

Bitcoin price forecasting method based on CNN-LSTM hybrid neural network model

eISSN 2051-3305

Received on 14th October 2019

Accepted on 19th November 2019

E-First on 6th July 2020

doi: 10.1049/joe.2019.1203

www.ietdl.org

Yan Li¹ ✉, Wei Dai¹¹Department of Financial Engineering, Central University of Finance and Economics, Beijing, People's Republic of China

✉ E-mail: 15600129388@163.com

Abstract: In this study, aiming at the problem that the price of Bitcoin varies greatly and is difficult to predict, a hybrid neural network model based on convolutional neural network (CNN) and long short-term memory (LSTM) neural network is proposed. The transaction data of Bitcoin itself, as well as external information, such as macroeconomic variables and investor attention, are taken as input. Firstly, CNN is used for feature extraction. Then the feature vectors are input into LSTM for training and forecasting the short-term price of Bitcoin. The result shows that the CNN-LSTM hybrid neural network can effectively improve the accuracy of value prediction and direction prediction compared with the single structure neural network. The finding has important implications for researchers and investors in the digital currencies market.

1 Introduction

Nowadays, electronic payment is mainly based on traditional electronic payment tools, while digital currency relies on the virtual currency which is on the basis of the blockchain technology. Compared with traditional currency and electronic trading, it has a fast transaction speed but low cost as it uses a decentralised peer-to-peer network and does not require the third-party payment platform. What is more, the transaction is more secure and transparent because it is difficult to crack or forge by using encryption algorithms and automatic authentication mechanisms. Consequently, since the birth of digital currency, it has quickly gained widespread attention.

In particular, Bitcoin is the world's first distributed super-sovereign digital currency. It was proposed and built by a Japanese programmer Satoshi Nakamoto in 2009, relying on an electronic payment system based on encryption technology and P2P (Point to Point) technology. In April 2010, Bitcoin was first publicly traded at a price of only \$0.03 and in January 2013 the price still did not exceed \$15 with an intermittent price peak close to \$1000 in December of the same year. Moreover, in December 2017, the price reached a new high of \$19,000 but then fell sharply to \$6700 in February 2018 and even reached at 3200 in December at the same year. It can be seen that the price of Bitcoin has climbed all the way, but gradually declined, fluctuating greatly. Bitcoin also carries out seven 24-h round-the-clock trading, non-government-led, having no price limit, consequently, the price trend is obviously characterised by skyrocketing and plunging.

As a new type of investment tool, its frequent price changes naturally raise a question: whether the price of Bitcoin can be predicted. This is a significant question, especially given Bitcoin's short history and the fact that its price is easily influenced by the attitudes of governments around the world.

This paper studies whether the price of Bitcoin can be predicted based on internal information such as the historical price of Bitcoin and external information such as market factors. The data used for prediction is not only limited to the historical transaction information of Bitcoin, but also includes the macroeconomic variables and investor attention towards Bitcoin. At the same time, artificial intelligence technology is introduced into Bitcoin price prediction. In this paper, convolutional neural network (CNN) is used to extract the characteristics that have a great influence on Bitcoin price in the data set, and then long short-term memory (LSTM) is used for price forecasting. A CNN-LSTM hybrid neural network-based price prediction model is proposed. The current

research focuses on the accuracy and direction of Bitcoin price forecasting.

Based on this, the structure of this paper is as follows: Section 2 reviews the relevant literature; Section 3 introduces the research methods used; Section 4 presents and analyses the results, and finally, Section 5 is the conclusion of this paper.

2 Related literature

At present, many domestic and international scholars have conducted a great deal of in-depth discussions on the price of digital cryptocurrencies especially Bitcoin.

Due to the dramatic changes in the price of Bitcoin, many scholars have studied the factors affecting the price of Bitcoin. Jermain Kaminski (2014) studies the correlation between investor's Twitter sentiment and Bitcoin price and trading volume. Eom *et al.* [1] claim investor sentiment can help explain changes in Bitcoin volatility for future periods significantly. Ladislav Kristoufek (2013) studies Bitcoin price, finding that there is a relationship between Bitcoin price and search engine. Additionally, the price of Bitcoin would be affected by investors' speculative behaviour, and there is a bubble in it. Vassiliadis *et al.* [2] point out that there is a strong correlation between the price of Bitcoin and trading volume and transaction cost, and there is a certain relationship with gold, crude oil and stock market index. Ciaian *et al.* [3] also draw the conclusion that there is a short-term guidance lag relationship between Bitcoin price and macroeconomic variables in the empirical test. Huang *et al.* [4] use 124 technical indicators based on the historical price of Bitcoin to build a return prediction model, showing that the combination of big data and technical analysis can help predict the return of Bitcoin.

