CNN-BiLSTM-Attention Algorithm-Based Stock Prices Prediction During COVID-19

Xiaoshuai Wang

School of Statistics and Data Science, Nankai University, Tianjin, 300071, China
email: 2111287@mail.nankai.edu.cn

Abstract. Stocks play a crucial role in the field of financial investment, and achieving more accurate stock price predictions is a key focus for today's investors and scholars. However, current algorithms predominantly focus on stable periods and are somewhat inadequate in studying stock price trends during exceptional periods, such as the COVID-19 pandemic and financial crises. Therefore, this paper focuses on stock price prediction during the COVID-19 pandemic. Centered on the LSTM algorithm, it employs CNN for feature extraction and introduces BiLSTM and Attention mechanisms to further enhance model robustness. By constructing a CNN-BiLSTM-Attention composite model, the paper achieves predictions for the stock prices of SZ002912, SZ300006, and SS603311 from different sectors. The study finds that the prediction accuracy of the CNN-BiLSTM-Attention model is significantly higher than that of the standard LSTM model, as it better captures short-term stock price fluctuations. However, it fails to provide effective warnings to investors in the event of sudden, sharp declines in stock prices, unlike the LSTM model.

Keywords: CNN-LSTM-Attention, LSTM, stock price prediction

1 Introduction

In recent years, China's economy has been thriving, and the number of individuals entering the financial investment field has been increasing. Stocks, due to their flexibility, liquidity, and substantial investment returns, have always been favored by a large number of investors. As of 2023, the number of investors in China has exceeded 221.4 million, with various types of professional institutional investors holding a total circulating market value of A shares of 15.9 trillion yuan, which is twice the level at the beginning of 2019 [1]. However, stock data are complex and volatile, influenced by various factors such as the economy, politics, and individual behavior, and exhibit characteristics such as non-stationarity, non-linearity, and high noise, posing significant challenges to time series data prediction. Despite these challenges, the flourishing development of big data analysis technology has brought about a new opportunity, with quantitative investment becoming popular. How to better meet investors' personalized needs and bring steady returns based on

Y. Wang (ed.), *Proceedings of the 2024 International Conference on Artificial Intelligence and Communication (ICAIC 2024)*, Advances in Intelligent Systems Research 185,
https://doi.org/10.2991/978-94-6463-512-6_47

quantitative models under controllable risks has become the focus of attention for scholars and investors both domestically and internationally.

Earlier stock predictions mainly used Autoregressive Conditional Heteroskedasticity (ARCH) models, Autoregressive Integrated Moving Average (ARIMA) models, and Stochastic Volatility (SV) models. However, due to the increasingly complex nonlinear relationships among stock data, traditional statistical models have become less effective, and machine learning algorithms have emerged as efficient alternatives. For instance, Deng et al. [2] achieved satisfactory results using the Back Propagation (BP) neural network for stock prediction, while Chen et al. [3] successfully predicted the stock prices of the new energy vehicle industry based on Random Forest. However, the above algorithms overly rely on sample selection and lack flexibility. Therefore, deep learning algorithms with stronger learning and generalization capabilities have been introduced. The Long Short-Term Memory (LSTM) neural network is a type of recurrent neural network that can better handle long input sequences while considering the time series and nonlinear nature of stock data. Nelson et al. [4] used LSTM networks to achieve a prediction accuracy of 55.9% for stock prices. Vidal and Christophe Yangbouler [5] proposed a method combining Convolutional Neural Network (CNN) and LSTM to predict gold price fluctuations, with prediction accuracy surpassing that of individual CNN or LSTM models.

However, current research is often conducted without considering political interventions or policies, and without taking into account major unexpected events. There is a lack of research on predicting stock prices during significant global events. This paper focus on the performance of three stocks in different sectors during the COVID-19 epidemic. Based on the LSTM algorithm, this paper constructed a CNN-BiLSTM-Attention hybrid model to explore stock price prediction issues during COVID-19 epidemic [6].

