

23.11.13

다변량 스팀 사용 이상 감지 및 영향 변수의 원인 분석

서울과학기술대학교 산업공학과/데이터사이언스학과

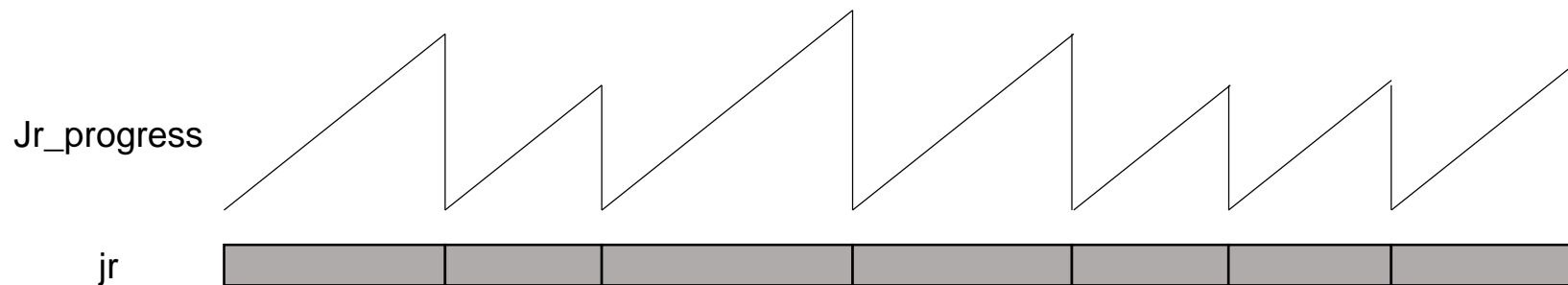
이성호 sean0310@seoultech.ac.kr

배소희 shbae2819@g.seoultech.ac.kr

심재웅 jaewoong@seoultech.ac.kr

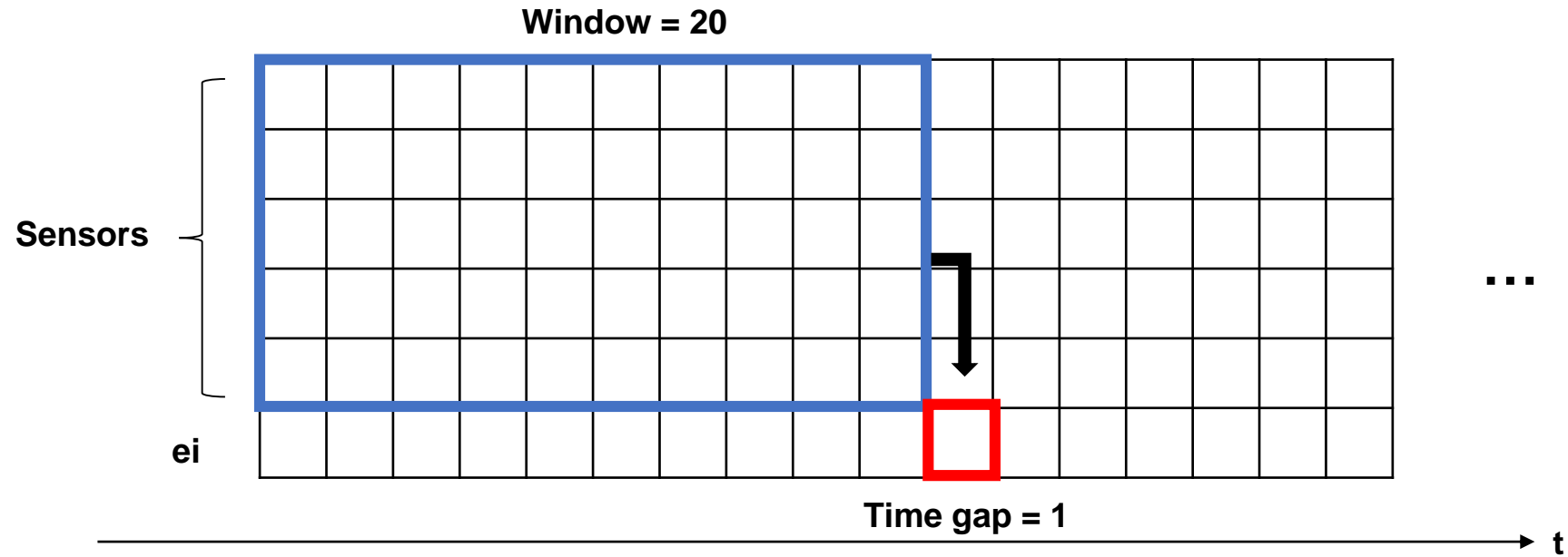
데이터 전처리

- tg02, tg04 제거 (tg02: 종이별 측정 무게, tg04: 스팀 순간값)
 - Ei값 산출에 직접 사용되는 변수인 것으로 파악되어 모델 input에서 제거
- Jr 변수 재구성
 - 각 제품의 생산 경과 시간 (분 단위)을 표현한 jr_progress 변수 도입
 - 각 제품의 생산 시작 시점을 0으로 하여 새로운 제품이 생산되기 전까지 1분에 1씩 증가



