Analysis and Research on Stock Price of LSTM and Bidirectional LSTM Neural Network

Mingzhu Jia [a], Jian Huang [b], Lihua Pang [c] and Qian Zhao [d]

College of Communication & information engineering, Xi'an University of Science & Technology, Xi'an 710054, China

[a]2447823795@qq.com, [b]huangjian_work@163.com, [c]lhpang.xidian@gmail.com,[d]52156950@qq.com

Abstract. The stock market has the characteristics of large fluctuations and high dimensions, and can be regarded as a nonlinear time series system, so the traditional time series method is not applicable. For the study of deep learning methods, this paper proposed a bidirectional long-short term memory (BLSTM) neural network. Compared with ARIMA model and LSTM neural network, the (BLSTM) neural network is used to predict the accuracy of GREE stock price. Firstly, the stock data is normalized and pre-processing, and then the processed data is input into one-way and two-way LSTM respectively. In the neural network, Dropout is used as an optimization term to prevent over-fitting of the network, and then the three evaluation criteria of RMSE, MAE and Loss are selected to comprehensively analyze the rationality of a single bidirectional LSTM neural network. The experimental results show that the RMSE and MAE are reduced by about 24.2% and 19.4% respectively, and the deviation accuracy is increased by 0.13%. The network complexity is not high, and the error is effectively reduced. It can provide reference value for short-term market investors.

Keywords: Bidirectional long-short term memory network; over-fitting; long-short term memory network; deviation error; stock price.

1. Introduction

The stock market is compared to the barometer of the national economy, so the rapid rise and fall of the stock market cannot be ignored[1]. With the rapid development of artificial intelligence and natural language processing, we urgently need to use the Internet to mine useful and reliable information and analyze and predict stock price movements. Stock data is a system with time series characteristics, but because of its large fluctuation, high dimension and nonlinear, neural network machine learning is better than traditional statistics and econometric prediction method.

In recent years, there have been machine learning methods for predicting stocks. Kim et al. proposed support vector machine SVM for stock price prediction[2], but the shortcomings of SVM algorithm only give the second-class classification algorithm, but in the practical application of data mining. Generally, it is necessary to solve the classification problem of many types.

In 2006, Professor Hinton of the University of Toronto first proposed the concept of deep learning[3]. He used an unsupervised learning strategy called layer-by-layer training to extract more abstract and representative features from raw data. The neural network called Deep Belief Network can be effectively trained. Later, Professor Hinton and other scholars proposed a method to solve the neural network over-fitting Dropout method, and proved that layer-by-layer training can effectively improve the generalization ability of deep network[4]; Chen et al. analyzed the impact of using LSTM network on prediction results with different number of characteristics and different data preprocessing conditions[5]; Nakhaei, a foreign scholar, Amirkabir University of Technology Fardis and Irannajad Mehdi used the LSTM neural network for the estimation of ore grade and recovery rate[6], which achieved good results.

Hammd, M. alhaj Ali and L. Hall used multi-layer BP neural network to analyze the stock price of Jordan stock market[7]. BP neural network has the characteristics of strong prediction ability and high accuracy, but traditional BP neural network is easy to fall into the problem of local minimum.

Hu Xinchen of Harbin University of Technology applied LSTM neural network deep learning technology to the task of semantic relation classification, and found that the neural network is more suitable for processing text sequence data[8].

In 2005, Graves first applied LSTM neural network model to classification problem[9]. Subsequently, the model was extended to various tasks of natural language processing; Graves applied it to speech recognition[10]in 2013; these applications have achieved good results; Literature[11]separately input text and data into the neural network to generate abstract features, and then sent them to LSTM for combination prediction; however, LSTM neural network has the shortcoming of insufficient information utilization.

Therefore, a new neural network algorithm, bidirectional LSTM, is proposed for time series stock forecasting. By comparing and demonstrating the accuracy of one-way and two-way LSTM neural network, Dropout method is added to prevent over-fitting. Three evaluation criteria are used to analyze the prediction error. The results show that the two-way LSTM neural network has better prediction effect, which provides a reference method for investors and effectively avoids bear market.

