Research on Stock Prediction Based on Bidirectional Long Short Term Memory Neural Network Model Based on Attention Mechanism Optimization

Zhaorui Li *

Department of Network Engineering, Shandong University of Science and Technology, 271019, Shandong China

* Corresponding Author Email: 17661229610@163.com

Abstract. The complexity and volatility of financial markets place higher demands on the accuracy of stock forecasts. This paper uses Ping An Daily Market Data as the source of experimental data for empirical analysis, and innovatively combines BiLSTM with attention mechanism for stock price prediction in the experimental process. Introduce an attention mechanism, AM, to reinforce important features by reassigning the weight of each data; Bidirectional LSTM can learn sequential and reverse time series data information to ensure adequate data utilization. After establishing BP (backpropagation), LSTM and CNN-LSTM models for comparison, it is found that the proposed AM-BiLSTM model has high accuracy, good feasibility and versatility.

Keywords: Attention mechanism, BiLSTM model, CNN model, Neural networks, Stock forecasts.

1. Introduction

The modern market economy has seen tremendous development and transformation, leading to the financial service mechanism system becoming increasingly standardized and organized. Consequently, people have become more and more invested in the financial market, and their need for financial information has grown daily. The stock market's position in the financial market is becoming increasingly significant due to the modern development of economic and social informatization, as well as the Internet's modernization, which has enabled computers and other technical systems to become increasingly comprehensive. Consequently, stock price forecasting has become particularly crucial with the stock market's expansion. Generations of researchers have labored ceaselessly, resulting in the creation of a multitude of algorithms for stock forecasting.

A neural network of long-term memory, known as Long-Short Term Memory (LSTM), employs input, output, and forget gates to manage data, store long-term information in other unit states, possess long-term memory capabilities, and demonstrate superior performance in dealing with intricate financial time series data. In 2019, Siami-Namini et al. found that BiLSTM models provide better prediction results than ARIMA and LSTM models [1].BiLSTM, in contrast to the one-way propagation of traditional LSTMs which can cause a mismatch or even the exclusion of context information, advances and broadens the traditional LSTM neural network model, thereby enhancing the financial long time series regression problem model's performance.

The attention mechanism assists the model to capture important information by assigning weight parameters, which can directly obtain the relationship between global and local, and the model is relatively simple and low complexity [2], and is now used to varying degrees in financial markets. Among them, Wu Xiaofei found that the attention mechanism can introduce the exchange rate into LSTM, and the RMSE and MSE of mixed LSTM are lower than other LSTMs [3]. Kim Sang-yeon et al. Various deep learning models, such as multilayer sensors, one-dimensional convolutional neural networks, superimposed long short-term memory, attention networks, and weighted attention networks, are compared for financial time series forecasting [4]. Based on LSTM, Wei, Dou established a stock prediction model and introduced an attention mechanism to capture the characteristics of stocks at important time points. It can be seen from practice that the LSTM neural network model has some disadvantages, such as the time delay of prediction. However, due to the attention layer, it can make predictions about stock prices [5]. Combining the advantages of

bidirectional LSTM and attention mechanism AM, this paper proposes an AM-BiLSTM model, and establishes BP, LSTM and LSTM-CNN models for comparison. Finally, the experimental results of four models on the Ping A dataset of China show that the AM-BiLSTM model proposed in this paper has good validity and versatility.

2. Model structure and principle

2.1. LSTM Principles and modeling

The problem of "vanishing gradient" that arises with increasing depth has yet to be adequately addressed. LSTM, a type of recurrent neural network, is one such network that can address this issue. It is suitable for temporal feature extraction and has the capability to learn the dependence of long-term series accurately. This is in contrast to the initial utilization of RNNs to learn sequence patterns on time series data.

Figure 1 shows the general structure of LSTM, it can be clearly seen that LSTM has three activation functions: input gate i_t, forget gate f_t,output gate o_t, these three gates through the control of information, can retain the information useful to the final training target, better solve the problem of gradient disappearance. c_t is the current cell, h_t represents the hidden layer and x_t represents the input data.

