

23.12.22

다변량 스팀 사용 이상 감지 및 영향 변수의 원인 분석

서울과학기술대학교 데이터사이언스학과

이성호 sean0310@seoultech.ac.kr

배소희 shbae2819@g.seoultech.ac.kr

심재웅 jaewoong@seoultech.ac.kr

데이터

■ 이용 데이터

– 제품 2종에 대한 6달간 센서 데이터

df_ext(2023-03,04)(5123,0385)_2023-11-16 seoultech

df_ext(2023-05,06)(5123,0385)_2023-11-16 seoultech

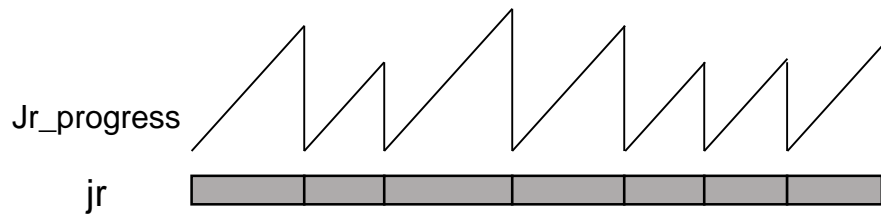
df_ext(2023-07,08)(5123,0385)_2023-11-16 seoultech

기간 : 2023-03-02 08:00:00 ~ 2023-08-27 03:00:00 (분)

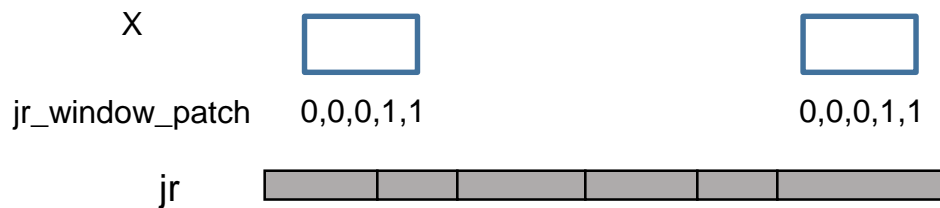
date	날짜
tg	Sensor (38개)
stop	공정 분석값 0 : 가동 1 : 중지 이벤트 발생 2 : 중지 복구
jr	단위 공정값 / 제품 생산 주기 (생산품 번호)

– 이용 데이터 전처리

- jr_progress : jr을 기준으로 시간에 따른 정수를 새로운 변수로 추가



- jr_window_patch : window 사이즈 만큼의 데이터 패치 후, 예측하고자 하는 시점의 jr과 다른지 같은지 표기하는 새로운 변수 추가

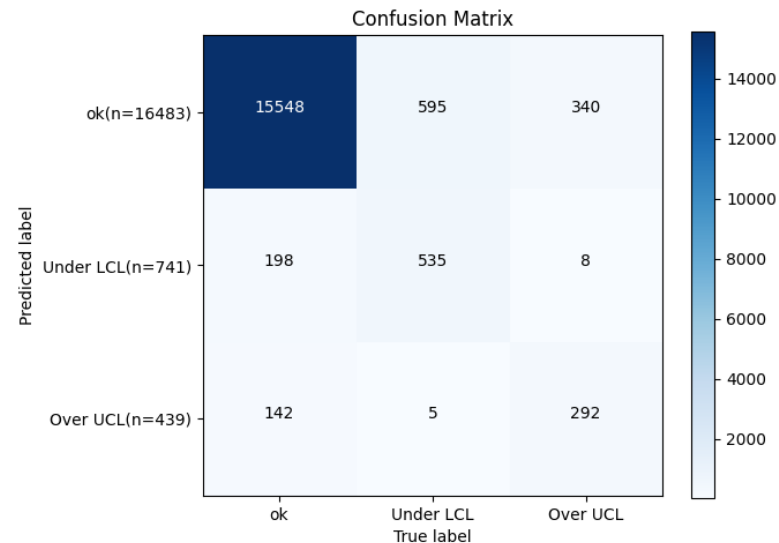
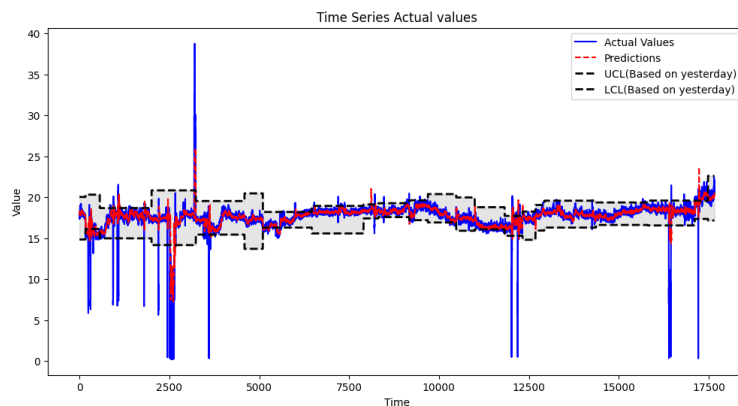


