Application of LSTM and Attention Mechanism for Stock Price Prediction and Analysis

Yingbing Li[1], Xue Zhang[2], and Xueyan Zhu[3](✉)

[1] School of Economics, Qingdao University, Qingdao 266071, China
liybing777@qdu.edu.cn
[2] College of Biological Science and Biotechnology, Beijing Forestry University, Beijing 100083, China
[3] School of Technology, Beijing Forestry University, Beijing 100083, China
xueyan0111@bjfu.edu.cn

Abstract. The relationship between stock prices and economic development is well-established, and the study of stock price prediction methods is crucial for gaining insights into the economy. This research aims to enhance the accuracy of stock price prediction by leveraging a combination of convolutional neural network (CNN), attention mechanisms, and long short-term memory (LSTM). Specifically, the stock data of Banco Bilbao Vizcaya Argentina (BBVA) was selected as the research object, and the stock price prediction results after incorporating efficient channel attention (ECA), channel attention module (CAM), spatial attention module (SAM), and convolutional block attention module (CBAM) in CNN + LSTM were analyzed. The findings demonstrate that the incorporation of attention mechanisms in CNN + LSTM has a positive impact on stock price prediction accuracy. Notably, the model that integrates CBAM attention mechanism in CNN + LSTM yields the best prediction results, with MAE, MSE, RMSE, and R^2 metrics achieving 58.68×10^{-4}, 7.94×10^{-5}, 8.91×10^{-3}, and 0.9673, respectively. These results have implications for improving the accuracy of stock price prediction and provide valuable insights for future research in this area.

Keywords: stock price · CNN · LSTM · attention mechanisms · CBAM

1 Introduction

Investors have shown great interest in stocks due to their convenience, high-risk, and high-yield investment potential [1]. The recent growth of the economy and financial awareness has resulted in an increasing number of small and medium-sized investors participating in stock trading. Various factors, including news, policies, and market sentiments, significantly impact stock prices, making it difficult for investors to analyze future trends and make informed decisions. Therefore, it is essential to explore and develop effective stock price forecasting methods.

Currently, stock price forecasting methods can be categorized into traditional, machine learning, and deep learning approaches [2–5]. Traditional methods, such as

K. Hemachandran et al. (Eds.): ICAID 2023, AHIS 9, pp. 553–561, 2023.
https://doi.org/10.2991/978-94-6463-222-4_60

generalized autoregressive conditional heteroskedasticity [6], vector autoregression [7], and autoregressive integrated moving average model [8], have limitations in capturing the complex relationships between historical stock data and future prices. As a result, their forecasting outcomes often diverge significantly from actual results. Machine learning (ML) techniques have gained immense popularity in the domain of stock price prediction owing to their ability to capture complex relationships in data, which often elude human observation [9]. Among the various ML techniques employed for this purpose, support vector machines [10] and random forests [11] have emerged as popular choices. Recent research endeavors have focused on improving the predictive accuracy of random forest algorithm for stock price prediction. Yan et al. [12] reporting an MAE of 0.610 and an MSE of 0.728 for predicting the stock price of Ping An of China using an improved random forest algorithm. Additionally, Chen et al. [13] have explored the use of feature-weighted support vector machines and feature-weighted K-neighborhood algorithms based on the former, for predicting the price of SSE Composite Index and SZSE Composite Index, respectively.

Recurrent neural networks (RNNs) are popular tools for predicting stock prices; however, they often suffer from the “gradient disappearance” and “gradient explosion” problems during the training process [14, 15]. To address these challenges, long-short term memory (LSTM) networks have been introduced, which have proven effective in stock price prediction. Yadav et al. [16] optimized the LSTM model to achieve accurate stock price prediction in India. However, stock price prediction with limited data samples remains a challenge. Nguyen et al. [17] proposed a deep migration learning technique to train the LSTM network, which improves the prediction accuracy by enabling effective training on limited data. Specifically, the authors predicted the stock prices of the top five companies in terms of market capitalization in the United States and South Korea from 2012 to 2018 and found that the proposed deep migration learning technique effectively improved the prediction accuracy.

Moreover, attention mechanisms have been widely used in image and natural language processing to improve the accuracy of models. Hence, to enhance the accuracy of stock price prediction by LSTM models, this study proposes the incorporation of an attention mechanism into LSTM models, which can focus on key features and improve prediction accuracy. Therefore, the authors design attention LSTM models for stock price prediction.

