Proceedings of the 4th International Conference on Mathematical Physics and Computational Simulation
DOI: 10.54254/2753-8818/2026.35018

Multi-step Stock Price Forecasting Based on an LSTM-Attention Hybrid Model

Shuheng Wang

International Business School, Xi'an Jiaotong-Liverpool University, Suzhou, China
wangshuheng25@163.com

Abstract. The stock price sequence has a low signal-to-noise ratio, strong nonlinearity and statistical characteristics drift over time. Once it is necessary to continuously predict multiple moments in the future, the error will continue to accumulate along the step length, making the difficulty of prediction significantly increased. In response to this problem, this paper embeds the attention mechanism into the long-term and short-term memory network to construct an LSTM-Attention hybrid model for multi-step prediction: first, LSTM encodes the timing dependence of historical sequences, and then assigns learningable weights to the hidden state of each time step through the attention layer, so as to highlight the key points that are more distinguishable for future trends. Finally, the full connection layer gives the price of several future moments at once. The experiment selects the daily frequency data of the Shanghai-Shenzhen 300 Index from 2015 to 2024 and compares it horizontally with models such as ARIMA, LSTM, GRU, CNN-LSTM and Transformer. The results show that the proposed model leads the overall accuracy of RMSE, MAE, MAPE and direction, and the root mean square error decreases by about 20% compared with the single LSTM; and with the predicted step length increases, the attention mechanism can significantly inhibit the accumulation of error. The average absolute percentage error of 10-step prediction is about 31% lower than that of LSTM, indicating that the hybrid structure has a good effect in multi-step prediction tasks.

Keywords: stock price forecasting, multi-step prediction, LSTM, attention mechanism, deep learning, time series

1. Introduction

In the field of financial engineering and quantitative investment, stock price forecasting has always been a fundamental topic of great concern. Reliable price prediction not only helps investors improve asset allocation and prevent risks, but also provides reference for market supervision and macro research and judgment. However, stock prices are also driven by multiple factors such as macroeconomics, policy orientation and market sentiment. Its sequence often fluctuates violently, is strongly nonlinear and lacks stability. In addition, the effective market hypothesis itself is controversial about the predictability of prices, making it difficult to achieve high-precision forecasting in the long run.

DOI: 10.54254/2753-8818/2026.35018

The early work mainly used statistical models such as ARIMA and GARCH, as well as traditional machine learning methods such as support vector regression to portray stock prices. However, the former is usually based on stability, while the latter relies heavily on the characteristics of artificial design, and it is difficult for both of them to fully express the complex nonlinear timing relationship within the price sequence. With the outstanding ability of deep learning in automatic feature extraction and nonlinear fitting, its application in financial time series prediction is becoming more and more extensive [1]. It has been reviewed and systematically summarized the evolutionary vein and application results of relevant methods [2, 3]. Among many structures, the long-term and short-term memory network (LSTM) has effectively alleviated the gradient disappearance of the cyclic neural network with the gate control design, and has gradually become a common scheme for timing modeling.

However, when processing long input, a single LSTM has limited ability to identify critical moments, and far-end information is also easy to decay; at the same time, most studies only stop at single-step forecasts, and real transaction decisions often need to make judgments about the direction of the future. Because the error of multi-step prediction will gradually increase with the step length, higher requirements are put forward for the robustness of the model. The attention mechanism provides ideas for breaking through the above bottlenecks by learning different weights in different time steps and guiding the model to pay attention to historical fragments with more information.

Based on the above analysis, this paper combines LSTM with the attention mechanism for multi-step prediction of stock prices: relying on the timing of the LSTM coding sequence, the multi-step results are output at one time after weighting the attention layer, and comparing the real transaction data of the Shanghai-Shenzhen 300 index with a variety of baseline model systems to test its performance in multi-step prediction accuracy and error control.

2. Literature review

2.1. Traditional methods and machine learning methods

In the early study of stock price prediction, statistical models dominate: ARIMA and its variants describe the sequence evolution with linear autoregression, and the GARCH series is mainly used to describe volatility. Kim and Won combined LSTM with multiple GARCH models to predict stock index fluctuations, which is more effective than simple statistical models [4]. After that, support vector regression, random forest and other methods introduced nonlinear modeling capabilities, but they still require a large number of artificial structural features, and the generalization is weak in the face of high-noise and unstable price sequences.

