Bitcoin Price Prediction Using LSTM with Google Trend

Jingwei Dai
Department of Marketing
The Chinese University of Hong Kong
jingweidai@link.cuhk.edu.hk

Abstract

This study intends to predict Bitcoin price using long short-term memory (LSTM), combined with Google Trend for keywords related to Bitcoin and commonly used financial indicators. It also compares the performance of the proposed model with other models, including LSTM with Google Trend but without financial indicators, LSTM without Google Trend and financial indicators, as well as traditional financial forecasting methods like the ARIMA model. The results show that combining Google Trend would increase the predicting accuracy significantly. However, integrating the financial indicators into the model does not influence the model performance much, and may even impose a negative effect on predicting accuracy.

1. Introduction

Cryptocurrency is a virtual currency that is secured by cryptography [4]. Among all kinds of cryptocurrencies, Bitcoin is the most famous one, which has a large trading volume each day. Currently, it is also of the highest value among all kinds of them, which increased around 370% in the year 2020. The profit of investment in Bitcoin in the past years due to the surging value attracts not only individual investors but also institutions like investment banks on Wall Street [8]. Investors have tried hard to predict the trend of Bitcoin so as to adopt corresponding appropriate actions. The practical use and promising applications in the real commercial world drive the research towards the prediction of Bitcoin price, which has been a hot topic in the business research field in the past years.

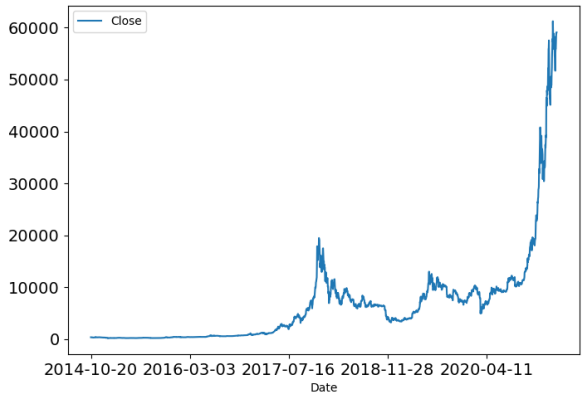

Figure 1. Bitcoin Close Price

However, it is not easy to predict the price of financial products due to the volatile property of them. They are rather sensitive to real commercial world situations, which are usually unpredictable. Traditional financial forecaseting methods, like the ARIMA model, are usually purely based on the previous price trend to linearly predict the future trend. Thus, the estimation results generally are not satisfying enough. Therefore, machine learning and deep learning techniques have gradually been introduced into business research in the past several years.

The main goal of the project is to predict Bitcoin price with LSTM-network combined with Google Trend. The reason for picking Google trend is that it summarizes how many people are searching Bitcoin and keywords related to it, such as "Buy Bitcoin", "Sell Bitcoin", "Bitcoin Low", indicating how much attention is being paid to Bitcoin. I consider the Google trend for neutral words like "Bitcoin", "Blockchain", positive words like "Buy Bitcoin", "Bitcoin High", as well as negative words like "Sell Bitcoin", and "Bitcoin Low".

This is one of the major innovations of the project, as although there has been research combining Google trend in predicting stock price [5], little research has paid attention to the property of words, and mainly use neutral words, like

"Bitcoin", "BTC", "Blockchain". These words, although can show the attention that the market pays on Bitcoin, cannot reflect investors' thoughts and mood, which can predict their action more accurately. For example, if many people are searching "Sell Bitcoin", it shows that many people are considering selling it, which could lead to the selloff of Bitcoin, and its price would slump. Using more accurate words can increase the accuracy rate of prediction.

