A Deep Neural Network for Stock Price Prediction

Hengwei Dai¹,a,†, Wenfei Wang²,b,†, Jiashang Cao³,c,†, Haotian Wu⁴,d,†

¹Southwest University, Chongqing, China;
²University of Toronto, Toronto, Ontario, Canada;
³Michigan State University, East Lansing, Michigan, United States;
⁴Beijing Jiaotong University, Beijing, China.

‡dzy201021@email.swu.edu.cn, ‡wenfei.wang@mail.utoronto.ca, ‡caojiash@msu.edu,
‡19711032@bjtu.edu.cn

†These authors contributed equally.

Abstract. The stock market is a large, complex system that consists of many human factors. Due to the human action's randomness to the market environment, the market behavior is usually erratic. In this case, many previous works were trying to find out an automatic prediction model that can capture relationships between stock market prices and the surroundings of the stock markets. Recently, a simple LSTM-attention mechanism was proposed to advance the prediction accuracy of previous works. However, the simple LSTM with attention mechanism only demonstrates how the past data influence the next. This forward-only relationship does not reflect a backward relationship, which is how the future data is related to the past data. To enforce the backward relationship on the attention-LSTM model, a multilayer bidirectional LSTM model with an attention mechanism is proposed. A bidirectional LSTM layer that can encode the data relationship in both directions is leveraged. Then attention mechanism is applied to introduce the model with the ability to focus on critical content and ignore the disturbance. Experimental result shows that the attention-BiLSTM model has a coefficient of the determinant of 0.9940, which is better than the baselines.

1. Introduction
Stock price prediction is always a critical topic for individuals and organizations in stock markets. Since the complicated stock markets involve various factors, it is difficult to predict future stock market prices. Given a timeline, stock market prices at different moments can be influenced by different events. In other words, the relationships among the data are not straightforward enough to be computed because many factors are entangling. Besides, these factors are too complicated to be computed.

Thus, it is necessary to predict stock prices automatically. Intuitively, statistical methods like least regression fit and personal experience are feasible approaches to compute the prediction on the stock. However, it is not practical for a human to generate an exact value prediction based on his experience with the stock prices, so the performance of the human prediction will be difficult to evaluate. Moreover, statistical methods [1] do not have a satisfying performance. In recent years, the development of neural networks has led to the popularity of its application in stock market prediction.
because neural networks are good automatic approximators on big data. A well-designed neural network can drive the prediction to generate more precise results.

Many different types of neural network models were applied to the stock market data predictions in the research fields. Because stock prices are changed over time, the data pattern of the stock market status is a time sequence. In this case, the recurrent neural network (RNN) [2] was widely used for the prediction. For example, Nabipour and Nayyeri [3] employed long short-term memory (LSTM) to predict stock prices. With the LSTM, the vanishing gradient problem, which forgets the long-term pattern, can be solved. Stock data with a longer sequence can be used with LSTM. However, with only LSTM, the data in each part of the sequence does not have attention on a specific past event that is critical to itself. It can only keep track of the aggregation of all the past data. Recently, Qiu et al. [4] implemented an attention mechanism to an LSTM model with the name LSTM-attention model, which reinforces the data relationship among the data by enabling each part of the data sequence to have a retrospect to each of the past events.

Nevertheless, the connection between each data in the sequence is not enough. In this paper, a model called the Attention-BiLSTM is proposed for the stock market prediction. The motivation behind this mechanism is by assigning an additional backward sequence layer to the data sequence, and the model now has a greater capability to encode the sequence data in different locations in the timeline. In this case, the connection among the sequence will be stronger, and better utilization of the sequence pattern will be achieved. These can improve the accuracy of the model to a certain degree. In the model of this paper, the process can be divided into three sections. The input data will be sent to a sequence processing layer first. Then, it will be fed into a section that consists of three bidirectional LSTM layers. Finally, all the output from the last section will be processed by a soft attention mechanism part, and the output from the attention architecture will be the result. The experiment was applied to the SSE composite index. The result shows that the Attention-BiLSTM model has a coefficient of the determinant of 0.9940, which is better than the baseline model simple LSTM-attention [4].

2. Related Work

With the breakthrough of deep learning (DL) in recent years, different neural network models were employed to predict stock market prices. Hiransha et al. (2018) leveraged four DL architectures - MLP, RNN, LSTM, and CNN to predict stock prices of NSE and NYSE, which were two leading stocks in the world. Their work shows that stock market price contained underlying dynamics, and DL models were capable of identifying such dynamics while statistical methods like ARIMA cannot. And their experimental results demonstrated that CNN performed best when using only a single model for prediction [5].

