Monitoring the internet public opinion of mountain flood disaster using big data analysis

R J Cao\textsuperscript{1}, R G Jiang\textsuperscript{1,2}, J C Xie\textsuperscript{1}, J Wang\textsuperscript{1}, T L Shi\textsuperscript{1} and C C Wen\textsuperscript{1}

\textsuperscript{1}State Key Laboratory of Eco-hydraulics in Northwest Arid Region of China, Xi’an University of Technology, Xi’an 710048, China

E-mail: jrengui@163.com

Abstract. The mountain torrent disasters have caused severe damage and loss of life and property to the society and the economy, which attract the public’s attention. Timely monitoring the characteristics and evolution of internet public opinion in mountain torrents is of significant impact on the emergency response and management. The primary objective of the paper is to analyze the time and spatial characteristics, attention key points, emotional trend and participants of the public opinion in mountain torrent disasters using big data and statistical analysis. The current situation is investigated and a prediction method using correlation analysis is proposed, which provides reference for decision makers to formulate response plans according to public opinion from internet. The data mining method is used to obtain relevant data based on microblog API interface, and the data is processed by statistical methods. Combined with the content analysis method, the characteristics of public opinion in mountain torrents are comprehensively analyzed, and the hotspots of public concern and the trends of public opinion are obtained. The results show that the eruption time of the public opinion coincides with the disaster occurrence time. From spatial characteristics analysis, the flood disasters in China mainly occur in the North China Plain, the middle and lower reaches of the Yangtze River, the Pearl River Basin, the Sichuan Basin, and the Northeast Plain. The public opinion attention degrees are highly positive correlated with the occurrence of mountain torrent disasters, and the correlation coefficient is 0.96. Citizens not only pay attention to the disaster situation and rescue work, but also focus on the improvement of the disaster prevention work. Public emotional trend is mainly based on expressing hope for a safe life, pity and sympathy for the victims, panic about the severity of the disaster, objective evaluation of the situation, as well as admiration and concern for rescuers, and suggestions on early warning education. As an opinion leader, People Daily has great influence on the internet public opinion. At different stages, participants should work together to maintain a healthy public opinion environment.

1. Introduction

Owing to its complex geography and climate, China is one of the countries most seriously threatened by floods [1]. Since 2014, there have been large-scale flash floods in China every year. Representatively, a flash flood occurred in Yunyang County and Fengjie County of Chongqing Province claimed 52 lives on September 1, 2014, while events on July 9 in Yunlong County and Fugong County of Yunnan Province claimed 31 lives [2]. Minqing county, Fujian Province caused a historical flood, with 95,000 people affected, 39 deaths and direct economic losses of more than 2.17 billion yuan on July 9, 2016 [3]. However, avoiding the occurrence of such a heavy loss of life and property from mountain torrents is a significant issue that valued by people.
With the rapid development of Internet technology recently, various types of network platforms have become the main channels for the public to express their opinions and obtain information. As the main social platform in China, Sina microblog’s public opinion big data under this platform has become the main carrier of public opinion analysis. In view of the research on monitoring internet public opinion analysis, domestic and foreign scholars have achieved fruitful results. Jiang et al proposed a new system for hydrologic monitoring, simulation and visualization, which can be used to analyze and simulate disasters including mountain torrents [4]. Cambria et al discussed the past, present, and future trends of sentiment analysis by delving into the evolution of opinion mining systems [5]. Okazaki et al employed data mining techniques to get Tweets about IKEA in order to identify social networks created by prosumers [6]. Bai et al proposed a structured framework to detect potential incidents implicated by victims’ negative emotions in the post-disaster situation, which is meaningful and useful to disaster relief process and emergency [7]. Zhang et al analyzed the data collected from microblog and WeChat of an actual news event to visualize the information spread process in the cross-network dissemination case of public opinion [8]. Nan et al studied the characteristics of microblog’s communication and the process of public opinion formation, the mechanism of communication, and so on, which provided a reference for the effective, intelligent, and real-time detection of microblog public opinion [9]. In recent years, through the analysis of big data public opinion to grasp the public’s comments on cases and the development trend of events has become an important basis for decision makers to formulate plans. Although, the studies on the public opinion in emergencies have been already developed, analysis specifically for mountain torrents has not been investigated well. Therefore, it is very necessary to analyze characteristics of internet public opinion in mountain torrents.

