Research on sentiment analysis of public opinion events based on human behavior dynamics

Shuai Wu[a], Tianshu Ren[a], Huan Xia[b], Xiuzhang Yang[a],*
[a]School of Information, Guizhou University of Finance and Economics, Guiyang 550025, Guizhou, China; [b]Guizhou Province E-commerce Big Data Marketing Engineering Research Center, Guizhou University of Finance and Economics, Guiyang 550025, Guizhou, China

ABSTRACT

With the rapid development of the Internet and social networks, how to analyze the behavior trajectory, emotional changes and relationship trends of online public opinion events have become an important research topic today. This article combines human behavior dynamics and sentiment analysis methods to study online public opinion events. This article uses Python to capture hot public opinion topics and comments on social platforms, and uses human dynamics to analyze public opinion events from time interval distribution and activity. Then this paper uses the maximum likelihood estimation method to evaluate the power exponent of its distribution, and finally uses BosonNLP and sentiment intensity to analyze the sentiment of the comment objects in the public opinion event. Experimental results show that the time interval of group public opinion events obeys a power-law distribution, and its activity is positively correlated with the power exponent. The sentiment analysis method of public opinion events based on human behavior dynamics performs well, and the dominant sentiment of the comment object is distributed with a power exponential. The number of likes and the sentiment value of follow-up comments effectively improve the results of sentiment analysis, and social platforms play an important role in the communication of sentiments.

Keywords: Public opinion events, sentiment analysis, human behavior dynamics, power law distribution

1. INTRODUCTION

With the rapid development of the Internet and social networks, network data has exploded. The variety of data, large amounts of information, complex semantics, and real-time nature have brought convenience and freedom to the public[1]. Netizens vent their emotions on the Internet, triggering the venting and resonance of social emotions. Such public opinion events have also caused harm. How to dig out the public's emotions and attitudes from these public opinion events, analyze public opinion trends, and discover potential social phenomena has become a hot topic of research today[2]. Human behavior is a complex phenomenon, and its research aims to understand the characteristics of its own behavior comprehensively and profoundly, expand the breadth and depth of human understanding of the world, and generate great value for real life and social development[3]. The traditional analysis of public opinion events is realized through machine learning or mathematical statistics methods. This paper proposes a sentiment analysis method of public opinion events based on human behavior dynamics. It analyzes public opinion events from another perspective and studies its internal laws.

2. RELATED RESEARCH WORK

2.1. Research on human behavioral dynamics

In 2005, Barabási published an article in Nature[4]. The study of task-based queuing theory models showed that the interval of human behavior is highly non-uniform, which is called obeying the power-law distribution, which is $P(k) \sim Cx^{-\alpha}$. New research direction of "behavioral dynamics". Subsequently, scientists[5] found through extensive empirical statistics that the power-law distribution characteristics widely exist in the time interval distribution of human behavior and spatial motion behavior. DEZSö et al.[6] verified that the time interval of web browsing obeys the power law distribution. Wang[7] found that the time intervals of the three online behaviors of blog posting, wiki modification and bookmark collection all showed power law distribution. Zeng et al.[8] used the human behavior dynamics method to

*472191973@qq.com

analyze the open-source software community from four aspects: activity, episodicity, time interval distribution, and long-range correlation. Song et al.[9] found that human travel distance and dwell time obey the power law distribution, which are two key characteristics that affect human space motion behavior. Song et al.[10] verified that the distribution of time intervals of human blogs and microblogs obeys $\alpha = 1.3$ and $\alpha = 2.0$ power law distributions. Goh et al.[11] proposed relevant indicators and verified that the length of the time interval and memory affect both paroxysmal activities in nature. Ni et al.[12] found that the travel distance distribution of humans has a certain effect on the spread of infectious diseases and the group infection rate.

2.2. Sentiment analysis

Sentiment Analysis aims to use natural language processing, text mining, and other technologies to identify subjective views of texts, analyze, process, reason, and classify texts with emotional colors, and determine the positive and negative tendencies of text emotions. Common methods include a method based on sentiment dictionary matching and a method based on machine learning.

