

# Microblog Public Opinion Communication Analysis Based on Data Mining

Xiang Li\*  
1286917369@qq.com

School of Management Shanghai University of Engineering Science, Shanghai, China

**Abstract**—With the rapid development of the Internet, most users are gradually accustomed to online social networking, so social platforms have become an important way of information exchange and dissemination. The unrest, loss and impact of secondary disasters transmitted by unconventional emergencies are incalculable. Sina Weibo is a social platform widely used by many users. The platform has a large number of users. Research on the use habits and information dissemination channels of microblog users is of great significance for the early warning and intervention of network public opinion in unconventional emergencies. This paper mainly studies the information dissemination of unconventional events from the microblog text analysis, network public opinion characteristics and time series analysis. It is found that the number of microblog posts significantly affects the number of fans, the number of fans significantly affects the number of followers, and the number of followers significantly affects the total number of Weibo retweets, and the time series evolution law of "Xinjiang cotton" predicted by SPSS is consistent with the reality.

**Keywords**- Data mining; text analysis; time series analysis; weibo public opinion;

## 1 Introduction

With the increasing social networking of modern human activities, the turmoil, loss and impact of secondary disasters transmitted by unconventional emergencies are incalculable. For example, due to the social networking of modern human activities, the movement of people and goods has become more frequent and convenient, which has greatly accelerated or expanded the epidemic of infectious diseases, as well as the spread of rumors and panic derived from this life.[1] The emergency management of secondary disasters in non-conventional emergencies constitutes an important part and research content of emergency management of non-conventional emergencies.

Based on the perspective of user behavior, Cha uses Twitter as a platform to analyze the characteristics of public opinion information, and compares and studies three indicators: the number of fans, the number of mentions and the number of retweets; pal used data on Twitter to calculate the influence of individuals and diffusion influence.[2] Public opinion dissemination in the network environment has obvious immediacy, and with the popularization and upgrading of the Internet and the innovation of smart terminals, Weibo has further strengthened this feature.[3] Weibo users can post information and express opinions on platforms or devices such as computers, mobile phones, tablets, and smart terminals, while largely not being limited by time and geography. At the same time, with the help of the Internet, a network public opinion

can be quickly formed in a short time, and an emotional opinion or hot event may trigger a public opinion. Various public opinion information through various channels for rapid interaction, mutual influence, and finally become a huge momentum.

This paper uses data mining technology to collect public opinion text information on the topic of "Xinjiang cotton" on Sina Weibo, uses SPSS PRO to deeply analyze the relationship between the number of Weibo posts and the number of fans, the number of followers and the total number of Weibo comments, and uses the quadratic exponential smoothing method to analyze the time series of the topic, summarizes the communication characteristics of the evolution of public opinion and gives corresponding intervention suggestions, which has certain practical guiding significance.[4]

## 2 Materials and Methods

Unconventional emergencies network public opinion mainly refers to the discussion and reporting of the event by netizens and online media, and a series of emotions, cognitions and evaluations generated by netizens and online media after an unconventional emergency occurs. Use crawler technology to obtain data such as clicks, comments and posts of the "Xinjiang Cotton Incident", an unconventional emergency that occurred in 2021, from Weibo. At the same time, the rapid increase in the number of users of social networks has made the amount of data information continue to expand and show explosive growth. The questions that need to be addressed in this article are: Use web crawling technology to crawl the data such as clicks, comments and posts of the unconventional emergency event of "Xinjiang cotton". Conduct descriptive statistical analysis and exploratory data analysis on the characteristics and influencing factors of the spread of the "Xinjiang cotton" incident. Build a "Xinjiang cotton" event transmission network, identify key users, and conduct time series evolution analysis.

Using network mining technology, the text information about "Xinjiang cotton" was extracted on Sina Weibo, based on 85301 original data, this study cleaned it through Python tools, and after modifying and deleting some incorrect formats, the sorted 84590 text data were uniformly stored as xlsx tables. Then, SPSS PRO was used to perform descriptive analysis and multiple regression analysis on the cleaned data, which clearly and clearly saw the relationship between the number of Weibo posts and the number of followers, the number of followers, and the total number of Weibo retweets. Secondly, based on the quadratic exponential smoothing method, the extracted number of microblog posts and forwards is analyzed by time series evolution, and the communication characteristics of public opinion evolution on microblog are summarized.