It is of great significance to understand the characteristics of public opinion in mountain torrent disasters. This paper adopts the data mining technology based on microblog API interfaces to obtain the public opinion data of mountain torrents during 2014-2018, and combines the content analysis method to show the current status of mountain torrents public opinion. At the same time, it uses correlation analysis to study the development trend, and provides reference for a series of work such as mountain torrents warning and forecasting.

2. Data and methods

2.1. Data collection

Based on the microblog API interfaces, the data of the public opinion in mountain torrent disasters are collected. According to the key word “Mountain torrent”, we got a number of microblog contents, the release time, the user name, the user type, the followers count, and the text-related content, such as comments, reposts, and attitudes.

Baidu Index (https://index.baidu.com) is an online search tool that can reflect the public concern. The data is available for users to view how frequent the specific keywords, subjects and phrases have been queried with different time units including daily, weekly, monthly or yearly [10]. The paper will use Baidu search index to analyze the time characteristics of public opinion in mountain torrent disasters.

Sina microblog provides API (Application Program Interface, https://open.weibo.com/wiki) for third parties developers to use. Through Sina microblog API interfaces, it can realize the social access to multiple types of terminals, introduction of external website, commercialization services, and acquisition of microblog users’ data. Users can obtain data through different interfaces according to requirements.

2.2. Method

The obtained text data contains a large amount of data redundancy and noise information, which seriously interfere with the study of text content. Data suitable for analysis is obtained by preprocessing the text, including noise removal, word segmentation, and removal of stop words.
3. Results and discussion

3.1. Time characteristics of mountain torrent disasters

According to the data analysis of the Baidu search index, the public search of the keyword “mountain torrents” continued from 2014 to 2018. The average annual search index is 208. The peak search period was concentrated in flood season of July and August. In China, the flood seasons are mostly from May to September, and more frequent from June to August. Therefore, the peak of search period coincides with the occurrence of frequent mountain torrents. From July 4, 2016 to July 10, 2016, the peak of search index reached 861 (point B in figure 1). In July 4, 2016, a flood event with long process and large intensity in historical record over 30 years occurred in the Huaihua city, Hunan Province. It resulted in 24183 acres of county’s crops destroyed and 491 houses collapsed. With the number of 486,000 affected people, it caused widespread public concern. The search index at point A is 703, reaching the peak of search in 2015. On August 5 in Xi’an, a sudden flash flood washed away a family of 8 while they were having a birthday dinner. The strong destructive power of flood caused the public to panic and showed sympathy for the victims.

Figure 1. Baidu search index trend of “mountain torrents” during 2014 – 2018.

3.2. Spatial characteristics of mountain torrent disasters

In order to analyze the relationship between the hotspot areas and the mountain torrent places, the attention degree of provinces and the number of mountain torrent events in 2017 are used for correlation analysis. The attention degree comes from the frequency of provinces in the microblog text data about mountain torrents in 2017. The number of mountain torrent disasters originated from the
geological disasters and prevention situation in China Statistical Yearbook-2018, and a total of 20 provinces data are collected in table 1.

**Table 1.** Statistical table of attention degree and number of mountain torrents in provinces.

| Province   | Hubei | Chongqing | Guangxi | Hunan | Yunnan | Jiangxi | Sichuan | Guizhou | Shanxi | Jilin |
|------------|-------|-----------|---------|-------|--------|---------|---------|---------|--------|------|
| Attention degree | 925   | 402       | 367     | 330   | 299    | 290     | 264     | 217     | 198    | 181  |
| Number of disasters | 752   | 454       | 353     | 307   | 283    | 253     | 140     | 101     | 71     | 57   |

After linear fitting (figure 2), the regression analysis is carried out on attention degree and the number of mountain torrent disasters. The Pearson correlation coefficient can reflect the linear correlation between two variables. According to the Pearson correlation analysis, the correlation coefficient is 0.960, indicating that the number of mountain torrents is positively correlated with the heats mentioned by the public at *% significance level. At the same time, the significance of p value is less than 0.01 indicates that there is a very significant correlation between them. Furthermore, the regression equation can predict the degree of public opinion after the occurrence of mountain torrents.

