Academic Journal of Humanities & Social Sciences
ISSN 2616-5783 Vol. 2, Issue 2: 111-121, DOI: 10.25236/AJHSS.040050

Public Emotion Analysis Based on the Influence of Opinion Leaders on Weibo

Xiyang Zhao*, Meihong Wu

Business School of Sichuan University, Chengdu 610065, China
**Corresponding author e-mail: 1063697870@qq.com*

ABSRTACT. *Sina Weibo is the most influential social media platform in China. Its large number of users and high activity make it a platform for information dissemination and public opinion fermentation. Taking the M-Cup panic buying incident as an example, this paper first reviews the process from the occurrence to the end of the whole event. Secondly, through emotional analysis and keyword analysis, it analyzes the public's emotional attitude and causes to this event. Then, it identifies opinion leaders by two-level clustering method. Finally, it explores the change of emotional values of opinion leaders and general users. Case analysis shows that S company has well handled the M-Cup panic buying incident and the opinion leader emotinal value fluctuations is greater than that of the general user emotional value.*

KEYWORDS: *Emotion analysis, Weibo, Opinion leader, Two-level clustering*

1. Introduction

In recent years, Weibo has become an important channel for people to disseminate information and express their views. The M-Cup panic buying incident quickly boarded the Weibo hot search and generated online public opinion. Due to the high degree of attention and the adverse effects of the incidents that some people even fought for M-cups in public, S company's statement and response have attracted high public attention. In response to this incident, the article has explored the public's emotional attitudes toward events, and simultaneously analyzed changes in opinion leader and general user emotional value, and finally, proposes some advices towards this event.

The article collects the relevant microblogs of the M-Cup on Weibo as data to explore the changes in opinion leaders and public sentiment values. There are many domestic and foreign literatures on the study of emotion analysis and opinion leaders. The usual emotional analysis categorizes the polarity of text, and judges whether the opinions expressed in the text are positive emotions, neutral emotions or negative emotions, and more finer-grained emotional analysis refers to dividing the text into

ISSN 2616-5783 Vol. 2, Issue 2: 111-121, DOI: 10.25236/AJHSS.040050

seven categories: "joy", "anger", "sorrow", "evil", "fear", "shock" and "no emotion" [1]. In terms of sentiment analysis, there are two main types of techniques: one is the method of using emotional lexicon [2], and the number of positive and negative emotional words in the text to be analyzed is counted with the help of emotional dictionary, and the emotional polarity of the text is analyzed according to the difference between them. The other is the use of machine learning methods [3], labeling training corpus and test corpus, using support vector machine, maximum entropy, KNN and other classifiers for sentiment classification. The concept of opinion leaders was first proposed by American scholar Paul Lazarsfeld in the book "The People's Choice". Those "activists" who often provide information or opinions to others in interpersonal communication and exert influence on others are called opinion Leaders. They play an important intermediary and filtering role in the formation of mass communication effects and spread information to the audience, forming a two-level communication of information transmission [4]. In response to the identification of opinion leaders, Wu Yuhui et al. proposed the opinion leader mining algorithm Topic-LeaderRank to mine opinion leaders [5]. Sun Naili et al. used the analytic hierarchy process to construct a user leader value scoring index system from the perspective of user influence and activity [6]. In response to the influence analysis of opinion leaders, Wang Ping used statistical analysis software to conduct descriptive analysis of opinion leaders, and to study the relationship between various variables to obtain the characteristics of opinion leaders and the impact on public opinion [7].

This article is based on big data for case analysis, collecting microblogs and user data, using sentiment analysis and opinion leader research methods. After identifying the opinion leaders, conduct emotional analysis of opinion leaders and general users and draw emotional time series maps to explore changes in opinion leaders and public sentiment values.

2. Research design

2.1 Data collection

Paper data comes from microblogs about M-Cup from February 18 to March 2. Data collection uses Octopus software to write collection rules and collect relevant data.

