UDC 681.513.7; 519.688

O. Gavrylenko, M. Miahkyi, Y. Zhurakovskyi

THE TASK OF ANALYZING PUBLICATIONS TO BUILD A FORECAST FOR CHANGES IN CRYPTOCURRENCY RATES

Abstract: An analysis of methods, areas of application, and approaches to analyzing publications and forecasting events based on the data collected, and also was given a concept of the impact of publications on the change of cryptocurrency course. The substantiation of urgency of a theme is presented and possibilities of corresponding application of results of work are described. The main stages of working with event forecasting data are identified, namely: preliminary data processing, their further analysis and forecasting. One of the key factors in the formation of preliminary requirements is the definition of target data, i.e. those in which there will be keywords and a certain emotional color. In addition, it is also necessary to identify target experts, i.e. to select people whose positions, in our opinion, will have a significant impact on the further formation of the information flow. The relevance of using the introduced ratio of emotional coloring of target posts and the interest of target persons has been revealed.

Keywords: cryptocurrency, revent forecasting, data processing, data analysis, forecast construction, target data, target experts, emotional coloring, information flow, content monitoring.

Introduction

Today you can get any information without leaving home, via the Internet. However, as is always the case, there are problems with accessibility - a large amount of non-informative data that makes no sense, in other words: "garbage data", which the current Internet is full of. This is why, every day, it becomes harder and harder to find the necessary information and analyze so much data.

This problem is inherent in all areas of human activity, whether building your own business or participating in society through governing bodies.

Also, the emergence of large amounts of uninformative data, as always, is the fault of humanity itself, as every entrepreneur seeks to maintain as much information about the state of his business and at the same time, seeks to remain competitive, as it is a vital factor for any business

Given all this, the methods of analysis of Internet content must be constantly evolving to provide an opportunity to conveniently and most importantly quickly find the necessary data.

© **O. Gavrylenko, M. Miahkyi, Y. Zhurakovskyi**

ISSN 1560-8956

As you know, the most expensive today is information, and at the moment, the main sources of information are messengers, in which people share their thoughts and plans. However, to use these publications for your own purposes, you need to analyze the data.

Analysis of recent publications

In general, the task of analyzing publications from the Internet is very important, because a well-analyzed publication can provide much more information than a simple question to the author, such as: "What did you mean in this publication?". The paper [1] reviews the process of computational definition and categorization of opinions expressed in a fragment of the text to determine whether the writer's attitude to a particular topic, product, etc. is positive, negative or neutral. A detailed study of the analysis of temporary moods and the causal relationship of moods. When using the analysis of temporary moods, we can learn a generalized event based on mood and time. On the other hand, the use of causation will be useful not only for determining the causes and consequences, respectively, but also for their further prediction. The main part of the publication is a review of the combination of these two approaches, which degenerates into a model that allows you to determine the mood for future events, as well as to create a time prediction about the time that will elapse between certain events. The following parameters were used to estimate accuracy: mean absolute error and root mean square error.

To view the publications, you need to choose the place where they are most and they are in a single text format, for this a good messenger - Twitter, in the publication [2], was considered in detail a special linguistic analysis and statistics of Twitter. This study aimed to identify criminal elements in the United States by modeling topics for discussion and further incorporating them into the crime prediction model. Thus, a study was conducted on the impact of social media publications on future crimes.

In the following work [3], methods for predicting the user's estimation of certain elements using probabilistic algorithms were considered. In fact, the article perfectly illustrates the existence of computational patterns as to what users of the network will like in certain circumstances. In other words, this study highlights the impact of probabilistic algorithms in the field of recommendation systems, as well as provides an overview of key methods that have been successfully applied. The object classification algorithms that have been considered allow us to solve the problem of predicting the evaluation of content by users and its categorization, as well as to improve existing methodologies for building information systems.

The publication [4] is quite relevant today due to the difficult epidemiological situation in the world. It analyzed microblogs on Twitter and suggested several methods for identifying messages. It was determined that in ten weeks out of more than five hundred thousand reports, their best model reaches a correlation of 0,78 with CDC statistics.

