Proceedings of the 3rd International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/4/20230510

An application of deep learning on bitcoin price forecasting

Rui Zhong

Queensland University, Sir Fred Schonell Drive St Lucia, QLD 4072 Australia.

s4722643@student.uq.edu.au

Abstract. The long short-term memory (LSTM) network and a cutting-edge method that combines wavelet decomposition and LSTM (W-LSTM) were applied to deep learning in this study's analysis of Bitcoin's price and movement. To be specific, it predicted next day's both price and price movement (trend) with historical data. The input of the model is close price itself, basic trading information, and technical indicators calculated solely on basic trading information. Large number of numerical experiments come to the same conclusion that: for price prediction, only close price as input obtains the best performance for regression, and minor improvement achieved after 1-order wavelet decomposition; for price movement, no improvement after changing the number of input features or with the model W-LSTM has been spotted for the same network structure and hyper-parameters, and enlarging time step and batch size will improve accuracy and Matthews correlation coefficient despite of number of input and model used in this paper.

Keywords: deep learning, Bitcoin price forecasting, price trend forecasting.

1. Introduction

In last several years, cryptocurrencies have drawn large quantity of attention from financial investors and fund managers. It is increasingly not surprising to spot cryptocurrencies in investment portfolios founded and managed by financial institutions. Compared with well-studied traditional assets, such as bonds, stocks, options and futures, cryptocurrencies, together with their own properties and characteristics as assets, are still in the process of understanding [1]. However, like traditional assets, cryptocurrencies have time series containing trading information, so the method investigating traditional financial time series is also applicable for them. Though there exits more than 5000 digital currencies in the world, Bitcoin is the most famous and prevalent one among them. It even has more than 5.8 million dynamic clients and more than 111 exchanges worldwide [2]. Therefore, the study object of this paper is Bitcoin as an asset.

For Bitcoin research, the most common research target is price [3-6] and volatility [7]. Though traditional statistical models, e.g. autoregressive integrated moving average mode, are still used in investigating Bitcoin, machine learning method, especially deep learning, is more effective and prevalent. Though dozens of machine learning method exist, recursive neutral network (RNN) based price/trend forecasting models dominates the area of time series prediction [5, 8]. And this trend will probably last for quite a long time, mainly because they are easy to adapt to most asset forecasting problems [8]. At the same time, improved variations of the original long short-term memory (LSTM) or RNN model are becoming popular and are typically included with hybrid learning systems[5, 8]. In

© 2023 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

this paper, original LSTM network and a novel technique combines wavelet decomposition and LSTM (W-LSTM) [9-11] are used.

Most financial time series forecasting applications do not place as much emphasis on price prediction accuracy as they do on directionality accuracy [5, 8]. However, the major ambition of the article is to investigate application of deep learning method on Bitcoin and its potential improvement, so this paper is intended to predict next day's both price and price movement (trend) with historical data, the former is a regression problem while the latter is a classification problem. The most prevalent number of class in asset price movement classification is 2-class, i.e. up or down movements; it's less common to see 3-class classification, i.e. up, down, or neutral movements, or any other number of class. In a two-class categorization, a price that is up indicates that it will be higher the following day than it is now, whereas a price that is down indicates the opposite.

2. Method

This subsection first introduces model LSTM and W-LSTM, and then lists metrics evaluating regression and classification performance.

2.1. Model selection

Among the family of recursive neural network (RNN), the LSTM based network structure is the most prevailing one. The dominance of LSTM lies on delicately and creatively constructing the "gate" structure. This allows it to properly learn the long-term dependence and avoids the gradient disappearance and gradient explosion caused by the traditional RNN [12]. The LSTM model thus has significant advantages when handling the prediction and classification of time series.

