Stock Market Prediction Using Deep Learning Based on Modified Long Short-Term Memory

Wenxuan Li[1], Meiying Huang[1](✉), and Yangqiu Pi[2]

[1] Finance and Economics, Qinghai University, Xining, China
1145953334@qq.com
[2] China Merchants Securities Co. Ltd., Shanghai, China
piyangqiu@cmschina.com

Abstract. The stock market is a key factor in financial field. It is affected by the current trends and other market factors. The stock market prediction can provide precise information for investment and maximize investors interests. Nowadays, with the development of artificial intelligence, there is an increasing trend of using intelligent technique to predict stocks' tendency, which is the main part of quantitative investment. The techniques applied to stock prediction include convolutional neural network (CNN), support vector machine (SVM) and other techniques. However, the performance of these methods are poor when dealing with the time series data. Therefore, we proposed a framework based on modified long short term memory (LSTM). In order to evaluate the effectiveness of proposed method, other mainstream methods are applied in comparative experiments. The results of the experiments reveal that the proposed method has higher prediction accuracy.

Keywords: Stock Market Prediction · Deep Learning · Data Processing · Long Short Term Memory

1 Introduction

The stock market has always been one of the most popular investments because of its high profits. According to the demand of investment and trading increases, people begin to look for tools and methods that can both increase returns and reduce risks [8]. In addition, stock market prediction is a classical problem in the intersection of finance and computer science [6]. The previous approaches are tending to measure different levels of efficiency in mature and emerging markets [4] and establish effective prediction models [5].

Therefore, based on computer science, many prediction models have been proposed to predict stock markets. Like SVR, BP, LSTM and other models. In this paper, we mainly discuss about SVR, BP and LSTM. Firstly, SVR is widely applied to stock market prediction, which is a very helpful method if we don't have much idea about the data. And it can be used to predict discrete values. However, due to the current situation of big date, we need to deal with the numerous date, but SVR does not perform very

Z. Zeng et al. (Eds.): ECIT 2022, AHE 11, pp. 522–530, 2023.
https://doi.org/10.2991/978-94-6463-005-3_52

well when the data set has more information [1]. When the number of features of each data point exceeds the number of training data samples, the training results of SVR performance not very good. Secondly, BP is also a popular model which is frequently used in stock market's prediction. By this algorithm, weights and biases are corrected by gradient descent, in order to minimize detected errors to ensure the model is ready to make predictions. However, it has high requirements for data sets, because the actual performance of BP training on specific problems depends on the input data, and in data mining, BP algorithm is particularly sensitive to the data with noise number [11]. Therefore, this method puts forward higher requirements for the accuracy and simplicity of the data, which increased the difficulty in data selection and processing.

In this paper we proposed a framework based on modify LSTM to predict the stock market, which are explicitly designed to avoid the long-term dependency problem. In order to verify the effectiveness of the proposed method, the main methods as SVR and BP. The result shows that the proposed method has better performance on realistic stock data. The main contributions of this paper are:

1) feature extraction model. The modified LSTM based method is very suitable for processing highly time series related problems, such as stock market prediction, dialogue generation, encoding, etc.;
2) Stock market prediction estimation. A modified LSTM is a highly time series related method which improved for the regression task to increase the prediction accuracy, in the prediction of stock market.;
3) Validation on realistic data set. The data sets contain realistic stock data of a listed company in the past 20 years from 1991 to 2020. On the one hand, we conducted the task of randomly extracting data, in order to prove the rationality of the experimental data.

The rest of this paper is organized as follows. In Sect. 1, an overall framework of the proposed method and mathematical expressions of the used algorithms are given. In Sect. 2, the details of the stock market data set, including total hand, turnover, maximum price, etc. are described. And the results of stock market prediction and comparison results with other methods are also shown in Sect. 3. Finally, Sect. 4 concludes this paper.

2 Methods

In order to show the structure of the work, this section includes the detailed descriptions of the overall framework, the mathematical expressions of data processing, modified LSTM and performance evaluation metrics, which are listed as follows.

2.1 SVR

SVR is a statistical regression method for structural risk minimization, originally proposed by Drucker et al. [2]. This method can achieve high generalization ability and prediction accuracy under the premise of low model complexity in the case of limited samples. This is a supervised learning technique based on Vapnik's concept of support vectors. SVR [10] aims to reduce errors through determining the hyperplane and minimizing the range between predicted and observed values. The values in the following formula to minimize ω are similar to those defined to maximize margin, as shown below:

$$\min\|\omega\|^2 + C\sum_{i}^{n}\left(\zeta_i^+ + \zeta_i^-\right) \tag{1}$$

where the summation part represents an empirical error.

