



# High Frequency Trading survey

Tae-Won Lee,  
leetee95@gmail.com  
2021. 10. 15.

# Introduction

---

- High Frequency Trading(HFT)
  - HFT(고빈도 트레이딩)는 특수한 컴퓨팅 장비를 이용하여 분,초 단위의 거래로 수익을 내는 트레이딩 방식
  - 한국의 경우 미국(0.1%) 등에 비해 높은 증권 거래세(0.25%)로 인해 HFT가 불가능한 국가로 여겨짐
  - 2023년 증권 거래세 개편(0.25% -> 0.15%)으로 HFT가 가능해질 것으로 예상.

# Related works

- High frequency trading with deep learning

| 학회 명                                               | IF    | 논문 명                                                                                                   | 발간 년도 | 인용 수 | 내용                                                                                                                                          |
|----------------------------------------------------|-------|--------------------------------------------------------------------------------------------------------|-------|------|---------------------------------------------------------------------------------------------------------------------------------------------|
| ACM Transactions on Management Information Systems | 2.312 | A Double-Layer Neural Network Framework for High-Frequency Forecasting                                 | 2017  | 38   | 1) Dnn 을 High-Frequency data에 적용함<br>2) High frequency data에서 <b>종목간의 상관성</b> 을 발견, 특정 주식을 예측하는데 다른 주식 데이터 이용<br>3) 주가 예측에 도움이 되는 기술적 지표 이용 |
| Journal of Forecasting                             | 1.57  | An ensemble of LSTM neural networks for high-frequency stock market classification                     | 2018  | 64   | 1) <b>기술적지표</b> 를 사용한 lstm ensemble model 이용하여 분류                                                                                           |
| Quantative finance                                 | 1.491 | Encoding of high-frequency order information and prediction of short-term stock price by deep Learning | 2019  | 25   | 1) <b>호가창의 정보를 이용하기위해 CNN을 이용하여 feature 탐색</b> 하여 neural network의 input으로 이용<br>2) 다음 주가의 상승/하락을 분류함                                        |
| Journal of Computational Science                   | 2.644 | Sequence classification of the limit order book using recurrent neural networks                        | 2018  | 63   | 1) 호가창 데이터를 이용해 sequence의 등락을 분류<br>2) <b>RNN을 사용하여 호가창 데이터의 시공간적 표현을 학습</b>                                                                |
| Mathematical Problems in Engineering               | 1.009 | Stock Market Prediction on High-Frequency Data Using Generative Adversarial Nets                       | 2018  | 98   | 1) Generative Adversarial Nets을 이용하여 1분봉의 종가 예측                                                                                             |

# Related works

- High frequency trading with deep learning

| 학회 명                                     | IF   | 논문 명                                                                                                                               | 발간 년도 | 인용 수 | 내용                                                            |
|------------------------------------------|------|------------------------------------------------------------------------------------------------------------------------------------|-------|------|---------------------------------------------------------------|
| journal of risk and financial management | 0.44 | Trend Prediction Classification for High Frequency Bitcoin Time Series with Deep Learning                                          | 2019  | 31   | 1) 분 단위 데이터의 가상화폐의 상승과 하락을 분류하는 모델                            |
| journal of risk and financial management | 0.44 | Deep LSTM with Reinforcement Learning Layer for Financial Trend Prediction in FX High Frequency Trading Systems                    | 2019  | 21   | 1) Lstm과 강화학습을 이용한 hft 시스템 구축<br>2) 환율데이터(fx)를 이용             |
| IEEE Transactions on Signal Processing   | 5.23 | DeepLOB: Deep Convolutional Neural Networks for Limit Order Books                                                                  | 2019  | 79   | Limit Order Books (호가창 데이터)의 공간적 구조 데이터 정보를 학습하기 위한 CNN모델 개발  |
| Journal of Risk and Financial Management | 0.44 | Trend prediction classification for high frequency bitcoin time series with deep learning                                          | 2019  | 32   | 30분간격 비트코인 데이터를 이용해 비트코인 방향분류                                 |
| Statistical Finance                      | 2.25 | CLVSA: A convolutional LSTM based variational sequence-to-sequence model with attention for predicting trends of financial markets | 2021  | 20   | LSTM Cell 내부에 CNN 구조를 결합한 모델로 금융 시장을 예측                       |
| Expert Systems with Applications         | 5.45 | Convolution on neural networks for high-frequency trend prediction of cryptocurrency exchange rates using technical indicators     | 2020  | 33   | CNN, CNN-LSTM, MLP와 비교하여 기술적 지표 데이터를 이용한 제안된 CNN 추세예측 알고리즘 비교 |