In recent years, domestic and foreign scholars have done a lot of research and practice on sentiment analysis. Li et al.[13] elaborated a review of text sentiment analysis; Wang et al. used sentiment classification algorithms and LDA models to study the subject of competing corporate news texts ; Wu et al. proposed a semantic analysis algorithm for web financial texts, Construct financial sentiment dictionary and semantic rules to improve the F value, recall and precision of the classifier ; Zhao et al.[14] applied CNN-SVM sentiment analysis model to analyze the satisfaction of Haitao APP user reviews; An et al.[15] researched on public health emergencies based on social network sentiment pictures; Guo et al.[16] combined sentiment analysis and user influence to build a user influence model for sentiment analysis.

In the process of dissemination of Internet public opinion events, because users' browsing, forwarding, commenting, and like operations have emotional tendencies, the influence of public opinion events under different emotions should be different and may be consistent with certain human behaviors. Dynamics. In response to the above problems, this article uses a Python custom crawler to capture four hot public opinion events, their comments, and praise information, and proposes a sentiment analysis method of public opinion events based on human behavior dynamics to mine the propagation rules of online public opinion events; using BosonNLP And sentiment intensity to conduct sentiment analysis on comment objects in public opinion events, and to explore the impact of different emotional tendencies on users.

3. EMOTIONAL ANALYSIS OF PUBLIC BEHAVIOR EVENTS IN HUMAN BEHAVIOR DYNAMICS

3.1. Algorithm overall process

The overall technical route of this paper is shown in Figure 1. The sentiment analysis for social network public opinion events is mainly divided into the following four steps:

• Use Python and Scrapy technology to build a custom crawler to grab online public opinion events, and store relevant information such as comment content, comment time, and number of likes into a local MySQL database.

• Perform preprocessing on the captured text corpus, including data cleaning, stopword filtering, Chinese word segmentation, and part-of-speech tagging.

• Use TF-IDF to construct a dictionary of sentimental feature words, and obtain the sentimental feature vectors of each public opinion event and its review information.

• Public opinion event analysis mainly includes time interval analysis, activity analysis and sentiment analysis, and experiments are compared and visualized for different analysis algorithms.

3.2. Data acquisition and preprocessing

This article captures the four hot public opinion topics and related comment information of the Tianya community. The captured fields include the event title, release time, release author, clicks, etc., as well as the content of the comment, the user name of the comment, the time of the comment, Comments and likes.

Table 1 is an example of the data format of four hot public opinion events, which are the "Huawei Meng Wanzhou" incident, the "DG insulting China" incident, the "Chang'e-4" incident, and the "after 70 house buying" incident.
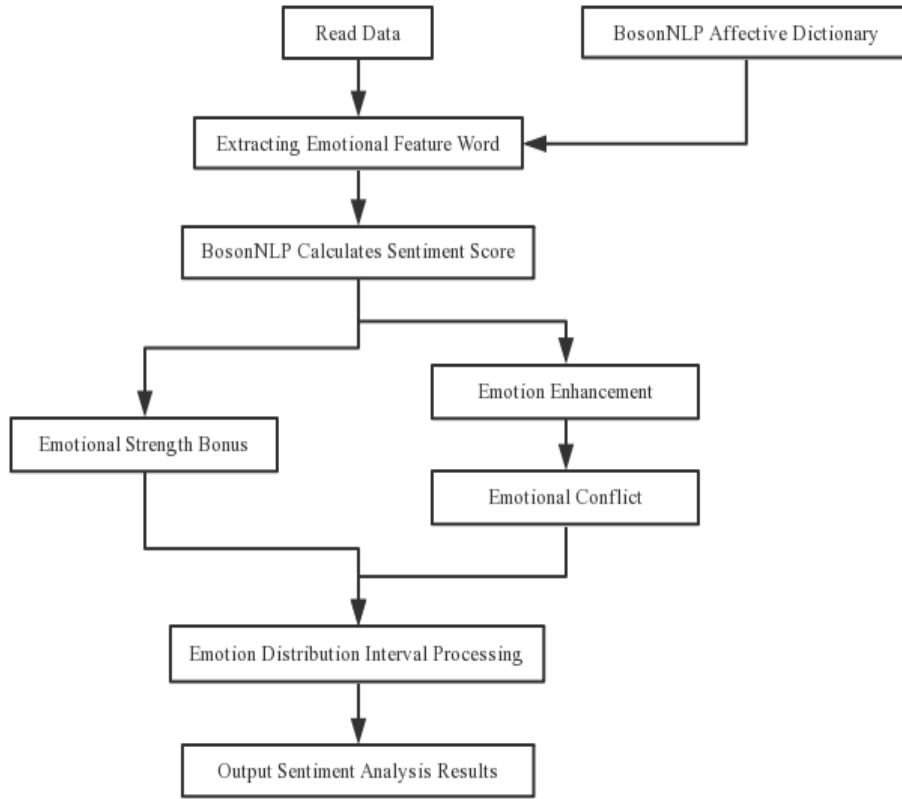
Figure 1. Road map of sentiment analysis for public opinion events based on human behavior dynamics.

