MSBD 5013 Final Report

Team Greenwich members:

Zhao, Cai

Wang Zhicong

Ge Mengye

Lu, Kuojian

Abstract

With the historical minute-level OHLC data of four major crypto currencies, BTC, BCH, LTC, and ETH, we aimed to write a minute-level trading strategy which can output the desired position of next minute. According to the position, we can determine the trading in next minute. In this report, we proposal three trading strategies and use vote ensemble to combine these three strategies together as the final strategy.

1. Exploratory of Bollinger Bands

To make better decisions about the timing of longing and shorting, we searched several currency trading strategies online. After studying each of these strategies, we decided to use Bollinger Bands.

Bollinger Bands is widely used as a technical analysis indicator. We used a 14-day moving average as centerline and then traced two bands, each one standard deviation wide, on either side of the moving average [1]. Bollinger Bands actually has three lines, the upper line, the centerline, and the lower line. And there are two different ways of using Bollinger Bands. Longing when the price is higher than the upper line and shorting when it is lower than the lower line. The other one is just the opposite, shoring when the price is higher than the upper line and longing when it is lower than the lower line.

By visualizing the price tendency of each crypto currency, we found that price only moves outside the bands a few times. It is better to use the second method, shorting when the price is higher than the upper line and longing when it is lower than the lower line. The final price would most probably higher than the price we bought them.

2. Trading Strategies

2.1 Trading Strategy 1

Bayesian regression is used for predicting price variation of BTC (Bitcoin), BCH (Bitcoin cash), LTC (Litecoin) and ETH (Ethereum). This idea is from the paper published by MIT

Professor Shah in Oct 2014 [2]. Based on this price prediction method, they devise a simple strategy for trading Bitcoin. The strategy is able to nearly double the investment in less than 60 day period when run against real data trace.

However, we have done some optimizations basing on this method from the paper. We choose to use the sample entropy to choose the effective centres, while the in the paper, the choosing method is not published. The whole details of this method can be obtained from the original paper. The differences are as follows.

Firstly, they only apply this method on Bitcoin price prediction; we apply this method on the price prediction of BTC, BCH, LTC and ETH. Secondly, the interval of time-series is different; we set it 60 seconds, but in the paper the interval is set as 10 seconds. Thirdly, the data is different; we do not have the vbid (the total volume people are willing to buy in the top 60 orders based on the current order book data) data and vask (the total volume people are willing to sell in the top 60 orders based on the current order book data).

The main steps of modified method from the paper are as follows:

1. use the historic time series to generate three subsets of time-series data of three different lengths: S_1 of time-length 180 minutes, S_2 of time-length 360 minutes, and S_3 of time-length 720 minutes.

2. at a given point of time, to predict the future change Δp, use the historical data of three length: previous 180 minutes, 360 minutes and 720 minutes - denoted x^1, x^2 and x^3

3.use x^j with historical samples S^j for Bayesian regression to predict average price change Δp^j for $1 \leq j \leq 3$.

4.The final estimation Δp is produced as

$$\Delta p = w_0 + \sum_{j=1}^{3} w_j \Delta p^j$$

where $\mathbf{w} = (w_0, \ldots, w_3)$ are learnt parameters.

For finding Sj,$1 \leq j \leq 3$ and learning w, the steps are as follows:

1. Utilize the first time period to find patterns S_j, $1 \leq j \leq 3$(previously divide the entire time duration into two, roughly equal sized periods)

2. The second period is used to learn parameters w. The learning of w is done simply by finding the best linear fit over all choices given the selection of S_j , $1 \leq j \leq 3$. Now selection of S_j, $1 \leq j \leq 3$.

3.take all possible time series of appropriate length (effectively vectors of dimension 180, 360 and 720 respectively for S_1, S_2 and S_3)

Each of these form x_i and their corresponding label y_i is computed by looking at the average price change in the 60 second time interval following the end of time duration of x_i.

4. we clustered patterns in 100 clusters using k−means algorithm. From these, they chose 20 most effective clusters and took representative patterns from these clusters by their sample entropy.

5. The one missing detail is computing 'distance' between pattern x and x_i , this is squared l2-norm. use $\exp(c \cdot s(x, x_i))$ in place of $\exp(-\|x - x_i\|22 * 0.25)$ with choice of constant c optimized for better prediction using the fitting data (like for w).

So in this strategy, when predicting the next minute's price, we should know the former 720 minutes' prices which is a time series; if the price change is larger than a threshold, we buy one BTC, BCH, LTC or ETH; if the price change is smaller than a threshold, we sell one BTC, BCH, LTC or ETH; else we do nothing.

2.2 Trading Strategy 2 Based on LSTM

2.2.1 Motivation

For the minute-level crypto currencies prediction, the data is in time series, so the prediction method is required to interpret the temporal information during a continuous period. Traditional statistical learning and machine learning methods, such as linear regression and tree-based methods, fail to take the temporal information into account, thus they are not much suitable for the time series data. While LSTM, which is a recurrent neural network, is more suitable due to its ability of handling sequential data. Using LSTM, it is more likely to discover the temporal pattern of the changing trend for the crypto currencies from the time series data.