데이터

■ 새로운 파생 변수 비교 (IMV-LSTM)

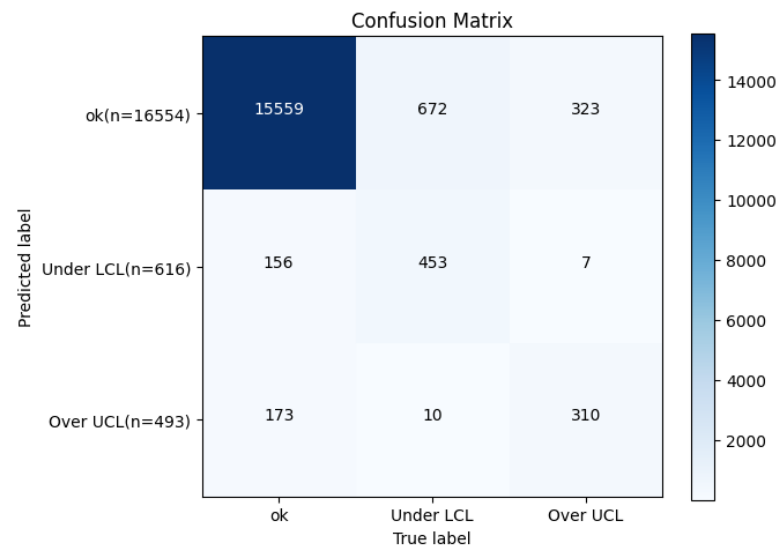
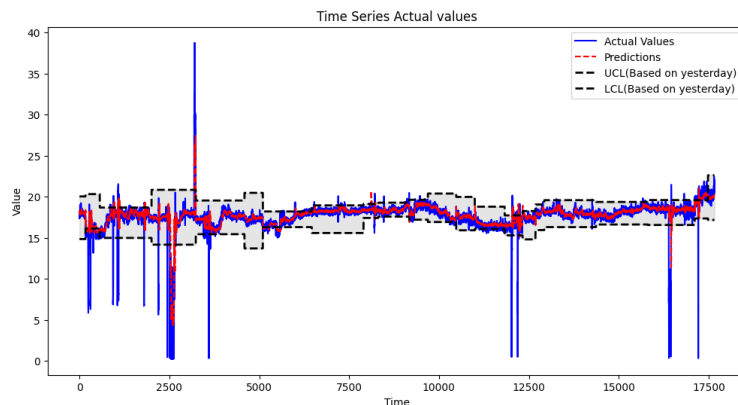
– 센서 데이터 + jr_progress

- 모든 테스트 데이터셋에 대해 (R2 : 0.5098, Acc : 0.9606, Macro F1 : 5704)



– 센서 데이터 + jr_progress + jr_window_patch

- 모든 테스트 데이터셋에 대해 (R2 : 0.4896, Acc : 0.9241, Macro F1 : 0.6746)

