Stock Price Prediction based on Long Short-Term Memory Model

Ziji Liu

South China Agricultural University, Guangzhou 510642, China

202025320116@stu.scau.edu.cn

Abstract. The stock market is a place that brings investors opportunity to gain profit. Meanwhile, the stock market also brings investors high risks, which requires machine learning methods to improve the accuracy of prediction. This paper uses the Long Short-Term Memory (LSTM) model to predict stock prices. In the study, the daily stock historical data of Sinopec, Moutai and SPD Bank in the past 21 years are used as samples, including the date, the trade volume, the highest price, the lowest price, and the opening and closing prices. After the LSTM model has been trained, the three companies' predictions of opening price have achieved good results, and the predicted opening price curve and the actual opening price curve seem to be quite consistent. In terms of evaluation indicators, MAPE of the three companies is less than 1%. These results can provide some help for investors to predict stocks and shed light on guiding further studies.

Keywords: Machine Learning; Long Short-Term Memory; Stock Market; Price Prediction.

1. Introduction

The stock market is a place for listed companies to trade stocks. Listed companies give part of their ownership to investors in exchange for part of their capital, which will help companies. In simple terms, the stock market is the gathering place of buyers and sellers [1]. The stock market plays an important role. Research shows that the stock market is closely related to economic growth, and the development of the stock market can promote economic growth [2]. In addition, the stock market can attract more capital to the company and expand its business and productivity. For investors, it also gives investors a good opportunity to make money, which can be achieved by selling stocks [3].

However, with the high returns of stocks, the stock market also brings huge risks to investors. The stock market is nonlinear and unpredictable. In other words, the stock market may disappoint investors while giving them hope. The stock market will be affected by many factors, such as corporate debt [4], politics, war. Among them, the war has a great impact on the stock market. According to research, the negative effect of World War II on the British stock market was very obvious. Even though there were some positive things to stimulate the market at that time, the negative effect was still greater. Meanwhile, by analyzing the daily data of 94 countries' stock returns during the Russian Ukrainian war, the war would have a huge negative effect on the world's stock indexes [5, 6]. In general, the stock market is high-risk and affected by many factors. Therefore, investment in stocks should be cautious and well prepared.

Many researchers obtain important information about the stock market through computer related knowledge and economics, make reasonable predictions, and finally obtain huge wealth. In an epoch when computer technology has become very powerful, investors should make good use of computers and use advanced analytical and prediction techniques to enhance the possibility of making money. In addition, forecasting the stock market can help them invest more confidently by considering risks and fluctuations [1, 7]. Previously, researchers have proposed many models and algorithms that can analyze stock market trends through machine learning methods, e.g., linear regression [8], neural networks. However, according to the research, using neural network to predict the stock market is more effective than linear regression algorithm. The neural network algorithm is executed in parallel and has higher fault tolerance [9]. SVM algorithm is a popular algorithm. In terms of hit rate, the stock market trend prediction method based on SVM has better generalization ability than traditional methods [10]. Besides, according to research, this algorithm performs well in predicting the future

trend of stocks. The long-term prediction accuracy based on random forest algorithm is between 85% and 95% [11]. Moreover, decision tree algorithm is also a good choice for forecasting stocks. In one study, the prediction algorithm based on decision tree and rough set had a high accuracy rate of 90%, while the accuracy rate based on artificial neural network was only 77.66%. This method can be selected for stock market forecasting [12].

This article will introduce another popular machine learning algorithm LSTM and use it to predict stocks, providing investors with more options to analyze the stock market. The rest of the article is structured as follows. The second part describes the data used in this article, the principle of the model and the selection of parameters, as well as the final indicators to evaluate the effect of the model. The third part describes the experimental results and the possible significance of this study. The fourth part states the limitations of this study and prospects for the future. The fifth part summarizes the whole article.

2. Data & Method

2.1 Data

The historical stock data of the three companies used in this paper are obtained from Shanghai Stock Exchange through the Tushare interface. The stock data used here are from October 1, 2001, to October 1, 2022, including date, opening price, closing price, highest price, lowest price and trading volume. In addition, these three companies are Sinopec, Kweichow Moutai and Shanghai Pudong Development Bank, all of which are Chinese companies.