I run a simple regression to test the effect of the commonly used financial indicators and searching keywords related to Bitcoin on the Bitcoin close price, and get the table below. We can see that the effect of the Google trend is overall significant, and the signs of most words are in line with our expectations. However, the sign for blockchain is significantly negative. One possible reason is that blockchain is the technology database of all cryptocurrencies, which contains many substitutes to bitcoin. When people search "blockchain" much, it reflects that people may want to seek alternatives. This predicts a price fall. Since the searching volume for "Bitcoin low" is 0 in many days, I use "Sell Bitcoin" as the keyword in the LSTM model. However, I have also tried "Bitcoin Low" into the model, the performance is very similar. So using each word does not change the results greatly. Similarly, using reverse-ordered words like "Bitcoin Buy", instead of "Buy Bitcoin" also does not change the model performance.

	term	estimate	p.value
1	(Intercept)	-391.12	0
2	volume	0	0.09
3	fin_sma	-1.76	0
4	fin_ema	2.7	0
5	fin_bbdn	0.07	0
6	fin_rsi	6.53	0
7	gt_bitcoin	6.33	0.07
8	gt_blockchain	-8.66	0
9	gt_bitcoinhigh	21.62	0
10	gt_bitcoinbuy	17.76	0
11	gt_bitcoinsell	-1.32	0.65
12	gt_bitcoinlow	-15.88	0

Table 1. Model-free Evidence of Google Trend on Bitcoin Close Price

2. Related Work

LSTM and RNN have been introduced to predict stock price in academic research. Ding and Qin proposed a multi-input LSTM model that can predict the open price, high price, low price of the stock, but does not consider the close price, which is in fact the most important price measurement that investors focus on [1]. Li and Dai found that introducing outside factors can increase Bitcoin price prediction model accuracy significantly, which supports our hypothesis of introducing Bitcoin to reflect the overall market situation [7]. But little research has integrated Google Trend and used different kinds of words for prediction. This research will fill in the gap.

3. Method and Model

3.1. Baseline Model

In the statistics and finance domain, the most frequently used method to predict time series data is linear autoregressive integrated moving average (ARIMA) [2]. The model has three parts. The AR part of ARIMA indicates that the evolving variable is regressed on its own lagged values. The MA part indicates that the regression error is a linear combination of error terms at the current equation and previous ones. The I indicates that the data values have been replaced with the difference between their values and the prior values. The process of differencing is to make the data stationary, which is the requirement of the AR and MA parts. The equation of ARIMA model is below.

$$\left(1-\sum_{i=1}^{p}\phi_i L^i\right)(1-L)^d X_t = \left(1+\sum_{i=1}^{q}\theta_i L^i\right)\varepsilon_t \quad (1)$$

Where d is the order of differentiation, p is for AR part, and q is for MA part.

However, if the data is rather non-stationary, it will need to be differenced multiple times, making the model very complicated and hard to predict the future value. Another limitation of ARIMA is that it can usually predict the future value in a short period. If the dimension of the future period increases, the predicted value would quickly converge to a constant, which is not satisfying. The third limitation is that its estimation is purely based on previous data pattern, and does not take outside factors into consideration. This may work for regular time series data but Bitcoin price is very vulnerable to outside factors, so usually it is rather non-stationary. If not integrating into other factors, the performance of the model would not be well.

After observing differencing result, ACF test and PACF test results, I use p = 3, d = 1, q = 3 for the ARIMA model.

