

Stock Prediction with Recurrent Models and Supplemental Information

CS2750

Brian Falkenstein

March 2020

1 Outline

Why someone may want to predict the future price of stocks is obvious: with an accurate way to predict the rise and fall of stock prices, one could net enormous and consistent profits. However, this is much easier said than done. Given the nature of a market, and its dependencies on a large number of factors, making predictions from previous quantitative performance alone is fruitless[1]. Other world events, such as elections, trade agreements, and even large-scale tragedies, as we are seeing now with the global pandemic, have massive impacts on stock prices. The current state of affairs in regards to global markets, as well as individual companies, must be taken into consideration when making a stock purchase or sale. That is to say, if one were to study only the stock prices of IBM prior to the Coronavirus outbreak, one would have no reason to suspect a drastic drop in price.

Applying machine learning to the problem of stock prediction is far from a new idea. Models exist which use SVM[2], ensemble models[3], and deep learning (including CNN and RNN models)[4]. [5] Provides a general review of recent methods. However, none of these attempt to use any supplemental information about other world events, such as news information.

2 Methods

I propose a method of including multi-modal information along with basic financial history information

to improve stock prediction. This can be done by during training and testing time, combining features from historical data and other technical indicators, with features gathered from large news sites (Forbes, CNN, etc.) which may have recent articles relating to the company in question. I am planning on using simpler feature methods at first, such as SVM for technical indicators and *Doc2Vec* for news information.

3 Evaluation

In order to maximize the probability of there being recent news articles about the companies I am studying, I will be using data on the SP 500 companies (<https://datahub.io/core/s-and-p-500-companies>), as these are the largest companies in the US. I plan on evaluating both the regression task (predicting future price) as well as the binary task of "buy" or "don't buy", which may be an easier task. Results can then be compared to previously mentioned methods to see if the added information aided in the task.

4 Timeline

I will plan on having the data pipeline (fetching news articles) ready by the end of March, have early, simpler models (SVM, Doc2Vec) in place by early April, and allow the rest of the semester to explore more advanced models (LSTM, CNN, etc.).

References

- [1] Stibel, Jeff *Why We Can't Predict Financial Markets*. Harvard Business Review. 2009;
- [2] Hegazy, Osman Soliman, Omar S. Abdul Salam, Mustafa. *A Machine Learning Model for Stock Market Prediction*. International Journal of Computer Science and Telecommunications. 2013; 4. 17-23.
- [3] Patel, Jigar Shah, Sahil Thakkar, Priyank Kotecha, K *Predicting stock market index using fusion of machine learning techniques*. Expert Systems with Applications. 2015; 42. 2162-2172
- [4] Hiransha M, Gopalakrishnan E.A., Vijay Krishna Menon, Soman K.P. *NSE Stock Market Prediction Using Deep-Learning Models* Procedia Computer Science, 132. 2018; pp. 1351-1362
- [5] Pahwa, Nirbhey Singh et al. *Stock Prediction using Machine Learning a Review Paper*. (2017).