# Related works

- Dataset

| 학회 명                                              | IF    | 논문 명                                                                                                            | 발간 년도 | 인용 수 | 사용 데이터                                                        | 비고                                                                                                                                               |
|---------------------------------------------------|-------|-----------------------------------------------------------------------------------------------------------------|-------|------|---------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------|
| journal of risk and financial management          | 0.44  | Deep LSTM with Reinforcement Learning Layer for Financial Trend Prediction in FX High Frequency Trading Systems | 2019  | 21   | EUR/USD, GBP/USD, EUR/GBP                                     | 1분 간격                                                                                                                                            |
| Journal of Computational Science                  | 2.64  | Stock price prediction using support vector regression on daily and up to the minute prices                     | 2018  | 118  | 브라질 주식 데이터                                                    | 1분간격                                                                                                                                             |
| International Conference on Intelligent Computing | 0.82  | High-Frequency Trading Strategy Based on Deep Neural Networks                                                   | 2016  | 67   | TAQ database of the NYSE. AAPL tick-by-tick transactions data | next one-minute average price. 2008/09/02~2008/11/07<br>TAQ database는 The Wharton School of the University of Pennsylvania과 협정이 되어있는 학교가 다운로드 가능 |
| IEEE Conference on Business Informatics           | 1.16  | Forecasting Stock Prices from the Limit Order Book Using Convolutional Neural Networks                          | 2017  | 184  | 5개 주식 종목의 <b>limit order book(호가창)</b> 데이터                    | 10일간의 데이터. the total number of messages is 4.5 million with equally many separate depth (450만개의 messages)                                        |
| Journal of Computational Science                  |       | Sequence classification of the limit order book using recurrent neural networks                                 | 2018  | 66   | <b>limit order book(호가창)</b> 데이터                              | E-mini S&P 500 데이터 이용. E-mini S&P 500은 S&P500주가지수 계약단위를 1/5사이즈로 축소하여 전산으로 거래 가능하도록 한 선물                                                          |
| Quantitative Finance                              | 1.491 | financial Trading as a Game: A Deep Reinforcement Learning Approach                                             | 2018  | 24   | tick-by-tick forex data. 12 currency pairs                    | resample the data into 15-minute. 15분 간격 예측                                                                                                      |

# Related works

- Dataset

| 학회 명                                     | IF    | 논문 명                                                                                      | 발간 년도 | 인용 수 | 사용 데이터                                    | 비고                                                                                                                                                                                                                                                     |
|------------------------------------------|-------|-------------------------------------------------------------------------------------------|-------|------|-------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| journal of risk and financial management | 1.557 | Trend Prediction Classification for High Frequency Bitcoin Time Series with Deep Learning | 2019  | 21   | BTCUSD, BTCCNY                            | <b>30분 간격</b> 데이터                                                                                                                                                                                                                                      |
| Computational Finance                    | 1.34  | Agent Inspired Trading Using Recurrent Reinforcement Learning and LSTM Neural Networks    | 2017  | 33   | USD/GBD                                   | We take 2000 data points from US Dollar/British Pound price series, we used 1000 <b>30 minutes interval</b> price points from 1/6/17 to 2/3/17 as training set and again take 1000 30 minutes interval price points from 2/3/17 to 3/6/17 as test set. |
| Acm transactions on information systems  | 2.889 | A Double-Layer Neural Network Framework for High-Frequency Forecasting                    | 2017  | 39   | S&P 500 (January 1, 2013 to May 31, 2013) | <b>5-minute interval</b> , large dataset that includes 100 stocks with the largest capitals in the S&P 500.                                                                                                                                            |
| Journal of Forecasting                   | 1.57  | An ensemble of LSTM neural networks for high-frequency stock market classification        | 2018  | 64   | 22 large cap US stocks (대형주)              | We use 1 year of rawtrade data which were cleaned and aggregated into <b>5-minute intervals</b> , amounting to roughly <b>19,000 observations</b> per stock.                                                                                           |

# Related works

---

- Dataset for High frequency trading
  - Limit order book (호가창) 데이터와 1분, 5분, 15분 30분 등이 High frequency data로써 사용됨을 알 수 있음
  - Limit order book (호가창) 데이터를 사용하는 경우 CNN, RNN과 같은 모델을 사용해 데이터의 시공간적 특성을 Encoding 하여 Encoding된 정보로 주가를 예측하는 연구가 이루어지고 있음
  - Limit order book (호가창) 데이터를 이용할 경우 증권사 api를 이용하면 초당 5~10회 정도의 데이터 제공
  - 개별 주식, (환율, 지수, 현물) 등을 이용한 선물(future) 데이터 등을 수집하여 이용가능

