News-oriented Stock Price Trend Prediction

Zecheng Zhang [*1] **Xinwei He** [*1] **Jiachen Ge** [*2]

Abstract

The prediction of stock prices in the market has always been an important task. However, due to the market volatility, it is difficult to make the correct prediction solely based on historical stock data. Therefore in this paper, with the analysis of daily news' impact on stock markets, we identify some key features that could be useful in stock price prediction and propose a deep learning model to capture the dynamics of stock price trend with rich news textual information.

1. Introduction

Stock price is closely associated with daily news. Positive news such as new acquisition opportunities, strong earnings reports and good management governance can increase the price of certain stocks. Negative news such as ineffective monetary policy, political uncertainty and natural disaster may negatively affect the whole stock market. Various news may lead to different reactions on stock market. Thus, when new information comes into public, it is important for investors to quickly identify its impact and make correct corresponding predictions on stock price.

Since for a fixed period of time, stock market prices are typically low-dimensional sequential values, it might not be suitable to apply many advanced machine learning models directly. However, textual data like daily news appears to be a suitable resource for us to exploit and it was verified by recent works (Lee et al., 2014) that news do play crucial roles in stock price prediction. Thus, it would be an interesting idea to explore in that direction.

In this report, we start by introducing related works in section 2. In section 3, we provide an overview of our dataset. In section 4, we first discuss NLP approaches to mine information from daily news, then we introduce and compare the performance of classical time series model. In the next step, we propose an extension of an existing deep learning based model (HAN) to combine both textual and sequential signals in dynamic stock price trend prediction. In section 5, we go over our experimental settings, baselines and experimental results. In Section 6, we give conclusion and discuss the future works.

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

2. Related Work

2.1. Text Representation and Sentiment Analysis

To learn the news information, NLP based methods are necessary. Recently, many text representation models have been reached state-of-art results, such as *word2vec*, *GloVe* and *BERT* (Mikolov et al., 2013; Pennington et al., 2014; Devlin et al., 2018). They generate representations of words or sentences in relatively low dimensional spaces. The representations thus can be directly applied to various downstream tasks. Similarly, the sentiment analysis, as discussed in (Zhang et al., 2018), representing sentence sentiment in numerical values provides richer text information for multiple downstream tasks.

2.2. Time Series Model

Time series prediction models have always played important roles in stock price prediction tasks. They incorporate time order into historical market data analysis, such as transaction prices, volumes, variations etc. The forecast of the future market movement is usually based on that. One of the most widely used model for time series forecasting is the ARIMA. ARIMA is a generalization of an autoregressive moving average model, which can be applied when data is non-stationary and where an initial differencing step (corresponding to the "integrated" part of the model) can eliminate the non-stationarity of data. However, due to the non-linear

[*]Equal contribution [1]Computer Science, Stanford University [2]ICME, Stanford University. Emails: Zecheng Zhang <zecheng@stanford.edu>, Xinwei He <xhe17@stanford.edu>, Jiachen Ge <j5ge@stanford.edu>.

and non-stationary nature of stock prices, classical models are not good enough to capture the underlying patterns. With the development of deep learning, more research efforts have been spent on utilizing deep learning models in financial market prediction [(Adebiyi et al., 2014), (Göçken et al., 2016), (jae Kim & Ahn, 2011)]. To further model the long-term dependency in time series, RNN approaches, such as LSTM and GRU, have also been employed in financial market prediction [(Akita et al., 2016), (Gao, 2016)].

3. Dataset and Features

In this section we introduce the dataset we are working on and some related features and patterns by case studies.

3.1. 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 on Reddit of each day from 2008 to 2016. The DJIA contains the core stock market information for each trading day such as Open, Close, and Volume. The label of the dataset represents whether the stock price is increase (labeled as 1) or decrease (labeled as 0) on that day. The total number of days in the dataset is 1989. During our preliminary experiments, we split the dataset into train, validation and test. The train dataset contains 1526 days, around 76% of the total dataset. The validation and test datasets contain 85 days and 378 days respectively.

3.2. An Overview of Stock Market Information

Stock data provides some important information, which reflects the market movement and helps with stock price prediction. Here, we give an overview of some important stock related features.