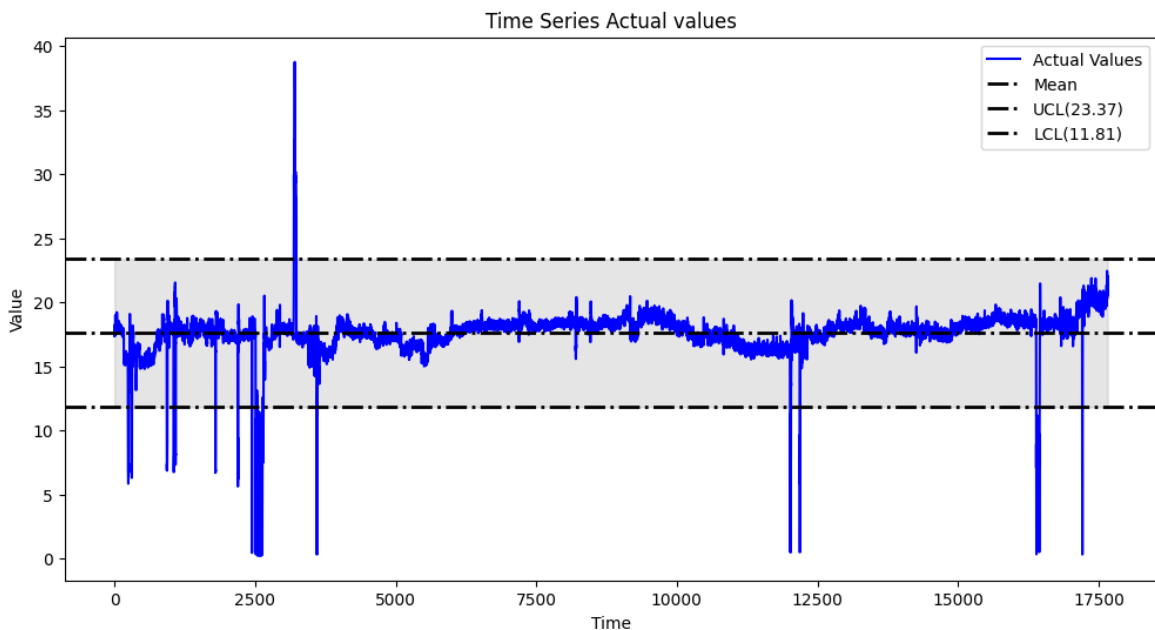


데이터

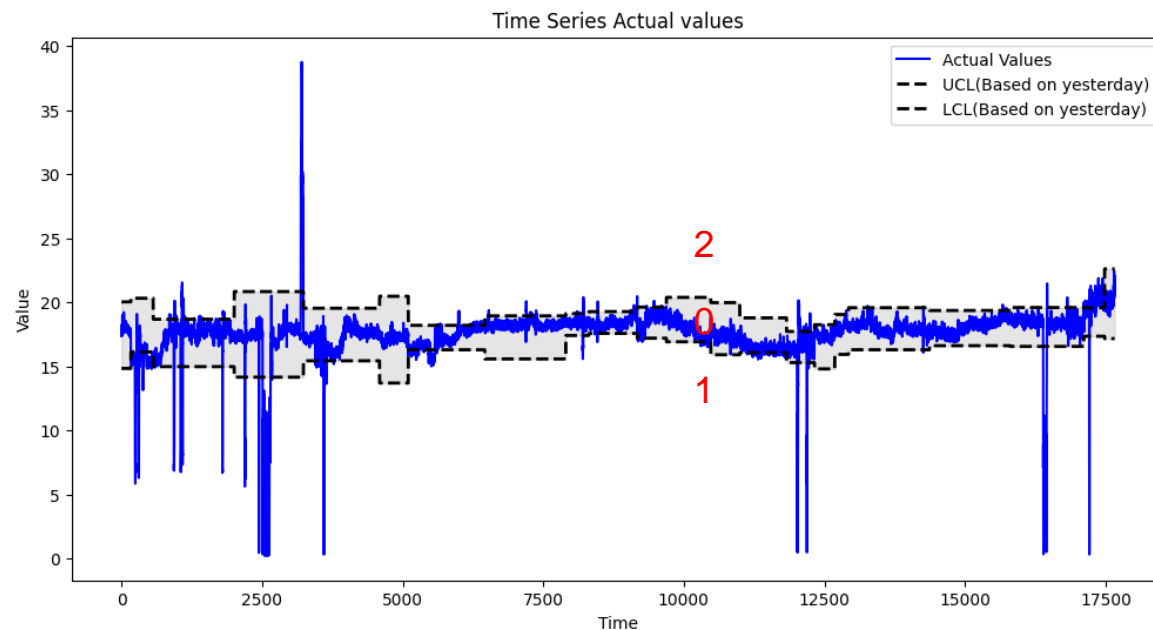
- 이용 데이터
 - 이용 데이터 정의
 - **input(X) : tg (37개 sensor 데이터), 공정진행도(jr_progress) , 공정변화도(jr_window_patch)**
 - Window size : 20
 - **output(y) : 10분 후의 tg 04(스팀 순간값)**
 - Data split
 - Train/Validation/Test = 60:20:20
 - Train : 2023-03-02 08:00:00 ~ 2023-06-16 18:22:00 (52980)
 - Validation : ~ 2023-07-16 12:37:00 (17660)
 - Test : ~ 2023-08-27 03:00:00 (17663)

이상에 대한 정의 및 탐지

- 전체 테스트 데이터셋 기준 $\text{mean} \pm 3\sigma$



- 최근 1일 동안의 데이터 (60 * 24개 시점) 기준으로 관리 한계선을 매일 업데이트
 - $\text{mean} \pm 3\sigma$ 기준
 - 이상으로 탐지되었던 시점은 계산에서 제외



