Cryptocurrencies Prediction & Forecast

Group 14:
Cheng-Hao Ho(ch3561)
Shuoting Kao(sk4920)
Wei-Ren Lai(wl2777)

12/17/2021

Overview

Novelty/Originality

- Reframe the problem - Predict the timing of **buy and sell** instead of price
- **Self-invented method** to predict the sweet buying/selling timing
- Monitor large transactions
- Real-Time whale alerts
- Event subscription

Big data challenge (volume, variety, and velocity)

1. **Volume**
   - The volume of transaction price/volume
   - Twitter posts
   - Reddit posts
   - Google trend
   - Wikipedia pageview
2. **Velocity**
   - Twitter streaming data
   - Transaction streaming data
3. **Variety**
   - Cryptocurrency transaction structural data
   - Cryptocurrency wallet in and out data
   - Twitter text data
   - Reddit text data
   - Google search popularity
   - Wikipedia Bitcoin topic popularity

Business value

- Provide valuable and accurate metric
- Offer a powerful tool for investors to increase their profit
- Help investors earn more money

Method & Challenges

Data labelling

Labels: Buy, Hold, Sell

Problem

To few buy and sell points. Only 3% of the data is labeled as buy or sell point, highly **imbalanced dataset**.

Problem

To few buy and sell points. Only 3% of the data is labeled as buy or sell point, highly **imbalanced dataset**.

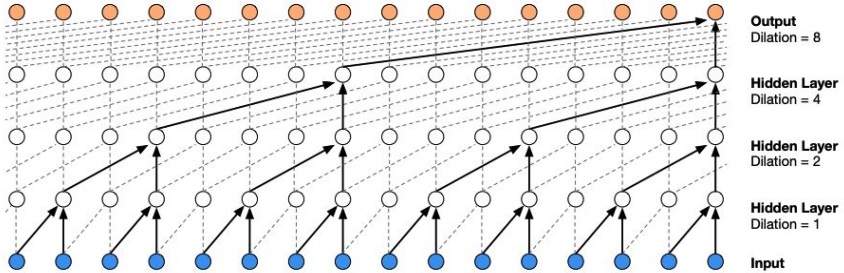
Solution: Loss Weight + Smooth data

Set a tolerant threshold (0.03). For example, if price 100 is one of our buy points, any price point that is smaller than 100*1.03 will also be labeled as “buy” point.

Deep Learning Models

WaveNet

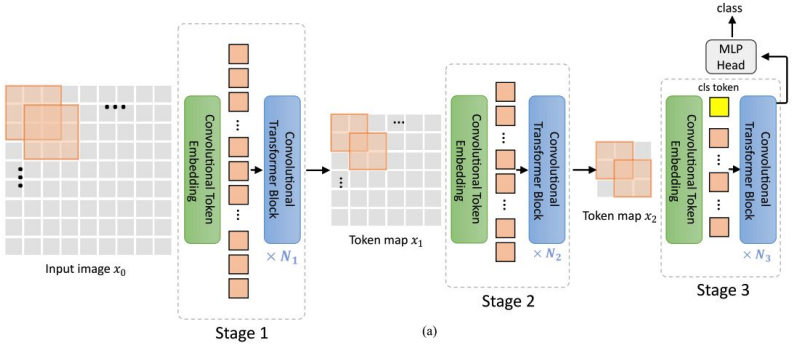
Parameter size ~ 40K



https://arxiv.org/pdf/1609.03499.pdf

CvT
(Convolution Vision Transformer)

Parameter size ~ 1M



https://arxiv.org/pdf/2103.15808.pdf

Problem

At the beginning, the result is too noisy. The model obviously doesn’t provide good strategy.

Solution

We use three strategies to solve the problem:

Probability Distribution:

(buy:0.8, sell: 0.1, hold: 0.1 v.s. buy:0.4, sell: 0.3, hold: 0.3)

1. Use entropy as a threshold.
2. Use highest percentage as threshold.
3. Use difference between the highest and the second highest percentage as threshold.

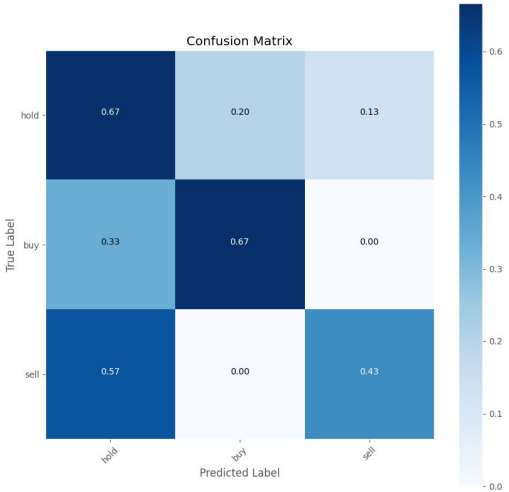
   In our observation, **result 1 and 2** usually have better performance.