The Figure 2(a) and Figure 2(b) show the price changes of DJIA from August 8th 2008 to July 1st 2016 with four stock price attributes (Open, Close, High and Low). The Figure 3(a) and Figure 3(b) show the stock price changes and transactions volume which offer a wider perspective on stock market information.

3.3. News-oriented Stock Price Patterns

Positive news, such as expansionary monetary policy and geopolitical peace, will create an optimistic view in market, leading to an upward movement in stock price. However, news such as trade tension and geopolitical uncertainty may lead to a downward trend in stock market. Here, we are looking at several cases to illustrate the possible impacts of news on stock market. **Case 1:** On September 29th, 2008, DJIA had the largest daily loss of 777.68 in the close market data that may be caused by the top headlines on that day's news "Huge European bank and insurance giant fails". **Case 2:** On October 13th, 2008, DJIA had the largest daily gain of 936.42 in the close market data, which accompanied with news like "Europe puts 2.3 trillion dollars on line for banks". **Case 3:** There is a sequence of news related to certain hot issue may have impact on stock markets for several days. During June 24th,2010 to July 2nd, 2010, the stock markets had 7-day consecutive loss. As the daily news headlines at that period show that "May Toronto's G20 be the last".

The general patterns and relationships of cases are shown in Figure 4. From these observations, we will summarize what capabilities of the models are needed in order to learn these patterns well in Section 4.

4. Methods

We analyze the approaches by news learning and stock price learning first. Then we propose what should a model needed in order to capture more complete patterns and achieve better prediction results. At last, we will exhibit our own approaches.

4.1. News Learning

News data always contains rich information in it. In this subsection, we discuss the information that can be mined from daily news and the corresponding methods that might be useful.

4.1.1. News Representation

The word representation models help to change the textual data into mathematical representations, which can be applied to machine learning models directly in the classification tasks.

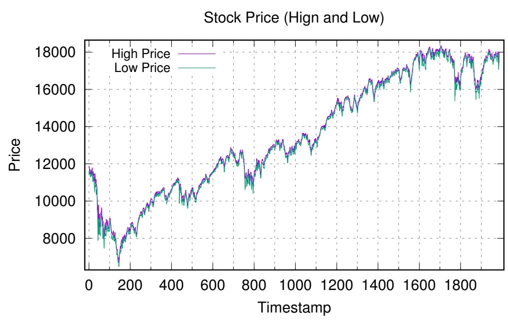
The most basic one is bag-of-words representation. Several more advanced word representation models include *word2vec* (Mikolov et al., 2013), which adopts skip-gram model idea to incorporate surrounding words in the objective function.

Another one is the *GloVe* (Pennington et al., 2014), which introduces a more general weighting function and makes the cost function becomes

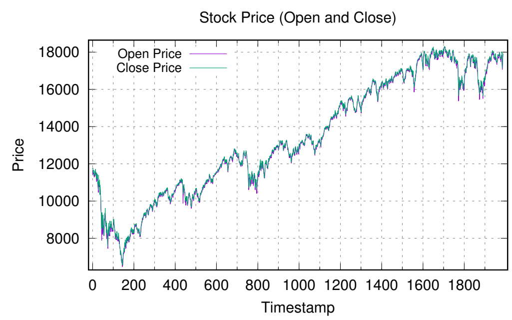
$$\hat{J} = \sum_{i,j} f(X_{ij})(w_i^T \tilde{w}_j - \log X_{ij})^2 \tag{1}$$

where this cost function leverages both the global representations and local context information.

Thus, by averaging the words' representations in a single piece of news, we can get representations of that news. This kind of representation is helpful to design machine learning based models because we can just feed the representations

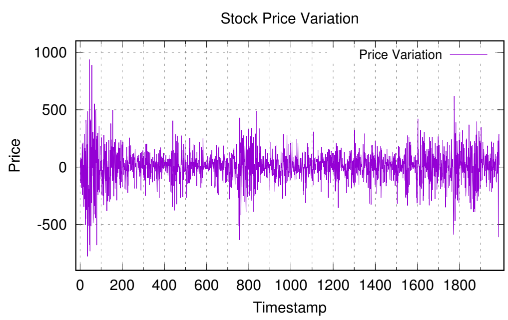


(a) DJIA Stock Price (High and Low).

