Bitcoin Price Prediction Based on CNN-Bi-LSTM-Attention Model

Qifei Yang [1,*], Yuhan Sun [2], Yuhao Wu [3]

[1] College of Software, Jilin University, Changchun, 130000, China

[2] Sichuan University Pittsburgh Institution, Sichuan University, Chengdu, 610225, China

[3] College of Science, China University of Mining and Technology, Beijing, 100083, China

* Corresponding Author Email: yangqf94605@gmail.com

Abstract. Due to many factors, Bitcoin has experienced huge price fluctuations since its emergence, and it has received extensive attention. Forecasting the price of bitcoin is of great significance for investors and for the country's future development. This paper collects the data of bitcoin price and indicator that may affect the price, and then use random forest algorithm for feature selection to remove all nonessential indicators. Then, CNN-Bi-LSTM-Attention model is built to train the data and predict the price of bitcoin. Finally, this model is compared with other models. It can be found that this model has higher prediction accuracy and better prediction effect than traditional models such as LSTM and CNN-LSTM.

Keywords: Convolutional Neural Network (CNN), Bidirectional LSTM, Attention mechanism, Bitcoin, Prediction Model.

1. Introduction

In recent years, the Internet has developed rapidly, and the operation mode of the real economy has changed dramatically. Among them, cryptocurrency is a new type of digital assets, which uses the decentralized network based on block chain technology and encryption technology to verify transactions [1]. At present, Bitcoin has attracted wide attention due to its highly unstable price fluctuations. Therefore, it is very important to study the factors affecting bitcoin price and predict its future price. From a macro point of view, the study of bitcoin is helpful to the development of digital currency and is of great significance to the whole country and society. From a micro perspective, the price prediction of bitcoin can help investors understand the market in depth and make more rational investment decisions.

So far, the research on bitcoin price prediction has achieved rich results. Cavalli et al. [2] predicted bitcoin prices based on one-dimensional convolutional neural networks (1D-CNN) and gave appropriate trading strategies. McNally et al. [3] used long-short term memory (LSTM) and historical price data to classify price movement patterns. Kristjanpoller et al. [4] tested the ANN-GARCH model for price forecasting, and used ANN to capture the nonlinear effect of price fluctuations. Indera et al. [5] constructed a nonlinear autoregressive exogenous input model based on multi-layer perceptron They used the past opening value, closing value, minimum and maximum value and moving average technical indicators to predict the bitcoin price. Dihn et al [6] used recurrent neural network and LSTM to predict bitcoin, and compared with the traditional multi-layer perceptron (MLP). The results show that the machine learning method is better because of its more advanced time characteristics. However, the existing prediction models are relatively single, and the prediction accuracy is not high. Therefore, this paper combines and improves the existing deep learning models and uses CNN-Bi-LSTM-Attention model to predict. CNN and LSTM are machine learning models that are often used to predict. CNN is a deep neural network with convolution structure, which can extract higher level and more abstract data features from raw data, and is widely used in image classification, face recognition and other fields. As a variant of RNN (recurrent neural network), LSTM neural network solves the problems of gradient disappearance and gradient explosion that are easy to occur in the training process of RNN. The forward LSTM and backward LSTM are combined into Bi-LSTM, which can not only obtain the past information of the input data, but also use the future information. It is very helpful for the sequence data task [7].

2. Methodology

2.1. Random Forest

Due to the large number of indicators selected in this paper, Random Forest Algorithm (RF) is adopted for feature engineering. Meanwhile, the algorithm can also prevent overfitting and significantly improve the generalization ability of the model [8].

2.2. The model of CNN-Bi-LSTM-Attention

2.2.1 Convolutional Neural Network (CNN)

Considering that the price data of Bitcoin in this paper is a time series, this paper uses the feature abstraction ability of CNN to extract features. It is mainly composed of convolutional layers, pooling layers, and fully connected layers [9]. The convolutional layer uses convolution kernels for feature extraction and feature mapping. The pooling layer performs down-sampling to sparse features. The fully connected layer is usually refitted at the end of the CNN to reduce the loss of feature information; The structure of CNN is shown in Figure 1:

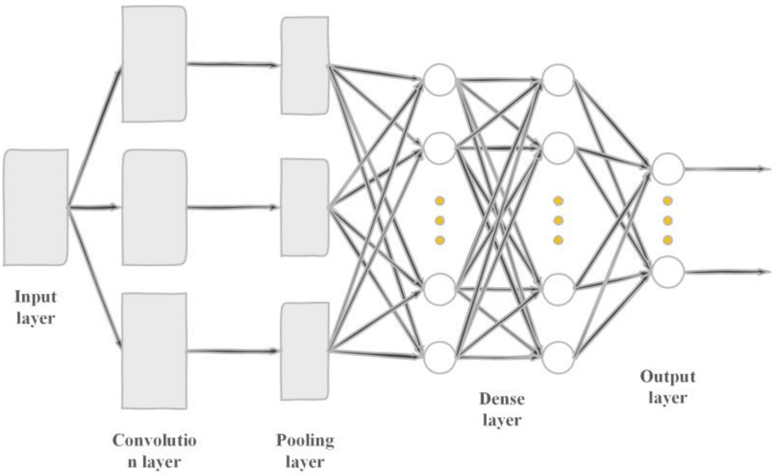

Figure 1. Structure of CNN

2.2.2 Bidirectional LSTM(Bi-LSTM)

LSTM is an improved time recurrent network. LSTM regulates the information flow by introducing the internal mechanism of the "gate" [10]. LSTM can only perform the forward propagation process, while Bi-LSTM consists of a forward LSTM and a reverse LSTM network. Bi-LSTM network can consider both historical information and future information, improve model efficiency and reduce the risk of overfitting [11]. The structure of Bi-LSTM is shown in Figure 2.

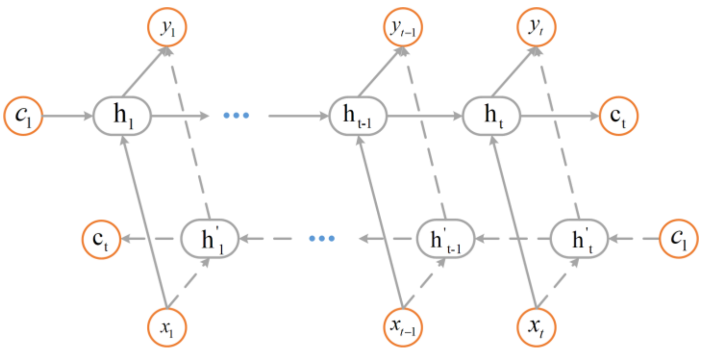

Figure 2. Bidirectional LSTM

2.2.3 Attention mechanism

The attention mechanism is widely used in deep learning, and this paper applies it to the analysis of the Bitcoin price time series. Its structure is shown in Figure 3.

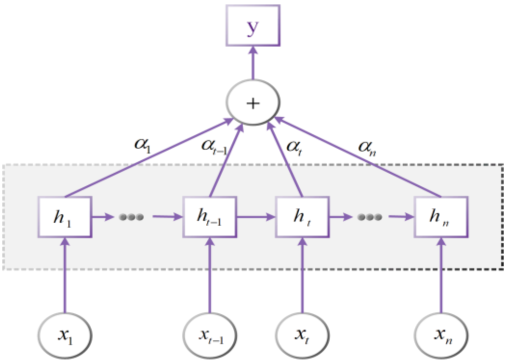

Figure 3. Schematic diagram of Attention mechanism structure

The attention mechanism can extract key features by weighting to allocate limited processing resources to key information. It learns the input sequence information better and improves the prediction performance effectively. The formula of the attention mechanism is:

$$a_t = \frac{\exp[h_t \overline{h_s}]}{\sum_{j=1}^{n} \exp[s(h_t \overline{h_s})]} \tag{1}$$

$$s(h_t, \overline{h_s}) = h_t^T \overline{h_s} \tag{2}$$

$$c = \sum_{i=1}^{t} a_i h_i \tag{3}$$

3. Results

3.1. Data sources and evaluation indicators

In order to prove the good effect of the bitcoin price prediction model in this paper, a total of 10,000 data is selected, which includes the opening price, closing price, maximum price, minimum price, and trading volume of bitcoin.

3.2. Modeling of CNN-Bi-LSTM prediction method based on Attention mechanism

3.2.1 Parameter optimization for sliding windows

This model uses the sliding window method to deal with the Bitcoin price sequence. The selection of the sliding window size will affect the model's performance, so the sliding window parameters must be analyzed and selected. Table 1 illustrates the prediction performance indicators of the CNN-Bi-LSTM-Attention model on the test set under different sliding window sizes. It can be seen from Table 1 that when the sliding window size is 4, the prediction result is the best.

Table 1. Comparison of prediction results with different sliding window sizes

window size	MAPE/%	RMSE	MAE	R^2
2	1.149	0.015	0.008	0.988
4	**1.134**	**0.008**	**0.006**	**0.991**
8	1.170	0.016	0.009	0.986
16	2.272	0.018	0.015	0.981
32	3.098	0.024	0.020	0.974

3.2.2 Hyperparameter Optimization of CNN

If the CNN network is too deep, it may cause gradient explosion or disappearance. Therefore, this paper sets one convolutional layer and one pooling layer for CNN. Too small a convolution kernel may lead to insufficient time series feature extraction. However, if the convolution kernel is too large, it may cause loss of some local features or overfitting. The CNN-Bi-LSTM-Attention model is tested five times under different values of the convolution kernel (K_c) and the pooling kernel (K_p), respectively. Moreover, the average of the evaluation indicators is compared. The results are shown in Table 2.