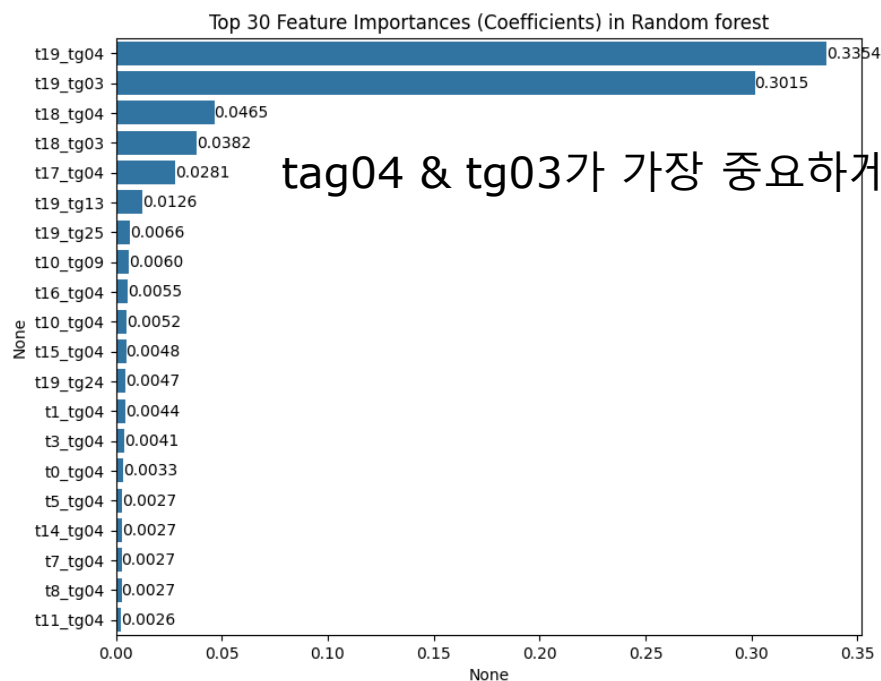
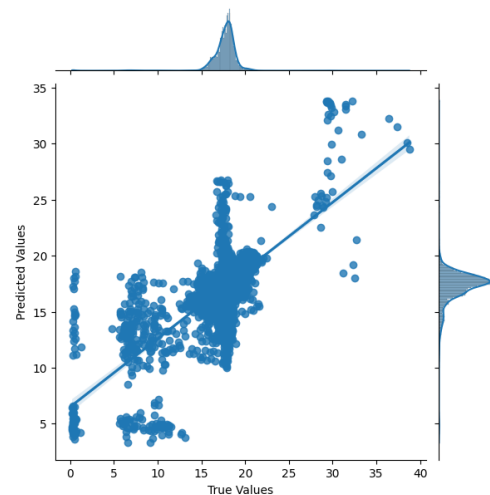
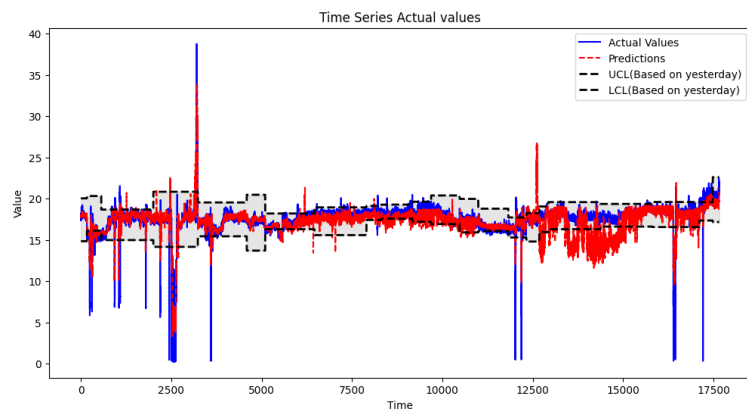
- 관리한계선에 따른 클래스 지정
 - 관리한계선 이상 : 2
 - 관리한계선 이하 : 1
 - 관리한계선 내부 : 0

머신 러닝 모델 (Random forest)