2.2. Deep learning method based on LSTM

LSTM relies on input gates, forgetting gates and output gates to regulate information flow, which can better capture long-term timing dependence [5] and has become the basic component of financial sequence modeling. Bao and others first use stacked self-encoder to denoise, and then hand over to LSTM prediction, which improves the stability of multi-step prediction [6]; the CNN-LSTM structure designed by Lu and others first extracts local features with a convolutional layer, and then uses LSTM modeling timing, which performs better than a single network [7]. Relevant reviews also point out that although deep learning has become an important paradigm for time series prediction, the model's ability to distinguish key points still needs to be strengthened [8].

DOI: 10.54254/2753-8818/2026.35018

2.3. Attention mechanism and mixed model

The attention mechanism enables the model to focus on key information through adaptive empowerment of historical hidden states. Qiu et al. introduced attention in LSTM to predict stock prices, which significantly reduced the error [9]; Lu et al. combined CNN and two-way LSTM with attention to further improved accuracy [10]; Lin et al. gave an attention-based closing price LSTM prediction model [11]; Zhang et al. built a CNN-BiLSTM-Attention model and verified its superiority on the Shanghai-Shenzhen 300 index [12]. In addition, Kanwal et al. proposed a prediction model that integrates multiple deep structures [13], Transformer is also used for long-sequence timing prediction [14], Md and others have improved the prediction effect by optimizing multi-layer LSTM, and the role of attention mechanism in investment risk prediction has also been confirmed. Overall, although the combination of attention and LSTM has been widely adopted, there are still few studies focusing on multi-step prediction and systematic error accumulation laws, which is the entry point of this article.

3. Methodology

3.1. Problem definition and data

The stock price sequence is $\{p_1, p_2, \ldots, p_T\}$. Multi-step forecast aims to estimate the price of the next H moment by using the characteristics of the past T time steps. The characteristic vector $x_t \in \mathbb{R}^d$ of each time step is composed of market data such as opening, high, low, closing, trading volume (OHLCV), as well as technical indicators such as moving average, relative strength index (RSI), MACD, etc. This paper takes the daily frequency data of the Shanghai-Shenzhen 300 Index from January 2015 to December 2024 as the research object. It is divided into three parts according to time: training, verification and testing (the ratio is about 7:1.5:1.5), and Min-Max normalizes each characteristic to eliminate the difference in the scale.

3.2. LSTM coding layer

LSTM uses the gate control mechanism to model the long-range dependence of the sequence [5]. In step t , the calculation of forgetting gate, input gate and output gate is as follows:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f), \quad i_t = \sigma(W_i[h_{t-1}, x_t] + b_i), \quad o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \tag{1}$$

The update of candidate memory and cell state and hidden state is:

$$\tilde{c}_t = \tanh(W_c[h_{t-1}, x_t] + b_c), \quad c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t, \quad h_t = o_t \odot \tanh(c_t) \tag{2}$$

In the formula, σ is a Sigmoid function, $\odot$ means multiplication by elements, and W and b are the parameters to be learned. From this, LSTM gradually gives the hidden state sequence $\{h_1, \ldots, h_T\}$.

3.3. Attention layer

Different moments in history have different effects on future trends. The attention layer calculates the weight of the hidden state of each time step and guides the model to focus on key information. The scoring function and normalization weight are defined as:

$$e_t = v^{\top}\tanh\left(W_a h_t + b_a\right), \quad \alpha_t = \frac{\exp(e_t)}{\sum_{k=1}^{T}\exp\left(e_k\right)} \tag{3}$$

Then get the context vector:

$$z = \sum_{t=1}^{T} \alpha_t h_t \tag{4}$$

Among them, α_t reflects the importance of the t time step, and z gathers the weighted whole timing information.

3.4. Multi-step prediction and training

The context vector is mapped to the predicted value of the next H moment at one time by the full connection layer, and adopts a direct multi-step method to avoid the problem of step-by-step error transmission in recursive prediction:

$$\hat{y} = W_z z + b_z, \quad \hat{y} \in \mathbb{R}^H \tag{5}$$

The training takes the mean square error as the target function, and the joint optimization of each prediction step:

$$\mathscr{L} = \frac{1}{H} \sum_{h=1}^{H} \left(y_{T+h} - \hat{y}_{T+h}\right)^2 \tag{6}$$

The optimizer selects Adam and combines Dropout with early stop suppression overfitting. The overall structure of the model is shown in Figure 1.