**Figure 2.** Correlation analysis between public opinion heat and mountain torrents.

The mountain torrent disasters in China mainly occur in the North China Plain, the middle and lower reaches of the Yangtze River, the Pearl River Basin, the Sichuan Basin, and the Northeast Plain [11]. According to the results, these areas have large rainfall during the flood season, and frequent disasters have caused a series of public opinion. The level of public opinion attention degree is higher than in other regions.

**3.3. Key points of mountain torrent disasters**

After the word segmentation and noise removal of the microblog text data, the keywords of the mountain torrent disasters are obtained. The keywords can intuitively display the public’s attention hotspots and the orientation of public opinion. The key words about mountain torrents of microblog are “Disaster”, “Storm”, “Debris flow”, “Death”, “Missing”, “Casualty”, “Direct economic loss” and names of multiple regions. The reports of mountain torrent disasters mainly focus on disaster news,
and spread the disaster situation to the public in real time. At the same time, the emergence of words such as “Rescue”, “Disaster relief”, and “Security” indicates that the dissemination of disaster relief work is also attracting public attention. In addition, vocabulary related to disaster prevention work such as “Flood prevention”, “Emergency”, “Monitoring” and “Safety education” also become key points of mountain torrent disasters.

The keywords can be divided into four aspects: disaster situation, hot area, disaster relief and disaster reduction. In figure 3(a), microblog’s public opinion most related to the theme of disaster situation, accounting for 65.88%. The real-time tracking report of the disaster situation is the most concerned by public, among which casualties and direct economic losses are the focus of attention. In figure 3(c), the mountainous areas such as Sichuan and Yunnan Province, due to the frequent occurrence of mountain torrents, 35% of the public opinion reports are related to these areas. In 2018, the occurrence of mountain torrents in Malipo County, Yunnan Province has been greatly affected and seriously damaged, which is widely discussed by netizens, accounting for 11% among hot areas. The impact of the disaster has made the public aware of the importance of emergency management and rescue work. In figure 3(d), 12% key words of disaster reduction are related to emergency drills. The public gradually pays attention to disaster control and reduction work, which further indicates that improving the accuracy of monitoring and forecasting, enhancing the ability of early warning are an important direction that needs to be improved and developed.

![Figure 3](image.png)

**Figure 3.** Key points of mountain torrent disasters. a Classification of keywords; b Key words of disaster situation; c Key words of hot area; d Key words of disaster relief and reduction.

3.4. Public opinion emotional trend of mountain torrent disasters
Study on emotional characteristics aims to analyze the public opinions, feelings, attitudes and moods in text and comments [12]. After filtering the obtained microblog data and extracting the words with emotional trend in hot comments, the emotional topics can be classified. It can visually show the
public attitude and emotional characteristics of disasters. The main attitudes of the public towards mountain torrents can be divided into six categories: panic about the severity of the disaster, pity and sympathy for the victims, hope for a safe life, objective evaluation of the disaster, as well as admiration and concern for the rescuers and suggestions for disaster reduction work. In figure 4, the public is mostly worried about the disaster’s situation, which caused major losses, and shows earnestly hope that the natural disasters will be reduced and the people will be safe. The words in high frequency of the emotional are “Peace”, as well as “Hope”, “Pray”, “Blessing” and other words that represent the people’s strong emotional desires. Adjectives words such as “Scare”, “Terrible” and “Horrible” express the people’s panic after the occurrence of the mountain torrent disasters, and the words “Rescue”, “Tribute”, “Caution”, “Attention”, “Refueling” and other words indicate the people’s concern for the rescue work and the personal safety of the rescuers. In addition, the emergence prevention words such as “Reduction of disasters” represents that the public starts to focus on the prediction, forecasting, early warning and disaster reduction of mountain torrents.

Figure 4. Classification of emotional trend.