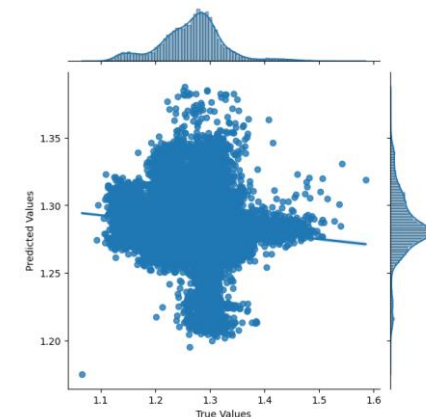
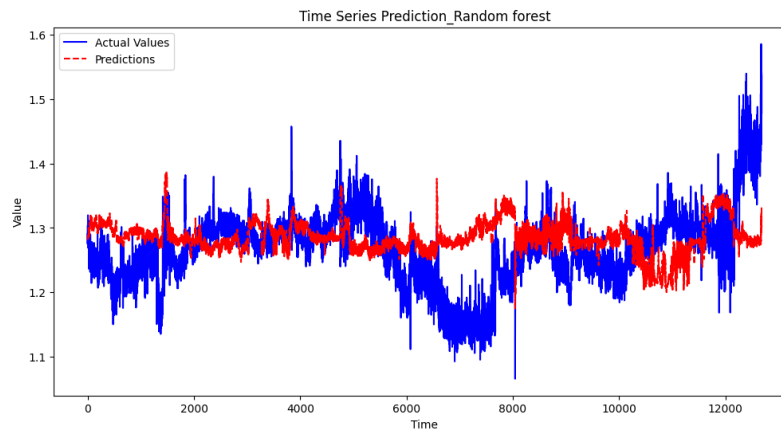
데이터 전처리

- 미래 시점에 대한 예측 모델



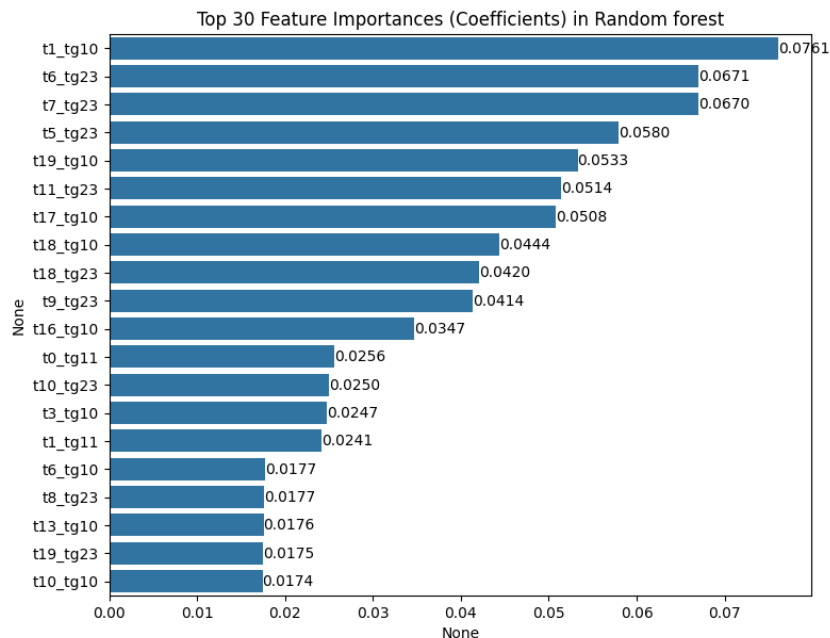
머신 러닝 모델 (Random forest)

- 예측 성능 지표
 - R2 : -0.3543
 - MSE : 0.0046



- 변수 중요도

일반 머신러닝 모델은 동시점을 제외했을때, 예측 성능이 제한됨



tag02 & tg04 제거 이후, tg10(설비B1 속도)과 tg23(설비 BE1 속도)이 가장 중요하게 작용

특정 시점의 센서 정보가 중요하다는 정보는 관찰되지 않음 (다양한 시점이 관측)

딥러닝 모델 (LSTM)