2.2 Data cleaning

We export the collected data to EXCEL table, then remove the empty value and missing data, finally, increase the required data. Microblogs that are sent in the same period are merged and sorted out. Moreover, irrelevant data are eliminated, and valid data are retained. We delete microblogs that are sent by the microblogging trumpet, in case that a large number of useless microblogs affect subsequent data analysis.

ISSN 2616-5783 Vol. 2, Issue 2: 111-121, DOI: 10.25236/AJHSS.040050

2.3 Emotion analysis design

First, each microblog data is preprocessed with text, that is, punctuation marks are used as split points, each blog post is divided into certain segments, and then the emotional words in each segment are extracted. Then, taking this emotional word as the center, looking for adjacent adverbs of degree, negative words, etc., and comprehensively scoring against the emotional vocabulary. Finally, we judge the rhetorical tone, sigh tone and other tones and change the corresponding degree of score based on the previous scores.

We adjust the emotional value so that the most positive emotional value in the sentiment dictionary is set to "1", and the most negative emotional value is set to "-1". Above 0 means that this microblog post has positive emotions, and below 0 represents this microblog post has negative emotions. Each blog post is scored through an emotional dictionary, and finally the average is taken to observe the emotional value of the microblog.

2.4 Keyword analysis design

Through word segmentation of all micro-blog posts in the same period, we get the segmented blog posts. Then we extract 300 high-frequency words from the blog posts, and generate a co-word matrix. Using the Aprior algorithm, the confidence of a word to other words is calculated, and all the confidences of the word to other words are summed, and the sum value is used as the weight of the keyword. We select the words with the weights of the top 10 as keywords, and the greater the weight, the more times mentioned in the microblog posts.

2.5 Identify opinion leaders

This article describe the characteristics of opinion leaders from the perspective of the personal charisma and influence of opinion leaders, and use these characteristics to cluster users and identify opinion leaders. The personal charisma will be represented by the attributes of “popularity”, “activity” and “authenticity”; and the influence of communication will be expressed by the “value”, “affected area” and “acceptance” of the published blog post [8] [9].

Table 1 Attribute value description table

Feature	Attribute	Data description
Personal charisma	Popularity	Number of fans
	Activity level	Number of microblogs, Number of followers
	Authenticity	Is it an authenticated user
Spreading influence	Value	Number of comments in blog posts
	Affected area	Number of microblog forwarded in blog posts
	Acceptance	Number of praises in blog posts

ISSN 2616-5783 Vol. 2, Issue 2: 111-121, DOI: 10.25236/AJHSS.040050

2.6 Emotional analysis and design of opinion leaders and the general users

The opinion leaders and the general users in the same priod are classified, and then the microblog articles of opinion leaders and the general users are analyzed respectively. Secondly, the emotional values of opinion leaders and the general users are drawn into emotional time series maps. Finally, the emotional values of opinion leaders and the general users are compared and analyzed, and the corresponding conclusions are drawn.

3. Empirical analysis

3.1 Data acquisition and preprocessing

In this paper, the Octoparse software is used to acquire data from microblog with M-cup as the topic. A total of 28528 blog data and user data are obtained.

The data are cleaned and pretreated by EXCEL software to we remove the blank and missing data and delete the obvious invalid data. After pretreatment, 26599 data were obtained. Table 2 shows microblog data for each period:

Table 2 Microblog data quantity distribution table

Period	Number of microblogs
Before 26-Feb	426
26-Feb	6250
27-Feb	5767
28-Feb	5431
1-Mar	5840
2-Mar	2885
Total	26599

3.2 Sentiment analysis

After the M-Cup panic buying incident on February 26, S Company responded to the first time and responded to the emergency. Then it launched a limited edition of the M-Cup online from the 28th. In the process of event fermentation, S Company issued a statement “rejecting speculation, not marketing hunger” to control the direction of public opinion, and the incident gradually subsided on March 2. Therefore, for the M-Cup panic buying incident, the four time periods before February 26, February 26, February 28, and March 2 were selected before the event occurred, when the public opinion rose, and when the S company responds to the action and when the event subsides. By calculating the emotional values of the four time periods, it is more intuitive to feel the emotional value of the whole event.