2. Algorithm Description

2.1 Long-term and Short-term Cyclic Neural Network LSTM

The LSTM neural network (Long-short-term memory neural network) model is a variant of RNN. Its network structure uses gates-control mechanism consisting of a forgotten gate, an input gate, a cell state, and an output gate. It has a long-term memory function[12], and its network structure is shown in Figure 1.

x_t :it indicates an the input of the current node; C_t :represents the candidate value vector in the calculation process; σ represents the feedforward sigmoid activation function of the network layer; tanh represents the tanh activation function of the feedforward network layer; the following detailed description of the working principle of each gate.

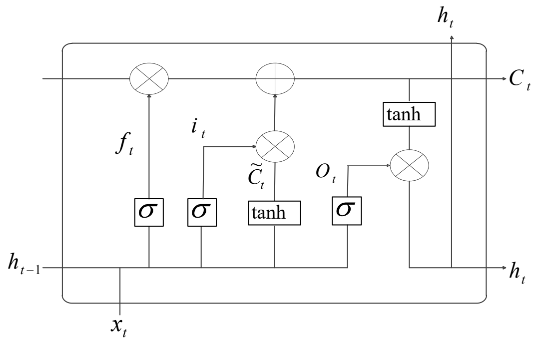

Fig.1. Long-term and short-term cyclic neural network LSTM

$\otimes$: Represents matrix point-by-point multiplication; $\oplus$:represents matrix addition;

First calculate the input gate, the formula is as follows:

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

Secondly, the calculation time T input candidate cell state value $\tilde{C}_t$, and the formula is as follows:

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{2}$$

Then calculate the activation function of the forgotten gate, the formula is as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{3}$$

According to the above calculation, the cell t state update value at time t is obtained, and the formula is as follows:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{4}$$

Finally, the value of the output gate is obtained. The formula is as follows:

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

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

2.2 The Principle and Structure of Bidirectional Long-short Term Memory Network

2.2.1 Bidirectional LSTM Principle

In the Forward layer, the forward calculation is performed from 1 moment to t moment, and the output of the forward hidden layer at each time is obtained and saved. In the Backward layer, the calculation is reversed along the time t to the time 1 to obtain and save the output of the backward hidden layer at each time. The six unique weights are repeatedly used in each time step, and the six weights are respectively used. Correspondence: Input to the forward and backward hidden layers (w1, w3), hidden layer to the hidden layer itself (w2, w5), forward and backward hidden layers to the output layer (w4, w6). Finally, at each moment, the final output is obtained by combining the output of the Forward layer and the Backward layer, as shown in Figure 2 is a bidirectional LSTM network diagram [13]. The mathematical expressions are as follows:

$$h_t = f(w_1 x_t + w_2 h_{t-1}) \tag{7}$$

$$h_t^{'} = f(w_3 x_t + w_5 h_{t+1}^{'}) \tag{8}$$

$$o_t = g(w_4 h_4 + w_6 h_t^{'}) \tag{9}$$

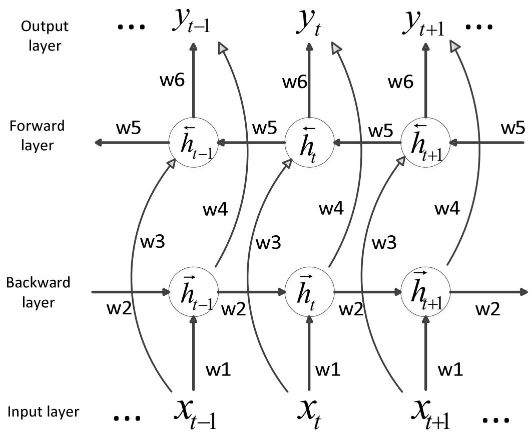

Fig.2. Bidirectional LSTM network diagram

2.2.2 Bidirectional LSTM Structure

The BLSTM uses the pre-processed data as input, passes through the forward and backward LSTM neural network layer, then goes to the full connection layer, and outputs the prediction result[13], as shown in Figure 3. In Figure 3, the weights and biases initialization of the full connection layer are all based on a random normal distribution.

