Article

Using Social Media to Mine and Analyze Public Sentiment during a Disaster: A Case Study of the 2018 Shouguang City Flood in China

Xuehua Han [1,2,3] and **Juanle Wang** [1,3,4,*]

1   State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China; hanxh@lreis.ac.cn
2   College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China
3   The International Knowledge Centre for Engineering Sciences and Technology (IKCEST) under the Auspices of UNESCO, Beijing 100088, China
4   Jiangsu Centre for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China
*   Correspondence: wangjl@igsnrr.ac.cn; Tel.: +86-010-6488-8016

check for updates

Abstract: Social media has been applied to all natural disaster risk-reduction phases, including pre-warning, response, and recovery. However, using it to accurately acquire and reveal public sentiment during a disaster still presents a significant challenge. To explore public sentiment in depth during a disaster, this study analyzed Sina-Weibo (Weibo) texts in terms of space, time, and content related to the 2018 Shouguang flood, which caused casualties and economic losses, arousing widespread public concern in China. The temporal changes within six-hour intervals and spatial distribution on sub-district and city levels of flood-related Weibo were analyzed. Based on the Latent Dirichlet Allocation (LDA) model and the Random Forest (RF) algorithm, a topic extraction and classification model was built to hierarchically identify six flood-relevant topics and nine types of public sentiment responses in Weibo texts. The majority of Weibo texts about the Shouguang flood were related to "public sentiment", among which "questioning the government and media" was the most commonly expressed. The Weibo text numbers varied over time for different topics and sentiments that corresponded to the different developmental stages of the flood. On a sub-district level, the spatial distribution of flood-relevant Weibo was mainly concentrated in high population areas in the south-central and eastern parts of Shouguang, near the river and the downtown area. At the city level, the Weibo texts were mainly distributed in Beijing and cities in the Shandong Province, centering in Weifang City. The results indicated that the classification model developed in this study was accurate and viable for analyzing social media texts during a disaster. The findings can be used to help researchers, public servants, and officials to better understand public sentiments towards disaster events, to accelerate disaster responses, and to support post-disaster management.

Keywords: social media; flood; public sentiment; disaster risk reduction; China

1. Introduction

Over the past few decades, an increase in natural hazards has imposed greater challenges on natural disaster management [1,2]. To reduce the impact of natural disasters on humanity, disaster management requires more human-centric information in addition to objective disaster information. Since disaster management demands a large amount of information in the face of low availability [3], social media (e.g., Twitter, Facebook, or Sina-Weibo) is an additional information source that is gaining increasing attention from geographic information scientists and disaster researchers [4–6]. Social media

is not only a platform for sharing people's personal lives but can also be used to examine public opinion and perceptions, which may be comparable to the public comments collected by traditional approaches(e.g., questionnaires) [7–9]. Combined with spatial-temporal information collected from social media, the public opinions and feelings on a disaster mined from social media can assist government decision-making and help people better understand the state of disaster events [8,10]. However, social media texts are unstructured and people's conversational content on social media varies in topic and tone. Thus, a significant challenge remains in accurately identifying public sentiments about a disaster from social media texts.

Existing studies mainly leverage the spatial and temporal characteristics of social media to assist disaster management and response while underutilizing the rich public sentiments in disaster-related social media content. Gelernter used named entity recognition and a machine learning method to geocode non-standard geographic content (e.g., abbreviated or highly localized place name) from disaster-related tweets [11]. Bakillah utilized a clustering algorithm based on semantic similarity to detect geo-located communities on Twitter during disaster situations [12]. Herfort identified the relationship between the geographical positions of social media texts and the geographic features of flood phenomena [5]. Crooks analyzed the spatial and temporal characteristics of earthquake-related tweets to enhance situational awareness capabilities [13]. Blanford found that the temporal changes of tornado relevant tweets corresponded to different stages of a tornado [14]. Some studies leveraged flood-related information (e.g., locational information, depth of water) from social media to support rapid inundation mapping [6,15–17]. In addition to research that highlights the spatial and temporal characteristics of social media, other scholars mined topics and created classifications from social media texts to improve knowledge about disaster situations. Imran built human-annotated Twitter corpora collected during 19 different crises to train machine learning classifiers to classify informative tweets and extract disaster-relevant information [18,19]. Wang investigated people's discussions on wildfire-related twitter feeds by identifying important terms [20]. Ye classified Weibo related to "dengue" into five categories and analyzed the relevance between the outbreak of the disease and people's discussion of it on Weibo [21]. Wang divided Weibo texts about the 2012 Beijing Rainstorm into five topics to investigate the distribution pattern of Weibo [22]. Zhu identified public sentimental attitudes on the "2014 Shanghai Stampede" as either positive, neutral, or negative from geo-tagged Weibo [7]. However, the extant literature on mining topics from disaster-relevant social media texts only extracts important terms or roughly divides the content into different topics and does not identify public opinions in depth.

