정상(89%), 이상(11%)
- UCR: 다양한 도메인의 데이터가 포함되어 있으며 2~6개의 데이터가 같은 도메인으로 구성



< UCR Time Series Anomaly Detection Dataset >

< Wafer Time Series Dataset >

Comparison of augmentation results

❖ Dataset: Wafer Time series Dataset

- 제안된 증강 방법은 잘 정제 되어 있는 시계열 데이터에서 좋은 성능일 보임
- 시계열 데이터 형태에 따라 최적 증강 기법이 다른 것으로 확인

	base	jitter	mwarp	wdba	rgw	spawner
F1 score	0.989	<u>0.991</u>	<u>0.992</u>	0.98	0.971	0.985
Accuracy	0.997	<u>0.998</u>	<u>0.998</u>	0.996	0.993	0.997
recall	0.982	<u>0.986</u>	<u>0.989</u>	0.965	0.946	0.979
precision	0.995	<u>0.997</u>	<u>0.995</u>	0.995	<u>0.998</u>	0.991

Performance comparison for training and test data ratios

❖ Dataset: UCR Time series anomaly detection 2021

- 시계열 이상탐지 문제에서 딥러닝 방법을 사용할 경우 다양한 이상 유형에 대한 학습이 부족할 경우 낮은 성능을 보임
- 데이터 증강기법을 사용한 모델에서는 부족한 학습에 대한 성능 감소 폭이 낮아짐을 확인할 수 있음

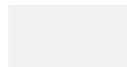
	Base				Base + Augmentation (jitter, mwarped)			
Train : Test	4 : 1	3 : 2	2 : 3	1 : 4	4 : 1	3 : 2	2 : 3	1 : 4
F1 score	0.991(± 0.019)	0.716(± 0.300)	0.638(± 0.333)	0.382(± 0.152)	0.991(± 0.019)	0.891(± 0.143)	0.775(± 0.131)	0.496(± 0.321)
Accuracy	0.997(± 0.007)	0.935(± 0.062)	0.920(± 0.060)	0.870(± 0.023)	0.997(± 0.007)	0.969(± 0.038)	0.938(± 0.033)	0.892(± 0.055)
recall	0.983(± 0.037)	0.631(± 0.350)	0.541(± 0.337)	0.245(± 0.117)	0.983(± 0.037)	0.827(± 0.208)	0.650(± 0.185)	0.381(± 0.302)
precision	1.0(± 0)	1.0(± 0)	0.9(± 0.316)	1.0(± 0)	1.0(± 0)	1.0(± 0)	1.0(± 0)	1.0(± 0)

❖ Conclusion

- 몇 개의 증강 기법에서 성능 향상
- 증강 기법 적용 후 train data 비율이 낮아 지더라도 성능 감소의 폭이 줄어들음

❖ Evaluation

- 증강 기법 이후 성능 증가에 의미를 부여하기 어려움
- Train-test data 비율에 대한 비교도 명확하지 않으며, 유의미하다고 보기 어려움



시계열 데이터의 예측 정확도 향상을 위한 상관 순환유닛

심성현¹, 김도희², 배혜림³

¹동의대학교 산업경영 빅데이터 공학과 조교수

²부산대학교 산업공학과 산업데이터공학융합전공 박사과정

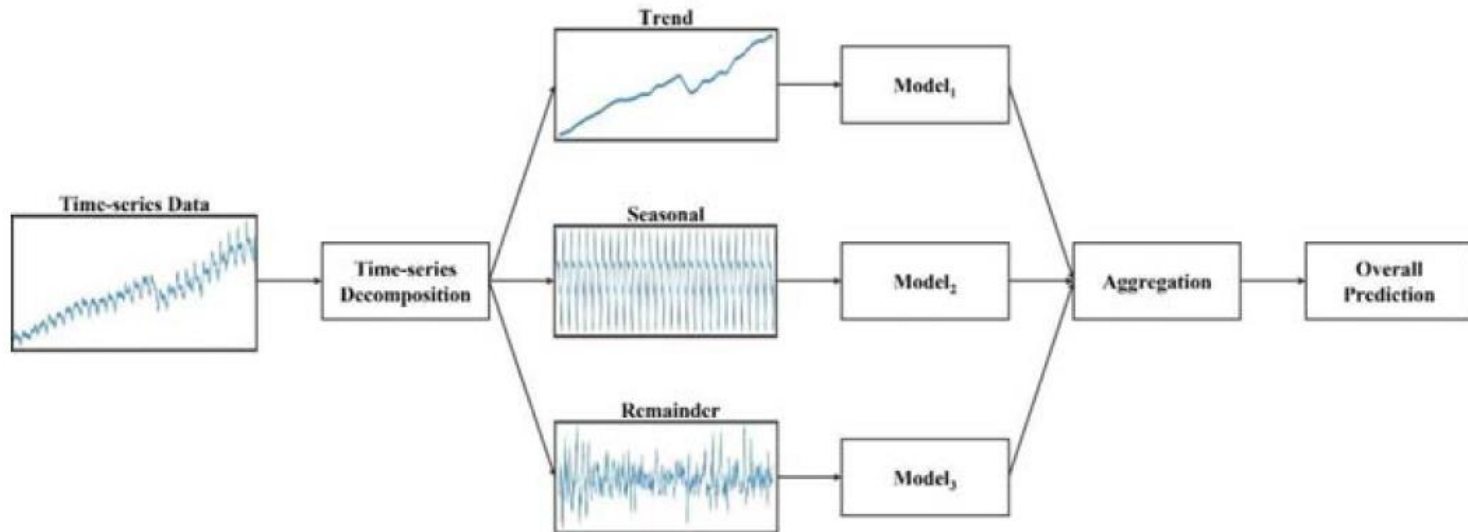
³부산대학교 산업공학과 산업데이터공학융합전공 교수

경영과학연구실 김윤석

Background

❖ 딥러닝 모델과 시계열 분석 기법이 결합

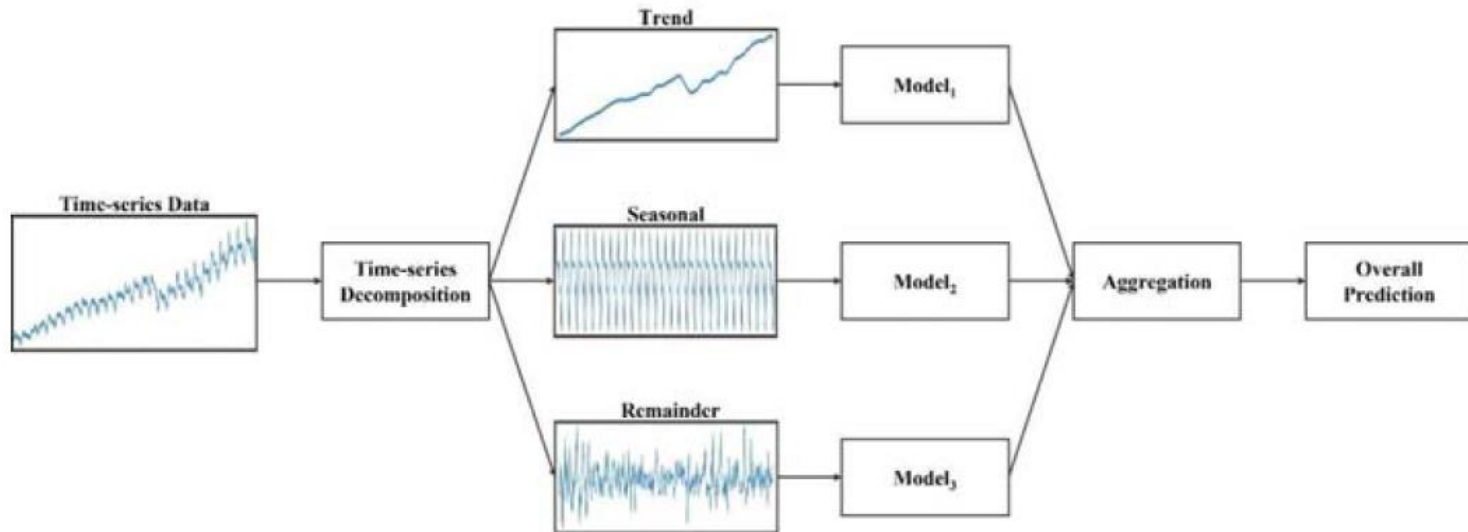
- 딥러닝 모델이 발전함에 따라 머신러닝 모델에 비해서 딥러닝 모델이 좋은 성능을 보임
- 최근 다양한 연구들을 통해서 그림과 같이 시계열 분석 기법과 딥러닝 모델을 결합하는 방법이 제안되고 있음



