**Random Forest combined with grid transaction**

1. Overall

Random forest is a kind of ensemble method. Its main idea derives from decision trees, which is a popular machine learning algorithm. Below is a sample of decision tree.

图片包含 文字



自动生成的说明

Among all features we have, decision tree method chooses one feature having highest entropy at every split of node. Following this, random forest randomly chooses a subset of features available (usually a square root of total number of features) when splitting a node. Repeat it n times, finally we obtain n different decision trees. At last we average the outputs of these trees as the final prediction result. So next the key part is preprocessing data and set suitable features.

1. Reasons to choose the model and principle of generating features
2. Model theory

In traditional trading theory, grid transaction is a typical kind of trading strategy. Its main principle is called “low long and high short”, which means when we notice price has decreased a lot, we buy; when price has increased a lot, we sell. The criterion is the price edge we set, which seems like a grid. An example is showing below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Trend | Price | Positions | [-3%buy, 5%sell] | [-5%buy, 10%sell] |
| Fall | Buy 3 | 90% | 9.1 | 8.5 |
| Buy 2 | 70% | 9.4 | 9.0 |
| Buy 1 | 40% | 9.7 | 9.5 |
| Plain | Base | No change | 10.0 | 10.0 |
| Rise | Sell 1 | 60% | 10.5 | 11.0 |
| Sell 2 | 30% | 11.0 | 12.0 |
| Sell 3 | 10% | 11.5 | 13.0 |

Source: https://www.joinquant.com/post/539?f=home&m=alg\_example

From the chart above, we can see that we will change the position when price touches certain edge.

In general, we can conclude that the strategy is according to the extent of price change in the past, we predict if it goes up or down next time stamp. Here the prediction rule is naïve, that is if price has risen a lot, we predict it will decrease. Then we long and vice versa.

Based on this, for a certain currency, we collect the extent of price and transaction volume change of itself and all other currencies in the past as features. Along with these features, we utilize Random Forest to predict if its price will go up or down next stamp. Here we choose one hour as a time stamp. If the prediction result is ‘up’, we long. If it is ‘down’, we short.

1. Feature engineering

Because all data of price is continuous, first we should discretize dataset.

For one time stamp (60 minutes), if price goes up over 1%, we set its category as ‘up’. If price goes down over 0.5%, we set it as ‘down’. Rest are set as ‘middle’. For transaction volume, if its absolute changing rate is larger than 201%, we set it as ‘sharp’. Rest are set as ‘flat’. For one prediction, features are price and volume change described above in past two time stamp (120 minutes).

For instance, if we want to predict BTC, features needed are listed below:

|  |  |
| --- | --- |
| BCH\_t0: price change of BCH 2 hours ago. | BTC\_t0: price change of BTC 2 hours ago. |
| ETH\_t0: price change of ETH 2 hours ago. | LTC\_t0: price change of LTC 2 hours ago. |
| BCH\_t1: price change of BCH 1 hour ago | BTC\_t1: price change of BTC 1 hour ago. |
| ETH\_t1: price change of ETH 1 hour ago. | LTC\_t1: price change of LTC 1 hour ago. |
| dV\_BCH\_t0: volume change 2 hours ago | dV\_BTC\_t0: volume change 2 hours ago |
| dV\_ETH\_t0: volume change 2 hours ago | dV\_LTC\_t0: volume change 2 hours ago |
| dV\_BCH\_t1: volume change 1 hour ago | dV\_BTC\_t1: volume change 1 hour ago |
| dV\_ETH\_t1: volume change 1 hour ago | dV\_LTC\_t1: volume change 1 hour ago |

Among them, for BCH\_t0, it has categories: ‘up’, ‘middle’, ‘down’. For dV\_BCH\_t0, it has categories: ‘flat’, ‘sharp’. Rest are so on so forth. Therefore, we can use all features above to train Random forest model.

1. Implementation
2. Encountered problems and corresponding solutions

The main problem is setting multiple thresholds including the length of time stamp, bound to discretize price and volume change. As for time stamp, because the target of the model is to predict trend not exactly price, it is not suitable to set a too short length of time. Here we choose 30 and 60 minutes as two candidates. Finally, we choose 60 minutes as it has higher average return across several weeks. Intuitively, it can stabilize exchange procedure in a high fluctuant market.

Another threshold is the bound for discretizing price change. Initially, we hope to lower risk. So we set large threshold. For example, we set -2% as the criterion for ‘down’ and 4% for ‘up’. However, the result is during a week, the program doesn’t even make a position change, which is meaningless in actual trading. After adjusting many times, we choose -0.5% and 1% as corresponding bound.

1. Performance

We train our model using data upon 2 months (201807, 201808). Here we use cross-validation to check the performance of the model. The accuracy of prediction is listed below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Currency | BCH | BTC | ETH | LTC |
| Accuracy | 0.66 | 0.85 | 0.72 | 0.72 |

From chart above, we can see our prediction can get average accuracy over 73%. In view of high fluctuation of market, such an accuracy can meet requirement.

Besides general test about accuracy, we also combine with complementary strategy to test the model in actual trading. For reducing risk, we add a condition that if cash balance is lower than 20000 USD, we close all positions to make sure we will not be forbidden. We backtest the model in several weeks. Performance are listed below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Time | 0916-0923 | 0923-0930 | 0930-1007 | 1007-1014 |
| Weekly return | -0.028 | 0.004 | 0.001 | 0.003 |

From the chart we can see that except first week we tested, the whole strategy has a small amount but stable earnings. Because of high fluctuation of virtual currency market, if we can be stable and make money sometimes, we can survive. Preliminarily, the model satisfies the requirement.

1. Potential optimization

Based on model above, we can see that several thresholds count a lot. Next we think that the model can be improved by adjusting these thresholds dynamically according to fluctuation analysis during recent one or two weeks. For instance, if recently the market is relatively plain, we can set the bound of discretizing price change smaller. On the contrary, we can set the bound larger to lower risk.