Also, it is impossible not to cover Internet blogs, in which many people express their own thoughts and perceptions of certain problems and more. Therefore, in the publication [5], a study was conducted to identify hate groups. The proposed approach is semi-automatic and consists of four modules, namely: spider blog, information retrieval, network analysis and visualization. The study was conducted on a blogging site - Xanga. The results of the analysis were the identification of some interesting demographic and topological characteristics in hate groups and identified at least two large communities in addition to the smaller ones. The proposed approach is also appropriate to use in the study of hate groups and other related communities in blogs.

For businesses, the process of analyzing large amounts of data and understanding the needs of most people is very important, as it directly affects the company's revenue. In [6], a constructive review of the problem of business accumulation of large amounts of data and problems with their intellectual processing was conducted. A clear wording and explanation of the terms "data mining" and "data mining" are provided. In the end, an objective conclusion was made about the feasibility of using data mining to increase the competitiveness of enterprises.

Determining the method of analysis of publications

Common to all previously considered publications is that they collect data from certain sources and after receiving them in sufficient quantities, begin to analyze them. However, in my opinion, the main problem with this approach is that the data was selected without prior preparation of a number of requirements that these data must meet, which, in turn, leads to the fact that we have a lot of "unimportant" effects on further analysis. Therefore, I think it is necessary to start with the definition of target data, which will be set by a number of criteria (for example: the data must contain the word bitcoin, etc.). After determining the target data, it should also be noted an equally important criterion, namely, who will act as experts in this area, as each person is engaged in a particular activity and achieves results in it, but, clearly, can not be an expert in absolutely everything. Thus, before searching and collecting data, we will develop certain criteria that will help collect more data that is more relevant to the subject of analysis and at the same time weed out data that will not be relevant to him.

In the analysis, we need to examine two main points: the interest of authors in creating a certain impact on people through their text, as well as to determine whether this data makes sense, given the structure of the message, ie to determine whether the text was actually written by man was done through automated systems with automatic text generation and inclusion of keywords.

Also, an important aspect of each publication is its emotional color. There are three categories: positive, negative and neutral. If it is impossible to determine the emotional color

ISSN 1560-8956

in the publication, it also means that the text was not written by a person and we need to remove this publication to free up resources of the analyzer.

After making all the necessary changes to the data collected, it is worthwhile to begin the process of creating a ratio of expert interest and behavioral coloring. This is necessary in order to have an idea of the direction in which an expert wants to influence certain events, for example, to determine whether a person wants to adjust public opinion to something negative or not. However, this is not the only option, as, with a neutral emotional color and high interest, the expert clearly tries to maintain a position for one reason or another, but he does not act in a leading role, in the case of positive or negative color, on the contrary, the expert wants to independently form public opinion about something in a certain way.

Therefore, after obtaining the ratio data, you can determine the direction of publications and based on the analyzed data to make certain hypotheses, such as whether the publication will succeed, ie, the expert will be able to influence the event as planned or not.

An example of the process described above can be found in Fig. 1 below.

The first stage in Fig. 1 is “Definition of target data”. This is an aggregation of data on cryptocurrency rates. The second stage is “Definition of target expert”. This is an aggregation of data on target experts. The third stage is “Search for publications that would contain key data”. After this stage, the following processes will proceed in parallel, namely: analysis of the interests of the authors of publications in the change of the exchange rate of cryptocurrencies and the selection of data that does not make sense. Further, on the basis of the received data, it is carried out “Creating a relationship of interest and extreme behavioral coloring”. Next stage is “Determining the direction of publications” and the last one is “Forecasting based on the obtained data”.

The first stage in Fig. 2 is “Definition of target data”: Aggregation of data about cryptocurrency rates.

When studying this problem, the question arises of using existing methods for analysis, however, in [7], it is well explained that the initial paradigms of search engines and content analysis, formed several years ago, no longer correspond to the real situation. Therefore, it is necessary to implement content monitoring technology, which will provide a number of advantages: obtaining operational quantitative and qualitative analytical sections, timely provision of necessary profile information, ensuring targeted work of employees, eliminating distractions inherent in the Internet, data protection, confidentiality.