A study carried out by Nabipour et al. (2020) employed six tree-based models (Decision Tree, Bagging, Random Forest, Adaboost, Gradient Boosting, and XGBoost) and three DL models (ANN, RNN, and LSTM). They predicted the stock market price of the Tehran stock exchange for 1, 2, 5, 10, 15, 20, and 30 days in advanced and used MAPE, MAE, RRMSE, and MSE as evaluation methods. LSTM was observed to have the lowest error and best fit of predictions of all stock market groups [3].

Khaled et al. (2018) performed a study to compare bidirectional and stacked LSTM with neural networks and basically formed LSTM. Bidirectional LSTM could make use of current future inputs, and stacked LSTM could capture complex patterns. Their experiments used Standard & Poor 500 Index data and predicted short-term (1 day) and long-term (30 days) stock prices. Their results showed that both BLSTM and SLSTM performed better than shallow neural networks and LSTM. They had better performance in short-term prediction over long-term prediction. In particular, BiLSTM achieved the best performance for both short-term and long-term prediction [6].

Pisut and Peerapon (2018) proposed a model that took historical stock price as along with news headlines as input. Event embedding was applied to long-term (past 30-days), mid-term (past 7-days), and short-term (past 1-day) events. Long-term and mid-term event vectors were fed into a pooling layer to extract important features then concatenated with short-term event vectors as the input to
hidden layers. The seven technical indicators proposed by [7, 8] were computed using historical stock prices then normalized before feeding into the LSTM network and stock price data. The prediction based on news headlines and historical data were concatenated as the input to the final hidden layer. Their work demonstrated that this hybrid model gave better prediction than models using event embedding vectors alone, headlines and indicators, news headlines alone, and indicator alone. They also suggested introducing attention mechanisms in order to give future improvements [9].

3. Methodology

3.1. The framework of Attention-BiLSTM

Attention-BiLSTM consists of three parts, the input layer, three layers of BiLSTM, and one layer of Attention mechanism. The framework of our model is shown in Figure 1:

![Figure 1. The overall flowchart of Attention-BiLSTM.](image)

3.2. The input layers

The input layer first obtains the input sequence consists of $X = (x_1, x_2, x_3, ..., x_T)$, the sequence length is $T$, $x_i$ represents the model input at the time $i$.

3.3. BiLSTM layers

Then there are three layers of BiLSTM, each layer (label m) includes $T$ number of LSTM units, form two sub-networks in opposite directions together. Both of sub-networks are the LSTM-based recurrent NN structure. The basic components of the LSTM are as follows: one in-put gate $i^t_m$ with weight matrix $(W_{ixm}, W_{ihm}, W_{cxm}, b_{im})$, one forget gate $f^t_m$ with weight matrix $(W_{fxm}, W_{fhm}, W_{cfm}, b_{fm})$, and one output gate $o^t_m$ with weight matrix $(W_{oxm}, W_{ohm}, W_{ocm}, b_{om})$. All of those gates are set to generate some degrees, use the $h^t_{m-1}$ input at the current time $t$ of the previous layer, the hidden layer state $h^t_{m-1}$ generated by the previous step $t-1$, and the current state $c^t_{m-1}$, to decide whether to use the current input, or forget the memory before, or output the state generated after.

So, the current state is determined by the state of the previous moment and the information generated by the input gate and forget gate at the current moment. As these following equations demonstrate [10]:

$$...$$
\[ i^t_m = \sigma(W^i_m \cdot [h^t_{m-1}, h^{t-1}_{m}]) \]  
\[ f^t_m = \sigma(W^f_m \cdot [h^t_{m-1}, h^{t-1}_{m}]) \]  
\[ c^t_m = \text{tanh}(W^c_m \cdot [h^t_{m-1}, h^{t-1}_{m}]) \]  
\[ c^t_m = i^t_m c^t_m + f^t_m c^{t-1}_{m} \]  
\[ o^t_m = \sigma(W^o_m \cdot [h^t_{m-1}, h^{t-1}_{m}]) \]  
\[ h^t_m = o^t_m c^t_m \]

At the same time, considering the hidden state transfer in the opposite direction. So, the output \( y^t \) of the last layer (M) at the \( t \)-th moment is determined by the results of the two directions \( \tilde{h}^t_m \) and \( \tilde{h}^t_m \), as shown in the following equations:

\[ h^t_M = [\tilde{h}^t_m \oplus \tilde{h}^t_m] \]  
\[ y^t = \text{softmax}(h^t_M) \]

### 3.4. Attention mechanism

The three-layer BiLSTM can obtain the state of the financial market at each moment. But according to experience, the information at different moments has a different influence on the state prediction at the current moment. Therefore, the attention mechanism is introduced to evaluate the influence weights at different moments [11].

