Risk Identification of Internet Public Opinion on Social Media in Public Crisis Events Based on the Five-stage Perception State - A Brief Analysis on the Impact of Emotional Characteristics on the Risk Identification of Online Public Opinion

Zhiyuan Sun*[a], Anwen Pan[b]

{[a]19163195801@163.com*, [b]13971195932@163.com}

Wuhan University of Technology School of Safety Science and Emergency Management, Hongshan District, Wuhan 430079, Hubei, China

Abstract. It was proposed during the 20th National Congress of the Communist Party of China that it was necessary to improve the public opinion supervision system and improve the public opinion guidance mechanism for major public opinions and emergencies. This paper focuses on public crisis events. In consideration of the characteristics of viral spread, emotional spread, and mixed spread of online public opinion in public crisis events, and the government's governance of public crisis events, where time pressure is extremely high, uncertainty risks are great, and risk causes are complex, it is of great importance to build an accurate identification system for public opinion risks on social media network in public crisis events. This paper carried out theme crawlers on multiple mainstream social platforms. According to the number of information releases and event node combinations, public crisis events were divided into five stages, namely the public opinion embryonic period, the public opinion expansion period, the public opinion hot discussion period, the public opinion fluctuation period and the public opinion fading period, and public opinion data were collected and organized; This paper carried out custom named entity recognition and emotion annotation to expand the public opinion corpus of public crisis events; it also built a fused BERT pre-training language model based on the Transformer framework based on classic cases of sudden public crises, conducted public opinion emotion recognition and emotion analysis, and explored emotional characteristics of online public opinion in the five stages of perception. In addition, this paper explored the correlation of various potential risks of online public opinion from the perspective of public opinion emotions so as to improve the identification system of online public opinion risks.

Keywords: online public opinion, big data application, big data search and information retrieval technology

1 Introduction

The development trend of network information is gradually becoming standardized and modernized. In order to ensure a safer network information environment, it was proposed in the report of the 20th National Congress of the Communist Party of China that it was necessary to improve the public opinion supervision system and improve the public opinion

IEDM 2024, February 23-25, Kuala Lumpur, Malaysia

DOI 10.4108/eai.23-2-2024.2345911

guidance mechanism for major public opinions and emergencies. Since the 21^{st} century, with the rise of social network platforms, short video traffic dominating the screen, and the popularity of mobile smart terminals, various public opinion fields have become diverse and complex. In addition, the frequency, radiation range, and damage of public crisis events have also changed. The intensity is far greater than before. In public crises such as the COVID-19 epidemic, heavy rainstorms in Henan, a 7.4-magnitude earthquake in Qinghai, and forest fires in Xichang, Sichuan, netizens in a state of panic and anxiety used social network platforms to express their emotions and attitudes, and even the Internet platform served as a channel for their venting anger, causing a large amount of true and false public opinion information to emerge instantly and spread geometrically to the outside world, resulting in a viral fission-like communication effect.

2 Research objectives

Based on the above background, this paper has the following research purposes:

1. Based on the public opinion data of sudden public crisis cases and the development trend of public opinion, construct a language model to analyze the emotional tendency of online public opinion on sudden public crisis events, and expand the public opinion corpus of public crisis events.

2. Build an risk identification system of online public opinion based on users' communication behavior characteristics, emotional characteristics, public opinion development trends, and public opinion communication mechanisms.

