A Stock Price Foresting Using LSTM Based on Attention Mechanism

Xiaofei Wu(✉)

Minzu University of China, Beijing, China
xfeiww@163.com

Abstract. Stock price prediction has been a hit subject in recent decades. Many researchers find different methods to predict stock price. LSTM is an excellent variant model of RNN, but single LSTM can only process a single form of data and lacks the ability to process multiple mixed forms of data. Considering that stocks represent the financial market, the exchange rate would have a particular impact on the financial market, so rate change affects stock price movement. Therefore, attention mechanism could introduce exchange rate into LSTM, so we produce a hybrid LSTM module based on attention mechanism to predict stock price. We find that the RMSE and MSE of hybrid LSTM are lower than others.

Keywords: Stock price prediction · LSTM · Attention mechanism · Rate change

1 Introduction

Since economy has grown by leaps and bounds in recent years, stock market has become the main place for investors to accumulate idle funds. Investors need to take huge risks because the stock market is volatile, and the data generated in stock market is enormous, highly nonlinear and complex. Besides, there are many factors that would affect investors' understanding and judgment of the stock market, such as company's operations, information asymmetry, policy change, market environment and considerable customer manipulation [1]. So it is difficult for stockholders to predict stock price. Many researches establish quantitative models based on mathematical statistical method, machine learning and artificial intelligence. Although there are many methods developed to forecast stock price, these methods have certain defects. The traditional mathematical statistical method is unsuitable for forecasting because stock price is highly nonlinear and random. Therefore, deep learning method is introduced in prediction. LSTM is an excellent variant model of RNN which inherits the characteristics of RNN models. Many pieces of research show that the currency of LSTM is better than other single deep learning modules, such as ARIMA, GRACH [2, 3]. However, with the growth of time series, LSTM cannot deal with the long series data, so research combines models to predict stock and find that the hybrid model is more accurate [4].

The relationship between stock price movement and rate change has been researched for many years. Bahmani-Oskooee M [5] show that the relationship between S & P 500 index and the USD exchange rate is Granger causality. Granger [6] find that the

F. Balli et al. (Eds.): ESFCT 2022, AEBMR 226, pp. 1467–1476, 2022.
https://doi.org/10.2991/978-94-6463-052-7_162

relationship is different in 9 countries. Therefore, it is clear that rate change has an influence on stock price movement.

Because rate change has an influence on stock price movement, we use attention mechanism to introduce rate change into LSTM of stock prediction. Attention mechanism is an optimized mechanism imitating people's attention proposed by Bahdanau D [7]. It can select some key information input for processing so that the network can grasp the important learning characteristics in the training process to improve the accuracy. Therefore, this paper is going to build an attention-based LSTM model which could employ the weight coefficients of LSTM output and rate change. We take the closing index of the Shanghai Securities Composite Index and the Shenzhen Securities Component Index as the stock characteristic index. Considering these indexes as the first tier LSTM input, we use the first tier LSTM output and RMB rate as attention tier input to find a nonlinear relationship between variables.

2 Related Works

2.1 Stock Prediction

The method of stock price forecasting is mainly divided into two methods: mathematical statistical method and machine learning model. Many research studies use Generalized Autoregressive Conditional Heteroskedasticity (GARCH), Auto Regressive Integrated Moving Average (ARIMA) and Autoregressive Conditional Heteroscedasticity (ARCH) as the common method to forecast stock price [8, 9]. Basel M.A. Awartani [10] examine the predictive ability of different GARCH models through S&P-500 stock index, with particular emphasis forecast outcome of the asymmetric component. Machine learning is a common method of predicting stock prices by learning historical data. Various deep neural network architectures are used in stock predicting and quantitative investment [11]. White [12] first used the back propagation neural network to treat multivariate nonlinear non-parametric data and predict daily rate of return of the IBM company. Jiawei Long [13] used convolutional neural network to make financial investment decisions. Bengio [14] pointed out that traditional RNNs have problems with vanishing gradients so it is difficult to capture long-term dependency. LSTM neural network is a special kind of RNN. D. M. Q. Nelson finds LSTM model is more accurate than others by comparing the accuracy of LSTM and other single models. Although the LSTM model can deal with time series more effectively than other single models, its effect becomes lower with time-series growth. So varieties of studies design the combined model based on LSTM. Kim [15] combined LSTM with one to three GRACH-type models and found the performance of the hybrid model is better. Varieties of studies [1, 16] show that the composite LSTM model is more accurate than the single model because single LSTM is difficult to solve long sequences. Thus, we combine the LSTM model with other tricks, the attention mechanisms, to better deal with the problem.