As shown in illustration. A number of isomorphic cells make up an LSTM unit, which can update its internal state to store information for an extended period of time [12]. Three cells share the same cell structure, as shown by the letter A. The input gate, forgetting gate, output gate, and cell state are the four basic components of each cell [13].

$$\mathrm{f}_t = \sigma\left(\mathrm{W}_f \times [\mathrm{h}_{t-1}, \mathrm{x}_t] + \mathrm{b}_f\right) \tag{1}$$

$$\mathrm{i}_t = \sigma(\mathrm{W}_i \times [\mathrm{h}_{t-1}, \mathrm{x}_t] + \mathrm{b}_i) \tag{2}$$

$$\tilde{\mathrm{C}}_t = \tanh(\mathrm{W}_c \times [\mathrm{h}_{t-1}, \mathrm{x}_t] + \mathrm{b}_c) \tag{3}$$

$$\mathrm{C}_t = \mathrm{f}_t \cdot \mathrm{C}_{t-1} + \mathrm{i}_t \cdot \tilde{\mathrm{C}}_t \tag{4}$$

$$\mathrm{O}_t = \sigma(\mathrm{W}_o \times [\mathrm{h}_{t-1}, \mathrm{x}_t] + \mathrm{b}_o) \tag{5}$$

$$h_t = \mathrm{o}_t \cdot \tanh(\mathrm{C}_t) \tag{6}$$

In which represents the unit state; The subscript represents the time; Is the activation function of sigmoid and tanh, respectively; Is the weight and deviation matrices, respectively; Is the input vector of LSTM unit; Is the cell output vector; Is a forgetting gate, an input gate, and Is an output gate, respectively.

The cell state, which is the essential component of the LSTM unit, is modified by the input gate and forgetting gate and preserves the memory of the cell state throughout time [12]. The forgetting gate's purpose is to allow the cell to remember or forget its previous state, the input gate's purpose is to permit or disallow the incoming signal from updating the cell state, and the output gate's purpose is to regulate the output and transmission of the unit state to the following cell [12]. The back-propagation algorithm is typically the most popular training method, and the internal structure of the LSTM unit is built of many perceptrons [12].

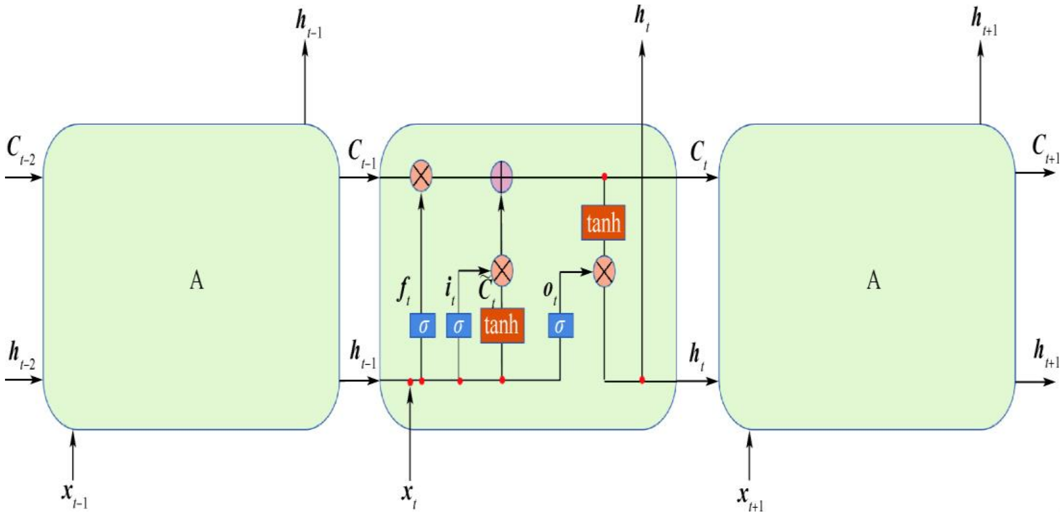

Figure 1. LSTM unit.