Result

Cleaned version:

How to evaluate our model?

- **Confusion matrix**: Confusion matrix can give us the preliminary concept of how good our model is. However, because our model is used to predict “buy” and “sell” points, the accuracy doesn’t fully explain our result at all.



How to evaluate our model?

- Confusion matrix
- **Backtesting**: Backtesting is one the most straightforward solutions. In our model, we just try to imitate the actions of real people: buy at buy points and sell at sell points. Here, we suppose we have one million US dollars in the beginning. And see how much we can earn at the end.

Training data and Test data

Training data: from 2015-07-20 to 2021-03-03

Test data: from 2021-03-04 to today

Why? Because the Bitcoin price on 2021-03-04 is about 49000 (USD), and the price today is about 47000 (USD). We choose similar start price and end price to avoid biased output.

Backtesting Result: every 1 day data

Model1: BTC price/volume

Model2: BTC price/volume + Google Trend

- Use entropy as threshold: $1369306
- Use percentage as threshold: $1215603
- Use difference as threshold: $1162341

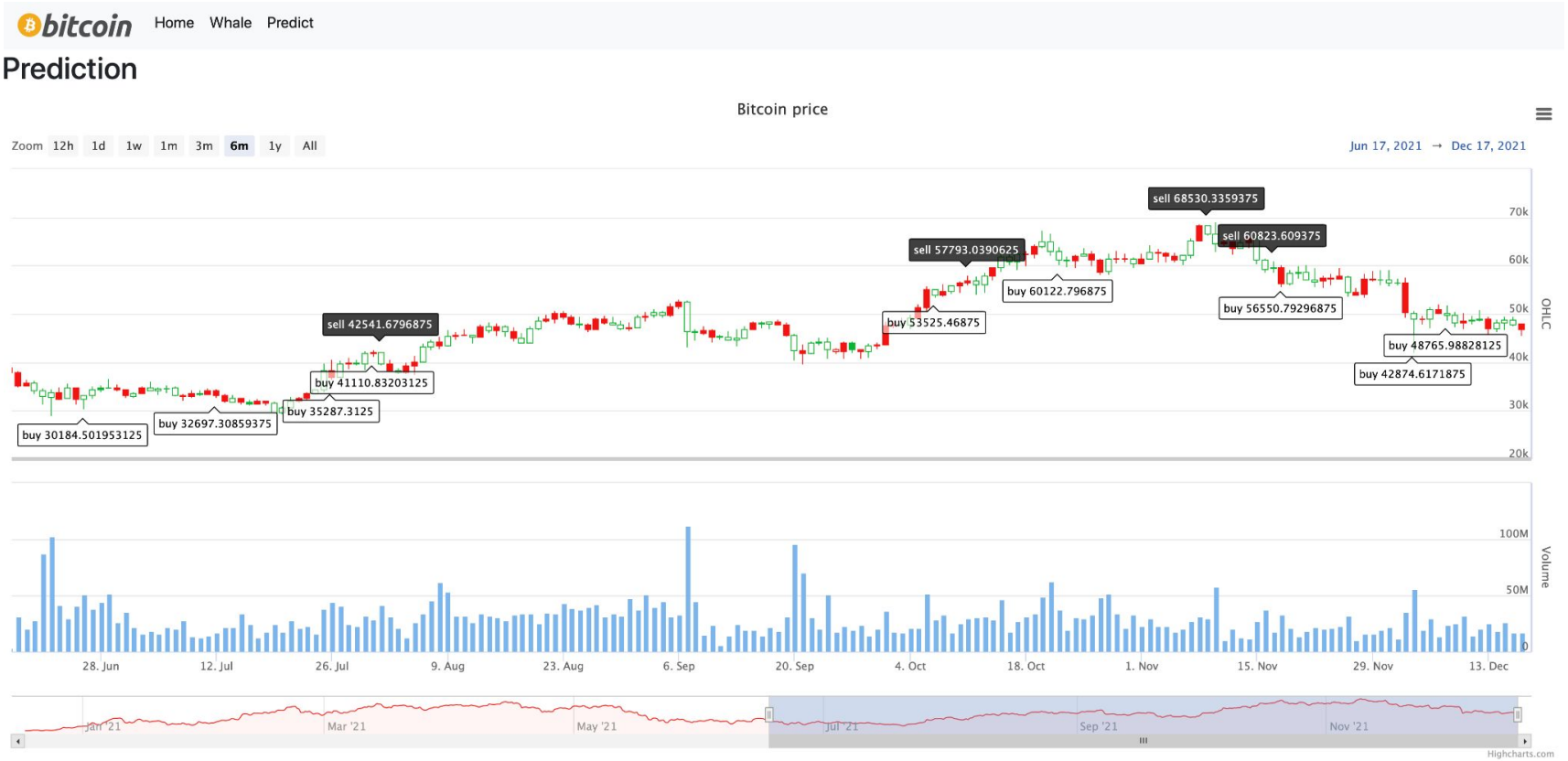
Model3: BTC price/volume + Wikipedia Pageview

