# Research on the Correlation between Hot News and Stock Index Volatility Based on Deep Learning

Yuting Zhang [1] and Zhaogong Zhang [2+]

[1, 2] School of Computer and Technology, Heilongjiang University, China

**Abstract.** With the development of technology, the representation of text semantics by computer is also improving day by day, and the application scope is more and more extensive. The correlation between news and market volatility has become one of the hot topics in recent years. Stock market fluctuation is affected by many factors, and news is an important channel of influence factors. In this paper, a neural network structure named 'BLA' is designed to study the correlation between daily hot news and DJIA fluctuations, to judge the rise and fall of the index of the day. Experiments show that the accuracy of the 'BLA' network in analyzing the fluctuations of hot news and stock market index is higher than that of the traditional network.

**Keywords:** news, stock market index, neural network, attention.

## 1. Introduction

At present, most researches on the correlation between news and stock prices focus on the study of fixed time nodes, such as taking the closing price of the predicted target date. In the research history of financial markets, Eugene Fama [1] proposed the efficient market hypothesis. The theory holds that in a strong efficient market, the price of a security fully reflects all information, both public and private. In this case, the investment cannot obtain excess return. However, in the real market, due to the delay of news transmission, people's expectations and irrationality, asset prices usually cannot express their intrinsic value completely and quickly, which provides a certain basis for the study of stock volatility. As an important medium of news dissemination, news plays an important role in the study of stock index volatility.

The extent to which external information affects asset prices can be traced back to 1994. Mitchell used the size of the New York Times headline to indicate the importance of news and found that the number of Dow Jones announcements is directly related to market activity. How the various media influence the stock market, how to link this effect with the stock market price, is still full of difficulties. But with the development of technology, we can already see many exciting results [2-3]. Mittermayer, MA (2006) and Knolmayer, G F [4] released an automatic text classification prototype that goes beyond the precision of stock analysis in the past. Robert P. Schumaker (2009) and Hsinchun Chen [5] released a quantitative stock forecasting system based on financial news. By collecting financial articles, the article is processed using the word bag method in natural language processing, and the stock price of the next 20 minutes is estimated by using the SVM to measure the stocks by region. The 20-minute selection is based on the previous work of Gidofalvi [6]. The team also believes that in addition to the views of the article, the emotions it exhibits will also affect the price of the stock. On the previous experimental results, emotional indicators were added to cluster the sentiment of the article, and the conclusion that the subjective article is easier to predict the price direction is obtained [7]. With the development of social media, people think that the emotions they express have a great influence on the stock market. The results of Johan Bollen et al. [8] show that collecting data from Twitter for processing, including specific public sentiment dimensions, can significantly improve the

---

[+] Corresponding author. Tel.: + 86 13836170316

*E-mail address*: zhaogong.zhang@qq.com

accuracy of predictions. Ryo Akita [9] and others used CNN to model the impact of multiple company news on opening prices and achieved good forecasts on the Tokyo Stock Exchange. Ghazanfar, Mustansar Ali et al. [10] confirmed the effectiveness of multiple machine learning methods in stock market analysis. Xiao Ding et al. [11] took into account the influence of past news, and used deep learning to comprehensively consider past long-term and short-term news, and obtained good prediction accuracy.

This article uses daily hot news as the research basis. Based on the rapid development of pre-training technology, a newer Bert method with attention mechanism and recurrent neural network is used to try to analyze the correlation between non-financial news and stock market fluctuation.

## 2. Methodology

### 2.1. Attentional Mechanism

The Attention mechanism is a way to solve problems by imitating human Attention, and its core lies in screening out high-value information. Generally, it is used to solve the problem that it is difficult to obtain the final reasonable representation of RNN model when the sequence is long. Different from the regular encoder-decoder model, it assigns different weights to the vector that comes into the decoder. To get better results. Our formula for calculating vector weight c is:

$$c_i = \sum_{j=1}^{T_x} a_{ij} h_j \tag{1}$$

Where, $h_j$ is the higher value of hidden state $a_{ij}$ of encoder, which means that the i output will allocate more attention to the j input, and will be more affected by the j input when generating the i output. The calculation formula is as follows:

$$a_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{T_x} \exp(e_{ik})} \tag{2}$$

Where $e_{ij}$ is the matching of the hidden state at encoder $h_j$ and the hidden state at decoder $S_{j-i}$. The calculation formula is as follows:

$$e_{ij} = a(S_{j-i} + h_j) \tag{3}$$

### 2.2. News Embedding

To convert news into usable and training vectors, we first need to perform Embedding processing. We use Bert for this task. BERT is an advanced preprocessing model that can generate deep biphasic language representations and get better results than traditional methods. We use the Bert Large model, which contains 24 multi-head self-attention layers with 16 heads in each layer. The incoming news will be converted into a vector with a dimension of 1024.