Then the data is pre-processed, Chinese word segmentation and part-of-speech tagging is performed by Jieba word segmentation tool, and missing values and outliers are modified, and stop words and punctuation are removed. In this way, high-standard and high-quality data are obtained, and the experimental results are improved.

3.3. Sentiment analysis algorithm

The sentiment analysis algorithm flow of public opinion events is shown in Figure 2. This article uses the sentiment analysis algorithm based on sentiment enhancement and sentiment analysis to analyze the comment information of public opinion events to obtain the sentiment score of each review text.

Table 1. Example of online comment data format for online public opinion events.

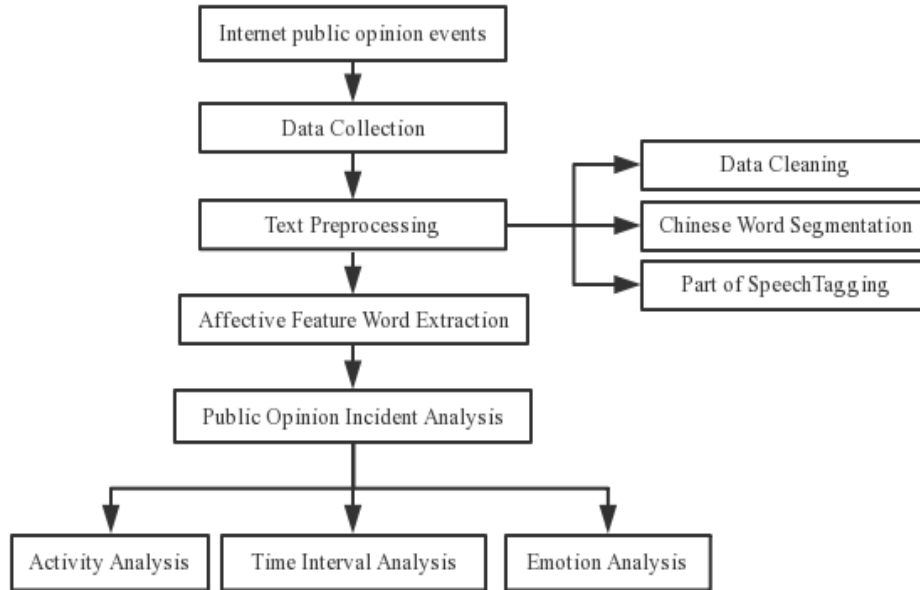| Public opinion event theme	Comment time	Review username	Comment Likes	Comments
"Huawei Meng Wanzhou" Incident	2018/12/9 14:59:13	Miss Bear 228	71	Get more frustrated! Huawei wins!
"DG insults China" Incident	2018/12/13 8:55:55	Happiness is calling	3	Being a strong Chinese can win the world's respect for the Chinese, which has nothing to do with being rich or poor.
"Chang'e Four" Incident	2019/1/3 17:10:50	Bi Di-pu	5	Congratulations to Chang'e Four for successfully landing on the back of the moon. Great, my country! Great!
"Buy a house after 70" Incident	2019/1/15 15:43:46	AA all the time	4	Life is very tiring, there is work to be done, and unsatisfactory desires.

Figure 2. Read the comment text after pre-processing of public opinion events.

- Read the comment text after pre-processing of public opinion events.

