Predicting Crypto Price Trends

Qiuyang Wang, Xinyu Jiang, Yan Zhan

University of Colorado, Boulder

Abstract

Long short-term memory (LSTM) networks are a important method in deep learning for time series forecasting. In this project, we perform a forecasting framework using LSTM with AR(2) to forecasting hourly bitcoin price. Our model predicts the tendency of bitcoin price well without large external factors.

Introduction

Ever since its launching in 2008, Bitcoin has become increasing popular among investors. As the most famous cryptocurrency, Bitcoin as also drawn crucial attention due to its volatile price change. The factors affecting cryptocurrency prices can be divided into internal factors and external factors. The former are driven by the demand and supply of cryptocurrency and the later contains the attractiveness of the cryptomarket, the macro-financial situation and political influences.

Researchers from different fields have been trying to predict the price of Bitcoin with various methods. In this project, we follow We et al(2018) and try to predict the price of Bitcoin in a short-term future with LSTM. As a time series approach, LSTM outperform others such as Autoregressive (AR), univariate Moving Average (MA) and Autoregressive Integrated Moving Average (ARIMA) in terms of the Bitcoin price prediction. Those methods are more suitable for data with seasonal trends pattern which is not observed in Bitcoin price. LSTM is favored over other time series method due to the temporal nature of the more advanced algorithms.

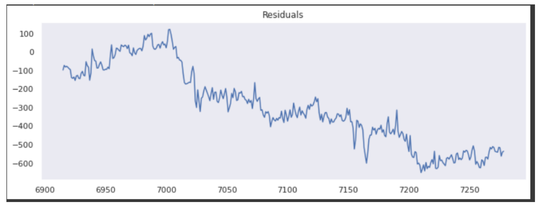
Approach

After discussing, we decided to use LSTM as our model and also to change the scope of the project, from predicting trend to predicting crypto price in the next hour. To reach the goal, we also need to reshape and edit the raw data to make sure we can feed the correct data to our model.In terms of predicting the crypto price in the next hour, we found a way to create the label which acts like the crypto price in the next hour. Then we fit the LSTM model with the pre-processed data and feed test data to the model as well. The model generated the predicted crypto prices in the next hour just after the last close time of the last row of the data. And we validated the predicted labels with our test y, and generated MSE, loss, residual plots to evaluate the performance of our model.

Results

Residuals

The graph below represents the residuals of difference between the predicted value and the actual Bitcoin closing price. Based on the graph we can conclude that the RNN LSTM network cannot predict a sudden price drop that is caused by external factors(like covid, or government policies). Prior to this, it can successfully handle forecasts that deviate from actual prices, ranging from 100 to -250 US dollars. In addition to that, the result shows that our false predictions generated by our model are more likely to be slightly below the actual bitcoin price.



Data

We used the data from Binance Data Collection which is a website provides free crypto data sets in daily and monthly with complete time resolutions(minutes, hours, days, months). The we used klines data sets in 1-hour resolution, the time range of the data sets are from December 2020 to November 2021. For the original data sets, each data set contains fixed numbers of features, but we only used close feature(the price of a crypto when the market is closed) as our main feature. Only one column of feature is definitely not enough to train our LSTM model. Therefore, we used 4 crypto close prices in total and we did data preprocessing on these 4 crypto close prices. The data from the website are zipped in month scale, we concatenated the data from December 2021 to October 2020 as four dataframes for each crypto.

Results

Prediction Result In order to test the prediction of bitcoin price and actual value, we use the mean square error regression loss (MSE) to measure. Based on the graph below, the test MSE result we got is 0.06682 which represents the lowest error measurement. The pot shows without the influence of particularly large external factors, there is not much difference between the predicted price and the actual price. However, after the interference of external factors, there are some differences between our predicted results(the blue line) and actual results(the red line), however, our model can still successfully predict the tendency of bitcoin price(based on the graph, the red line shows the same tendency with the blue line).

Discussion

As you can see in the graphs and images we listed here, our model is able to predict the Bitcoin price in short term(3 hour) when there is no external factor or the external factor is not able to cause rapid increase and decrease of Bitcoin prices(and other crypto prices as well). However, when there are external factors that they are large enough to result the rapid validation of crypto prices, our model is not be able to predict the accurate prices(with in certain range). Instead, it will predict the similar trend of the Bitcoin price with large gap between real Bitcoin price and predicted Bitcoin price. And according to the residual graph we have generated, the observed price difference between real price and predicted price is ranged from 100 to -600, considering the Bitcoin price is nearly 50k usd, our model is "accurate" enough for predicting the trend in the next hour but not the exact price. However, we used a year long data set to predict the next hour Bitcoin price with probability of high error, with this efficiency, I don't think we can say that we have reached our previous goal which is our model has the ability to predict Bitcoin price in the next hour.

The reason why our model can't perform its job largely is due to unpredictable external factor and we can't really forecast how these external factors are going to affect the market. In terms of external factors, they could be political orders, like banning crypto as valid currencies, banning mining crypto, etc. These actions are complete break-down the investors' confidences and largely affect crypto prices negatively. However, there are many other external factors like Elon Musk's tweet, he can tweet whatever he wants about crypto and these tweets will result in market shock. Although the shock could be negative or positive, not like political actions which most of them are negative. And there is no build-in logic to analyze how these tweets/news will affect the market, considering currently, crypto are built upon nothing- there is no actual asset to guarantee the value of crypto.