- Use entropy as threshold: $1143567
- Use percentage as threshold: $1032401
- Use difference as threshold: $982564

Model4: BTC price/volume + Google Trend + Wikipedia Pageview + (Sentiment analysis + huge amount transaction information)

- Use entropy as threshold: $2122670
- Use percentage as threshold: $1570829
- Use difference as threshold: $1177225

Backtesting Result: 1 day data



Backtesting Result: every 1 hour data

Model1: BTC price/volume

Model2: BTC price/volume + Google Trend

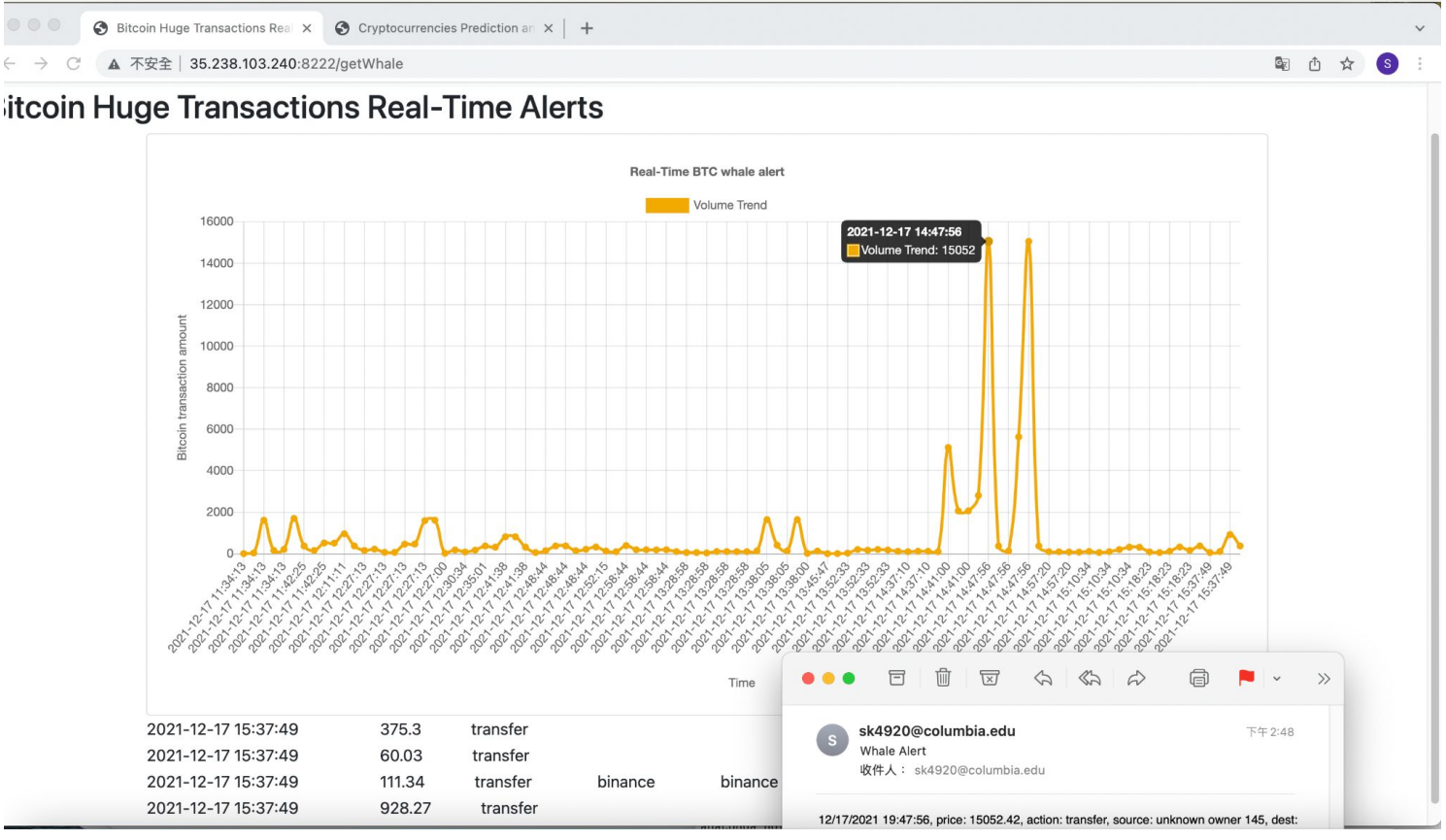
- Use entropy as threshold: $1145321
- Use percentage as threshold: $1032654
- Use difference as threshold: $9821547