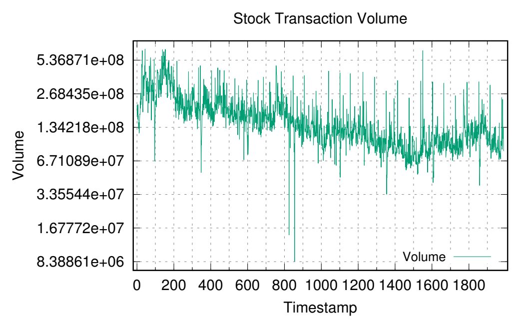


(b) DJIA Stock Price (Open and Close).

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.

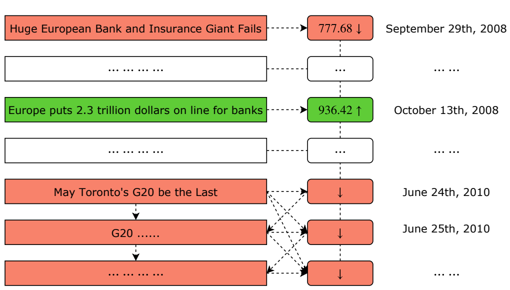


Figure 4. Examples of how news and stock price influence with each other and the impacts might remain for several days.

directly into the model.

4.1.2. Sentiment Analysis

Sentiment analysis is another prevailing technique that has been studied a lot in recent years. The sentiment of words or sentences can reveal peoples' opinions, emotions and attitudes behind the words (Zhang et al., 2018). These attributes play significant roles in the stock price predictions. For example, negative sentiment in a news may reflect some degree of possible negative atmosphere in stock market. The bad atmosphere would then cause possible decline in stock price.

4.1.3. News with Time Series Model

To learn news patterns or information better, time series model is necessary. Because it is possible that there are dependencies or associations among news across a range of timestamps. We will further introduce our selection of time series models in Section 5.

4.2. Stock Price Learning

In this section, we present some time series models that can learn useful information from the stock data.

4.2.1. Time Series Model

One approach is called ARIMA that 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

$$(1 - \sum_{i=1}^{p} \phi_i L^i)(1 - L)^d X_t = (1 + \sum_{i=1}^{q} \theta_i L^i)\epsilon_t \qquad (2)$$

Double Exponential Smoothing (DES): Another interesting model we tried out is double exponential smoothing model

$$s_t = \alpha x_t + (1+\alpha)(s_{t-1} + b_{t-1}) \tag{3}$$
$$b_t = \beta(s_t - s_{t-1}) + (1-\beta)b_{t-1} \tag{4}$$

where α is "data smoothing factor", where $0 < \alpha < 1$ and β is the "trend smoothing factor", where $0 < \beta < 1$.

4.2.2. Deep Learning Based Time Series Model

LSTM and GRU are the most widely used forms of Recurrent Neural Network Model (RNN). Compared to classical time series models like ARIMA, which only utilizes historical data within fixed length sliding window, they are able to capture richer and more complex long-term dependencies.

4.3. Joint Learning with Time Series

Analyzing the news and stock data together, we found there are several interesting patterns, as mentioned in Section 3. In this subsection, the requirements of the model in order to capture these patterns. Then, we introduce an existing framework for stock price trend prediction and extend that to capture richer information which gives a state-of-art result.

4.3.1. Model Capacities

Following is the list of capacities that required to capture richer information in news-oriented stock price trend prediction task. **(1)** The model should identify which news is important and which is not. **(2)** The model should capture the relationships between news and stock price. **(3)** The model should learn both stock and news data with time series information.

4.3.2. Hybrid Attention Learning

To learn the news-oriented stock price prediction well, the model should have three key capabilities which is listed previously. One popular approach contains these capabilities is Hybrid Attention Learning (HAN) (Hu et al., 2018). Following is the general structure of HAN.