2 Method

2.1 Data Preparation

Data Description. This study selected stock data from three different industry sectors, they are Goldensea Hi-Tech (Stock Code: 603311), Sinovatio Technology (Stock Code: 002912), Lummy Pharmaceutical (Stock Code: 300006). The data covers the daily closing prices from October 11, 2018, to April 30, 2024, totaling 2028 trading days. All data was downloaded from Yahoo Finance [7].

Data Preprocessing. To avoid issues such as gradient explosion and model accuracy degradation caused by large data scales, and to promote rapid convergence of the loss function, this study employs the Min-Max normalization method to normalize the original data. The formula for this method is as follows:

$$x_i^* = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{1}$$

In the above equation, x_i represents the original data, and x_{max} and x_{min} represent the maximum and minimum values of the original data, respectively. Since the model's output data is also normalized, this study uses the method of inverse normalization to map it back to the original value range. The calculation formula is as follows:

$$x^* = y^*(x_{max} - x_{min}) + x_{min} \quad (2)$$

Where x^* represents the de-normalized value, y^* represents the original data, and the rest of the variables are defined as above.

2.2 Convolutional Neural Network

The Convolutional Neural Network is a type of feedforward neural network proposed by Lecun et al. in 1998 [8], which has demonstrated excellent performance in tasks such as object detection and semantic segmentation. CNN consists of input layers, convolutional layers, pooling layers, fully connected layers, and output layers. The convolutional and pooling layers are the core of CNN, responsible for feature extraction and dimensionality reduction. In this study, CNN is mainly utilized to extract non-linear features from the local spatiotemporal characteristics of stock data, generating effective feature information input for LSTM model training.

2.3 Long Short-Term Memory Networks

In tasks such as natural language processing and stock prediction, where previous information is essential for the current task, Recurrent Neural Networks is a good choice. RNNs are artificial neural networks designed to handle time series and sequential data, making them suitable for capturing temporal dependencies in data sequences.

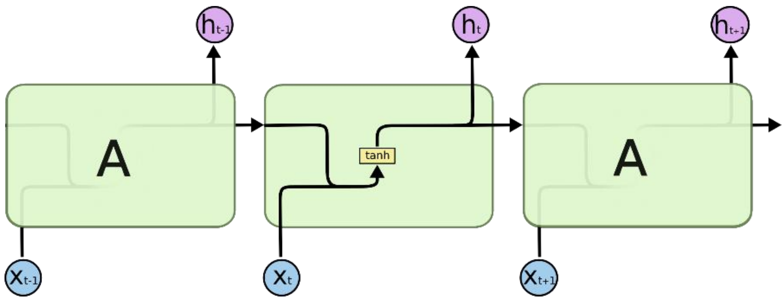

Fig. 1. LSTM structure [9].

Although theoretically Recurrent Neural Networks are capable of handling long-term dependencies, in practice, they are often affected by short-term memory issues, leading to difficulties in retaining information from earlier time steps in long sequences. To address this issue, Long Short-Term Memory networks were introduced.

LSTM shown in Fig. 1, as a variant of RNN, was first proposed by Hochreiter and Schmidhuber in 1997 and later improved and popularized by Alex Graves. Compared to RNNs, LSTMs do not have significant structural differences, but they incorporate a gated unit system consisting of input gates, forget gates, and output gates. This improvement in the hidden layer structure effectively mitigates problems such as gradient vanishing and exploding, allowing LSTMs to more effectively retain long-range temporal dependencies. The structure is depicted in Fig. 2.

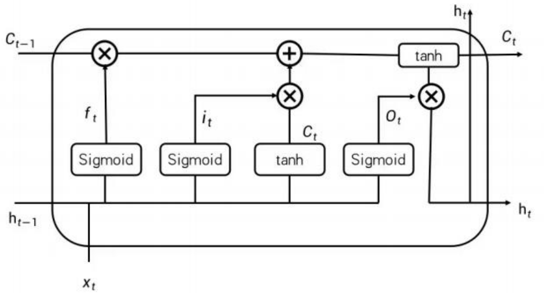

Fig. 2. LSTM memory cell [6].