2.2 Attention Mechanism

Treisman and Gelade [17] propose the theory of attention mechanism, an optimized mechanism imitating people's attention. Dzmitry Bahdanau [18] built a new RNN model based on attention assigning calculating weights on different hidden outputs corresponding to different words. Assigning attention weights on neural networks has achieved great success in various machine learning tasks [16]. So the attention mechanism changes the conventional encoder-decoder structure and the traditional decoder gives the same weight to each input.

Ma C K [19] showed two possible influences of rate change on stock price movement. Muhammad Kamran KHAN [20] indicated the exchange rate has an influence on stock Returns based on the autoregressive distributed lag model. Bahmani-Oskooee M [5] show the connection between exchange rate and stock price is dual causal in a short term. So this paper considers rate change as another input and uses attention-based LSTM model to predict stock.

This section introduces LSTM which is one of machine learning models, attention mechanism and the impact of rate change on stock price. Considering that rate change would affect stock price, we input two factors into LSTM model and then calculate the weight coefficient of each factor through attention mechanism. The conclusion is drawn by comparing RMSE and MSE of LSTM and the attention based LSTM.

3 Methods

Many pieces of research show exchange rate could affect stock price [5,119,200], but single LSTM cannot handle multiple forms of information. In order to combine the rate change into LSTM and ensure that our model can adequately handle both forms of data, this paper proposes an LSTM model based on attention to predict stock price movement. Specifically, we use stock price series and exchange rate price series of corresponding dates as input and use multiplicative attention to calculate the weight coefficients, then output the predicted value. Figure 1 shows the proposed framework containing several units.

We use the min-max scaling method to preprocess two different data sequences, then input processed data of stock price and rate change into LSTM. Next, taking the output of LSTM as the input of attention to calculating weight coefficient, after that we got the final result.

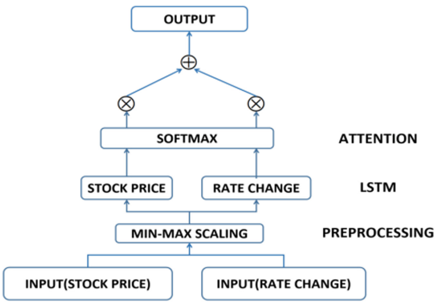


Fig. 1. Illustration of the hybrid module.

3.1 Dataset

Since the sample of the Shanghai Securities Composite Index (Stock code:000001) is all the stocks in Shanghai Stock Exchange, it can represent the range of all the stock trends in the Shanghai Stock Exchange. Although the sample of the Shenzhen Securities Component Index (Stock code:3990010) is the most representative of 40 stocks, it can still represent the stock trend of the whole Shenzhen Stock Exchange. So we collect stock price data from the closing index of Shanghai Securities Composite Index and the Shenzhen Securities Component Index from NetEase Finance and commensurable RMB exchange rate from Sina Finance.

3.2 Data Preprocessing

The pretreated dataset is stored in the CSV file. Due to the different dimensions of the sample data, it is necessary to normalize data to improve the experiment's reliability.

Min-max Scaling: The data is processed through linear transformation, and the data of different orders of magnitude are mapped to the [0, 1] interval. The formulation of min-max scaling is shown in Eq. (1).

$$X_{\text{new}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{1}$$

Where X is the original data, X_{new} is the normalized data, $X_{\max}$ and $X_{\min}$ is the maximum and minimum values of the original data, respectively.

