Research on the Impact of UGC Based on Fluctuation Mode on Cryptocurrency Market

Kun Jia[1,2], Yizhen Zhu[2] and Yuxin Zhang[1,*]
[1]Institute of Artificial Intelligence and Change Management, Shanghai University of International Business and Economics, Shanghai, China
[2]School of Statistics and Information, Shanghai University of International Business and Economics, Shanghai, China

Keywords: Blockchain, Cryptocurrency, Fluctuation Pattern, Sentiment Analysis, Investment Advice, Financial Public Opinion.

Abstract: As the investment properties and financial properties of cryptocurrencies are accepted or recognized by major listed companies and even sovereign countries, they are increasingly sought after by ordinary investors around the world. The frantic influx of a large number of investors into the cryptocurrency market has also led to more and more market analysis and predictions for retail investors and new investors, as well as popular science teaching videos on the YouTube website. However, their sentiment towards cryptocurrency is not must be "right". To ascertain the attitude of video producers in YouTube videos towards the market, this article uses sentiment analysis to analyze the videos and compare the similarity with the Bitcoin fluctuation data before and after the video upload. We found that in most cases, video producers ’attitudes have a high degree of similarity with the ups and downs of the period before the upload of the video, but the low similarity to the trend for some time in the future. This shows that the cryptocurrency-related videos on YouTube merely reflect past ups and downs and cannot be used as investment advice, even if the video uploaders are doing so.

1 INTRODUCTION

In January 2021, the "GameStop" incident in the US stock market suddenly broke out. It was based on the retail accounts of the largest US forum website Reddit and the stock topic forums with millions of fans on the website. and concentrated buying, which was strongly shorted by hedge funds. The company’s GameStop stock, and used the power of the group to push up the price of the stock, forcing hedge funds Citron Research, Melvin Capital, and other short positions to surrender. In the end, the hedge funds ended disastrously, wrote a historic page of Wall Street.

Reddit netizens who are proud to carry forward the MEME picture will naturally not miss the Dogecoin born because of the Doge emoji. Elon Musk can be described as the best "cargo carrier" for Dogecoin. The price of Dogecoin has skyrocketed many times before, in many cases, it is derived from his tweets, which can be traced back to April 2019. At that time, the official Dogecoin account launched a vote for the Dogecoin CEO on Twitter, and Musk was elected with high votes. Subsequently, Musk tweeted that Dogecoin is his favorite digital currency and changed his Twitter account information to "Former Dogecoin CEO". On April 15 this year, Musk posted a picture of "dogs roaring on the moon" on social platforms, implying that he will bring Dogecoin to the moon, and therefore Dogecoin has a "To the Moon" Slogan. And his series of actions have also attracted investors' pursuit, and the price of Dogecoin has risen all the way.

Since the "GameStop" incident, people seem to be more willing to believe in the power of self-media and social media than large companies or institutions. In particular, a large number of new cryptocurrencies continue to emerge. In the 24-hour cryptocurrency market where prices soar and plummet at any time, everyone hopes to reproduce the myth of Bitcoin in themselves and get a share in this frenetic market. However, we compared the actual fluctuations before the video was released with the binary data generated after sentiment analysis on the subtitles of the video released on YouTube, and found that:

In the English video, the similarity between the two reached 78% and comparing the actual fluctuations after the video was released, it was found