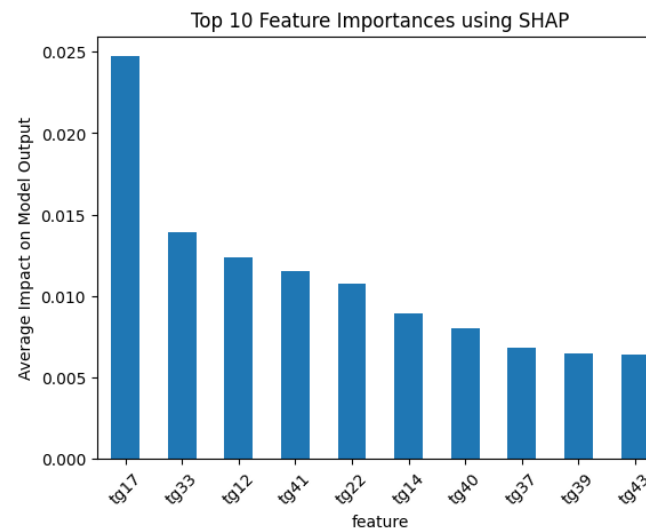
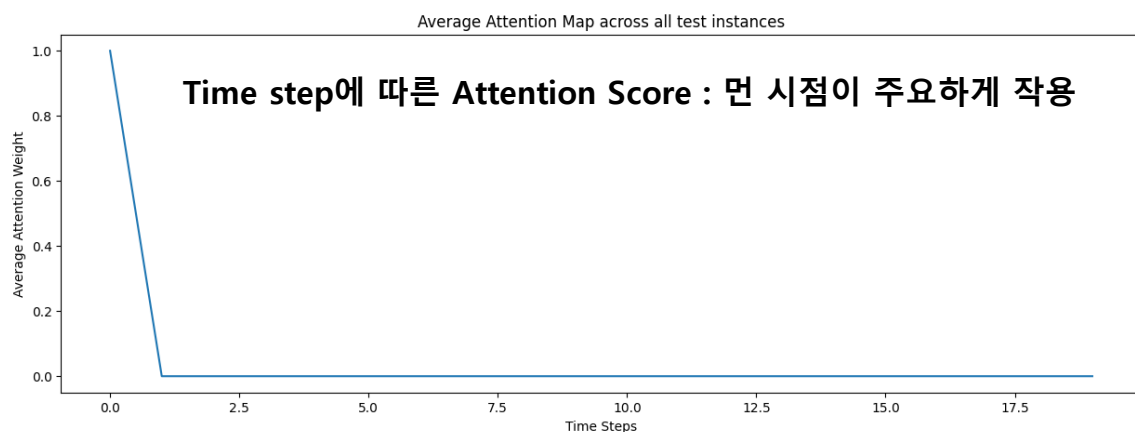
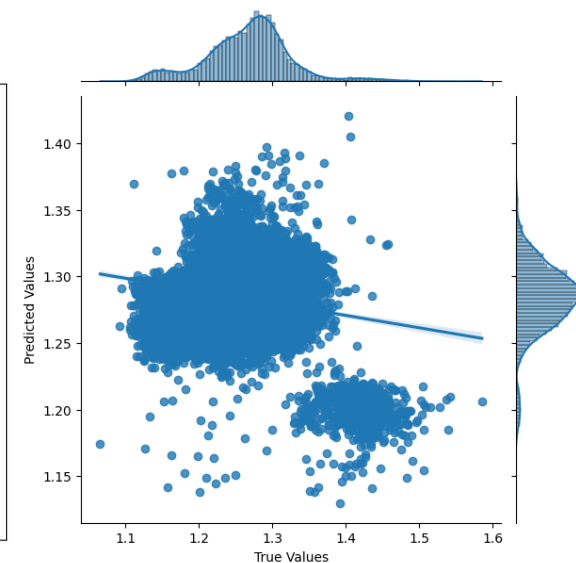
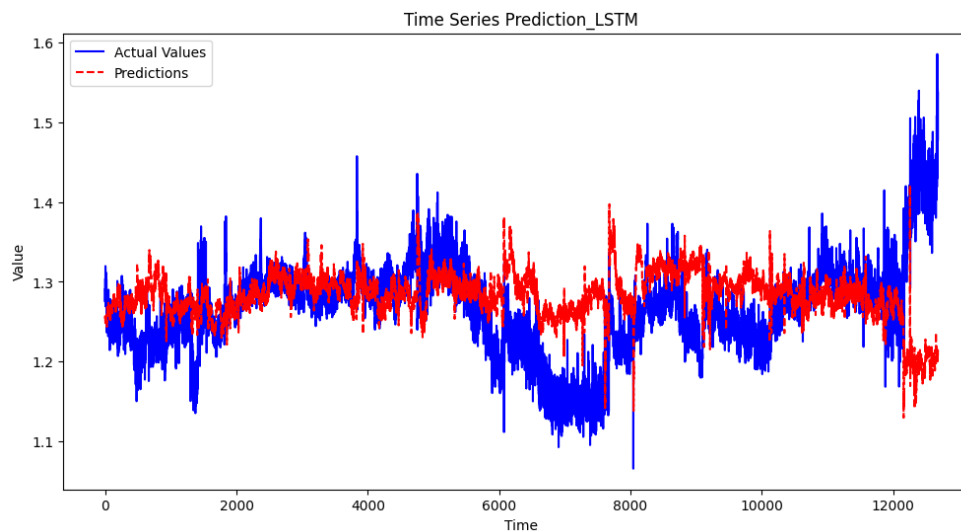
Experiment setting

- LSTM layer(hidden=256, layer=8) + attention layer
- Epoch : 100
- optimizer : Adam(lr=1e-4)

Result

- R2 : -0.5000
- MSE : 0.0052

✦ 모델 해석



tg17: 스팀 누적값
tg33: 수분값
tg12: 설비B2 속도
tg41: 스팀 온도
tg22: 설비 PE1 속도

딥러닝 모델 (1D CNN)

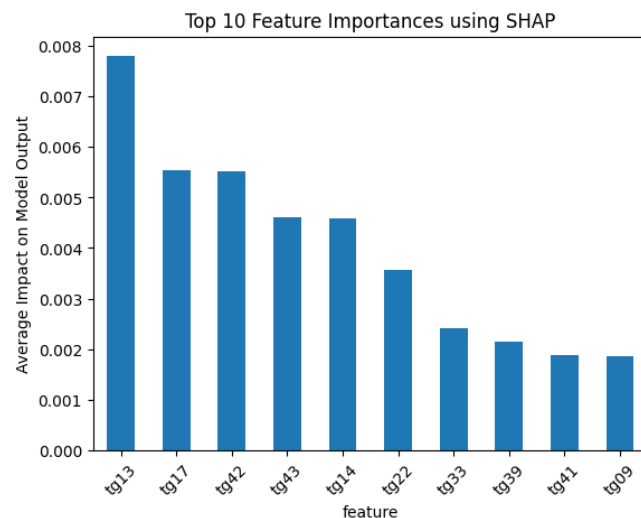
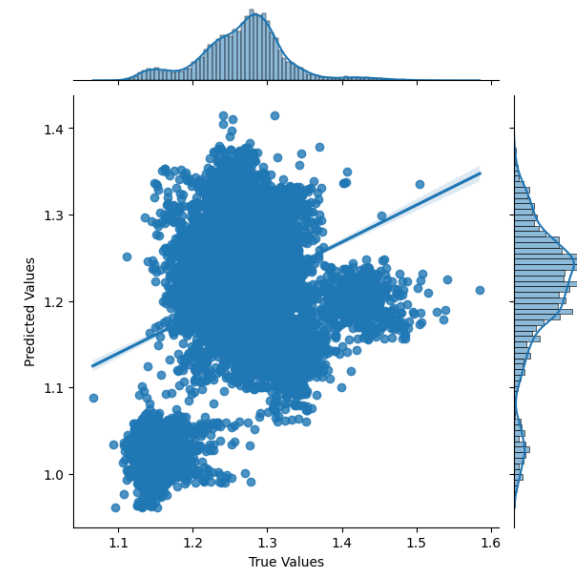
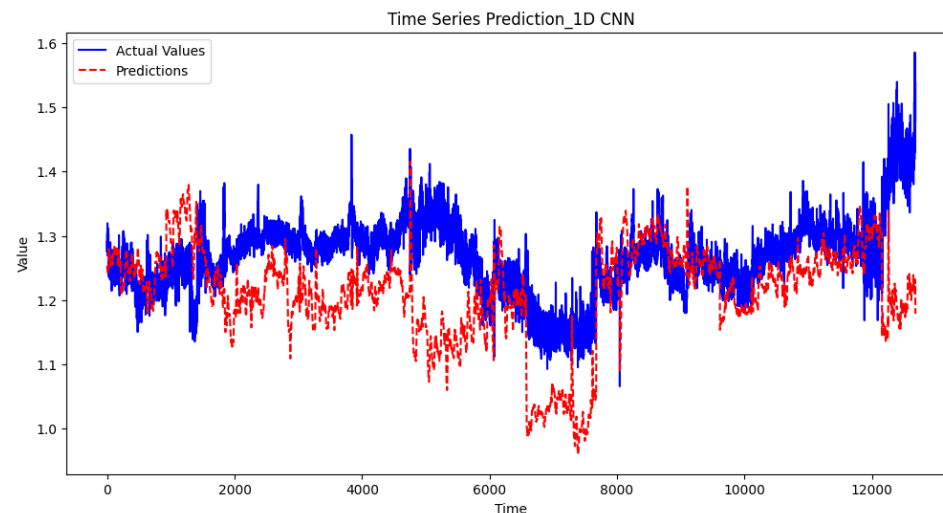
Experiment setting