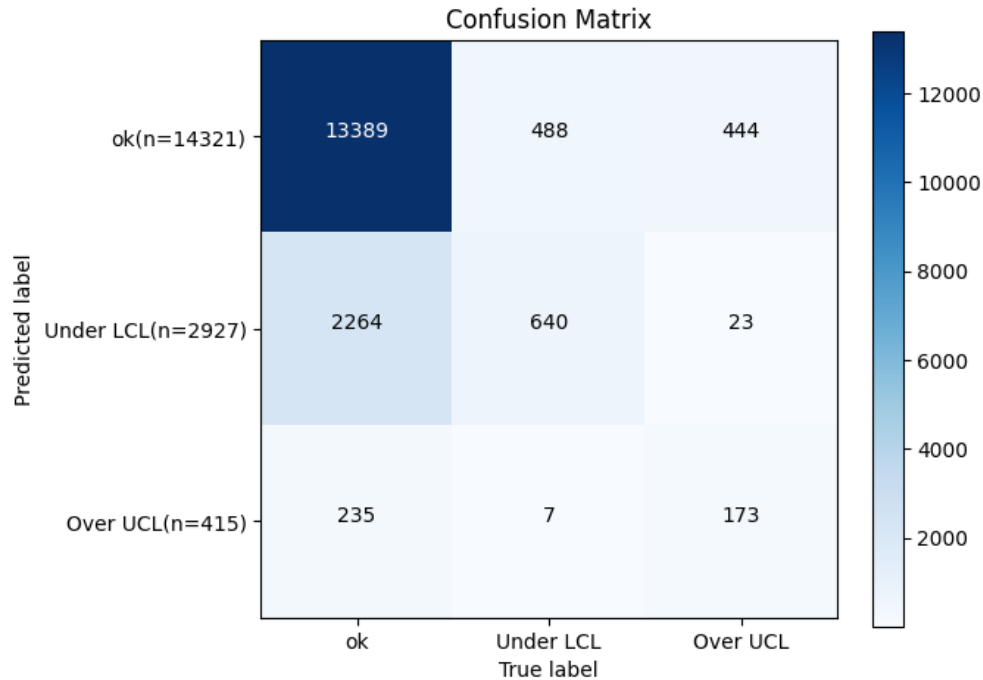
■ 예측 성능 지표

- R^2 : 0.4051
- MSE : 2.2104
- Acc : 0.8041
- F1 : 0.5098

■ 변수 중요도



tag04 & tg03가 가장 중요하게 작용



딥러닝 모델 (LSTM)

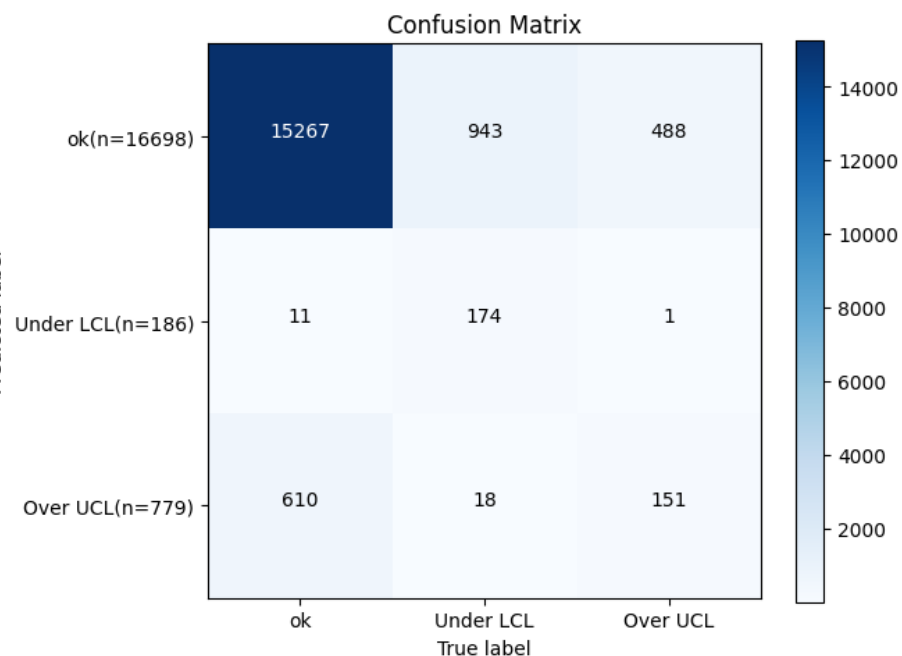
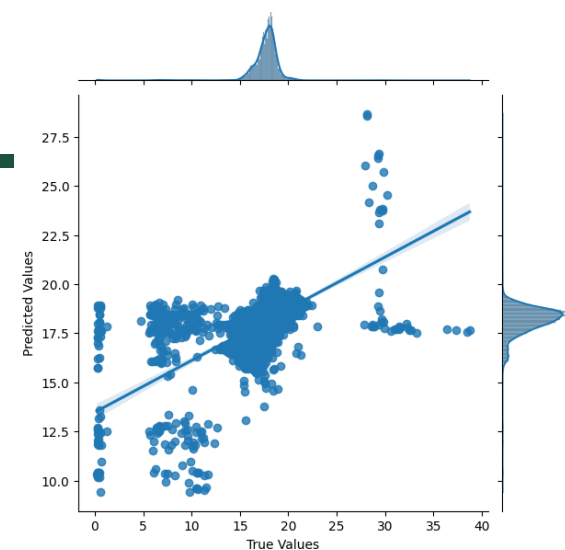
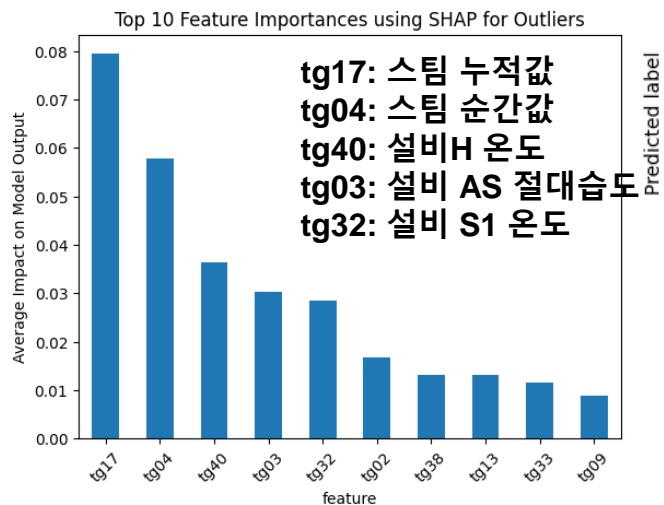
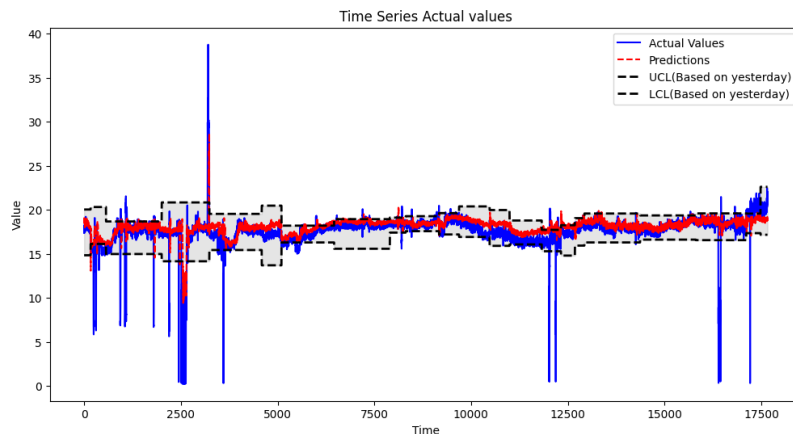
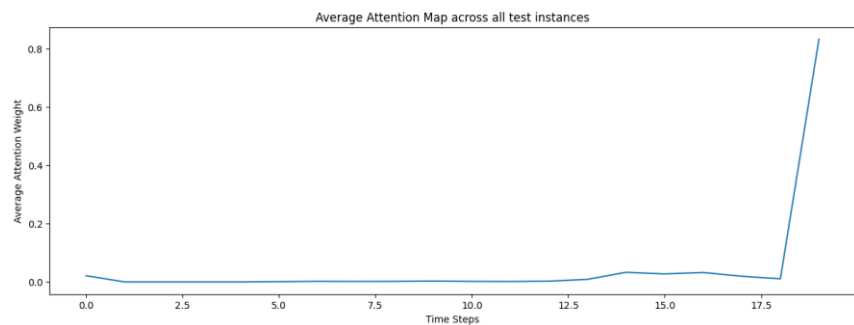
Experiment setting