BLSTM is a variant of recurrent neural network, which solves the long-term dependence of RNN and LSTM. It combines LSTM in two different directions and extracts forward and reverse information data at the same time. Stocks are data with strong time series characteristics. Selecting cycling neural network can make better use of historical information. Compared with LSTM, BLSTM can simultaneously make use of temporal information in both directions, so it is easier to mine potential unused data [14].

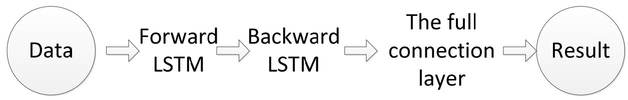

Fig.3. Bidirectional LSTM structure diagram

3. Data Analysis

Data collection: Stock data can be regarded as time series data. Take the financial data GREE stock 000651 as an example. Get 568 days from January 1, 2017 to May 14, 2019 through ts.get_hist_data in python tushare. there are 14 characteristics, namely: date, open, high, close, low, volume, price_change, p_change, ma5, ma10, v_ma5, v_ma10, v_ma20. Since there are many factors affecting stocks, the daily closing price close is selected as the benchmark research object, in which the first 364 sets of data are selected as the training set, and the last 204 sets of data are the verification set. The closing price of GREE stocks is shown in Figure 4.

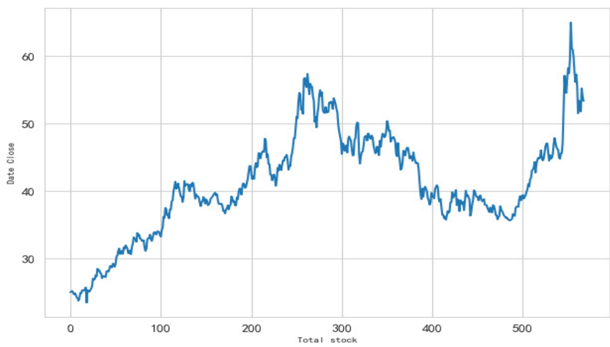

Fig.4. Daily closing price of GREE stock

Data preprocessing: Since stock data is a nonlinear time series with large fluctuations and high dimensionality, it is no longer applicable to use traditional linear processing data. in view of the different characteristics of stock data, normalization and standardization are carried out.

4. Experiment Analysis

4.1 Experimental Environment

The environment is configured as KINGSTON SA400S37480G solid state drive, processor AMD A12-9700P RADEON R7, 10 COMPUTE CORES 4C+6G 2.50 GHz, operating system is Windows 10, 64-bit operating system, x64-based processor, and use Python 3.6.3 Interpreter (64-bit), the network framework is TensorFlow 1.13.1.

4.2 Judging Criteria

The evaluation criteria for stock time series prediction use root mean square error (RMSE), mean absolute error (MAE), and deviation accuracy (Loss).

RMSE: The square root of the ratio of the square sum of the observed and true deviations to the number m of observations. it is used to measure the deviation between the predicted value and the true value. The smaller the value, the more accurate it is.

MAE: The average of the absolute errors. It can better reflect the actual situation of the predicted value error. The smaller the value, the more accurate it is. The calculation formula is as follows.

Loss: Deviation accuracy. Used to measure the degree of dispersion of the data itself.

4.3 Experimental Results

In view of the different structures of unidirectional LSTM and bi-directional LSTM neural networks, the stock price prediction of GREE is analyzed and studied respectively.

In the machine learning model of Dropout, if the parameters of the model are too many and the training samples are too few, the trained model is prone to over-fitting. When training neural networks, we often encounter the problem of fitting. Over-fitting is manifested in the following: the loss function of the model is smaller on the training data, and the prediction accuracy is higher; but the loss function is relatively larger on the test data, and the prediction accuracy rate is low[16].

Dropout can effectively alleviate the occurrence of over-fitting and achieve regularization to some extent.

The i and ii experiments respectively analyzed the experimental results of unidirectional LSTM and bi-directional LSTM in forecasting stock price:

i: One-way LSTM. 364 stock closing prices from January 1, 2017 to June 30, 2018 were selected as training sets, and 204 stock closing prices from July 1, 2018 to May 14, 2019 were selected as validation sets for network training. Dropout is added to prevent over-fitting. Hidden layer unit = 50, iteration number epochs = 100, optimizer = adam. Figure 5 (left) shows the forecast of closing price of GREE stock without Dropout. Table 1 shows the corresponding RMSE, MAE and Loss values when Dropout = 0.2, 0.5 and 0.8.