Wavelet transform and LSTM are combined to generate the W-LSTM. A popular analysis technique in the field of signal processing is the Fourier transform. It may change signals from the time domain to signals from the frequency domain. In the temporal domain, the Fourier transform has no power to discriminate. So, the wavelet transform is developed in view of the shortcomings of Fourier transform. It uses wavelet and a family of band-pass filters to decompose the original time-domain function and decompose the signal into two-dimensional time-frequency information, which greatly enhances the performance of local signals and improves the noise resistance of the model.

A technique for data decomposition and reconstruction is the wavelet transform. First, a low-pass filter and a high-pass filter are used to separate the original data into low-frequency and high-frequency wavelet coefficients, respectively. It is possible to further deconstruct certain of them, namely the low frequency wavelet coefficients, and to do so repeatedly up to the point at which the maximum decomposition times are attained. The low frequency signal and high frequency signal are obtained after decomposing the original data and reconstructing the low frequency wavelet coefficients and high frequency wavelet coefficients, respectively. The low frequency signal denotes approximation information, and the high frequency signal denotes detail information. To complete data restoration, all high frequency and low frequency signals are added. For W-LSTM, the obtained signals, rather than the raw time-series, are the input of LSTM; then for regression problem, the predicting result are summed together, and for classification problem, only low frequency signal are used to predict the labels.

2.2. Evaluation metrics

To measure how well the model predicts prices, mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) are selected as assessment indices [14]. The greater the divergence between the projected outcomes and the actual values, and the less accurate the findings are, the larger the values of MSE, RMSE, MAE, and MAPE are [14] The details of the formula are as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(\hat{y}_n - y_n)^2} \tag{7}$$

$$\mathrm{MAPE} = \frac{1}{N}\sum_{n=1}^{N}\frac{|\hat{y}_n - y_n|}{y_n} \tag{8}$$

DOI: 10.54254/2755-2721/4/20230510

$$\text{MAE} = \frac{1}{N}\sum_{n=1}^{N}|\hat{y}_n - y_n| \tag{9}$$

$$\text{MSE} = \frac{1}{N}\sum_{n=1}^{N}(\hat{y}_n - y_n)^2 \tag{10}$$

There are the required number of samples in this. is the true value; it corresponds to what the model predicted.

For predicting price movement, accuracy (ACC) and Matthews correlation coefficient (MCC) are used. In a two-class classification, samples can be divided into true positive (TP), true negative (TN), false positive (FP), and false negative (FN) categories based on the interaction between the actual values of the samples and the projected values of the model. Accuracy is a statistic for all samples that compares the proportion of correctly identified samples to all samples. MCC is a thorough indicator that considers TP, TN, FP, and FN. As a fairly balanced indication, it can also be used when the sample is unequal.

The MCC's value range is [- 1, 1]. A score of 1 means the predicted outcome is entirely consistent with the actual value, a value of 0 means it is not as good as the random prediction, and a value of -1 means it is entirely inconsistent with the actual value. The details of the formula are as follows:

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{TP + TN + FP + FN} \tag{11}$$

$$\text{MCC} = \frac{\text{TP} \cdot \text{TN} - \text{FP} \cdot \text{FN}}{\sqrt{(TP + FP) \cdot (TP + FN) \cdot (TN + FP) \cdot (TN + FN)}} \tag{12}$$

3. Experiment

This section contains following subsections, i.e. data description, which describes data source and its characteristics; determinants, which lists indicators that are fed into the model; implementation of LSTM and W-LSTM, declares model structure and settings of hyper-parameter about training; regression performance and classification performance show the performance respectively.

3.1. Data description

The trading information of Bitcoin are collected from website 'www.investing.com', which offers free real time financial news, quotes, historical and live market data about stock and other financial assets, such as Bitcoin. The collected information includes the open, high, low, and close prices of Bitcoin as well as the daily trading volume, or the fundamentals of Bitcoin trading. The sample period of these data lasts from June 2010 to June 2022.