Hence, to minimize this error, we use the following equation.

$$g(x) = \sum_{i}^{n}\left(\alpha_i^* + \alpha_i\right)K(x, x_i) + B \tag{2}$$

The α term represents the Lagrange multiplier, the value of which is greater than or equal to K represents the kernel function, and β is the bias term. In this paper, we used a polynomial kernel: each element in the extracted features is mapped by a nonlinear activation function called Leaky ReLU, which is expressed as:

$$K(x, x_i) = \gamma(x * x_i + 1)^d \tag{3}$$

where d is the polynomial degree and is the polynomial constant.

2.2 BP

It is a standard form of artificial network training, which helps to calculate gradient loss function with respect to all weights in the network. The idea of BP algorithm is that the learning process includes two processes, signal forward propagation and error back propagation. If the actual output of the output layer does not match the target expectation, the error will be entered. The back-propagation of the network adjusts the weights of neurons in each layer. This process continues until the training network has a present number of learning times or is less than the error value output by the network. At present, the most used three-layer BP network [7, 9], including input layer, hidden layer and output layer. Set up a three-layer BP network, the input vector is $X = \left(x_1, \cdots, x_j, \cdots, x_m\right)^T$, the output vector of the hidden layer is $Y = \left(y_1, \cdots, y_j, \cdots, y_m\right)^T$, the layer output vector is $O = (o_1, o_2, \cdots, o_k, \cdots, o_l)^T$, and the expected output vector is $d = (d_1, d_2, \cdots d_k, \cdots, d_l)^T$, the weight between the input layer and the hidden layer is

$V = (V_1, V_2, \cdots, V_J, \cdots, V_M)$, the weight between the hidden layer and the output layer is

$W = (W_1, W_2, \cdots W_k, \cdots, W_l)$.

(1) The output of the hidden layer in the forward propagation process can be shown as:

$$y_i = f\left(\sum_{i=0}^{n} v_{ij}x_i\right) j = 1, 2, \cdots, m \tag{4}$$

Among them, $x_0 = -1$ is set by the threshold value introduced by the hidden layer neurons.

The output of the output layer can be expressed as:

$$o_k = f\left(\sum_{j=0}^{m} w_{ij} y_i\right) k = 1, 2, \cdots, l \tag{5}$$

Among them, $y_0 = -1$ is set by the threshold value introduced by the output layer neurons.

(2) The weight modification of the output layer in the backpropagation process can be expressed as:

$$\begin{aligned} &\Delta_{wjk} = \eta(d_k - 0_k)o_k(1 - 0_k)y_i \\ &k = 1, 2, \cdots, l, m; j = 0, 1, \cdots, m \end{aligned} \tag{6}$$

The weight modification of the hidden layer can be expressed as:

$$\Delta v_{jk} = \eta\left(\sum_{k=1}^{l} \delta_k^0 w_{ik}\right) y_i(1 - y_i)x_i \tag{7}$$

Among $\begin{aligned} &\delta_k^0 = (d_k - o_k)o_k(1 - o_k) \\ &k = 1, 2, \cdots, l; i = 0, 1, \cdots, n \end{aligned}$;η is the study rate.

2.3 LSTM

Long Short-term Memory Neural Network (LongShort-term Memory Net-works, LSTM) is an advanced RNN [3], a sequential network, that allows information to persist. is a special type of RNN that can learn long-term dependencies information. RNN neural network models have been widely used in language recognition and text classification and many other research fields. Compared with artificial neural network model (ANN). In other words, the RNN neural network model can recycle the weight parameters of neurons, it can well apply information related to historical data to forecasting. However, the error back-propagation algorithm of the RNN neural network model is just like the ANN neural network. As simple as in the network model, the reuse of weights can bring benefits, but also there are great disadvantages, such as gradient explosion and gradient disappearance problems, that is, the impact of historical data. The problem of long-term dependencies cannot be effectively solved. To solve these two problems, machine learning. The researchers in the study area have developed a long- and short-term memory neural network model.

Compared with the RNN model, the most obvious improvement of the LSTM model is an increase of 1 a cell state C and three valves, the three valves are the forget gate f, the output gate o and input gate i. When the LSTM model error back propagates to correct the weights, some error can be passed directly to the next layer of neurons through the input gate. Some errors can be forgotten through the forgetting gate, which solves the problem of gradient explosion and the problem of disappearing is to effectively deal with the redundancy of relevant information in historical data, etc. question.