News Embedding: The Hybrid Attention model first train a news representation n_i' for news n_i by using a word embedding layer. Then, for example, if on date t there are news n_i and n_j, then they will have corresponding news representations n_{ti}' and n_{tj}'.

News Attention Layer: Following the news embedding is the news attention learning. This layer basically learns which news is important to the stock price changes for each day. The attention for day t and news i, α_{ti}, is learned as follows:

$$\alpha_{ti} = \frac{\exp\left(\sigma((Wn_{ti}' + b))\right)}{\sum_{j=1}^{c} \exp\left(\sigma((Wn_{tj}' + b))\right)}; d_t = \sum_{i=1}^{c} \alpha_{ti} n_{ti}' \tag{5}$$

where c is the number of news for each day and σ is activation functions such as ReLU and Sigmoid. With appropriate attentions, we can thus generate the embedding for each day d_t.

Sequential Modeling: After learning the representation for each day as d_t, deep time series model, such as Gated Recurrent Units (GRU) can be used to capture news dependencies with time series. To capture both the previous and future dependencies, bi-directional GRU is used. The embedding generated by the bi-directional GRU as:

$$h_i = \left[\overrightarrow{GRU}(d_j), \overleftarrow{GRU}(d_k)\right] \tag{6}$$

where $j \in [1, L]$ and $k \in [L, 1]$. Therefore, the vector h_i incorporates the information from surrounding context and itself.

Temporal Attention Layer: This attention layer learns both the inherent temporal patterns and news information. The attention β_i and the final embedding Z are defined as:

$$\beta_i = \frac{\exp\left(\sigma((W_h h_i + b_h))\right)}{\sum_{j=1}^{c'} \exp\left(\sigma((W_h h_j + b_h))\right)}; Z = \sum_{i=1}^{c'} \beta_i h_i \tag{7}$$

Stock Price Trend Prediction: At last, we add a fully connected layer to train with the final embedding Z to predict whether the stock price will increase of decrease.

It is true that the Hybrid Attention Learning can learn the news importance and patterns by attention and GRU, it has several drawbacks: **(1)** The news embedding should not only consider the text contextual information, but also some other important information such as sentiment of news. **(2)** This model does not incorporate the information from the stock data.

4.3.3. Extend and Combine

From our analysis, the news sentiment is crucial for news comprehension. Also, stock market information itself plays roles in stock price trend prediction. Hence, we first extend the HAN to capture the news sentiment scores and then combine the attention learning with the stock information which will capture important interaction patterns.

To make the attention learning more concise, we inject the sentiment score s and stock information p after the news embedding layer by making

$$n_{ti}'' = [n_{ti}', s_{ti}, p_t] \tag{8}$$

where n_{ti}'' is the input to the next layer.

With this design, shown in Figure 5, we can learn the importance for both the richer news information and the stock data. The following GRU and temporal attention layers will therefore capture interaction dependencies among the news and the stock prices within the time series context.

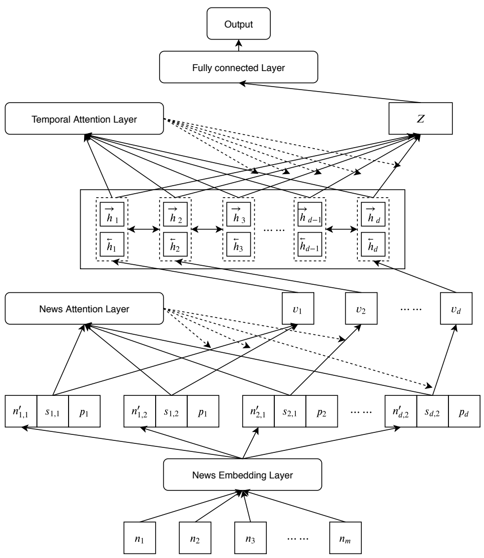


Figure 5. An overview of the framework.

5. Experiments

We conduct thorough experiments on our approaches and compare the results with multiple baselines. Our method eventually outperforms all other methods. The code is available at https://github.com/RagnaroWA/news_oriented_stock.