2 Materials and Methods

2.1 Stock Data Acquisition and Dataset Construction

In this study, we investigate the behavior of the stock of Banco Bilbao Vizcaya Argentina (BBVA) on the New York Stock Exchange using historical trading data spanning from December 15, 1988 to December 12, 2022. The data includes the opening, lowest, highest, and closing prices of the stock, and was collected in 1 business day timing units, resulting in a total of 8565 data points. To build our predictive model, we divided the collected data into a training set (5996 items) and a test set (2569 items) using a 7:3 ratio. Our model takes as input a time period of 5 working days, each containing the 4 aforementioned features.

2.2 Stock Price Method

The present study proposes a novel approach to enhance the precision of stock price prediction. The proposed approach leverages a hybrid architecture that integrates a convolutional neural network, an attention mechanism, and an LSTM network. This architecture is depicted in Fig. 1. Specifically, the convolutional neural network is employed to extract features from the raw stock data. Next, the attention mechanism is utilized to identify the salient features among the extracted features that are relevant to stock prediction. Lastly, the LSTM network is utilized to process the attended features and make accurate predictions of the stock prices.

Attention Mechanism

The incorporation of attention networks in deep learning has been shown to effectively enhance the accuracy of prediction models in various fields, such as image recognition, image segmentation, speech recognition, and text translation. These attention networks are capable of weighting the features extracted by convolutional neural networks according to their relevance, enabling models to focus on crucial features and disregard irrelevant ones during the training process. In the context of stock price prediction, this study aims to improve the accuracy of prediction models by integrating attention networks.

The efficient channel attention (ECA) and convolutional block and attention module (CBAM) have gained popularity in a range of prediction networks due to their remarkable performance. This study investigates the potential of these mechanisms in enhancing the accuracy of stock price prediction models. The ECA attention module, in particular, is an efficient and small parametric attention module that implements a local cross-channel interaction strategy using fast one-dimensional convolution. It also employs an adaptive channel dimension function to determine the size of the one-dimensional convolution kernel. Figure 2 depicts the network structure diagram of the ECA attention module.

As demonstrated in Fig. 2, the ECA attention mechanism employs a one-dimensional convolution to capture information interaction among local channels. The effectiveness of this module is critically dependent on the kernel size k of the one-dimensional convolution, which directly determines the coverage of ECA attention. Notably, a mapping relationship exists between the kernel size k and the channel dimension C within the ECA attention module. Specifically, as the channel dimension C increases, the long-term interaction of ECA attention tends to be stronger, whereas a smaller channel dimension

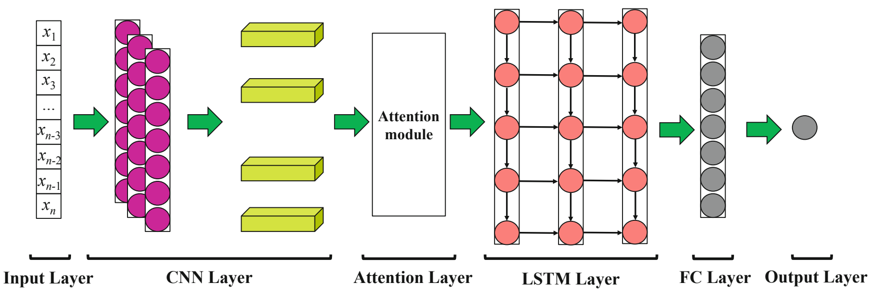


Fig. 1. The architecture of proposed stock price predicton method

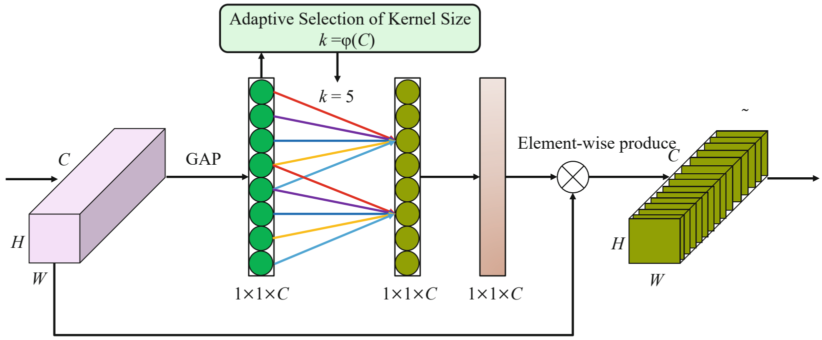


Fig. 2. Network diagram of efficient channel attention module

results in a stronger short-term interaction of ECA attention. This mapping relationship is formulated mathematically in Eq. (1).

$$k = \varphi(C) = \left| \frac{log_2 C}{\gamma} + \frac{b}{\gamma} \right|_{odd} \tag{1}$$

where both γ and b are parameters in the model training process, and $|x|_{odd}$ denotes the nearest odd number to x.