Problem statement

❖ STL decomposition과 Deep learning 결합

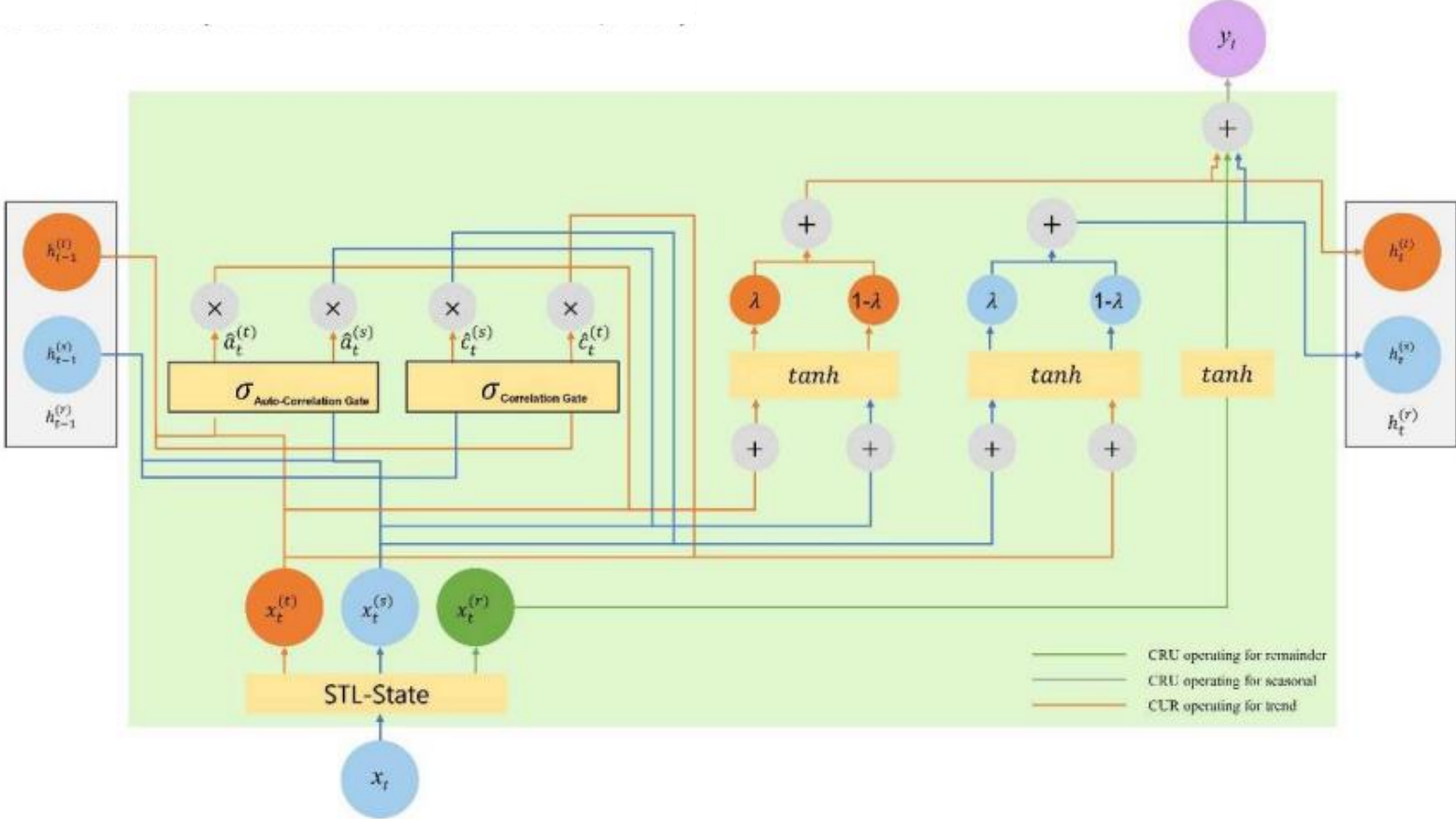
- STL decomposition은 시계열 분석, 분해에서 많이 사용되는 방법
- Trend와 Seasonality가 서로 간에 영향을 줄 수 있는 요인임에도 독립적인 데이터로 학습하는 문제가 있음