Due to such characteristics, content monitoring technology can significantly improve the efficiency and quality of information and analytical work.

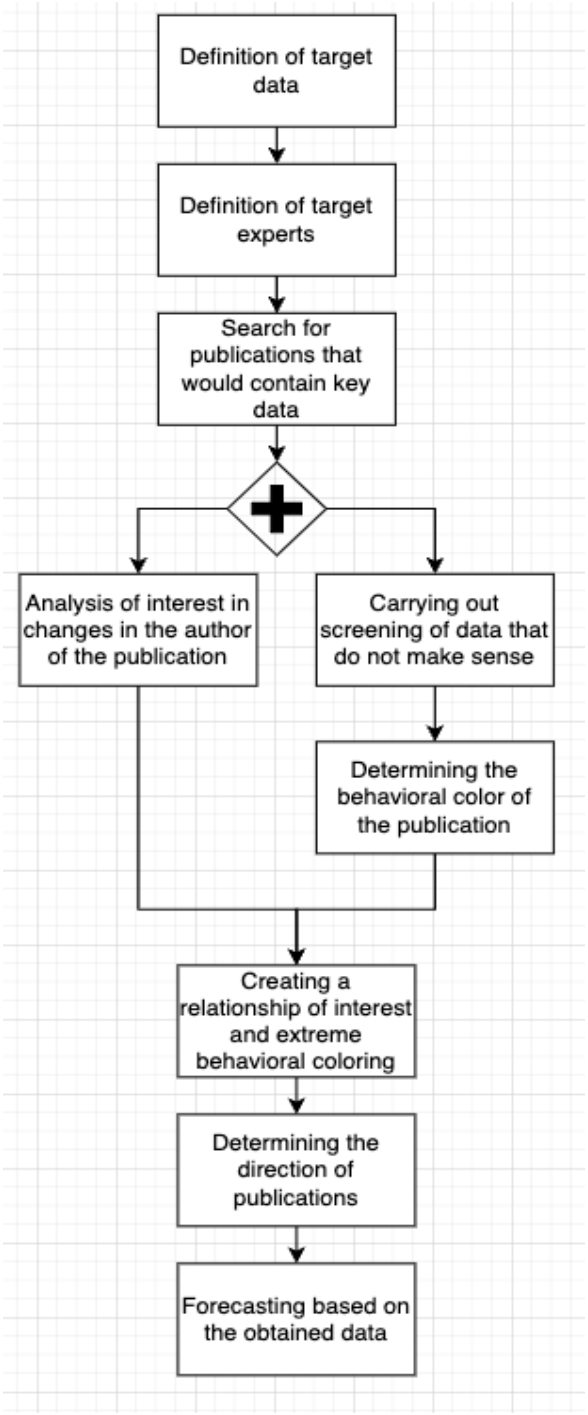

Figure 1. The sequence of stages of analysis processes

The concept of the impact of publications on the change of cryptocurrency course

Based on our previous analysis we provide the next concept of the impact of publications on the change of cryptocurrency course.

In order to determine the impact of a publication on changes in cryptocurrency rates, you need to take a few steps:

- Data search;

ISSN 1560-8956

- Data research;
- Forecasting theoretical changes;
- Data comparison;
- Forecast adjustment;

At the data search stage, you need to pre-configure search filters, such as:

- Target person (famous person whose publications will be selected for consideration);
- Target place to search for posts;
- Target cryptocurrency for research;
- Target place to search for changes in the selected cryptocurrency;
- Creating / selecting a list of keywords;
- Target dates for analysis.

After determining all the above data, we will have access to the desired publications of the selected celebrity for a certain period. Based on these data, you must also check the results of the selected cryptocurrency on a particular exchange for the selected period of time.