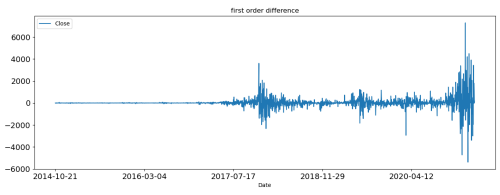

Figure 2. First Order Difference

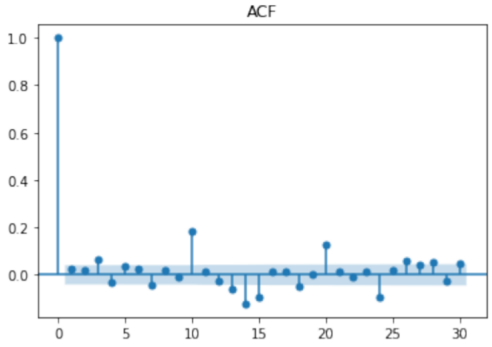

Figure 3. ACF Test

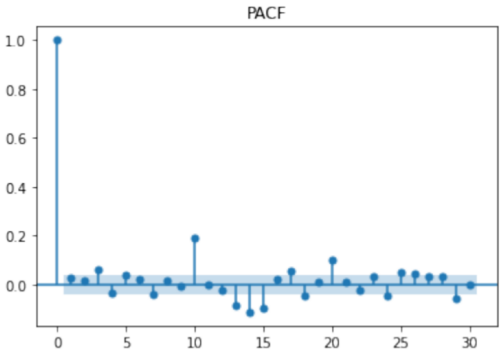

Figure 4. PACF Test

3.2. LSTM Networks

The long short-term memory network or LSTM is designed to address the problem of vanishing gradients in the recurrent neural network (RNN). LSTM would enable the network to learn more about previous time steps, with the error more steady. This helps the recurrent neural network to capture and learn long-term trends. LSTM is also suitable to deal with time-series data. The key to LSTM is cell state, which have three gates: the input gate, forget gate, and output gate. The first one decides which information should be disregarded. The second one decides which new information would be stored. The last decides which information to output. Gates are composed of a sigmoid neural net layer and a pointwise multiplication operation [6].

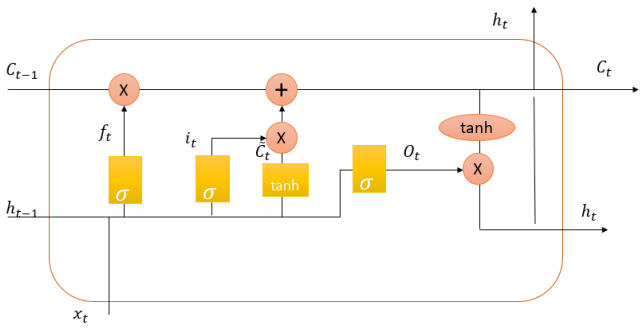

Figure 5. LSTM Structure

4. Experiments

4.1. Experiment Design

The LSTM model will mainly have three inputs: the past price data, trading volume, Google Trend for different keywords, and financial indicators. I collected daily Bitcoin price and trading volume data from Yahoo Finance between October 20, 2014, to April 1, 2021. So, in total, I have 2356 observations. I use the first 60% of the data as the training set, the following 20% of the data as the validation set, and the last 20% of the data as the testing set. Google trend data for frequently searched keywords related to Bitcoin comes from the Pytrend library of Python. The words that I use include neutral words - "Bitcoin", "Blockchain", positive words – "Bitcoin Rise", "Buy Bitcoin", and negative words - "Sell Bitcoin". I only use one most frequently used financial indicator - Relative Strength Index (RSI), which can measure the magnitude of recent price changes in the price of the stock or other assets [3]. Investors rely heavily on this indicator to decide whether to buy or sell their Bitcoin and other financial products. I only use this one because most financial indicators just reflect the past price trend, which in fact has been captured by LSTM. Also, too many input variables would lead LSTM to overfit. For other parameters when running LSTM model, I use epochs = 30, batch size =25, drop out rate = 0.01 to train the model. Below are the loss value rate changes.

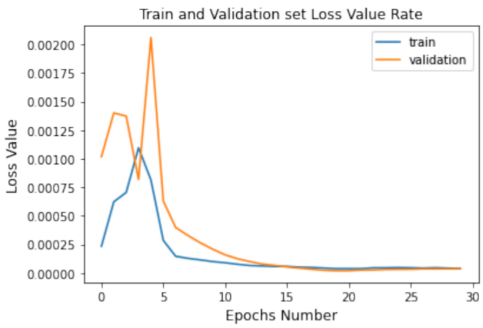

Figure 6. Train and Validation Set Loss Value Rate

4.2. Experiment Results and Evaluation

The result of the proposed model's predicted value and real value is below. We can see that overall the prediction is very accurate. This shows the current model, integrated with Google trend and financial indicators, would have generally good performance in predicting future values of Bitcoin.

