Research on Microblog Public Opinion Event Management Based on Logit Regression

Shiyuan Zhang[1], Baiping Chen[2](✉), Yajuan Deng[1], and Xiaohong Deng[3]

[1] Department of Business Administration, Sichuan University, Chengdu, China
[2] Innovation Institute, Science and Technology for Development Research Center of Sichuan Province, Chengdu, China
251458382@qq.com
[3] Agricultural Bank of China Tianfu New District, Chengdu, China

Abstract. This paper is based on the data of social public opinion events that occurred on Weibo within 24 h of a certain day in 2017. Through the establishment of Logit model to explore the influencing factors that affect the popularity of online public opinion. The empirical results show that the number of fans and likes has a significant positive effect on the popularity of online public opinion. When the number of fans of a platform that publishes social public opinion events increases, the social public opinion events released by the platform are easier to warn. The more likes, the more netizens pay attention to the social public opinion event, and the content of the event can attract enough attention, and the public is more interested in the event. Appropriate control of the number of fans and likes on large-scale Weibo platforms has important implications for government supervision and protection of the healthy development of online public opinion.

Keywords: Public opinion · government supervision · Logit

1 Introduction

According to the statistical report on China's Internet Development released by CNNIC, it shows the development of China's Internet in many aspects. By June 2020, the number of mobile Internet users in China had reached 932 million, an increase of 35.46 million over March 2020, and the proportion of network names using mobile Internet had reached 99.2%. The public opinion analysis briefing of the top ten hot events in the third quarter of 2020 shows that the four hot events broke out with microblog as the source. For example, Dushan County, Guizhou Province has a high debt to do image engineering, Weiya and Li Ziqi have been elected members of the all China Youth Federation, and CCTV has criticized the "big stomach king" for serious waste of food and broadcasting. The event was released and amplified through the microblog network platform, and the information spread in a fission mode. The depth and breadth of information dissemination are enough to change the trend of network public opinion [7].

Nowadays, the sudden risk of network public opinion is higher and higher, and once the network public opinion is formed, similar news reports will emerge in an endless

A. Luqman et al. (Eds.): PMIS 2022, AHCS 6, pp. 323–332, 2023.
https://doi.org/10.2991/978-94-6463-016-9_34

stream and ferment rapidly in a short time. The rapid development of microblog public opinion not only increases the pressure of the government to deal with the crisis, but also increases the public opinion risk of the media, which makes the prediction model of microblog public opinion evolution become a subject of widespread concern by the government and academia [1]. The government can do a good job in monitoring network public opinion information. Due to the wide dissemination channels of network information and the diversification of public opinion risk sources, in order to ensure the comprehensiveness of information monitoring, the network public opinion monitoring system such as Yifang software can automatically carry out all-weather real-time monitoring on the whole network platform, so as to improve the discovery rate of network public opinion information, facilitate the early discovery of public opinion, deal with it as soon as possible and avoid potential public opinion risks [8]. Government departments and some industry workers can find hot issues of concern to the people through big data processing and analysis. After that, relevant departments can take effective prediction methods to predict the development trend of network public opinion events, find potential public opinion crises, and scientifically and quickly take appropriate measures to guide the correct development of events, which is of great practical significance [9].

The situation of network public opinion has had a very important impact on social stability and political order. This paper collects relevant data of social public opinion events on microblog within 24 h of a day in 2017. Based on the two dimensions of fans and praise, this paper analyzes the influencing factors of network public opinion heat by establishing logit binary regression model, hoping to find the key factors affecting the network epidemic to help the government supervise the network public opinion and ensure the healthy development of network public opinion [2].

2 Data Collection

2.1 Establish Assumptions

2.1.1 Hypothesis of Whether the Number of Fans has an Impact on Online Public Opinion Early Warning

Microblog has the function of "+ attention". The number of fans refers to the number of other users who pay attention to the microblog user. Fans are the direct recipients of posting media. Fans' reflection and participation in information directly affect the development trend of online public opinion. Peng zhihui believed that the netizen force played an important role in the evaluation system of network public opinion heat of unconventional emergencies [5]; Li Ming constructed the netizen subsystem module based on the system dynamics model and studied the emergency strategy of network public opinion crisis in dangerous chemical water pollution emergencies [3]. In conclusion, the following assumptions are put forward:

H1: after controlling other factors, the number of fans has a positive impact on whether the online public opinion early warning exists.