- 1D Conv layers(64-128-256-512-1024 / kernel=3) + linear layers (1024-512-128-64-1)
- Epoch : 100
- optimizer : Adam(lr=1e-4)

Result

- R2 : -1.7503
- MSE : 0.0095

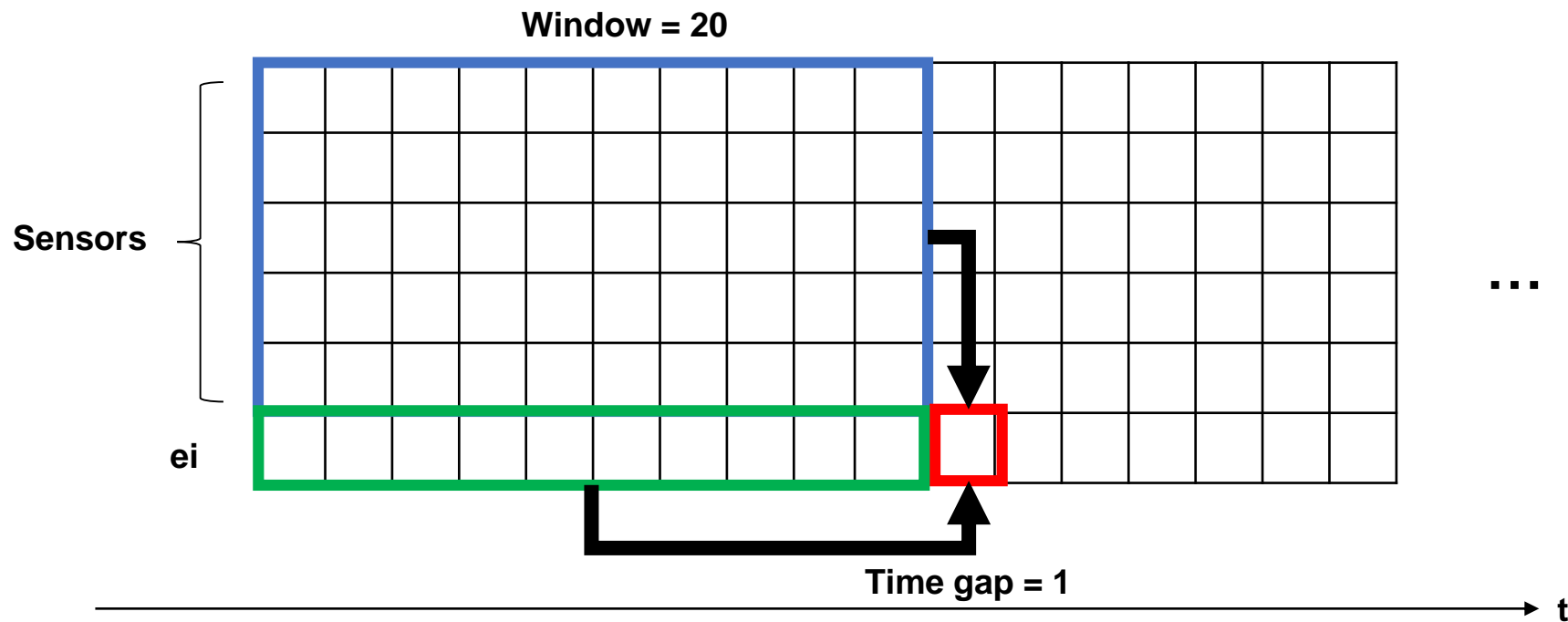
✦ SHAP을 통한 모델 해석



tg13: 설비B1 절대습도
tg17: 스팀 누적값
tg42: 설비 PE1 온도
tg43: 설비 PE2 온도
tg14: 설비B2 절대습도

데이터 전처리

- 이전 타겟 정보까지 사용한 미래 시점에 대한 예측 모델
 - 과거의 e_i 값을 모델 input으로 사용



딥러닝 모델 (LSTM)