This paper aims to expand prior research by identifying public sentiments from social media and analyzing their spatial-temporal characteristics. Sina-Weibo, often referred to as "Weibo", is one of the most popular social media platforms in China. Weibo had over 445 million active users each month in 2018. Using the 2018 Shouguang flood in China as a case, this study analyzed the Weibo texts related to the flood in terms of space, time, and content. More specifically, this study analyzes the temporal changes within six-hour intervals and the spatial distribution on sub-district and city levels of flood-related Weibo. In addition, a topic extraction and classification model was built to identify topics of flood-related Weibo and uncover public sentiments in response to the flood. Using this model, we classified the Weibo texts into six topics and nine sentiments related to the Shouguang flood and analyzed their spatial-temporal characteristics.

2. Data and Methods

2.1. Study Area

The study area is Shouguang, a county-level city belonging to Weifang city, Shandong Province, China. Figure 1 shows the geographical position of Shouguang City, located in the coastal plain area of north-central Shandong Province. There are five sub-districts and nine towns in Shouguang City, with a total population of 1,084,870. The Mihe River flows through the city. The area belongs to the

warm temperate monsoon zone of continental climate, which is suitable for vegetable cultivation. Shouguang is one of China's important producers of vegetables, mainly for the Beijing market.

On 18–19 August 2018, there was torrential rain in Shouguang caused by typhoon "Rumbia." Affected by the heavy rain, three reservoirs upstream simultaneously released floodwater to protect dams from collapsing, which was considered one of the reasons for the flood. By August 20, widespread flooding submerged many villages, houses, and farmland, leaving 13 people dead and at least three people missing. Beyond the direct loss of lives, this flooding disaster also affected vegetable prices in the surrounding cities. Hence, this flood aroused widespread public attention.
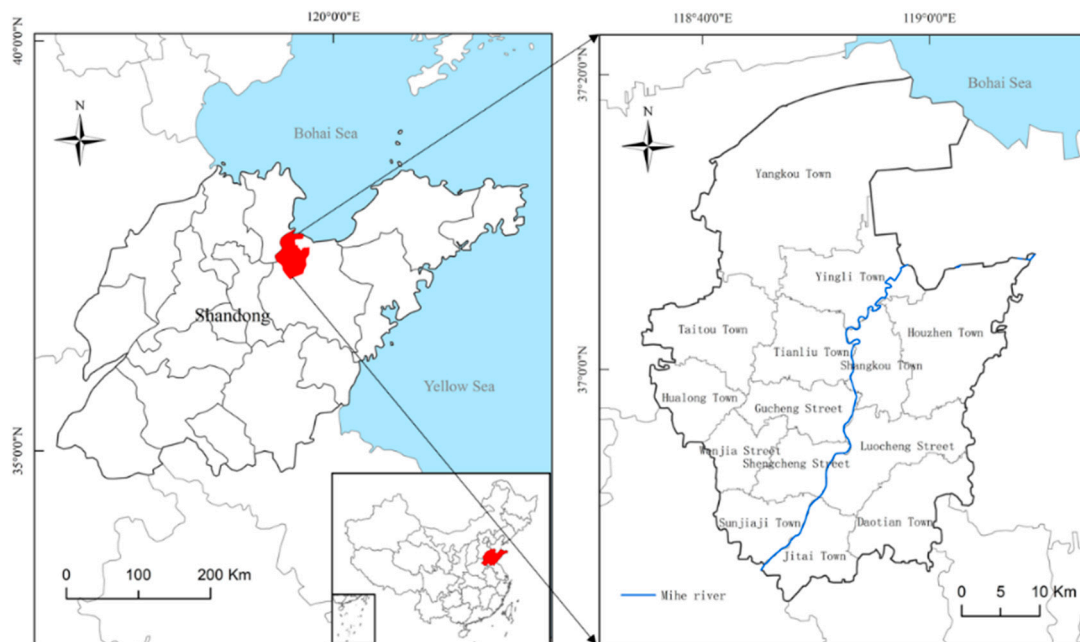

