Information Visualization Analysis of Public Opinion Data on Social Media

Feng Chen and Shi Zhang
Luxun Academy of Fine Arts, Liaoning 116650, China
E-mail: chengfang09583@163.com

Public opinion data on social media contains much useful information, which can be visually displayed through visualization. This study mainly focused on Weibo and analyzed the keyword extraction of text and the analysis of emotional tendency. Keywords were extracted using the term frequency-inverse document frequency (TF-IDF) method, and the emotional tendency of the text was calculated based on the HowNet emotion dictionary and BosonNLP emotion dictionary. Finally, relevant data were collected by taking "Jiang Ziya" as the keyword for visualization analysis. It was found that the discussion on "Jiang Ziya" gradually reduced in the research period, and the extracted keywords were relatively positive. The visualization results of word cloud showed that there were many positive comments on "Jiang Ziya", but there were also negative comments. Finally, the calculation of emotional tendency showed that 69% of the texts showed a positive emotional tendency, and 31% of the texts were negative, indicating that netizens' emotional tendency towards "Jiang Ziya" was mainly positive. The study results make some contributions to the visualization of public opinion data and can be further applied in practice.

Povzetek: Razvita je metoda za vizualizacijo mnenj v socialnih omrežjih, tj. na Weibo.

1 Introduction

With the development of network technology, the popularity of networks has improved, and the number of network users is also growing. It is not only a tool for people to learn and work. Because of its anonymous, timely, and interactive characteristics, the network has become a new platform for people to obtain, share, and exchange opinions. Compared with traditional media, network-based social media plays a great role in information spreading and exchange. Everyone can express their opinions and spread the news through social media, promoting the rapid development of online public opinion. Moreover, the particularity of social media makes network public opinions have concealment and abruptness [1]. It is an important issue to guide and monitor public opinions correctly. Social media includes Weibo, WeChat, forums, etc., which contains many public opinion data. These data contain netizens' emotion and attitude towards events, which has a strong influence and will change with the development of events. However, if the public opinion of some events develops to some scale or the trend is not conducive to social stability, it may cause chaos of public opinion and lead to the occurrence of adverse public opinion events. Therefore, to create a stable public opinion environment, it is necessary to monitor, manage, timely warning, and guide public opinion, thereby establishing a harmonious network environment. At present, the research on these data includes hot spot discovery [2], crisis early warning, public opinion prediction [3], etc. Tan et al. [4] analyzed the campus network public opinion, explored technologies, such as Chinese word segmentation and topic recognition, and combined analytical hierarchy process (AHP) with wavelet neural network to monitor the network public opinion. Taking the Sade event as an example, they found that the method had a good estimation accuracy. Tang et al. [5] studied the role of fuzzy sets in network public opinion analysis, compared different functions and advantages of different fuzzy sets, and discussed the future trend of this field. Chai and Cheng [6] proposed an improved AHP-entropy method to evaluate the risk of network public opinion, which comprehensively considered the subjective weight and objective weight, and verified the effectiveness of the method by experiments. Zhang et al. [7] predicted network public opinion with the gray model, corrected the results of the gray model with the back-propagation neural network, and carried out a simulation experiment on the model by taking a hot topic as an example. The results showed that the method could effectively and accurately predict public opinion. The public opinion data on social media involves a lot of content, but the current research is mostly static and one-sided, which can not show the information dynamically and stereoscopically. As an effective way of data expression, visualization technology has good application in information processing, but it has less application in processing public opinion on social media. Thus, this study took Weibo as the research subject, analyzed the keyword extraction and emotional tendency, displayed the information through visualization, with the intention of understanding the reliability of the

visualization method in the public opinion data processing.

2   Public opinion and visualization

The public opinion data on social media is the information collection of a specific event, and it will change with the development of time. It is ① massive: in the network, an event can cause a great deal of discussion in a very short time, thus generating a huge amount of information; ② diverse: public opinion data has a variety of forms, and the concept of Internet users can also be expressed by videos and pictures in addition to text; ③ dynamic: public opinion is in rapid dynamic evolution.