At the stage of data research, the behavioral nature of the collected publications will be determined, which will provide an opportunity to identify the significance of certain posts regarding theoretical changes in the cryptocurrency rate.

At the stage of forecasting theoretical changes, the calculation of the change in the cryptocurrency exchange rate will be calculated, taking into account the impact of publications. To do this, enter the significance factor c_t, which is calculated by the formula:

$$c_t = k_t \cdot \mathrm{ch}_t \tag{1}$$

where ch_t- assessment of the behavior of the post,

$$\mathrm{ch}_t = \begin{cases} 1, & \text{if post is positive,} \\ 0, & \text{if post is neutral,} \\ -1, & \text{if post is negative.} \end{cases} \tag{2}$$

where k_t - the correctness of the previous forecast,

$$k_t = |P_t - F_t|, \tag{3}$$

where P_t - the predicted value of the cryptocurrency exchange rate is obtained using time series, F_t - the actual value of the cryptocurrency exchange rate, t - moment in time.

After determining the coefficient c_t from (1)-(3), a forecast of the change in the cryptocurrency exchange rate will be created, based on existing data for the period of time from the moment when the publication was posted up to 1 hour.

$$P'_{t+1} = P_{t+1} + c_t. \tag{4}$$

At the stage of comparing the data, we compare the forecast data with the actual ones and calculate the deviation from the norm if it is less than 25% of F_t. This value was chosen because the cryptocurrency rate changes not only due to the publications of famous people, but also a large number of other factors, such as the number of financial transactions in the last hour, the weakening of other currencies, and so on.

At the stage of adjusting the forecast, it is necessary to make changes to the significance factor to make the analysis of further changes more accurate.

Concept of the impact of publications on the change of cryptocurrency course

An example of the process diagram described above can be found in Fig. 2 below.

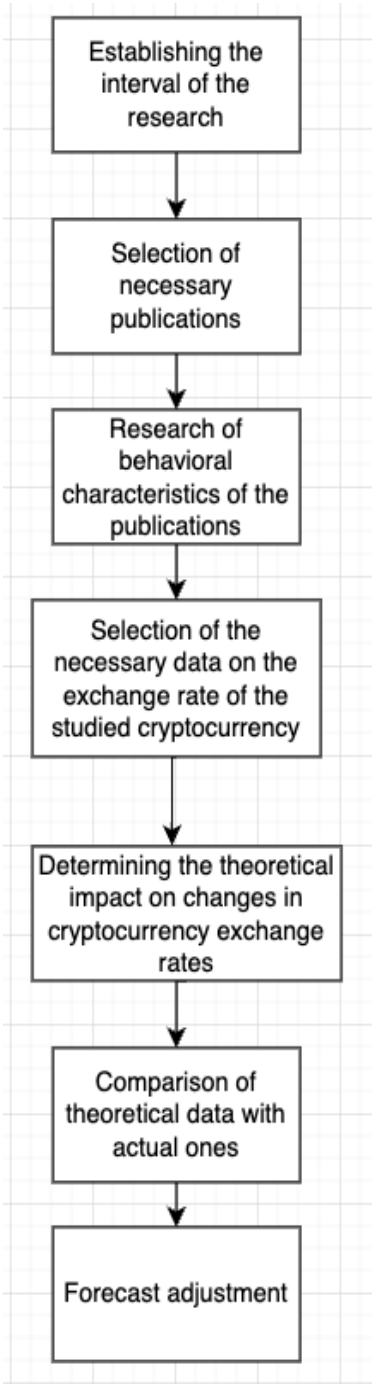

Figure 2. Stages for calculating the impact
of publications on changes in cryptocurrency exchange rates

ISSN 1560-8956

This diagram describes the process of analyzing the impact of a particular publication on the rate of the selected cryptocurrency, adjusting the significance factor for further analysis, which improves the accuracy for future forecasts.

Based on the formulated concept, a study was conducted of changes in the exchange rate of the "Dogecoin" cryptocurrency in the period from May 12 to 17, 2021. They also provide a scheme for calculating the influence of famous people's publications on the course of cryptocurrencies. Note that at this time, Elon Musk made three publications about the cryptocurrency "Dogecoin".