ISSN 2616-5783 Vol. 2, Issue 2: 111-121, DOI: 10.25236/AJHSS.040050

Table 3 Microblog data sentiment distribution table

Period	Number of positive microblogs	Number of neutral microblogs	Number of negative microblogs	Total number of microblogs	The proportion of negative microblogs	emotional value
Before 26-Feb	105	227	94	426	22.07%	0.522
26-Feb	1154	2573	2523	6250	40.37%	-0.018
28-Feb	1204	2708	1519	5431	27.97%	0.300
2-Mar	846	1204	835	2885	28.94%	0.323

According to Table 3 analysis, public opinion emerged and broke out on the 26th. As a result of the panic buying incident and fight on the 26th, the emotional value dropped sharply and generated a large number of negative comments on Weibo. On the 28th, S Company took replenishment measures to restrict online purchasing and sales, making the emotional value rise at a certain level, while reducing the number of negative microblogs. On March 2, the incident gradually subsided, the number of microblogs decreased and the emotional value increased, but still lower than that before the event, indicating that the M-Cup panic buying incident still brought some negative effects.

According to the daily emotional value during the event occurrence process, the emotional time series diagram is drawn, as shown in Figure 1.

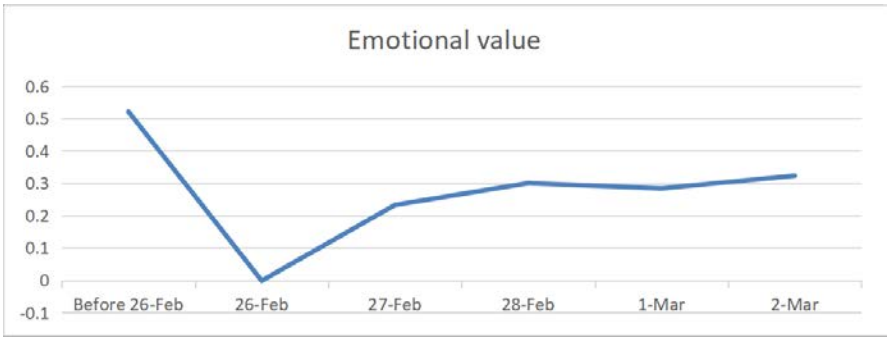

Figure. 1 Emotional value chart

According to the emotional trend chart, it can be seen that on February 26, the emotional value dropped sharply due to fighting events, resulting in network public opinion on Weibo, which has a negative impact. As S Company issued a timely statement and took measures to deal with it, the emotional value continued to rise over time, but it was still lower than the emotional value before February 26.

ISSN 2616-5783 Vol. 2, Issue 2: 111-121, DOI: 10.25236/AJHSS.040050

3.3 Keyword analysis

We select three event segments of microblog that were sent before the event, when public opinion broke out and when the event subsided for keyword analysis, then calculate the keyword weight. The higher the keyword weight, the more times the microblog mentions.

Table 4 Weibo data keyword weight table

Before 26-Feb		26-Feb		2-Mar	
M-cup	2.714	M-cup	3.446	M-cup	2.932
S Company	2.562	S Company	3.122	S Company	1.989
Girl	1.954	Cup	2.221	Cup	1.124
Cat claw	1.72	Fight	1.962	Good-looking	0.921
Limit	1.48	Queue	1.911	Queue	0.904
Lovely	1.439	Good-looking	1.866	Lovely	0.86
Cup	1.383	Lovely	1.835	Marketing	0.853
Weibo	1.246	Holy Grail War	1.275	Cherry blossoms	0.846
Cherry blossoms	1.215	Cat claw	0.853	Limit	0.634
Good-looking	1.171	Cherry blossoms	0.762	Weibo	0.608