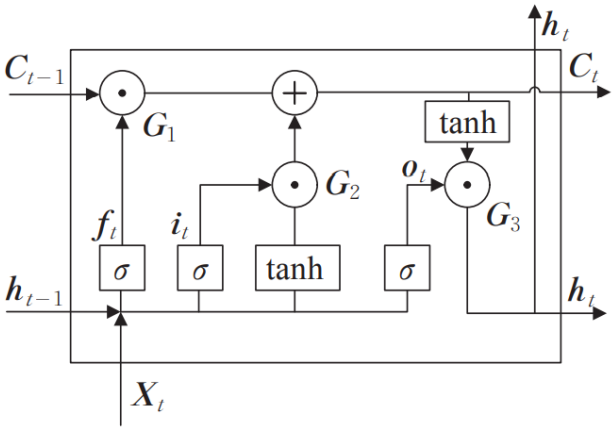


Fig 1. LSTM structure

By analyzing the structure, it can be seen that the working principle of LSTM: as shown in Equation (1) is the forget gate in the three activation functions, where the function of the sigma function is to output 0 or 1, 0 means that the information is completely discarded, and 1 means that the output of the structure is completely preserved. Equations (2) and (3) are the main functions in the input gate, and their role is to control how much input information is retained, and the function of cell renewal in this activation function is the core of LSTM. Equation (4) shows the process of cell renewal. The output gate in the activation function is responsible for determining the value of the next hidden state, as shown in Equations (5) and (6), through the sigma function, the output gate weight matrix, and the output gate offset term, the final unit output is obtained. In Figure 1 and the formula mentioned above, the operator "·" represents the cross product of vectors, and "+" represents the vector superposition operation.

$$f_t = \sigma(W_f[h_{t-1}, X_t] + b_f) \tag{1}$$

$$i_t = \sigma(W_i[h_{t-1}, X_t] + b_i) \tag{2}$$

$$C_t = \tan\mathrm{h}(W_c[h_{t-1}, X_t] + b_c) \tag{3}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \overline{c_t} \tag{4}$$

$$o_t = \sigma(W_o[h_{t-1}, X_t] + b_o) \tag{5}$$

$$h_t = o_t \cdot \tanh(C_t) \tag{6}$$

The extraction of series features of stock time series data is made possible by the construction of a LSTM model. In the LSTM model, the data is converted into a stacked data type, which can easily solve the problem of data input at the same time. The model has good effectiveness when using data from T-30 to T and predicting the closing price at the moment of T+5[6].

The LSTM modeling, depicted in Figure 2 and constructed with two layers of LSTM and two layers of fully connected layers, enhances the nonlinear prediction capability of the model, as a result of this.

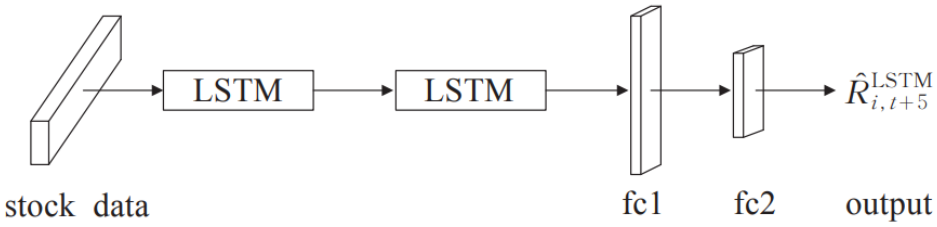


Fig 2. LSTM model

2.2. AM-BiLSTM Short-term stock price prediction model construction

Traditional LSTMs can only transmit data forward or backward, and cannot encode back-to-front information, and cannot notice the interaction between data information.BiLSTM, which can transmit and encode data in both directions, allows us to comprehend it as taking into account both past and future data. To better capture bidirectional data dependence, we should combine forward and backward LSTM. The forward and reverse inverse processing sequences represent two different dimensions of the data, as shown in equation (7) representing the hidden state H of BiLSTM at t, containing forward and reverse h[7].

$$\begin{aligned} \overrightarrow{h_t} &= \overrightarrow{\mathrm{LSTM}}(h_{t-1}, x_t, c_{t-1}), t \in [1, T] \\ \overleftarrow{h_t} &= \overleftarrow{\mathrm{LSTM}}(h_{t+1}, x_t, c_{t+1}), t \in [T, 1] \\ H_t &= [\overrightarrow{h_t}, \overleftarrow{h_t}] \end{aligned} \tag{7}$$

BiLSTM's capability to capture the connection between long-term series through forward and backward combination, as well as its capacity to remember long-term information through the "gate" structure, provides a more robust temporal information characteristic [8]. Consequently, this article will forecast short-term stock prices based on BiLSTM.