Figure 1. The administrative region of the Shouguang study area.

2.2. Data and Pre-Processing

(1) Weibo data. Weibo (http://us.weibo.com), a Twitter-like microblogging system, is the most popular microblogging service in China. Using web crawlers and Weibo API, 28,608 original Weibo messages were collected with "Shouguang" as the keyword with timestamps between 00:00 on August 19 and 00:00 on August 28. The following information was extracted: user ID, timestamp (i.e., the time when the message was posted), text (i.e., the text message posted by a user), and location information.

(2) Data pre-processing. The original Weibo texts contain interference information, for instance, http hyperlink, spaces, punctuation marks, hashtag, and @user. Text filtering is necessary to eliminate noise and improve the efficiency of word segmentation. These types of interference information were removed by the regular expression operations ('re' module) in Python. Very short Weibo texts (less than four words) and duplicated Weibo texts were deleted. That left 26,963 Weibo messages, including 4528 texts with location information, 2088 of which were located in Shouguang.

2.3. Method

2.3.1. Time Series Analysis

Time-Series analysis of the Weibo texts was necessary to investigate the temporal diversification of the amount of Weibo texts at all the stages of the flood. The overall temporal trend of the Weibo texts related to the Shouguang flood was analyzed by Seasonal-Trend decomposition procedure based on Loess (STL). STL was utilized to extract a seasonal trend from the time series of the number of

flood-related Weibo, using statistical product and service solutions (SPSS) software. As expressed in Equation (1), time series can be considered the sum of three components: a trend component, a seasonal component, and a remainder in STL:

$$x_t = T_t + S_t + R_t. \tag{1}$$

Here, x_t is the original time series of interest. T_t is the trend component. S_t is the seasonal component. R_t is the residual component.

2.3.2. Topic Extraction and Classification

A topic extraction and classification model, combining the Latent Dirichlet allocation (LDA) model and the Random Forest (RF) algorithm, was used to process flood-related Weibo texts. The first step was to mine and generalize the topics from the flood-related Weibo sample using the LDA model. Then, the topic extraction results were utilized as training samples for the RF algorithm to classify the whole Weibo data. The flood-related Weibo were generalized into six topics: "weather warning", "traffic conditions", "rescue information", "public sentiment", "disaster information", and "other." A secondary classification was implemented to divide the broad topic of "public sentiment" into nine more detailed sentiments, including "the disaster situation", "questioning the government and media", "seeking help", "praying for the victims", "feeling sad about the disaster", "making donations", "thankful for the rescue", "worrying about vegetable prices", and "other". Some additional steps (e.g., Chinese word segmentation) were needed to implement the model, which are shown in Figure 2.
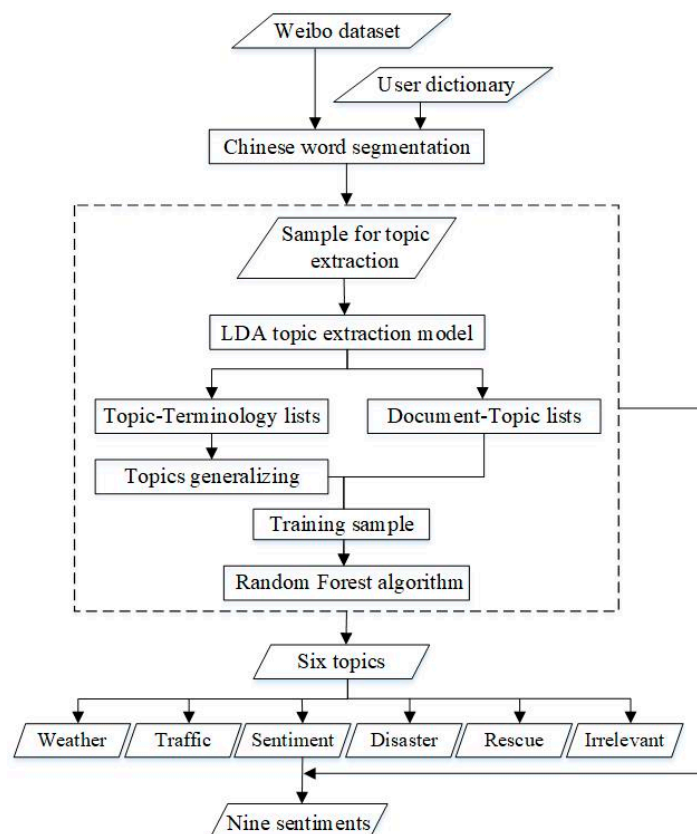