2.1.2 Hypothesis of the Influence of Praise Quantity on Network Public Opinion Early Warning

Microblog users can like other users' blogs. If you need to like a blog post, the premise is that you need to click it. When you click it, we may know the main content of the blog post, which also increases the browsing and forwarding volume of the blog post to a certain extent. To sum up, the following assumptions are put forward:

H2: after controlling other factors, the number of likes has a positive impact on whether the network public opinion early warning exists.

2.2 Data Sources

The data source of this paper is the data of the main relevant variables of social public opinion events on microblog within 24 h to estimate the parameters of the model.

2.3 Model Selection and Variable Design

2.3.1 Model Selection

According to the types of dependent variables and the characteristics of the two models, this paper selects logit binary regression model. The logistic model uses logistic regression and the maximum likelihood estimation method to fit the sample data, so as to estimate the value of each parameter. After the regression calculation of the final model, the probability of whether public opinion early warning can be obtained. Logit model is an effective method to solve the 0–1 regression problem. At this point, we assume that probability π_i Logit transformation (not probability) π_i Itself) obeys the linear model, i.e.

$$\text{logit}(\pi_i) = ln\left(\frac{\pi_i}{1-\pi_i}\right) = x_i{}'\beta \tag{1}$$

Among them, x_i construct vectors for explanatory variables, β is the coefficient vector.

Since the logit transformation is one-to-one correspondence, we can obtain the probability value from the logit inverse by calculating the inverse logarithm (commonly known as antilogit). It can be solved by Eq. (1)

$$\pi(x_i) = \frac{\exp(x_i{}'\beta)}{1+\exp(x_i{}'\beta)}$$

Further, the explanatory variables are expressed as:

$$y_i = \pi(x_i) + \varepsilon_i$$

Among them, ε_i it is a random interference term with two possible values. If $y_i = 1$, then $\varepsilon_i = 1 - \pi(x_i)$, the corresponding probability is π_i; if $y_i = 0$, then $\varepsilon_i = -\pi(x_i)$, the corresponding probability is $1 - \pi(x_i)$.

2.3.2 Variable Design

2.3.2.1 Dependent Variable

Whether public opinion warning (1 = yes, 0 = no). According to Liu Jianming of Tsinghua University, "if 61.8% of people hold a certain opinion within a certain range, such opinion will form a dominant public opinion within this range". Therefore, taking 61.8% as the segmentation point, this paper divides microblog public opinion into two categories. 0 is a general public opinion event. When the popularity of public opinion prediction is insufficient, you can choose to ignore it.1 is a typical warning event. When the public opinion exceeds 61.8%, the relevant departments shall strictly monitor the trend of the public opinion and control the negative public opinion in the incubation period. This paper takes the number of times a microblog is clicked as the reference index of microblog public opinion, and then determines whether the public opinion event needs early warning.

2.3.2.2 Independent Variable

X1 blogger identity: blogger identity refers to the hierarchical identity made by Sina Weibo according to the activity and influence of users. It is divided into four levels: ordinary users, members, microblog experts and + v authentication. In order to facilitate the data collection of microblog initial information, this paper does not consider the status authentication such as microblog obtaining headline qualification by paying membership fees.

X2 number of fans: microblog has the function of "+ attention". The number of fans refers to the number of other users who pay attention to the microblog user.

X3 attention: attention refers to the number of microblog users paying attention to other users. The trends of microblog users who are concerned will be directly pushed to the fans' home page for browsing.

X4 number of historical posts: the number of historical Posts refers to the number of Posts issued by the microblog user since registration.

X5 topic length: the topic is composed of topics, the essay is ## enclosed, and the microblog information is classified according to the topic. Topic length refers to the number of characters of the topic.