Given the characteristics of public opinion data, traditional information processing methods can not effectively deal with them [8]. As a relatively mature information processing method, visual analysis combines computer graphics, data mining, etc., and it can show massive and abstract information in a visual way [9] to help people find the hidden information in the data. For public opinion data, visualization can display the distribution, development, and change of public opinion through images so that people can analyze the public opinion data dynamically and globally. A chart is a basic form of visualization. At present, the commonly used visual chart include ① bar chart, which displays the difference of data through rectangles with different lengths; ② histogram, which is used for understanding the distribution of things; ③ pie chart, which is used for showing the proportion of items; ④ trend chart, which can display the development trend of time; ⑤ theme river chart, which is used for displaying the change of events in a period; ⑥ word cloud, which is the visualization of keywords and can directly display the main idea of the text.

3   Processing of Weibo public opinion data

3.1   Keyword extraction of Weibo text

Weibo is a very widely used social media. Users can participate in the discussion of events as long as they register the account. Everyone can be both the publisher of information and the receiver of information. With the development of the network, the influence of Weibo has become increasingly larger. More and more events arouse the widespread concern of people across the country through the form of hot search on Weibo. Due to the characteristics of Weibo, the Weibo text is limited to 140 words, which is concise and comprehensive and is a fragmented description of an event. These texts include users' emotions and attitudes towards the event. Keywords in the text is a summary of the theme, which contains the key information of the text. The extraction of keywords is conducive to understand the opinions and emotions of users.

Before keyword extraction, first of all, text segmentation is needed. Text segmentation refers to dividing a complete sentence into separate words. Weibo text is mainly in Chinese, with few words and prominent colloquialism. In this study, natural language processing and information retrieval (NLPIR) (Institute of Computing Technology, Chinese Lexical Analysis System) [10] was used for word segmentation. It is a Chinese word segmentation system developed by the Chinese Academy of Sciences. In addition to word segmentation, it will also mark properties of words, which is more conducive to subsequent text processing.

After word segmentation, due to the existence of punctuation, space, and function words in the text, in order to improve the efficiency of text processing, it is necessary to remove stop words. The specific steps are as follows: ① function words, such as adverbs, prepositions, and conjunctions, and notional words, such as numerals and quantifiers, have no practical significance; therefore, these words with little public opinion information should be filtered; ② in the process of word segmentation, some symbols, numbers, and separate words will be segmented; therefore, separate words, numbers, and symbols with a length of 1 should be filtered; ③ Weibo texts usually carry some special words, such as "comment", "link", "forward", etc., which contains few public opinion information and cannot reflect the emotion of users, and these words should also be filtered.

In this study, keywords were extracted by the term frequency-inverse document frequency (TF-IDF) model [11]. In a text, if a word has high TF (high frequency in the same text) and high IDF (rarely appears in other documents), it can be used as a keyword. For word t, the calculation method of TF is: $TF = \frac{count(t)}{count(d_i)}$, where $count(t)$ refers to the number of t in document d_i and $count(d_i)$ refers to the total number of words in document d_i, and the calculation method of IDF is: $IDF = \frac{num(corpus)}{num(t)+1}$, where $num(corpus)$ refers to the number of documents in corpus and $num(t)$ refers to the number of documents containing word t in the corpus.

3.2   Analysis of emotional tendency

In analyzing public opinion data, emotional analysis is a key part, reflecting the users' emotional tendency to an event. The extraction of keywords can help understand users' key emotions to the event, but it is difficult to judge the specific emotional tendency. Based on HowNet emotional dictionary [12] and BosonNLP emotional dictionary, this study analyzed the emotional tendency of Weibo text. HowNet emotional dictionary includes 836 positive emotion words and 1254 negative emotion words. BosonNLP emotional dictionary takes the emotional value as the expression of emotional tendency, positive number as positive emotion word, and negative number as negative emotion word. BosonNLP emotional dictionary includes 114767 words.