3.3 Attention-Based LSTM

A single LSTM module consists of three gates, Input gate, Output gate and Forget gate. The input gate decides which information to add to the memory state. The output gate determines the output according to the input and memory unit. Forget gate decides which data needs to be deleted. The internal memory unit adds candidate memory status information. Equation (2) indicates the output result of LSTM layer.

$$H_t = LSTM\,(X_{\text{t}}, H_{t-1}) \tag{2}$$

Where H_t is the output at time t and X_{t} is the input at time t, H_{t-1} is the output at time t − 1.

To calculate the corresponding weight coefficient of stock price and exchange rate, we use attention layer to make it. The input of the attention layer is the output of the LSTM. First, it is encoded by the encoder, then the corresponding weight of the encoding vector is calculated by attention (α_1, α_2). The calculated weight is used to weight the coding vector as the input of the decoder and finally get the output through the decoder. Through continuous learning and optimization of the corresponding weights, the critical information in the input features is highlighted, and the nonlinear features between variables are further mined. We choose multiplicative attention to calculate attention score. Equation (3) of the similarity of the attention layer is shown as below.

$$e_i = s^T W h_i \tag{3}$$

While s^T is the query vector, W is the weight matrix, h_i is the values vector.

Then input e_i into the softmax function for normalization and get the normalized weight α_j as Eq. (4).

$$\alpha_i = \frac{\exp(e_i^T u_w)}{\sum_{i=1}^{T} \exp(e_i^T u_w)} \tag{4}$$

Where u_w is randomly initialized attention weight matrix.

Then we take the weighted sum to get the final output as Eq. (5).

$$a = \sum_{i=1}^{T} a_i h_i \tag{5}$$

3.4 Evaluation Index

To ensure the consistency of data, we chose 80% of the data as the training set and 20% of the data as the test set. Within 728 trading days, 588 days were used as training data and 140 days as test data. In order to evaluate the accuracy of the model, we choose Root Mean Squared Error (RMSE) and Mean Squared Error (MSE) as evaluation criteria. Equation (6) and (7) of the calculation formula of RMSE and MSE are shown as below.

$$RMSE = \sqrt{\frac{1}{n}\sum\nolimits_{i=1}^{T}(X_{pre} - X_{act})^2} \tag{6}$$

$$MSE = \frac{1}{n}\sum\nolimits_{i=1}^{T}(X_{pre} - X_{act})^2 \tag{7}$$

Where X_{pre} is the predicted data and X_{act} is the actual data.

4 Discussion

Considering that stock price movement is affected by rate change and single LSTM module is challenging to process long data sequences, we make a hybrid LSTM module based on attention to improve the above problems. Attention is used in this hybrid module, which could introduce rate change as another model characteristic to improve the accuracy of forecasting.

We first import sequence of stock price index and RMB exchange rate into LSTM layer. Then taking the output of LSTM layer as the input of attention layer to calculate the corresponding weight coefficient. Finally, the model output the prediction results. To verify the performance of the hybrid LSTM, we set the parameters of LSTM and hybrid LSTM to be consistent to maintain the variables. Parameters settings are shown in Table 1.

Table 2 shows the normalized weight α_j of two sequences output of LSTM in attention layer of the Shanghai Securities Composite Index and the Shenzhen Securities Composite Index.

Table 1. The setting of model parameters

Learning rate	0.01
Optimization function	Adam
Activate functions	ReLU
Number of hidden layer neurons	16

Table 2. The weight coefficient of stock price and rate change

	Shanghai Securities Composite Index	Shenzhen Securities Composite Index
The weight of stock price	0.504298	0.467301
The weight of rate change	0.495702	0.532699