In recent years, AM has developed rapidly, and the combination model established by integrating with other models has quickly become one of the effective methods to solve practical problems. Scholars have long advocated for the utilization of global and local attention models in machine translation [9]; since then, the attention mechanism has been extensively utilized in a variety of natural language processing areas, including deep neural networks, and has yielded positive outcomes. Let us assume that each component of the AM input set is a key and value, with a target of T, and the outcome is attention weight.

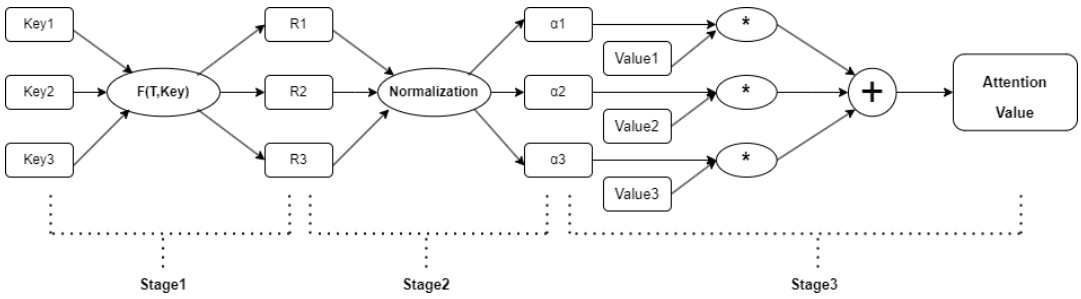


Fig 3. AM Calculation process

In Figure 3, the process of computing AM is elucidated: the correlation R is determined by the correlation calculation function F (T, Key). The attention weight a of the input feature αt is calculated after calculating it: i.AM can be implemented in a neural network by introducing it, which can be expressed as:

$$\begin{aligned} e_t &= F(T, Key) \\ \alpha_i &= \frac{\exp(e_i)}{\sum_{j=1}^{n} \exp(e_i)} \\ att &= \sum_{i=1}^{n} \alpha_i \cdot value \end{aligned} \tag{8}$$

The figure below displays the deep learning structure diagram that was acquired through the amalgamation of the attention mechanism and BiLSTM in this paper.

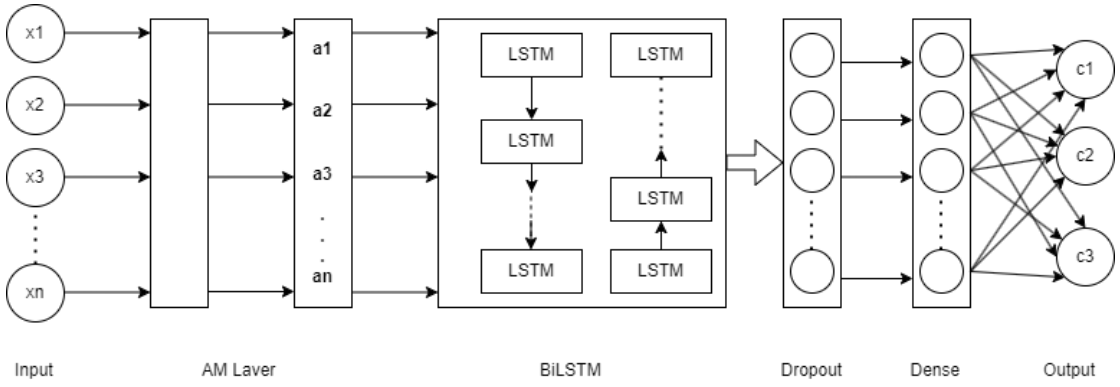


Fig 4. Deep learning

The backpropagation algorithm of LSTM is employed in this paper, with equation (9) as the error function. If the predicted value fails to meet the model output conditions, the error is then backpropagated to each layer node, and the model parameters are continually modified to reduce the error value until convergence is achieved and training is concluded.