Key idea

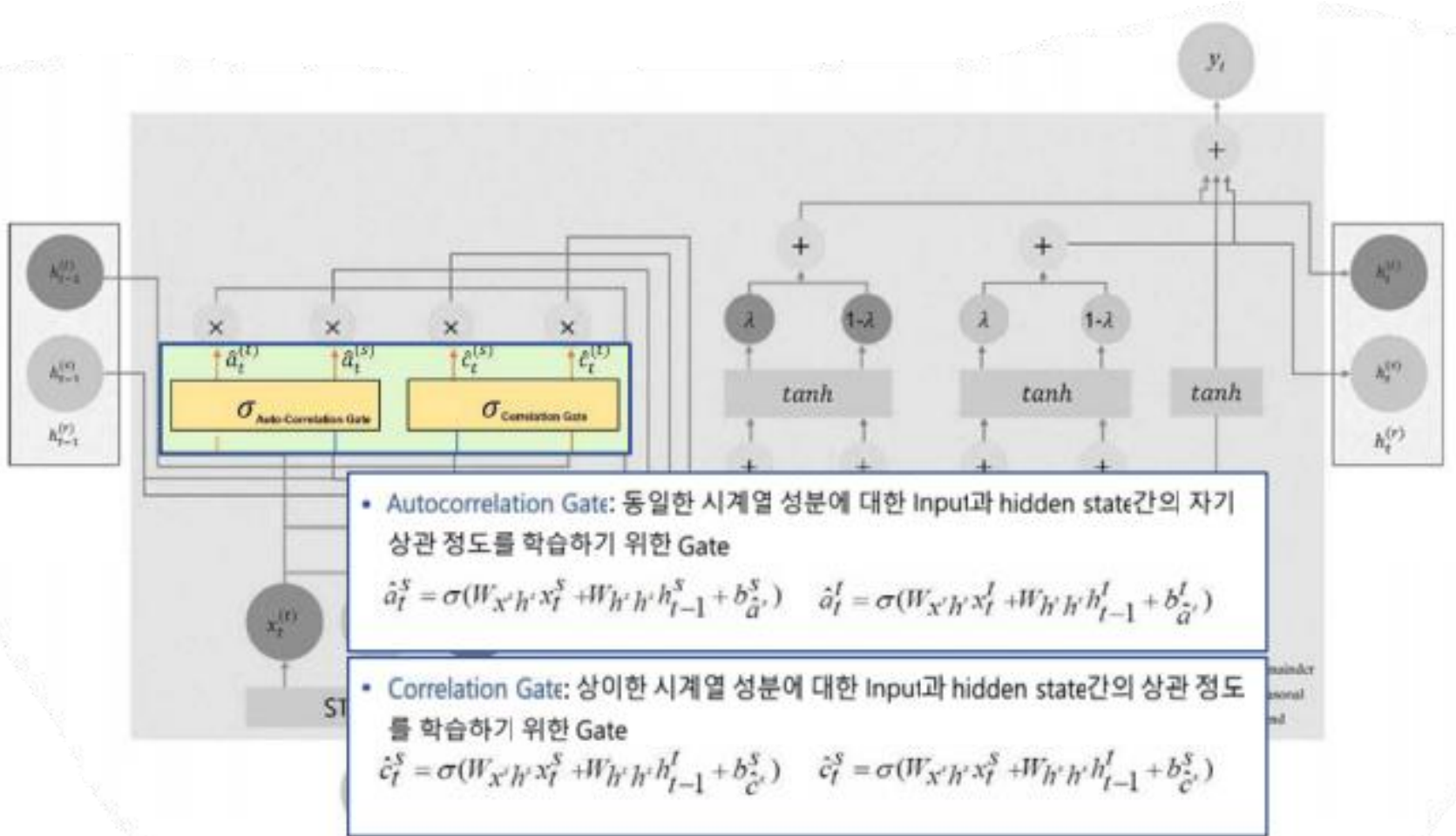
❖ Correlation Recurrent Units, CRU)

- 시계열 분해 요소 간에서 관련성(Autocorrelation과 Correlation)을 학습하기 위해 새로운 Gate를 제안



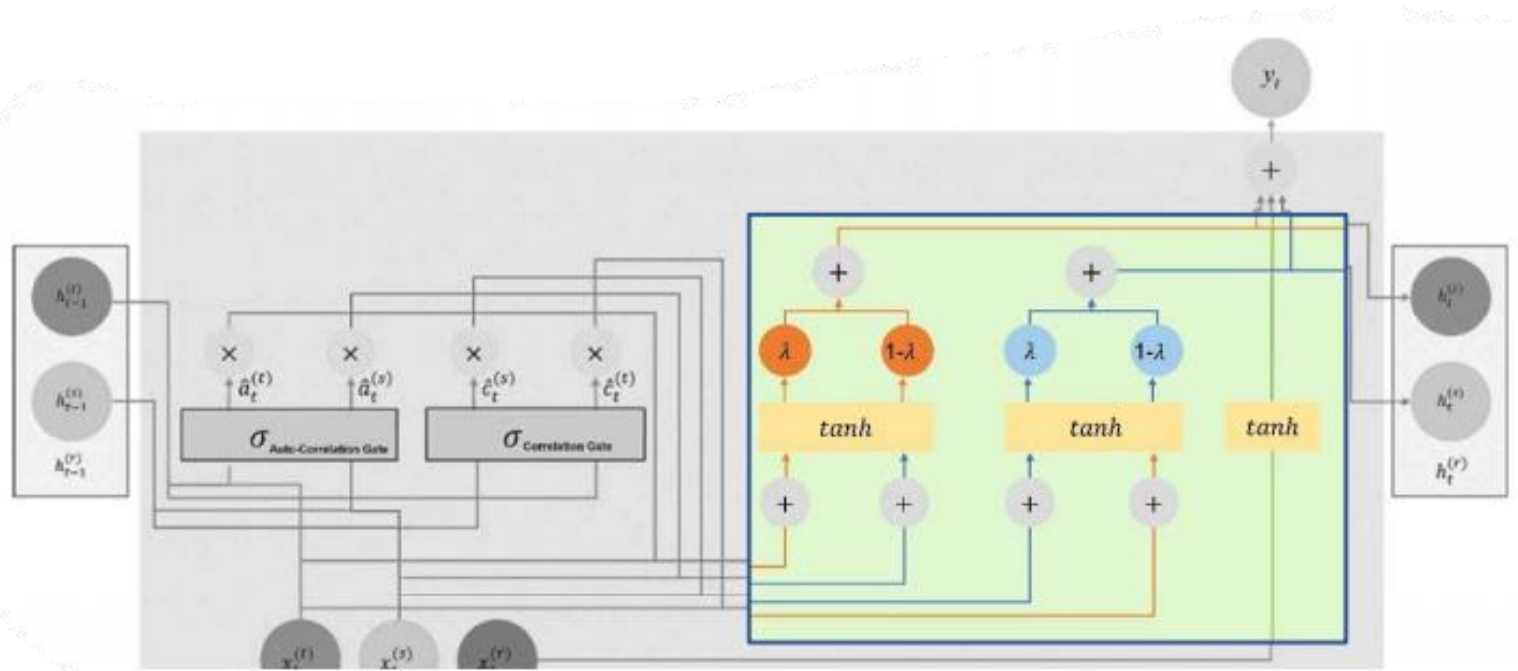
Autocorrelation, Correlation Gate

❖ Gate structure



Hidden State update

❖ State update method



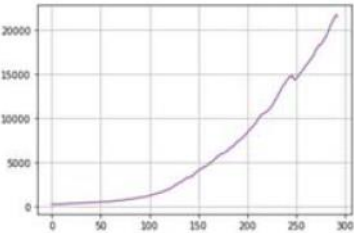
$$h_t^S = \lambda \times \tanh(W_{x'h'} x_t^S + W_{h'h'} (\hat{a}_t^S \odot h_{t-1}^S) + b_{h'}^S) + (1 - \lambda) \times \tanh(W_{x'h'} x_t^S + W_{h'h'} (\hat{c}_t^S \odot h_{t-1}^S) + b_{h'}^S)$$

$$h_t^I = \lambda \times \tanh(W_{x'h'} x_t^I + W_{h'h'} (\hat{a}_t^I \odot h_{t-1}^I) + b_{h'}^I) + (1 - \lambda) \times \tanh(W_{x'h'} x_t^I + W_{h'h'} (\hat{c}_t^I \odot h_{t-1}^I) + b_{h'}^I)$$

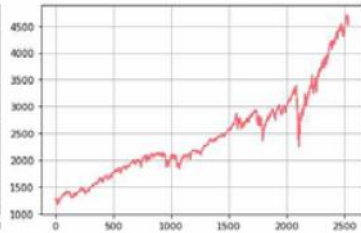
$$h_t^R = \tanh(W_{x'h'} x_t^R + b_{h'}^R)$$

