Sentiment-based Bitcoin Movement Prediction with Deep Learning

Yiyang Li

Information & Computer Sciences University of California, Irvine Irvine, United States

Yiyal12@uci.edu

Abstract—Bitcoin movement prediction has been a research topic, especially with the Bitcoin bubble period during the COVID-10 pandemic era. It has been discussed that sentiment factors are influential in the prediction process. In this study, to validate the helpfulness of these emotion-related features, three open datasets with sentiment are used, and three types of neural networks, including LSTM, GRU, and TCN, are chosen as the prediction models, using a regression model to study the relationship between them. The result shows that the sentiment-based factors did help in two of the three datasets.

Keywords: Bitcoin; sentiment; neural networks; deep learning.

1. Introduction

Blockchain technology gave birth to Bitcoin, a new generation of cyber currency, which has drawn the public's attention in the past decade. Bitcoin trading itself has become a new trading market in the past, becoming another choice other than the stock market or the golden trading market. Besides bitcoin, many digital currencies are traded on particular crypto exchanges, which most traditional investors cannot access. During the COVID-19 pandemic, more and more people used Bitcoin as a trading tool to prevent macroeconomic risks. Some countries have begun to accept Bitcoin as a new investment method in law. For example, El Salvador has become the first country to adopt Bitcoin as a legal tender in 2021. Some Internet companies have also begun to accept Bitcoin as a payment method, such as Tesla and Amazon. As a new financial tool, Bitcoin is becoming more and more critical in modern financial markets.

In 2020, the change in the market value of Bitcoin was 318%, according to Messari. While the change in the market value of gold is only 21%, and the change in the market value of the equities is only 15%. As the most prominent digital asset in market capitalization, Bitcoin has shown an exponential increase in its growth rate. Its performance is significantly better than most major asset classes. Some people attribute this deviation to the economic recession triggered by the COVID-19 pandemic, which has also triggered inflationary pressures brought about by the increase in the money supply.

Driven by the potential high profit with a precise movement prediction of Bitcoin, more and more studies have been conducted in this direction, both from academia and the industry. Some

ICIDC 2022, June 17-19, Qingdao, People's Republic of China

DOI 10.4108/eai.17-6-2022.2322739

previous studies are based on technical analysis of historical prices, transaction volumes, and data like Blockchain statistics. However, this analysis often fails to consider related events and emotional factors towards Bitcoin. Therefore, the prediction effect is not satisfactory. More and more studies have recently begun to consider sentiment analysis based on Tweets by extracting the sentiment factors contained in tweets related to Bitcoin transactions as input variables of the prediction model, hoping to achieve better prediction results. Even though researchers have been exploring these different research topics, there are not enough comparisons between different approaches. It remains unknown which model is the best choice to predict the change of bitcoin price.

Instead of predicting the specific price values, we aim to predict the price movement direction in this study, determined by the daily or hourly change of close prices. In practice, the movement triggers the transaction signal. It is modeled as a binary classification problem, and three different deep learning models are used to solve this problem, namely, LSTM, GRU, and TCN. To make a comprehensive comparison, three open datasets with sentiment factors collected in different periods are leveraged to compare the performances of these models. Including sentiment factors abstract from Twitter, Google trend, and Reddit, all are the most widely used social media platforms and can reflect perspectives about bitcoin from users all over the world.

Our main results are as follows:

(1) We find no single winner for all the three datasets we use in this study. Interestingly, LSTM, GRU, and TCN are the best model for each of the three datasets.

(2) We find that for two of three datasets, the emotion factors are helpful. However, this conclusion cannot be extended to all three datasets.

The related work is discussed in Section 2; Datasets are described in Section 3; Models are introduced in Section 4; Experiments and results are discussed in Section 5; The conclusion is drawn in Section 6.

