

2025 09 04
발표 자료

광운대학교 로봇학과
FAIR Lab

김한서

1D CNN 모델

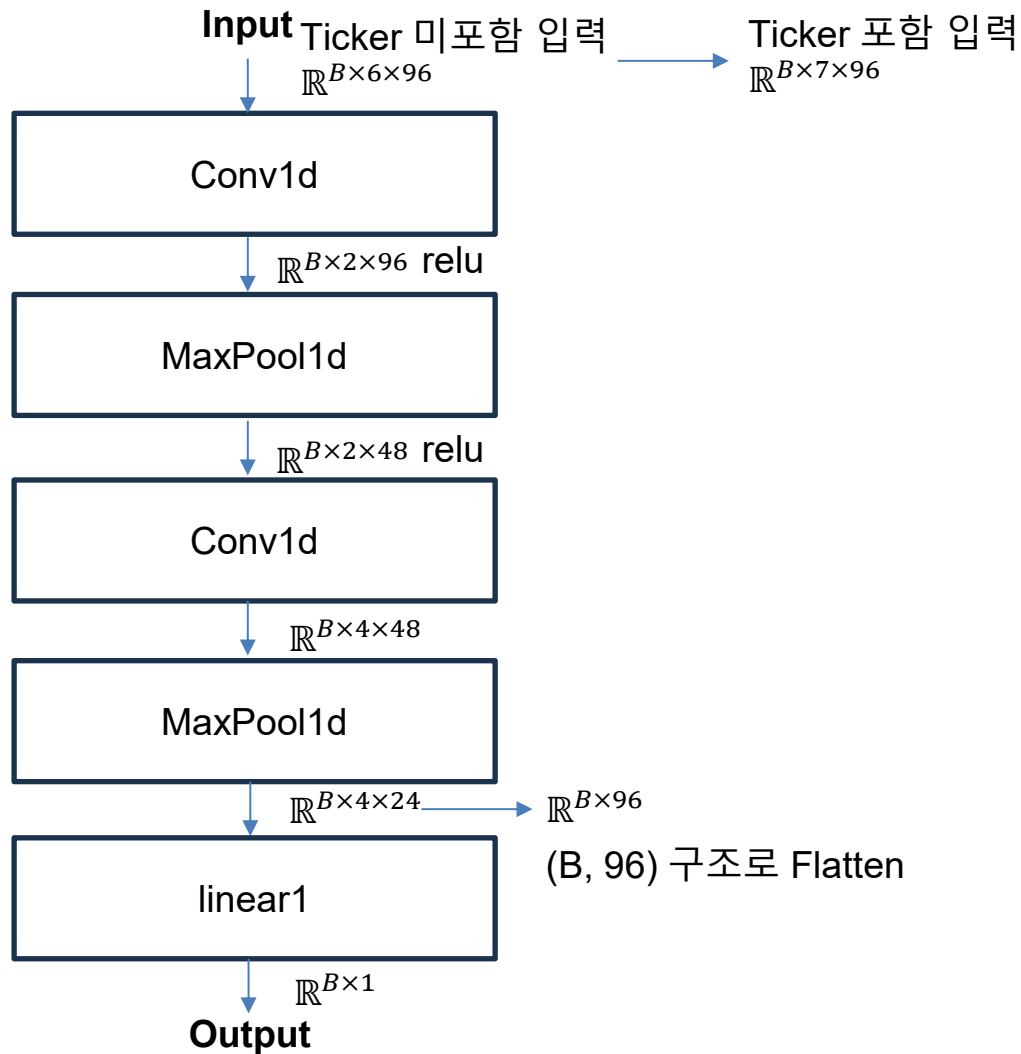
이번 주 진행사항

- 1D CNN 모델 구성
- 종목별 StandardScaler 적용 확인
- Train, Valid, Test 각각의 종목 개수 확인
- Train, Valid, Test 날짜 설정 확인 및 샘플 수 확인
- 제외된 종목명과 개수 확인
- 제외된 종목이 Train 기간 이후 언제 상장됐는지 확인
- Ticker 포함/미포함 실험 결과 비교
- Ticker 포함/미포함 시각화 비교

Quant data 모델 학습

현재 진행 중인 실험 설명

- Input / Label 전 종목 사용
- Input feature → Open, Close, High, Low, Volume, Vwap
Label feature → 20_day_return_rate
- Ticker 포함/미포함 실험 결과 비교
- 상위 3개 종목 샘플 시각화 비교
각 종목마다 80개 샘플의 GroundTruth, Prediction 화면에 표시