2.4 Bi-directional Long Short-Term Memory

Considering the heightened uncertainty and volatility in stock markets during the COVID-19 pandemic, this study aims to enhance the robustness of the model. To achieve this, this study introduced the BiLSTM model. BiLSTM is an improvement over LSTM [10]. A single BiLSTM layer consists of two LSTM layers, which process input sequences in both forward and backward directions. The outputs of both LSTM layers are concatenated after all time steps, allowing BiLSTM to capture information from both past and future time steps. This structure enables BiLSTM to effectively learn long-term dependencies in time series data. The structure is depicted in Fig. 3.

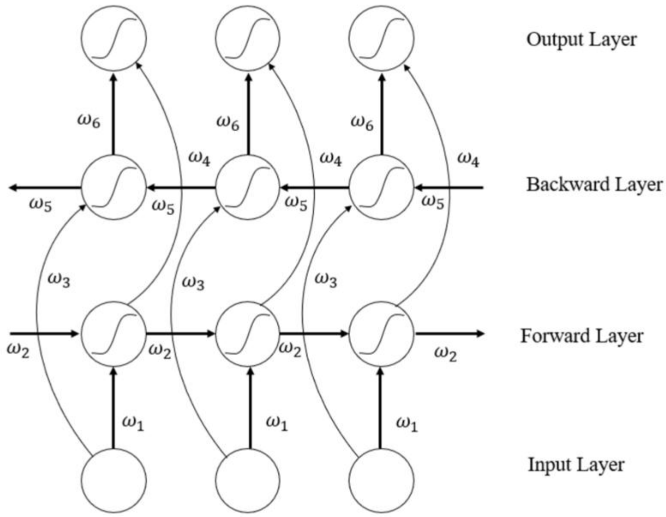

Fig. 3. BiLSTM structure [11].

2.5 Attention Mechanism

The emergence of attention mechanisms is inspired by research on human visual perception. When humans observe an image, they often focus more on certain prominent local features. This allows us to concentrate the limited attention on essential information, thereby conserving resources and quickly obtaining the most effective information. Based on this principle, attention mechanisms can allocate greater weights to key information in long sequence data and selectively discard less important information, effectively addressing the problem of information loss.

2.6 CNN-BiLSTM-Attention

To better enhance the predictive accuracy of the model, this study introduced a hybrid model called CNN-BiLSTM-Attention [6]. The specific structure is as follows:

Step1 Normalize the data and split it into training and testing sets.

Step2 Utilize CNN to extract local spatiotemporal features from the data. The CNN layer consists of a Conv1D layer for feature extraction from 1D time-series data and a Dropout layer to prevent overfitting. The features extracted by CNN are then fed into a BiLSTM layer for training. Additionally, incorporate an attention mechanism to enhance the BiLSTM's ability to capture complex spatiotemporal dependencies by assigning higher weights to key information. Finally, flatten the output using a Flatten layer and input it into a fully connected layer to generate the final result. This neural network structure is shown in Fig. 4.

Step 3 Denormalize the obtained data to obtain the prediction results in the original value range.

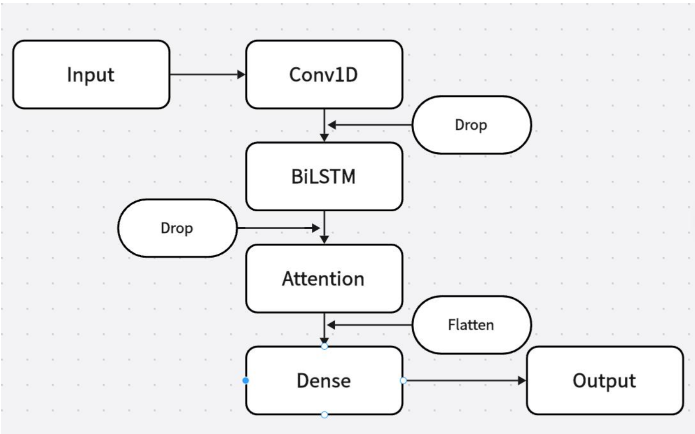

Fig. 4. CNN-BiLSTM-Attention (Photo/Picture credit : Original).