2. Related Work

The author in [1] uses Twitter sentiments and Google Trends data, LSLR, and the Bayesian ridge regression model to forecast short-term prices of primary cryptocurrencies, BTC, ETH, ETN, XRP, and ZEC. Results show a significant relationship between Google Trends data with crypto data and tweet frequency data. After using the bot to conduct 1 to 3 transactions per day for a month, the account balance changed from $100 to $114.82.

The authors in [2] use historical data from 1 January 2017 to 31 October 2020 of BTC, ETH, and XRP. They compare a MICDL model with two CNN-LSTM models. They find that MICDL has better-exploited training data and can predict the price movement with higher accuracy and reliability than the CNN-LSTM models.

The authors in [3] use a dataset consisting of four parts, namely, general market statistics, Blockchain statistics, social network statistics, and volatility. They use the Black-Scholes model and LSTM to predict BTC option pricing. The length of the prediction window size is also thoroughly studied.

The authors in [4] use the cryptocurrency transaction database and daily user reviews on online forums from the third quarter of 2018 to the first quarter of 2019. They use models including BPNN, RBF, SVM, CNN, and LSTM, for the price prediction of BTC. The result shows that LSTM has the best prediction performance with the smallest MAE, and the addition of comments' sentiment can significantly improve prediction accuracy.

The authors in [5] use the trade data to compute intraday BTC-USD returns and StockTwits data to measure investor sentiment and investor attention. They use the OLS regression and Granger causality tests to predict BTC returns. They find that investors' sentiment allows predicting the evolution of returns, but only for high frequencies (up to 15 minutes).

The authors in [6] use historical trading data of Bitcoin collected from five bitcoin exchanges that traded the most Bitcoins. Attentive LSTM and embedding networks are used for Bitcoin fluctuation prediction. The authors in [7] use bitcoin data from Coindesk and collect tweets and Textblob sentiment polarity. MLR, price assumption. The authors in [8] use the daily closing price of bitcoin and ten types of internal (Bitcoin trading data) and external (macroeconomic variables and investor attention) information. They use the combination of MRC and LSTM to predict price.

The authors in [9] use bitcoin returns data from CryptoCompare, and EPU (economic policy uncertainty), EURQ (economic uncertainty related queries), and EGARCH model to predict bitcoin returns. The authors in [10] use historical market data was obtained from the top-performing 65 cryptocurrency exchanges and social data obtained from raw tweets from Twitter. MLP, SVM, RF are used to predict price movement.

We summarize the related work in Table 1. More relevant studies about financial market prediction can be found in recent surveys [11, 12].

Table 1. Related Work Summary

Reference	Dataset	Model
[1]	Historical Price Data; Twitter sentiments; Google Trends	LSLR. Bayesian ridge regression models
[2]	Historical Price Data	2 CNN-LSTM model and 1 MICDL model
[3]	General Market Statistics; Blockchain Statistics; Social Network Statistics; Volatility	BS Model and LSTM
[4]	Cryptocurrency Dataset;	BPNN, RBF, SVM, CNN, and LSTM
[5]	Trade Data; StockTwits data	OLS regression and Granger causality tests.
[6]	Historical trading data	Attentive LSTM and embedding network
[7]	Bitcoin data; Tweets	MLR
[8]	Daily closing price of bitcoin and 10 types of internal and external information.	MRC-LSTM
[9]	Bitcoin returns data; EPU; EURP	EGARCH model
[10]	Historical market data; raw tweets from twitter	MLP, SVM, RF

3. Datesets

In this study, we use three open datasets from the literature. We use the extracted sentiment scores for each dataset instead of extracting the score by ourselves since the raw text data are not available. For Dataset #1 [11], the sentiment score is extracted by the Vader method from the Twitter scraper library. For Dataset #2 [12], the sentiment score is extracted by the Vader method from all the tweets the authors collected through the python module Twitter-Scraper. For Dataset #3 [13], the sentiment score is extracted by the Vader method from daily basis tweets collected by fetching tweets from Twitter streaming API and saving them in a time-series database.