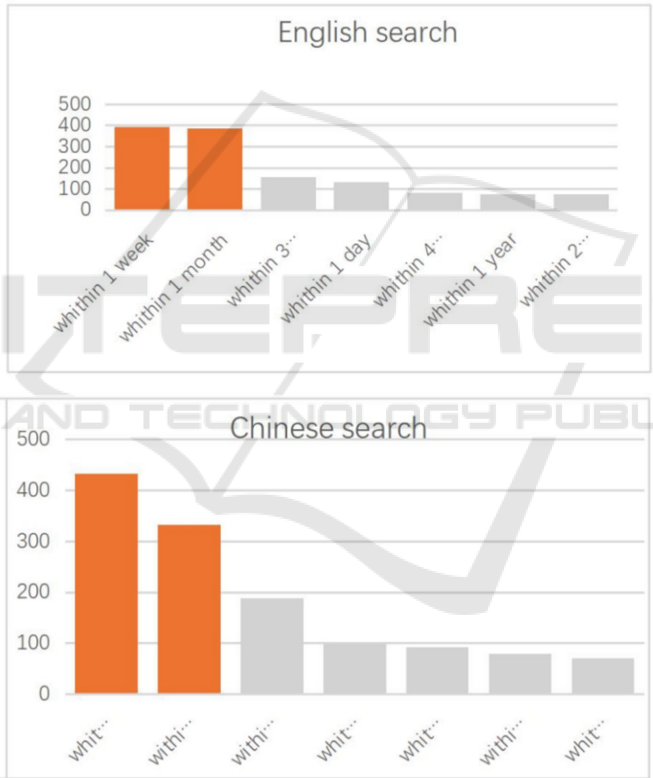
Jia, K., Zhu, Y. and Zhang, Y.
Research on the Impact of UGC Based on Fluctuation Mode on Cryptocurrency Market.
DOI: 10.5220/0012042400003620
In *Proceedings of the 4th International Conference on Economic Management and Model Engineering (ICEMME 2022)*, pages 707-712
ISBN: 978-989-758-636-1
Copyright © 2023 by SCITEPRESS – Science and Technology Publications, Lda. Under CC license (CC BY-NC-ND 4.0)

that the similarity was only 51.1%. This is almost the same as the probability of blindly guessing whether the rise or fall. It is also for investors. It does not constitute any investment advice. In Chinese video, the similarity of the former is as high as 99.2%, and the similarity of the latter is 64.2%. This may be due to the small amount of data and the fact that the video content is mostly positive.

2 DESCRIPTIVE STATISTICS

2.1 YouTube Data Descriptive Statistics

Using data mining techniques to search for videos on YouTube based on the relevance of keywords such as "Bitcoin", "Ethereum", "cryptocurrency", "Dogecoin" in Chinese and English, and obtain 1408 valid data in Chinese search results. There are 245 pieces of subtitles that can be generated; 2131 pieces of valid data in English search results are obtained, of which 1,691 pieces of subtitles can be generated. The following does not give special instructions, which means that the data is as of June 10th (The data source will be available at https://github.com/SUIBE-jk/YouTube-DATA).



SCITEPRESS
SCIENCE AND TECHNOLOGY PUBLICATIONS

Figure 1: Chinese and English retrieval to get the number of videos in each period.

As you can see, with the support of the YouTube video recommendation system, most of the data we get has been uploaded recently. Of course, this is also related to the surprising increase in cryptocurrency in the fourth quarter of last year and the first quarter of this year.

Make word cloud diagrams for the titles of all the search results in both Chinese and English. It can be seen that keywords such as "analysis" and "forecast" which guide trading appear more frequently; while "big rise" and "big drop" appear Frequent words that are eye-catching; in the cryptocurrency market, Musk, who has "one call to a hundred responses", appears very frequently in video titles, usually in the same field as Dogecoin.

Figure 2: Comparison of Chinese and English word cloud diagrams.

2.2 Cryptocurrency Data Descriptive Statistics

The data on the rise and fall of the cryptocurrency used in this article comes from investing.com. Because of the predecessors' understanding of cryptocurrency and the research on the risk of cryptocurrency, it has been very mature and complete. However, domestic research on cryptocurrency mainly focuses on Bitcoin, the original cryptocurrency, and at the same time, there is not enough attention to the practical application value of later new cryptocurrencies. At present, there is no research on whether there is a phenomenon of self-media misleading investors with general emotional bias in YouTube videos. Therefore, this article takes the overall emotional bias in YouTube videos as the research object and compares the fluctuation patterns obtained through Bitcoin's daily fluctuations.

Table 1: Data example.