M-Cup and S Company are the topic centers in Weibo, and the whole event revolves around S Company and M-Cup. Before the rushing event, most of the keywords in the blog post were “girl”, “good-looking”, “lovely” and so on. These keywords indicate that the attention of netizens to the M-Cup before the rise of public opinion is mostly positive. Keywords such as “cherry blossoms”, “limited” and “cup” indicate that the M-Cup belongs to the cherry blossom season limited cup launched by S Company and has become the target of mass buying.

February 26 was the time when public opinion rose and broke out. Some netizens edited the M-Cup panic buying incident into a blog post and posted it on Weibo, which immediately caused a hot discussion among netizens and generated online public opinion. In response to the M-Cup panic buying incident, words such as “fighting” and “Holy Grail War” are frequently appearing on Weibo, triggering negative emotions of the public and lowering public sentiment values. At the same time, the purchase of the M-Cup requires a long queue of lines, so "queue" also becomes a keyword.

As time goes by, the heat of events is declining, and the key weights of “M-Cup” and “S Company” are declining. Some netizens think that this incident is the hunger marketing behavior of S company, so "marketing" has become one of the key words. Although the incident had a negative impact, according to the keyword analysis, when the netizen mentioned the M-Cup, it was still associated with “good-looking” and “lovely”, indicating that the appearance of the M-Cup is still popular among most people.

ISSN 2616-5783 Vol. 2, Issue 2: 111-121, DOI: 10.25236/AJHSS.040050

3.4 Identify opinion leaders

In this paper, the two-level clustering method is used to identify the opinion leaders. According to the identification of the opinion leader attribute values described in Table 1, the data is identified by the opinion leader. All data is pre-processed. In order to reduce the interference of invalid data, we delete the blog post data which the number of likes, comments, and forwardings is 0, and 12566 pieces of data are finally obtained.

Table 5 Cluster Distribution

		N	% of Combined	% of Total
Cluster	1	437	3.5%	3.5%
	2	12065	96.5%	96.0%
	Combined	12502	100.0%	99.5%
Excluded Cases		64		0.5%
Total		12566		100.0%

According to the sample data shown in Table 5, 3.5% of the users are groups with obvious opinion leader characteristics, and the remaining 96% are general users. Among them, whether the user is authenticated is an important indicator for judging whether it is an opinion leader. From the results of clustering, most of the users who have not authenticated belong to the general user, and most of the clustered opinion leader groups are certified user. At the same time, the opinion leader has a large difference in other attribute values compared with the general user, and has the characteristics of high forwarding number, high number of comments, and high number of praises.

3.5 Opinion leaders and general users emotional value changes

Two kinds of users are obtained through the two-level clustering method, then we use the emothion analysis towards opinion leaders and the general users, and the number of microblog posts and emotion values of the two types of users in each time period are obtained. The emotional time series graph is drawn by the emotional values of the two types of users, and compared with all the microblog emotional values, and the corresponding conclusions are obtained.

ISSN 2616-5783 Vol. 2, Issue 2: 111-121, DOI: 10.25236/AJHSS.040050

Table 6 Opinion leader and general user emotional value table

Period	Number of opinion leaders on Weibo	Number of general User on Weibo	Opinion leader emotional value	General User Emotional Value	Emotional value
Before 26-Feb	15	221	0.675	0.479	0.522
26-Feb	102	2878	-0.290	-0.061	-0.002
27-Feb	95	2156	-0.086	0.231	0.233
28-Feb	72	2429	-0.017	0.126	0.300
1-Mar	90	2861	-0.124	0.191	0.284
2-Mar	63	1520	0.137	0.213	0.323

As time goes by, the popularity of public opinion is gradually decreasing. On February 26, when the M-Cup panic buying incident and fight broke out on microblog, it triggered a heated discussion among netizens and generated online public opinion. After the outbreak of public opinion, the emotional value of opinion leaders is mostly negative, indicating that the emotions of opinion leaders about the M-Cup panic buying incident are always negative. The average user's emotional value for events is slightly higher than that of opinion leaders, but their emotional value also decreases significantly after the event, and produces negative emotions when public opinion erupts.

