An Attention-Based LSTM Model for Stock Price Trend Prediction Using Limit Order Books

Yunhao Li1, 2, Liuliu Li3, Xudong Zhao3, Tianyi Ma2, 3, Ying Zou3 and Ming Chen2, 3

1Harbin Institute of Technology, 92 West Dazhi Street, Harbin, China
2Zhejiang University, No. 38 Zheda Road, Hangzhou, China
3Hithink RoyalFlush Information Network Co., Ltd., Hangzhou, China
Email: crimsonjhu@gmail.com, {liliuliu, matianyi, zouying}@myhexin.com, sdu_zxd@163.com, chm@zju.edu.cn

Abstract. Stock price trend prediction has been a hot issue in the financial field, which has been paid much attention by both academia and industry. It is a challenging task due to the non-stationary and high volatility of the stock prices. Traditional methods for predicting stock price trends are mostly based on the historical OHLC (i.e., open, high, low, and close prices) data. However, it eliminates most of the trading information. To address this problem, in this paper, another type of stock price data, i.e., limit order books (LOBs), is used. For better exploring the relationship of the LOBs and stock price trend, inspired by the successful application of deep learning-based methods, an attention-based LSTM model is applied. The trend of stock price can be predicted by using the LOBs data of the previous day. By using the real stock price data of the China stock market, the effectiveness of the proposed model is validated by experimental results.

1. Introduction
Stock investors intend to explore and summarize the underlying trading patterns for predicting stock price trends to maximize obtained profit [1]. Therefore, it has become a hot topic to predict price trends in the financial field accurately. However, because the stock market is non-stationary and easily influenced by various uncertain factors, predicting stock price is challenging and difficult. With the accumulation of financial data and the update of data analysis technologies, more and more financial economists even data analysts focus on the study of various historical data, including macroeconomic data, the stock trading price and so on, in the hope of discovering patterns of stock price movements by building quantitative and prediction models.

Concerning stock price prediction, most of the existing works [2, 3] are based on the historical OHLC (i.e., open, high, low, and close prices) data of the stock. Although the OHLC data retains the trend characteristics of market movements to a certain extent, it eliminates most of the detailed trading information. Therefore, the limit order books (LOBs) data instead of traditional OHLC data are used in this paper. Besides, when the stock prices change dramatically, the states of the LOB data are significantly different. For example, when the stock prices rise/fall sharply, the bid-ask spread narrows/widens [4]. It suggests a significant correlation between stock prices and the LOB data.

In the financial stock markets, one LOB data contains ask and bid limit orders (LOs). Notably, the ask LO is a sell quote no less than a specified price, while the bid LO is a buy quote no more than a specified price. Figure 1 shows an example of 10 price levels of LOB data. It should be noted that there are one or more LOs of one specified price at a time point. Therefore, the LOBs presented in...
Figure 1 consists of price and the corresponding volumes. Compared with the OHLC data, the LOB data contains more information, and it can be used to reflect the fluctuation of the stock price better. In this paper, the 10 price levels of LOB data are used to forecast the stock price trend.

![LOB Example](image)

**Figure 1.** An example of 10 price levels of LOB data. The green and red bars represent the ask and bid orders, respectively. The length of the bar represents the volume of each price.

Traditional approaches for stock price prediction are based on statistical analysis models [5] and machine learning models [6]. Recent years have witnessed the rapid development of deep learning methods and its successful application in various fields such as image classification [7-9], speech recognition [10-12] and natural language processing (NLP) [13-15]. Recurrent Neural Network (RNN) is proposed to process sequential data; however, it causes vanishing gradient problems. To alleviate the problem, Long Short Term Memory (LSTM) network [16], which is a variant of RNN, is proposed to learn the long-term dependency by using the gating mechanism. Besides, the attention mechanism [17] is proposed to make the model focus on the salient features and improve the overall performance.

In this paper, an attention-based LSTM model is applied to study the relationship between LOBs and stock price and predict the stock price trend. Mainly, considering the time series data are used, the LSTM model is applied to extract features from a series of LOB data. And then, we use the attention mechanism to make the model focus on the salient and informative features for better representing the change of stock price. Finally, the whole model is applied to predict the stock price trend by using the LOBs of the previous day. By using the real LOBs and stock price data, we have shown the proposed attention-based LSTM model has achieved a relatively good result for predicting stock price trends.