Date	Close**	Open*	High	Low	Volume	Chg
2021/6/29	36,342.30	34,477.30	36,410.30	34,247.60	102.51K	5.41%
2021/6/28	34,475.90	34,682.20	35,231.20	33,944.90	112.00K	-0.58%
2021/6/27	34,678.50	32,247.10	34,685.50	32,041.70	148.80K	7.55%
2021/6/26	32,243.40	31,592.10	32,643.00	30,206.90	156.67K	2.06%

3 FLUCTUATION PATTERN

By observing the morphology of the time series, it is found that the morphology of the time series is composed of three states: rising, maintaining, and falling. In this article, to dig out the morphological similarity between the sequences in the massive time-series data set, we propose a fluctuation pattern. The concept of FP) saves the fluctuation trend of the original sequence through the extended Boolean time series, uses {1, 0,-1} as the value range of the Boolean time series data to save the original time series data change trend, and only pays attention to the original time series. The morphological characteristics of the ordinal data, regardless of its numerical value, are represented by 1 as increasing, 0 as maintaining, and -1 as descending. In the subsequent similarity calculation, the generated Boolean time-series data is used for similarity measurement.

For $X < d_1, d_2, \cdots, d_n >$, there are Boolean time-series data $B < b_1, b_2, \cdots, b_{n-1} >$ and:

$$b_i = \begin{cases} 1 & d_i < d_{i+1} \\ 0 & d_i = d_{i+1} \\ -1 & d_i < d_{i+1} \end{cases} \quad (1)$$

B records all the fluctuations of X, so B is called the fluctuation pattern of X.

4 EMOTION ANALYSIS

Text sentiment analysis can be roughly divided into two categories: one is based on the sentiment dictionary to get the sentiment score of the text. The second is based on machine learning. First, use the manually labeled text (whether the comment has been marked as a positive comment or a negative comment), and put in various algorithm models (such as Naive Bayes, SVM, etc.) for training, and finally realize the new comment Classification is essentially text classification. In this case, the data content is the subtitle content of the video crawled on the YouTube platform, that is, the video producer's product comment on the encrypted currency. The useful variable in the data is the "comment". We use this data to obtain a sentiment of the up master on the rise and fall of cryptocurrencies. The framework diagram of the sentiment analysis method used in this article is as follows:

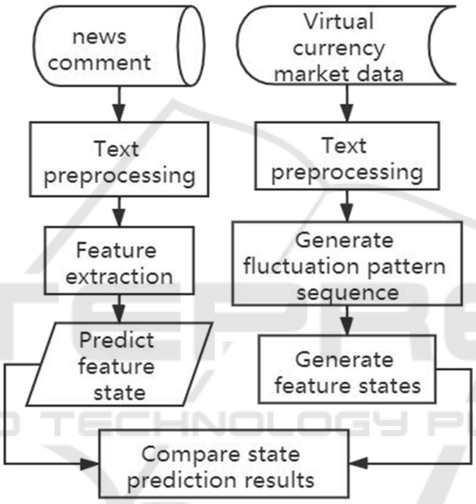


Figure 3: Sentiment analysis method framework image.

4.1 Chinese Text

Based on the sentiment analysis of Chinese text, we realized it by using snownlp. Through this case, we learned about snownlp and directly called SnowNLP(txt).sentiments to calculate the sentiment score, which filled the gap that the installed library could not be called, and successfully used the function to realize the code. The idea of reuse, this case directly calls the pre-trained algorithm model of Snownlp, without actually training and tuning the algorithm model, and the accuracy may need to be improved. Snownlp can mainly perform Chinese word segmentation (the algorithm is Character-Based Generative Model), part-of-speech tagging (the principle is TnT, 3-gram Hidden Markov), sentiment analysis, text classification based on the principle of naive Bayes, pinyin conversion, traditional to simplified, Extract text keywords and abstracts, segment sentences, and text similarities based on TextRank. It is predicted from the model that the up subject of each video corresponds to the emotional tendency of the cryptocurrency problem. The emotional tendency is divided into positive and negative, which is a typical two-category problem.