2.2.2 Feature engineering

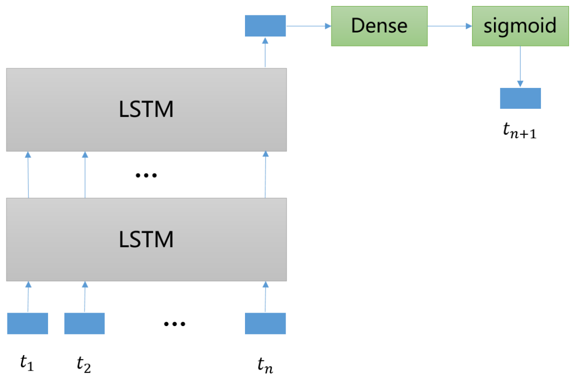
The crypto currencies data are in minute level, which is too frequent for us to handle. We use resampling method to generate lower frequency data from the high frequency data. Specifically, we combine every continuous ten minutes into one time period based on the following rules: open price of this time period is the open price at the first minute; close price is the close price at the last minute; high and low prices are the corresponding highest and lowest prices during the ten minutes; volume is the average result of volumes in this time period.

Since predicting the concrete values for the crypto currencies is very difficult because their changing trend is quite unstable, in other words, it can have a dramatic increase in one time period and then a sharp decrease in the next time period. So instead, we try to predict whether their prices will increase or decrease rather than the concrete values. Specifically, we want to predict the probability of price increasing in the next time period.

We use open, close, high, low and volume as features.

2.2.3 Architecture

The architecture of LSTM network is shown in Fig. 1. For the prediction of a specific currency at a specific time period, the inputs of the network are the features in time order during the last n time periods. Taking the sequential data as input, LSTM network with multiple layers tries to learn the pattern of the changing trend for this specific currency. We take the the output of this LSTM layer at the last time step and transform it by one fully connected layer. Finally, the probability of price increasing at the next time period is obtained by sigmoid function.



2.2.4 Trading strategy

Once we get the probability of pricing increasing at the next time period, we can use it for trading. Specifically, from this probability, we can get three signals for trading, which are increasing, holding on and decreasing, through two thresholds.

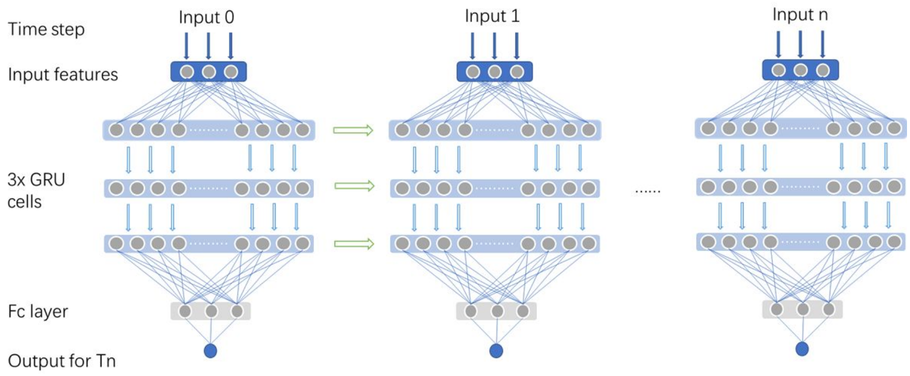
2.3 Trading Strategy 3 based on GRU

2.3.1 Advantage of GRU

Same as LSTM, GRU is also created to solve the vanish gradient problem in RNNs. The overall performance of them is quite similar but it seems that GRU will converge quicker than LSTM. The performance on different classes still need to be test separately.

2.3.2 Model Structure

Here is the rough structure of the GRU model.



2.3.3 Training Data Selection

The trend for different currency will change rapidly with the time. It is easy to find that the trend of last month will largely changed compared with this month. So, we only use the data of the most recent weeks as the training set. The time far from now may mislead the model to do the wrong prediction.

2.3.4 Feature and Label

The only useful data here are the price and the volume, so we choose these two values as input and the prediction of next time step as the output. The label for the training set is the changing percent from this time to the next time.

2.3.5 Prediction Procedure

Due to the existence of the trading tax, we could not simply use the increasing or decreasing as signals. Here, in order to take the tax into consideration, we use the 5% or more increment or decrement as two kinds of signal and the value in between will be viewed as the third one. Here we simply define them as positive,negative and neutral signal. We will may change the action on each signal for different strategy.

For directly buy or sell, if the change exceeds the 5% threshold, we will buy or sell certain fixed amount of bitcoins.

For the statistical prediction, the action of origin strategy may lag behind the real curve and may reduce the overall benefit. With the help of the model, we can have action one step earlier so that we may gain higher profit. If the predicted value of next time exceeds the statistical threshold, then we will perform the action.

2.4 Final Trading Strategy: Voting

We use vote ensemble to combine these three strategies together, which is our final strategy.

References

[1] Investing Answers, Bollinger Bands

https://investinganswers.com/financial-dictionary/technical-analysis/bollinger-bands-854

[2] Devavrat Shah, Kang Zhang Bayesian regression and Bitcoin https://arxiv.org/abs/1410.1231