# Related works

- High frequency trading 평가 방식

| 학회 명                                              | IF    | 논문 명                                                                             | 발간 년도 | 인용 수 | 비고                                                                                                                                                                                    |
|---------------------------------------------------|-------|----------------------------------------------------------------------------------|-------|------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| International Conference on Intelligent Computing | 0.82  | High-Frequency Trading Strategy Based on Deep Neural Networks                    | 2016  | 65   | DNN 모델의 performance를 확인하기 위해 MSE(regression)와 Prediction Accuracy (classification(up/down))측정.<br><br>또한, test data에 대한 trading에 대한 평가로써 return(수익률) 사용<br>수익률과 누적수익률을 평가 metric으로 사용 |
| The Journal of Financial Data Science             | 0     | Deep Reinforcement Learning for Trading                                          | 2020  | 47   |                                                                                                                                                                                       |
| Mathematical Problems in Engineering              | 1.009 | Stock Market Prediction on High-Frequency Data Using Generative Adversarial Nets | 2018  | 98   | 예측 정도를 나타낼 수 있는 RMSRE와 방향을 예측했는지에 대한 Direction Prediction Accuracy (DPA)를 사용                                                                                                          |

- High frequency **Trading** 모델을 목적으로 삼는다면, return(수익률)과 누적수익률을 평가방식으로 사용 가능
- High frequency data **prediction**을 목적으로 삼는다면, MSE 계열과 Classification accuracy 등을 사용 가능



# Related works

- Limit Order Book datasets

- 국내 증권사 계정 API로 실시간으로 제공되는 해외주식의호가 정보가 존재하지 않음
- 국내 증권사 API를 이용하여 국내주식 실시간호가, 해외 선물 실시간호가 수집 가능
- 해외선물의 경우 증권사에 추가 사용 요청 후 이용가능

- 진행상황

- 요청 시 실시간 정보를 불러오는 과정까지 진행
- 연속적으로 실행하여 데이터화 가능
- 증권사 승인 시 해외선물 호가데이터 까지 수집 가능

|        |                     |
|--------|---------------------|
| 10,797 | 94,900              |
| 7,726  | 94,800              |
| 9,532  | 94,700              |
| 5,072  | 94,600              |
| 8,758  | 94,500              |
| 5,834  | 94,400              |
| 15,106 | 94,300              |
| 12,967 | 94,200              |
| 10,220 | 94,100              |
| 8,110  | 94,000              |
|        | 94,000 -1700 -1.78% |
|        | 93,900 20,270       |
|        | 93,800 16,259       |
|        | 93,700 26,003       |
|        | 93,600 75,439       |
|        | 93,500 110,892      |
|        | 93,400 18,872       |
|        | 93,300 28,078       |
|        | 93,200 25,809       |
|        | 93,100 30,516       |
|        | 93,000 70,866       |

```
array([[92300, 92400, 92500, 92600, 92700, 92800, 92900, 93000, 93100,
        93200],
       [92200, 92100, 92000, 91900, 91800, 91700, 91600, 91500, 91400,
        91300],
       [ 5437,  6699,  2830,  2609,  4152, 11484,  5217,  7331,  1254,
         1195],
       [ 2269,  4688, 17622, 13294, 12763,  6713,  5863, 10373,  5876,
        7610]])
```

# Related works

---

- 진행상황
  - 국내 증권사 계정 API로 실시간으로 제공되는 해외주식의 호가 정보가 존재하지 않음
  - 해외 주식 호가 데이터
  - 국내 증권사 API를 이용하여 국내주식 실시간 호가, 해외 선물 실시간 호가 수집 가능

# Related works

---

- 데이터셋 수집 종료
  - 해외 증권사 계좌를 외국인이 가입하는 것은 불법
  - 국내 증권사 api로는 국내 주식 호가, 해외 선물 호가 데이터 이용가능.
  - 국내 증권사 호가 데이터의 경우 외부 요인에 의해 예측이 무의미한 경우가 많음 ex) 해외상황, 국내정책 등
  - 해외 선물 호가 데이터의 경우 분기별 평균 자산이 1000만원 이상일 경우 다음 분기부터 데이터 수집 가능
  - 해외 선물 호가데이터 구매시 한 종목당 약 130만원 으로 많은 비용이 소모되어 현실적인 어려움
  - 따라서 데이터셋 수집을 종료함

# Reference

---

- 1 Siyu Yao, Linkai Luo and Hong Peng. High-Frequency Stock Trend Forecast Using LSTM Model. [International Conference on Computer Science & Education \(ICCSE\)](https://doi.org/10.1109/ICCSE.2018.8468703). <https://doi.org/10.1109/ICCSE.2018.8468703>
- 2 [Svetlana Borovkova](#), [Ioannis Tsiamas](#). An ensemble of LSTM neural networks for high-frequency stock market classification. Journal of Forecasting (2019). <https://doi.org/10.1002/for.2585>
- 3 Takuya Shintate and Lukáš Pichl. Trend Prediction Classification for High Frequency Bitcoin Time Series with Deep Learning. Journal of Risk and Financial Management (2019). <https://doi.org/10.3390/jrfm12010017>