❖ 데이터 종류

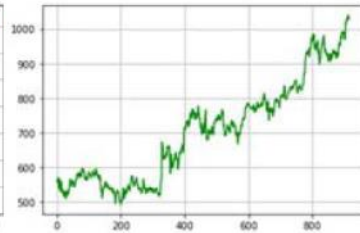
데이터 종류	표기	데이터 수	변수 수	기간	단위	데이터 설명
단변량 시계열 데이터	UTD1	294	1	1947.01.01~ 2020.01.01	분기	세계은행에서 제공하는 미국의 국내총생산 데이터
	UTD2	2268	1	2011.11.03~ 2021.12.02	일	S&P500지수 데이터로써, 500개의 대형기업의 주식을 포함하는 지수
	UTD3	917	1	2014.03.27~ 2017.11.10	일	구글의 주식 데이터
	UTD4	4567	1	1999.11.01~ 2018.02.09	일	발틱해운거래소에서 제공하는 건화물 운임지수 중 BDI 지수 데이터
	UTD5	3298	1	2008.07.04~ 2021.10.20	일	EU 트랜잭션 로그에서 제공하는 EU-ETS(EU Emissions Trading System) 가격 데이터



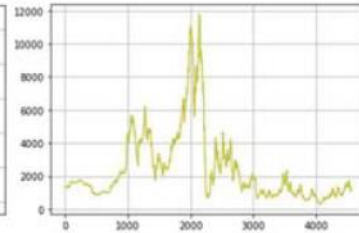
UTD1



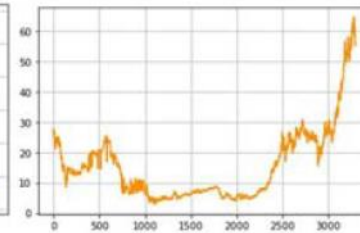
UTD2



UTD3



UTD4



UTD5

Experiment

❖ 예측 성능 결과 평가

모델	UTD1		UTD2		UTD3		UTD4		UTD5	
	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
CRU*	104.76 (1.904)	0.491 (0.012)	9.889 (0.010)	0.456 (0.001)	7.296 (0.001)	0.615 (0.001)	20.093 (0.030)	1.595 (0.003)	0.676 (0.006)	1.841 (0.022)
RNN-STLD	112.88 (0.118)	0.539 (0.012)	11.337 (0.789)	0.645 (0.031)	8.228 (0.002)	0.706 (0.001)	26.537 (0.006)	1.720 (0.007)	0.796 (0.001)	2.074 (0.006)
LSTM-STLD	115.30 (0.457)	0.552 (0.008)	12.018 (1.494)	0.687 (0.021)	8.230 (0.001)	0.707 (0.001)	26.539 (0.001)	1.716 (0.001)	0.808 (0.001)	2.105 (0.006)
GRU-STLD	113.88 (0.533)	0.531 (0.005)	11.304 (0.652)	0.641 (0.010)	8.230 (0.001)	0.705 (0.001)	26.539 (0.001)	1.716 (0.001)	0.808 (0.001)	2.104 (0.011)
RNN	120.70 (3.183)	0.872 (0.124)	12.283 (0.012)	0.579 (0.001)	9.016 (0.001)	0.734 (0.001)	30.689 (0.016)	2.110 (0.007)	0.863 (0.004)	2.279 (0.021)
LSTM	121.46 (3.476)	0.876 (0.012)	12.333 (0.022)	0.583 (0.002)	9.046 (0.001)	0.736 (0.001)	30.623 (0.008)	1.984 (0.005)	0.868 (0.001)	2.291 (0.001)
GRU	121.43 (3.455)	0.775 (0.101)	12.363 (0.015)	0.584 (0.001)	9.048 (0.001)	0.736 (0.001)	30.570 (0.006)	2.079 (0.007)	0.862 (0.006)	2.272 (0.032)

1-Step후 예측 결과

모델	UTD1		UTD2		UTD3		UTD4		UTD5	
	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
CRU*	307.03 (1.529)	1.607 (0.003)	21.661 (0.024)	1.031 (0.001)	16.776 (0.008)	1.506 (0.001)	93.578 (0.024)	6.745 (0.004)	1.515 (0.009)	4.320 (0.024)
RNN-STLD	344.13 (1.566)	2.056 (0.041)	24.792 (0.514)	1.310 (0.010)	21.390 (0.006)	1.900 (0.001)	113.88 (0.406)	7.974 (0.119)	1.853 (0.020)	5.315 (0.108)
LSTM-STLD	316.45 (3.314)	1.847 (0.006)	26.108 (0.356)	1.421 (0.093)	20.818 (0.005)	1.896 (0.001)	113.94 (0.508)	8.145 (0.212)	1.803 (0.016)	5.026 (0.115)
GRU-STLD	339.51 (6.784)	1.900 (0.004)	25.834 (0.415)	1.402 (0.064)	20.821 (0.003)	1.897 (0.001)	113.90 (0.032)	7.943 (0.022)	1.812 (0.018)	5.172 (0.103)
RNN	366.48 (2.136)	1.892 (0.031)	28.783 (0.019)	1.407 (0.001)	22.616 (0.041)	1.980 (0.004)	118.03 (0.066)	8.155 (0.029)	1.845 (0.005)	5.192 (0.027)
LSTM	358.72 (1.992)	1.798 (0.034)	28.906 (0.244)	1.417 (0.027)	23.017 (0.001)	2.008 (0.001)	118.15 (0.086)	8.172 (0.033)	1.906 (0.003)	5.400 (0.019)
GRU	374.23 (2.052)	1.949 (0.041)	29.551 (0.172)	1.484 (0.016)	22.762 (0.015)	1.986 (0.001)	118.13 (0.059)	8.149 (0.019)	1.860 (0.007)	5.243 (0.022)

6-Step후 예측 결과

모델	UTD1		UTD2		UTD3		UTD4		UTD5	
	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
CRU*	180.89 (0.736)	0.943 (0.002)	14.671 (0.003)	0.693 (0.001)	11.331 (0.001)	1.002 (0.001)	54.862 (0.021)	3.952 (0.002)	0.939 (0.001)	2.657 (0.006)
RNN-STLD	203.11 (2.728)	1.163 (0.038)	17.255 (0.967)	0.925 (0.053)	14.258 (0.003)	1.255 (0.001)	67.539 (0.027)	4.648 (0.069)	1.184 (0.001)	3.292 (0.012)
LSTM-STLD	208.23 (2.252)	1.128 (0.015)	17.646 (1.335)	0.952 (0.075)	14.245 (0.001)	1.255 (0.001)	67.677 (0.001)	4.646 (0.001)	1.183 (0.001)	3.239 (0.005)
GRU-STLD	209.51 (1.837)	1.145 (0.052)	18.528 (1.300)	1.008 (0.073)	14.244 (0.001)	1.255 (0.001)	67.433 (0.020)	4.643 (0.023)	1.184 (0.002)	3.239 (0.004)
RNN	221.55 (1.804)	1.087 (0.034)	20.742 (0.042)	1.012 (0.001)	16.065 (0.001)	1.402 (0.001)	69.428 (0.052)	4.862 (0.068)	1.356 (0.001)	3.636 (0.010)
LSTM	223.41 (1.909)	1.039 (0.033)	20.806 (0.076)	1.014 (0.003)	16.247 (0.001)	1.428 (0.001)	69.400 (0.049)	4.846 (0.033)	1.388 (0.001)	3.738 (0.014)
GRU	223.32 (1.144)	1.049 (0.031)	20.941 (0.039)	1.022 (0.001)	16.217 (0.009)	1.423 (0.001)	69.388 (0.038)	4.843 (0.034)	1.377 (0.002)	3.680 (0.008)