We show the summary of the three datasets in Table 2 and the attributes in Table 3-5.

Table 2. Summary of Datesets Used in This Study

Dataset	Time Range	Frequency
Dataset#1 [11]	From 2017-04-01 to 2019-10-31	Daily
Dataset#2 [12]	From 2013-04-28 to 2020-02-14	Daily
Dataset#3 [13]	From 2018-02-13 09:00:00 to 2018-04-14 17:00:00	Hourly

Table 3. The Details of Dataset #1

Column	Description
Timestamp	Date
Sentiment	Average sentiment of tweets
Sentiment_volumes	Total quantity of tweets related to BTC
BTC volumes	Total quantity of shares traded
BTC weighted price	Average price of the BTC

Table 4. The Details of Dataset #2

Column	Description
day	Date
open_value	Open price
high_value	Highest price
low_value	Lowest price
close_value	Close price
volume_value	Trading volume
marketcap_value	Market capital
compound	Compound score
pos_score	Positive score
neg_score	Negative score
tweets_number	Number of tweets

Table 5. The Details of Dataset #3

Column	Description
timestamp	Timestamp
neg	Average of neutral sentiments
neu	Average of negative sentiments
norm	Sum of the valence scores of each word
pos	Average of positive sentiments
pol	Geometric mean of pos and neg
close	Close price
high	High price
low	Low price
open	Open price
volumefrom, volumeto	The trading volumes
mid	Middle price

4. Models

Three deep learning models are explored in this study, namely, LSTM, GRU, TCN. These models have been proven effective in other prediction problems [16-18].

Long short-term memory (LSTM) [19] is a particular type of RNN model, which can learn long-term dependent information. As shown in Figure 1, the data sent into the LSTM cell are the hidden state data from the previous time step and the input data of the current time step. Three fully connected layers then process them with activation sigmoid function to calculate the value of the forget gate, input gate, and the output gate, so the result of these three gates are in range (0,1). The candidate memory cell is similar to the other three gates, but using the tanh activation function, makes its range (-1, 1).

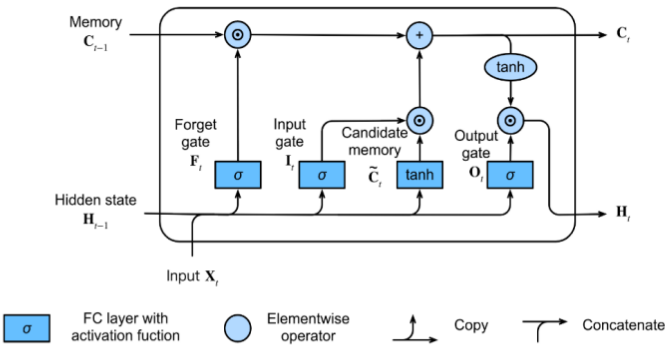

Figure 1. The LSTM cell

The computation process of the states can be formulated as follows (1)-(4):

$$I_t = \sigma(X_t W_{xi} + H_{t-1} W_{hi} + b_i) \tag{1}$$

$$F_t = \sigma(X_t W_{xf} + H_{t-1} W_{hf} + b_f) \tag{2}$$

$$O_t = \sigma(X_t W_{xo} + H_{t-1} W_{ho} + b_o) \tag{3}$$

$$C_t = F_t \odot C_{t-1} + I_t \odot C_t \tag{4}$$

where $W_{xi}, W_{xf}, W_{xo} \in \mathbb{R}^{d\times h}$ and $W_{hi}, W_{hf}, W_{ho} \in \mathbb{R}^{h\times h}$ are weight parameters $b_i, b_f, b_o \in \mathbb{R}^{1\times h}$ are bias parameters.