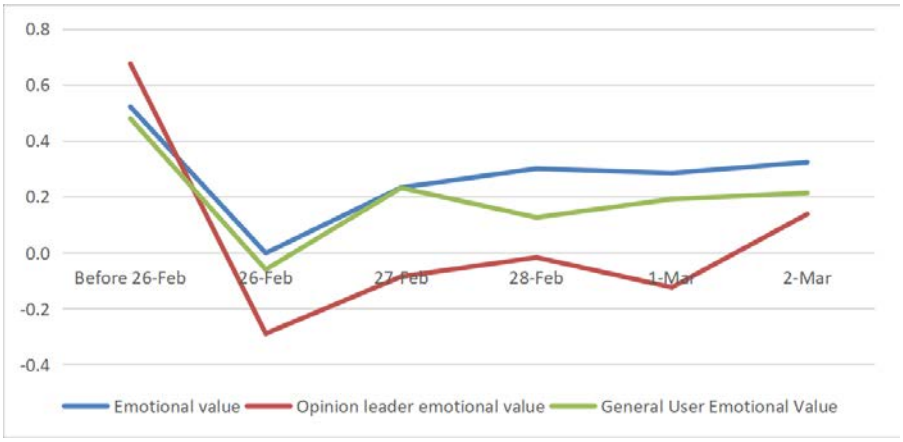

Figure. 2 Opinion leader and general user emotional value chart

According to the analysis in Figure 2, the sentiment value trends of opinion leaders, general users and all users are generally the same. When the public opinion breaks out, all the values are under 0, indicating that the M-Cup panic buying incident has a large negative impact, therefore reduce the user's emotional value. After the incident, S company issued a statement and took countermeasures to save

ISSN 2616-5783 Vol. 2, Issue 2: 111-121, DOI: 10.25236/AJHSS.040050

the negative impact, so that the emotional value is gradually increasing, but still lower than the emotional value before the incident, still affecting the reputation of S company in China.

The opinion leader's emotional value fluctuated more violently than the average user. Before the incident, the opinion leader's emotional value was 0.675, which showed a positive emotion for the M-Cup. After the incident, the emotional value dropped to -0.290, and the emotional fluctuations changed greatly, with obvious negative emotions. The average user's emotional value fluctuates less. When public opinion occurs, it has negative emotions. When the event gradually subsides, the emotional value is constantly rising, slightly higher than the opinion leader's emotional value.

The emotional value of the blog of all users is the emotional value calculated when the the number of likes, the number of comments, and the number of forwardings are 0. According to the analysis of Figure 2, the emotional values of all users are slightly higher than the emotional values of the general users. Most of the time, it is higher than the emotional value of the opinion leader, indicating that the microblogs which the number of likes, comments, and forwardings are 0 tends to be positive, and improves the emotional value of all users. It infers that S company has take action to control public opinion, leading a positive emotional development.

4. Conclusion

The paper takes the M-cup panic buying incident as a case, after crawling the Weibo data, preprocessing the data, useing sentiment analysis, keyword analysis, two-level clustering method to identify opinion leaders, analyze opinion leaders and general users. Then it draws a conclusion that the emotional value of the opinion leader changes greatly compared with the emotional value of the general user, and the enterprise needs to pay attention to the conclusion of the emotional value of the opinion leader in the public relations crisis. In the course of this incident, the following conclusions were drawn:

Before the incident occurs, it is necessary for enterprises to do a good job in monitoring public opinion and discover network publicity in a timely manner. Before the M-Cup panic buying incident, the M-Cup was positively evaluated on the Internet. At this time, public opinion monitoring should be conducted to check the product sales system. The limited purchase problem is the trigger for the M-Cup to buy and fight.