A time series is a chronological sequence of interrelated random variables whose observations need to be at the same time interval. Time series analysis is to find all the potentially valid information in a set of time series on the basis of time series, find out the statistical laws of this set of data in the process of changing over time through observation, analysis and forecasting, and understand the system to be studied by revealing the laws, and then predict and control the events that have not yet occurred in advance, and ultimately achieve the purpose of improving the level of business decision-making.[5] A set of random variables that are usually arranged in chronological order is called a time series and is denoted  $\{X_t\}$ . To know the nature of this time series, it is necessary to analyze the observations  $X_1, X_2, X_3, \dots$  the nature of  $X_n$ , inferred from the part of the population. Time series analysis methods can be divided into two modules:

descriptive time series analysis and statistical time series analysis. Descriptive time series analysis generally refers to intuitive data comparisons such as mean, variance, etc., or the degree of trend dispersion in a plotted observation set. Usually this method is simple and easy to understand, and you don't need to have strong professional knowledge to find the hidden rules, and it is still one of the common methods in people's daily life.

### 3 Results & Discussion

In the era of big data, data mining and collection is the foundation and important part of doing a good job in research. After obtaining the data of clicks, comments and releases of the unconventional emergency event of "Xinjiang cotton" crawled by web crawler technology, python was used to preprocess the text of the crawled data, remove duplicate and missing data, use SPSS PRO to conduct descriptive statistical analysis of the characteristics and influencing factors of network public opinion transmission, and analyze the number of microblog posts, fans, followers, total number of Weibo retweets and other factors and their relationships. For the analysis of the characteristics of public opinion propagation, based on the data on the topic of "Xinjiang cotton" crawled by crawling technology, the time series analysis of microblog public opinion is carried out by using the quadratic exponential smoothing algorithm to obtain the curve of topic propagation volume on time, and finally the characteristics and laws of public opinion propagation are analyzed based on this curve. The following equation is the built model.

$$RET = \mu_0 + \mu_1 FAN + \mu_2 FOL + \mu_3 PUB + e \quad (1)$$

where  $\mu_0$  is a constant term;  $\mu_i$  ( $i=1,2,3$ ) is the regression coefficient in the model;  $e$  is the residual term. RET stands for Weibo retweet ratings; FAN stands for the number of fans; FOL stands for number of concerns; PUB stands for Weibo posts.

This article will use exponential smoothing for time series analysis, and the following is the code to implement exponential smoothing

Algorithm	Quadratic exponential smoothing
1.	<code>import numpy as np</code>
2.	<code>def secondaryExponentialSmoothingMethod(list,n_average,alpha,day):</code>
3.	<code>def fangChengZu(a1,b1,a2,b2,c1,c2):</code>
4.	<code>    a = np.array([[a1,b1],[a2,b2]])</code>
5.	<code>    b = np.array([c1,c2])</code>
6.	<code>    x,y = np.linalg.solve(a,b)</code>
7.	<code>    return x,y</code>
8.	<code>list_left = list[0:n_average]</code>
9.	<code>list_right = list[n_average+1:len(list)]</code>
10.	<code>list_left_average = np.mean(list_left)</code>
11.	<code>list_right_average = np.mean(list_right)</code>

---

```

12.     x1 = (n_average+1)/2
13.     x2 = (len(list) - x1) + 1
14.     a1,b1 =
        fangChengZu(1,x1,1,x2,list_left_average,list_right_average)
15.     S11, S12 = fangChengZu(2,-1,a1,b1,-b1,(alpha/(1-alpha)))
16.     b_tao = 0
17.     for i in range(len(list)):
18.         S1 = alpha*list[i] + (1-alpha)*S11
19.         S2 = alpha*S1 + (1-alpha)*S12
20.         S11 = S1
21.         S12 = S2
22.         a_tao = 2*S1-S2
23.         b_tao = (alpha/(1-alpha))/(S1-S2)
24.     H = a_tao + b_tao * day
25.     return H
26. if __name__ == '__main__':
27.     data = [0, 1, 2, 3, 4, 5, 6, 7]
28.     prediction_day1 =
        secondaryExponentialSmoothingMethod(data, 3, 0.5, 1)
29.     prediction_day2 =
        secondaryExponentialSmoothingMethod(data, 3, 0.5, 2)
30.     print(prediction_day1,prediction_day2)