$$E = \frac{1}{n}\sum_{i \in inputs} (y_i - h_i)^2 \tag{9}$$

2.3. CNN Principles and modeling

CNN's convolutional capability can draw out distinct components of the input to create layer-by-layer convolution of the input data, and its pooling process can subsample the feature map acquired by the convolutional layer. As an important role in feature extraction, the convolutional layer iterates on the input data to extract the deep features of the data by local perception characteristics and weight sharing mechanism [10]. Equation (10) displays the continuous convolution process, while Equation (11) displays the formula for the discrete convolution.

$$s(n) = (f * \omega)(n) = \int_{-\infty}^{\infty} f(x)\omega(n-x)dx \tag{10}$$

$$s(n) = (f * \omega)(n) = \sum_{x=-\infty}^{\infty} f(x)\omega(n-x) \tag{11}$$

The input function, f, is weighted by function ω, and the mapping of features, s, is the output.

Maximum or average pooling is employed by the pooling layer to subsample the convolutional layer's output, preserving the main features and minimizing the input dimension of the following network layer to avoid overfitting.

The advantage of CNN is that high-dimensional data can be effectively processed by sharing convolution kernels and combined layer by layer to generate abstract high-level features. Despite the lack of a memory function and the disregard of temporal correlation of time series data [11], CNNs remain.

The convolutional neural network, a type of feedforward neural network with convolutional computation, is widely known for its deep structure, which is usually composed of an input layer, an

alternating convolutional layer and a pooling layer, a fully connected layer, and an output layer [12]. The convolutional and pooling layers usually appear in an alternating fashion. Once the convolutional and pooling layers are finished, the output layer and the fully connected layer are joined, thus allowing us to create a general schematic diagram of the convolutional neural network's structure.

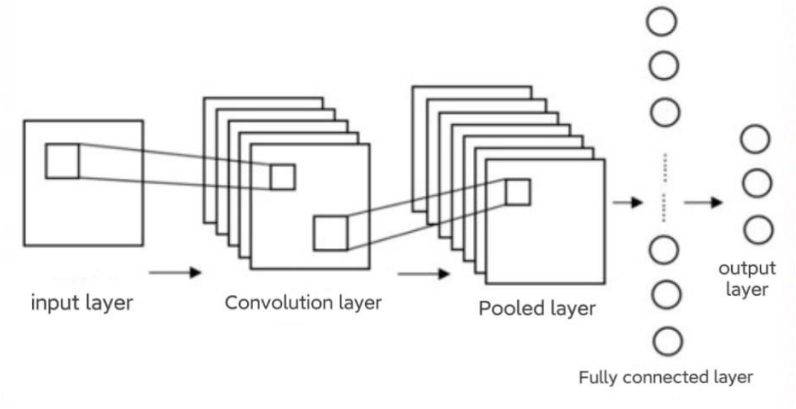


Fig 5. CNN Structural Diagram

3. Empirical analysis

3.1. Data Origin and Processing

From September 02, 2022 to March 03, 2023, this paper utilized the daily market data of Ping An stocks as the research object, selecting five basic indicators such as opening price, closing price, highest price, and trading volume for experiments, as depicted in Figure 6 - a total of 606 pieces of data were collected and utilized during the experiment.