*B : Batch Size

Learning rate	0.00001
Epoch	100/300
Batch size	64
Loss function	MSE Loss
Sequence Length	96
Output_window	1

종목별 StandardScaler 적용 확인

StandardScaler 적용 전 →

[A.csv] Feature means (before):

[37.7684 37.7808 38.1302

37.4034 3647757.6631 37.7707]

[A.csv] Feature stds (before):

[11.848 11.8392 11.8793

11.8104 2344796.1332 11.842]



StandardScaler 적용 후 →

[A.csv] Feature means (after):

[-0. -0. -0. 0. -0. -0.]

[A.csv] Feature stds (after):

[1. 1. 1. 1. 1. 1.]

StandardScaler를 통해 데이터의 feature 값들을
평균이 0, 표준편차가 1이 되도록 변환합니다.

A.csv 종목으로 스케일링 적용 전, 후의 평균과
표준편차 결과입니다.

Train, Valid, Test 각각의 종목 개수 확인

출력 결과 →

Train count: 514

Val count: 514

Test count: 514

only val: set()

only test: set()

Ticker 포함/미포함 종목 개수 비교

Train: 514개

Valid: 514개

Test: 514개

Only val: 없음

Only test: 없음

Only val과 Only test는 train에 없는 종목이 val 또는 test에 들어가는 것을 체크한 것이며, Ticker 추가와는 관계없이 종목 개수는 동일합니다.

Train, Valid, Test 날짜 확인, 샘플 수 확인

출력 결과 →

Train date range: 2009-12-31 ~ 2018-07-09

Valid date range: 2018-07-09 ~ 2021-03-24

Test date range: 2021-03-24 ~ 2023-12-11

Total samples before split: 3414

0 train: 2048

0 valid: 682

0 test: 684

첫 번째 종목인 A.csv 파일의 Train, Valid, Test

각각 설정된 날짜 출력과 6:2:2로 나누기 전
전체 샘플 수와 나눈 뒤의 샘플 수 출력입니다.

제외된 종목명 및 개수 확인 / Train 기간 이후 언제 상장됐는지 확인

출력 결과 →

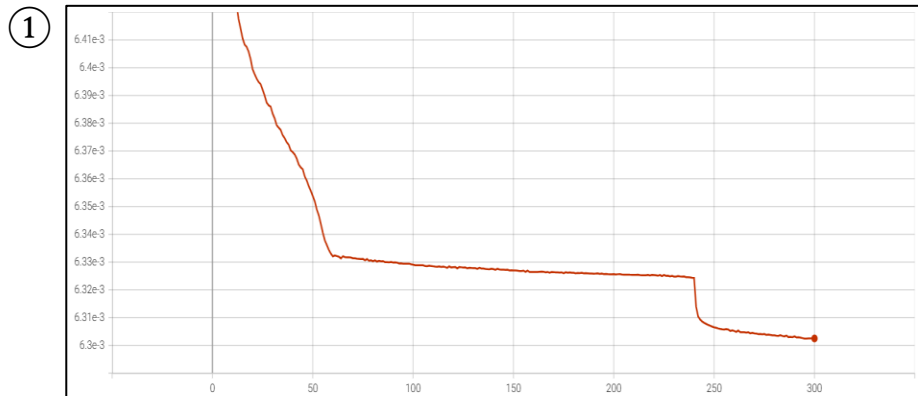
ABBV.csv: first trade, 2013-01-02 (after train date stock
ABNB.csv: first trade, 2020-12-10 (after train date stock
ALLE.csv: first trade, 2013-11-18 (after train date stock
AMCR.csv: first trade, 2012-05-15 (after train date stock
ANET.csv: first trade, 2014-06-06 (after train date stock
APTV.csv: first trade, 2011-11-17 (after train date stock
...
ZTS.csv: first trade, 2013-02-01 (after train date stock