3.5. Participants of the public opinion in mountain torrent disasters

In the process of internet public opinion, the participants mainly include three types: public opinion triggers, makers and responders. The triggers of public opinion mainly issue the event information. The makers refer to the individuals or organizations that express opinions or take action on the event, while the responders are the government and relevant departments that take corresponding measures for public opinion [13]. According to the data obtained from microblog, the main triggers of the public opinion in mountain torrents are microblog certificate authorities, newspapers, TV stations, etc. In mountain torrent events, the main opinion leaders in the process of mountain torrent disasters are News Head, People Daily, People Net, CCTV News, Xinhua Viewpoint, City Express and other media. Their large number of followers can formulate the information spreads rapidly, which is likely to cause a public opinion outbreak. In figure 5, the influence of the above six opinion leaders is comprehensively analyzed by the number of followers, comments count, reposts count, and attitudes count on the same microblog content. The yellow color indicates a higher influence and the red color indicates a lower influence. The impact factor of six opinion leaders refers to the proportion of each influencing indicator. The influence of the City Express is slightly lower than the other five. People Daily is at a high value among the four indicators, indicating that People Daily has a high impact in the mountain torrents, and it releases more authoritative and accurate information, enough to lead the direction of public opinion.
4. Conclusions

- The outbreak period of public opinion in mountain torrents mainly detected in July and August, which coincides with the occurrence of frequent disasters. During 2014-2018, the search peak value occurred on July 4, 2016 reaching 861, and the annual search average is 208. The mountain torrents occurred frequently the North China Plain, the middle and lower reaches of the Yangtze River, the Pearl River Basin, the Sichuan Basin, the Northeast Plain and other places due to large rainfall in flood seasons. By linearly fitting the relationship between the attention degree and the occurrence of mountain torrent disasters, the correlation coefficient is 0.960, which indicates the two are highly positive related.

- From the key points and emotional trend analysis, the public’s concerns in mountain torrents mainly include three aspects. One is the public’s attention to the disaster situation in affected areas, the second is the concern for the rescue and disaster relief work, and the third is the emphasis on flood control and disaster reduction. Emotional trend can be divided into six categories including panic about the severity of the disaster, sympathy and pity for the victims, hope for a safe life, objective evaluation of the disaster, as well as admiration and concern for rescuers, and suggestions on early warning education. By analyzing the keywords and emotional trend of public opinion, it is significant to grasp the trend of public opinion.

- There are three types of participants in the whole process public opinion: triggers, makers and responders. The public opinion triggers such as certification authorities, news media, newspapers, satellite TV to spread information firstly in small areas among online media, ordinary netizens and followers. The internet ordinary users and opinion leaders as makers of public opinion to outbreak the public opinion. In mountain torrents, there are six main opinion leaders, and the most influential media is People Daily. The public opinion responders including the government and relevant departments finally cooperate to calm down the public. In the whole process, the relevant participants at different stages should respond in time to make the public opinion healthy.

Acknowledgments
The study was partly funded by the National Key Research and Development Program of China (2016YFC0401409), National Natural Science Foundation of China (51509201, 51679188 and...
71774132), Natural Science Basic Research Plan in Shaanxi Province of China (2018JM5031). The microblog data were obtained from microblog API interfaces at https://open.weibo.com/wiki. The search index data were obtained from Baidu Index at https://index.baidu.com. The authors thank the editor and anonymous reviewers for their valuable comments and suggestions.

References
[1] Chen S, Luo Z and Pan X 2013 Nat. Hazards 69 1597-605
[2] Wei L, Hu K H and Hu X D 2018 J. Mt. Sci. 15 2492-504
[3] Ma M H, He B S, Wan J H, Jia P F, Guo X R, Gao L, Maguire L W and Hong Y 2018 Water 10 704-16
[4] Jiang R G, Xie J C and Li J X 2013 Disa. Adv. 6 33-47
[5] Cambria E, Schuller B, Xia Y Q and Havasi C 2013 IEEE Intell. Syst. 28 15-21
[6] Okazaki S, Diaz-Martin A M, Rozano M and Menendez-Benito H D 2015 Internet Res. 25 416-34
[7] Bai H and Yu G 2016 Nat. Hazards 83 1177-96
[8] Zhang L F, Su C, Jin Y F, Goh M and Wu Z Y 2018 Inform. Sciences 451 240-52
[9] Nan F, Suo Y N, Jia X Y, Wu Y Y and Shan S J 2018 IEEE Access 6 76502-15
[10] He G Y et al 2018 Sci. Rep. 8 9038-47
[11] Duan W L, He B, Nover D, Fan J L, Yang G S, Chen W, Meng H F and Liu C M 2016 Nat. Hazards 82 401-13
[12] Bai H and Yu G 2016 Nat. Hazards 83 1177-96
[13] Ridings C M, Gefen D and Arinze B 2002 J. Strategic Inf. Syst. 11 271-95