3 Research status and development trends at home and abroad

Information is a resource with a life cycle, and online public opinion also has its life cycle. According to the evolution rules and characteristics, identifying and foreseeing possible future risks at the early stage of public opinion are the ultimate goal of early warning research on online public opinion. As early as the 20^{th} century, educational circles began to study the propagation rules and mechanisms of online public opinion. In the existing literature, for example: Wen Zhitao et al. [1] built an online public opinion situation awareness model based on the Gompertz population growth model and divided the online public opinion evolution cycle into three stages. By combining with the three levels of subject, theme and emotion to describe the embryonic period, expansion period and fading period of the development of public opinion, nine independent perceptual states were constructed; Ma Yue et al. [2] adopted the classic five-stage model , divides the development process of public opinion on the Internet for public health emergencies into the embryonic period, outbreak period, spread period, and fading period, and summarizes the characteristics of each stage of the evolution life cycle of public opinion on the Internet for public health emergencies; Wang Lin et al. [3] crawled the number of blog posts with the keywords "Fangcang Hospital" and "Cross-infection" was used to divide the stages based on the number of information releases, and the cycle of this public opinion event was divided into three stages: the generation period, the dissemination period and the fading period; Huang Shijing[4] et al. From the perspective of sentiment analysis, combined with sentiment scores and search index, public opinion was divided into five stages: embryonic period, rapid rise period, concentrated outbreak period, steady decline period, and decline and fluctuation period. Through search index and sentiment analysis, it was found that the risk perception of netizens had a cumulative effect. Grasping the various evolutionary stages of online public opinion and understanding the characteristics of each stage are of great significance to the identification and early warning of online public opinion risks.

Regarding the research on risk identification of online public opinion and early warning, domestic and foreign scholars generally extract information elements to construct a risk identification index system based on the detection of online public opinion. For example: Lian Zhixuan et al. [5] selected bloggers' identities as a qualitative indicator for Microblog public opinion, the number of views was used as a representation index, and then 11 quantitative indicators were extracted from bloggers' characteristics and blog post characteristics to construct a first-level index system based on logistic prediction; Chen Peiyou et al. [6] started from the main body of public opinion and constructed a system that includes the destructive power of events, The social network public opinion risk early warning indicator system with four first-level indicators, media influence, netizen influence, and government guidance and control, as well as 17 second-level indicators, is more systematic and hierarchical. However, to a greater or lesser extent, these indicator systems suffer from imperfect or redundant indicators, or some of the indicators are difficult to obtain quantitative data for in practice.

With the development of cutting-edge computer technology, some scholars have used topic crawlers, web page cleaning, and other technologies to conduct data mining on multi-modal information in data acquisition and processing, who used classic classifiers of machine learning and in-depth learning neural networks to conduct text analysis. Classification and clustering processing uses natural language processing and similar text aggregation analysis for topic identification and tracking, and uses natural language processing, machine learning, and in-depth learning for emotional tendency analysis. For example, Nemes L et al. [7] analyzed the negative sentiment tweet and, sorted based on entity type and mention frequency, which could highlight the main dissatisfaction of the people during the epidemic; Cai Yang [8] conducted research based on building a model based on the attention mechanism, and improved the structure of Transformer to solve the problem sentence-level and aspect-level text review sentiment analysis problems, and for fine-grained aspect-level sentiment analysis problems. The Light-Transformer-ALSC model was proposed to better evaluate the impact of aspect words on the overall text sentiment polarity; Wang Gang[9], based on the transfer learning method, proposed a relationship-based emotional knowledge learning and transfer model R-EKLT, who combined the self-attention mechanism to realize the visualization of emotional knowledge, thereby enable himself to perform trainings on emotional classifiers more easily in new areas without rich labels; Yadav et al. [10] proposed a locationless embedding model based on attention mechanism for aspect level sentiment classification Analyze and conduct experiments on datasets Restaurant 14, Laptop 14, Restaurant 15, and Restaurant 16,The accuracy rates reached 81.37%, 75.39%, 80.88%, and 89.30%, respectively.

