Quantifying the Wow Surprise in Cryptocurrency Exchange

Linqing Liu
Tongji University
4800 Caoan Road
Shanghai, China
likicode@gmail.com

Yao Lu
Tongji University
4800 Caoan Road
Shanghai, China
luyao@ieee.org

1. INTRODUCTION

1.1 What is Cryptocurrency

A cryptocurrency is a medium of exchange like normal currencies such as USD, but designed for the purpose of exchanging digital information through a process made possible by certain principles of cryptography. Cryptography is used to secure the transactions and to control the creation of new coins. The first cryptocurrency to be created was Bitcoin back in 2009. Litecoin was announced in 2011 with the goal of being the "silver" to Bitcoin's "gold". At the time of writing, Litecoin has the highest market cap of any mined cryptocurrency, after Bitcoin.

1.2 Cryptocurrency Trading

1.2.1 Trading Platform

People compete to "mine" cryptocurrency using computers to solve complex math puzzles. But with the limited total amount of cryptocurrency, mining became increasingly difficult to gain profit. More people buy and sell cryptocurrency online just like the way of stock deal. Several marketplaces called "cryptocurrency exchanges" allow people to buy or sell them freely. Huobi[1] and OKcoin[2] are one of the most largest cryptocurrency exchanging platforms.You can see more details on this technical review[3].

1.2.2 Details of Trading

- **The order book**
  Often you'll see order books displayed as tables showing open buy orders and sell orders at price levels below and above the last market price (Figure 1). The next trade in the example above will execute either at the buy of CNY 2745.80 in case of a seller stepping in, or at CNY 2745.95 if a buyer is willing to pay the sell, more on this later. It's important to note that transactions between sellers and buyers only occur if there are open bid and ask orders at the same price level.

- **Limit orders vs market orders**
  Orders in the book are all 'limit orders', meaning they will only be executed if anyone is willing to trade at the requested price level. On the other hand, 'market orders' are orders without a price tag attached. For this reason they don't show up in order books. Instead, they are immediately executed at the next best limit order in the book.

Figure 1: Order Book of Huobi Platform

1.3 What is our target?

We intend to make an extreme utilization of Litecoin historical data, try to predict future price variation and help develop profitable quantitative strategy.

2. DATASET

We use data related to price and order book obtained from Huobi and OKcoin, the top tier digital currency exchangers in the world.

- **Collection Range** The Huobi dataset concerns time period between August 2015 to December 2015, the OKcoin dataset ranges from August 2015 to March 2016.
- **Market Conditions** Our dataset comprises all kinds of market conditions. During the data collection period, it has not only been a sharp decline in price, but also doubled or tripled investment. It's convincing that what we collected is a true portrayal of the real market.

[1]Huobi Website https://www.huobi.com
[2]OKcoin Website https://www.okcoin.com
[3]http://www.coindesk.com/makes-bitcoin-exchanges-tick/

ACM ISBN 978-1-4503-2138-9.
DOI: 10.1145/1235

- **Market Depth** The order book depth of Huobi is 10 and that of OKcoin is 50. Depth means the amount of best price one is willing to buy or sell at the given point of time.
- **Size of Dataset** All data is acquired per second in consecutive days. The total raw Huobi data is more than 6.5 million. To the best of our knowledge, this is the most complete dataset of such minimal granularity and covering all kinds of market conditions.

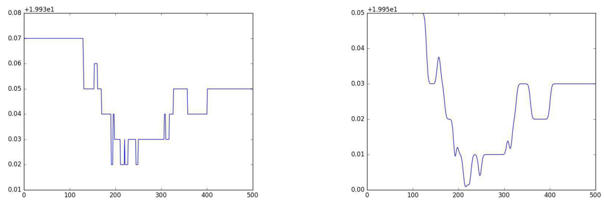
3. METHODOLOGY

3.1 The Importance of Orderbook

A stable market will automatically balance the sell orders and buy orders of order book by fluctuating current price. Let's assume for a moment the imbalance between the two sides of order book trigger the change of price. In a manner of speaking, the order book disparity is the accelerates of market changing. But this kind of effect can only last for 15 seconds to 30 seconds. So it's reasonable to depend on the amount of orders in specific time period to predict future price change.

3.2 Data Processing

To remove tangential noises, we first apply Gaussian filter algorithm on the price values(show in Figure 2).Then we label each data point as UP, DOWN, or STABLE according to the rise and fall trend of the following price in 15 seconds. Here, UP indicates the following prices are all higher than the current one or a subtle fall may occur but others rise sharply more than 0.02.