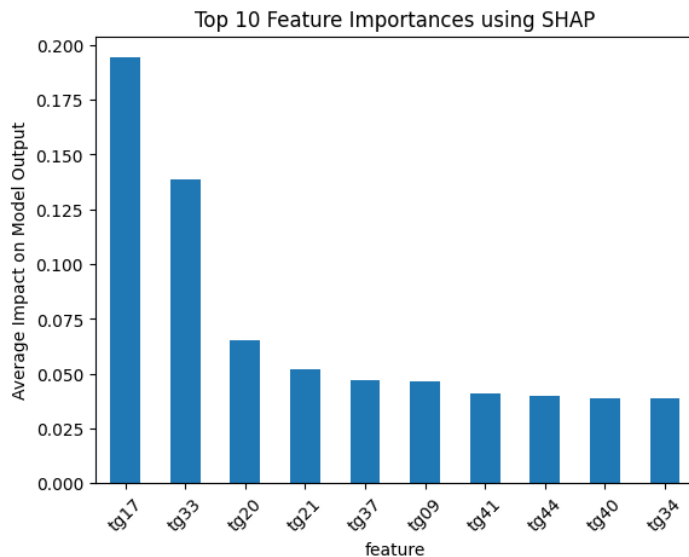
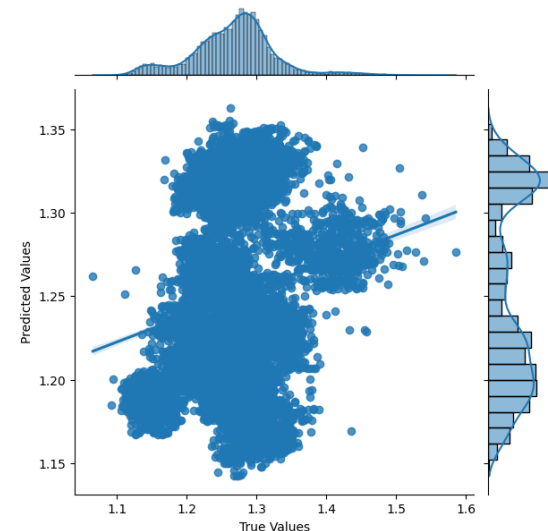
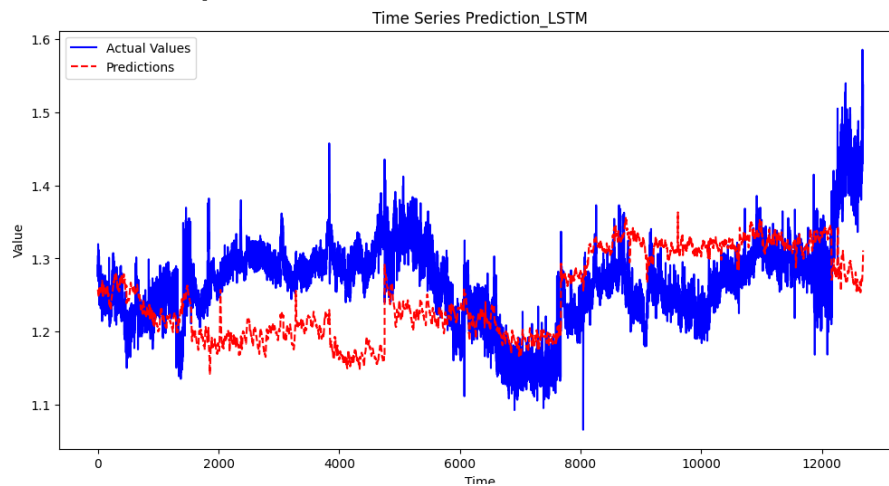
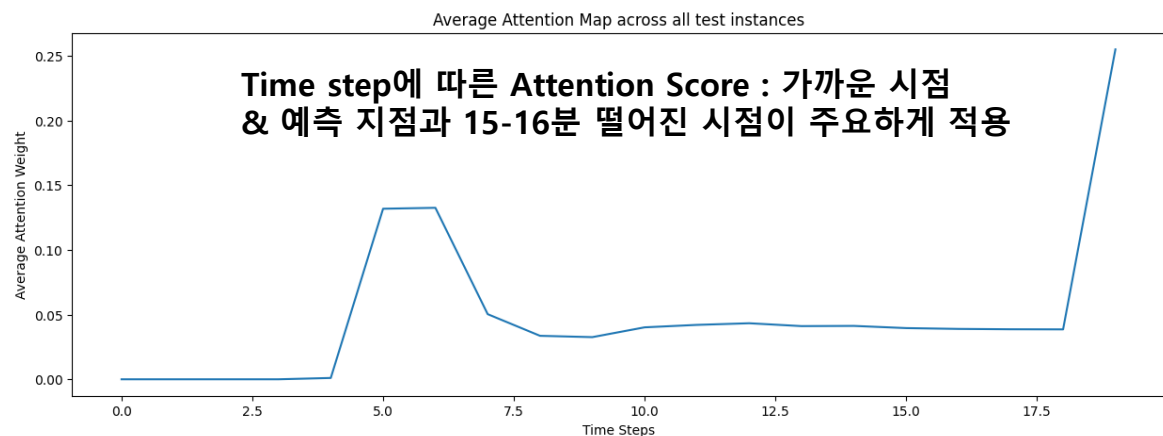
Experiment setting

- LSTM layer(hidden=256, layer=8) + attention layer
- Epoch : 100
- optimizer : Adam(lr=1e-4)
- 이전 시점의 ei 변수 적용