The value of cryptocurrency exchange rates in points 1-3 can be seen in tab. 1.

In Fig. 3, 4 and tab. 1, we can observe that at the beginning of the study, the algorithms behave differently, but in the end they almost completely "merge", that is, they give results as close as possible to the actual values, and the points (1,2,3) mark the time after posts by Elon Musk on Twitter. And as we can see, the difference in the results of the algorithm at these points will be more significant, which proves the importance of researching similar situations with other famous people and other cryptocurrencies, since in addition to the research nature, this work has an applied nature - with the help of these forecasts, you can navigate in the future prices of cryptocurrencies and carry out relevant financial transactions.

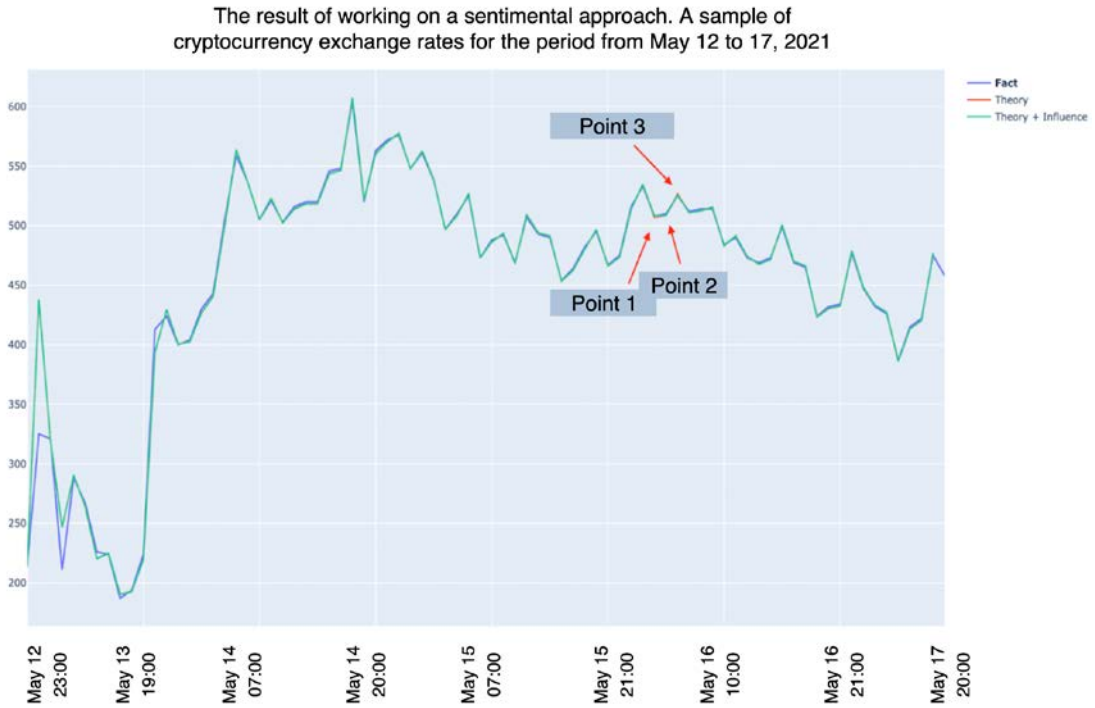

Figure 3. The result of the "standard prediction" algorithms and taking into account the impact of the publication