```

---

### 3.1 Descriptive Statistical Analysis

In the process of network dissemination of unconventional emergencies, the public obtains events, selects event information through traditional media and network media, and disseminates and shares event information through the network, on this basis, forming cognition, emotion, attitude and evaluation about events, that is, network public opinion about unconventional emergencies. On this basis, we will further use the various communication methods of the Internet to spread information with cognition, emotion, attitude and evaluation again to form new network public opinion. Table 1 is a descriptive statistical analysis of the characteristics and influencing factors of network public opinion transmission.

**Table 1** Descriptive statistics of variables

Descriptive statistics	Describe the results overall					
	N	Min	Max	Mean	Standard Deviation	Variance
Number of blogs	84590	1	20	8.86	4.993	24.931
Number of forwards	84590	0	169657	13.25	939.680	882999.318

Number of comments	84590	0	19138	5.83	157.460	24793.591
Number of likes	84590	0	724592	73.68	4514.838	20383765.012

Among them, the maximum and minimum values of the number of blogs are 1 and 20 respectively, indicating that the difference in the number of blogs is not large. The mean value of the number of forwards is 13.25, and the standard deviation is 939.680, indicating that the difference in the number of forwards is relatively large. The minimum value of the number of comments is 0 and the maximum value is 19138, and the difference in the number of comments is very large, as is the number of likes.

### 3.2 Regression Analysis Results

In order to deeply explore the relationship between the number of posts, retweets, comments and likes, the model in this paper is profoundly analyzed.

**Table 2** Regression analysis results

Regression analysis	Regression analysis metrics						
	Beta	t	Saliency	Lower limit	Upper limit	Tolerance	VIF
Constant	\	495.276	.000	8.825	8.896	\	\
Number of forwards	.004	.918	.049	-.014	.085	.593	1.687
Number of comments	.001	.167	.007	-.066	.096	.281	3.563
Number of likes	.002	.376	.017	.110	.073	.354	2.826

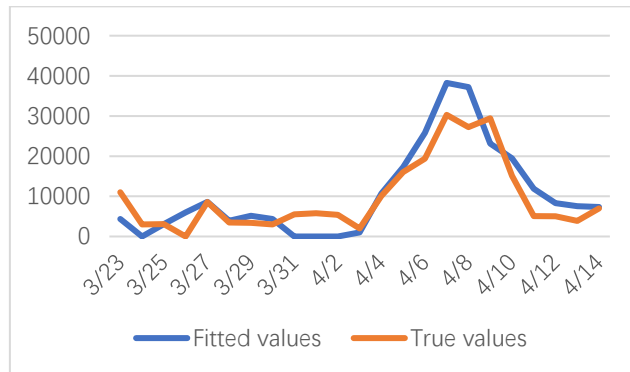
According to the results of regression analysis in Table 2, the number of Weibo posts significantly affects the number of fans, the number of fans significantly affects the number of followers, and the number of followers significantly affects the total number of Weibo retweets. At the same time, we can see the significance at the level of 1% and 5%, respectively. In addition, the influence of the event has increased→ the level of netizens' attention to the event has increased→ the number of microblogs has increased→ the participation of netizens has increased→ the popularity of online public opinion on public emergencies has increased→ the influence of the event has increased. With the increase of the influence of the event, it will attract more attention from netizens, netizens express their views and opinions through posting and other behaviors, and the interactive process of views will increase the degree of participation of netizens in this event, thereby promoting the further increase in the popularity of Internet public opinion.