5.1. Experimental Setting

5.1.1. Baselines

We perform various NLP approaches on the news to predict the stock price trend. We utilize the word representation and sentiment analysis and feed those representations into the downstream classification models. We mainly use the logistic regression and Naive Bayes models as the baseline downstream classification models. For convenience, we first we denote Bag-of-words as BOW, Logistic Regression as LR, sentiment as SEN, and Naive Bayes as NB. For time series DES method, we add the differencing approach which is denoted as DES(w/ difference).

5.1.2. Implementation Details

We set the news embedding size to 100 by using the *GloVe* representation learning and get the sentiment score for each news through *Google NLP API*. For deep learning model training, we utilize the Adam optimizer.

5.2. Evaluation and Discussion

All of the experiment results are shown in Table 1. Our approach Extended HAN achieves the highest accuracy and outperforms other methods a lot. From the experiment results, we have several findings: **(1)** The sentiment of news is helpful for predicting the trend of the stock price. **(2)** Simple models cannot fully learn the complex patterns from both stock and news data. Both the simple classification and time series models do not give good results, which means that deep learning models are required. **(3)** Our extended model outperforms the original HAN model by more than 3% of accuracy. This potentially means that our model captures more complex patterns among the news data and stock data by importance attention in time series context.

Models	Accuracy%	Macro-F1%
ARIMA	46.03	46.00
DES	46.15	46.12
DES(w/ difference)	46.94	46.94
BOW + LR	44.71	43.60
BOW + SEN + LR	42.86	49.53
GloVe + LR	49.21	46.15
GloVe + SEN + LR	53.17	50.68
BOW + NB	50.53	49.12
SEN + LR	48.68	44.82
HAN	53.17	54.22
Extended HAN	**56.88**	**59.95**

Table 1. Experiment results.

6. Conclusion and Future Work

In this project, we first explored multiple ways to do the prediction on stock price trend by news and stock price separately. Due to the limitation of direct news-based prediction and non-stationary features of the stock data, we need a more advanced model to combine news and stock data which can provide a more complete and accurate way to capture richer relationships and dependencies. With analysis on required capacities of models in the trend prediction task, we extended the HAN approach to learn more hidden and complex patterns and achieved state-of-the-art results. However, it is still promising if we can incorporate richer information from the news, such as adding the phrase embedding, which will be remained as the future work.

7. Contribution

Zecheng Zhang: dataset preprocessing, NLP related experiments, Extended HAN model construction and training, and report writing.

Xinwei He: dataset preprocessing, Time Series related experiments, Helped with HAN model construction, and report writing.

Jiachen Ge: dataset preprocessing, baselines models, cases study, and report writing.

References

Adebiyi, A. A., Adewumi, A. O., and Ayo, C. K. Comparison of arima and artificial neural networks models for stock price prediction. *J. Applied Mathematics*, 2014: 614342:1–614342:7, 2014.

Akita, R., Yoshihara, A., Matsubara, T., and Uehara, K. Deep learning for stock prediction using numerical and textual information. *2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)*, pp. 1–6, 2016.

Devlin, J., Chang, M., Lee, K., and Toutanova, K. BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805, 2018. URL http://arxiv.org/abs/1810.04805.

Gao, Q. Stock market forecasting using recurrent neural network. 2016.

Göçken, M., Özçalici, M., Boru, A., and Dosdogru, A. T. Integrating metaheuristics and artificial neural networks for improved stock price prediction. *Expert Syst. Appl.*, 44:320–331, 2016.

Hu, Z., Liu, W., Bian, J., Liu, X., and Liu, T.-Y. Listening to chaotic whispers: A deep learning framework for news-oriented stock trend prediction. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pp. 261–269. ACM, 2018.

jae Kim, K. and Ahn, H. Simultaneous optimization of artificial neural networks for financial forecasting. *Applied Intelligence*, 36:887–898, 2011.

Lee, H., Surdeanu, M., MacCartney, B., and Jurafsky, D. On the importance of text analysis for stock price prediction. In *LREC*, 2014.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pp. 3111–3119, 2013.

Pennington, J., Socher, R., and Manning, C. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1532–1543, 2014.

Zhang, L., Wang, S., and Liu, B. Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4): e1253, 2018.