The follow-up computation process for updating the memory candidate is as follows (5):

$$C_t = F_t \odot C_{t-1} + I_t \odot \tilde{C}_t \tag{5}$$

Furthermore, the hidden state is updated as follows (6):

$$H_t = O_t \odot \tanh C_t \tag{6}$$

Gated Recurrent Unit (GRU) [20] is simplified and improved RNN compared with LSTM. The three gates are reduced to two. Thus the computation of a GRU cell is more straightforward, as shown in Figure 2. The whole process is as follows (7) (8):

$$R_t = \sigma(X_t W_{xr} + H_{t-1} W_{hr} + b_r) \tag{7}$$

$$Z_t = \sigma(X_t W_{xz} + H_{t-1} W_{hz} + b_z) \tag{8}$$

where $W_{xr}, W_{xz} \in \mathbb{R}^{d\times h}$ and $W_{hr}, W_{hz} \in \mathbb{R}^{h\times h}$ are weight parameters, $b_r, b_z \in \mathbb{R}^{1\times h}$ are bias parameters.

$$\widetilde{H}_t = \tanh(X_t W_{xh} + (R_t \odot H_{t-1}) W_{hh} + b_h) \tag{9}$$

where $W_{xh} \in \mathbb{R}^{d\times h}$ and $W_{hh} \in \mathbb{R}^{h\times h}$ are weight parameters, $b_h \in \mathbb{R}^{1\times h}$ is the bias parameter.

$$H_t = Z_t \odot H_{t-1} + (1 - Z_t) \odot \widetilde{H}_t \tag{10}$$

Other than the RNN family, another type of CNN model is also used in this study, namely, the Temporal Convolutional Network (TCN) [21], as shown in Figure 3. Compared with the traditional convolutional operations, TCN is featured by the causal convolution and dilation convolution. The causal convolution is conducted with only the input from previous steps, to keep the causality relationship held in the time series. Moreover, the dilation convolution is proposed for capturing the long-term dependency.

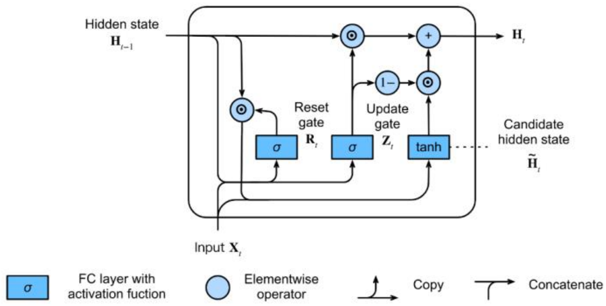

Figure 2. The GRU cell

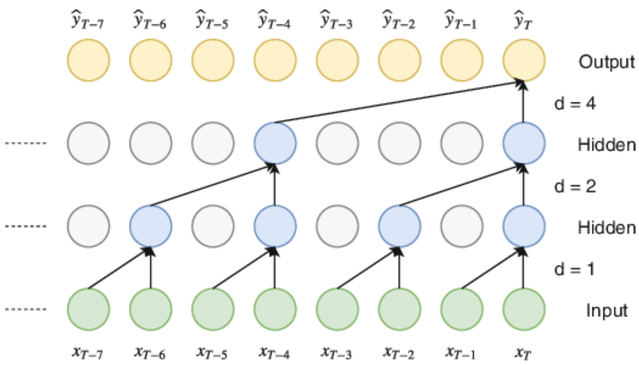

Figure 3. The TCN Structure

5. Experiments and results

The experiments of this study are conducted with Python and its packages, including scikit-learn and PyTorch. The datasets are divided into training, and test sets with a ratio of 80%:20%, in which the training set is used for the model training and the test set is used for performance evaluation and comparison. Four evaluations are used in this study, namely, accuracy, F1 score, precision, and recall. Since we are dealing with the price movement prediction, it can be defined as a binary classification problem with two possible outputs, e.g., 1 for up and 0 for down. As shown in figure 4. Then we can define the confusion matrix between the actual and predicted values, as shown in Figure 4. Accuracy is further defined as (11); Precision is further defined as

(12); Recall is further defined as (13), and F1 score is the geometric average of precision and recall.

$$(TP+TN)/(TP+FN+FP+TN) \tag{11}$$

$$TP/(TP+FP) \tag{12}$$

$$TP/(TP+FN) \tag{13}$$

We further divide a validation set for hyper-parameter tuning of each model. In the experiment process, we conduct hyperparameter tuning, choose different layers and neutron numbers of the models, try the bidirectional version of the models and use early stopping to prevent overfitting. Two groups of input features are compared, namely, with and without the emotion-related features. The evaluation results are summarized in Table 6.