### 2.3. Bi-LSTM

Bi-LSTM is a combination of forward LSTM and backward LSTM. LSTM models can be used to analyse long-range dependencies. However, modelling sentences using LSTMs cannot encode back-to-front information. Bi-LSTM can better capture bidirectional semantic dependencies. We can get the vector group of forward propagation $\{hf_1, hf_2 \dots, hf_n\}$ and the vector group of backward propagation $\{hb_1, hb_2 \dots, hb_n\}$. In the end, we used $\{hf_i, hb_i\}$, which contains all the information of forward and backward.

### 2.4. BLA Network

Based on the above technologies, we have designed a new network architecture based on daily hot news. Embedding the daily hot news sections, and sequentially transmitting them to the Bi-LSTM network, assigning different weights through the attention mechanism to calculate the output of the recurrent neural network. And the results are transferred to the discriminant network to obtain the analysis results of the stock market's rise and fall on that day. Its network structure is shown in Fig. 1.
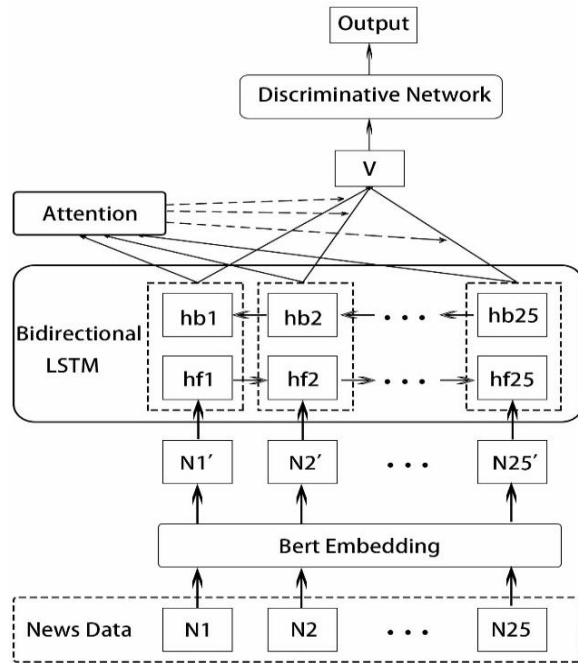
Fig. 1: BLA Network

# 3. Experiment

## 3.1. Data Sources

We used the Kaggle dataset [12]. This data set contains DJIA's index data from 2008 to 2016 and 25 hot news headlines for each day corresponding to the date, from the selection of Reddit World News Channel. The total contains data for 1989 days. Sample stock data and news data are as follows.

Table 1: Sample stock price data

| Date | Adj Close | Label |
|------|-----------|-------|
| **2016/7/1** | 17949.37 | 1 |
| **2016/6/30** | 17929.99 | 0 |
| **…** | … | … |
| **2008/8/8** | 11734.32 | 0 |

Table 2: News data sample

| Date | News |
|------|------|
| **2016/7/1** | IMF chief backs Athens as permanent Olympic host |
| **2016/7/1** | The president of France says if Brexit won, so can Donald Trump |
| **…** | **…** |

## 3.2. Experimental Design

We first compare the daily closing price in the stock index data set with the previous day, mark stock indexes up or down and transfer it to the model in the form of one-hot. The news data is divided into a training set and a test set at a ratio of 8: 2. The daily hot news is trained into a 1024-dimensional vector through Bert, making it a 1989 * 1024 * 25-dimensional matrix. Next, we transfer the 25 embedding results of each day to the Bi-LSTM in order. The results generated by Bi-LSTM will be processed through an additional attention mechanism to determine which news information we should pay more attention to. Next we pass the results to the discriminative neural network. Finally, the two-dimensional vector obtained through the softmax function was compared with the ups and downs to obtain the accuracy result. For this classification problem, we use cross entropy to represent the loss function.

### 3.3. Experimental Result

We tested the trained model multiple times. The average accuracy of the BLA model was 54.61%, of which the highest accuracy rate was 59.26%, which has improved the accuracy of the fluctuation analysis. In addition, we use the traditional Word2vec-LSTM and Bert (base) -LSTM networks for comparison. The specific accuracy is shown in Table 3.