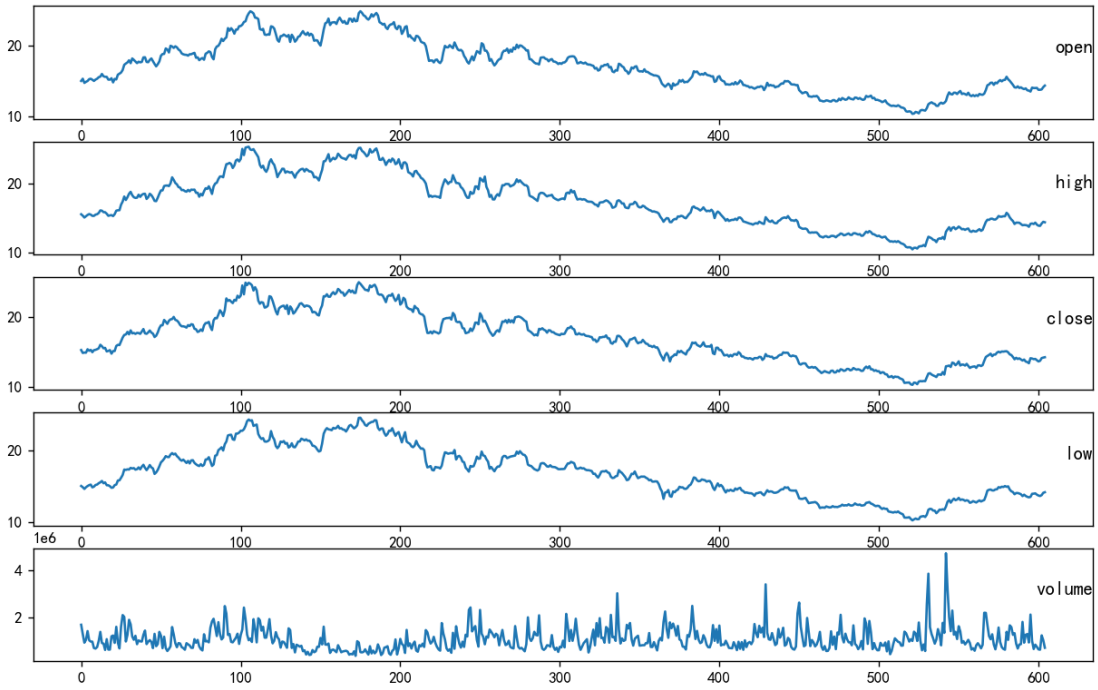


Fig 6. Trends in each indicator

The Max-Min normalization method was used to normalize the data before training, and the formula is shown in (11). In order to ensure the effectiveness of the experiment, the established model adopts the method of cross-validation, and the collected data is continuously input into the model according to the time sliding window, and then the model is trained.

$$z_{i,j} = \sum_{i=1}^{13} \sum_{j=1}^{2627} \frac{x_{i,j} - \min_{x_{i,j}}}{\max_{x_{i,j}} - \min_{x_{i,j}}} \tag{12}$$

On day j, the ith indicator's stocks contain the data of $x_{i,j}$, with $\max x_{i,j}$ and $\min x_{i,j}$ being the maximum and minimum data, respectively. After training, the model results are denormalized, and finally the error analysis of the prediction results is performed.

3.2. Experimental process

Figure 7 illustrates the process of preprocessing data, setting up experiments, training the model, fine-tuning parameters, predicting the model, and comparing the model - all of which are the main components of the experiment in this paper.

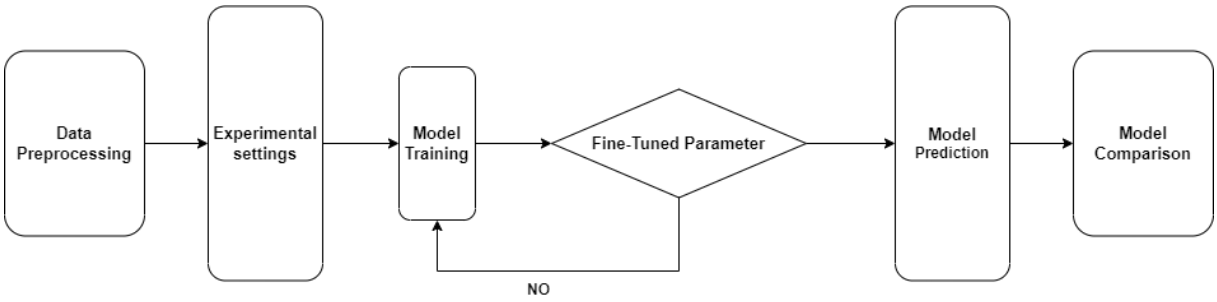


Fig 7. Experimental process

3.3. Parameter settings and evaluation indicators

To maximize the learning rate and keep the response to gradient updates faster, the optimizer chose Nadam. In order to accelerate the discovery of extreme values, reduce memory, and improve training speed, the batch_ size is set to 256, and the debugging results show that the model has the most favorable optimization when Dropout is located between BiLSTM and the fully connected layer. Thus, in subsequent experiments, the Dropout layer was located between the BiLSTM and the fully connected layer. In general, when the Dropout value is set too low, it may have no practical use; the Dropout value is set too high, which can lead to underfitting, so this article sets the Dropout to 5% after debugging the experiment.