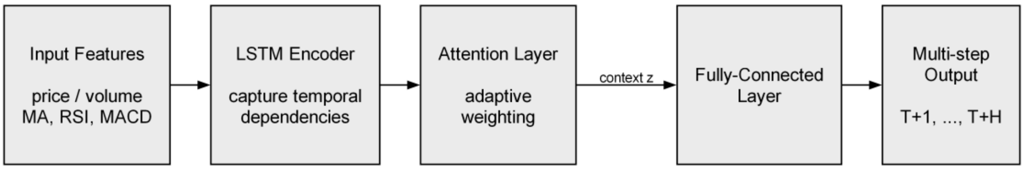

Figure 1. Research framework

4. Results

4.1. Experimental settings

In order to measure the performance of the model, this paper takes ARIMA, support vector regression (SVR), LSTM, GRU, CNN-LSTM and Transformer as comparison objects, and evaluates them together with the LSTM-Attention model under the same data division. The back window takes 30 trading days, and the forecast step length H is set to 1, 3, 5 and 10 respectively. The average square error (RMSE), the average absolute error (MAE), the average absolute percentage error (MAPE), the direction accuracy rate (DA) and the decision coefficient (R^2) are used for evaluation.

DOI: 10.54254/2753-8818/2026.35018

4.2. Results and Analysis

The single-step prediction results of each model are listed in Table 1. ARIMA is constrained by linear assumptions and has the largest error; the depth model is generally superior to the traditional method, and the model in this article ranks first in all indicators: RMSE is 28.3, down about 20.5% compared with the single LSTM (35.6), MAPE decreased from 1.74% to 1.37%, and the direction accuracy rate increased to 61.4%. It can be seen that the attention mechanism improves the discrimination of LSTM by strengthening the attention to critical points.

Figure 2 shows the trend of MAPE under the asynchronous length. With the increase of the step length, the error of each model increases, reflecting the accumulation of error inherent in multi-step prediction; among them, the increase of the model in this article is the smallest, and the MAPE of 10-step prediction is only 4.21%, which is about 30.8% lower than the 6.08% of LSTM. The examination of the attention weight also shows that the model is more inclined to give greater weight to the segment of the adjacent moment and the obvious inflection point of the price, which coincides with the short-term momentum characteristics of the financial sequence and also gives a certain interpretability to the prediction results.

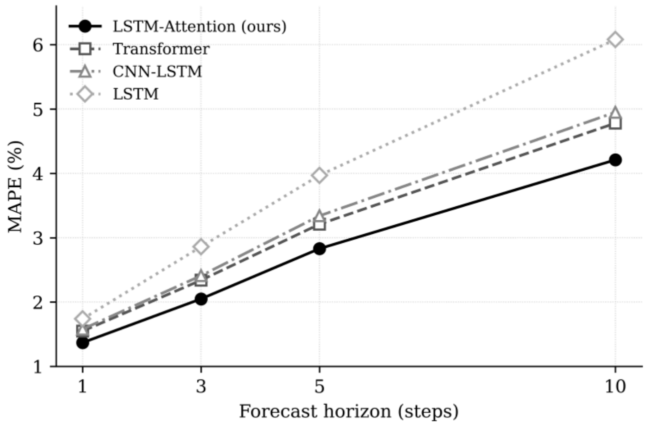

Figure 2. Main results

Table 1. Comparison of single-step forecasting performance on the CSI 300 index

Model	RMSE	MAE	MAPE (%)	DA (%)	R^2
ARIMA	48.7	37.2	2.41	51.3	0.872
SVR	44.1	33.8	2.18	52.6	0.891
LSTM	35.6	27.1	1.74	55.8	0.936
GRU	34.9	26.5	1.70	56.2	0.939
CNN-LSTM	32.4	24.7	1.58	57.9	0.948
Transformer	31.8	24.1	1.55	58.6	0.951
LSTM-Attention	28.3	21.4	1.37	61.4	0.962

5. Discussion

The experimental results show that the integration of LSTM with the attention mechanism does have an advantage in the multi-step prediction of stock prices. The intrinsic reason is that LSTM

DOI: 10.54254/2753-8818/2026.35018

undertakes the task of capturing the long-term dependence of the sequence, and the attention layer reweights the time step hidden state on this basis, so as to alleviate the problem of remote information attenuation and the cover-up of critical points in the long sequence. In the multi-step prediction scenario, this design is particularly critical - direct multi-step output avoids the disadvantages of recursive prediction passing errors along the step length, and the focus of attention on key fragments further reduces the error accumulation speed, so that the model can still maintain relatively stable accuracy in a longer prediction field of view.

Compared with the existing research, this paper not only confirms the advantages of the attention-LSTM hybrid structure in single-step prediction, but also more systematically portrays its ability to suppress the accumulation of multi-step prediction errors, making up for the shortcomings of most previous work to focus on single-step prediction; the interpretability brought by attention weight is also conducive to understanding the decision-making process of the model.

