



News-oriented Stock Price Trend Prediction

Zecheng Zhang, Xinwei He, Jiachen Ge
{zecheng,xhe17,j5ge}@stanford.edu,

Computer Science, Stanford University ICME, Stanford University

Motivation

For stock market traders, creating price prediction trend models on stock market is always very important. However, due to the market volatility, it is difficult to make the correct prediction solely based on historical stock data. Therefore in this project, we analyzed the impact of daily news on stock markets. Specifically, we explored the dynamics of stock price movements by incorporating the news data.

News and Stock Price

Stock price is closely associated with the market news. Either stock-specific or overall economy news may lead to different impacts on stock market. Thus, when new information comes into the market, it is important to quickly analyze its impact and make correct future prediction on the market.

Since for a fixed period of time, stock market prices are typically low-dimensional sequential values, it might not be suitable to apply static machine learning models directly. The daily news, containing rich semantic (as shown in Figure 1) and sentiment information, need to use fine-grained model and make more exploitation. In this project we modeled and analyzed the dynamics of stock markets by utilizing both daily news and stock prices in our joint model. With the power of deep learning on capturing dependencies, text information and time series, our model provides a promising approach.

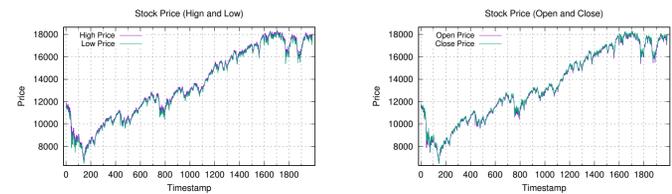


Figure 1: The wordcloud of the entities in the Reddit news dataset.

News Learning and Stock Price Learning

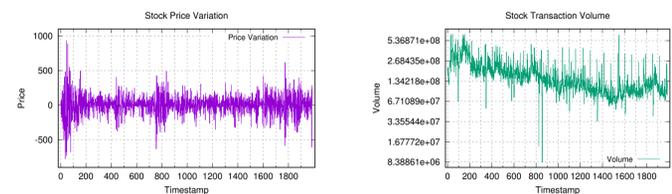
Both of news data and stock market data are crucial for stock price prediction.

The news text data contains rich information for our stock price prediction task. Specifically, the text representation, such as [2, 3], and sentiment of the news are very helpful. Both of them can reflect possible stock market change in different ways. But to capture the dependencies of news and its sentiment across different time, deep time series models are needed.



(a) DJIA Stock Price (High and Low). (b) DJIA Stock Price (Open and Close).

Figure 2: DJIA stock price changes (Open, Close, High and Low) from 2008-08-08 to 2016-07-01.



(a) DJIA Stock Price Variation. (b) DJIA Transaction Volume.

Figure 3: DJIA stock price variation and transaction volume changes from 2008-08-08 to 2016-07-01.

Stock market information contains useful data, shown in Figure 2 and 3. To learn from the it, we can use simple models such as ARIMA. It is formulated as follows: given a time series of data X_t , where t is an integer index and X_t are real numbers, an ARIMA(p,d,q) is given by

$$(1 - \sum_{i=1}^p \phi_i L^i)(1 - L)^d X_t = (1 + \sum_{i=1}^q \theta_i L^i) \epsilon_t \quad (1)$$

Notice that valuable dataset which has strong correlation with stock prices could not easily be fed into classical time series models. It is necessary to use deep learning based prediction models to resolve the problems.

Joint Learning with Time Series

Joint learning on news and stock data thus is a more comprehensive approach and provide a better prediction by incorporating both information directly. As discussed in the previous section, it is necessary to include timestamps information. In addition, we also want the model to capture following interesting patterns between news and stock:

- News affects stock market.
- Stock market has impact on News.
- Influences usually remain a period of time.

To capture aforementioned relationships with time information, the model need to have following capabilities:

- The model should identify which information is important.
- The model should capture the aforementioned patterns.
- The model should learn both data with time series information.

One popular model is Hybrid Attention Learning [1]. It utilizes attention and GRU, fulfilling the mentioned capabilities. We extend the model, shown in Figure 4, to further include sentiment and stock data. With attention and time series, our approach captures the patterns and achieve state-of-art result.

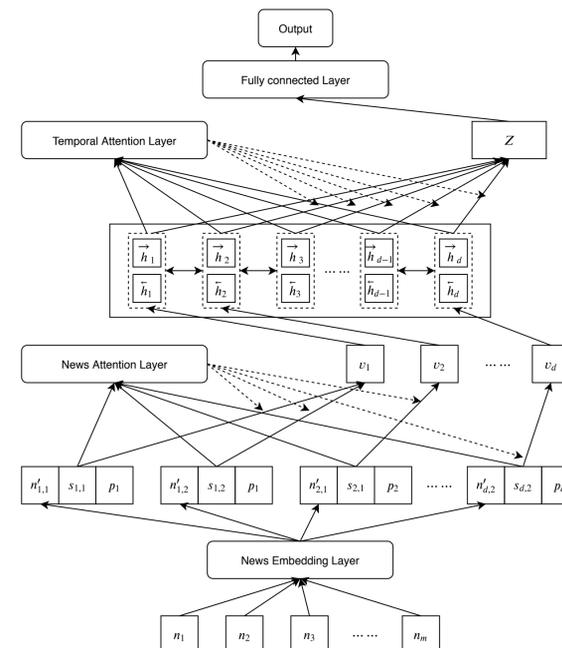


Figure 4: An overview of the framework.

Dataset

The dataset we are working on is a combination of Reddit news and the Dow Jones Industrial Average (DJIA) stock price from 2008 to 2016. The news dataset contains the top 25 news from Reddit on each day from 2008 to 2016. The DJIA contains the core stock market information for each day such as Open, Close, and Volume. The label of the dataset is whether the stock price is increase (labeled as 1) or decrease (labeled as 0) on that day. We perform thorough experiments and evaluation based on this dataset with selected baselines and our extension to the Hybrid Attention Learning.

Conclusion

Our model provides a promising approach for future work on stock market price trend prediction with richer news information. At first, We found that for news data both the news representation and sentiment play crucial roles. Hence it is necessary to combine them into one prediction model. However, for stock price data, simple time series model does not work. Thus, deep learning based model is required to learn through timestamps. At last, by experiments on joint learning with stock and news information and specific design of models, our approach give a state-of-art result with capturing interesting patterns.

References

- [1] Listening to chaotic whispers: A deep learning framework for news-oriented stock trend prediction. *CoRR*, abs/1712.02136, 2017. Withdrawn.
- [2] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119, 2013.
- [3] J. Pennington, R. Socher, and C. Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014.