In order to verify the legitimacy and scientificity of the model, the mean absolute error (MAE), mean squared error (MSE), coefficient of determination R2 and root mean square error (RMSE) were selected as its evaluation indicators, among which the smaller the RMSE and MAE values, the more accurate the prediction results of the model and the smaller the deviation from the actual value. The closer the coefficient of determination R2 is to 1, the better the goodness of fit of the model and the better the prediction effect. The equation is as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^{n} \left| y_i - \tilde{y}_1 \right| \tag{13}$$

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \tilde{y}_1)^2 \tag{14}$$

$$R^2 = 1 - \frac{\mathrm{MSE}\{y_i - \tilde{y}_i\}}{\frac{1}{n} \sum_{i=1}^{m} (y_i - \tilde{y}_1)^2} \tag{15}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{n-1}^{N} (y_n' - y_n)^2} \tag{16}$$

3.4. Model results and comparisons

To assess the efficacy of the BiLSTM and attention mechanism combined model in forecasting multi-feature and time-series long stock data, the test set in this paper's prediction results are contrasted

to the BP, LSTM, CNN-LSTM, and AM-BiLSTM models; however, the BP and LSTM models' prediction results are unsatisfactory and are not featured in this paper.

Comparing Figures 8 and 9, it can be seen that the closing price predicted by the AM-BiLSTM combination model is significantly better fitting than the CNN-LSTM combination model and the fitting effect is getting better and better. Observing MSE reveals that the AM-BiLSTM model's estimator and estimated amount are much more similar than the CNN-LSTM model.