		Predication Outcome	
		1	0
Actual Value	1	True Positive (TP)	False Negative (FN)
	0	False Positive (FP)	True Negative (TN)

Figure 4. The confusion matrix for a binary classification

Table 6. A Summary of Results in Different Datasets

Dataset	Model	Accuracy	F1 score	Precision	Recall
#1	LSTM (with emotion)	0.5245	**0.6146**	0.5362	**0.7616**
	LSTM (w/o emotion)	0.5330	0.6126	0.5442	0.7253
	GRU (with emotion)	0.5191	0.5470	0.5215	0.6444
	GRU (w/o emotion)	0.5074	0.5108	0.5309	0.5030
	TCN (with emotion)	0.5202	0.5989	0.5458	0.7152
	TCN (w/o emotion)	**0.5351**	0.5974	**0.5517**	0.7394
#2	LSTM (with emotion)	0.4945	0.6014	0.5075	0.7510
	LSTM (w/o emotion)	0.5089	0.6365	0.5157	0.8428
	GRU (with emotion)	0.4895	0.4094	**0.5293**	0.4825
	GRU (w/o emotion)	**0.5137**	**0.6720**	0.5164	**0.9626**
	TCN (with emotion)	0.5000	0.4312	0.5123	0.4942
	TCN (w/o emotion)	0.4948	0.4916	0.5169	0.5899
#3	LSTM (with emotion)	0.5401	0.5340	0.5420	0.5540
	LSTM (w/o emotion)	0.5083	0.5209	0.5070	**0.6734**
	GRU (with emotion)	0.5610	0.5136	**0.5780**	0.4719
	GRU (w/o emotion)	0.5466	0.4697	0.5681	0.4014
	TCN (with emotion)	**0.5646**	**0.5393**	0.5769	0.5122
	TCN (w/o emotion)	0.5329	0.4654	0.5644	0.4691

We first point out the best model for each dataset. We find that there is no single winner for all three datasets. For dataset #1, the best model is LSTM (with emotion), and the F1 score is 0.6146. For dataset #2, the best model is GRU (without emotion), and the F1 score is 0.6720. For dataset #3, the best model is TCN (with emotion), and the F1 score is 0.5393. Relatively, we would agree that dataset #3 is more difficult to predict when the hourly data are used instead of the daily data as used in datasets #1 and #2. In practice, hourly trading may be more profitable if the transaction costs can be ignorable.

Then we compare the cases with and without emotion features. For dataset #1, adding emotion as input features positively affects all three models, LSTM, GRU, and TCN. For dataset #2, adding emotion as input features has a negative effect on all three models. We find that the models that are using BTC data only, however, have better F1 scores. Moreover, for dataset #3, the result is identical to dataset #1 and adding emotion as input features improve the F1 scores.

6. Conclusion

In this study, while our main objective is to evaluate the influence of adding emotional factors for Bitcoin price movement prediction, we conclude that this problem is still challenging for deciding when to use emotional factors as external inputs. Two possible directions are worth trying. The first direction is to collect more data and analyze the emotion extraction techniques in more detail. The second direction is to attempt more models, especially those with the more robust ability to learn the patterns from high-dimensional input data.

References

[1] K. Wołk, “Advanced social media sentiment analysis for short-term cryptocurrency price prediction,” Expert Systems, 2020, 37(2): e12493.