A Weibo text is decomposed into several sentences, and then the emotional polarity of every sentence is calculated. In a known document named D , there are n sentences. The document can be written as $D = \{s_1, s_2, \cdots, s_n\}$. Firstly, the emotional value of every

sentence is calculated: $F(s_i) = \sum s_{w_i}$, where s_{w_i} stands for the emotional value of word w_i in the sentence. Then, the emotional value of the whole text is: $F(s) = \sum F(s_i)$. If $F(s) > 0$, it indicates that the text has a positive emotional tendency; if $F(s) < 0$, it indicates that the text has a negative tendency; $F(s) = 0$, it indicates that the text is neural.

The verbs and adjectives are separated from the text as emotional words for calculation using the word segmentation system: $s_{w_i} = \frac{fp_{w_i}}{(fp_{w_i}+fn_{w_i})} \times \frac{N_p}{(N_p+N_n)} - \frac{fn_{w_i}}{(fp_{w_i}+fn_{w_i})} \times \frac{N_p}{(N_p+N_n)}$, where fp_{w_i} stands for the ratio of w_i to positive emotional words, fn_{w_i} stands for the ratio of w_i to negative emotional words, N_p stands for the number of positive emotional words in the emotional dictionary, and N_n stands for the number of negative emotional words. For the calculated result, $s_{w_i} > 0$ is determined as the positive emotional word, $s_{w_i} < 0$ as the negative emotional word, and $s_{w_i} = 0$ as the neural word.

4 Analysis of Weibo information visualization

The film "Jiang Ziya" was released in the mainland of China and North America on October 1, 2002. The film was originally scheduled to be released at the Spring Festival in 2020 but was canceled due to the influence of the epidemic. After its release, the film has caused extensive discussion on Weibo. The reading quantity of "Jiang Ziya" on Weibo has reached 430 million, and there are about 227000 discussions. In this study, "Jiang Ziya" was taken as the keyword. Through the octopus data collector, Weibo texts were collected as public opinion data from September 30, 2020, to October 15, 2020. First of all, the top ten heat searches related to "Jiang Ziya" in this period are shown in Table 1.

It was seen from Table 1 that the release of "Jiang Ziya" had caused extensive discussion on Weibo. The highest ranking of "Jiang Ziya" on the hot search list is 5, and the lowest is 35. In this period, the spread trend of "Jiang Ziya" is shown in Figure 1.

It was seen from Figure 1 that the popularity of "Jiang Ziya" was the highest on the day of its release, and the number of Weibo texts related to it reached 2261. Then, discussions on "Jiang Ziya" began to decline. On October 8, according to the box office of the National Day, the box office reached nearly 3.7 billion yuan; the box office of "Jiang Ziya" was 1.324 billion yuan, which aroused a new round of discussion. Subsequently, with the extension of the showing time, the number of Weibo texts related to "Jiang Ziya" gradually decreased.

Keywords were extracted from the collected Weibo texts using the TD-IDF method. The top 10 keywords are shown in Table 2.

It was seen from Table 2 that most netizens' comments on "Jiang Ziya" were positive and they thought that "Jiang Ziya" was a breakthrough in animated films and had high expectations for it. In order to display

Ranking	Topic of conversation	Maximum heat	The highest ranking
1	"Jiang Ziya" extended showing	985577	11
2	The box office of "Jiang Ziya" exceeds 1.5 billion	227522	35
3	The box office of "Jiang Ziya" exceeds 1 billion	424056	20
4	Details in "Jiang Ziya"	1089140	6
5	The box office of "Jiang Ziya"	655630	16
6	Imitated makeup of Daji Su in "Jiang Ziya"	1429278	5
7	"Jiang Ziya" renews the 1st-week box office record of animation film	1124068	9
8	Post-credit scenes of "Jiang Ziya"	913289	22
9	The box office of "Jiang Ziya" exceeds 0.6 billion	309196	30
10	Poster of Dujie City for "Jiang Ziya"	415238	20