This research is implemented using the Python programming language, with TensorFlow and Keras versions 2.11.0. The 1D convolutional layer contains 64 convolutional kernels and uses the ReLU activation function. The model loss function adopts Mean squared error (MSE), and the Adam algorithm is used to update the parameters of each layer during iteration. Drop layers are introduced after the convolutional layer and BiLSTM layer to help prevent overfitting and improve the generalization ability of the model.

2.7 Evaluation Indicators

This study adopts three metrics, namely the coefficient of determination R^2, Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE), to evaluate the predictive performance of the model. They are calculated by the following equations, respectively:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(y_1 - y_i'\right)^2}{\sum_{i=1}^{n} \left(y_1 - y_i'\right)^2} \tag{3}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(y_1 - y_i'\right)^2} \tag{4}$$

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_1 - y_i'}{y_i} \right| \times 100\% \tag{5}$$

3 Results and Discussion

3.1 The Performance of the Model

The prediction results for the three stocks are shown in Figs. 5-7. The horizontal axis represents the number of days, and the vertical axis represents the stock price. The blue curve represents the predicted stock prices, while the green curve represents the actual stock prices. The closer the two curves are to each other, the better the prediction performance. It can be observed that the CNN-BiLSTM-Attention model used in this study accurately fits the long-term trend of these three stocks and, to a certain extent, reflects the short-term fluctuations in stock prices.

Meanwhile, as shown in the results in Table 1, the prediction accuracy for SZ300006 and SS603311 is significantly higher than for SZ002912 as also evidenced in Figs. 5-7. By examining the original data, it can be observed that SZ002912 exhibits significantly greater short-term price volatility, which poses a greater challenge for the model's predictions and reduces the accuracy of the final results.

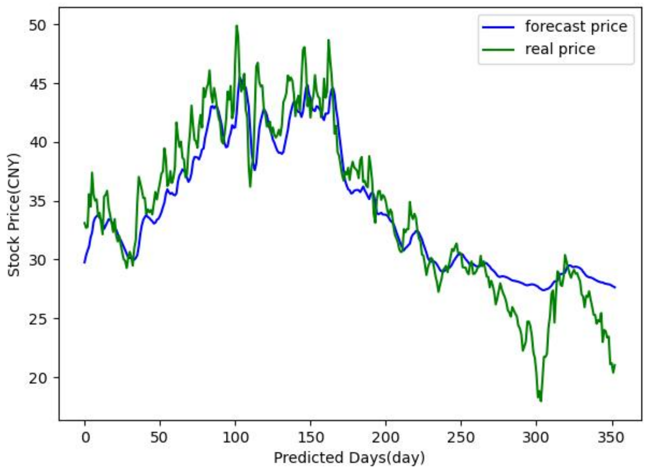

Fig. 5. SZ002912 Prediction results based on CNN-BiLSTM-Attention (Photo/Picture credit : Original).

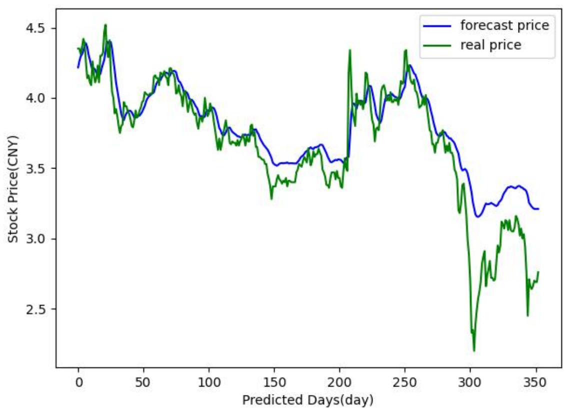

Fig. 6. SZ300006 Prediction results based on CNN-BiLSTM-Attention (Photo/Picture credit : Original).

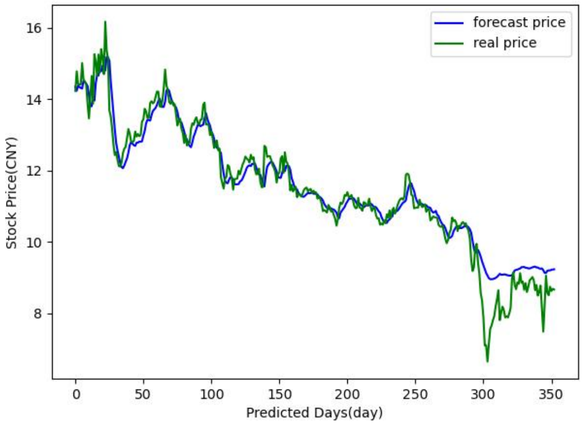

Fig. 7. SS603311 Prediction results based on CNN-BiLSTM-Attention (Photo/Picture credit : Original).