[2] I. E. Livieris, N. Kiriakidou, S. Stavroyiannis, and P. Pintelas, “An advanced CNN-LSTM model for cryptocurrency forecasting,” Electronics, 2021, 10(3): 287.

[3] L. Li, A. Arab, J. Liu, J. Liu, and Z. Han, “Bitcoin options pricing using LSTM-based prediction model and blockchain statistics.” 2019 IEEE international conference on Blockchain (Blockchain), IEEE, 2019: 67-74.

[4] Y. Wang and R. Chen, “Cryptocurrency price prediction based on multiple market sentiment,” Proceedings of the 53rd Hawaii International Conference on System Sciences, 2020.

[5] D. Guégan and T. Renault, “Does investor sentiment on social media provide robust information for Bitcoin returns predictability?”, Finance Research Letters, 2021, 38: 101494.

[6] Y. Li, Z. Zheng, and H. N. Dai, “Enhancing bitcoin price fluctuation prediction using attentive LSTM and embedding network,” Applied Sciences, 2020, 10(14): 4872.

[7] A. Jain, S. Tripathi, H. D. Dwivedi, and P. Saxena, “Forecasting price of cryptocurrencies using tweets sentiment analysis,” 2018 eleventh international conference on contemporary computing (IC3). IEEE, 2018: 1-7.

[8] Q. Guo, S. Lei, Q. Ye, Z. Fang, “MRC-LSTM: A Hybrid Approach of Multi-scale Residual CNN and LSTM to Predict Bitcoin Price,” arXiv preprint arXiv:2105.00707, 2021.

[9] E. Bouri, R. Gupta, “Predicting Bitcoin returns: Comparing the roles of newspaper-and internet search-based measures of uncertainty,” Finance Research Letters, 2019: 101398.

[10] F. Valencia, A. Gómez-Espinosa, and B. Valdés-Aguirre, "Price movement prediction of cryptocurrencies using sentiment analysis and machine learning," Entropy, 2019, 21(6): 589.
[11] W. Jiang, "Applications of deep learning in stock market prediction: recent progress," Expert Systems with Applications, 2021: 115537.
[12] V. M. Hao, N. H. Huy, B. Dao, T. Mai, and K. Nguyen-An, "Predicting Cryptocurrency Price Movements Based on Social Media," 2019 International Conference on Advanced Computing and Applications (ACOMP). IEEE, 2019: 57-64.
[13] G. Serafini, P. Yi, Q. Zhang, M. Brambilla, J. Wang, Y. Hi, et al. "Sentiment-Driven Price Prediction of the Bitcoin based on Statistical and Deep Learning Approaches," 2020 International Joint Conference on Neural Networks (IJCNN). IEEE, 2020: 1-8.
[14] S. Cavalli and M. Amoretti, "CNN-based multivariate data analysis for bitcoin trend prediction," Applied Soft Computing, 2021, 101: 107065.
[15] F. Valencia, A. Gómez-Espinosa, and B. Valdés-Aguirre, "Price movement prediction of cryptocurrencies using sentiment analysis and machine learning," Entropy, 2019, 21(6): 589.
[16] W. Jiang and L. Zhang, "Geospatial data to images: A deep-learning framework for traffic forecasting," Tsinghua Science and Technology, 2018, 24(1): 52-64.
[17] W. Jiang and J. Luo, "Graph neural network for traffic forecasting: A survey," arXiv preprint arXiv:2101.11174, 2021.
[18] W. Jiang, "Internet Traffic Prediction with Deep Neural Networks," Internet Technology Letters, 2021.
[19] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural computation, 1997, 9(8): 1735-1780.
[20] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, et al. "Learning phrase representations using RNN encoder-decoder for statistical machine translation," arXiv preprint arXiv:1406.1078, 2014.
[21] C. Lea, M. D. Flynn, R. Vidal, A. Reiter, and G. D. Hager, "Temporal convolutional networks for action segmentation and detection," proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 156-165.