제외된 종목 개수 – 74개

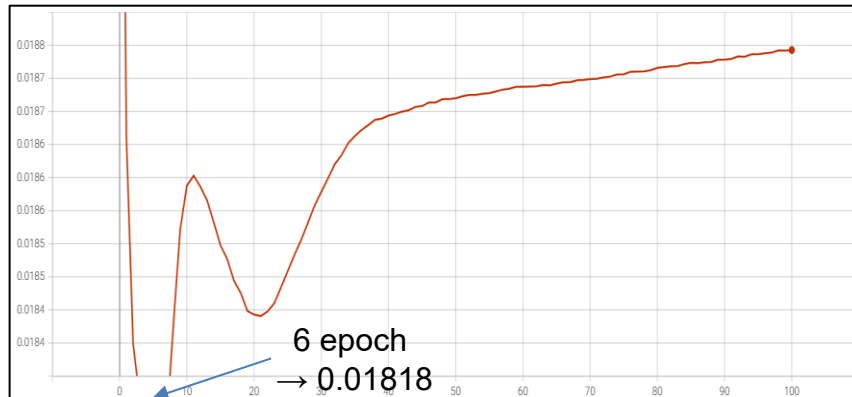
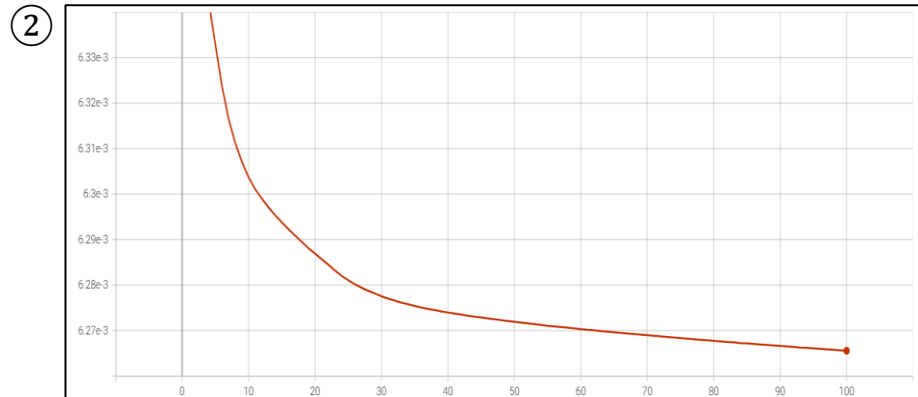
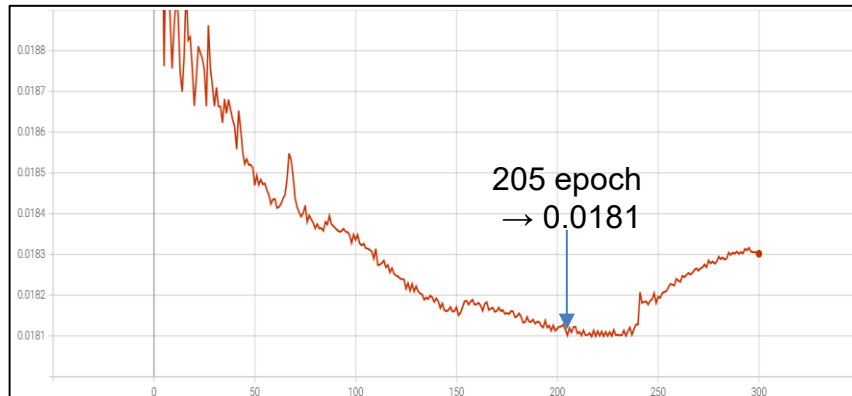
앞에서 지정한 데이터셋 2009-12-31~2023-12-31
조건을 충족하지 못하여 제외 되었으며,
Train 기간 이후 상장된 날짜 확인은 first trade 뒤의
날짜를 통해 확인합니다.