Among them, researches with emotional tendency analysis as the main body accounted for a relatively small proportion of the total. Emotional tendency is one of the most valuable data of online public opinion, which can reveal the development trend of online public opinion. Existing research results have shown that the optimization of technical means such as topic identification and sentiment analysis, etc. can effectively improve the early warning effect. He Jiejun et al. [11] studied the emotional evolution rules of public opinion reversal events from two dimensions: time and space, used the LDA topic clustering method for modeling, and analyzed the emotional characteristics of each stage of evolution; An Lu et al. [12] studied the emotional evolution of public opinion, carried out fine-grained divisions, calculated emotional intensity, and analyzed the impact of emotional types and emotional intensity on the evolution of online public opinion communication in emergencies; Fu P et al. [13] constructed an online public opinion communication model that considered individual emotional attitudes and found that the emotion like panic, sadness, etc. has an important impact on the formation and diffusion speed of online public opinion.

After collecting the relevant online public opinion sentiment analysis literature in the past five years, a keyword co-occurrence map has been drawn as shown in Figure 1.

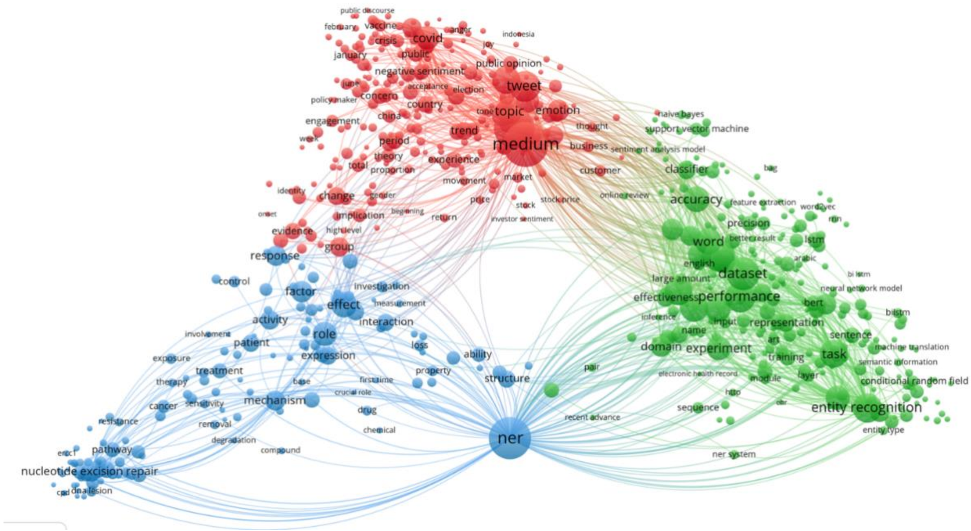

Fig. 1. Keyword co-occurrence map

Judging from the literature published in the web of science in the past three years, among the research directions of social media network public opinion sentiment analysis, the research on NER is relatively popular. Based on the existing literature, there are the following two issues to be added: On the one hand, based on In the sentiment analysis of machine learning, the classification and measurement of emotions mostly focus on the simple binary classification of positive and negative, and there is a lack of more scientific and detailed multi-emotion classification. Sentiment analysis based on in-depth learning is relatively rare and has only emerged in recent years.

On the other hand, in existing research, research on the evolution life cycle of online public opinion and research on online public opinion trends are often conducted separately. In fact, online public opinion trends are the external manifestation of the evolution and progress of online public opinion. Therefore, based on the evolution laws and communication mechanisms of online public opinion, emotional analysis of online public opinions was conducted, which enriched emotional division dimensions, improved risk identification system indicators, and better improved the efficiency of risk warning.

In addition, some existing literature is too segmented in the sub-research direction of risk identification of online public opinion, so that its research results are often not universally applicable to the entire online public opinion environment or a series of common emergencies.