Table 2. Evaluation indicators for different K_c and K_p values

K_c	RMSE		R^2	
	K_c=1	K_c=2	K_c=1	K_c=2
1	0.015	0.014	0.967	0.973
2	0.011	0.012	0.985	0.979
3	**0.009**	0.011	**0.991**	0.982

The experimental results show when the convolution kernel size is 3, and the pooling kernel size is 1, the model has the best prediction effect. Therefore, this paper sets the convolution kernel size of CNN to 3 and the pooling kernel size to 1.

3.2.3 Parameter optimization of Bi-LSTM

The number of layers and the size of the hidden layer of Bi-LSTM have an impact on the performance of the model in Bitcoin price prediction. The larger the hidden layer of Bi-LSTM, the better the time-series correlation features can be mined. However, if it is too large, it will cause redundancy and reduce efficiency. It is evident from the data in Table 3 that when the number of Bi-LSTM network layers is 2 and the hidden layer size is 32 and 64, respectively, the prediction accuracy is the best among the above 10 models.

Table 3. Model prediction accuracy under different network structures

Layers number	units	MAPE/%	RMSE	MAE	K_p
1	32	2.446	0.015	0.014	0.968
	64	1.471	0.019	0.011	0.983
	128	1.377	0.009	0.007	0.988
	256	1.310	0.009	0.007	0.988
2	32-32	1.395	0.010	0.007	0.988
	32-64	**1.134**	**0.008**	**0.006**	**0.991**
	64-64	1.526	0.010	0.008	0.985
	64-128	2.640	0.016	0.014	0.965
	128-128	2.716	0.016	0.014	0.963
	128-256	1.999	0.012	0.009	0.981

3.2.4 Parameter optimization for batch size

In order to choose the optimal batch size, this paper will select the value of the batch size at {64, 128, 256, 512, 1024}, and perform 10-fold cross-validation in turn. The average of ten MAE results with different batch sizes is shown in Table 4 below. It can be seen from the table that when the batch size is 256, there is the lowest MAE.

Table 4. Average of MAE under different batch sizes

Batch size	Average of MAE
64	0.0261
128	0.0372
256	**0.0255**
512	0.0370
1024	0.0263

3.3. Structure of CNN-Bi-LSTM-Attention

The structure of the CNN-Bi-LSTM-Attention model proposed in this paper is shown in Figure 4. After inputting the normalized data, a one-dimensional convolutional layer (Conv1D) is used for feature extraction, which contains 64 convolution kernels of size 3. Then, filter the data through a one-dimensional maximum pooling layer with a pooling kernel size of 2; The filtered data is trained in the Bi-LSTM network. There are two layers of Bi-LSTM in the Bi-LSTM network, and the number of neurons in the hidden layer is 32 and 64, respectively; After the weight of the feature vector is allocated and optimized by the Attention layer, the data passes through the Dense layer to output the prediction result.

This article uses Tensorflow to build a CNN-Bi-LSTM-Attention model. The value of epochs is set to 100, and the value of batch size is set to 256. The model adopts the EarlyStopping mechanism to improve training efficiency. The loss function uses mean squared error (MSE), and the optimizer is Adam.

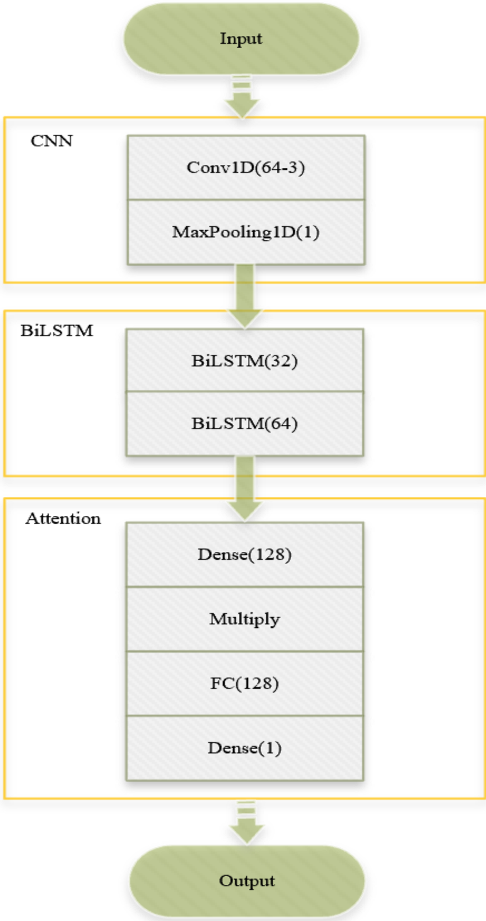

Figure 4. Structure and parameters of CNN-Bi-LSTM-Attention model

To reflect the optimal result of the CNN-Bi-LSTM-Attention model proposed in this paper to predict the trend of Bitcoin's price further, this article selects CNN-LSTM-Attention, CNN-Bi-LSTM CNN-LSTM, LSTM, SVM, BP model and the method proposed in this paper to compare on the test set. Figure 5 compares the data predicted by each model after inverse normalization. It can be seen that the prediction result of the CNN-Bi-LSTM-Attention model for the Bitcoin price is the closest to the actual value.