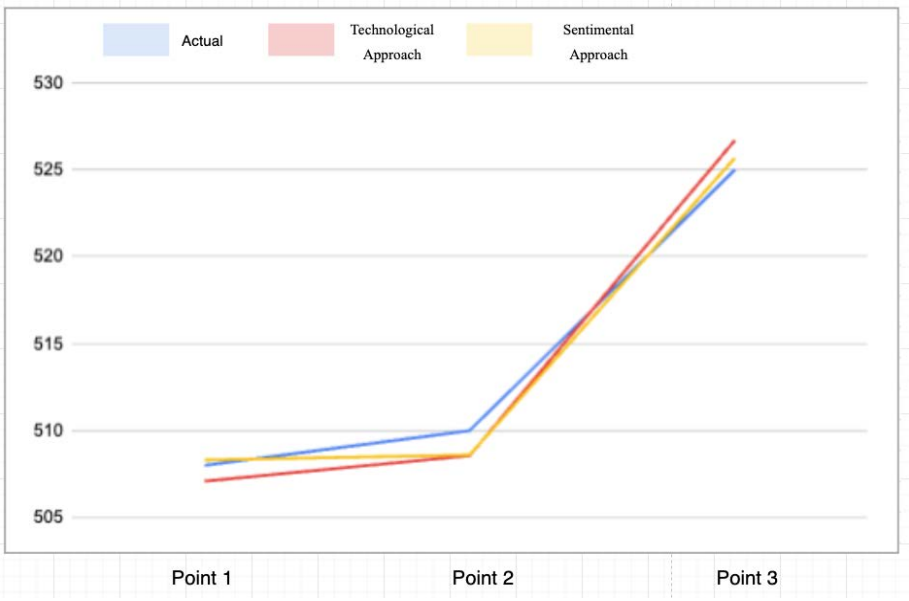

Figure 4. Comparison result of the prediction
for the “Technological Approach” and “Sentimental Approach”

Table 1. Points description for predicted value
by “Technological Approach” and “Sentimental Approach”

Name	Actual value	Predicted value by Technological Approach	Predicted value by Sentimental Approach
Point 1	508	507,09	508,31
Point 2	510	508,56	508,6
Point 3	525	526,69	525,67

Conclusions

The task of analysis of publications is formulated and defined, which is based on determining its direction and further forecasting of events in relation to these publications. It is clear that the task is complex and should be divided into several parts, such as: preliminary preparation of requirements for the required data, creating a relationship of behavioral color and interest, as well as determining the direction of the publication. The last step will be to predict the success of the publication, which will be to shape public opinion in the way necessary for the author. A review of similar publications on the topic. Their common shortcoming has been identified, which is eliminated through preliminary preparation of requirements. The decision was made and decided that for effective work it is necessary to

ISSN 1560-8956

implement the technology of content monitoring. The concept of calculating the impact of publications of famous people on changes in cryptocurrency exchange rates is presented.

REFERENCES

1. Preethi P. Temporal Sentiment Analysis and Causal Rules Extraction from Tweets for Event Prediction / P. Preethi, V. Uma, A. Kumar. // Procedia Computer Science. – 2015. – №48. – С. 84–89.

2. Gerber M. Predicting crime using Twitter and kernel density estimation / Matthew S. Gerber. // Decision Support Systems. – 2014. – №61. – С. 115–125.

3. Gavrilenko OV Research of tasks of estimation of objects in computer social networks by means of reliable algorithms [Text]: materials of the 8th international scientific and technical conference "Modern directions of development of information and communication technologies and means of management", Kharkiv, April 26-27, 2018 . / Gavrilenko OV, Kuptsova IV /. - 108 p. - p.67- 68

4. A. Culotta, B. Huberman, Towards detecting influenza epidemics by analyzing Twitter messages, Proceedings of the First Workshop on Social Media Analytics. // ACM, 2010, pp. 115–122.

5. M. Chau and J. Xu. Mining communities and their relationships in blogs: A study of online hate groups. // *Int. J. Hum.-Comput. Stud.*, 65(1):57--70, 2007.

6. Gavrilenko OV Review of methods of hybridization of intellectual information processing // Scientific Bulletin of Kherson State University, series "ECONOMIC SCIENCES", issue 8, part 3, p. 1, Kherson, 2014 / Gavrilenko OV,. - 108 p. - p.204- 207

7. A.Yu. Berko, J.P. Kis, VI Sukhoversky. Content monitoring system of news Internet resources. // SISN. 2011; Volume 699, number 1: cube. 13 - 20