Result

- R2 : -0.7034
- MSE : 0.0059

✦ 모델 해석



tg17: 스팀 누적값
tg33: 수분값
tg20: 스팀 압력

※ ei값의 중요도는 예상외로 낮음

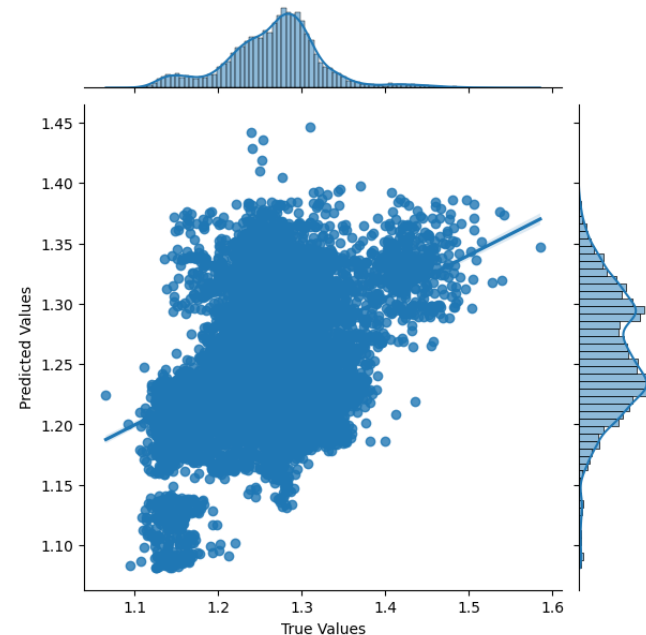
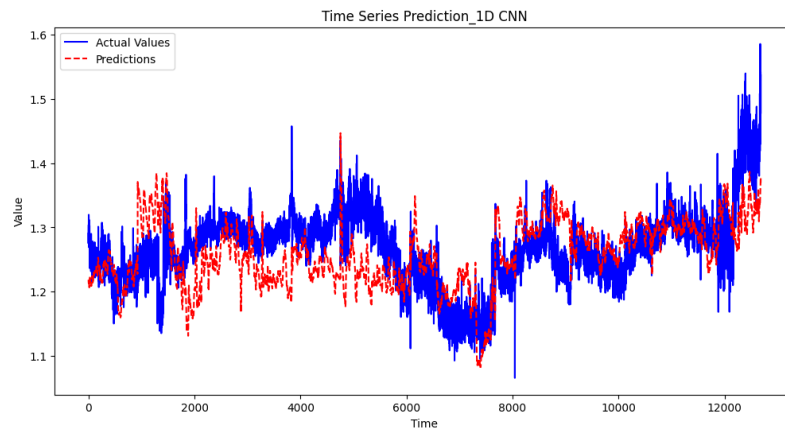
딥러닝 모델 (1D CNN)

Experiment setting

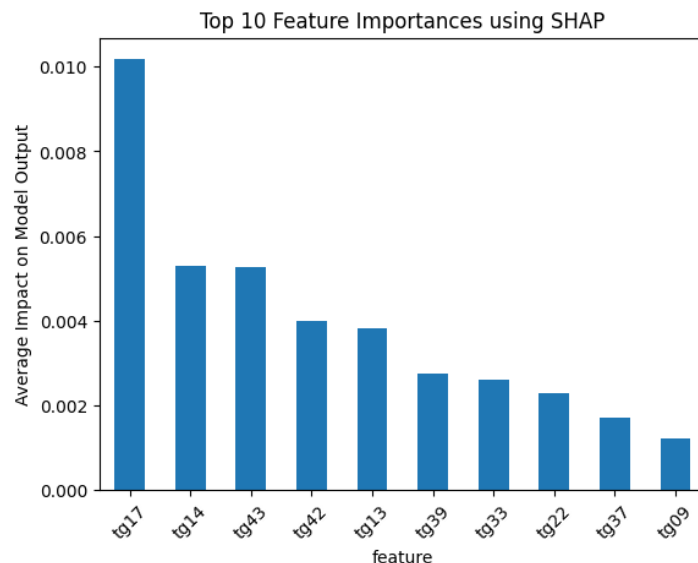
- 1D Conv layers(64-128-256-512 / kernel=3) + linear layers (512-128-64-1)
- Epoch : 200
- optimizer : Adam(lr=1e-4)
- 이전 시점의 ei 변수 적용

Result

- R2 : -0.0838
- MSE : 0.0037



✦ SHAP을 통한 모델 해석

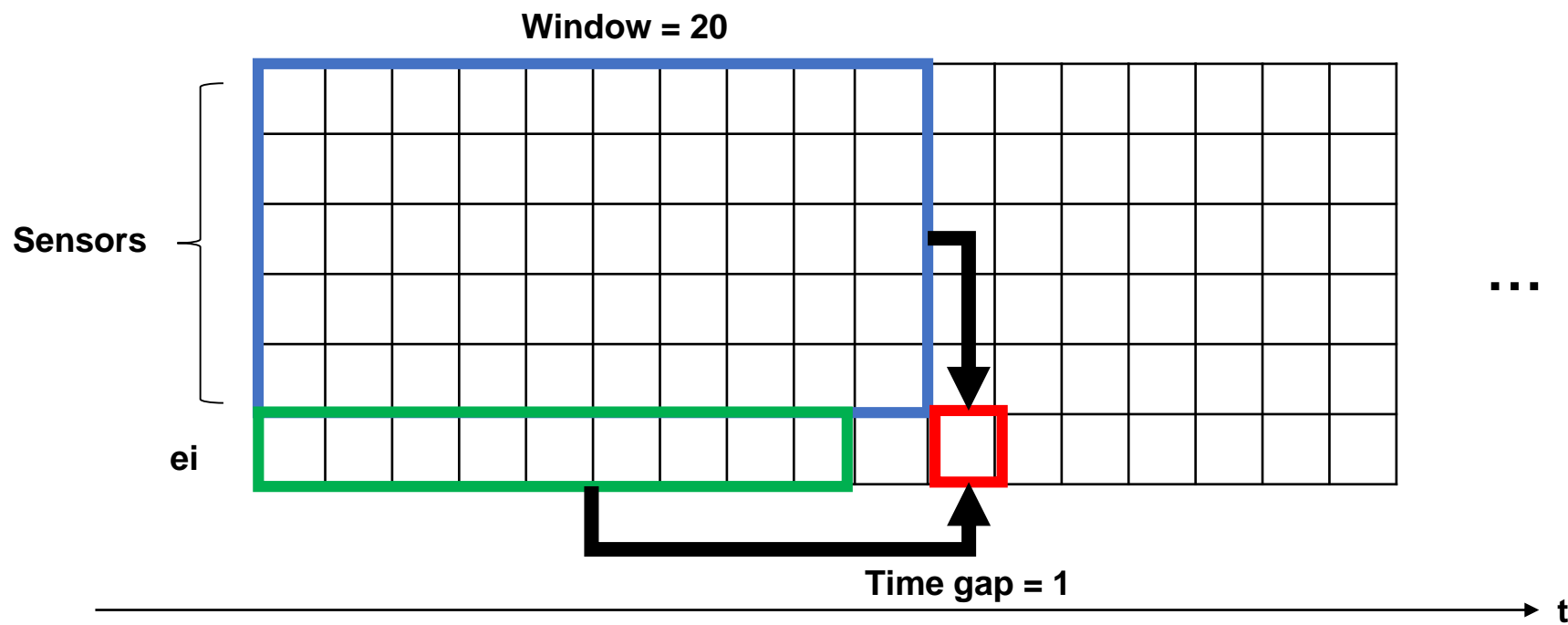


tg17: 스팀 누적값
tg14: 설비B2 절대습도
tg43: 설비 PE2 온도
tg42: 설비 PE1 온도
tg13: 설비B1 절대습도

※ ei값의 중요도는 예상외로 낮음

데이터 전처리 – DARNN (Dual Attention RNN)