3-Step후 예측 결과

모델	UTD1		UTD2		UTD3		UTD4		UTD5	
	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
CRU*	498.98 (2.067)	2.816 (0.004)	35.493 (0.193)	1.780 (0.014)	27.065 (0.036)	2.455 (0.002)	160.23 (0.056)	11.397 (0.042)	2.375 (0.041)	6.868 (0.030)
RNN-STLD	528.86 (5.199)	3.167 (0.008)	37.584 (0.210)	2.041 (0.026)	29.483 (0.010)	2.659 (0.001)	185.80 (0.243)	13.135 (0.178)	2.568 (0.045)	7.569 (0.067)
LSTM-STLD	532.43 (3.678)	3.001 (0.020)	37.200 (2.019)	2.019 (0.034)	28.887 (0.008)	2.591 (0.001)	184.57 (0.275)	13.223 (0.180)	2.558 (0.044)	7.692 (0.030)
GRU-STLD	538.54 (3.847)	3.067 (0.019)	39.261 (2.185)	2.185 (0.090)	29.089 (0.006)	2.579 (0.001)	185.46 (1.087)	13.159 (0.060)	2.542 (0.025)	7.606 (0.041)
RNN	557.91 (4.167)	2.955 (0.060)	38.659 (0.012)	1.907 (0.004)	30.890 (0.080)	2.676 (0.018)	210.23 (0.058)	14.948 (0.056)	2.750 (0.023)	8.059 (0.159)
LSTM	541.19 (4.061)	2.763 (0.042)	39.419 (0.306)	1.910 (0.016)	30.762 (0.045)	2.712 (0.005)	210.17 (0.303)	15.019 (0.138)	2.824 (0.020)	8.531 (0.040)
GRU	528.48 (4.109)	2.723 (0.066)	44.963 (1.137)	2.383 (0.073)	30.824 (0.029)	2.698 (0.016)	210.65 (0.605)	15.051 (0.125)	2.781 (0.016)	8.398 (0.033)

12-Step후 예측 결과

Conclusion

❖ Conclusion

- CRU 새로운 Cell 구조를 제안
- CRU는 sigmoid function을 이용한 Correlation Gate, Autocorrelation Gate를 구성
- 단기 및 장기 예측에서 제안된 CRU가 가장 좋은 성능을 보여주었으며, RNN/LSTM/ GRU에 비해 최대 20%, Multiple-RNN/LSTM/GRU에 비해 최대 10%이상 개선된 성능을 보여주었음

❖ Evaluation

- 실험 데이터의 trend, seasonality가 관계가 있는 데이터인지 설명이 부족
- Trend, seasonality 상관관계를 학습 후 성능 향상이 있음을 보여줌

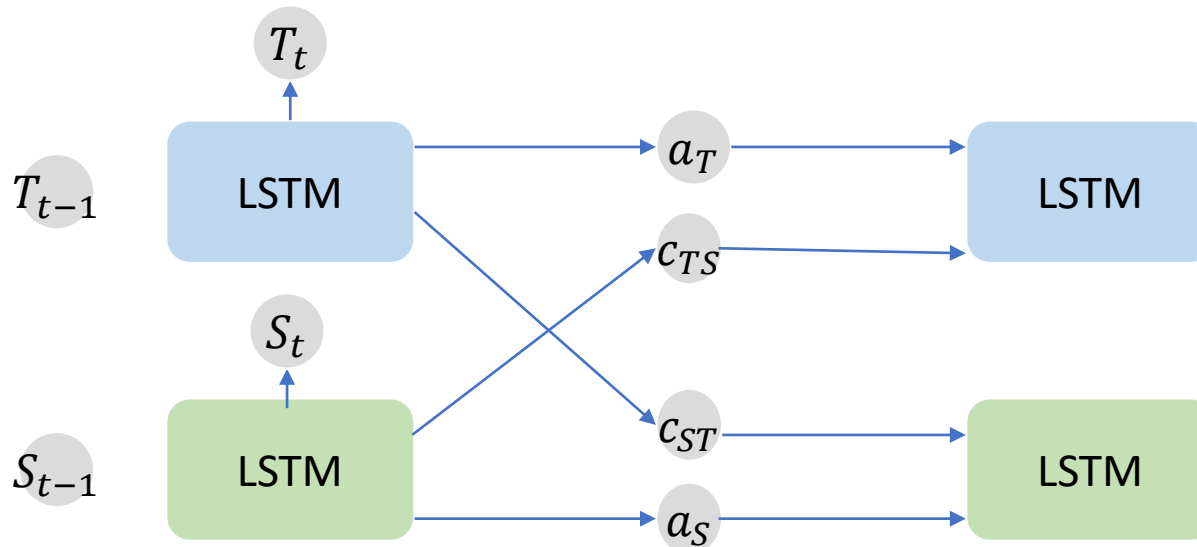
Limitation and Improvement

❖ Limitation

- Gate의 출력이 1 과 0일 때 상관관계에 대한 설명이 부족함

❖ Improvement

- Sigmoid를 추가한 unit이 아닌 상관계수를 입력 받는 RNN (LSTM) 모델



변수 간 어텐션 메커니즘을 활용한 트랜스포머 기반 다변량 시계열 이상치 탐지

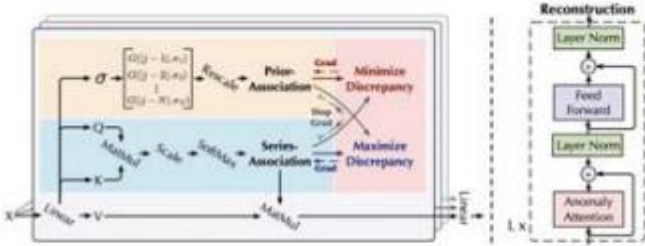
강형원, 정의석, 박경찬, 강필성*
고려대학교 산업경영공학과

경영과학연구실 김윤석

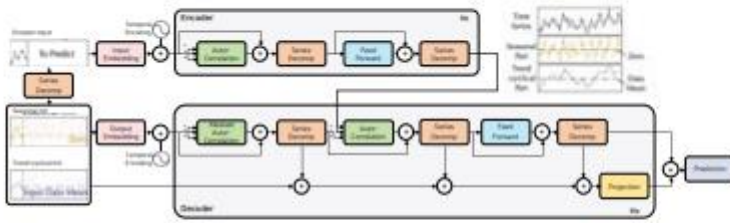
Background

❖ Transformer for Time-series Analysis

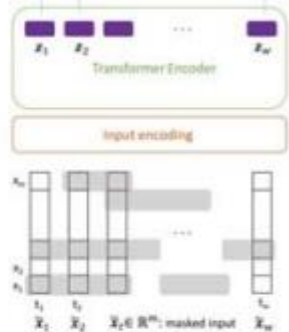
- Sequence를 입력으로 받아 sequence를 출력하는 구조이므로, 시계열에도 적용 가능함
- 기계 번역을 위해 제안된 모델이므로 복잡한 다변량 시계열에 적용하는 연구들이 제안



[Anomaly Transformer]



[Autoformer]