Table 1. The evaluation error indexes of three stock data.

Stock Code	MAPE (%)	RMSE	$\boldsymbol{R^2}$
SZ002912	0.065	2.725	0.854
SZ300006	0.047	0.227	0.751
SS603311	0.033	0.490	0.931

3.2 The Comparison Between LSTM and CNN-BiLSTM-Attention

Although the CNN-BiLSTM-Attention model generally predicts the long-term trends of the three stocks well, it can be observed that all three stocks experienced a significant decline between days 250 and 350, and the model did not accurately predict this drop. Then this study introduced a simple LSTM model and compared its prediction results with those of the CNN-BiLSTM-Attention model. The comparison results are shown in Figs 8-10 and Table 2.

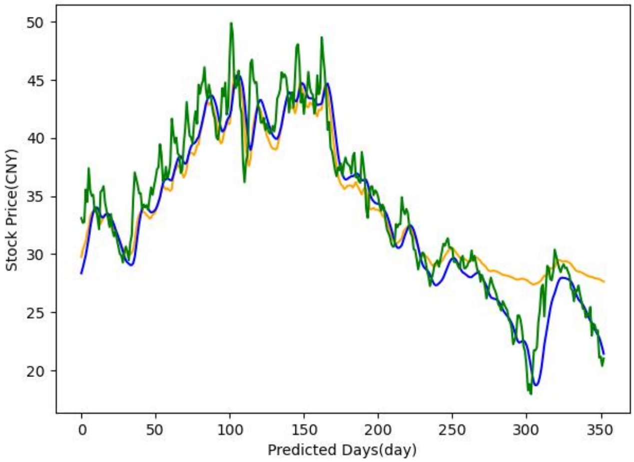

Fig. 8. Comparison between LSTM and CNN-BiLSTM-Attention (SZ002912) (Photo/Picture credit : Original)

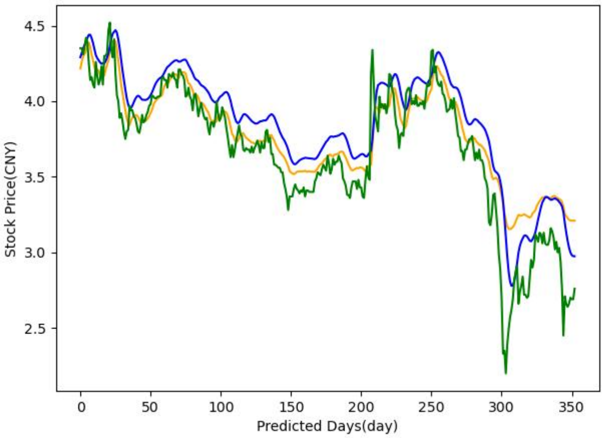

Fig. 9. Comparison between LSTM and CNN-BiLSTM-Attention(SZ300006) (Photo/Picture credit : Original).

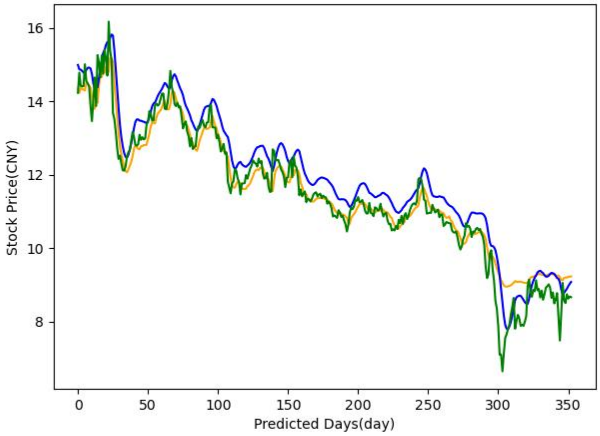

Fig. 10. Comparison between LSTM and CNN-BiLSTM-Attention (SS603311) (Photo/Picture credit : Original)

Table 2. Comparison of evaluation error indexes between LSTM and CNN-BiLSTM-Attention.