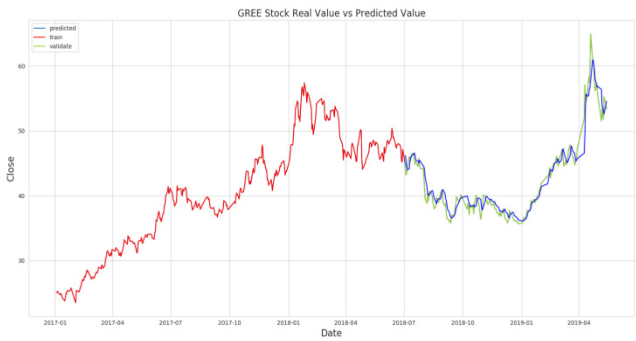

Fig.5.One-way LSTM without Dropout predicts stock closing price(left)

Table 1. Dropout=0.2, 0.5, 0.8. RMSE, MAE and Loss

	Dropout	RMSE	MAE	LOSS
One-way LSTM	no	1.3824	0.9355	0.0007
	0.2	1.6592	1.1413	0.0017
	0.5	1.7314	1.1825	0.0023
	0.8	2.1057	1.3818	0.0031

It can be seen from Table 1 that the accuracy of RMSE and Loss is better than that of No Dropout, but there may be over-fitting; when adding Dropout=0.2, 0.5, and 0.8, it is verified that when Dropout = 0.2, The root mean square error is the smallest and the prediction effect is the best.

ii: Bidirectional LSTM. The data training method is the same as the one-way LSTM neural network, which combines two different directions of LSTM to simultaneously extract the forward and reverse information of the data and output the predicted results together at the output layer. Figure 6 (left) shows the forecast of the closing price of GREE stock under the condition of no Dropout; Table 2 shows the corresponding RMSE, MAE and Loss values when Dropout=0.2, 0.5, and 0.8.

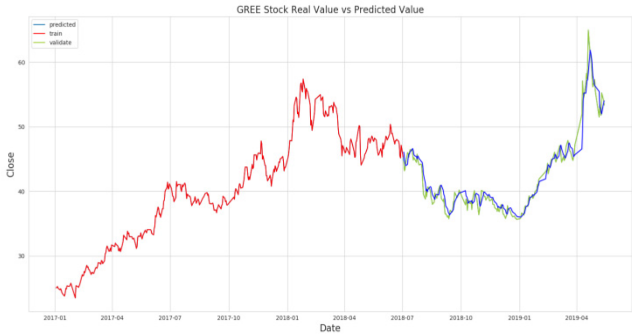

Fig.6.Two-way LSTM without Dropout predicts stock closing price(left)

Table 2. Dropout=0.2, 0.5, 0.8. RMSE, MAE and Loss

	Dropout	RMSE	MAE	LOSS
Two-way LSTM	no	1.3382	0.8565	0.0008
	0.2	1.4082	0.9398	0.0015
	0.5	1.5515	1.0008	0.0022
	0.8	1.8105	1.1834	0.0030

After theoretical research and experimental analysis, compared with experiments i and ii, it is found that the two-way LSTM prediction accuracy is higher and the curve coincidence is better. From Table 1 and Table 2, it can be seen that the RMSE and MAE of bidirectional LSTM are decreased by an average of 24.2% and 19.4%, respectively, and the prediction accuracy of bidirectional LSTM is improved by about 0.13% compared with unidirectional LSTM. In addition, comparing Dropout with or without Dropout, it is found that when Dropout = 0.2, the accuracy is higher, and the RMSE error value is at least 1.4082, and the MAE error value is 0.9398. It shows that the prediction accuracy of bidirectional LSTM neural network is higher than that of unidirectional LSTM neural network, and the root mean square error is smaller.