The main contributions of the paper are as follows:
- Instead of using historical OHLC data, LOB data is used to predict the stock price trend.
- Inspired by the successful application of deep learning methods and considering the temporal properties of the LOB and stock price data, an attention-based LSTM model is applied for stock price trend prediction.
- By using real stock data in the China stock market, we demonstrate the effectiveness of the proposed attention-based LSTM model on stock price trend prediction.

The remainder of this paper is organized as follows. Section 2 describes the proposed attention-based LSTM model for stock price trend prediction by using the LOB data in detail. Section 3 presents the experimental results to validate the effectiveness of the model by using the real China stock market data. Section 4 gives the conclusion of the paper.

2. The Attention-Based LSTM Model for Stock Price Trend Prediction

In this section, the proposed attention-based LSTM model for stock price trend prediction is described in detail. We firstly introduce the overview of the attention-based LSTM model, and then each component of the whole model is described.
2.1. Overview of the Attention-Based LSTM Model

Instead of directly predicting the stock price at the next moment, the goal of this paper is to predict the stock price trend of the next day by using the LOB data. Considering the “T+1” trading rule for ensuring the stability of the stock market, in other words, stocks are bought by investors on one trading day can only be sold until the next trading day. Therefore, it is also meaningful to predict the stock trend of the next day for assisting the investors in making sound decisions.

In this paper, we use the attention-based LSTM model to predict the stock price trend based on the LOB data. Figure 2 shows an overview of our model.

![Figure 2. Overview of the attention-based LSTM model](image)

As shown in Figure 2, the LOB data of one day is used as the input of the model. Then the LSTM layer is used to extract features from the sequential LOB data, and the attention layer is utilized to learn the salient features corresponding to the stock price trend and improve the whole model performance. Finally, the learned feature representations are fed into several dense layers to give the predicted stock price trend of the next day.

2.2. LOB Data of the Model

For one LOB data, there are multiple levels of price and volume data. For the level \( l = 1, \ldots, 10 \) of one LOB data at the time \( t \), we use the \( pb^l_t \) and \( vb^l_t \) to represent the price and volume of the bid order, which can be obtained every 3 seconds. Correspondingly, the \( pa^l_t \) and \( va^l_t \) are used to describe the price and volume of the ask order. Therefore, at time \( t \), the LOB data of level \( l \) contains both the bid order and ask order, which can be expressed as \( LOB^l_t = (pb^l_t, vb^l_t, pa^l_t, va^l_t) \). As described above, the LOB data, which includes 10 price levels, is used in the paper as shown in Figure 1. Hence, the LOB data at the time \( t \) is 40 dimensions in total, which can be described as:

\[
LOB_t = (LOB^1_t, ..., LOB^{10}_t) = (pb^1_t, vb^1_t, pa^1_t, va^1_t, ..., pb^{10}_t, vb^{10}_t, pa^{10}_t, va^{10}_t)
\] (1)

With respect to the input of the whole model, it is unnecessary to use all LOB data obtained every 3 seconds of one day due to the needs of enormous processing capacity for processing these massive amounts of data in order to predict the stock trend of next day. Therefore, the LOB data used for training the model can be sampled at the same time interval. For example, in the following experiments of the paper, we sample the LOB data every 15 minutes from open to close on the China stock market. There are 23 LOB data in one trading day in total is used as the input of the model, which can be expressed as:

\[
Input = (LOB_1, LOB_2, ..., LOB_{23})
\] (2)

2.3. LSTM Layer of the Model

RNN is used to learn non-linearity features by processing sequential data. However, the vanishing gradient problem, which results from the backpropagation through time (BPTT) process, makes it difficult for RNN to learn long-term dependencies in a sequence effectively. As one of the variants of RNN, LSTM is proposed to mitigate the above problem by applying the gating mechanism, i.e., in the cell of LSTM, it contains multiple gates, including the input gate, forget gate, and output gate.