- BosonNLP Chinese Semantic Open Platform was used to extract sentiment feature words of public opinion event review text and calculate its sentiment score. Assume that the comment sentence of the public opinion event contains a series of emotional feature words $\{f_{i_1}, f_{i_2}, \cdots, f_{i_n}\}$, and each feature word corresponds to an emotional intensity containing an emotional word, a modified adverb, and a negative adverb, and recorded as $\{w_{i_1}, w_{i_2}, \cdots, w_{i_n}\}$. The calculation equation (1) of emotion score is as follows. Among them, represents the sentiment score of the i-th comment or, which is the sum of the scores of all sentimental feature words of the sentence.

$$Score_i = \sum_{j-1}^{n} w_{i_j} \times num_{i_j} \tag{1}$$

- Use the number of likes to add the emotional intensity. The calculation formula is as follows. $Score_i$ represents the emotional score of the i-th comment, $Weight_i$ represents the weight after the i-th comment is normalized by the like, and T_i represents the score after the emotional intensity is added. Its calculation equation (2) is as follows:

$$T_i = Score_i \times Weight_i \tag{2}$$

- Combined with the content being reviewed, the emotion score at the current moment is enhanced and conflicted. When using positive emotional texts to review, enhance the emotions and vice versa. As shown in equation (3), H_i represents the emotion score after being affected by the review.

$$H_i = \begin{cases} T_i + Score_{Positive} \ \text{Positive Emotional Review} \\ T_i + Score_{Negative} \ \text{Negative Emotional Review} \end{cases} \tag{3}$$

- Interval processing is performed on the emotion distribution, and the overall emotion score at the current moment is output.

4. EXPERIMENTAL RESULTS AND ANALYSIS

4.1. Time interval distribution

Time interval distribution is an important topic in the study of the temporal characteristics of human behavior, and an important parameter to characterize human behavior. Studying users' repetitive activities or participating in public opinion topics in daily life can find their inherent behavioral rules, which is of great significance for studying human behavior and public opinion events. The interval time in this article is defined as the time difference between two consecutive responses in a public opinion event. The time interval distribution of its four public opinion events is shown in Figure 3. The time interval distribution map of each public opinion event includes two parts. The abscissa in the lower part represents the time interval, and the ordinate represents the distribution of the response behavior. The blank area in the upper part represents the residence time of the individual reply at a certain time The black vertical lines indicate that the response behavior occurred at different times. The more black vertical lines, the more active the response event at that moment.

As can be seen from Figure 3, the time interval distribution of user responses in the four public opinion events is a power law distribution with obvious fat-tail characteristics, showing long-term silence and short-term burst characteristics in frequency. The results show that most of the public opinion events are low in activity and remain silent for a long time, and a small number of users are active and actively comment on public opinion events. Among them, Figure 3a shows the time interval distribution of the "Huawei Meng Wanzhou" event. The results of the figure confirm that the time interval distribution of user response behavior of public opinion events conforms to the power law distribution of the power index $\alpha \approx 1.94$. Calculated by the maximum likelihood estimation method; Figures 3b-3d represent the time of the "DG shame China" event, the "Chang'e 4" event, and the "after 70 house purchase" event, respectively Interval distribution, they all obey the power law distribution, and their power exponents are $\alpha \approx 2.00$, $\alpha \approx 1.73$, and $\alpha \approx 2.80$.
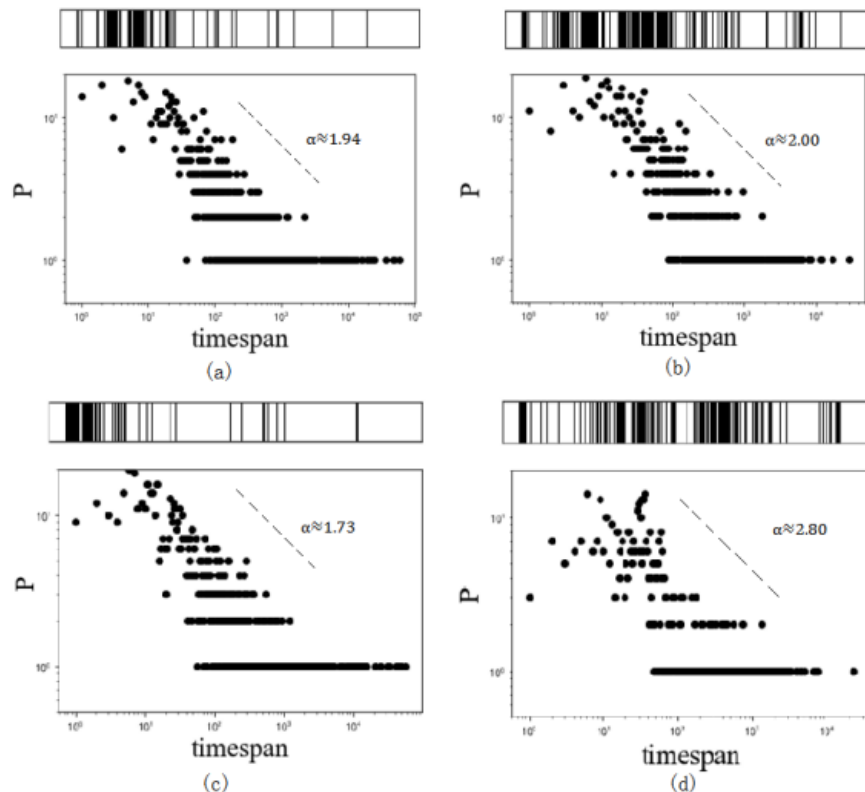

Figure 3. Time interval distribution of user reviews in public opinion events.