Ticker 포함/미포함 실험 결과

train loss



validation loss

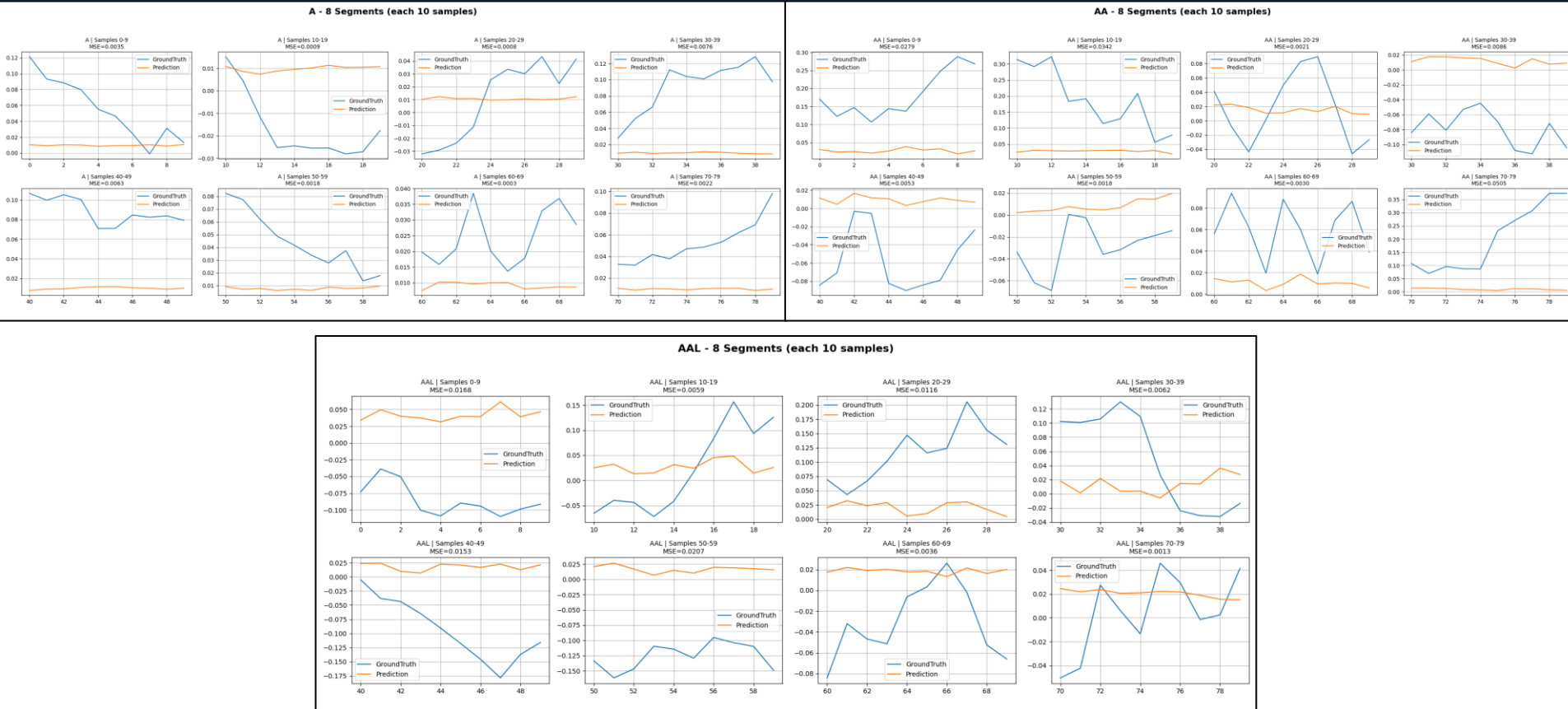


Test Loss ①	Test Loss ②
0.437039	0.437610

x: epoch y: loss

① → Ticker 포함
② → Ticker 미포함

Ticker 포함 샘플 시각화 결과



x: 각 샘플의 index y: 해당 샘플의 값
종목 3가지 (A.csv, AA.csv, AAL.csv)

Ticker 미포함 샘플 시각화 결과



x: 각 샘플의 index y: 해당 샘플의 값
종목 3가지 (A.csv, AA.csv, AAL.csv)

실험 결과 정리

- Ticker feature 포함/미포함 실험 결과
 - 두 결과의 Train 러닝커브는 잘 감소하지만, Ticker 미포함 실험 결과의 Valid Loss는 계속해서 증가하였습니다. 두 결과 모두 비슷한 Loss들이 나오지만 일반화 성능이 더 좋은 Ticker feature를 포함한 상태를 기준으로 앞으로 테스트할 예정입니다.
- 샘플 시각화 결과
 - Ticker feature 포함과 Ticker feature 미포함 결과, GroundTruth와 Prediction은 모두 동일하게 나왔습니다.

Ticker feature	Train Loss	Valid Loss	Test Loss	학습 소요 시간
미포함	0.006266	0.018793	0.437123	1시간 43분 42초
포함	0.006302	0.018302	0.437039	4시간 36분 48초

이후 계획

- 학습률, 하이퍼파라미터 변경 후 학습
- 모델 구조 변경 후 학습