When the event begins and quickly ferments, enterprises need to issue a statement to respond first. When the Internet generated a public opinion about fight against the M-Cup on the 26th, S Company responded at the first time and issued a microblog statement at 16:18 pm on the 26th that it is stepping up replenishment and will sell it online. S company's timely response saves the public's emotional value and greatly reduces the occurrence of negative emotions.

ISSN 2616-5783 Vol. 2, Issue 2: 111-121, DOI: 10.25236/AJHSS.040050

When the event develops, enterprises should guide the direction of public opinion and reduce the emergence of negative public sentiment. In response to the M-Cup panic buying incident, S Company issued a statement stating that “not hunger marketing” and “rejecting speculation”, such discourses control the trend of public opinion and gradually increase the public's emotional attitude value. At the same time, enterprises should pay attention to the emotional value of opinion leaders, increase the number of positive emotional microblogs, and improve the public's emotional value.

As the event subsides, the enterprises should appease customers and use other products to divert public attention. When the panic buying incident subsided, S Company used the heat to promote other products of the cherry blossom season theme, divert public attention and promote marketing for other products. The information is time-sensitive, so when the heat of the network decreases, the other products can be used to divert public attention, reducing the negative emotions of the public and improving the corporate image.

There are two inadequacies in the article. One is that using the sentiment dictionary to analyze the emotions of microblog posts is kind of one-sided and reduces the accuracy. Second, the data source is a little singular, only microblog data is collected. Future research can collect data from emerging social platforms, such as Tik Tok, Xiao Hong Shu and so forth for sentiment analysis research, and do the time difference correlation analysis about sentiment values between opinion leaders and general users to explore the impact of opinion leaders on the general user's emotions.

References

[1] Ren Juwei, Yang Liang. Prediction of Public Emotional Trend in Weibo Events Based on Emotional Common Sense [J]. Journal of Chinese Information Processing, 2017, 31 (2): 169-178.

[2] Lunwei Ku, Tungho Wu, Liying Lee, et al. Construction of an Evaluation Corpus for Opinion Extraction [C] // NTCIR-5 Japan.2005: 513-520.

[3] Dasgupta S, Ng V. Mine the Easy. Classify the Hard: S Semi-Supervised Approach to Automatic Sentiment: Classification [C] //ACL’09: 701-709.

[4] Lazarsfield P F, Berelson B, Gauset H. The People’s Choice: How the Votes Makes up His Mind in a Presidential [M]. NewYork: Columbia University Press, 1948.

[5] Wu Xianhui, Zhang Hui, Zhao Xujian, et al. Mining Algorithm of Microblogging Opinion Leaders Based on User-behavior Network [J]. Application Research of Computers, 2015, 32 (9): 2679-2683.

[6] Sun Naili, Wang Yulong, Shen Qiwei. Microblog Opinion Leader Identification [J]. Telecommunications Technology, 2012 (12): 78-80.

[7] Wang Ping, Xie Yungeng. An Empirical Study of Micro-blog Opinion Leaders in Public Emergencies——A Case Study of Wenzhou Motor Train Crash [J]. Modern Communication: Journal of Communication University of China, 2012, 34 (3): 82-88.

ISSN 2616-5783 Vol. 2, Issue 2: 111-121, DOI: 10.25236/AJHSS.040050

[8] Wang Di, He Yue. The Network Structure of Opinion Leaders Based on Social Network Analysis [J]. Statistics & Information Forum, 2013, 28 (10): 84-89.
[9] He Yue , Zhu Can. Sentiment Analysis of Weibo Opinion Leaders—— Case Study of “Illegal Vaccine” Event [J]. Data Analysis and Knowledge Discovery, 2017 (9): 65-73