Of course, there are also some limitations in this article: the experiment only takes the Shanghai-Shenzhen 300 Index as an example, and has not been fully tested in multiple markets and multiple types of assets; the model mainly uses price and technical indicators, and does not include external information such as news text and investor sentiment; at the same time, the impact of transaction costs on actual income is not included. These problems are left for further research and improvement.

6. Conclusion

For the multi-step forecast of stock prices, this paper constructs a hybrid model that integrates LSTM and attention mechanism. The model is dependent on the timing of the LSTM coding history sequence, which uses the attention layer to adaptively weight the critical point, and outputs the price of multiple future moments at once with a direct multi-step strategy. The experiment on the daily frequency data of the Shanghai-Shenzhen 300 index shows that the model is superior to the baselines such as ARIMA, LSTM, GRU, CNN-LSTM and Transformer in terms of RMSE, MAE and direction accuracy, and the mean square error is about 20% lower than that of a single LSTM. When the predicted step length increases, the attention mechanism can effectively delay the accumulation of error, and the 10-step predicted MAPE is about 31% lower than that of LSTM, which confirms the effectiveness and robustness of the hybrid model in multi-step prediction.

The follow-up work is planned to be promoted from three directions: first, integrate news text, investor sentiment and other multi-source information to enrich the expression of characteristics; second, test the generalization ability of the model in multiple markets and multiple assets; third, combine transaction costs and risk constraints to explore the application of the model in quantitative investment strategies.

References

[1] Fischer T, Krauss C. Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research, 2018, 270(2): 654-669. https: //doi.org/10.1016/j.ejor.2017.11.054

[2] Sezer O B, Gudelek M U, Ozbayoglu A M. Financial time series forecasting with deep learning: A systematic literature review: 2005-2019. Applied Soft Computing, 2020, 90: 106181. https: //doi.org/10.1016/j.asoc.2020.106181

[3] Jiang W. Applications of deep learning in stock market prediction: Recent progress. Expert Systems with Applications, 2021, 184: 115537. https: //doi.org/10.1016/j.eswa.2021.115537

[4] Kim H Y, Won C H. Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models. Expert Systems with Applications, 2018, 103: 25-37. https: //doi.org/10.1016/j.eswa.2018.03.002

[5] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Computation, 1997, 9(8): 1735-1780. https: //doi.org/10.1162/neco.1997.9.8.1735
[6] Bao W, Yue J, Rao Y. A deep learning framework for financial time series using stacked autoencoders and long-short term memory. PLoS ONE, 2017, 12(7): e0180944. https: //doi.org/10.1371/journal.pone.0180944
[7] Lu W, Li J, Li Y, Sun A, Wang J. A CNN-LSTM-based model to forecast stock prices. Complexity, 2020, 2020: 6622927. https: //doi.org/10.1155/2020/6622927
[8] Lim B, Zohren S. Time-series forecasting with deep learning: a survey. Philosophical Transactions of the Royal Society A, 2021, 379(2194): 20200209. https: //doi.org/10.1098/rsta.2020.0209
[9] Qiu J, Wang B, Zhou C. Forecasting stock prices with long-short term memory neural network based on attention mechanism. PLoS ONE, 2020, 15(1): e0227222. https: //doi.org/10.1371/journal.pone.0227222
[10] Lu W, Li J, Wang J, Qin L. A CNN-BiLSTM-AM method for stock price prediction. Neural Computing and Applications, 2021, 33(10): 4741-4753. https: //doi.org/10.1007/s00521-020-05532-z
[11] Lin Y, Huang Q, Zhong Q, Li M, Li Y, Ma F. A new attention-based LSTM model for closing stock price prediction. International Journal of Financial Engineering, 2022, 9(3): 2250014. https: //doi.org/10.1142/S2424786322500141
[12] Zhang J, Ye L, Lai Y. Stock price prediction using CNN-BiLSTM-Attention model. Mathematics, 2023, 11(9): 1985. https: //doi.org/10.3390/math11091985
[13] Kanwal A, Lau M F, Ng S P H, Sim K Y, Chandrasekaran S. BiCuDNNLSTM-1dCNN — A hybrid deep learning-based predictive model for stock price prediction. Expert Systems with Applications, 2022, 202: 117123. https: //doi.org/10.1016/j.eswa.2022.117123
[14] Zhou H, Zhang S, Peng J, Zhang S, Li J, Xiong H, Zhang W. Informer: Beyond efficient transformer for long sequence time-series forecasting. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(12): 11106-11115. https: //doi.org/10.1609/aaai.v35i12.17325