In terms of forecasting, machine learning has made breakthrough progress in recent years. More and more scholars use deep learning technology to predict Bitcoin price. Qiu *et al.* [5] introduce the wavelet analysis to the prediction of the trend of Bitcoin price over a quarter, using the time series of Bitcoin price. Jing [6] collects Bitcoin transaction data from January 2009 to March 2016, establishing a Bitcoin market forecasting model which uses the data of the previous day, week and month of Bitcoin market on the back propagation (BP) neural network. The author suggests that the longer the prediction period, the larger the prediction error, and the transaction volume is more difficult to predict compared with the price. Mallqui and Fernandes [7] employ artificial neural networks (ANN), support vector machine (SVM), and ensemble algorithms (based on recurrent neural networks (RNN) and K-means clustering methods) to predict the

Table 1 List of input attributes

Transaction information	Technical indicators	Macroeconomic variables	Investor attention
lowest price	RSI	gold price	Baidu Index
closing price	MFI	exchange rate	—
highest price	OBV	NYSE Index	—
opening price	—	NASDAQ Index	—
trading volume	—	S&P 500 Index	—
transaction amount	—	federal funds rate crude oil futures price	—

direction of Bitcoin price, and analyses the behaviour of ANN and SVM for the maximum, minimum and closing prices predictions. The study concludes that the combination of RNN and a Tree classifier can better predict the direction of Bitcoin price, meanwhile, the SVM algorithm obtains a more precise prediction than ANN in forecasting the Bitcoin price.

Therefore, this study indicates that the price of Bitcoin is affected by a variety of factors with high volatility, which is difficult to predict in traditional methods. Nevertheless, machine learning can train and learn complex non-linear data, which is suitable for Bitcoin price prediction. At the same time, considering that the price of Bitcoin changes dramatically, this paper mainly forecasts the short-term price of Bitcoin.

3 Proposed methodology

3.1 Data collection and pre-processing

Sources of information can be divided into internal information (different parameters of Bitcoin) and external information (macroeconomic factors and investor attention).

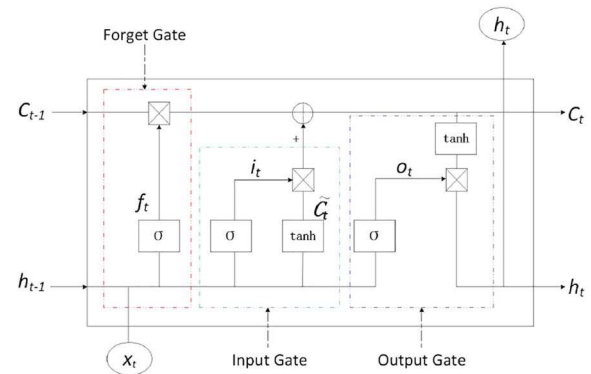
Among them, the internal information includes the opening price, the highest price, the lowest price, the closing price, the trading volume, and the transaction amount of Bitcoin. They are derived from Kraken Bitcoin exchange trading data provided by Quandl (<https://www.quandl.com/>). Huang *et al.* [4] declare that technical indicators can be used to predict the price of Bitcoin, so this paper chooses three technical indicators: relative strength index (RSI), money flow index (MFI) and on balance volume (OBV). As a new type of investment tool, the price of Bitcoin is considered to be related to macroeconomic variables. Hence, the following indicators are selected in this paper: crude oil futures price, gold price, S&P 500 Index, NYSE Index, NASDAQ Index, Federal Funds Rate, and Yuan–Dollar exchange rate. The data are all from the Wind Database.

Da *et al.* [8] point out that search activities reflect investors' concerns. When an investor searches for virtual currency in a search engine, we think he has paid attention to the virtual currency. The search index is based on the number of times a keyword is searched in a search engine. For this reason, this paper uses the Baidu Index to accurately quantify investors' attention to Bitcoin [9]. The data comes from the Baidu website (<https://index.baidu.com/#/>).

According to [6], the long prediction period may lead to enormous prediction error. So the forecast period used in this paper is 3 days, namely, the characteristic parameter data of the previous 3 days is used to predict the closing price of Bitcoin on the fourth day. Each set of samples in the constructed data set has a 51-dimensional feature that contains all the features of the previous 3 trading days. And the time interval of all data in this paper is considered, ranging from 30 December 2016 to 31 August 2018. The first 588 samples are used as the training set for model training, and the last 20 samples are applied in the test set to verify the feasibility of the model. Table 1 shows the attributes which compose the data set.