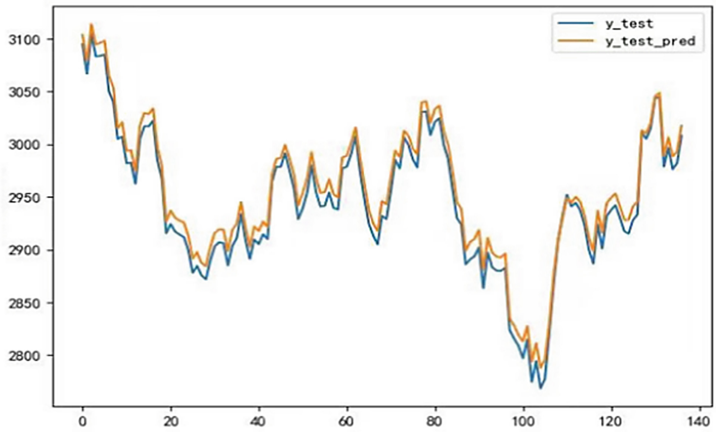
Figure 2 shows the comparison between the predicted results of hybrid LSTM and the actual data on the test set of the Shanghai Securities Composite Index and the Shenzhen Securities Composite Index.

Similar to the results of the majority of papers [15, 16], the composite LSTM model can improve the prediction accuracy by inputting new stock characteristics compared with single LSTM. During the experiment, each model was tested 10 times on the same training suite and test suite to ensure the experiment's objectivity, and the average value of RMSE was taken as the final result of the model. The experimental results of RMSE and MSE are shown in Table 3 and Table 4.

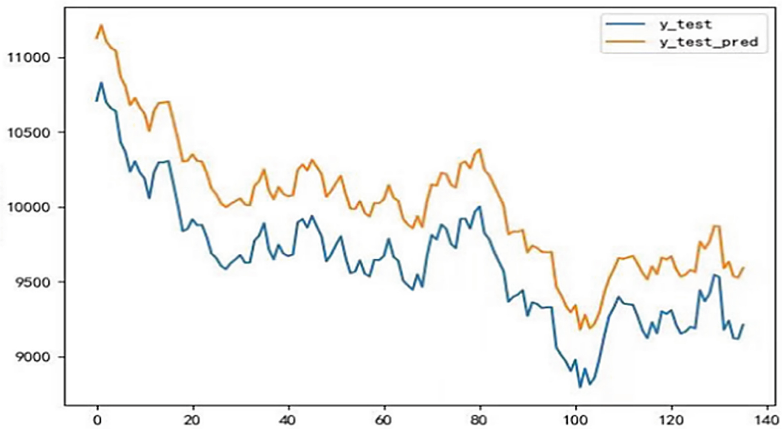
It is clear that the RMSE and MSE composite LSTM is less than that of single LSTM, which means the prediction result of hybrid LSTM model is closer to the real data. Therefore, the LSTM model based attention mechanism is more accurate than the single LSTM model.

The exchange rate could affect the import and export prices of relevant enterprises, which will have a certain positive correlation with the level of foreign trade competitiveness of these enterprises. Once the price changes, the operating performance and the income of the enterprises participating in international trade will continue to be volatile, which will affect the stock price of these companies. Indeed, the results show that introducing exchange rate into the model features can effectively improve the accuracy.

However, not only does exchange rate affect stock price fluctuation, but also many other factors affect it, such as market conditions. So only introducing the exchange rate into the model can improve the accuracy a little and have specific limitations. In the future, other factors which would affect stock price could be considered as the model input and test such more complex models. By adding more kinds of data sequences, the accuracy can be improved more effectively.



(a)



(b)

Fig. 2. The comparison between the predicted results of hybrid LSTM and the actual data on the test set of two indexes. (a) Shanghai Securities Composite Index. (b) Shenzhen Securities Composite Index.

Table 3. The RMSE results of different methods

module	Shanghai Securities Composite Index	Shenzhen Securities Composite Index
Single LSTM module	0.33	0.18
LSTM module based on attention mechanism	0.03	0.168

Table 4. The MSE results of different methods

module	Shanghai Securities Composite Index	Shenzhen Securities Composite Index
Single LSTM module	0.106	0.033
LSTM module based on attention mechanism	0.001	0.0284

5 Conclusion

This paper proposes an LSTM module based on attention mechanism and use the Shanghai Securities Composite index and the Shenzhen Securities Composite index as the training data in LSTM module. Then input the stock price and exchange rate into the attention layer, calculate the weight coefficient and finally get the predicted value. By comparing the result values of RMSE and MSE, we find that the RMSE and MSE of the composite model are smaller than that of single LSTM, which means that the prediction result of hybrid LSTM model is closer to the actual data. Therefore, it can be concluded that the LSTM based on attention mechanism is more accurate than single model. Considering that there are varieties of factors which affect the stock price, we can introduce these elements to establish more complex forecasting models in the future.