- 이전 타겟 정보까지 사용한 미래 시점에 대한 예측 모델
 - 과거의 ei값을 모델 input으로 사용
 - input data: 20 timestep 동안의 Sensor 값들과 19 timestep 동안의 ei로 이루어진 텐서
 - Output : 21 timestep의 ei($\hat{y}_{T+1} = F(y_1, y_2, \dots, y_{T-1}, x_1, x_2, \dots, x_T)$)



딥러닝 모델 (DARNN)

Experiment setting

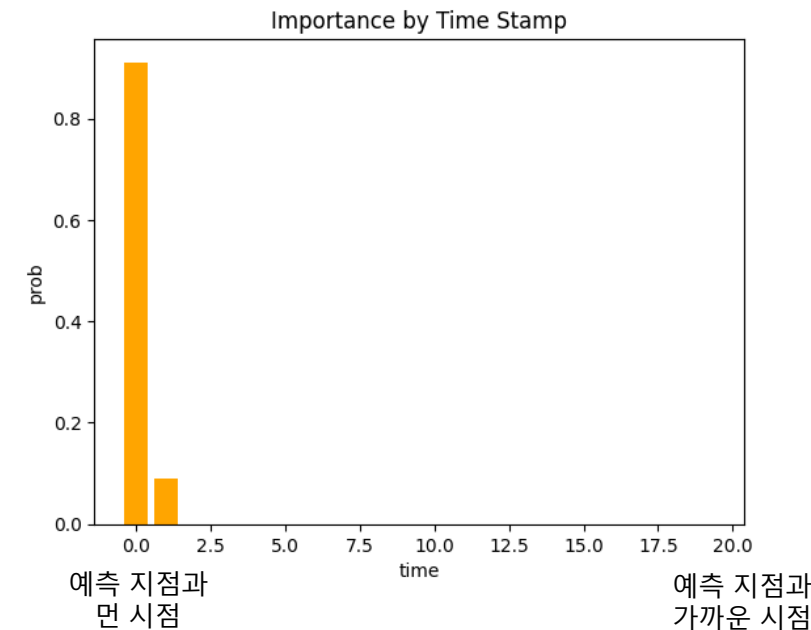
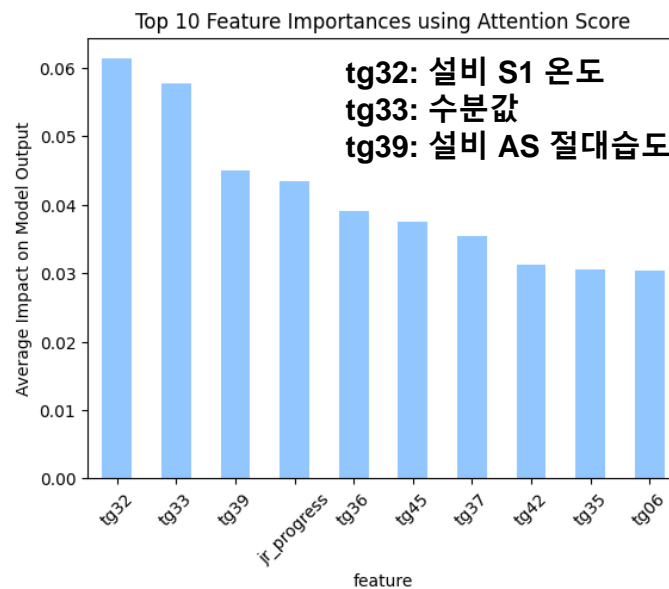
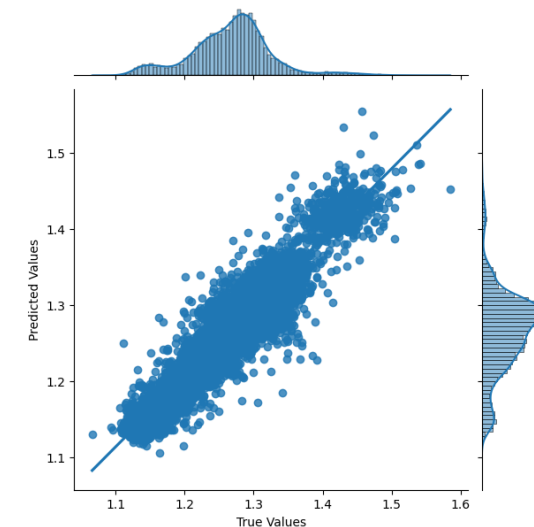
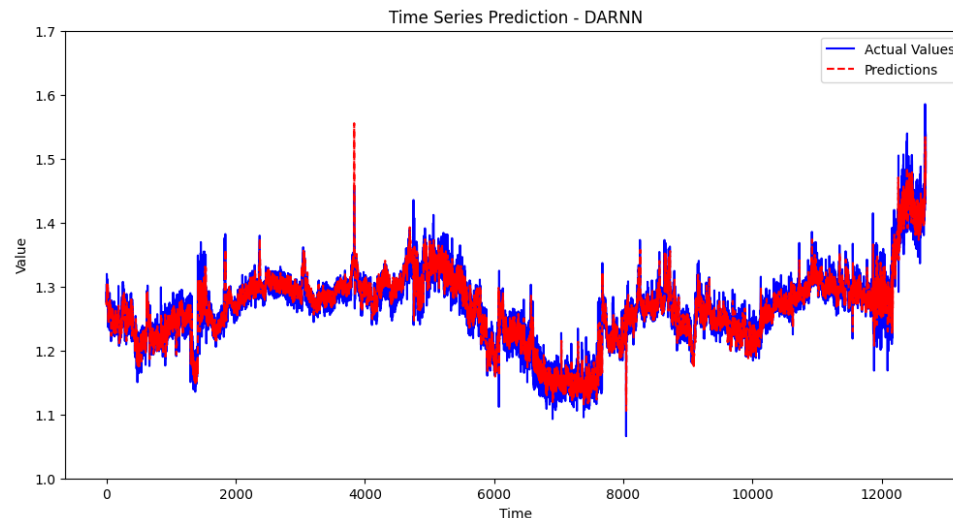
- Encoder & Decoder Layer : 16
- Epoch : 200
- Optimizer : Adam(lr=1e-3)