The following is the background of the three companies. Sinopec is a large petrochemical enterprise group established in July 1998. The left panel of Fig. 1 shows the historical opening price of Sinopec. Kweichow Moutai is a company that produces Moutai liquor. Moutai liquor is a product of national geographical indication in China and one of the three largest distilled liquors in the world. The middle panel of Fig. 1 shows the company's historical opening price. Shanghai Pudong Development Bank opened on January 9, 1993, which is liked by people for its reputation of integrity and good performance. It is also one of the six major commercial banks in China. Similarly, its historical opening price is presented in right panel.

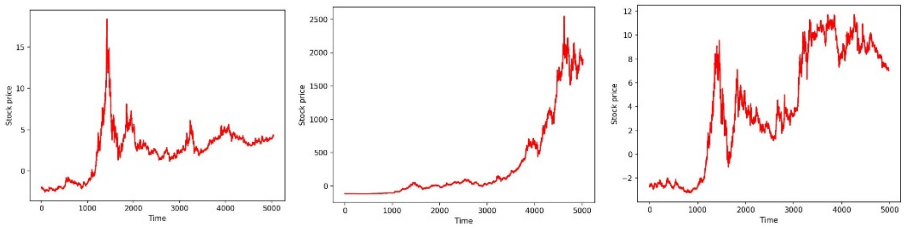


Fig 1. The opening price of Sinopec, Moutai and SPD Bank (from the left to the right).

2.2 Models

This article adopts the LSTM model. A unique RNN model called LSTM is employed to address the issue of gradient explosion and disappearance during lengthy sequence training. LSTM outperforms traditional RNN in longer sequences. Similar to the traditional neural network model, it has three layers, the input layer, the hidden layer and the output layer.

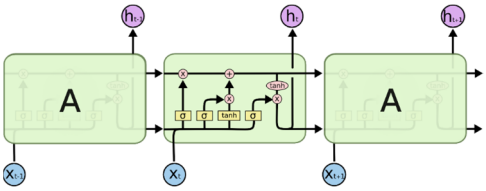


Fig 2. A sketch of LSTM [13].

Table 1. Selection and value of parameters.

Parameter	Value
Hidden layers size	128
LSTM layers	2
Dropout rate	0.2
Time step	30
Batch size	64
Learning rate	0.001
Epoch	20

The structure diagram of LSTM model is as follows (seen from Fig. 2). The forget gate, input gate, and output gate are the three gates in this model. The initial step of LSTM is to choose which data to remove from the cell state. Forget gate can complete this step. The forget gate will read the last output(h_{t-1}) and the current input (x_t), activate the function through the sigmoid to output a vector (the value of each dimension of the vector is between 0 and 1, 1 means completely reserved, 0 means completely discarded, which is equivalent to remembering the important and forgetting the unimportant), and finally multiply with the cell state. Choosing how much additional data to add to the cell status is the second stage. The input gate's sigmoid layer chooses which data needs to be updated, and the tanh layer creates a vector. Afterwards, the newly updated information should be multiplied by the newly obtained vector to play the role of information filtering and add it to the information filtered by the forgetting gate above. Finally, one needs to decide which portion of the cell state will be output, a sigmoid layer will be executed. The tanh processes the cell state (resulting in a value between - 1 and 1) and multiplied by the sigmoid layer output to produce the result, h_t. Model parameters are summarized in Table 1. Moreover, the anticipated opening price is produced using the input features of the opening price, closing price, highest price, lowest price, and trading volume. The training set, which comprises 95% of the data, is used, and the test set, which comprises 5% of the data.

2.3 Metrics

In order to evaluate the fitting effect of LSTM model on the historical opening prices of the three companies, The paper will introduce four indicators: MSE, RMSE, MAE, and MAPE. The MSE can be expressed as:

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

Here, y_i represents the real opening price, $\hat{y}$ represents the predicted opening price, and n represents the total days predicted in the test set. RMSE is calculated as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)^2} \quad (2)$$

The meaning of other variables is the same as that of the first formula, except that a root sign is added outside. The MAE can be obtained as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Y_i - \widehat{Y}_i| \quad (3)$$

Where N represents the total number of days in the test set.

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\left(Y_i - \widehat{Y}_i\right)/Y_i\right| \quad (4)$$

Similarly, the test set's overall length is indicated by the letter n. The difference between this and MAE is that each item has to be divided by a y_i, and then multiplied by 100%.