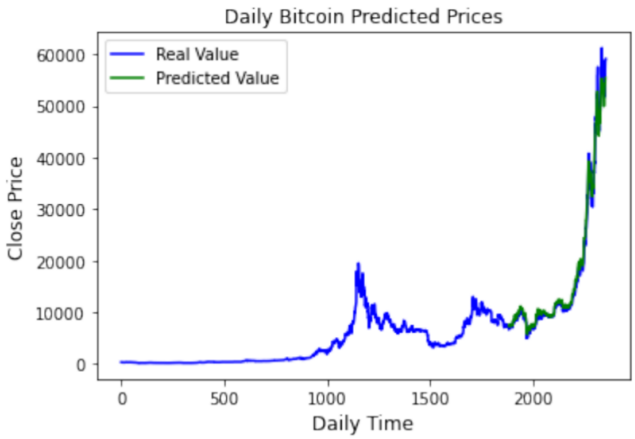

Figure 7. Daily Bitcoin Predicted Prices

To see how this model outperforms traditional models and how it is compared with variant models, I compare the RMSE of LSTM with Google trend and financial indicators, LSTM with Google trend only, LSTM without Google trend and financial indicators, and ARIMA model. The results are below.

Model	RMSE
LSTM with Google trend only	1140.70
LSTM with Google trend and financial indicators	1294.68
LSTM without Google trend and financial indicators	3094.80
ARIMA	18403.34

Table 2. Root Mean Square Error (RMSE) Comparison

From the table, we can see that LSTM with Google trend and without financial indicators performs marginally better than with financial indicators, but the difference is very minor. This demonstrates that Google trend alone can predict the trend of Bitcoin's future price better. It is also in line with our hypothesis that LSTM has already captured the hidden trend information that financial indicators intend to provide, so financial indicators are not needed to be input into the model. However, we can also see that without combining Google trend data, the performance of the LSTM network would be much worse. This is because Bitcoin price is very volatile to the outside factors, and without having known the market situation via Google trend, LSTM itself cannot predict the future trend well. That being said, all of these deep learning techniques still perform much better than the ARIMA model. The main reason is that, based solely on the past price data, ARIMA cannot predict long-term trends well. It largely can only capture the short-term linear trend as it does not have market information and the model is too restricted by its fixed form.

References

[1] G. Ding and L. Qin. Study on the prediction of stock price based on the associated network model of lstm. *International Journal of Machine Learning and Cybernetics*, 11, 2020.

[2] J. Fan and Q. Yao. Arma modeling and forecasting. *Nonlinear Time Series: Nonparametric and Parametric Methods*, 2013.

[3] J. Fernando. Relative strength index (rsi). Accessed: 17 May 2021. https://www.investopedia.com/terms/r/rsi.asp, 2021.

[4] J. Frankenfield. Cryptocurrency. Accessed: 17 May 2021. https://www.investopedia.com/terms/c/cryptocurrency.asp, 2021.

[5] M. Y. Huang, R. R. Rojas, and P. D. Convery. Forecasting stock market movements using google trend searches. *Empirical Economics*, 59, 2020.

[6] H. Li. Introduction to deep learning. Accessed: 17 May 2021. http://dl.ee.cuhk.edu.hk/, 2021.

[7] Y. Li and W. Dai. Bitcoin price forecasting method based on cnn-lstm hybrid neural network model. *The Journal of Engineering*, 2020.

[8] R. Ungarino. The crypto talent war is heating up as big money managers warm to digital assets. Accessed: 17 May 2021. https://www.businessinsider.com/asset-managers-wall-street-cryptocurrency-digital-assets-hiring-trends-recuiting-2021-5, 2021.