4.2 English Text

TextBlob is a Python library for processing text data. It provides a simple API for common natural language processing (NLP) tasks, such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, etc. In this case, we use the textblob library to do sentiment analysis on English text. The results are obtained through word extraction, sentence sentiment value

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

calculation, and syntactic analysis. TextBlob has been looking for words and phrases that can be assigned polarity and subjectivity and averaged them together. This is the principle when processing long texts, that is, simple average.

Table 2: English sentiment analysis example.

	title-ID	post_title	post_content	sentiment	sentiment_index
0	0	NO. 1	***	Negative	-1
1	2	NO. 2	***	Positive	1
2	3	NO. 3	***	Negative	-1
3	4	NO. 4	***	Positive	1
4	5	NO. 5	***	Positive	1
5	6	NO. 6	***	Negative	-1

5 RESULTS & DISCUSSION

We compared the actual fluctuations of a certain period before the video was released and the time series generated after emotional classification of the related video subtitles released on YouTube at this time using the longest common subsequence (LCS) and found that two The similarity of the video producers is greater, that is, the emotional tendency of the video producer is largely based on the previous rise and fall of the market, and the actual rise and fall of the video after the release of the video is compared with this time series data, and it is found that the similarity is relatively high. Small, that is, the emotional tendency of the video publisher and cannot play a role in predicting the rise and fall of the market.

Table 3: Comparison of results.

	Chinese data	English data
Before upload LCS	**99.2%**	**78.0%**
After upload LCS	**64.2%**	**51.1%**

6 CONCLUSIONS

This research takes the content of YouTube video subtitles as text data as an example, uses word cloud graphs and topic model feature analysis to analyze the text data of video subtitles content, and uses a python-based emotional dictionary method to calculate the emotional score of each comment one by one. The research analyzes the sentimental tendency of up owners towards cryptocurrencies and provides suggestions for investors based on the research results to help meet the expectations and needs of investors, thereby reducing the risk of investment. In the process of text mining, this study failed to identify and eliminate false comments, and the constructed dictionary was not complete, and there was missing vocabulary. Future research will pay more attention to the quality and authenticity of the data itself and improve the construction of the dictionary. However, this case is based on real-world real data, using data mining methods to analyze the emotional tendency of video uploaders to cryptocurrency, and provides a data science research paradigm for investors to make corresponding investment behaviors.

ACKNOWLEDGMENTS

This work was supported by Shanghai University of International Business and Economics Postgraduate Research Innovation Cultivation Program.

REFERENCES

Fang Y, Sugano K, Oku K, Kawagoe K. (2015) Applying a Multi-dimensional Time-Series Similarity Method to Typhoon-track Prediction. In: Proceedings of the 2015 IEEE 11th International Conference on e-Science (E-SCIENCE’ 15). Washington. 259–262.

Fu M, Wang D. (2020) Service quality evaluation of fresh agricultural products cold chain logistics based on principal component and neural network. In: IOP Conference Series: Earth and Environmental Science. Hulun Buir. 585: 012103.

Li Y, Lu L. (2019) Research on B2C Reverse Logistics Service Quality Evaluation System. In: Proceedings of the 2019 5th International Conference on E-Business and Applications, Bangkok. 10-15.

Liu B, Zhang Y. (2011) Research on evaluation of third-party logistics service quality based on dynamic fuzzy sets. In: MSIE 2011. Harbin. 833-837.

Qian C, Wang Y, Hu G, Guo L. (2015) A novel method based on data visual autoencoding for time series similarity matching. In: Proceedings of the 27th Chinese in Control and Decision Conference. Qingdao. 2551-2555.

Wu J, Lu K. (2019) Chinese weibo sentiment analysis based on multiple sentiment lexicons and rule sets. Computer Applications and Software, 36(09): 93-99.

Zhang J, Liu X. (2017) Evaluation of Integrated Logistics Service Based on SERVQUAL Model. In: International Conference on Computer Systems, Electronics and Control (ICCSEC). Dalian. 100-104.

SCITEPRESS
SCIENCE AND TECHNOLOGY PUBLICATIONS