X6 blog post length: blog post length refers to the number of characters in the microblog body.

X7 number of pictures: microblog users have the function of distributing pictures, and can distribute up to 9 pictures. The number of pictures refers to the number of pictures contained in microblog.

X8 includes small videos: microblog has the function of distributing 9s small videos.

X9 includes links: microblog has the function of web link forwarding of other platforms. Generally speaking, when there is a link to a web page under the microblog platform, the possibility of browsing the web page will also increase.

X10 forwarding: microblog users can forward other users' microblogs, pictures, videos and links. When microblog users publish an event, if multiple people forward it, it can expand the influence of the situation and increase the number of views of the event.

X11 comments: microblog users have the function of commenting on other users' blogs. The more comments, the more people can discuss the content of this blog post, and the more likely this blog post will be read by more people.

X12 like: microblog users can like other users' blogs. If you need to praise a blog post, the premise is that you need to click it. When you click it, we may be able to know the main content of the blog post and increase the browsing volume of the blog post to a certain extent.

3 Data Analysis

3.1 Descriptive Statistics and Variable Correlation Analysis

3.1.1 Descriptive Statistics

Descriptive statistical analysis was conducted on the original data of each variable, and the results are as follows.

From Table 1, we can see that the correlation coefficients between the variables do not exceed 0.75. Therefore, there is no problem of multicollinearity between the variables and can be incorporated into the model for analysis.

Table 1. Descriptive statistics of variables

	yuqin identity		ln_fans	ln_fol~w	ln_post	title	ln_len~h	Mean	Std
yuqin	1.0000								
identity	0.1836	1.0000						2.17	0.842
ln_fans	0.3971	−0,1789	1.0000					3626827.77	10209919.3
ln_follow	0.1201	−0.0244	0.4733	1.0000				1389.15	1121.521
ln_post	0.0399	−0.0984	0.4957	0.7168	1.0000			36435.77	36724.176
title	−0.1469	0.1368	−0.1691	0.1062	0.1397	1.0000		7.89	2.425
ln_length	0.0501	0.1536	−0.2322	−0.0837	−0.0719	−0.0399	1.0000	136.83	46.093
picture	0.0518	−0.0631	0.1007	−0.0973	0.0148	−0.0331	−0.0031	3.55	3.855
video	0.0596	0.0511	−0.3295	−0.3436	−0.4784	−0.0610	0.0165	0.06	0.247
link	−0.0843	−0.0559	0.1002	0.2420	0.2743	−0.0928	−0.0793	0.38	0.491
ln_repost	0.5586	0.0573	0.1690	0.2108	0.0985	0.2440	0.1044	2211.91	7986.315
ln_coiranent	0.5038	0.0024	0.1890	0.3526	0.1346	0.2089	0.1775	130307.21	861481.858
ln_like	0.5393	0.0301	0.1081	0.1429	−0.0170	0.1546	0.2836	4435	12794.618
	picture	video	link	ln_rep-*t	ln_com~t	ln_like			
picture	1.0000								
video	−0.2433	1.0000							
link	−0.7340	−0.2057	1.0000						
ln_repost	−0.0355	−0.0463	0.0518	1.0000					
ln_comment	−0.0899	−0.1536	0.0635	0.8813	1.0000				
In like	0.0109	−0.0445	−0.0712	0.9189	0.9056	1.0000			

3.2 Logit Regression Analysis of Fans and Likes of the Model

Because there is a large gap between the collected data, in order to reduce the error, the variables with large orders of magnitude are logarithmically processed. The forward selection method is used for logit regression, that is, the variables are added to the model one by one, and the selected variables are determined by observing the p value. Finally, the selected variables are the number of fans and praise. Therefore, model 1 performs logit regression on the number of fans and praise;

Logit regression was carried out for these two indicators and the dependent variable of public opinion early warning. The results are shown in Tables 2 and 3. The coefficients of independent variables are given in Table 2 and the odds ratio is given in Table 3.