- LSTM layer(hidden=128, layer=4) + attention layer
- Epoch : 100
- optimizer : Adam(lr=1e-4)

Result

- R^2 : 0.2547
- MSE : 2.7691
- Acc : 0.8827
- F1 : 0.4771

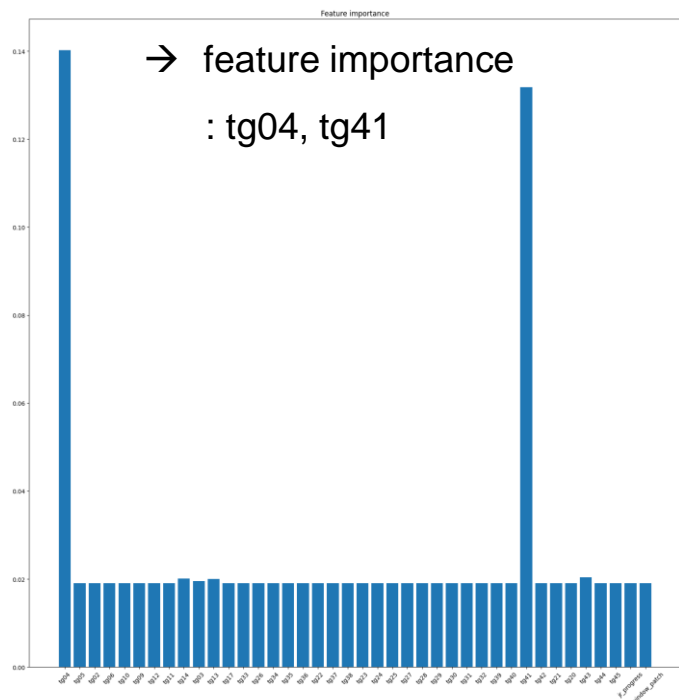
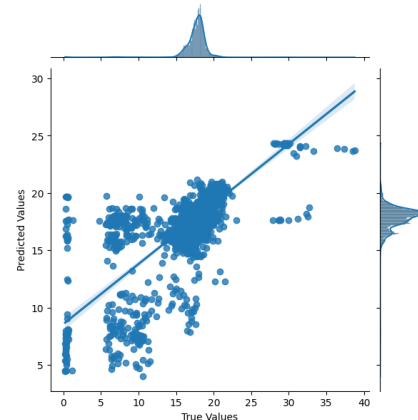
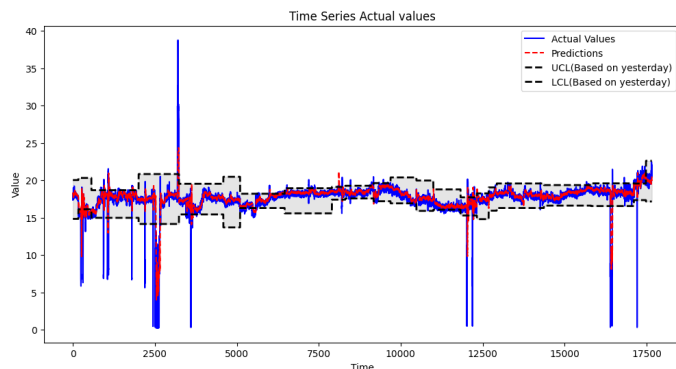
모델 해석