4 Research content

This paper is mainly divided into five parts of research. The research content and logical framework are shown in Figure 2:

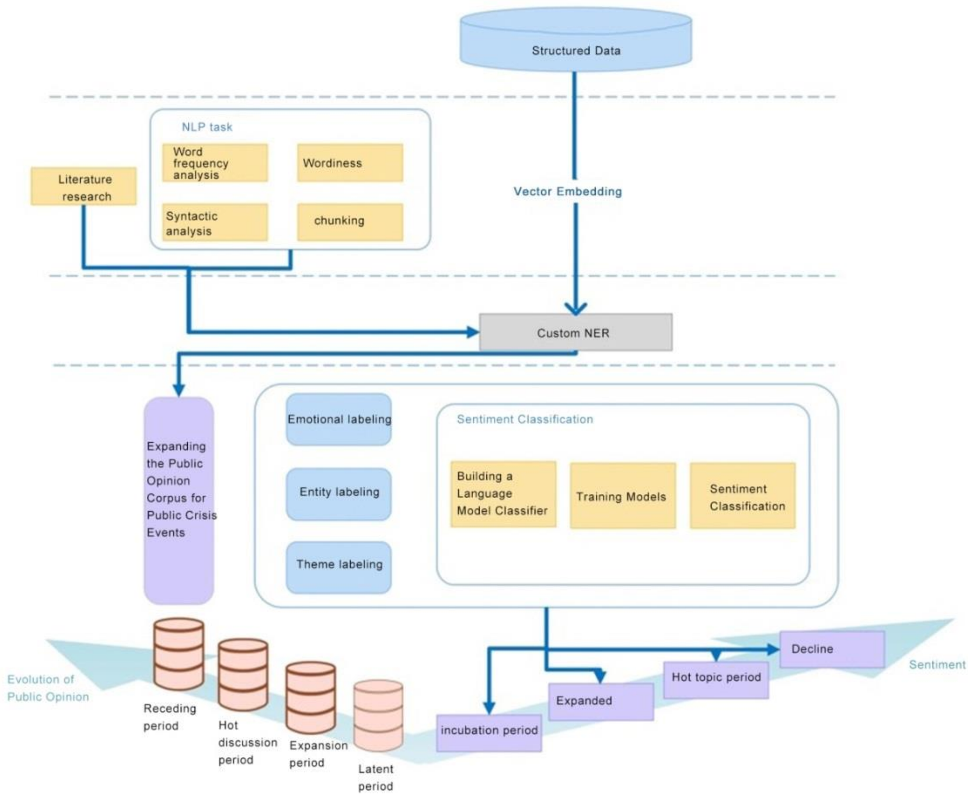

Fig. 2. Research content diagram

Public opinion risk management based on traditional media can generally be led by the government. The government has a variety of deployable government resources and administrative means to intervene and manage public opinion risks. Compared with traditional media, social media breaks the constraints of time and space and turns traditional information audiences into the role of actively selecting and producing information (Users Generate Content, UGC). This convenience makes it difficult to distinguish the authenticity of online information dissemination content. The source of online information is uneven. There are countless numbers on the channel. Relevant research has shown that online speeches tend to be bolder and more emotional. In 2003, American scholar Sunstein also proposed the phenomenon of emotional polarization of online public opinion groups in the book “Network Republic·Democracy Issues in Network Society”. When mixed information is spread, subjective and emotional remarks are also included. In emergencies of public crises, emotional speech is particularly obvious. For example, during the “7·20” heavy rain disaster in Zhengzhou, Henan Province in 2021, the local government late reported and concealed the number of people who died and were missing due to the disaster, which triggered online public opinion. The 5000 data on the microblogging and Zhihu platform are collected using common microblog business crawling software (such as Gisocket, Octopus and others). Then the text data processing is conducted, including Chinese word segmentation, feature word extraction, sentiment word extraction, text vector construction and sentiment calculation. The Chinese word segmentation is carried out using the jieba word segmentation tool based on Python. The platform obtained a word cloud diagram as shown in Figure 3. Preliminary analysis found that high-frequency words such as “man bao” (concealing reports) , “wen ze” (accountability) , “ren huo” (man-made disasters) , “guan yuan” (officials) and other entries all showed negative emotional tendencies; In April 2022, during Shanghai’s epidemic prevention and control work, “Disputes between community and medical staff in Pujin Street, Shanghai” triggered online public opinion. After collecting about 6,000 data on Microblog and Zhihu platforms, the word cloud diagram was obtained as shown in Figure 4. Preliminary analysis found that high-frequency words such as Entries such as “you qian” (rich) , “muo huan” (magic) , “sheng pa”(afraid) , “zhi neng” (only) and “xiang xia ren” (countryman) all showed negative emotional tendencies.