CBAM is a novel attention module that effectively enhances the performance of a model with minimal increment in the number of module operations and parameters. The proposed module combines both channel attention module (CAM) and spatial attention module (SAM), which are sequentially applied to the input feature map F, as depicted in Fig. 3.

As shown in Fig. 3, when the feature map F is input to CBAM attention, the channel attention module leverages global maximum pooling to extract channel-wise features from the input feature map F. The attention weights of each channel are learned using a shared perceptron and subsequently normalized using the Sigmoid activation function. The rescaled attention weights are then applied to the original input feature map F in a channel-wise manner to obtain the final output. The corresponding calculation process for channel attention is presented in Eq. (2).

$$M_c(F) = \sigma\left(W_1\left(W_0\left(F^c_{avg}\right)\right) + W_1\left(W_0\left(F^c_{max}\right)\right)\right) \tag{2}$$

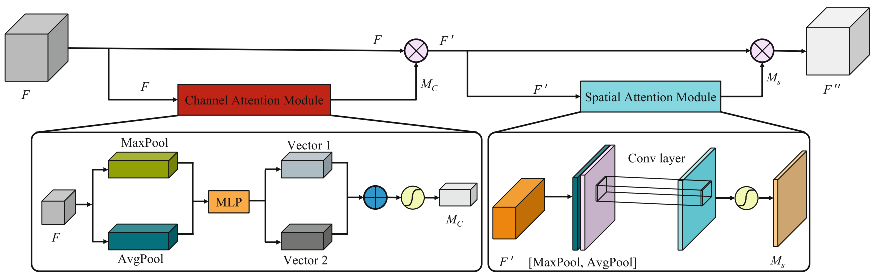


Fig. 3. Network diagram of convolutional block attention module

where $M_c(F)$ denotes the channel attention feature, σ denotes the Sigmoid activation function, W_1 and W_0 denote the weights of the shared weights, F^c_{avg} denotes the global average pooling feature, and F^c_{max} denotes the global maximum pooling feature.

To obtain the attentional features in spatial dimensions, the feature map *F'* generated by the channel attention module was input into the spatial attention module. The spatial attention module employed global maximum pooling and global average pooling to transform the input feature dimension from H × W to 1 × 1 based on the width and height of the feature map *F'*. In order to reduce the dimensionality of the feature map, a 7 × 7 convolution kernel and Sigmoid activation function were applied. The feature maps outputted by the channel attention module were then combined with those outputted by the spatial attention module to obtain rescaled input feature maps *F* over both space and channels. The calculation process of the spatial attention was represented by Eq. (3).

$$M_s(F) = \sigma\left(f^{7\times7}\left[F^c_{avg}; F^c_{max}\right]\right) \tag{3}$$

where $M_s(F)$ denotes the spatial attention feature and $f^{7\times7}$ denotes the convolution kernel size of 7 × 7.

LSTM Network

LSTM is a type of neural network that overcomes the challenge of long-term dependency in Recurrent Neural Networks. The basic structure of LSTM is illustrated in Fig. 4, comprising an input layer, hidden layer, and output layer. The hidden layer contains unique memory units that enable the network to remember information from the distant past and selectively forget irrelevant information. By integrating current and historical information, LSTM can effectively predict stock prices and address the issue of long-term dependence in stock price forecasting. Additionally, the forgetting gate mechanism in LSTM can alleviate the problem of gradient disappearance and explosion, thereby presenting significant benefits for stock price prediction.