In order to prevent the influence of different dimensions of the original variables on the prediction accuracy, the original data is standardised. The calculation formula is:

$$y_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (1)$$

**Fig. 1** Basic unit of LSTM neural network

where y_{ij} is the data after standardisation, \bar{x}_j is the mean and s_j is the standard deviation of each dimension component.

3.2 CNN-LSTM hybrid neural network

3.2.1 Convolutional neural network: CNN model is one of the most typical and widely used ANN in recent years. It mimics the perception of local information by biological vision cells. Local connection and layer-by-layer calculation are used to extract the data features, and finally, the global information is synthesised through the full connection. Its basic structure includes convolution layer, pooling layer and full connection layer.

The convolution layer uses a convolution operation instead of a matrix multiplication, and each convolution kernel can extract a feature of the input data. The weight sharing method is adopted in the convolution operation, which effectively reduces the number of parameters, decreases the complexity of neural network training, and improves the training speed. The pooling layer can reduce the dimension of input data and the size of data volume. Commonly used pooling methods include mean pooling, maximum pooling etc. The formula is as follows:

$$y_j^l = f\left(\sum_{i \in M_j} x_i^{l-1} \cdot k_j + b_j^l\right) \quad (2)$$

$$x_j^l = g(\beta_j^l \text{pooling}(x_j^{l-1}) + b_j^l) \quad (3)$$

where f and g are the activation function, M is the convolution kernel, b^l is the bias, k_j is the weight matrix of the convolution kernel, β_j^l is the coefficient of the channel corresponding to the pooling layer and $\text{pooling}()$ is the pooling function.

3.2.2 LSTM neural network: LSTM [10] is an improved RNN model, which can effectively solve the problem of gradient disappearance and gradient explosion in the RNN model. It is suitable for processing long-term sequence data and solving long-term dependence. Its basic unit is the memory module, containing the memory unit and three gates controlling the memory unit, namely Input Gate, Output Gate and Forget Gate. The gate is the structure that determines the selective passage of information. If the output value of the sigmoid function is 0, it is discarded completely, while if it is 1, it passes completely. Fig. 1 shows the basic unit of LSTM neural network.

(i) *Forget Gate*: The purpose of Forget Gate is to determine what information will be discarded. Reading in the output of the previous layer h_{t-1} with the current input x_t , the gate outputs f_t and assigns the current cell C_{t-1} , the calculation formula of f_t is as follows:

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (4)$$

where σ presents the sigmoid function.

(ii) *Input Gate*: The role is to update based on existing information. First, run the sigmoid function to get i_t and decide which values to enter. Then, according to the tanh function, a candidate value vector \tilde{C}_t is obtained, which is multiplied with i_t and added to the state C_t . The formula for this part is as follows:

$$i_t = (W_i x_t + U_i h_{t-1} + b_i) \quad (5)$$

$$\tilde{C}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c) \quad (6)$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \quad (7)$$

(iii) *Output Gate*: Output the information of the current point. After running a sigmoid function to get o_t and determining which parts will be output, C_t is processed by the tanh function to obtain a value between -1 and 1 . Finally, the value is multiplied with o_t to decide the ultimate output:

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (8)$$

$$h_t = o_t \cdot \tanh(C_t) \quad (9)$$

3.2.3 CNN-LSTM hybrid neural network: The structure of the CNN-LSTM hybrid neural network model proposed in this paper is shown in Fig. 2. The model consists of two parts. The first part is the CNN part, which is mainly responsible for data input and feature extraction. And the input is a feature graph with the size of 3×17 , arranged in time series. There are two convolution layers (Conv2D) in the CNN part. The number of the convolution kernels is 15 and 10 in turn, and both of the sizes of the two convolution kernels are 3×3 . The pooling layers (MaxPooling2D) all adopt the maximum pooling method. After two successive convolution and pooling operations, there is the full connection layer (Dense), extracting the characteristic data as a one-dimensional vector array whose length is 50.