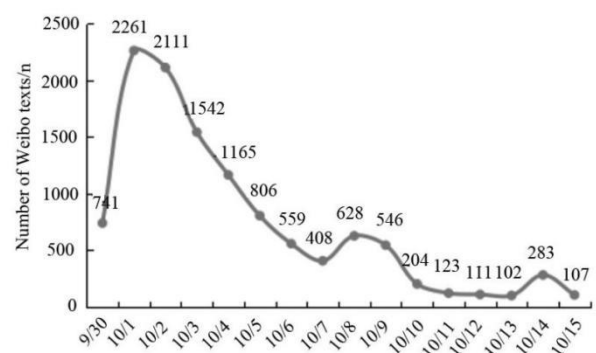Table 1: Hot searches related to "Jiang Ziya" on Weibo.

Figure 1: The spread trend of "Jiang Ziya".

keywords more intuitively, they were displayed in the form of the word cloud, and the results are shown in Figure 2.

1	Jiang Ziya
2	Film
3	Animation
4	Plot
5	Good
6	China
7	Expect
8	Hot blood
9	Special effects
10	Post-credit scene

Table 2: Keyword extraction results.

Figure 2: Keyword visualization.

It was seen from Figure 2 that the keywords in the texts could be seen directly through the word cloud. The larger the font was, the more times it appeared. The smaller the font was, the fewer times it appeared. The most prominent words in Figure 2 are "Jiang Ziya", "film", "plot", "animation", "good", "expect", etc., which showed that netizens generally had great expectations on the film. The plot of "Jiang Ziya" has been widely discussed by netizens. However, it was also seen from Figure 2 that there were words such as "polarization", "not good", "low", and "poor", which indicated that some netizens had negative comments on the film and they thought that the film was not good. As shown in Table 3, netizens mainly had two emotional tendencies.

By analyzing the emotional tendency of Weibo text, the number of positive and negative emotional tendencies is obtained by calculating the emotional value, as shown in Figure 3.

It was seen from Figure 3 that texts with a positive emotional tendency ($F(s) > 0$) accounted for 69% (8071) among the collected Weibo texts and texts with a negative emotional tendency ($F(s) < 0$) accounted for 31% (3626), which showed that the emotional tendency of netizens towards "Jiang Ziya" was mainly positive. Generally speaking, netizens love and praise the film.

Negative	We have waited for its release so long, but it is not good.
	"Jiang Ziya" is not good. I slept twice in the process of watching the film. I fell asleep in the cinema for the first time.
	Our expectations of "Jiang Ziya" are very high, and we have other feelings, even more than our expectations of "Nezha: Birth of the Demon Child". Under this expectation, even if the two films are equally good, we will still feel that "Jiang Ziya" is not good because it does not meet our psychological expectations.
	"Jiang Ziya" is not good. It may be because we have waited for its release so long. It is very general.
	How bad is "Jiang Ziya"? I played with my mobile phone when watching the film and left before the post-credit scene.
Positive	"Jiang Ziya" that was released recently is very good.
	I watched the film! Jiang Ziya!!!!!!!!!!!!!!!!!!! So gooooooood! Nezha in the post-credit scene, hahaha! Chinese animation is getting better and better! "Jiang Ziya" expected # Jiang Ziya's post-credit scene, deep sea# it has a feeling of Chinese style, and the hotel makes me remember the film of Miyazaki Hayao.
	The picture is super beautiful. Every frame is super beautiful and fluent, and the plot is wonderful too. I read the comments before watching the film, and someone said that the character of Jiang Ziya is unlikable and the film will not be good with Jiang Ziya as the main character. But now, I think such a comment is wrong. He is likable, the main characters are likable, the supporting roles are likable, and the plot is likable. It doesn't matter what you think, and it matters what I think.
	After watching "Jiang Ziya", I want to say that it is great. I hope it will get better and better. The animation effect is excellent. The adoption of plots is also good, but it can be better. Such excellent works should be encouraged as the revitalization of Chinese animation just start. I am looking forward to the rest of the story in Dujie City.
	"Jiang Ziya" is good, and the elk is so cute!