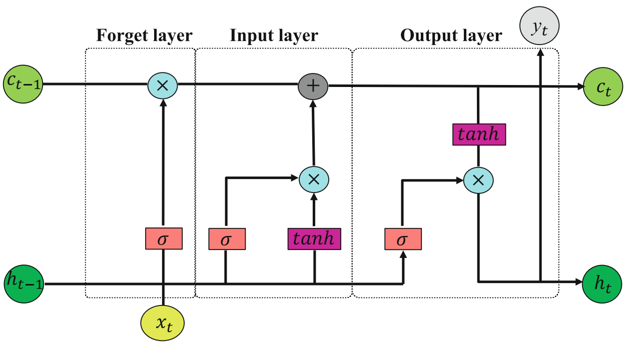
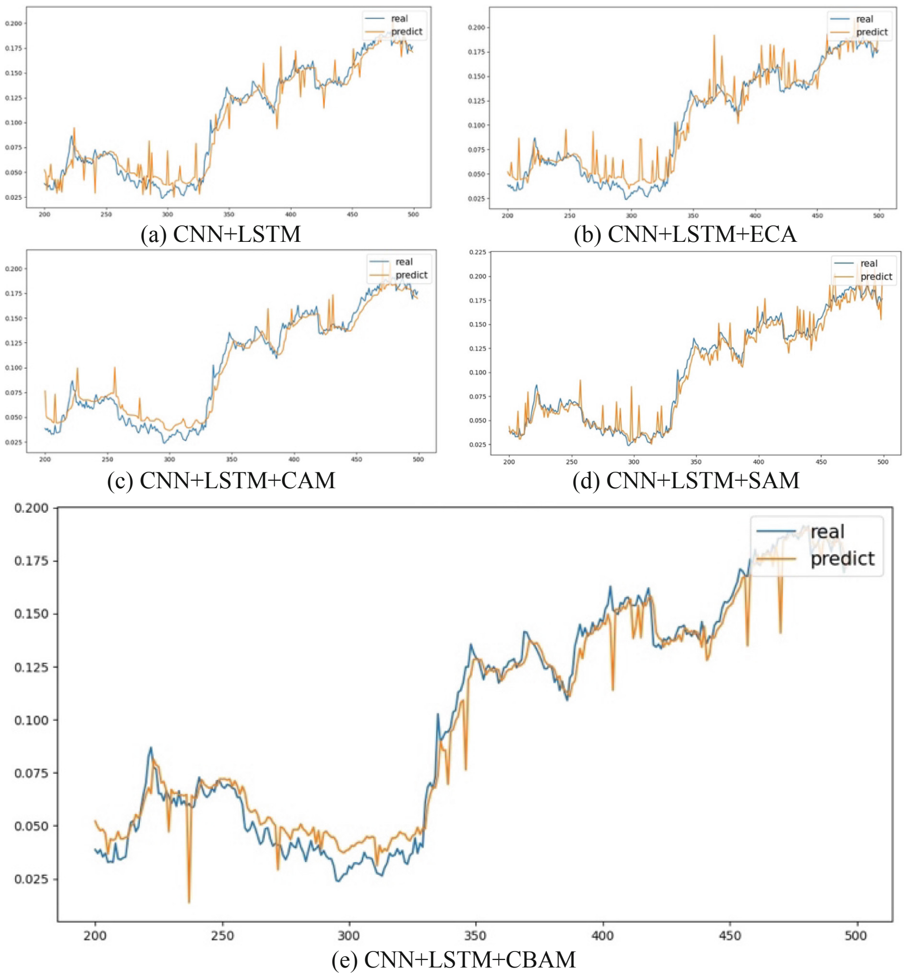


Fig. 4. Network diagram of LSTM

3 Results and Discussion

To fully verify the effects of different attention methods on stock price prediction results, a comparison experiment was designed. Figure 5 shows the stock price prediction results of different methods in the comparison experiment.

The findings from Fig. 5, depicting the stock price prediction results of various models, reveal a significant enhancement in prediction accuracy when attention is introduced in the CNN + LSTM model. However, there are variations in the effectiveness of different attention models in improving stock price prediction. Among them, the model has the best prediction following effect on stock prices after introducing CBAM attention in CNN + LSTM.The mean absolute error (MAE), mean square error (MSE), root mean square error (MSE) and correlation coefficient (R_2) were used to quantitatively evaluate



(a) CNN+LSTM (b) CNN+LSTM+ECA

(c) CNN+LSTM+CAM (d) CNN+LSTM+SAM

(e) CNN+LSTM+CBAM

Fig. 5. Comparison of stock price prediction results with different methods

Table 1. Quantitatively evaluate results of stock price prediction with different method

Method	MAE	MSE	RMSE	$R2$
CNN + LSTM	89.67 × 10–4	1.59×10^{-4}	1.26×10^{-2}	0.9346
CNN + LSTM + ECA	84.53×10^{-4}	1.55×10^{-4}	1.25×10^{-2}	0.9362
CNN + LSTM + CAM	78.13×10^{-4}	1.06×10^{-4}	1.03×10^{-2}	0.9565
CNN + LSTM + SAM	71.99×10^{-4}	9.67×10^{-5}	9.83×10^{-3}	0.9602
CNN + LSTM + CBAM	58.68×10^{-4}	7.94×10^{-5}	8.91×10^{-3}	0.9673

the stock price predicting results of different methods in the comparison experiment, as shown in Table 1.

