TradeAide: Stock Market Prediction

Daniel Rozenzaft, Yaroslava Shynkar, Jonathan Kelaty

Problem Description

- Use machine learning methods to predict stock movements and maximize investor profit margins
- Create a model to predict optimal times to buy and sell a stock
- Stock price prediction is challenging because the market is so unpredictable
  - If it wasn't, we would all be millionaires!
- Implement several algorithms (ANN, LSTM, RL, RF, SVM) to make extrema and compare results with those obtained in recent papers that describe the state of the art
- Incorporate news sentiment analysis as a feature
- **State of the Art:** *Nabapour et al.* and *Yang et al.* suggest that ANN, LSTM, and RL perform the best. Our hypothesis is that this will be our result, and that the news sentiment feature will improve model performance.
  - SVM and RF are also effective (Hiba Sadia *et al.)*.
  - We expect solid, but suboptimal results from these algorithms.

Approach

- **SVM** - Our SVM will predict whether the stock price will decrease or increase on a given day. Each "direction change" implies a local extreme: a point at which we will buy or sell.
- **Random forests** - Useful in this domain for preventing overfitting on one (type of) stock or time period.
- **LSTM/ANN** - LSTM/ANN can make decisions based on past information (previous stock prices), making it perfect for large-scale time series data analysis. We will train neural networks to predict the local minima and maxima in future prices.
- **NLP** - We will use NLP techniques to explore the relationship between relevant tweet sentiments and concurrent stock movements.
- **Reinforcement learning** is predicated on maximizing a reward function, which we will develop by applying our profit metric and use to measure the accuracy of our extrema predictions.

Novelty

Two ideas for the novel approach

- Weight daily stock price changes on average volatility
  - Anticipate major ebbs and flows
  - Avoid buying during short price falls (wait for bigger decrease)
  - Avoid selling during short price jumps (wait for bigger increase)
- User-specified parameters for acceptable risk tolerance and capital constraints (stretch goal)
  - Limit purchases and number of shares bought/held based on constraints
  - Goal: create **comprehensive portfolio recommendations** during a specified time period
    - Tell the users what (and how much) to buy and sell at what time

Roles

- Daniel - will work on building and training reinforcement learning agent and contribute to the development of NLP component
  - Focus on problem domain, industry-specific evaluation methods from *Yang et al.*
- Yaroslava - will build LSTM/ANN and contribute to the NLP
  - Focus on theoretical aspects in *Bing et al.*
- Jonathan - will build SVM and RF and compare results to *Hiba Sadia et al.*
  - Emphasis on coding and evaluation of sentiment analyzer, comparison to *Bing et al.*

Evaluation

- Our reward function is simple: *profit*
- Evaluate model fit using standard performance metrics (F1, ROC, etc.)
  - Compare to results in *Patel et al.* and *Nabipour et al.*
- Measure impact of sentiment analyzer and novel volatility weighting approach on model performance
- Observe if the models perform better on certain stocks
  - *Bing et al.* found that sentiment analysis performed best on media and tech stocks, but this conclusion was based on 2014 tweet data
    - This conclusion was based on 2014 tweet data
    - We would like to see if it holds in the present day

Timeline

Week 1: Look into relevant libraries, attempt to apply familiar algorithms (SVM, RF, etc.) onto data

Week 2: Initial version of ANN/LSTM (Y), RL (D); complete SVM/RF and gather Tweet data (J)

Week 3: Flesh out ANN, LSTM, and RL code; run preliminary Tweet analysis

Week 4: Finish ANN, LSTM, RL; collaborate to complete/integrate NLP feature

Week 5: Fix bugs, begin performance evaluation, explore novel volatility weighting and portfolio allocation strategies

Week 6–7: Complete performance evaluation, compare with references; complete final presentation on report

References

Bing, L., Chan, K. C. C., Ou, C. (2014). Public Sentiment Analysis in Twitter Data for Prediction of a Company's Stock Price Movements, presented at IEEE 11th International Conference on e-Business Engineering, Guangzhou, 2014. Guangzhou: IEEE Publishing. https://ieeexplore.ieee.org/document/6982085.

Hiba Sadia, K., Sharma, A., Paul, A., Padhi, S., Sanyal, S. (2019). Stock Market Prediction Using Machine Learning Algorithms. *International Journal of Engineering and Advanced Technology*, *8*(4), 25–31. https://www.ijeat.org/wp-content/uploads/papers/v8i4/D6321048419.pdf.

Nabipour, M., Nayyeri, P., Jabani, H., Shahab S., Mosavi, A. (2020). Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data; a Comparative Analysis. *IEEE Access*, *8*(1), 150199–150212. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9165760.

Patel, J., Shah, S., Thakkar, P., Kotecha, K. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Systems with Applications*, *42*(1), 259–268. https://www.sciencedirect.com/science/article/abs/pii/S0957417414004473

Yang, H., Liu, X., Zhong, S., Walid., A. (2020). Deep Reinforcement Learning for Automated Stock Trading: An Ensemble Strategy, presented at ACM International Conference on AI in Finance, New York, 2020. New York, NY: Association for Computing Machinery. Available at SSRN.