References

1. C. Xiao, W. Xia, J. Jiang, Stock price forecast based on combined model of ARI-MA-LS-SVM, in: Neural Computing and Applications, 2020, 32(10): 5379-5388.
2. A.H. Bukhari, MA.Z. Raja, M.Sulaiman, Fractional neuro-sequential ARFIMA-LSTM for financial market forecasting, in: Ieee Access, 2020, 8: 71326-71338.
3. C.N. Babu, B.E. Reddy, Selected Indian stock predictions using a hybrid ARIMA-GARCH model, in: 2014 International Conference on Advances in Electronics Computers and Communications. IEEE, 2014: 1-6.
4. S. Selvin, R. Vinayakumar, E. A. Gopalakrishnan, et al, Stock price prediction using LSTM, RNN and CNN-sliding window model, in: 2017 international conference on advances in computing, communications and informatics (icacci). IEEE, 2017: 1643-1647.
5. M. Bahmani-Oskooee, A. Sohrabian, Stock prices and the effective exchange rate of the dollar, in: Applied economics, 1992, 24(4): 459-464.

6. C.W.J. Granger, Essays in econometrics: collected papers of Clive WJ Granger, in: Cambridge University Press, 2001.
7. D. Bahdanau, K. Cho, Y. Bengio, Neural machine translation by jointly learning to align and translate, in: arXiv preprint arXiv:1409.0473, 2014.
8. P.F. Pai, C.S. Lin, A hybrid ARIMA and support vector machines model in stock price forecasting, in: Omega, 2005, 33(6): 497-505.
9. R.D. Brooks, R.W. Faff, McKenzie M D, et al, A multi-country study of power ARCH models and national stock market returns, in: Journal of International money and Finance, 2000, 19(3): 377-397.
10. B.M.A. Awartani, V. Corradi, Predicting the volatility of the S&P-500 stock index via GARCH models: the role of asymmetries[J], in: International Journal of forecasting, 2005, 21(1): 167-183.
11. A. Thakkar, K. Chaudhari, A comprehensive survey on deep neural networks for stock market: the need, challenges, and future directions, in: Expert Systems with Applications, 2021, 177: 114800.
12. H. White, Economic prediction using neural networks: the case of IBM daily stock returns, in: IEEE 1988 International Conference on Neural Networks, 1988, pp. 451-458 vol.2, doi: https://doi.org/10.1109/ICNN.1988.23959
13. J. Long, Z. Chen, W. He, An integrated framework of deep learning and knowledge graph for prediction of stock price trend: An application in Chinese stock exchange market, in: Applied Soft Computing, 2020, 91: 106205.
14. Y. Bengio, P. Simard and P. Frasconi, Learning long-term dependencies with gradient descent is difficult, in: IEEE Transactions on Neural Networks, vol. 5, no. 2, pp. 157-166, March 1994, doi: https://doi.org/10.1109/72.279181.
15. H.Y, Kim, C. H. Won, Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models, in: Expert Systems with Applications, 2018, 103: 25-37.
16. H. Li, Y. Shen, Y.Zhu, Stock price prediction using attention-based multi-input LSTM, in: Asian conference on machine learning. PMLR, 2018: 454-469.
17. A.M. Treisman, G.Gelade, A feature-integration theory of attention, in: Cognitive psychology, 1980, 12(1): 97-136.
18. D. Bahdanau, K. Cho and Y.S. Bengio, Neural machine translation by jointly learning to align and translate. CoRR, abs/1409.0473, 2014.
19. C.K. Ma, G.W. Kao, On exchange rate changes and stock price reactions, in: Journal of Business Finance & Accounting, 1990, 17(3): 441-449.
20. M.K. Khan, Impact of exchange rate on stock returns in Shenzhen stock exchange: Analysis through ARDL approach, in: International Journal of economics and management, 2019, 1(2): 15-26.