(a) Price

(b) Price with Gaussian Filter

Figure 2: Processing of Time Series Data Points

3.3 Orderbook Amount Feature

The order volume reflects the market transaction tendency. For each data point, we extract the volume of top 10 sell orders and top 10 buy orders. (In other words,we select the amount of 20 best orders of the market.)

3.4 Surprise Related Features

3.4.1 Prior of Orderbook

As shown in Figure 3, the data from every feature set of each price point is assumed to have a Poisson distribution.

3.4.2 Temporal Surprise Feature

- **T+1 Time-Series Feature**
  Surprise is computed for every price data point of top 20 orders which is in chronological order. When new data is observed with each new incoming price, the beliefs established thus far are used as prior,and Bayes' rule is applied to compute the posterior. The posterior at one price point is applied to compute the posterior.
- **T+N Time-Series Feature**
  Moreover, the interval of accumulative surprise learning is 1 second now, we plan to test a few more sets of values like 5, 10 or 15 to check the performance. **This part has not been tested yet.**

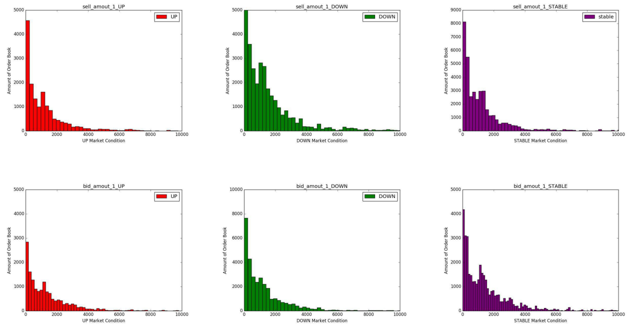


Figure 3: The distribution of orderbook amount(sell_1 and bid_1)

4. EXPERIMENT

We only use the Huobi Data from August 2015 to December 2015 (containing 6 million data points) to validate the features. All the features are measures with Precision, Recall and F1-Measure per-day.

- **Raw Data** We first classify each price as "up", "down" or "stable" with their 20 raw feature data. RandomForest achieved best performance among several basic classification algorithms.
- **Surprise Feature** We apply temporal surprise value of each feature groups to classify price label. It's obvious that Surprise feature is more powerful than raw data in discrimination.

All the experiment result are shown in Figure 4.

5. CONCLUSION

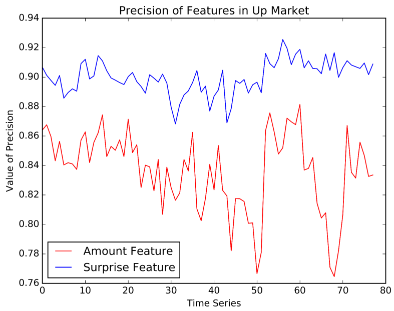
5.1 The Surprise Feature is Powerful

From the above pictures, it's obvious that for the "stable" price, surprise can barely be a predictive advantage since it nearly reach the same effect with raw amount feature on precision, recall and F-measure. But we aim at predicting the "down" and "up" label for each price to decide buy or sell. The three indicator values of surprise are all much higher than amount thus indicating surprise feature surely have an cutting edge on amount feature. Our Surprise feature is valid for predicting price change.

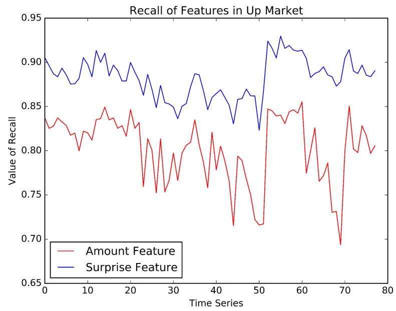
5.2 90% Accuracy != Huge Profit

However, such high value of indicators does not imply an affirmatory profitable transaction[4]. The actual transaction entails delay which is adverse to our strategy in 15 seconds.