Figure 2. The processes of the topic extraction and classification model.

(1) Word Segmentation

Chinese word segmentation was necessary after the data acquisition and pre-processing because there is no obvious separator between Chinese words. A Python package for Chinese text segmentation,

called "Jieba" was utilized. By building a user dictionary including keywords related to the Shouguang flood and names of the administrative places of Shouguang City, the package segmented words efficiently. After this process, stop words were removed as they are the most common words and lack valuable information.

(2) The LDA model

The topics of the Weibo texts were extracted by the LDA model. LDA is a Bayesian probability model that has three layers of "document-topic-word" [23], with which to identify the semantic topic information in large-scale document sets or corpus. In LDA, documents are represented as random mixtures over latent topics, each of which is characterized by a distribution of words [24]. This unsupervised machine learning technique has recently emerged as a preferred method for working with large collections of text documents.

The "Gensim" package in Python was used to implement the LDA model. The optimal number of initial topics was 20, for the LDA model repeated experiments. The topic-terminology lists and the document-topic lists were obtained from the LDA model. The topic-terminology lists are the vocabularies of each initial topic and the frequency with which those vocabularies occur. The Document-Topic lists show the probability that each Weibo text is associated with each of the initial 20 topics. We assigned each Weibo text to the topic it most closely resembled according to the probabilities in the document-topic lists. Based on the topic-terminology lists, the initial 20 topics were generalized into six ("nine" in the secondary classification) by merging similar topics and discarding the irrelevant ones.

(3) The RF algorithm

The RF algorithm was used to classify the Weibo texts into different topics. Random forests are a combination of tree predictors, wherein each tree depends on the values of a random vector sampled independently and there is the same distribution for all trees in the forest [25]. The RF classifier is considered a top-notch supervised algorithm in a wide variety of automatic classification tasks [26].

The RF algorithm was implemented by a machine learning package named "scikit-learn" in Python. Based on the document-topic lists obtained from the LDA model, 3000 annotated Weibo texts were used as training samples and 600 annotated Weibo texts were used as test sets. The number of classification trees (n_estimators) was an important parameter for the classification accuracy [24]. We used the out-of-bag (OOB) outputs to determine the optimized values of the parameters to 200.

2.3.3. Evaluation of Results

Precision, recall, and the F1-measure were used to evaluate the classification accuracy. Precision is the fraction of correctly classified positive items among the classified items. Recall measures the proportion of actual positives that are correctly identified. The F1-measure is a weighted harmonic mean of precision and recall. Higher values of the F1-measure indicate that the classification method is more effective (Piskorski et al., 2013). Precision (P), recall (R), and the F1-measure (F1) are defined as follows:

$$P = \frac{T_P}{T_P + F_P} \tag{2}$$

$$R = \frac{T_P}{T_P + F_n} \tag{3}$$

$$F1 = \frac{P \times R \times 2}{P + R} \tag{4}$$

where T_P is the number of correctly classified positive items. F_P is the number of wrongly classified positive items. F_n is the number of wrongly classified negatives.