3. Results & Discussion

The subject of this study is the prediction of stocks by LSTM model. Finally, the prediction results of the three companies are as follows: Sinopec (given in Fig. 3), Moutai (seen from Fig. 4) and SPD Bank (depicted in Fig.5). From Fig. 3, one can visually obtain some basic information about the

Sinopec's test set. The red line represents the original stock opening price, and the blue line represents the stock opening price predicted by LSTM. Then, the predicted maximum opening price is close to 4.4, the lowest price is about 3.6, and the data days are about 250 days. In the same way, the Fig. 4 gives us similar information. The red line represents the real opening price of Moutai, while the blue line represents the predicted opening price. The highest price exceeds 2100, the lowest price is about 1600, and the total number of days is close to 250. Finally, Fig. 5 also clearly shows the basic information of SPD Bank. For example, the red line is the original opening price, and the blue line is the predicted opening price. The predicted maximum price is slightly lower than 9, the minimum price is about 7, and the total number of days is still close to 250.

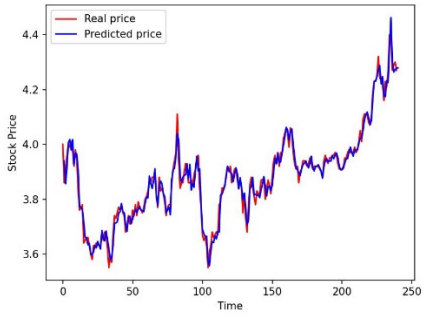


Fig 3. The opening price forecast results of Sinopec.

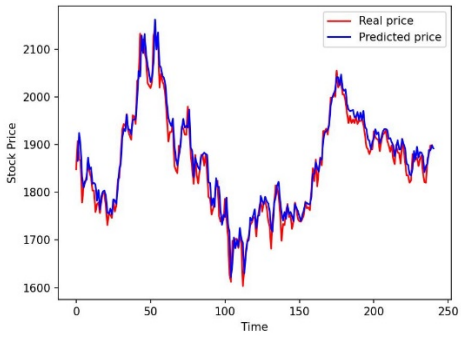


Fig 4. The opening price forecast results of Moutai.

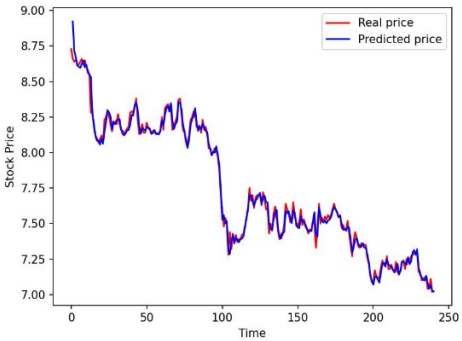


Fig 5. The opening price forecast results of SPD Bank.

The Table. 2 will show a more detailed evaluation of these three companies. From the results, there are four indicators. The calculation of these four indicators has been explained above, and only the most noteworthy points will be mentioned here. It is not difficult to find that the MSE, RMSE and MAE of Moutai are much higher than those of the other two companies, because the opening price of Moutai is much higher than that of the other two companies. Therefore, when the real data of the three companies are relatively matched with the predicted data, the three indicators of Moutai will be much higher. In addition, Table 2 also gives the MAPE of the three companies. The MAPE of Sinopec is 0.55%, that of Kweichow Moutai is 0.89%, and that of SPD Bank is 0.47%. The prediction results

of the three companies are very good. SPD Bank has the best prediction effect, while Moutai has the worst prediction effect. This may be due to the difference in the number of days. Although the selected data are from October 1, 2001, to October 1, 2022, the stock will sometimes be suspended for some reasons, so the total number of days of the three companies is actually different. In addition, this difference may also be caused by the same LSTM model parameters selected by the three companies, and the prediction effect of Moutai may be better by fine-tuning the time step and other parameters.

The results of this study may have a certain role and value for stock speculators. The basic information and historical trend of stocks can be obtained quickly through this study. It is known from Table 2 that the prediction error of the three companies is less than 1%, which is a very accurate model. The general direction of the stock can be predicted through this model, which may reduce the risk of mistakes or failures in stock speculators' decision-making. This may provide more options for stock speculators, who can use their own knowledge and combine this model to make a final decision to reduce the risk of stock speculation and improve the possibility of making money, when analyzing the rise and fall of stocks.

Table 2. Evaluation index results of three companies after using LSTM.