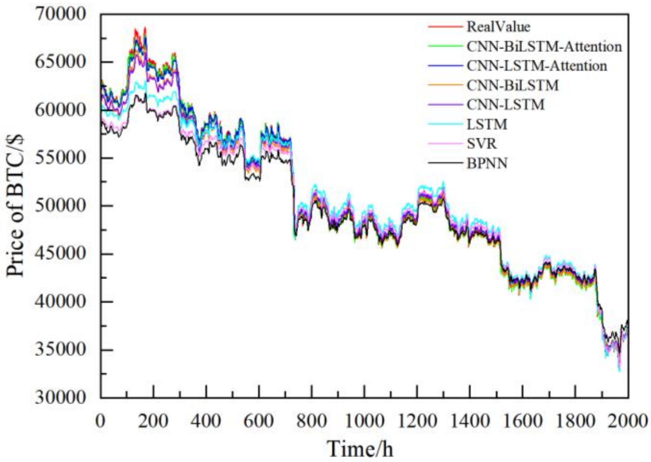

Figure 5. Predicted results of Bitcoin's price by different models

4. Conclusion and discussion

In this paper, a combined CNN-Bi-LSTM model based on the Attention mechanism is constructed. First, the useful feature information is extracted through the CNN network. Being trained by a two-layer Bi-LSTM network, the data is optimized by the Attention mechanism to the weight of the feature vector. 10,000 pieces of bitcoin price data were verified and analyzed. In the comparative experiments with different prediction methods, this model has the optimal prediction performance. In a follow-up study, more complex influencing factors to need to be considered for Bitcoin price prediction. In the future, an artificial intelligence-based bitcoin price prediction system will be studied and developed to support government decision-making and help bitcoin investors reduce risks.

References

[1] Guo H, Zhang D, Liu S, Wang L, Ding Y. Bitcoin price forecasting: A perspective of underlying blockchain transactions. Decision Support Systems. 2021; 151: 113650.

[2] Cavalli S, Amoretti M. CNN-based multivariate data analysis for bitcoin trend prediction. Applied Soft Computing. 2021; 101: 107065.

[3] McNally S, Roche J, Caton S. Predicting the price of bitcoin using machine learning. 2018 26th euromicro international conference on parallel, distributed and network-based processing (PDP): IEEE; 2018. p. 339-43.

[4] Kristjanpoller W, Minutolo MC. A hybrid volatility forecasting framework integrating GARCH, artificial neural network, technical analysis and principal components analysis. Expert Systems with Applications. 2018; 109: 1-11.

[5] Indera NI, Yassin IM, Zabidi A, Rizman ZI. NONLINEAR AUTOREGRESSIVE WITH EXOGENEOUS INPUT (NARX) BITCOIN PRICE PREDICTION MODEL USING PSO-OPTIMIZED PARAMETERS AND MOVING AVERAGE TECHNICAL INDICATORS. Journal of Fundamental and Applied Sciences. 2017; 9: 791-808.

[6] Thuy-An D, Kwon Y-K. An Empirical Study on Importance of Modeling Parameters and Trading Volume-Based Features in Daily Stock Trading Using Neural Networks. Informatics-Basel. 2018; 5.

[7] Xia T, Song Y, Zheng Y, Pan E, Xi L. An ensemble framework based on convolutional bi-directional LSTM with multiple time windows for remaining useful life estimation. Computers in Industry. 2020; 115.

[8] Tang X, Machimura T, Liu W, Li J, Hong H. A novel index to evaluate discretization methods: A case study of flood susceptibility assessment based on random forest. Geoscience Frontiers. 2021; 12: 101253.

[9] Xie G, Shangguan A, Fei R, Ji W, Ma W, Hei X. Motion trajectory prediction based on a CNN-LSTM sequential model. Science China Information Sciences. 2020; 63: 1-21.

[10] Liu Y, Dong S, Lu M, Wang J. LSTM based reserve prediction for bank outlets. Tsinghua Science and Technology. 2018; 24: 77-85.

[11] Wang S, Wang X, Wang S, Wang D. Bi-directional long short-term memory method based on attention mechanism and rolling update for short-term load forecasting. International Journal of Electrical Power & Energy Systems. 2019; 109: 470-9.