3. Results

3.1. Spatial-Temporal Analysis

Figure 3 shows the results of the time series analysis of the flood-related Weibo texts. Figure 3a is the original time series of the number of flood-related Weibo, which began to increase slowly on August 21 and then rose sharply on August 23. Split by day, it shows that the lowest point on the curve for each day appeared at 04:00, and then the curve began to rise sharply. Figure 3b shows the part of the cyclical change in the number of flood-related Weibo. The lowest point occurred at 05:00 every day and there were two daily peaks around 10:00 and 22:00. Figure 3c is the seasonally adjusted time series, which shows the trend of the number of flood-related Weibo after eliminating the seasonal factor. Figure 3d shows the trend component, which reflects the trends of the number of disaster-related Weibo. After the flood occurred, a slight increase appeared for a short time on August 21 and then the amount increased substantially on August 23. The fluctuation reached a peak at 04:00 on August 24 and it began to consistently decline after August 27.

Figure 4 shows the sub-district spatial distribution of the Shouguang flood-related Weibo. The Weibo were mainly concentrated in the south-central and eastern parts of Shouguang. The sub-district with the highest number of flood Weibo was the Shengcheng Sub-district, which is the political, economic, cultural, and information center of Shouguang. The Luocheng Sub-district, located to the east of the Shengcheng Sub-district, had the second highest number of flood-related Weibo.