Notably, the CBAM attention, when incorporated in CNN + LSTM, exhibits the most favorable impact on stock price prediction, followed by SAM, CAM, and ECA attentions, respectively, according to the quantitative evaluation results presented in Table 1. The evaluation metrics, including mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE), and correlation coefficient (R^2), collectively support the improved stock price prediction efficacy of the models with the incorporation of ECA, CAM, SAM, and CBAM attentions in CNN + LSTM. Specifically, for predicting BBVA stock prices, the MAE, MSE, RMSE, and R^2 of CNN + LSTM are measured as 89.67×10^{-4}, 1.59×10^{-4}, 1.26×10^{-2}, and 0.9346, respectively. Remarkably, the introduction of CBAM attention in CNN + LSTM leads to a reduction of 30.99×10^{-4}, 7.96×10^{-5}, and 3.7×10^{-3} in MAE, MSE, and RMSE of stock price prediction, respectively, while improving R^2 by 0.0327.

4 Conclusion

This study endeavors to enhance the accuracy of stock price prediction by integrating a deep learning convolutional neural network with attention mechanism and LSTM. The experimental validation of the proposed stock price prediction model demonstrates its efficacy, and the analysis of different attentional modules sheds light on their impact on stock price prediction. The results highlight the potential of attention mechanism in improving the accuracy of stock price prediction. In future research, the inclusion of sentiment analysis data, such as financial news and stock review analysis, in the stock prediction model will be explored to further enhance its prediction performance.

References

1. Hossain E, Hossain M S, Zander P. Machine learning with Belief Rule-Based Expert Systems to predict stock price movements. Expert Systems with Applications, 2022, 206, 117706.
2. Song Z, Yu C. Investor sentiment indices based on k-step PLS algorithm: A group of powerful predictors of stock market returns. International Review of Financial Analysis, 2022, 83, 102321.

3. Li A, Liu M, Sheather S. Predicting stock splits using ensemble machine learning and SMOTE oversampling. Pacific-Basin Finance Journal, 2023, 78, 101948.
4. Ai-sulaiman T. Predicting reactions to anomalies in stock movements using a feed-forward deep learning network. International Journal of Information Management Data Insights, 2022, 2(1): 100071.
5. Bhandari H N, Rimal B, Pokhrel N R, et al. Predicting stock market index using LSTM. Machine Learning with Applications, 2022, 9, 100320.
6. Adam P, Cahyono E, Millia H, et al. Statistical analysis on the effect of exchange rate on stock price in Indonesia: an application of ARDL and IGARCH models//Journal of Physics: Conference Series. IOP Publishing, 2021, 1899(1): 012117.
7. Bouri E, Gupta R, Hosseini S, et al. Does global fear predict fear in BRICS stock markets? Evidence from a Bayesian Graphical Structural VAR model. Emerging Markets Review, 2018, 34: 124-142.
8. Zhang X, Zhai R, Gao W. Analysis and economic prediction of stock index based on index tracking and ARIMA model//Journal of Physics: Conference Series. IOP Publishing, 2021, 1903(1): 012015.
9. Perry S. Using machine learning to predict clean energy stock prices: How important are market volatility and economic policy uncertainty? Journal of Climate Finance, 2022, 1, 100002.
10. Mallqui D C A, Fernandes R A S. Predicting the direction, maximum, minimum and closing prices of daily Bitcoin exchange rate using machine learning techniques. Applied Soft Computing, 2019, 75: 596-606.
11. Ribeiro M H D M, dos Santos Coelho L. Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series. Applied soft computing, 2020, 86: 105837.
12. Yan Z X. Stock price prediction of random forest model based on Pearson feature selection. Computer Engineering and Applications, 2021, (15):286-296.
13. Chen Y, Hao Y. A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction. Expert Systems with Applications, 2017, 80: 340-355.
14. Mustaffa Z, Sulaiman M H. Stock price predictive analysis: An application of hybrid Barnacles Mating Optimizer with Artificial Neural Network. International Journal of Cognitive Computing in Engineering, 2023, 4, 107-117.
15. Yan X, Cai Q, Zhang S H, et al. Exploring machine learning in stock prediction using LSTM, binary tree, and linear regression algorithms. International Core Journal of Engineering, 2021, 7(3): 373-377.
16. Yadav A, Jha C K, Sharan A. Optimizing LSTM for time series prediction in Indian stock market[J]. Procedia Computer Science, 2020, 167, 2091-2100.
17. Nguyen T T, Yoon S. A novel approach to short-term stock price movement prediction using transfer learning. Applied Sciences, 2019, 9(22): 4745.