\( H = (h^1_M, h^2_M, h^3_M, \ldots, h^T_M) \) is a set of states at each moment output by the third layer of BiLSTM. Using \( H \) as the input of the attention mechanism, the degree \( e^t_{t'} \) of hidden state needs to pay attention at time \( t \) can be calculated. As shown in the following equations:

\[ e^t_{t'} = \text{tanh}(W_h[h^{t-1}_{M}, h^t_{M}]) + b_h \]  

After obtaining \( e^t_{t'} \), calculating the weight \( \alpha^t_{t'} \) of the hidden state at time \( t' \) and the total weight at time \( t \) that the output sequence at time \( t \) focuses on. As shown in the following equations:

\[ \alpha^t_{t'} = \frac{\exp(e^t_{t'})}{\sum_{t'=1}^{T} \exp(e^t_{t'})} \]  
\[ r_t = \sum_{t'=1}^{T} \alpha^t_{t'} h^t_{M} \]

Finally, the output value at time \( t \) is obtained by:

\[ y_t = BiLSTM(r_t, h_{t-1}, c_{t-1}) \]

### 4. Experiments

#### 4.1. Dataset

The dataset is collected from the Shanghai Securities Composite Index from 1990.12 to 2020.3, the volume of which is 7146*4. The detailed explanation of each field in the dataset is shown in table 1.

| Field          | Description              |
|----------------|--------------------------|
| Timestamp      | Time index               |
| Price          | The price of the stock   |
| Stock_Volume   | The number of shares issued |
| Amount_Volume  | The total amount of shares issued |

The data set into training/test sets is divided according to the sequence order. Specifically, 80% of the data is used as the training set and 20% of the data is used as the test set to test the performance. And the Min-Max data normalization method is used to pre-process the data:
where $x$ represents the original value, and $x'$ represents the value in the interval [0,1] after compression.

### 4.2. Experimental Settings
Setting the batch size to 32. There are 3 hidden layers in the model, and each layer uses 64 neurons. And a dropout layer (dropout rate=0.2) is introduced to prevent overfitting caused by the reduction of model generalization ability.

In addition, Adam (original learning rate=0.001) is used as the optimizer, which has the advantages of high efficiency, adaptive adjustment of the learning rate, and small memory requirements. It is suitable for application scenarios that require a large number of parameters [12].

### 4.3. Evaluation Criteria
In order to test the predictive potential of the model, $R^2$, MAE, MedAE and RMSE are selected as the model evaluation criteria, and the functions are:

\[
R^2(y, \hat{y}) = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}
\]  
\[
MAE(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|
\]  
\[
RMSE(y, \hat{y}) = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}
\]  
\[
MedAE(y, \hat{y}) = median(|y_1 - \hat{y}_1|, \ldots, |y_n - \hat{y}_n|)
\]

where $y$ is the true value, $\hat{y}$ is the predicted value, $\bar{y}$ is the average of the true values, $n$ is the total number of samples.

MAE, MedAE, and RMSE are used to assess the error level of the prediction results. The smaller the value, the more accurate the prediction effect. $R^2$ is used to evaluate the degree of fitting of the predicted results to the overall raw data. The larger the predicted value, the higher the fitting degree of the data and the better the prediction effect.

### 4.4. Baselines
In order to compare the performance of the Attention-BiLSTM model, the deep learning model without introducing the attention mechanism, and the traditional statistical model on the stock price prediction task, three classical statistical models and three deep learning models are constructed as baselines in this work. Three classical statistical models include linear regression, ARIMA, and Prophet, and three deep learning models include LSTM and BiLSTM. The same normalization and split methods for the datasets are adopted on different baselines in terms of variable controlling. The same hidden layer structure of 3 layers and 64 neurons, the Adam optimizer, 500 epoch, and 32 batch size are used on LSTM and BiLSTM models.

### 4.5. Experimental Results and Analysis
Evaluating the performance of seven models through four different evaluation indexes, namely MAE, RMSE, $R^2$ and MedAE (These four indicators are calculated on the test set). The experimental results are shown in Table2:

| Table 2. Comparison of the Performance of BiLSTM-Attention and Baseline models. |
|-------------------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| Attention-BiLSTM                | LSTM              | BiLSTM            | Linear Regression | Prophet           | ARIMA             |
| MAE                            | 0.0085            | 0.0096            | 0.0092            | 0.0130            | 0.0795            | 0.0971            |
| RMSE                           | 0.0139            | 0.0148            | 0.0146            | 0.0195            | 0.1027            | 0.1212            |
| $R^2$                          | 0.9940            | 0.9931            | 0.9934            | 0.7783            | -0.1493           | -0.6718           |
| MedAE                          | 0.0051            | 0.0066            | 0.0061            | 0.009             | 0.06              | 0.094             |
It can be observed from table 2 that the values of MAE (0.0085), MedAE (0.0051) and RMSE (0.0139) in the Attention-BiLSTM model are all the minimum in the seven groups of experiments, and $R^2$ (0.994) is the maximum in the seven groups of experiments. Therefore, it can conclude that the Attention-BiLSTM model outperforms the other 6 baselines on the stock price prediction task.