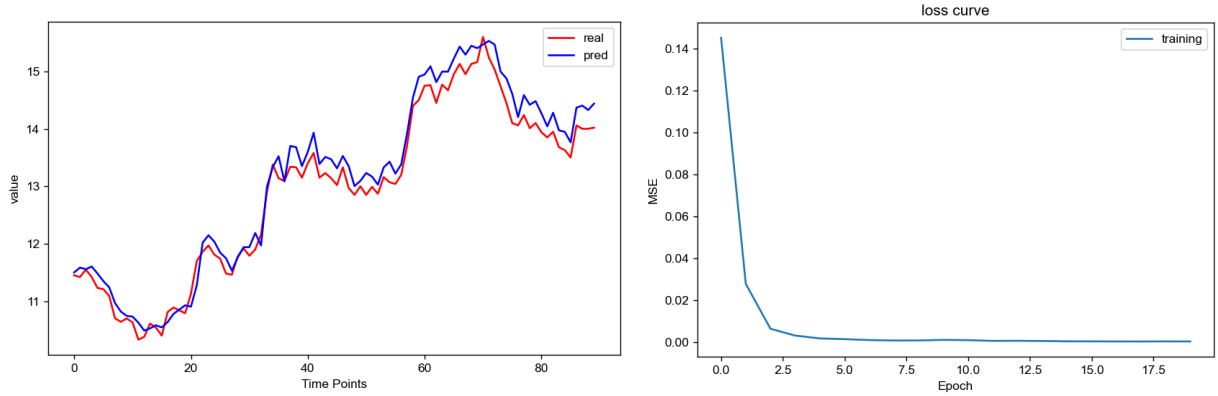


Fig 8. CNN-LSTM Prediction result graph

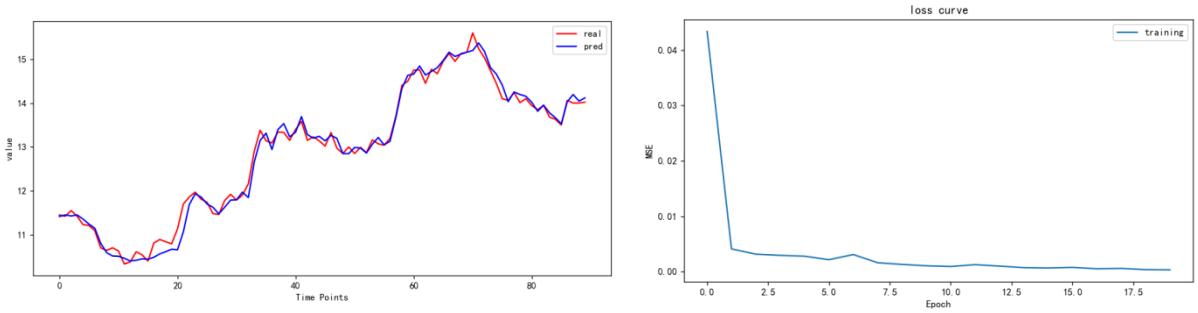


Fig 9. AM-BiLSTM Prediction result graph

Table 1's comparative analysis of the prediction results reveals that the AM-BiLSTM model predicts MAE to drop by 23.56%, 10.50%, and 4.35%, RMSE to decrease by 28.35%, 15.54%, 3.84%, and R2 to rise by 37.73%, 14.63%, and 2.01%, respectively, when compared to the BP, LSTM, and cnn-LSTM models. It can be seen that the AM-BiLSTM model constructed in this paper has the highest prediction accuracy, and its value basically fluctuates in a small range around the true value.

Table 1. Evaluation indicators

Models	MAPE	RMSE	MAE	R^2
BP	3.3646	0.3434	0.2794	0.6109
CNN-LSTM	1.4320	0.0983	0.0873	0.9681
LSTM	0.6458	0.2153	0.1488	0.8419
AM-BiLSTM	0.7851	0.0599	0.0438	0.9882

4. Conclusions

This paper combines the advantages of BiLSTM model, introduces attention mechanism, and constructs a stock prediction model based on BiLSTM and attention mechanism. Using the time sliding window, the research object was set as the model input and output, taking into account the daily opening, closing, highest, trading volume, and other 13 basic index data of Ping An stock market data from September 02, 2022 to March 03, 2023. Compared with the prediction results of BP model, LSTM model, CNN-LSTM model and AM-BiLSTM model, the proposed model shows superior performance.

References

[1] Siami-Namini S, Tavakoli N, Namin A S. The performance of LSTM and BiLSTM in forecasting time series[C]//2019 IEEE International Conference on Big Data (Big Data). IEEE, 2019: 3285-3292.

[2] Niu Z, Zhong G, Yu H. A review on the attention mechanism of deep learning[J]. Neurocomputing, 2021, 452: 48-62.

[3] Wu Xiaofei. A Stock Price Foresting Using LSTM Based on Attention Mechanism[P]. 2022 International Conference on Economics, Smart Finance and Contemporary Trade (ESFCT 2022),2022.

[4] Kim S, Kang M. Financial series prediction using Attention LSTM[J]. arXiv preprint arXiv:1902.10877, 2019.

[5] Wei D. Prediction of stock price based on LSTM neural network[C]//2019 international conference on artificial intelligence and advanced manufacturing (AIAM). IEEE, 2019: 544-547.

[6] Liu S, Liao G, Ding Y. Stock transaction prediction modeling and analysis based on LSTM[C]//2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA). IEEE, 2018: 2787-2790.

[7] Yang M, Wang J. Adaptability of Financial Time Series prediction based on bilstm[J]. Procedia Computer Science, 2022, 199: 18-25.

[8] Wang Haiyao,Wang Jianxuan,Cao Lihui,Li Yifan,Sun Qiuhong,Wang Jingyang. A Stock Closing Price Prediction Model Based on CNN-BiSLSTM[J]. COMPLEXITY,2021,2021.

[9] Nath, Basab,Sarkar, Sunita,Das, Surajeet,Mukhopadhyay, Somnath. Neural machine translation for Indian language pair using hybrid attention mechanism[J]. Innovations in Systems and Software Engineering,2022(prepublish).

[10] Keren He,Qian Jiang. Research on stock prediction algorithm based on CNN and LSTM[J]. Academic Journal of Computing & Information Science,2022,5(12).

[11] Brim Andrew,Flann Nicholas S. Deep reinforcement learning stock market trading, utilizing a CNN with candlestick images.[J]. PloS one,2022,17(2).

[12] Jiyao Wei,Jian Liao,Zhenfei Yang,Suge Wang,Qiang Zhao. BiLSTM with Multi-Polarity Orthogonal Attention for Implicit Sentiment Analysis[J]. Neurocomputing,2020,383(C).