2.4 Evaluation Metrics

In this paper we adopt three metrics to evaluate the prediction performance of stock market for each transfer record, which were also used in: 1) The mean absolute error (MAE) described in Eq. (8) is used to reflect the average absolute errors of stock market prediction of each test part; 2) The mean square error (MSE) described in Eq. (9) is used to reflect the difference between the predicted stock market and actual stock market of each test; 3) The mean absolute percentage error (MAPE) described in Eq. (10) is used to evaluate the ratio of stock market prediction error to actual prediction value of each test.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\left(\hat{y}_i - y_i\right)\right| \tag{8}$$

$$MSE = \frac{1}{n}\sum_{i}^{n}\left(Y_i - \widehat{Y}_i\right)^2 \tag{9}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\left(y_i - \hat{y}_i\right)}{y_i}\right| \times 100\% \tag{10}$$

3 Stock Market Prediction

For this part we are going to demonstrate the experiment process.

3.1 Data Set Description

The data is from a listed company, which includes the data from 1991 to 2020. The data has 10 indexes, consisting of opening price; closing price; maximum price; lowest price; up and down; stock amplitude; total hand; turnover; price and trading volume.

The opening price is the beginning transaction price for one share of a certain security after the trading days' opening of the stock exchange. Most of the world's stock exchanges. Most of the world's stock exchanges use the principle of maximum turnover to determine the opening price. Closing price is the ending price of the stock market, which is the volume-weighted average price of all transactions one minute before the last transaction of the security on that day. It is often used as a reference point by investors to compare a stock's performance and closing price from the previous day, and is often used to construct a line graph describing historical prices over time. Maximum price is the highest value of all trading prices of shares on the day. Lowest price is the lowest value of all trading prices of stocks on that day. Up and down, which through the closing price of the stock price for two consecutive trading days, it is judged whether the stock price has risen or fallen sharply. When the latest closing price is higher than the previous closing price, it shows the possibility of growth, otherwise it represent the possibility of fall. Stock amplitude is the absolute value of the difference between the highest price and the lowest price of the day after the stock opens and the percentage of yesterday's closing price. It shows the activity of the stock to a certain extent. Total Hand is the volume which from the opening to the instant situation. Turnover is the description of the purchase or sale of an equal share of futures from one person to another. Price is the total price of all the transactions. Trading volume is a generic term for the number of transactions of a security or contract during a specific period of time.

3.2 Data Preparation

The data have many different evaluation indicators, often having different dimensional and dimensional units, which will affect the results of data analysis. The existence of singular sample data will increase the training time, and may also lead to the failure of convergence. Therefore, the preprocessed data should be normalized before the training, so that the pre-processed data can be limited within a certain range.

Then, in order to analysis the data better, we reorder the data, which are changed to Opening price, Maximum price, Lowest price, Up and down, Stock Amplitude, Total Hand, Price, Turnover, Trading Volume and Closing price, as shown in Table 1. In addition, processing data in order to transform raw data into the form which suitable for LSTM network's inputting.

The next step is to preprocess the data, including windowing and dividing the training set and test set.

After that, the training set and test set are divided, part of the data are selected for model training, and the latter data are used for model testing. Specifically, we use high and low opening close and volume data to predict the next closing price data set.

3.3 Compared Methods

3.3.1 SVR Parameter Setting

In this method, we initialize the SVR radial basis kernel function, which is an important step in SVR.

Table 1. The stock market prediction data sets

No.	Indexes					
	Opening price	Maximum Price	Lowest Price	Up and down	Stock Amplitude	Total Hand
3320	0.707387	0.709322	0.713040	0.473203	0.175506	0.228606
3321	0.736287	0.732364	0.739406	0.421501	0.152542	0.220065
3322	0.730262	0.748318	0.740632	0.404161	0.250957	0.207867
3323	0.758560	0.750680	0.762707	0.380517	0.119738	0.124489
3324	0.734481	0.730592	0.743085	0.311791	0.110990	0.126211

3.3.2 BP Parameter Setting

The epoch and batch size are equally important in neural network. An Epoch refers to inputting all data into the network to complete a forward calculation and backpropagation. Batch is the part that each input network is trained on, and batch size is the number of training samples in each batch. In this method, we set 'Epoch' to 50, and set 'batch-size' to 256.

3.4 Proposed Method

Similar to BP, in our proposed method, we set Epoch to 50, and set 'batch-size' to 1, To achieve an optimal balance between memory capacity and memory efficiency to optimize the performance and speed of the network model. And the process of this work is shown as Fig. 1.

3.5 Comparison and Discussion

As shown in Table 2, we also compare the proposed method with other two domain adaptation methods which are commonly used in prediction field, like BP, SVR, where a MAPE metric is used for evaluating the performance.

The overall comparison results of different methods are visualized in Fig. 2. In the results view, we can know that the proposed method performs better than others.