The city-level spatial distribution of flood-related Weibo was mapped for the surrounding disaster area and included Beijing City, Tianjin City, Shandong Province, Hebei Province, and Liaoning Province. As shown in Figure 5, the flood-related Weibo were primarily concentrated in Beijing, Shandong Province, and Weifang City. The highest numbers of flood-related Weibo were in the cities of Weifang, Qingdao, Jinan, and Beijing.
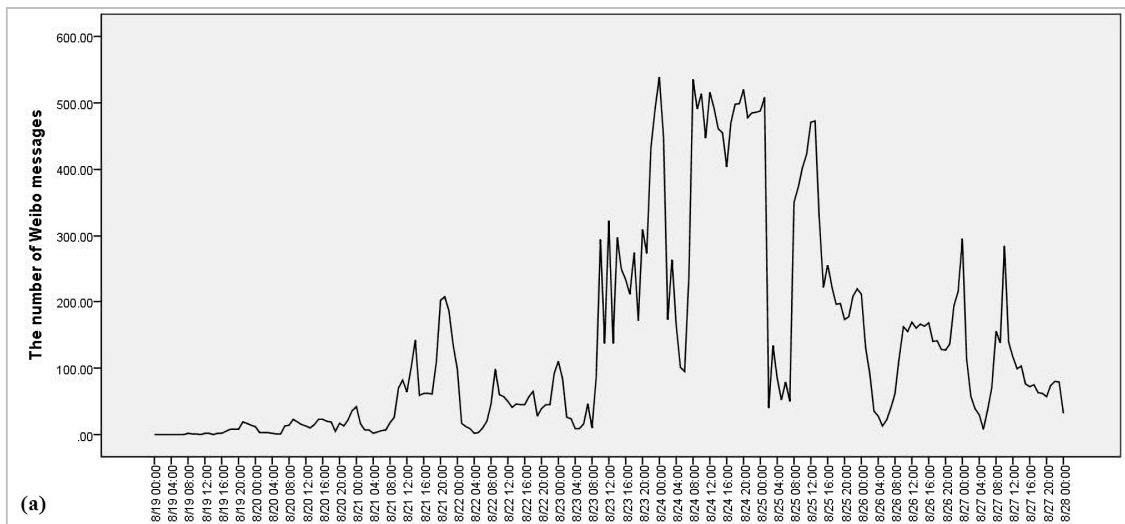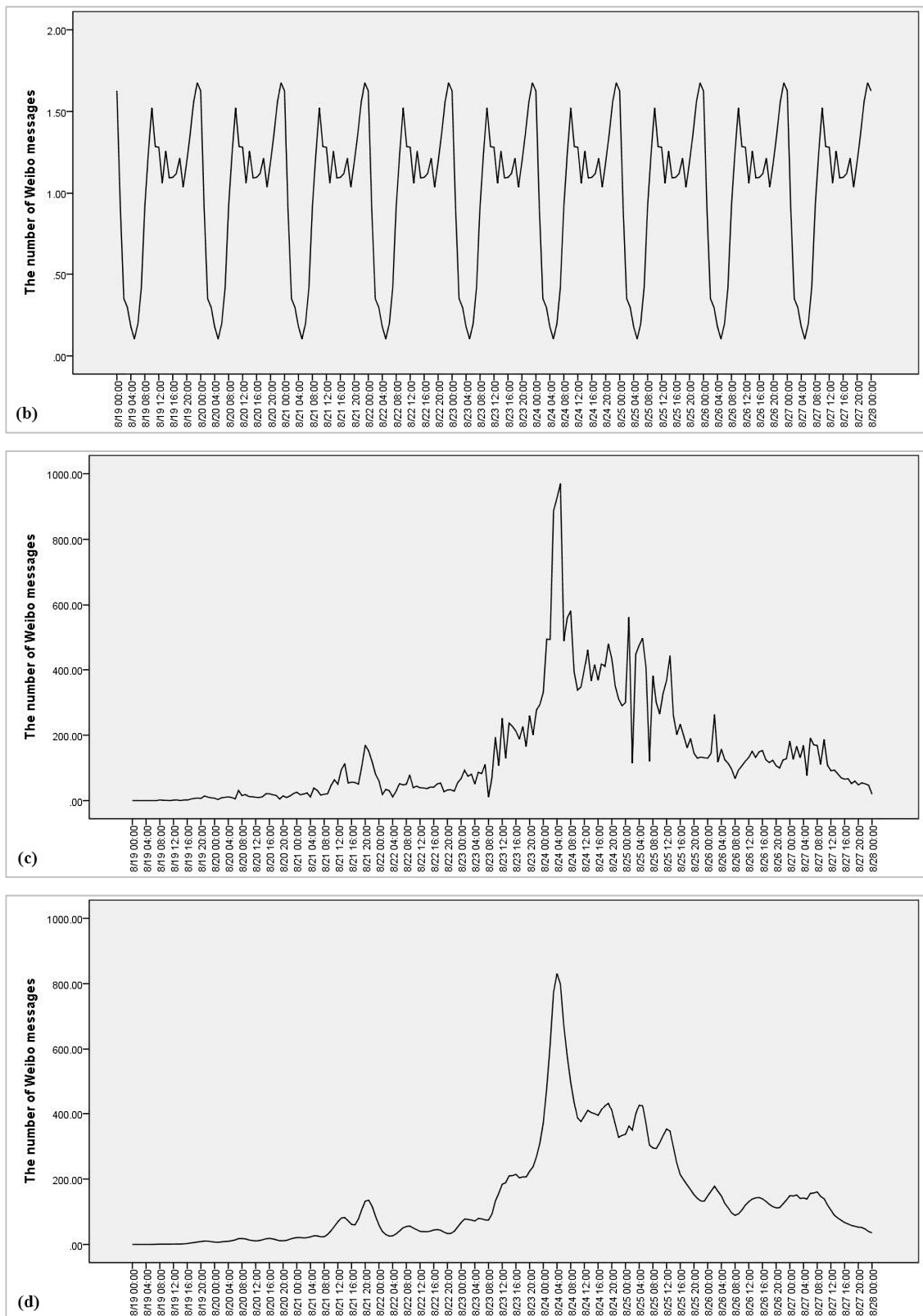