3.2. Determinants

As has mentioned above, feature engineering is not the focus of this paper, so it did not plan to obtain determinants that contain more information affecting trading, such as macro and micro economic indicators, exchange rate between Bitcoin and major currencies, public attention that is often quantified by the searching volume of Bitcoin-related keywords on major search engines, daily production of Bitcoin, etc. However, technical indicators calculated from the basic trading information are selected for that they are directed obtained from basic trading information, widely used in financial asset transaction, and can be deemed as a method of data process on basic trading information. The technical indicators obtained in this paper are shown in Table 1 [15]. After dropping the missing data, 4338 days of data spanning from August 14th 2010 to June 29th 2022 are obtained. For diminishing the volatility of the magnitude of data, these data are scaled to [0, 1].

DOI: 10.54254/2755-2721/4/20230510

Table 1. Basic trading information and technical indicators calculated on them.

Indicator	Description
OP	Open price
HP	High price
LP	Low price
CP	Close price
Volume	Trading volume
CCI	Commodity Channel Index
SAR	Stop and reverse index
ADX	Average directional movement
MFI	Money flow index
RSI	Relative strength index
SK	Slow stochastic %K
SD	Slow stochastic %D
RSI-S	RSI indicator signal
BB-S	Bollinger bands indicator signal
MACD-S	MACD indicator signal
SAR-S	SAR indicator signal
ADX-S	ADX indicator signal
S-S	Stochastic indicator signal
MFI-S	MFI indicator signal
CCI-S	CCI indicator signal
V-S	Sign (Volume-Avg (last 5 days))
CPOP-S	Sign(CP-OP)
CPCPY-S	Sign(CP-Closing price yesterday)

3.3. Implementation of LSTM and W-LSTM

This subsection consists of three categories, i.e. input data setting, network setting, and training setting. All parameters not mentioned are set by default. LSTM and W-LSTM shares the same network structure and hyper-parameter settings. In implementation, python package Tensorflow and Keras, and PyWavelets are used.

Input data setting: training set consists of first 90% of the dataset, and testing set contains the last 10%, and data are scaled to 0-1 by maximum and minimum value normalization, and restore the predicted value after obtaining the predicted value; Assuming that the time step is set to 5, that is, price on t day is predicted with information on $\mathrm{t}-5, \mathrm{t}-3, \ldots \mathrm{t}-1$ day, in this paper time step is set to 100, and the data is pre-processed to meet the data input requirements of Tensorflow recurrent neural network.

Network setting: LSM uses four-layer network structure settings, including the initial input and output layer, the LSTM layer, the dropout layer, and the LSTM layer. The activation function is set to tanh, the recursive activation function is set to sigmoid, and the dropout ratio is set to 0.2 for the first LSTM layer and 64 neurons for the second layer. For classification, a softmax activation function is employed; for regression, no activation function is specified for the output layer. The LSTM's structure is shown in Fig.2.