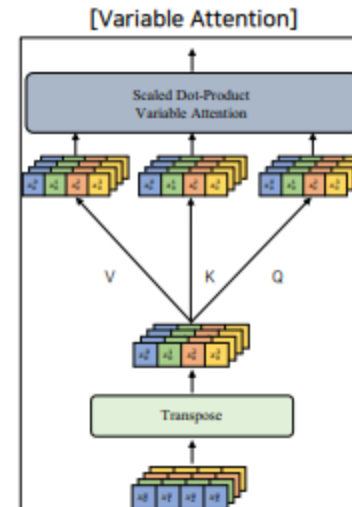
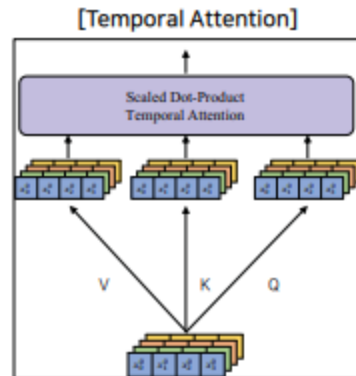


[TST]

Problem statement

❖ Variable Temporal Transformer

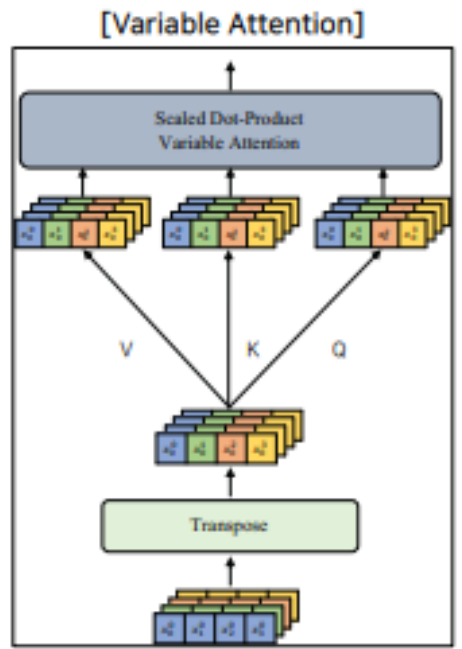
- Transformer block 내의 attention을 수정하여 변수와 시간과의 상관관계를 파악
- 변수의 정보, 상관관계 설명이 가능한 transformer가 필요



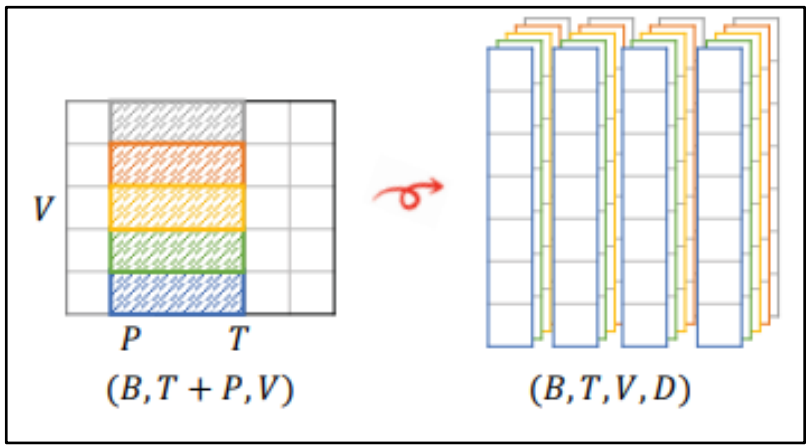
Key idea

❖ Embedding, Attention 수정

- Variable Temporal attention이라는 새로운 attention 방법을 제안
- 변수간 정보와 시간성을 유지하기 위한 임베딩 방법을 제안



제안된 Attention

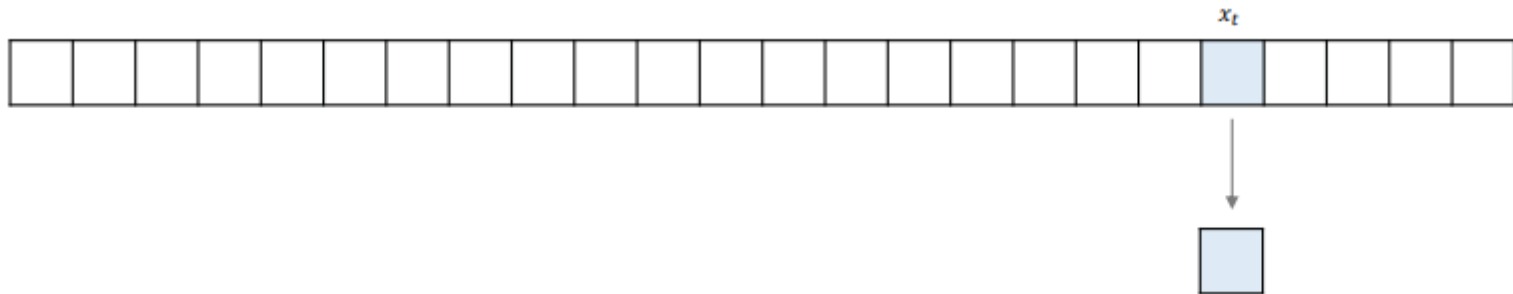
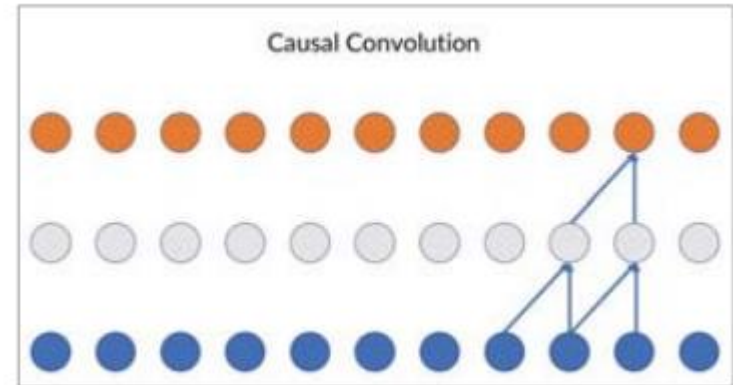
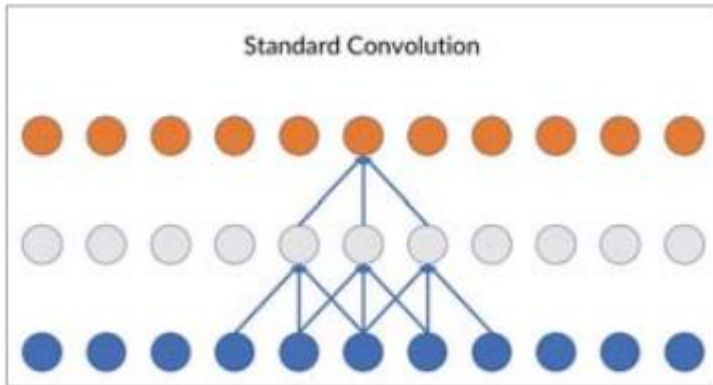