Result

- R2 : 0.8757
- MSE : 0.0004

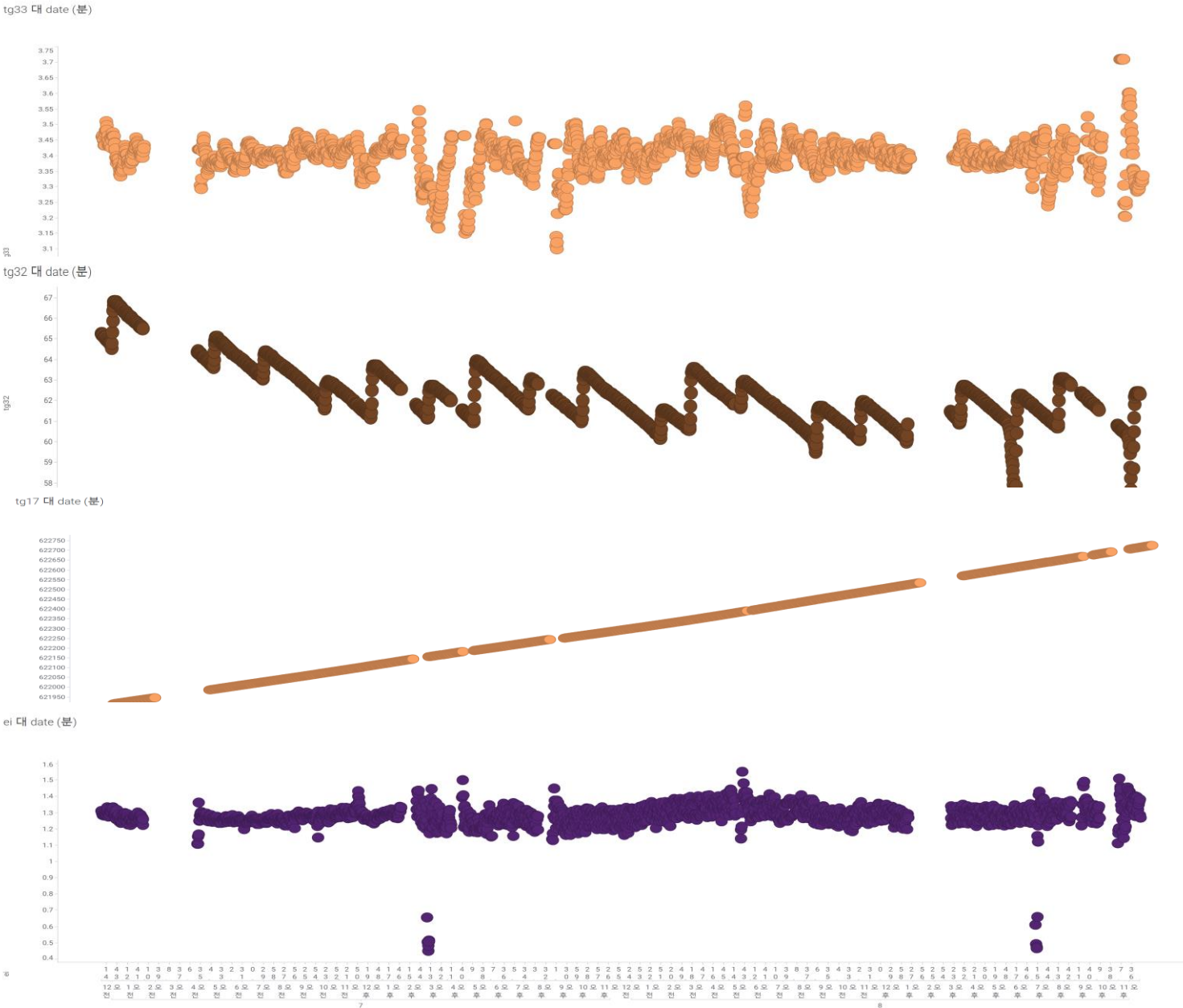
✦ Attention Score를 통한 모델 해석

- 먼 시점의 데이터가 더 주요하게 작용
- tg32, tg33이 중요하게 작용
- 새로 전처리한 jr이 중요하게 작용



tg33, tg32, tg17

- tg33: 수분값
- tg32: 설비 S1 온도
- tg17: 스팀누적값



요약 & 논의

■ 요약

- 모델 input 변수에 변화를 줌 (tg02, tg04, jr_progress)
- Dual attention RNN 도입 시도

■ 논의

- 구체적인 문제 정의 필요
 - 사전 예측(Forecasting)에 대한 해석 모델 vs 일어난 일에 대한 사후 해석 모델
 - 지난 ei 급등 사례의 경우 사후 해석
 - 탐지의 대상
 - Ei의 급격한 변화를 탐지하는 것이 목표?
 - 어떤 시점에서 detect가 요구되는지
 - 의미있는 Time gap의 정도, sensor값과 ei값이 획득되는 시점
- 변수 설명
 - tg04: 스팀 순간값
 - tg05: 추가 스팀값
 - Tg17: 스팀누적값

감사합니다