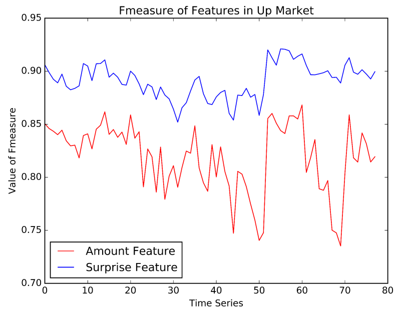
[4] Empirically, 75% of accuracy can balance profit and loss



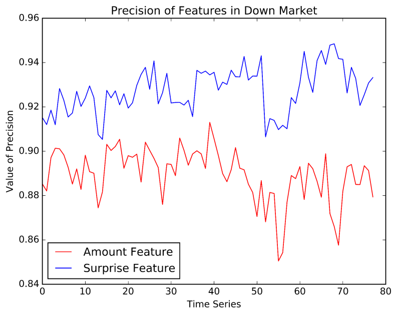
(a) up-Precision



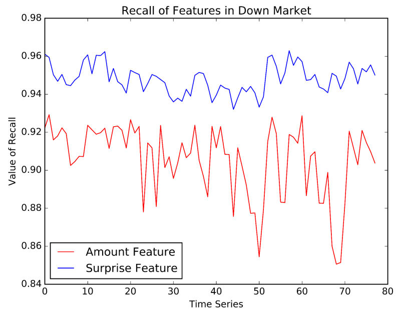
(b) up-Recall



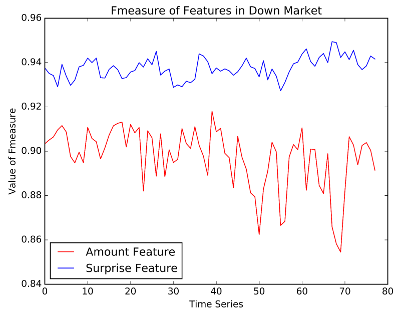
(c) up-Fmeasure



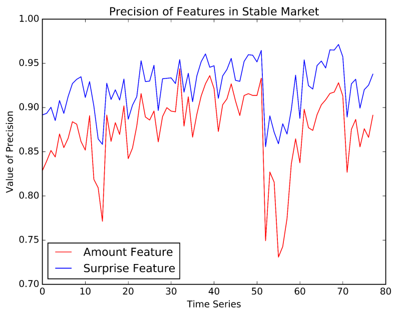
(d) down-Precision



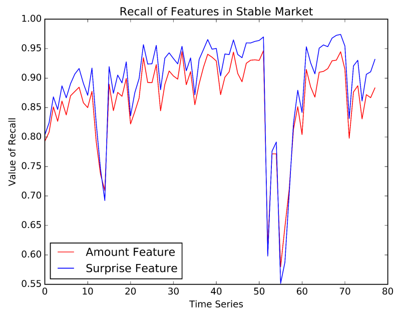
(e) down-Recall



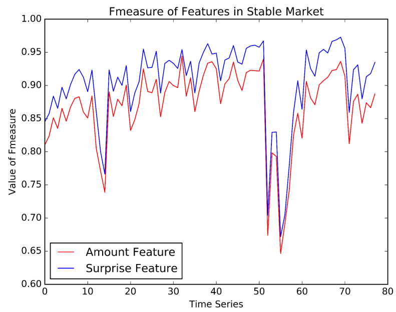
(f) down-Fmeasure



(g) stable-Precision



(h) stable-Recall



(i) stable-Fmeasure

Figure 4: Prediction Comparison of 2 Features in different Market Conditions

Sharp ratio is a measure for calculating risk-adjusted return, and this ratio has become the industry standard[5]. Our sharp ratio is mostly around 5 (MIT Bitcoin Prediction achieves sharp ratio of 4.5[6]) which means it only takes little risk in obtaining profit.

6. FUTURE WORK

- **Improve the adaption of Surprise algorithm in our scenario**
  We only consider the surprise between T+1 (the next second), we also need to test the performance of T+5 (after 5 seconds), T+10, T+15.

  The paper[7] has mentioned spatial surprise for a single neuron to derive its prior from the compound instantaneous activity of the surrounding neurons in the map. We plan to apply this thinking into our local sample feature space.

- **Classify the surprise points**
  We intend to locate the extreme "up" and "down" price point among data which have already been calculated with a surprise value. The steep rise and fall may help users gain more profit or avoid more risk.

- **Compare with another dataset**
  For now we only used one huobi dataset, later we will compare the surprise performance with that of the price on OKcoin dataset at same timestamp.

[5]Strategies with sharp ratio higher than 5 are extremely rare in quantitative finance.
[6]Paper Link: http://arxiv.org/abs/1410.1231v1
[7]Of bits and Wow: http://ilab.usc.edu/publications/doc/Baldi_Itti10nn.pdf