Model3: BTC price/volume + Google Trend + (Sentiment analysis + huge amount transaction information)

- Use entropy as threshold: $1183372
- Use percentage as threshold: $1282376
- Use difference as threshold: $1014512

Backtesting Result: 1 hour data

Real-Time Whale Alert and Event Notification



DEMO

Reference

1. https://medium.com/general_knowledge/watch-the-whales-101-guide-to-wallet-tracking-8ff5799f3dc4
2. https://github.com/manthanthakker/BitcoinPrediction
3. https://col-jung.medium.com/how-to-make-realistic-cryptocurrency-price-predictions-436f3f6f54e3
4. D. R. Pant, P. Neupane, A. Poudel, A. K. Pokhrel, and B. K. Lama, “Recurrent neural network based bitcoin price prediction by twitter sentiment analysis,” *IEEE 3rd International Conference on Computing, Communication and Security (ICCCS). IEEE*, 2018, pp. 128–132.
5. Jiawei Han, Jian Pei, Yiwen Yin. “Mining frequent patterns without candidate generation“ *ACM SIGMOD Record, Volume 29, Issue 2*, June 2000, pp 1–12
6. Jian Pei, Jiawei Han, B. Mortazavi-Asl, Jianyong Wang, H. Pinto, Qiming Chen, U. Dayal, Mei-Chun Hsu. “Mining sequential patterns by pattern-growth: the PrefixSpan approach” *IEEE Transactions on Knowledge and Data Engineering, Volume 16, Issue 11*, Nov. 2004
7. Oord, Aaron van den, et al. “Wavenet: A generative model for raw audio.” *arXiv preprint arXiv:1609.03499* (2016).
8. Chun-Hung Cho , Guan-Yi Lee , Yueh-Lin Tsai , Kun-Chan Lan “Toward Stock Price Prediction using Deep Learning” *UCC '19 Companion: Proceedings of the 12th IEEE/ACM International Conference on Utility and Cloud Computing CompanionDecember* 2019 Pages 133–135
9. Liu, Jialin, et al. "Stock prices prediction using deep learning models." *arXiv preprint arXiv:1909.12227* (2019).
10. Marco Cerliani “Corr2Vec: a WaveNet architecture for Feature Engineering in Financial Market.” https://towardsdatascience.com/corr2vec-a-wavenet-architecture-for-feature-engineering-in-financial-market-94b4f8279ba6
11. Hu, Zexin, Yiqi Zhao, and Matloob Khushi. "A survey of forex and stock price prediction using deep learning." *Applied System Innovation* 4.1 (2021): 9.
12. Asutosh Nayak. “Stock Buy/Sell Prediction Using Convolutional Neural Network” https://towardsdatascience.com/stock-market-action-prediction-with-convnet-8689238feae3
13. Xinpeng Yu and Dagang Li. “Important Trading Point Prediction Using a Hybrid Convolutional Recurrent Neural Network.” *Appl. Sci.* 2021, 11, 3984. https://doi.org/10.3390/app11093984
14. van den Oord, Aaron, Kalchbrenner, Nal, and Kavukcuoglu, Koray. “Pixel recurrent neural networks.” *arXiv preprint arXiv:1601.06759*, 2016a.
15. Arvind Kalia, N. W. (2019). Association Rule Mining for Stock Data. *International Journal of Advanced Science and Technology, 28(19), 796 - 802*. Retrieved from http://sersc.org/journals/index.php/IJAST/article/view/2665
16. S. M. Idrees, M. A. Alam and P. Agarwal, "A Prediction Approach for Stock Market Volatility Based on Time Series Data," in *IEEE Access*, vol. 7, pp. 17287-17298, 2019.
17. Kristoufek, L. BitCoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the

Q&A