4.2. Activity analysis

Activeness is often used to evaluate user behavior, and its equation (4) is defined as follows:

$$A_i = \frac{n_i}{T_i} \tag{4}$$

Among them, i represents the user, n_i represents the total number of active times of the user, and T_i is the time difference between the first activity and the last activity of the user. This article analyzes the relationship between user comment time intervals and activity intensity for public opinion events, and obtains the distribution map shown in Figure 4. In the figure, the abscissa is the time interval between the user comment time and the outbreak of the public opinion event, the unit is hour; the ordinate is the number of comments at a moment after the outbreak of the public opinion event. It can be seen from Figure 4 that the user comment information of the four public opinion events is distributed in a power law, and the comments are concentrated in a certain period of time, and the comments are calm in another period of time.

4.3. Emotion analysis

This article uses the number of likes and follow-up sentiment values to perform sentiment enhancement and sentimental processing on sentiment analysis. In the process of counting the number of likes in public opinion event comments, we found that the frequency of the likes and the order of the position frequency showed a power law distribution, as shown in Figure 5.
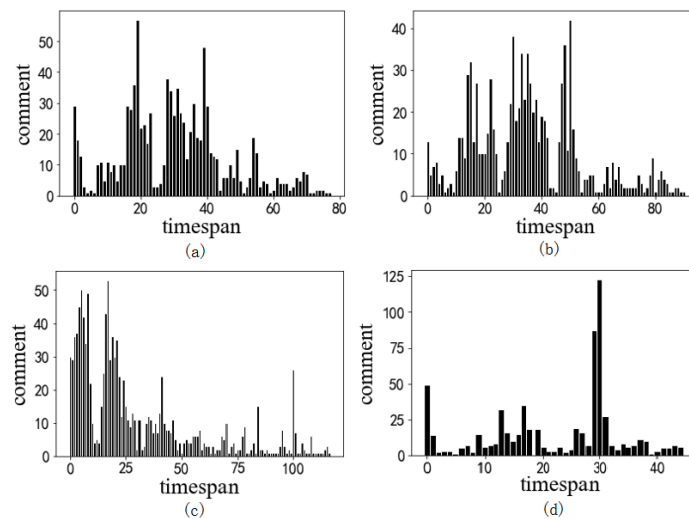

Figure 4. Analysis of the activity of user reviews in public opinion events.

In Figure 5, the abscissa represents the number of likes corresponding to the comments in the public opinion event, and the ordinate represents the frequency of the point likes appearing in the entire public opinion event. It can be seen from the figure that the number of likes of the four public opinion events has a clear fat tail distribution characteristic in double logarithmic coordinates. In Figure 5a, the number of comments from netizens in the "Huawei Meng Wanzhou" incident obeys the power law distribution of $\alpha \approx 2.80$. In Figure 5b, the number of likes from netizens' comments in the "DG Humiliating China" incident obeys a power law distribution of $\alpha \approx 3.06$. In Figure 5c, the number of "Chang'e 4" incident netizens commented on the number of likes following a power law distribution of $\alpha \approx 3.01$. In Figure 5d, the number of likes from netizens' comments in the "post-70s buying a house" event obeys a power law distribution of $\alpha \approx 5.10$. This result shows that the netizen's comment-like phenomenon in public opinion events obeys a power-law distribution. A small number of people are active and actively respond to public opinion events, while most of the netizens are low in activity.