Table 3: Accuracy rate

| Algorithm | Accuracy |
| --- | --- |
| Random | 50.08% |
| Word2vec-LSTM | 50.25% |
| Bert(base)-LSTM | 52.46% |
| **BLA(Bert Large)** | **54.61%** |

The experimental results of BLA are better than the general model combination. We plotted the ROC curve of this result, as shown in Fig. 2.
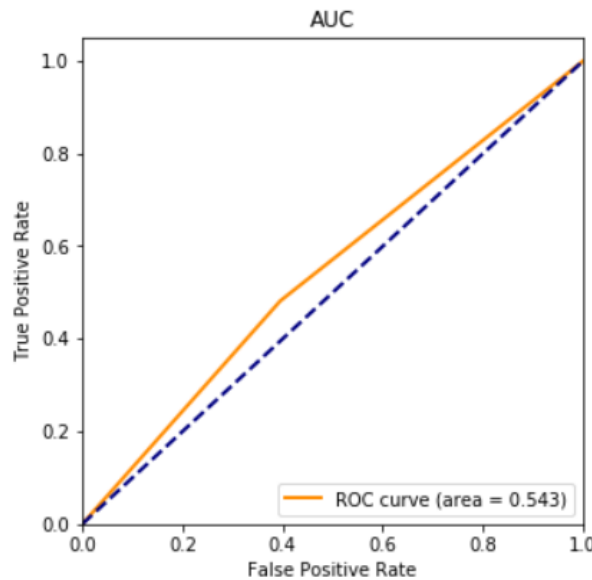


Fig. 2: ROC curve

## 4. Conclusion

This film proposes a new neural network architecture based on hot news analysis of stock volatility. The experiments show that the BLA network we designed to a certain extent verifies that there is a corresponding relationship between hot news and the stock market's rise and fall, and the accuracy of the analysis of the stock market's rise and fall is higher than traditional technology. Because the fluctuation of the stock market is determined by many factors, the influence of a single factor on the stock index is limited. And hot news also contains a lot of weak related information, its performance in accuracy is not dazzling. We will add sentiment analysis of the news and more stock market indicators in future work to further improve the accuracy of market analysis.

## 5. Acknowledgements

## 6. References

[1]   Malkiel, Burton G., and Eugene F. Fama. "Efficient capital markets: A review of theory and empirical work." *The journal of Finance* 25.2 (1970): 383-417.

[2]   Schumaker, Robert P. , and H. Chen . "Evaluating a news-aware quantitative trader: The effect of momentum and Contrarian stock selection strategies." *Journal of the American Society for Information Science and Technology* 59.2(2008):247-255.

[3]   Schumaker, Robert P. , and H. Chen . Textual analysis of stock market prediction using breaking financial news: The AZFin text system. *ACM*, 2009.

[4]   M.-A. Mittermayer, and G.F. Knolmayer. "NewsCATS: A news categorization and trading system." *Data Mining, 2006. ICDM '06. Sixth International Conference on* 2007.

[5]   Schumaker, Robert P. , and H. Chen . "A quantitative stock prediction system based on financial news." *Information Processing & Management* 45.5(2009):571-583.

[6]   Gidofalvi, Gyozo , and Győző Gidófalvi. "Using News Articles to Predict Stock Price Movements." *department of computer science & engineering* (2001).

[7]   Schumaker, Robert P. , et al. "Evaluating sentiment in financial news articles." *Decision support systems* 53.3(2012):p.458-464.

[8]   Bollen, Johan , H. Mao , and X. Zeng . "Twitter mood predicts the stock market." *Journal of Computational Science* 2.1(2011):1-8.

[9]   Akita, Ryo , et al. "Deep learning for stock prediction using numerical and textual information." *IEEE/ACIS International Conference on Computer & Information Science* IEEE, 2016.

[10]  Ghazanfar, Mustansar Ali, et al. "Using machine learning classifiers to predict stock exchange index." *International Journal of Machine Learning and Computing* 7.2 (2017): 24-29.

[11]  Ding, Xiao , et al. "Deep Learning for Event-Driven Stock Prediction." *International Conference on Artificial Intelligence* AAAI Press, 2015.

[12]  Sun, J. (2016, August). Daily News for Stock Market Prediction, Version 1. Retrieved [Date] *from https://www.kaggle.com/aaron7sun/stocknews*