Figure 3. *Cont.*

Figure 3. The seasonal-trend decomposition of the temporal trends of flood-related Weibo. (**a**) Original temporal trend of the Shouguang flood; (**b**) seasonal component; (**c**) the seasonally adjusted time series; (**d**) trend component.
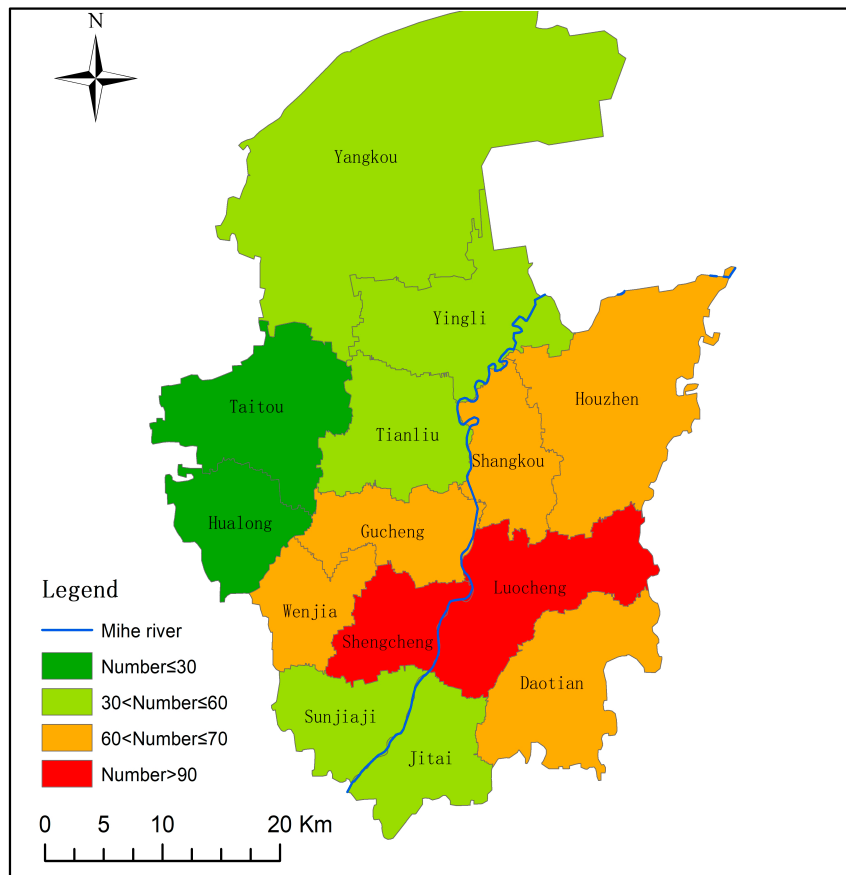
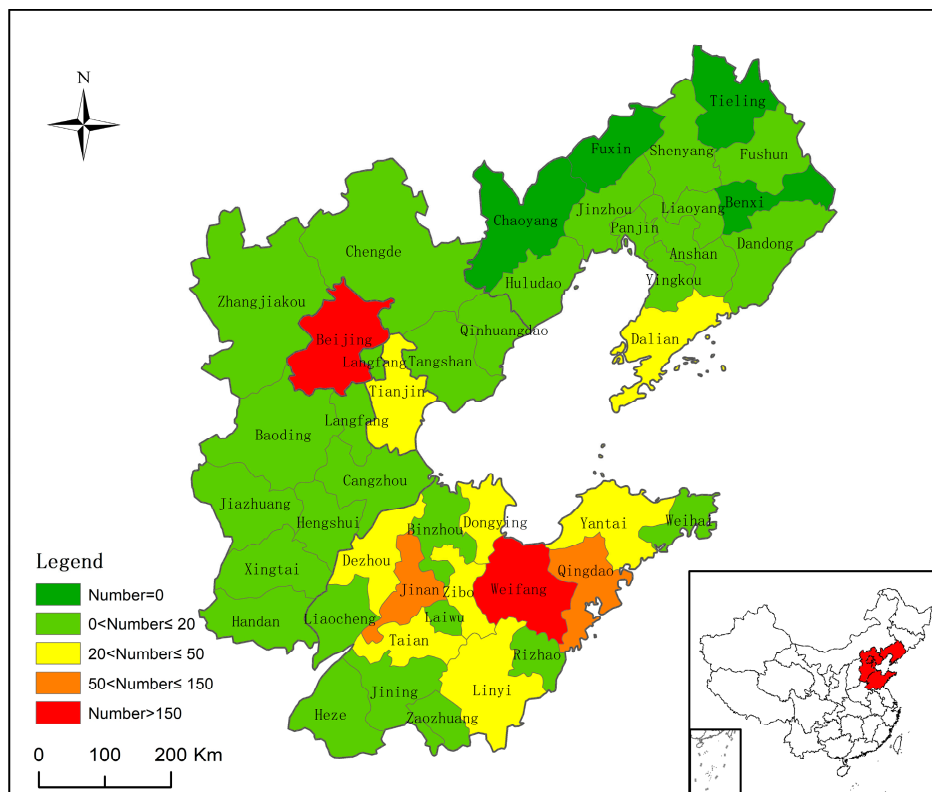**Figure 4.** The sub-district spatial distribution of flood-related Weibo in Shouguang.