Figure 6 is a result analysis diagram of comment information analysis in four public opinion events based on a sentiment analysis algorithm based on sentiment enhancement and sentiment collision. In the figure, the abscissa is the time interval between the user comment time and the outbreak of the public opinion event, the unit is hour; the ordinate is the overall emotional score at a certain moment after the outbreak of the public opinion event.
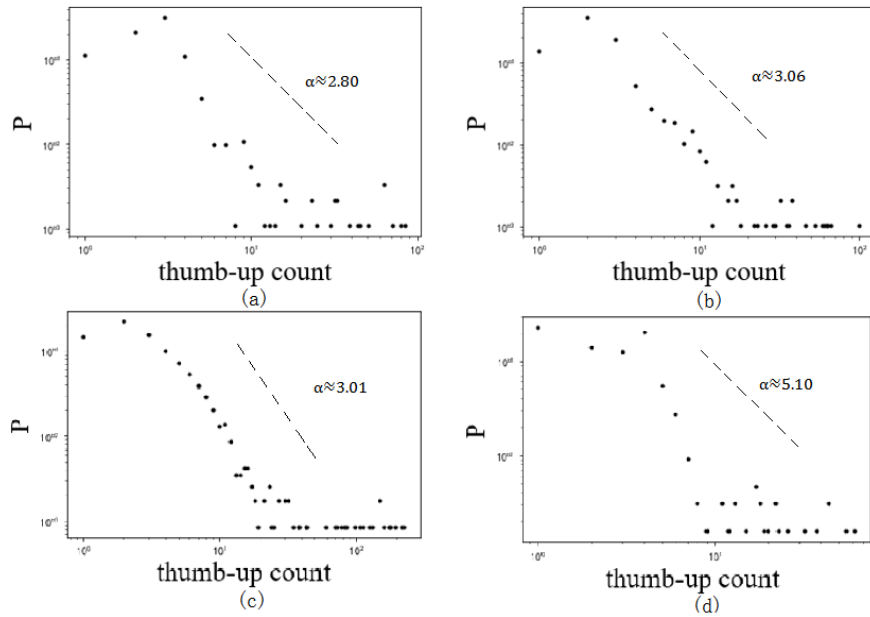
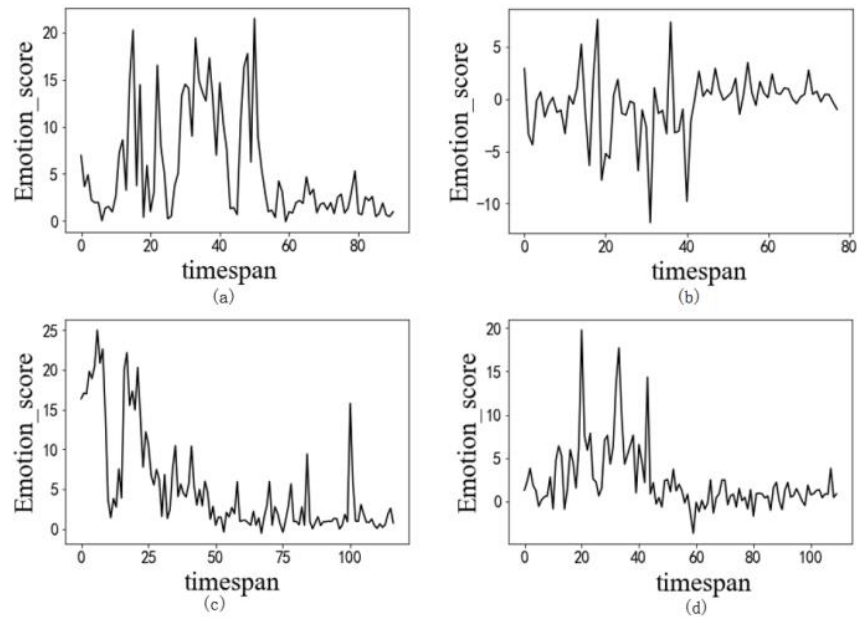Figure 5. Frequency distribution of comment likes.

Figure 6. Emotional time series of public opinion events.

Among them, Figure 6a is the emotional time series distribution diagram of the "Huawei Meng Wanzhou" event. The 15 hours, 32 hours, and 50 hours after the incident were the peak periods of user reviews. The comments on the event as a whole showed positive sentiment. Figure 6b is the emotional time series distribution diagram of the "DG humiliating China" event. The comments on the event have positive emotions and negative emotions. Among them, the negative comments were more active in the 30 hours and 40 hours after the incident. Figure 6c shows the sentiment analysis result of the "Chang'e-4" incident. The user's overall comments are expressed as positive emotions. Figure 6d shows the sentiment analysis result of the "post-70s buying a house" event. Except for the node 58-62 hours after the event, the overall user comments showed positive sentiment.