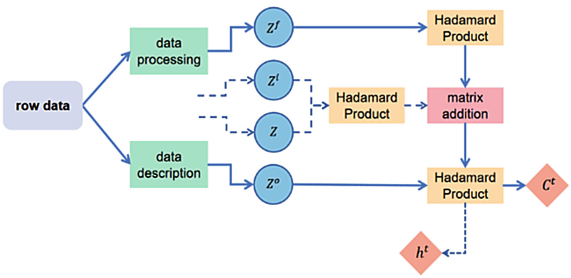

Fig. 1. Stock market prediction process based on modified LSTM.

Table 2. MAPE of stock market prediction using different domain adaptation methods.

Metrics	Methods		
	LSTM	BP	SVR
MAE	0.07525658463756	0.02566025574007	0.02613915264320
MSE	0.00707350497421	0.00096164570377	0.00107980412088
MAPE	9.28422799251	3.275612194305	4.198882162732

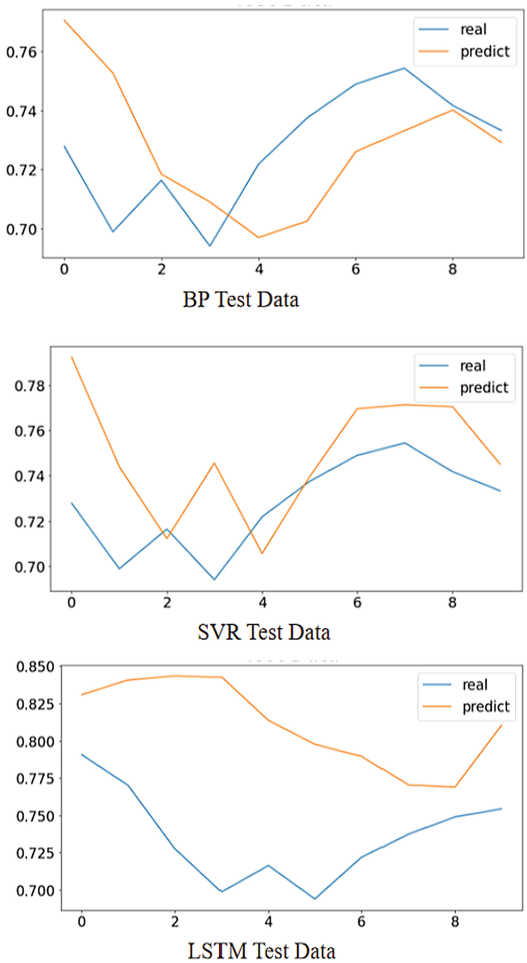

Fig. 2. Comparisons among different prediction methods.

4 Conclusions

In this paper, we propose a modified LSTM method for stock market prediction. The experimental results show that LSTM model has the best performance among all model data. The experimental results show that the model has high prediction accuracy and the ability to predict the extreme value accurately. However, the stock market changes overtime, and the data we used is the past data, we cannot rely solely on predictions based on past data. Therefore, in future work of stock market prediction, we should consider the influences at different view as much as possible, to make the prediction more suitable for investors to use.

References

1. Bathla G (2020) Stock price prediction using LSTM and SVR. 2020 Sixth international conference on parallel, distributed and grid computing (PDGC). IEEE, pp 211–214
2. Drucker H, Drucker H, Burges CJC, Kaufman L, CCJ KB (1996) Support vector regression machines advances in neural information processing systems. 9
3. Fischer T, Krauss C (2018) Deep learning with long short-term memory networks for financial market predictions. Eur J Oper Res 270(2):654–669
4. Gandhmal DP, Kumar K (2019) Systematic analysis and review of stock market prediction techniques. Comput Sci Rev 34:100190
5. Hiransha M, Gopalakrishnan EA, Menon VK, Soman KP (2018) NSE stock market prediction using deep-learning models. Procedia Comput Sci 132:1351–1362
6. Jiang W (2021) Applications of deep learning in stock market prediction: recent progress. Exp Syst Appl 184:115537
7. Misra D (2019) Mish: a self regularized non-monotonic activation function. arXiv preprint arXiv:1908.08681
8. Nivetha RY, Dhaya C (2017) Developing a prediction model for stock analysis. In 2017 international conference on technical advancements in computers and communications (ICTACC). IEEE, pp 1–3
9. Ramachandran P, Zoph B, Le QV Searching for activation functions (2017). arXiv preprint arXiv:1710.05941
10. Ruta D, Cen L, Liu M, Vu QH (2021) Automated feature engineering for prediction of victories in online computer games. In 2021 IEEE international conference on big data (big data). IEEE, pp 5672–5678
11. Zhang L, Wang F, Xu B, Chi W, Wang Q, Sun T (2017) Prediction of stock prices based on LM-BP neural network and the estimation of overfitting point by RDCI. Neural Comput Appl 30(5):1425–1444. https://doi.org/10.1007/s00521-017-3296-x