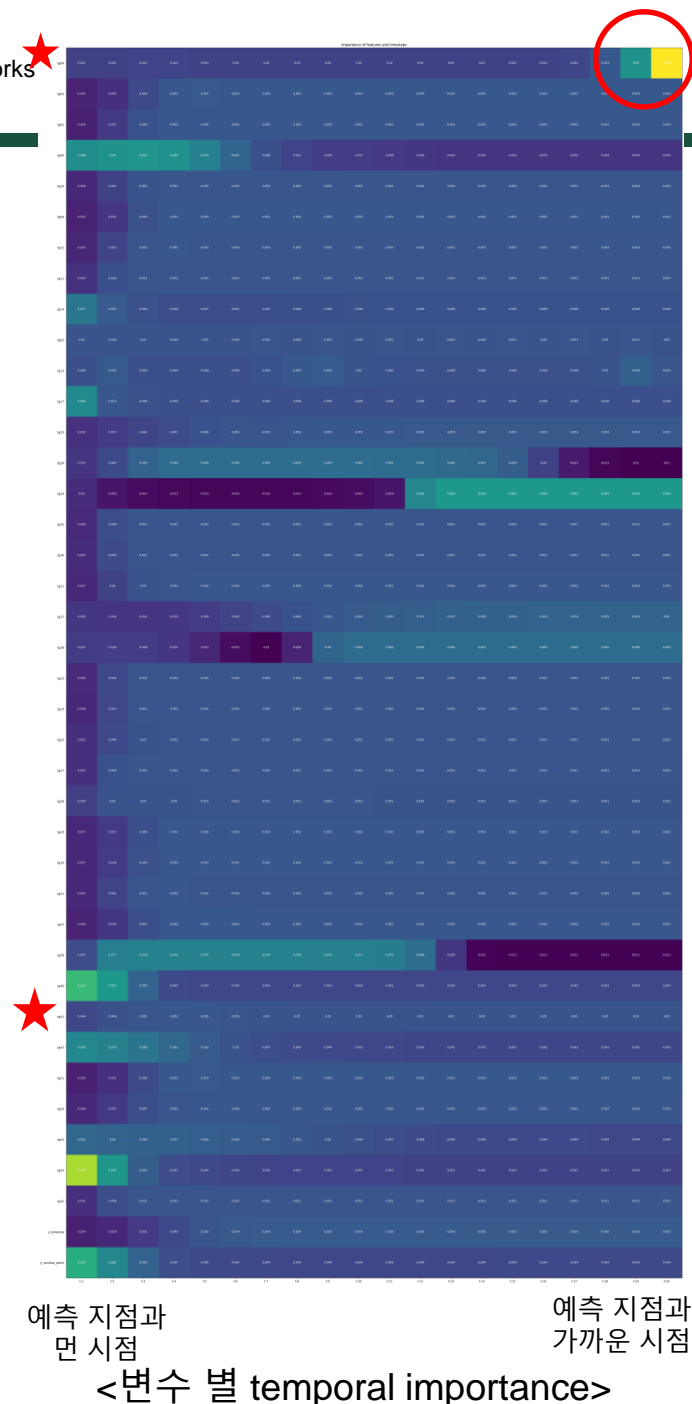
딥러닝 모델 (IMV-LSTM)

Guo, Tian, Tao Lin, and Nino Antulov-Fantulin. "Exploring interpretable LSTM neural networks over multi-variable data." *International conference on machine learning*. PMLR, 2019.