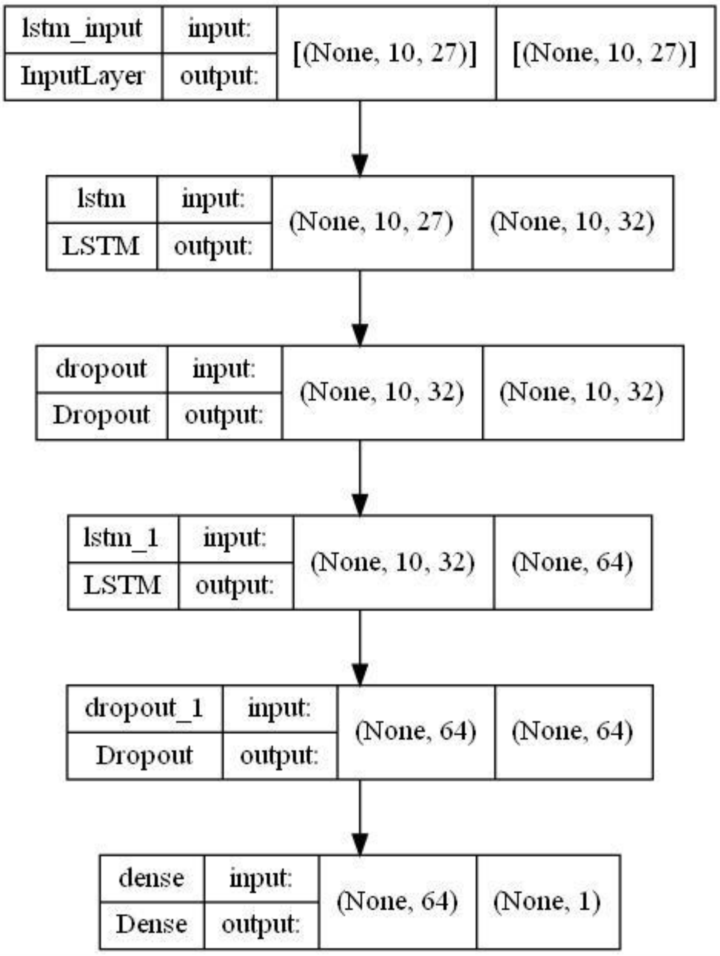

Figure 2. LSTM structure.

Training parameters include Adam as the optimizer, 0.001 as the learning rate, sparse categorical cross-entropy as the loss function for classification, 20 epochs, 100 batches, and 20% of the training set for cross validation. Mean square error is set as the regression loss function. The validation frequency is set to one.

3.4. Regression performance

The Table2 shows the performance of Bitcoin price prediction with model LSTM and W-LSTM. In Table 2, 'all' denotes all indicators listed in Table 1 are the input of the model; 'basic' denotes that the basic trading information are fed into the model; 'only CP' means only close price are used to predict next day's close price. In LSTM, for the current model structure and hyper-parameter settings, only close price as input obtains the smallest MSE, RMSE, MAE and MAPE, i.e. best performance, while basic trading information has the worst performance. Even though all indicators are exploited, LSTM model gains no better performance. W-LSTM with close price as input has a slightly better performance than LSTM with close price as input. Large number of numerical experiments, including setting different random seed, and changing network structure and other settings, have been done. Nearly all the experiments are consistent with the current finding. The conclusion holds: only close price as input obtains the best performance for regression; minor improvement achieved after 1-order wavelet decomposition. With the increase of number of features, it's more time-consuming and difficult to fine train a model. However, the practice of more features as input is not a guarantee for better performance, at least for the prediction of Bitcoin price.

Table 2. Regression performance of model LSTM and W-LSTM.

	LSTM			W-LSTM
	all	basic	only CP	Only CP
MSE	10255760	89238803	5947275	5479748
RMSE	3202	9447	2439	2340.886
MAE	2523	4149	1860	1786.549
MAPE	0.0598	0.1027	0.0456	0.043667

3.5. Classification performance

Compared with regression, classification problem witnessed no improvement after changing the number of input features or with the model W-LSTM. As is demonstrated in Table 3, for the same network structure and hyper-parameters, ACC and MCC proves the incapability of model to predict price movement. As a matter of fact, the finding that ACC and MCC remain the same, is just an accident, the truth is that enlarging time step and batch size will improve ACC and MCC despite of number of input and model used. However, large number of numerical experiments have proved that ACC is no larger than 0.52, and MCC is always around 0.

Table 3. Classification performance of model LSTM and W-LSTM.

	LSTM			W-LSTM
	all	basic	only CP	Only CP
ACC	0.4906	0.4906	0.4906	0.4906
MCC	-0.0013	0.0000	0.0000	0.0000

4. Conclusion

This paper implemented an application of deep learning method, i.e. LSTM and W-LSTM, on Bitcoin price and price movement. To be specific, it predicted next day's both price and price movement (trend) with historical data. The input of the model is close price itself, basic trading information, and technical indicators calculated solely on basic trading information. Large number of numerical experiments, including setting different random seed, and changing network structure and other settings, have been done. Nearly all the experiments are consistent with the conclusion that: for price prediction, only close price as input obtains the best performance for regression, and minor improvement achieved after 1-order wavelet decomposition; for price movement, no improvement after changing the number of input features or with the model W-LSTM has been spotted for the same network structure and hyper-parameters, and enlarging time step and batch size will improve ACC and MCC despite of number of input and model used. With the increase of number of features, it's more time-consuming and difficult to fine train a model. However, the practice of more features as input is not a guarantee for better performance, at least for the prediction of Bitcoin price and price movement.