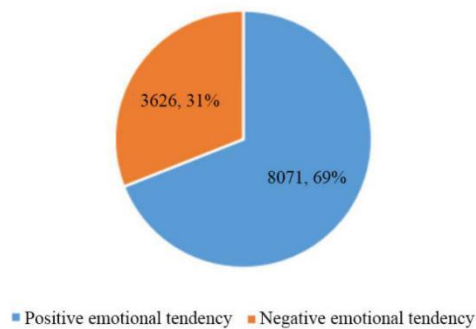Table 3: Netizens' emotional tendency towards the film.

Figure 3: Analysis results of emotional tendencies.

5 Conclusion

In this study, the visualization of public opinion data was analyzed. Taking Weibo texts as the subject, the text data were processed through keyword extraction and emotional tendency analysis and visualized. Then, Weibo texts were collected by taking "Jiang Ziya" as the keyword. The results showed that this method proposed in this study could extract the keyword of the text, visualize texts in the form of the word cloud, and calculate the emotional tendency of texts, which plays a very good role in correctly grasping public opinion data and understanding the emotional attitude of netizens.

References

[1] Fei Y, Qian Z, Xiao G (2017). Evolution mechanism and countermeasures of network public opinion of group emergencies based on data mining method. Boletin Tecnico/Technical Bulletin, 55, pp. 196-202.

[2] Chen Y C, Hui L, Wu C I, Liu H Y, Chen S C (2017). Opinion leaders discovery in dynamic social network. 2017 10th International Conference on Ubi-media Computing and Workshops (Ubi-Media).

[3] Dai J, Li Y (2017). Modeling and Simulating of Network Public Opinion Evolution Based on Dynamic Reference Point of Prospect Theory. 2017 6th International Conference on Measurement, Instrumentation and Automation (ICMIA 2017).

[4] Tan Y, Lin Q, Luan Y, Chen T, Qiao Y, Luan Y (2019). Campus Network Public Opinion Monitoring System Based on Reptile Technology. IOP Conference Series: Earth and Environmental Science, 252, pp. 052136 (8pp). https://doi.org/10.1088/1755-1315/252/5/052136.

[5] Tang J, Wang J, Li F (2020). Research Progress of Network Public Opinion Based on Fuzzy Set from the Perspective of Big Data. Journal of Physics: Conference Series, 1631, pp. 012108 (6pp). https://doi.org/10.1088/1742-6596/1631/1/012108.

[6] Chai W L, Cheng M (2016). The Research on the Network Public Opinion Risk Assessment based on the CWAHP-Entropy Method. International Journal of Security & Its Applications, 10, pp. 197-208. https://doi.org/10.14257/ijsia.2016.10.4.19.

[7] Zhang X (2016). Network Public Opinion Data Mining Model of Hierarchical Multi Level. Journal of Computational and Theoretical Nanoscience, 13, pp. 9498-9501. https://doi.org/10.1166/jctn.2016.5872.

[8] Yuan F, Yang J, Zheng Q (2019). Research on Network Public opinion Analysis Platform Architecture Based on Big Data. IOP Conference Series: Earth and Environmental Science, 252, pp. 032014 (6pp). https://doi.org/10.1088/1755-1315/252/3/032014.

[9] Wang L (2015). Big Data and Visualization: Methods, Challenges and Technology Progress. Canadian Journal of Electrical & Computer Engineering, 34, pp. 3-6. https://doi.org/10.1109/CJECE.2009.5443861.

[10] Kay S, Zhao B, Sui D (2015). Can Social Media Clear the Air? A Case Study of the Air Pollution Problem in Chinese Cities. Professional Geographer, 67, pp. 351-363. https://doi.org/10.1080/00330124.2014.970838.

[11] Chen K, Zhang Z, Long J, Zhang H (2016). Turning from TF-IDF to TF-IGM for term weighting in text classification. Expert Systems with Applications An International Journal, 66, pp. 245-260.

[12] Jiang X, Qiu L (2013). A Tibetan Ontology Concept Acquisition Method Based on HowNet and Chinese-Tibetan Dictionary. 2013 International Conference on Asian Language Processing.