- Experiment setting
 - Lstm layer node : 32
 - Epoch : 100
 - Optimizer : Adam(lr=1e-3)
- Result
 - R^2 : 0.5399
 - MSE : 1.7094
 - Acc : 0.9265
 - F1 : 0.6810



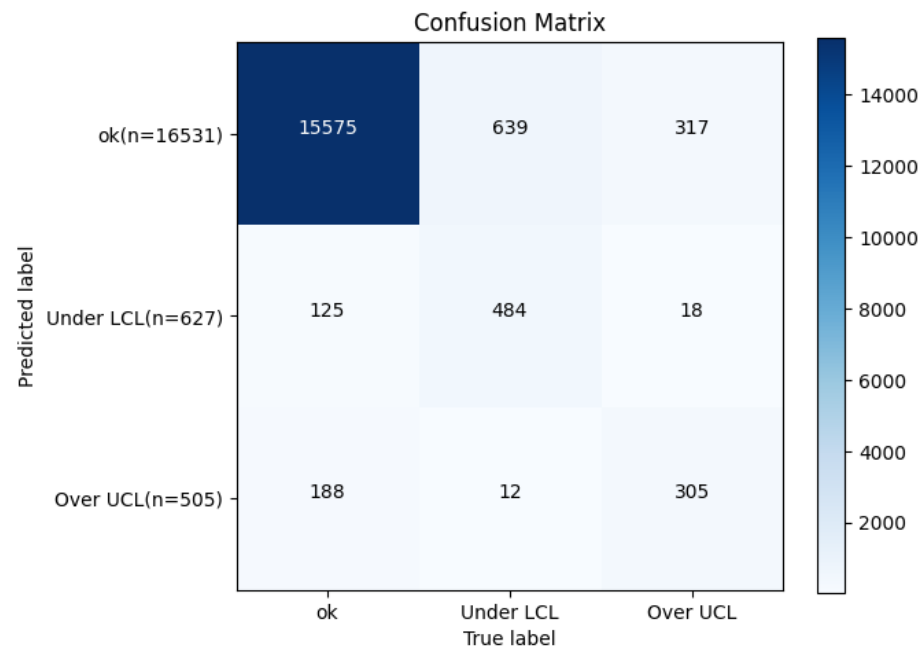
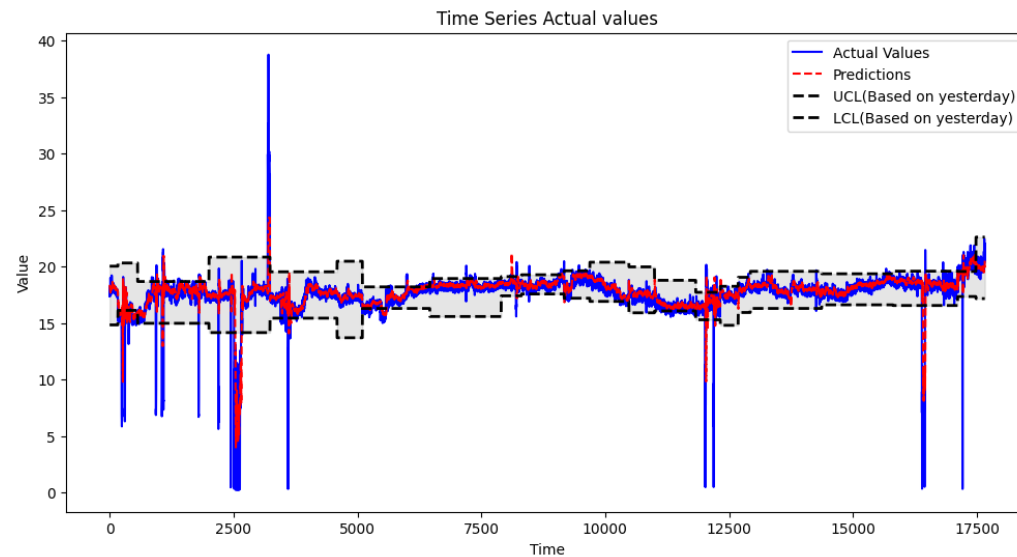
- Attention Map을 통한 모델 해석



딥러닝 모델 (IMV-LSTM)

Guo, Tian, Tao Lin, and Nino Antulov-Fantulin. "Exploring interpretable LSTM neural networks over multi-variable data." *International conference on machine learning*. PMLR, 2019.

- Experiment setting
 - Lstm layer node : 32
 - Epoch : 100
 - Optimizer : Adam(lr=1e-3)
- Result
 - R^2 : 0.5399
 - MSE : 1.7094
 - Acc : 0.9265
 - F1 : 0.6810



계획

- 각 이상 탐지 결과에 대한 해석 (Global interpretation → Local interpretation)
- 적절한 성능 지표 개발

감사합니다