In the second part, that is to say, the LSTM part, the output of the CNN part is used as the input of the LSTM neural network. This section consists of one LSTM layer and a full connection layer. Among them, the number of nodes is 30 and the learning rate is 0.01. Mean squared error (MSE) is used as the loss function and Adam is used as the optimisation method. Meanwhile, to avoid over-fitting, the Dropout layers are added to randomly inactivate some neurons. Moreover, with the increase of training times, the model will produce the phenomenon of over-fitting. On the contrary, the fitting effect is not ideal if the training times are not sufficient. Hence, for this model, the number of iterations is set to 130, and the size of the training batch is 5, which can obtain a better performance. This algorithm is based on the Keras deep learning framework.

3.3 Performance metrics

The ability of different neural networks to predict Bitcoin price is reflected in two aspects: accuracy of value prediction and direction prediction. In this paper, the mean absolute error (MAE), the root mean squared error (RMSE) and the mean absolute percentage error (MAPE) are used as performance indexes to quantify the ability of neural network to predict the value, meanwhile, the Precision, Recall and F_1 -Measure are introduced to measure the ability of price direction prediction.

Calculation formula is:

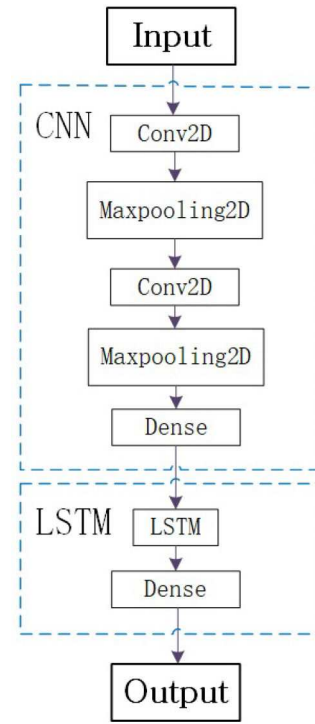


Fig. 2 Structure of the CNN-LSTM hybrid neural network model

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - f_i| \quad (10)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - f_i)^2} \quad (11)$$

$$MAPE = \frac{100}{N} \sum_{i=1}^N \left| \frac{y_i - f_i}{y_i} \right| \quad (12)$$

where y_i is the i th true value, f_i is the i th predicted value, and N is the number of data to be evaluated. MAE, RMSE and MAPE reflect the degree of deviation of the predicted value from the true value. The smaller the MAE, RMSE and MAPE, the higher the accuracy of the prediction.

$$\text{precision} = \frac{TP}{TP + FP} \quad (13)$$

$$\text{recall} = \frac{TP}{TP + FN} \quad (14)$$

where TP = True Positive; TN = True Negative; FP = False Positive; and FN = False Negative. F_1 is the harmonic mean of Precision and Recall, which is used to comprehensively reflect the classification effect of the prediction model.

$$F_1 = \frac{2 \times \text{precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (15)$$

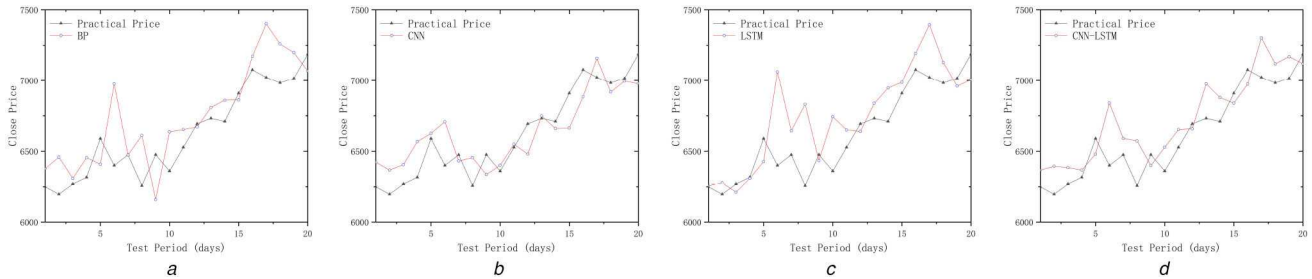
In order to obtain statistically significant values, each model was evaluated (trained and tested) 30 times and averaged as the performance metrics scores.