We use the following equations to represent the process of LSTM layer:

\[
F_t = \sigma(W_F \cdot [h_{t-1}, LOB_t] + b_F)
\] (3)

\[
I_t = \sigma(W_I \cdot [h_{t-1}, LOB_t] + b_I)
\] (4)
\[
C_t' = \tanh(W_C \cdot [h_{t-1}, \text{LOB}_t] + b_C) \\
C_t = F_t \cdot C_{t-1} + I_t \cdot C_t' \\
O_t = \sigma(W_O \cdot [h_{t-1}, \text{LOB}_t] + b_O) \\
h_t = O_t \cdot \tanh(C_t)
\] (5)  (6)  (7)  (8)

In the above equations, at the time \( t \), the input includes the previous hidden state \( h_{t-1} \) and the LOB data \( \text{LOB}_t \), the \( F_t \), \( I_t \) and \( O_t \) are the output of the forget, input, and output gates, respectively. \( C_t \) is the cell state. Also, \( W \) (i.e., \( W_f, W_i, W_C, W_O \)) and \( b \) (i.e., \( b_F, b_I, b_C, b_O \)) represent weight and bias, respectively. By applying these three gates, it can decide which parts of the input LOB data \( \text{LOB}_t \), the previous state \( h_{t-1} \), and the previous cell state \( C_{t-1} \) will be used to obtain the hidden state \( h_t \) of time \( t \).

2.4. Attention Layer of the Model

The attention-mechanism is applied to learn the informative and salient features and generate discriminative representations for better predicting stock price trend. Therefore, in the paper, an attention layer is added after the LSTM layer. The process of the attention layer is shown in Figure 3.

![Figure 3. Process of the attention layer after the LSTM layer](image)

The calculation process of the attention layer is shown as the following equation (9) and (10) according to the output of the LSTM layer at the time \( t \).

\[
a_i = \frac{\exp(W_a \cdot h_t)}{\sum_j \exp(W_a \cdot h_j)} \\
\text{Att}_t = \sum_j a_j h_j
\] (9)  (10)

Where \( a_i \) is the normalized weight to decide the importance of the hidden state \( h_t \) at the time \( t \) by using softmax function. \( W_a \) is the parameter vector of the attention layer. The generated representation after the attention layer, i.e., \( \text{Att}_t \), is calculated by performing a sum of weighted \( h_t \) according to \( a_i \).

The representation \( \text{Att}_t \) is then fed into multiple dense layers (i.e., fully connected layers) to obtain a high-level representation for giving the final predicted trend of the stock price.

2.5. The Predicted Trend of the Model

Although the stock price changes dynamically as the transaction progresses during a trading day, we intend to forecast the trend of stock price by the day rather than by the minute or even the second. Because the open price may be different from the close price of the previous day, the real trend, i.e., the label of the model, is defined by using the close price of adjacent days for ensuring continuity:

\[
K_{\text{real}} = \frac{p_{\text{close}}^d - p_{\text{close}}^{d-1}}{N-1}
\] (11)
Where \( p_{\text{close}}^d \) and \( p_{\text{close}}^{d-1} \) are the close price of two adjacent trading days (i.e., day \( d \) and day \( d-1 \)). \( N \) is the amount of LOB data used in the input. For example, \( N \) is 23 in the experiments.

For the attention-based LSTM model, it will directly give the predicted trend of the stock price \( K_{\text{pre}} \). The parameters of the model will be updated by minimizing the error between the \( K_{\text{pre}} \) and \( K_{\text{real}} \).

3. Experimental Results

In this section, we randomly select SSEC (Shanghai Stock Index) and other three stocks, including Boton Tec (Boton Technology), SINOSTEEL and Renhe PI (Renhe Pharmaceutical Industry), to validate the effectiveness of our model. As described in Section 2.2, we use 23 LOB data of one trading day as the input of the model. About 50000 data are used during the training. The Adam optimizer with a learning rate of 0.001 is applied, the final model is obtained until convergence or reaches a maximum of 1000 iterations.

Figure 4 shows the experimental results of the trading day from January 10, 2020, to January 20, 2020. As shown in the figure, only the corresponding stock prices of LOB data are presented in the figure to indicate the fluctuation of the stock price in a trading day (i.e., the red curve). The line between the two close prices of adjacent days represents the real trend \( K_{\text{real}} \) of the trading day (i.e., the blue line). The green arrow shows the predicted trend of the model \( K_{\text{pre}} \). It should be noted that the starting point of both the blue line and the green arrow is the close time point for ensuring continuity. Through comparing the blue line and the green arrow, it is intuitively that the obtained stock price trend by our model is basically in line with the reality for all experiments in most of the trading days.