ARIMA and Prophet rely excessively on historical data, which is prone to overfitting. Linear regression cannot fully mine the sequence features, nor can it solve the problem of potential long-term dependence among data. In deep learning, varieties of RNN such as LSTM can store effective sequence information through the gating mechanism and self-adaptively extract sequence features, thus solving the problem of long-term dependence among data to some extent. These factors may make the depth models overperform the traditional statistical model in the stock forecasting task in this paper. Based on LSTM, Attention-BiLSTM introduces a bidirectional operation and attention mechanism, which can synthesize historical and future sequence information and mine sequence features more fully and pay different degrees of attention to sequence information at different moments, thus improving model accuracy.

4.6. The impact of different parameter

![Figure 2. Comparison of $R^2$ values of Attention-BiLSTM models with different number of hidden layers.](image)

As can be observed from figure 2, when the number of hidden layers of BiLSTM is increased to 3 layers, the test set $R^2$ (0.994) reaches the maximum value and the fitting effect of the model is optimal. With the further increase of the number of hidden layers, the model's accuracy decreases gradually, and the calculation of parameters increases gradually. Therefore, the BiLSTM network structure with 3 hidden layers is selected.

5. Conclusion

In this paper, the Attention-BiLSTM model is proposed to improve prediction on stock market prices. The model consists of three layers of BiLSTM and one attention layer. Model performance on the stock price of the Shanghai Index was superior to both deep learning methods and statistical methods. The future work would focus on long-term stock market characteristics learning. A limited number of layers of a BiLSTM model, using a transformer or residual connection and an attention layer, may help predict stock market prices.

References

[1] Xia Y, Liu Y, Chen Z. Support Vector Regression for prediction of stock trend[C]/2013 6th International Conference on Information Management, Innovation Management and Industrial Engineering. IEEE, 2013, 2: 123-126.

[2] Medsker L R, Jain L C. Recurrent neural networks[J]. Design and Applications, 2001, 5.

[3] Nabipour M, Nayyeri P, Jabani H, et al. Deep learning for stock market prediction[J]. Entropy, 2020, 22(8): 840.

[4] Qiu, Jiayu, et al. “Forecasting Stock Prices with Long-Short Term Memory Neural Network
Based on Attention Mechanism.” *PLOS ONE*, vol. 15, no. 1, 2020, doi:10.1371/journal.pone.0227222.

[5] Hiransha M, Gopalakrishnan E A, Menon V K, et al. NSE stock market prediction using deep-learning models[J]. Procedia computer science, 2018, 132: 1351-1362.

[6] K. A. Althelaya, E. M. El-Alfy and S. Mohammed, "Evaluation of bidirectional LSTM for short-and long-term stock market prediction," 2018 9th International Conference on Information and Communication Systems (ICICS), Irbid, Jordan, 2018, pp. 151-156, doi: 10.1109/IACS.2018.8355458.

[7] M. R. Vargas, B. S. L. P. De Lima and A. G. Evsukoff, "Deep learning for stock market prediction from financial news articles", in Proceedings of the 2017 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA), pp. 60-65, 2017.

[8] Y. Zhai, A. Hsu and S. K. Halgamuge, "Combining news and technical indicators in daily stock price trends prediction", *International symposium on neural networks*, pp. 1087-1096, 2007.

[9] P. Oncharoen and P. Vateekul, "Deep Learning for Stock Market Prediction Using Event Embedding and Technical Indicators," 2018 5th International Conference on Advanced Informatics: Concept Theory and Applications (ICAICTA), Krabi, Thailand, 2018, pp. 19-24, doi: 10.1109/ICAICTA.2018.8541310.

[10] Zhou P, Shi W, Tian J, et al. Attention-based bidirectional long short-term memory networks for relation classification[C]//Proceedings of the 54th annual meeting of the association for computational linguistics (volume 2: Short papers). 2016: 207-212.

[11] LIU Chong, DU Jun-ping. Financial Data Prediction Method Based on Deep LSTM and Attention Mechanism [J].Computer Science,2020,47(12):125-130.

[12] Kingma D P, Ba J. Adam: A method for stochastic optimization[J]. arXiv preprint arXiv:1412.6980, 2014.