Stock Code	**MAPE (%)**		**RMSE**		R^2	
	LSTM	**CNN-BiLSTM-Attention**	**LSTM**	**CNN-BiLSTM-Attention**	**LSTM**	**CNN-BiLSTM-Attention**
SZ002912	0.068	0.065	2.744	2.725	0.852	0.854
SZ300006	0.066	0.047	0.264	0.227	0.665	0.751
SS603311	0.676	0.033	24.241	0.490	0.558	0.931

From the above results, it can be observed that the LSTM model is significantly less accurate than the CNN-BiLSTM-Attention model in overall prediction precision. The LSTM model produces smoother predictions, which means it does not capture short-term stock fluctuations as well. However, the LSTM model accurately predicts the steep declines that occurred between days 250 and 350 for the three stocks. Given the frequency and high risk of such steep declines during exceptional periods like COVID-19 pandemics, this article suggests that the LSTM model may better serve investors by providing warnings and helping to mitigate losses, compared to the CNN-BiLSTM-Attention model.

4 Conclusion

This study centers on the LSTM deep neural network, examining its strengths and weaknesses, and combines its capabilities with CNN to extract spatiotemporal features of stock data. By incorporating BiLSTM and Attention mechanisms, the model's ability to learn long-term dependencies in time series data is enhanced. Through the development of the CNN-BiLSTM-Attention composite model, this research achieves high-accuracy predictions for three stocks from different sectors in China during the pandemic. The findings indicate that, despite the high accuracy of this hybrid model, the LSTM model is more advantageous for providing early warnings to investors during exceptional periods, addressing a gap in stock prediction during such times in China.

Although this study has constructed an effective model, it still has shortcomings in capturing short-term stock price fluctuations. Additionally, whether this model can maintain its accuracy across different contexts and stocks in the future remains uncertain. Future research may delve further into the following areas: 1) Data Integration: Integrate multimodal data and consider more factors that influence stock price movements to enhance the model's generalization ability. 2) Model Development: Further develop models that can better capture short-term stock price trends and provide more effective warnings to investors.

References

1. Xueqiu. https://xueqiu.com/9835226687/282296819 (2024).
2. Deng, X.K., Wan, L., Huang, N.N.: Stock Prediction Research Based on DAE-BP Neural Network. Computer Engineering and Applications 55(3), 126-132 (2019).
3. Chen, M.L., Fan, C., Wu, Z.P.: Stock Price Prediction Research in the New Energy Automobile Industry—Based on Machine Learning Algorithms. Journal of Jilin Business and Technology College 40(01), 93-100 (2024).
4. Nelson, D., Pereira, A., Oliveira, R.: Stock market's price movement prediction with LSTM neural networks. In: Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN), pp. 1419. IEEE, New York, NY, USA (2017).
5. Vidal, A., Kristjanpoller, W.: Gold volatility prediction using a CNN-LSTM. Expert Systems with Applications 157, 113481 (2020).
6. Zhang, J., Ye, L., Lai, Y.: Stock Price Prediction Using CNN-BiLSTM-Attention Model. Mathematics 11, 1985 (2023).
7. Finance Yahoo. https://hk.finance.yahoo.com (2024).
8. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proceedings of the IEEE 86, 2278–2324 (1998).
9. Hochreiter, S., Schmidhuber, J.: Long Short-Term Memory. Neural Computation 9, 1735–1780 (1997).
10. Xu, Y., Prince, H., Wu, Z.: Predicting Stock Trends with CNN-Bi LSTM Based Multi-Feature Integration Model. Data Analysis and Knowledge Discovery 7, 126–137 (2021).
11. Lee, R.S.: Transfer Learning and Transformer Technology. In: Natural Language Processing: A Textbook with Python Implementation, pp. 175-197. Springer Nature Singapore (2023).