제안된 Embedding

Embedding

❖ 1D Casual Convolution

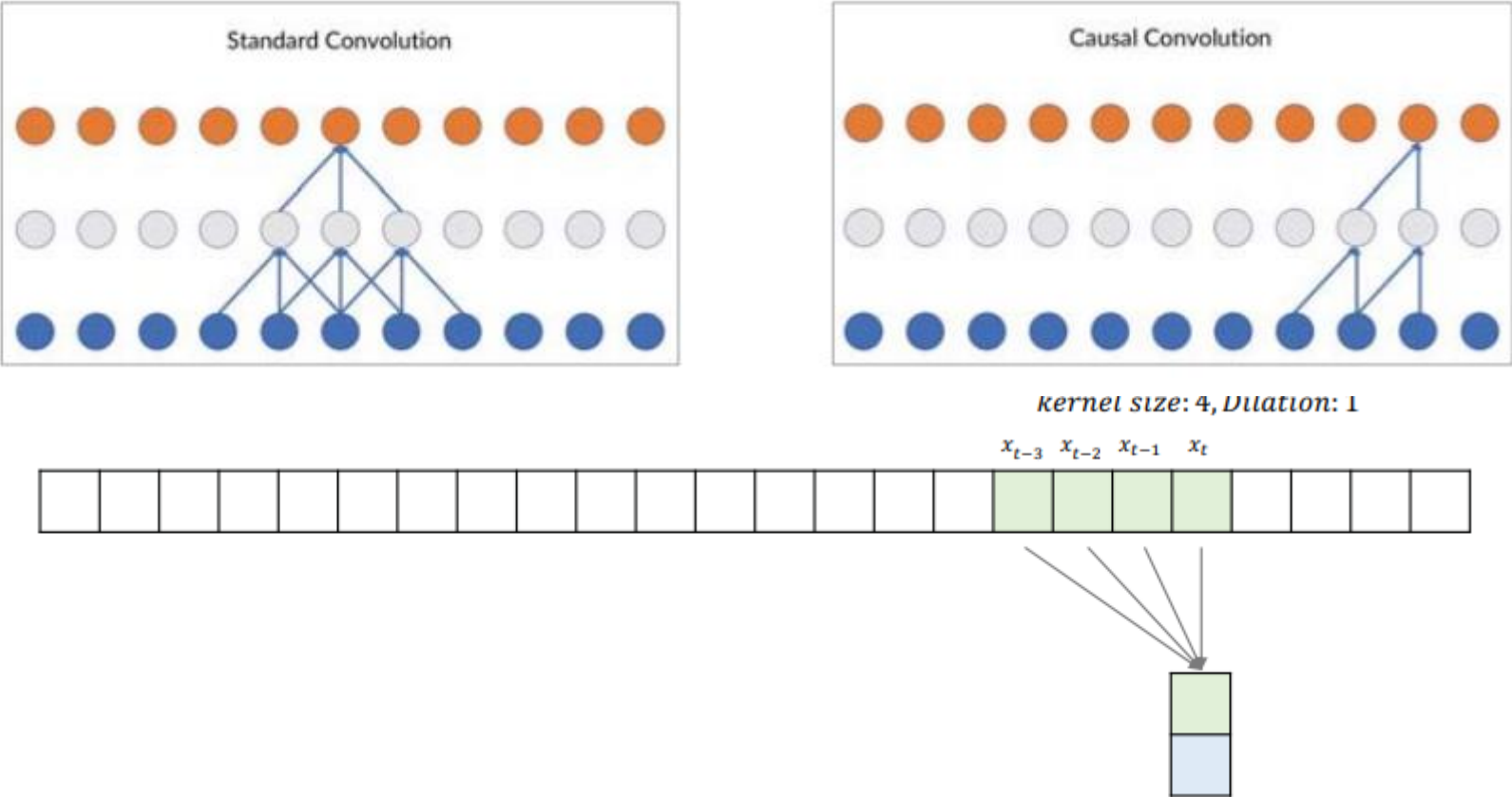
- Time step t 의 output을 내기 위해서 time step t 까지의 데이터만 고려
- Dilated casual convolution layer를 사용하여 up-sampling을 수행



Embedding

❖ 1D Casual Convolution

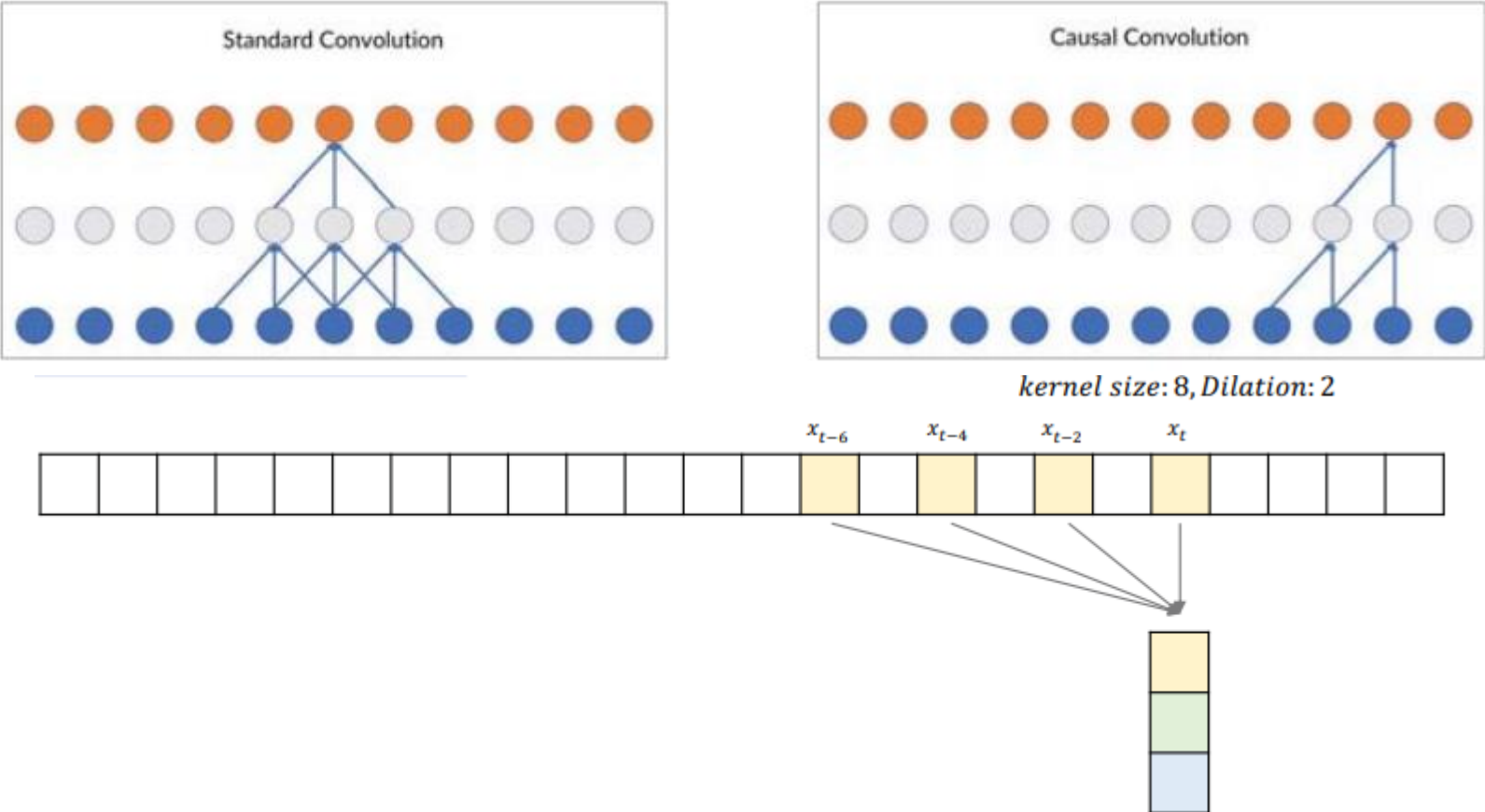
- Time step t 의 output을 내기 위해서 time step t 까지의 데이터만 고려
- Dilated casual convolution layer를 사용하여 up-sampling을 수행