Model 1 includes the number of fans and praise into the regression model at the same time. Through the forward method, we get the estimator of MLE.LR chi square = 26.02, degree of freedom = 2, P = 0.0000 < 0.1, the results of the whole model are significant.

Secondly, we analyze the regression results of the dependent variables fan number and praise number on whether public opinion early warning. The p value of the number of fans = 0.015 < 0.05, that is, whether the significant influence of the number of

Table 2. Regression results of fans and likes (1)

yuqln	Coef.	5td. Err.	z	P > |2|	(95* Conf. Interval]
ln_like	.8488465	.2920541	2.91	0.004	.2764309 1.421262
In fans	.3397388	.14021S9	2.42	0.01S	.0649206 .6145569
cons	−7.670706	2.625479	−2.92	0.003	−12.81655 −2.524861
N	47				
LR i2(2)	26.02				
pseudo R2	0.4027				

Table 3. Regression results of fans and likes (2)

	Odds Ratio	Std. Err.	z	P > |z|	[95% Conf.	Interval]
In fans	1.404581	.1969446	2.42	0.015	1.067074	1.848837
In like	2.33695	.6825158	2.91	0.004	1.318416	4.142345
_cons	.0004663	.0012242	−2.92	0.003	2.72C-06	.0800694
N	47					
LR i2(2)	26.02					
pseudo R2	0.4027					

fans at the level of 5% is an early warning of public opinion. At the same time, the regression coefficient of the number of fans is > 0, indicating that the number of fans has a significant positive impact on whether public opinion early warning exists. Hypothesis 1 is confirmed. The meaning of regression coefficient shows that in the case of a given number of praise points, the advantage estimate of public opinion early warning becomes 1.404581 times of the original for each additional number of fans.

P value of praise amount = 0.004 < 0.01. That is, whether the significant impact of praise at the level of 1% has a public opinion early warning. At the same time, the regression coefficient of praise amount is > 0, indicating that praise amount has a significant positive impact on whether public opinion early warning exists. Hypothesis 2 is confirmed. The meaning of regression coefficient shows that under the condition of a given number of fans, the advantage estimate of public opinion early warning becomes 2.33695 times of the original value every time the number of likes increases.

The regression model is:

$$\text{logit}(\pi_x) = 0.3397388\text{Number of Fans} + 0.8488465\text{Number of Likes} - 7.670706$$

3.3 Evaluation of Model 1

3.3.1 Model Prediction Efficacy Judgment

3.3.1.1 Probability Prediction Efficacy

From the below Fig. 1, the estimated sensitivity = 88.46%, specificity = 80.95%. The proportion of correct classification reached 85.11%. It shows that the accuracy of the prediction effect of the whole model is still very high.

3.3.1.2 Probability Prediction Efficacy ROC Curve

The receiver operating characteristic (ROC) curve is for possible critical values π_0, and the ROC curve can provide more information than the classification table because it summarizes all possible π_0, corresponding predictive efficacy. The area under the ROC

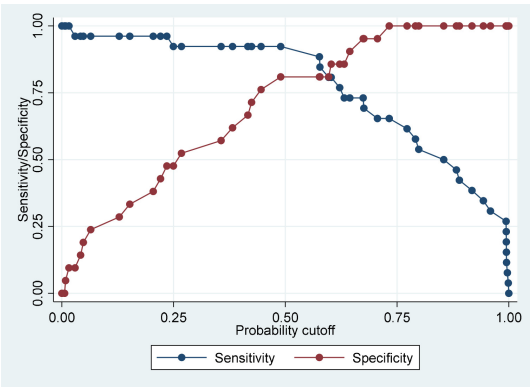

Fig. 1. Relationship among classification probability points, sensitivity and specificity