Figure 3 Word cloud of public opinion on the government’s concealment of the number of deaths and missing persons during the “7·20” incident in Zhengzhou, Henan

Figure 4 Word cloud of public opinion on the Shanghai epidemic prevention incident "Disputes between community and medical staff in Pujin Street, Shanghai"

Looking at the various forms of media platforms, the rapid development of instant messaging on mobile smart terminals has enabled the rapid spread of emergency text messages. Social platforms such as Microblog, WeChat, Tencent QQ, Zhihu, Douyin, and Douban rely on the relationship network to rapidly spread emergencies. Information has also become an important platform for the public to promote the development of public opinion. Professional forums in various industries focus on discussing emergencies, showing obvious cluster effects and celebrity effects. When the government fails to respond to public opinion promptly and correctly, online public opinion may quickly turn into a public opinion crisis, and then evolve into offline mass incidents that threaten social security. Data from the 50th "Statistical Report on China's Internet Development Status" released by the China Internet Network Information Center (CNNIC) showed that as of June 2022, the number of Internet users in China had registered 1.051 billion, and the Internet penetration rate had reached 74.4%. China's's netizen groups have built a huge online public space with a huge group base, which has also made social media risk management under emergency situations an important part of emergency management research. The academic and industrial circles have focused on the online public opinion of emergencies. Research on risk monitoring, risk identification, and risk early warning is also gradually advancing.

This paper was based on previous studies on the life cycle of public opinion on the Internet during emergencies. After public opinions enter the hot discussion period, due to the complexity and dynamic changes of public opinions, the event public opinions often go in different directions. It divided the evolution of public opinion into an embryonic period, an expansion period, a hot discussion period, a fluctuation period and a fading period. It was based on the theory of human behavior and the characteristics of risk management of sudden public crisis events. At the emotional level, a language model of the Transformer framework was built and the emotional tendencies of the five-stage perception states were analyzed. In the self-attention mechanism module, the self-attention mechanism first initializes the query matrix (Query, Q), key matrix (Key, K) and value matrix (Value, V). Each set of Q, K, V is called a head, and there can be multiple heads, which is also the source of the name multi-head attention. Assuming the input matrix is S, the calculation of the matrices mentioned above is as shown in equation (1).

$$
\begin{aligned}
Q &= XW^{Q} \\
K &= XW^{K} \\
V &= XW^{V}
\end{aligned}
\tag{1}
$$

Among them: W^Q, W^K, W^V are transformation matrices.

In order to better capture the effect of aspect words on the emotional polarity of the text, the model models the aspect word vector and the text context vector with different attention modules, respectively, the modeled eigenvector is summed and averaged as a vector (Query, Q) for calculating the next round of attention, specifically, the context eigenvector for attention computation uses the aspect word vector of the Query, and vice versa. After this round of attention calculations, the two vectors are spliced together and the emotional polarity is calculated using Softmax.

It is found that during the evolution of public opinion events, some subject affective values change from positive to negative, some subject affective values change from negative to positive, and some subject negative affective values continue to increase. For the governance of public opinion, we should pay special attention to those themes where the polarity of emotions changes from positive to negative or continues to be negative. These themes are the main expressions of negative emotions in public opinion events, it shows that when people participate in commenting on these topics, they show negative emotions such as sadness and anger, which will lead to the expansion of public opinion.