### 3.3 Time Series Analysis

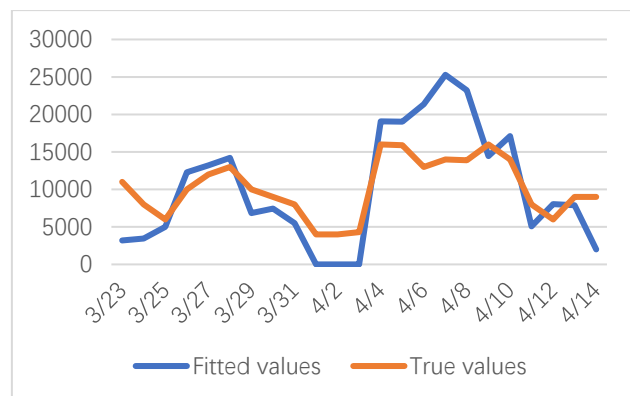
Establish an emergency management mechanism for unconventional emergencies from the

perspective of network public opinion to guide network public opinion to first carry out network public opinion monitoring, and establish a network public opinion monitoring mechanism, which is also a necessary premise and preparation for responding to network crises. Improve the technology and quality of network public opinion monitoring, and fully grasp the operational characteristics of network public opinion. Establish an information management system from the perspective of network public opinion, and do a good job in information collection, analysis and reporting.[6] Network public opinion monitoring includes continuous dynamic monitoring, collection, comparison and measurement, discovering hot public opinion and unconventional emergencies that cause public opinion, forming public opinion reports, and predicting the impact that public opinion may have on society as accurately as possible.[7]

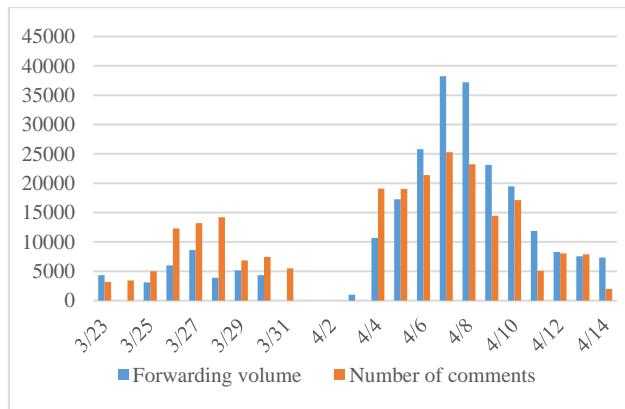
According to the curve change law, we selected the data from March 23 to April 14, 2021 on the day of the "Xinjiang cotton" incident for analysis, based on the data after cleaning, taking the sum of the number of Weibo retweets and comments as the basis for topic dissemination, and taking time as the basis for the visual analysis of the spread of the "Xinjiang cotton" incident, as shown in Figures 1, 2 and 3.



**Figure 1** Dissemination curve of "Xinjiang cotton" topic - number of comments forwarded



**Figure 2** Number of microblogs posted on the topic of "Xinjiang cotton"



**Figure 3** Visualization and timing curve of "Xinjiang cotton" topic propagation

From the topic communication visualization and timing curve of the above two figures, it can be seen that in the middle and late stages of public opinion communication, the number of microblog releases and the amount of dissemination are showing a decreasing trend, after the number of microblogs and the volume of dissemination reached the highest on March 27, the popularity of public opinion gradually decreased, and after reaching a new high heat on April 4, the topic of "Xinjiang cotton" tended to subside in mid-April, which shows the attention of netizens to the "Xinjiang cotton" incident.

When we conduct time series analysis, we select appropriate time intervals and make the time intervals as continuous as possible in order to better analyze the time transmission characteristics of public opinion. Weibo-based communication presents the characteristics of a circle-centric group, and Weibo users post most of the content they see and feel in their lives. With the continuous gathering and differentiation of Weibo audiences, users pay more attention to friends, classmates, colleagues, peers, etc. around them, and the information they receive is mainly topics with circle characteristics, the life dynamics of people around them, and the views of industry insiders. Under the influence of the autonomy of microblog user information release, opinion leaders can guide a large number of microblog user groups, so that microblog users and microblog content show obvious aggregation and group characteristics. At the same time, Weibo based on the "follow-fan" structure has the characteristics of multi-layer nested "fan" groups, which can gather more people through interest and effectively expand the interpersonal circle.