For making a clear and better comparison, Table 1 gives the errors between the real and predicted stock price trend of all trading days shown in Figure 4. Notably, the absolute value of the error is calculated as follows:

\[
\text{err} = |K_{\text{pre}} - K_{\text{real}}|
\]  

In Table 1, the smaller the absolute value of the error, the closer the predicted trend is to the real trend. It should be noted that the sign of the error (i.e., \( \pm \) before the absolute value) indicates the consistency of the predicted and real trend. Especially, the sign + represents that the predicted trend is in line with the real trend, and the sign – indicates that the predicted trend is the opposite of the real trend. From the table, in most of the trading day, the model can predict the right trend, although some values of error are relatively large. It indicates that our model can be able to assist stock investors in making appropriate decisions according to the predicted trend to some extent. The experimental results validate the effectiveness of our model.
4. Conclusion

Stock price trend prediction is a challenging task due to non-stationary and high volatility of the stock prices. To address this problem, in the paper, an attention-based LSTM model is applied to predict the stock price trend by using the LOB data instead of traditional OHLC data. Especially, the LSTM layer is used to learn the temporal relationships in the sequential LOB data. And then, the attention layer is applied to produce a better representation to make the model focus on the salient features related to the stock price trend. Finally, through using the real LOB data and stock price data in the China stock market, experimental results have validated the effectiveness of the proposed model.

5. References

[1] Hu Y, Feng B, Zhang X, Ngai E.W.T., and Liu M 2015 Stock Trading Rule Discovery with an Evolutionary Trend Following Model. Expert Systems with Applications 42 (1): 212-222.
[2] Roondiwala M, Patel H, and Varma S 2017 Predicting Stock Prices using LSTM. International Journal of Science and Research (IJSR) 6 (4): 1754-1756.
[3] Zhuge Q, Xu L, and Zhang G 2017 LSTM Neural Network with Emotional Analysis for Prediction of Stock Price. Engineering letters 25 (2).
[4] Chan Y C 2000 The Price Impact of Trading on the Stock Exchange of Hong Kong. Journal of Financial Markets 3 (1): 1-16.
[5] Wilmott P 2013 Paul Wilmott on Quantitative Finance, John Wiley & Sons.
[6] Cavalcante R C, Brasilheiro R C, Souza V LF, Nobrega J P, and Oliveira A LI 2016 Computational Intelligence and Financial Markets: A Survey and Future Directions, Expert Systems with Applications 55: 194-211.
[7] Ma B, LI X, Xia Y, and Zhang Y 2020 Autonomous Deep Learning: A Genetic DCNN Designer for Image Classification. Neurocomputing 379: 152-161.
[8] Chen W, Xie D, Zhang Y, and Pu S 2019 All You Need Is a Few Shifts: Designing Efficient Convolutional Neural Networks for Image Classification. *CVPR 2019*: 7241-7250.

[9] Chen Q, Zhang W, Yu J, and Fan J 2019 Embedding Complementary Deep Networks for Image Classification. *CVPR 2019*: 9238-9247.

[10] Lam M W.Y., Hu S, Xie X, Liu S, Yu, J, Su R, Liu X, and Meng H 2018 Gaussian Process Neural Networks for Speech Recognition. *INTERSPEECH 2018*: 1778-1782.

[11] Li K, Xu H, Wang Y, Povey D, and Khudanpur S 2018 Recurrent Neural Network Language Model Adaptation for Conversational Speech Recognition. *INTERSPEECH 2018*: 3373-3377.

[12] Zhang W, Cui X, Finkler U, Kingsbury B, Saon G, Kung D S, and Picheny M 2019 Distributed Deep Learning Strategies for Automatic Speech Recognition. *ICASSP 2019*: 5706-5710.

[13] Chai J and Li A 2019 Deep Learning in Natural Language Processing: A State-of-the-Art Survey. *ICMLC 2019*: 1-6

[14] Liu X, He P, Chen W, and Gao J 2019 Multi-Task Deep Neural Networks for Natural Language Understanding. *ACL (1) 2019*: 4487-4496.

[15] Malykh V Robust to Noise Models in Natural Language Processing Tasks. *ACL (2) 2019*: 10-16.

[16] Hochreiter S and Schmidhuber J 1997 Long Short-Term Memory. *Neural Computation* 9(8): 1735-1780.

[17] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser L, and Polosukhin I 2017 Attention is All You Need. *NIPS 2017*: 5998-6008.