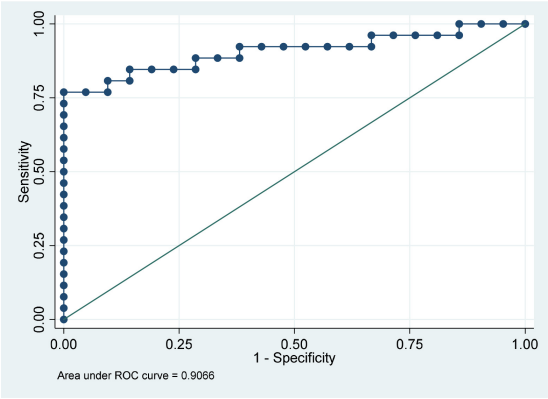

Fig. 2. ROC curve of logit regression model for public opinion early warning

curve is equal to the value of a measure of predictive efficacy, which is called consistency index. Figure 2 shows the ROC curve of logit regression model for public opinion early warning [4, 6].

It can be seen from the above table that the area below the ROC curve is 0.9066. As a measure of prediction efficiency, regional area shows that this logit model has good prediction efficiency. The prediction model has good discrimination ability.

4 Conclusion

By establishing logit binary regression model between public opinion early warning and its independent variables, we can get the following conclusions and suggestions.

The number of fans has a significant positive impact on whether public opinion early warning exists. Hypothesis 1 is demonstrated. When the number of fans on the platform that publishes social public opinion events is more, the social public opinion events published by the platform are easier to early warning. The reason is that the more fans on a platform, the more likely the number of fans is to forward the social public opinion event. Therefore, in a larger range to attract people's attention to the event.

There is a significant positive impact between the number of likes and whether the public opinion early warning, and hypothesis 2 is confirmed. It can be explained that the more likes, the more social public opinion events are concerned by more Internet users, and the content of the event can attract people's enough attention, and the public is more interested in the event. In this case, the public will praise the event, making it easy to read the number of people reading the event increased exponentially, which makes the event easier to become a public opinion early warning event.

With the change and development of computer technology and Internet technology, microblog has come into the publics' vision. The impact of microblog on social public opinion events has led to great changes in all links of the dissemination of social public opinion events from collection, review to release and tracking. Different from the traditional way of public opinion dissemination, microblog public opinion spreads rapidly with the help of the network, with obvious immediate impact characteristics

of sex, congregation, crack change, etc. Relevant state departments should establish a public opinion monitoring mechanism to control the speed of public opinion dissemination. Therefore, we can establish countermeasures to microblog public opinion from the following three aspects: government, platform and individual. For the government, we should first establish a public opinion analysis and research team for hot events to study the causes, communication speed and development trend of social public opinion. The government should solve the hot events in the early stage of public opinion this morning. After the full outbreak, it will be difficult to control the situation and waste human, material and financial resources. For the platform, we should strengthen self-regulation and establish a public opinion monitoring system. The system can be used to monitor the speech status of its own users. When expectations are reoriented and negative, they can be supervised at the first time. For users themselves, strengthen network moral literacy and look at network events rationally. It is not easy to make a conclusion on one thing, and then be affected by bad media, resulting in serious consequences.

References

1. Liu T (2018) Government network public opinion response strategy based on big data. Changchun University of technology
2. Liu DY, Ying Y (2019) Prediction and analysis of network public opinion based on big data. Modern Information Technology 3(12):100–101+103
3. Li M (2016) Research on network public opinion cloud governance of major public events in China. Huazhong University of science and technology
4. Lin ZM, Zhou JF (2019) Logistic regression combined with ROC curve to explore the application value of event-related potential P300 in patients with post-stroke cognitive impairment. Chin Foreign Med (32):1–4
5. Peng ZH (2016) Research path analysis of network public opinion. Libr J 35(12):63–68
6. Pen YQ, Zheng TY, Zhao CJ (2018) Application of logistic regression analysis and ROC curve analysis in the analysis of piperacillin tazobactam allergic reaction. Chin J Hosp Pharm 38(24):2571–2575
7. Shen WY (2018) Research on network public opinion management based on big data. Hunan University
8. Wang GH (2016) Research on supervision and prediction of network public opinion based on big data analysis. Comput Modernization (12):62–66
9. Sun LL (2016) Research on several types of network public opinion research and judgment models and coping strategies. Southeast University