Company	MSE	RMSE	MAE	MAPE
Sinopec	0.0009	0.0292	0.0213	0.55%
Moutai	449.6324	21.2045	16.5969	0.89%
SPD Bank	0.0023	0.0484	0.0359	0.47%

4. Limitations & Prospects

The disadvantage of this study is that if the number of samples is too large, the accuracy of model prediction will be reduced, and the error will become larger. This problem can be found by referring to the prediction results of Moutai Company. The sample size of Moutai Company is larger than that of the other two companies. At the same time, Mean Absolute Percentage Error is also significantly larger than the other two companies. These results can also provide investors with more stock information for reference and analysis, such as volatility, Moving Average Convergence and Divergence. In addition, the accuracy of the model can also be further improved. Stocks are affected by many factors including politics, the company's financial situation, war, natural disasters, etc. Nevertheless, the input characteristics of this article do not consider the factors that may affect the stock, and only take the opening price, closing price and trading volume of the stock as the input characteristics. For future work, research can increase the stock k-line chart, volatility, MACD and other information. For LSTM model, its parameters should be further adjusted to improve the accuracy and the ability to process large sample data. For example, the time step and the number of LSTM layers will have a great impact on the results. When inputting the model features, more features that have an impact on stocks should be input, such as policy features and features that obtain information from financial news.

5. Conclusion

In summary, the aim of the study e is to forecast the stock price based on LSTM model. In this study, three companies, Sinopec, Moutai and SPD Bank, were selected to evaluate the prediction effect of LSTM with their historical stock opening prices. The results show that the LSTM model has an excellent prediction effect on the three companies, and the Mean Absolute Percentage Error is less than 1%. It indicates that LSTM has a good effect on forecasting the stock market and is very suitable for forecasting stock prices. However, the limitation of this paper is that the effect is not so good when the number of samples is too large, and only the basic information of stocks (e.g., opening price and closing price) is included in the input characteristics. In the future, parameters can be adjusted,

and more features can be considered to increase accuracy. These results provide investors with a method to predict stock prices and offer a guideline for stick price prediction based on LSTM.

References

[1] Murkute, A., Sarode, T.: Forecasting market price of stock using artificial neural network. International Journal of Computer Applications, 124(12), 11-15 (2015).

[2] Masoud, N. M.: The impact of stock market performance upon economic growth. International Journal of Economics and Financial Issues, 3(4), 788-798 (2013).

[3] Fauzi, R., Wahyudi, I.: The effect of firm and stock characteristics on stock returns: Stock market crash analysis. The Journal of Finance and Data Science, 2(2), 112-124 (2016).

[4] Tan, C., Yan, Y..: Does Default Risk of Corporate Debt Affect Stock Liquidity Journal of Jinan: Philosophy and Social Sciences Edition, 42 (9), 84-102 (2020).

[5] Hudson, R., Urquhart, A.: War and stock markets: The effect of World War Two on the British stock market. International Review of Financial Analysis, 40, 166-177 (2015).

[6] Boungou, W., Yatié, A.: The impact of the Ukraine–Russia war on world stock market returns. Economics Letters, 215, 110516 (2022).

[7] Gharehchopogh, F. S., Bonab, T. H., Khaze, S. R.: A linear regression approach to prediction of stock market trading volume: a case study. International Journal of Managing Value and Supply Chains, 4(3), 25 (2013).

[8] Seber, G. A., Lee, A. J.: Linear regression analysis. John Wiley & Sons, 2012.

[9] Ahangar, R. G., Yahyazadehfar, M., Pournaghshband, H.: The comparison of methods artificial neural network with linear regression using specific variables for prediction stock price in Tehran stock exchange. arXiv preprint arXiv:1003.1457, (2010).

[10] Lin, Y., Guo, H., Hu, J.: An SVM-based approach for stock market trend prediction. In The 2013 international joint conference on neural networks (IJCNN) pp. 1-7 (2013, August).

[11] Khaidem, L., Saha, S., Dey, S. R.: Predicting the direction of stock market prices using random forest. arXiv preprint arXiv:1605.00003 (2016).

[12] Nair, B. B., Mohandas, V. P., Sakthivel, N. R.: A decision tree-rough set hybrid system for stock market trend prediction. International Journal of Computer Applications, 6(9), 1-6 (2010).

[13] Sophic Chen. Blogs from CSDN. Retrieved from: https://blog.csdn. net/ sophicchen/ article/ details/ 1033 06545.