4 Result and discussion

This paper also uses the BP neural network, CNN and LSTM after parameter tuning to train the same data as well as predict the closing price of Bitcoin. Table 2 shows the model parameters used by BP, CNN and LSTM neural network to predict the closing price of Bitcoin. All models are trained 30 times. Moreover, the predictions of the single structure neural network models are

Table 2 Model parameters used by single structure neural network

Neural network	Structure	Number of nodes/filters	Learning rate	Loss function	Number of iterations
BP	single hidden layer	60	0.01	MSE	100
CNN	Conv2D Maxpooling2D Conv2D Maxpooling2D	1,510,201	0.001	MSE	100
—	Conv2D Maxpooling2D Dense	—	—	—	—
LSTM	LSTM LTSM LSTM Dense	2,510,151	0.01	MSE	100

**Fig. 3** Comparison diagram of prediction and practical curves

(a) BP and practical curves, (b) CNN and practical curves, (c) LSTM and practical curves, (d) CNN-LSTM and practical curves

Table 3 Comparison of model prediction result

Neural network	MAE	Value prediction			Direction prediction	
		RMSE	MAPE, %	Precision	Recall	F_1
BP	416.09	515.35	2.85	0.59	0.58	0.55
CNN	215.98	261.90	2.03	0.64	0.63	0.60
LSTM	229.78	297.97	2.75	0.58	0.73	0.63
CNN-LSTM	209.89	258.31	2.35	0.64	0.81	0.69

compared with those predicted by the CNN-LSTM hybrid neural network.

Fig. 3 shows the comparison between the Bitcoin price predicted by different models and the practical price, respectively. It can be seen that the predicted results of CNN-LSTM have a good fit with the practical price trend of Bitcoin.

In Table 3 is presented the average results of performance metrics for value prediction and direction prediction, respectively. It indicates that the BP neural network has the worst performance among all the models, and CNN has more advantages in value prediction than LSTM, while LSTM performs better in direction prediction. As we can see, CNN-LSTM performs best in both value prediction and direction prediction. The hybrid neural networks have a much better forecasting effect than neural networks with a single structure. Although these performance metrics scores are still not very low, owing to the high volatility of Bitcoin, and therefore this is still a challenge for our future efforts.

5 Conclusion

As previously presented, this paper uses deep learning technology to predict the price of Bitcoin with CNN-LSTM hybrid neural network as the core. Different from traditional research, this paper not only considers Bitcoin's own transaction information, but also includes external factors such as macroeconomic variables and investor's attention, comprehensively employing various factors that may affect Bitcoin prices. This hybrid neural network is compared with the single structure neural network in terms of the value prediction and direction prediction. The results show that

CNN-LSTM hybrid neural network performs well in Bitcoin forecast and is more suitable for Bitcoin prediction.

In quantifying investor attention, we use the Baidu Index of the keyword 'Bitcoin'. Therefore, one of the possible improvements is to incorporate the popularity of Bitcoin on other but worldwide social software, such as Twitter and Instagram into the data.

6 References

- [1] Eom, C., Kaizoji, T., Kang, S.H., *et al.*: 'Bitcoin and investor sentiment: statistical characteristics and predictability', *Physica A, Stat. Mech. Appl.*, 2019, **514**, pp. 511–521
- [2] Vassiliadis, S., Papadopoulos, P., Rangoussi, M., *et al.*: 'Bitcoin value analysis based on cross-correlations', *J. Internet Bank. Commer.*, 2017, **22**, (S7), p. 1
- [3] Ciaian, P., Rajcaniova, M., Kancs, D.: 'The economics of BitCoin price formation', *Appl. Econ.*, 2016, **48**, (19), pp. 1799–1815
- [4] Huang, J.Z., Huang, W., Ni, J.: 'Predicting Bitcoin returns using High Dimensional technical indicators', *J. Finance Data Sci.*, 2018, **5**, pp. 140–155, <http://www.sciencedirect.com/science/article/pii/S2405918818300928>
- [5] Qiu, L., Fanbin, W., Longxiu, Y., *et al.*: 'Prediction of Bitcoin price based on wavelet analysis', *Neijiang sci. Technol.*, 2015, **36**, (5), pp. 129–130
- [6] Jing, L.: 'Using BP neural network to construct Bitcoin market forecasting model', *Finance Account. Month.*, 2016, **536**, (21), pp. 33–36
- [7] 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', *Appl. Soft Comput. J.*, 2019, **75**, pp. 596–606. <http://www.sciencedirect.com/science/article/pii/S1568494618306707>
- [8] Da, Z., Engelberg, J., Gao, P.: 'In search of attention', *J. Finance.*, 2011, **66**, (5), pp. 1461–1499
- [9] Yizhe, L.: 'The study of investor attention's impact on the stock performance based on Baidu Index' (Hunan University, People's Republic of China, 2015)
- [10] Hochreiter, S., Schmidhuber, J.: 'Long short-term memory', *Neural Comput.*, 1997, **9**, (8), pp. 1735–1780