Embedding

❖ 1D Casual Convolution

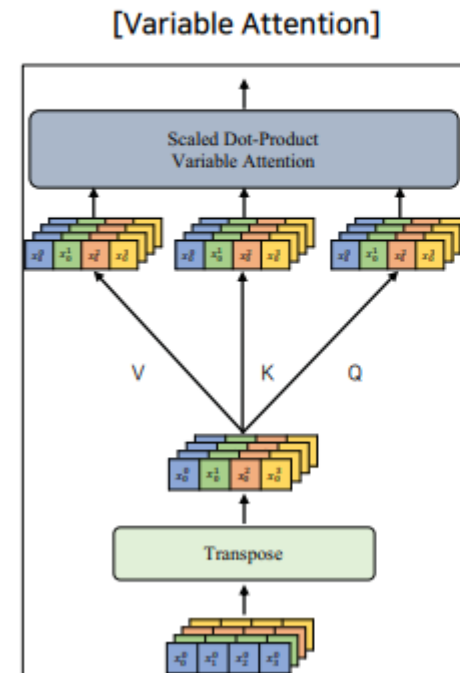
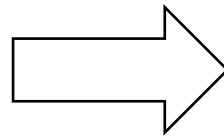
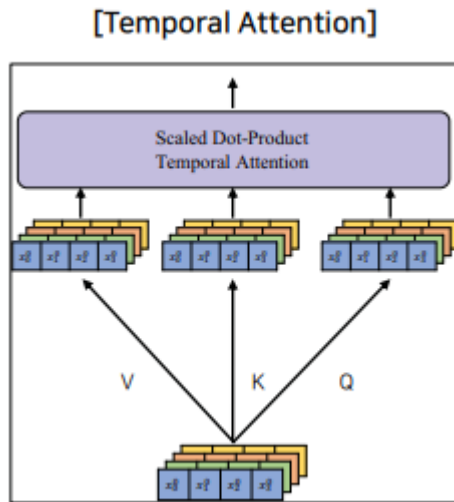
- Time step t 의 output을 내기 위해서 time step t 까지의 데이터만 고려
- Dilated casual convolution layer를 사용하여 up-sampling을 수행



Attention module

❖ Variable self-attention

- 기존의 입력을 전치하여 attention을 수행
- 변수 간의 상관관계를 고려함



Datasets

❖ 다변량 시계열 데이터 벤치마크 데이터셋

- SWaT (the Secure Water Treatment)
 - 수질 관련 시계열 데이터
- PSM (Pooled Server Metrics)
 - Ebay에서 Application server node를 기록함
- SMD (Server Machine Dataset)
 - 인터넷 회사에서 수집, 총 28개의 server에서 5주간 수집한 33개의 metric 정보임

	SWaT	PSM	SMD
Applications	Water	Server	Server
Variables	51	25	38
Sub-dataset	X	X	28
#Train (0.8)	396000	105984	566724
#Valid (0.2)	99000	26497	141681
#Test (labeled)	449919	87841	708420
Anomaly (%)	11.98	27.8	4.16

Comparison of results of the overall model

❖ 전체 모델과 비교 했을 때의 결과

Dataset	SWaT				PSM				SMD			
	P	R	F1	AUROC	P	R	F1	AUROC	P	R	F1	AUROC
LSTM-AE	0.907	0.8237	0.8534	0.906	0.9848	0.9709	0.9778	0.9828	0.9804	0.9057	0.9364	0.9525
LSTM-VAE	0.9848	0.6896	0.8112	0.8441	0.8866	0.8944	0.8905	0.9265	0.9810	0.9036	0.9356	0.9515
BeatGAN	0.8895	0.8569	0.8729	0.9210	0.9838	0.8948	0.9371	0.9447	0.9815	0.9046	0.9364	0.9520
USAD	0.8842	0.8569	0.8703	0.9206	0.966	0.8949	0.9291	0.9417	0.9810	0.9036	0.9356	0.9515
DAGMM	0.9295	0.7906	0.8545	0.8911	0.9979	0.8512	0.9187	0.9253	0.9826	0.9014	0.9353	0.9504
OmniAnomaly	0.9998	0.6764	0.8069	0.8382	0.6046	0.9592	0.7417	0.8662	0.9802	0.9037	0.9352	0.9515
VT-FAT	0.931	0.8952	0.9127	0.943	0.9725	0.9811	0.9768	0.9855	0.9861	0.8728	0.9166	0.9361
VT-EAT	0.8999	0.847	0.8727	0.917	0.9434	0.9035	0.923	0.9419	0.9809	0.9101	0.9391	0.9547
VT-SAT	0.9299	0.9849	0.9565	0.9873	0.9607	0.97	0.9653	0.9778	0.9804	0.9103	0.9390	0.9549
VT-PAT	0.9504	0.9463	0.9484	0.9698	0.9504	0.936	0.9484	0.9747	0.9831	0.9071	0.9387	0.9533

Comparison of results of the Transformer model

❖ 기존 Transformer 모델과 비교 진행

- 기존 Transformer보다 제안된 방법에서 더 좋은 성능을 볼 수 있음

Dataset	SWaT				PSM				SMD			
	P	R	F1	AUROC	P	R	F1	AUROC	P	R	F1	AUROC
TF	0.9054	0.8339	0.8682	0.9109	<u>0.9885</u>	0.8948	0.9393	0.9493	0.9802	<u>0.9103</u>	0.9389	0.9548
VT-FAT	0.931	0.8952	0.9127	0.943	0.9725	<u>0.9811</u>	<u>0.9768</u>	<u>0.9855</u>	<u>0.9861</u>	0.8728	0.9166	0.9361
VT-EAT	0.8999	0.847	0.8727	0.917	0.9434	0.9035	0.923	0.9419	0.9809	0.9101	<u>0.9391</u>	0.9547
VT-SAT	0.9299	<u>0.9849</u>	<u>0.9565</u>	<u>0.9873</u>	0.9607	0.97	0.9653	0.9778	0.9804	<u>0.9103</u>	0.9390	<u>0.9549</u>
VT-PAT	<u>0.9504</u>	0.9463	0.9484	0.9698	0.9504	0.936	0.9484	0.9747	0.9831	0.9071	0.9387	0.9533

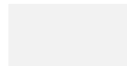
Conclusion

❖ Conclusion

- Temporal attention과 variable attention을 통해 해석 가능한 모델을 제안
- Kernel 사이즈를 다양하게 한 1d casual convolution을 통해 multi resolution 정보를 활용함
- 기존 모델 6가지와 비교했을 때 더 좋은 성능을 보임

❖ Evaluation

- 몇 개의 성능 지표에서 유의미한 성능 증가를 보임
- 임계치 설정 방법이 Best F1-Score는 이상 탐지 결과 비교에 적합하지 않아 보임



QnA