5 Conclusion

Emotional tendency is one of the most valuable data of online public opinion, revealing the development trend of online public opinion. Public emergencies are crisis events that may endanger public security and normal order due to natural disasters and the failure of social operating mechanisms in the process of social operation. Such events often threaten the basic values of the social system and the code of conduct of the social system. structure, etc. The governance of such emergencies requires the government to make key decisions under extremely high time pressure and uncertainty. Compared with other types of emergencies, netizens are more vulnerable to public crisis events. The psychological demands for security, trust, fairness, and justice are more urgent, which poses a challenge for the government to quickly and accurately identify risks, predict risks, and intervene in risks under the circumstances of limited governance means, limited network resources, and limited response time.

After preliminary research on the data, based on the emotional characteristics of classic cases of sudden public crises, the public opinion corpus of public crisis events was expanded in various types of online public opinions and the existing risk identification of online public opinion and early warning indicator system was improved, making its risk management of online public opinion on sudden public crisis events universal, thereby greatly enhancing the monitoring and prediction of online public opinion.

Acknowledgments. This work was supported by national innovation and entrepreneurship training program for college students S202310497007 Research on Accurate Identification and Early-warning Model of Public Opinion Risk in Social Media Network under Public Crisis

References

[1] Wen Zhitao, Xia Yixue. *Analysis of emergency network public opinion situation awareness based on evolutionary modeling* [J]. Journal of Intelligence , 2022, 41(09): 71-78.

[2] Ma Teng, Yin Yue, Zhao Shukuan, Li Jiawei. *Research on the evolution characteristics of online public opinion on public health emergencies based on multi-dimensional data fusion* [J]. Information studies: Theory & Application, 2022, 45(12): 170-177.

[3] Wang Lin, Zhang Mengxi, Wu Jiang. *Research on the spread and evolution analysis of online public opinion on the COVID-19 epidemic from the perspective of information ecology* [J]. Information Science, 2022, 40(01): 31-37+50.

[4] Huang Shijing, Wu Chuanhui, Yuan Qinjian, Xia Jingran. *Research on spatiotemporal evolution differences of public opinion on public health emergencies based on sentiment analysis* [J]. Information Science, 2022, 40(06): 149-159.

[5] Lian Zhixuan, Lan Yuexin, Xia Yixue, Liu Mo, Zhang Shuangshi. *Research on multi-dimensional dynamic classification and prediction model of online public opinion based on big data* [J]. Journal of Intelligence, 2018, 37(05): 123-133 +140.

[6] Chen Peiyou, Hou Tiantian. *Research on social network public opinion risk early warning based on ANP-gray fuzzy - taking the "Chongqing Bus Falling into the River Incident" as an example* [J]. Information Science, 2019, 37(05): 115-120 .

[7]Nemes L, Kiss A. Information Extraction and Named Entity Recognition Supported Social Media Sentiment Analysis during the COVID-19 Pandemic. Applied Sciences. 2021,11(22):11017.

[8] Cai Yang. *Research on methods of sentiment analysis of user comments based on in-depth learning* [D]. Zhejiang Sci-Tech University, 2022.

[9] Wang Gang. *Research and application of Chinese text sentiment analysis methods based on transfer learning* [D]. University of Electronic Science and Technology of China, 2022.

[10] Yadav R K, Jiao L, Goodwin M, et al. *Positionless aspect based sentiment analysis using attention mechanism* [J]. Knowledge-Based Systems, 2021,226:1-10.

[11] He Jiejun, Li Yang. *Research on the emotional evolution characteristics of public opinion reversal events based on a spatiotemporal perspective* [J]. Journal of Information Resources Management, 2022, 12(02): 88-100.

[12] An Lu, Wu Lin. *Analysis of the evolution of Microblog public opinion on emergencies that integrates themes and emotional characteristics* [J]. Library and Information Service, 2017, 61(15): 120-129.

[13]Fu P, Jing B, Chen T,et al. *Modeling Network Public Opinion Propagation with the Consideration of Individual Emotions. International Journal of Environmental Research and Public Health*,2020,17(18):6681.