## 4 Conclusions

In the traditional mode of communication, the media is at the center, and for a long time the audience did not have the right to choose and was accustomed to this non-reciprocal relationship with the media.[8] However, the emergence of microblogs has broken through the traditional "top-down" form of information transmission, and netizens are not only the recipients of information, but also the publishers and disseminators of information, stimulating the public's enthusiasm for creation and desire for expression. Weibo-based public opinion dissemination not only reduces the time cost, but also completely changes the way of information dissemination, and even changes people's lifestyles and interpersonal communication patterns.

Weibo-based microblog public opinion dissemination strengthens the concept of equality between information publishers and recipients, and shows the characteristics of decentralization with considerable affinity.[9]

Weibo users are huge and complex, and the threshold is low, and users' information literacy and information screening ability are different. In the process of microblog public opinion dissemination, users are both information recipients and information publishers, which brings convenience to public opinion dissemination but also buries hidden dangers. In short, in the case of the outbreak of unconventional emergencies, the government should first improve its understanding and scientifically treat network public opinion for the management of network public opinion.[10] Secondly, we should pay attention to the dredging and guidance of Internet public opinion, take the initiative to listen to the voice of the people through the Internet, conduct equal dialogue with netizens, and treat netizens sincerely; finally, it is necessary to strengthen the coordination among all government departments, speak out in a unified manner, and strengthen the ability to respond to non-routine emergencies at the network level.[11]

## References

- [1] Y. Zhang, F. Long and L. Bin, "Identifying Opinion Sentences and Opinion Holders in Internet Public Opinion," 2012 International Conference on Industrial Control and Electronics Engineering, 2012, pp. 1668-1671.
- [2] M. Tang, "Design of Visual Model and Solutions for Public Opinion Monitoring and Analysis for Big Data," 2021 Asia-Pacific Conference on Communications Technology and Computer Science (ACCTCS), 2021, pp. 295-301.
- [3] Jianfeng Tang, Xiang Xu. Research on Detection of Chinese Microblog Public Opinion Analysis System[C]//Proceedings of the 11th International Conference on Computer Engineering and Networks(CENet2021)Part I.,2021:775-782.
- [4] Lu Chen, Yang Liu, Yudong Chang, Xinzhi Wang, Xiangfeng Luo. Public Opinion Analysis of Novel Coronavirus from Online Data[J]. Journal of Safety Science and Resilience, 2020, 1(2).
- [5] Yanxia Yang. Research and realization of internet public opinion analysis based on improved TF-IDF algorithm[C]//Proceedings of 2017 16th International Symposium on Distributed Computing and Applications to Business, Engineering and Science (DCABES 2017), 2017:89-92.
- [6] G. Zhan, M. Wang and M. Zhan, "Public Opinion Detection in an Online Lending Forum: Sentiment Analysis and Data Visualization," 2020 IEEE 5th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA), 2020, pp. 211-213.
- [7] Yixuan Wang. Public Opinion Analysis of Machine Learning based on Big Data technology[C]//Proceedings of 6th International Symposium on Social Science (ISSS 2020), 2020:70-76.
- [8] X. Tan, M. Zhuang, X. Lu and T. Mao, "An Analysis of the Emotional Evolution of Large-Scale Internet Public Opinion Events Based on the BERT-LDA Hybrid Model," in IEEE Access, vol. 9, pp. 15860-15871, 2021.
- [9] Dai Lili, Shi Lei, Xie Gang. Public opinion analysis of complex network information of local similarity clustering based on intelligent fuzzy system [J]. Journal of Intelligent & Fuzzy Systems, 2020, 39(2).



[10] Jianfeng Tang, Xiang Xu. Research on Detection of Chinese Microblog Public Opinion Analysis System[C]//Proceedings of the 11th International Conference on Computer Engineering and Networks(CENet2021)Part I.,2021:775-782.

[11] Ye Pinghao, Liu Liqiong, Tan Joseph. Influencing Factors on College Students' Willingness to Spread Internet Public Opinion: Analysis Based on COVID-19 Data in China [J]. Frontiers in Public Health, 2022,10.