5. CONCLUSION

This article studies the Internet public opinion events through human behavior dynamics and sentiment analysis methods. Using the "Huawei Meng Wanzhou" incident, "DG insulting China" incident, "Chang'e 4" incident and "70s buying a house" incident as data sources, the comments, likes, and other information of each event were automatically collected through preprocessing. Time interval analysis, activity analysis, and sentiment analysis are used to comprehensively study the behavior rules and sentiment trends of public opinion events, and perform experimental comparisons and visualizations for different analysis algorithms.

The experimental results show that the time interval of the group public opinion events obeys the power law distribution. The user comment information of the four public opinion events is distributed in the power law. The comments are concentrated in a certain period of time and the comments are relatively calm in another period of time. The time interval distribution has an important effect. At the same time, the sentiment analysis method of public opinion events based on human behavior dynamics performs well. The dominant sentiment of the comment subject is exponentially distributed. The number of likes and follow-up sentiment values effectively promote the results of sentiment analysis. Social platforms have played a role in the transmission of sentiment. Important ties. The method of this paper has important theoretical significance and practical value. In subsequent research, the author will try to use deep learning, knowledge maps and other technical means to refine the granularity of sentiment analysis, combined with the semantic web to understand the sentiment of public opinion events, so as to better conduct public opinion warning and human behavior dynamics research.

REFERENCES

[1] He, J. M. and Li, X., "A hidden Markov model research in the microblog public opinion evolutionary analysis," Information Science, 34(4), 7-12 (2016).

[2] Liang, X. M. and Jian, X. U., "Sentiment analysis of objects in public opinion events and their relation network research," Information Science, 36(2), 37-42 (2018).

[3] Fan, C., Guo, J. L., Han, Y., et al., "A review of research on human dynamics," Complex Systems and Complexity Science, 8(2), 1-17 (2011).

[4] Barabási, A. L., "The origin of bursts and heavy tails in human dynamics," Nature, 435, 207-211 (2005).

[5] Yan, X. Y., "Empirical statistics on individual human travel behavior," Dianzi Keji Daxue Xuebao/Journal of the University of Electronic Science and Technology of China, 40(2), 168-173 (2011).

[6] Dezsö, Z., Almaas, E., Lukács, A., et al., "Dynamics of information access on the web," Physical Review E, 73(6), 066132 (2006).

[7] Zeng, J. Q., Yang, J. M. and Chen, Q., "Behavior of individual differences in human knowledge creation in the open source software community," Mathematics in Practice & Theory, 46(13), 1-13 (2016).

[8] Song, C. M., Koren, T., Wang, P., et al., "Modelling the scaling properties of human mobility," Nature Physics, 6(10), 818-823 (2010).

[9] Goh, K. I. and Barabasi, A. L., "Burstiness and memory in complex systems," Europhysics Letters, 81, 48002-48006 (2008).

[10] Ni, S. J. and Weng, W. G., "Impact of travel patterns on epidemic dynamics in heterogeneous spatial metapopulation networks," Physical Review E (Statistical, Nonlinear, and Soft Matter Physics), 79(1), 16111-0 (2009).

[11] Chen, L., Guan, Z. Y., He, J. H., et al., "A survey on sentiment classification," Journal of Computer Research & Development, 54(6), 1150-1170 (2017).

[12] Hakak, N. M., Mohd, M., Kirmani, M., et al., "Emotion analysis: A survey," 2017 International Conference on Computer, Communications and Electronics (Comptelix), IEEE, (2017).

[13] Wang, S. Y., Liao, H. T. and Wu, C. K., "Mining news on competitors with sentiment classification," Data Analysis and Knowledge Discovery, 2(03), 70-78 (2018).

[14] Wu, J., Tang C J, Li, T. Y., et al., "Sentiment analysis on web financial text based on semantic rules," Journal of Computer Applications, 34(2), 481-485 (2014).

[15] Zhao, Y., Li, Q. Q., Chen, Y. H., et al., "Examining consumer reviews of overseas shopping app with sentiment analysis," Data Analysis and Knowledge Discovery, 2(11), 19-27 (2018).

[16] Yu, D. M., Han, X. X., Li, D., et al., "Research on computable model of emotional interaction based on q-learning algorithm," Computer Engineering, 38(10), 277-279 (2012).