5. Conclusion

There are many factors affecting the stock market, and the stock data has the characteristics of large fluctuation and non-linearity. The traditional linear time series processing method is not applicable. in this paper, unidirectional LSTM and bidirectional LSTM neural networks in deep learning are used for comparative demonstration. After two groups of comparative experiments, the conclusions are as follows.(a) Deep Neural Network is easy to fall into the state of over-fitting, and Dropout is added to ensure the accuracy of data to prevent over-fitting and achieve good prediction results; (b) Bidirectional LSTM uses the combination of forward LSTM and reverse LSTM to predict, fully obtain historical data and future data for analysis, and the prediction degree is more accurate; (c) The evaluation criteria are RMSE, MAE and Loss. Factor comprehensive analysis, the accuracy is better. Deficiencies: The verification process is carried out in a very ideal environment, without considering the actual transaction fees, national policies and other factors, so there is still a big gap with the actual situation. In the future, we will consider the admission fees and changes in the international situation.

References

[1]. Yu Kai, Jia Lei, Chen Yuqiang, Xu Wei. Yesterday, Today and Tomorrow of Deep Learning [J]. Computer Research and Development, 2013, 50 (09): 1799-1804.

[2]. KIM K J. Financial time series forecasting using support vector machines. [J]. Neurocomputing. 2003,55(1/2): 307- 319.DOI:10.1016/S0925-2312(03)00372-2.

[3]. Hinton G E, Osindero S, Teh Y W. A Fast Learning Algorithm for Deep Belief Nets[J]. Neural Computation, 2014, 18(7):1527-1554.

[4]. Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: A Simple Way to Prevent Neural Networks from Overfitting[J]. Journal of Machine Learning Research, 2014, 15(1):1929-1958.

[5]. CHEN Kai, ZHOU Yi, DAI Fangyan. A LSTM-based method for stock returns prediction: A case study of China stock market[C] //IEEE International Conference on Big Data. Santa Clara: IEEE Press,2015:2823-2824. DOI :10.1109/BigData.2015.7364089.

[6]. Nakhaei Fardis, Irannajad Mehdi. Application and comparison of RNN, RBFNN and MNLR approaches on prediction of flotation column performance[J]. International Journal of Mining Science and Technology,2015,25(06):983-990.

[7]. Graves,A.,Schmidhuber, J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures[J]. Neural Networks, 2005, 18(5-6). IJCNN '05. Proceedings. 2005 IEEE International Joint Conference on. [10.1109/IJCNN.2005.1556215].

[8]. Martin Wöllmer, Metallinou A, Eyben F, et al. Context-sensitive multimodal emotion recognition from speech and facial expression using bidirectional LSTM modeling[C]// INTERSPEECH 2010, 11th Annual Conference of the International Speech Communication Association, Makuhari, Chiba, Japan, September 26-30, 2010. DBLP, 2010.

[9]. Graves A. Supervised Sequence Labelling with Recurrent Neural Networks[J]. Studies in Computational Intelligence, Springer, 2008, 385.

[10]. Wang P, Xu B, Xu J, et al. Semantic Expansion using Word Embedding Clustering and Convolutional Neural Network for Improving Short Text Classification[J]. Neurocomputing, 2016, 174(PB):806-814.

[11]. Akita R, Yoshihara A, Matsubara T, et al. Deep learning for stock prediction using numerical and textual information[C]// IEEE/ACIS International Conference on Computer & Information Science. IEEE, 2016.

[12]. Wang Jinping. Time series research and application based on financial data [D]. University of Electronic Science and Technology of China, 2017.

[13]. Gao LL, Guo Z, Zhang HW, et al. Video captioning with attention-based LSTM and semantic consistency. IEEE Transactions on Multimedia, 2017, 19(9): 2045–2055. [doi: 10. 1109 /TMM.2017.2729019].

[14]. HUANG Qiu-ping, ZHOU Xia, GAN Yu-jian, WEI Yu. Application of SVM and neural network model in stock forecasting[J].Microcomputer & Applications,2015,34(05):88-90.

[15]. Chen Bin, Zhou Yong, Liu Bing. Event triggered word extraction based on convolutional two-way long-term and short-term memory network [J]. Computer Engineering, 2019, 45 (01): 153-158.

[16]. WANG Xin, WU Ji, LIU Chao, YANG Haiyan, DU Yanli, NIU Wensheng. Fault time series prediction based on LSTM cyclic neural network[J].Journal of Beijing University of Aeronautics and Astronautics,2018,44(04):772-784 .