References

[1] Fang F., Ventre C., Basios M., Kanthan L., Martinez-Rego D., Wu F. and Li L., 2022, Cryptocurrency trading: a comprehensive survey, Financial Innovation , 8(1) doi:10.1186/s40854-021-00321-6.
[2] Rizwan M., Narejo S. and Javed M. 2019, Bitcoin price prediction using Deep Learning Algorithm14-15 Dec. 2019) p 1-7
[3] McNally S., Roche J. and Caton S., 2018, Predicting the Price of Bitcoin Using Machine Learning, 339-343 doi:10.1109/pdp2018.2018.00060
[4] Lamothe-Fernández P., Alaminos D., Lamothe-López P. and Fernández-Gámez M. A., 2020, Deep Learning Methods for Modeling Bitcoin Price, Mathematics, 8(8): 1245 doi:10.3390/math8081245

[5] Ozbayoglu A. M., Gudelek M. U. and Sezer O. B., 2020, Deep learning for financial applications : A survey, Applied Soft Computing, 93: 106384 doi:10.1016/j.asoc.2020.106384
[6] Liu M., Li G., Li J., Zhu X. and Yao Y., 2021, Forecasting the price of Bitcoin using deep learning, Financ. Res. Lett., 40: 101755 doi:10.1016/j.frl.2020.101755
[7] D'Amato V., Levantesi S. and Piscopo G., 2022, Deep learning in predicting cryptocurrency volatility, Physica A, 596: 127158 doi:10.1016/j.physa.2022.127158
[8] Sezer O. B., Gudelek M. U. and Ozbayoglu A. M., 2020, Financial time series forecasting with deep learning : A systematic literature review: 2005–2019, Applied Soft Computing, 90: 106181 doi:10.1016/j.asoc.2020.106181
[9] Liu Y., Guan L., Hou C., Han H., Liu Z., Sun Y. and Zheng M., 2019, Wind Power Short-Term Prediction Based on LSTM and Discrete Wavelet Transform, Applied Sciences, 9(6): 1108 doi:10.3390/app9061108
[10] Liang X., Ge Z., Sun L., He M. and Chen H., 2019, LSTM with Wavelet Transform Based Data Preprocessing for Stock Price Prediction, Mathematical Problems in Engineering, 1340174 doi:10.1155/2019/1340174
[11] Zeng Y., Chen J., Jin N., Jin X. and Du Y., Air quality forecasting with hybrid LSTM and extended stationary wavelet transform, Building and Environment, 2022, 213: 108822 doi:10.1016/j.buildenv.2022.108822
[12] Ruan G., Wang Y., Wang S. L., Zheng Y., Guo Q. and Shulga S. N., AUTOMATIC RECOGNITION OF RADAR SIGNAL TYPES BASED ON CNN-LSTM, 2020, 79(4): 305-321 doi:10.1615/TelecomRadEng.v79.i4.40
[13] Liao W., Ye G., Yin Y., Yan W., Ma Y. and Zuo D. 2020 The Inventory of Spare Parts Prediction based on LSTM for Large Data and a Short Replacement Cycle2-5 Nov. 2020) p 1-5
[14] Ma Y., Tang Y., Li B. and Qi B. 2020 Residential High-Power Load Prediction Based on Optimized LSTM Network23-25 Oct. 2020) p 538-541
[15] Jafari A. and Haratizadeh S., GCNET: graph-based prediction of stock price movement using graph convolutional network, arXiv preprint arXiv:2203.11091, 2022.