Figure 5. The spatial distribution of Weibo related to the flood on a city level.

3.2. Topic Analysis

3.2.1. Topic description

Table 1 illustrates the statistical results of the six topics of flood-related Weibo, including the quantity and percentage for each topic. "Public sentiment" accounted for 84.05% of all flood-related Weibo. "Disaster information" and "rescue information" were the second and third most frequent, respectively, at 11.59% and 3.62%. "Traffic conditions", "weather warning", and "other" accounted for much lower shares (Table 1).

Table 1. Classification statistics of Weibo under different topics.

	All Texts	Weather Warning	Traffic Conditions	Rescue Information	Public Sentiment	Disaster Information	Other
Number	26,963	53	89	976	22,662	3124	59
Percent	100%	0.20%	0.33%	3.62%	84.05%	11.59%	0.22%

A more in-depth analysis of public sentiment is presented in Table 2, including the frequency and percentage for each sentiment. "Questioning the government and media", "praying for the victims", and "making donations" were the three most widespread public sentiments during the Shouguang flood. The sentiment "questioning the government and media" which challenges the flood discharge decision, feels the government was responding too slowly or considers the media reports were untimely, accounted for 61.81% of the Weibo texts. The proportion of "praying for the victims" and "making donations" comprised 15.27% and 9.13%, respectively. The proportion of other topics was less than 5 % (Table 2).

Table 2. Classification statistics of nine sentiments.

Topic	Number	Percent
Concerned about the disaster situation	250	1.10%
Questioning the government and media	14,007	61.81%
Seeking help	401	1.77%
Praying for the victims	3461	15.27%
Feeling sad about the disaster	844	3.72%
Making donations	2068	9.13%
Thankful for the rescue	480	2.12%
Worrying about vegetable prices	1042	4.60%
Other	109	0.48%
Total	22,662	100.00%

Through computing precision, recall, and F1-measure values, Table 3 shows the classification accuracy of the topics and sentiments. For the six topics, the precision was found to be 89% and the F1 was 88%. For the nine sentiments, the precision and F1 values were 78% and 72%, respectively.

Table 3. Evaluation results of topic classification.

	Topics	Sentiments
Precision (P)	89%	78%
Recall (R)	88%	75%
F1	88%	72%

3.2.2. Temporal Trend of Topics

To display accurate temporal changes in the different topics, the number of Weibo texts for each topic and sentiment were counted using six-hour time intervals. With the exception of "other", the

time series of each topic is shown in Figure 6. The number of Weibo texts for the topics "public sentiment" and "disaster information" show great fluctuation. The texts related to "public sentiment" began to increase on August 21, rose sharply on August 23 and peaked on August 24, which is similar to the overall trend of the flood-related Weibo texts. The "disaster information" topic appeared on August 20 simultaneously with the beginning of the flood and peaked on August 24. The "weather warning" topic mainly occurred on August 19 while the "rescue information" topic was concentrated to August 26–27. On August 19–22, the number of "Traffic conditions" topics was much higher than in other time periods.
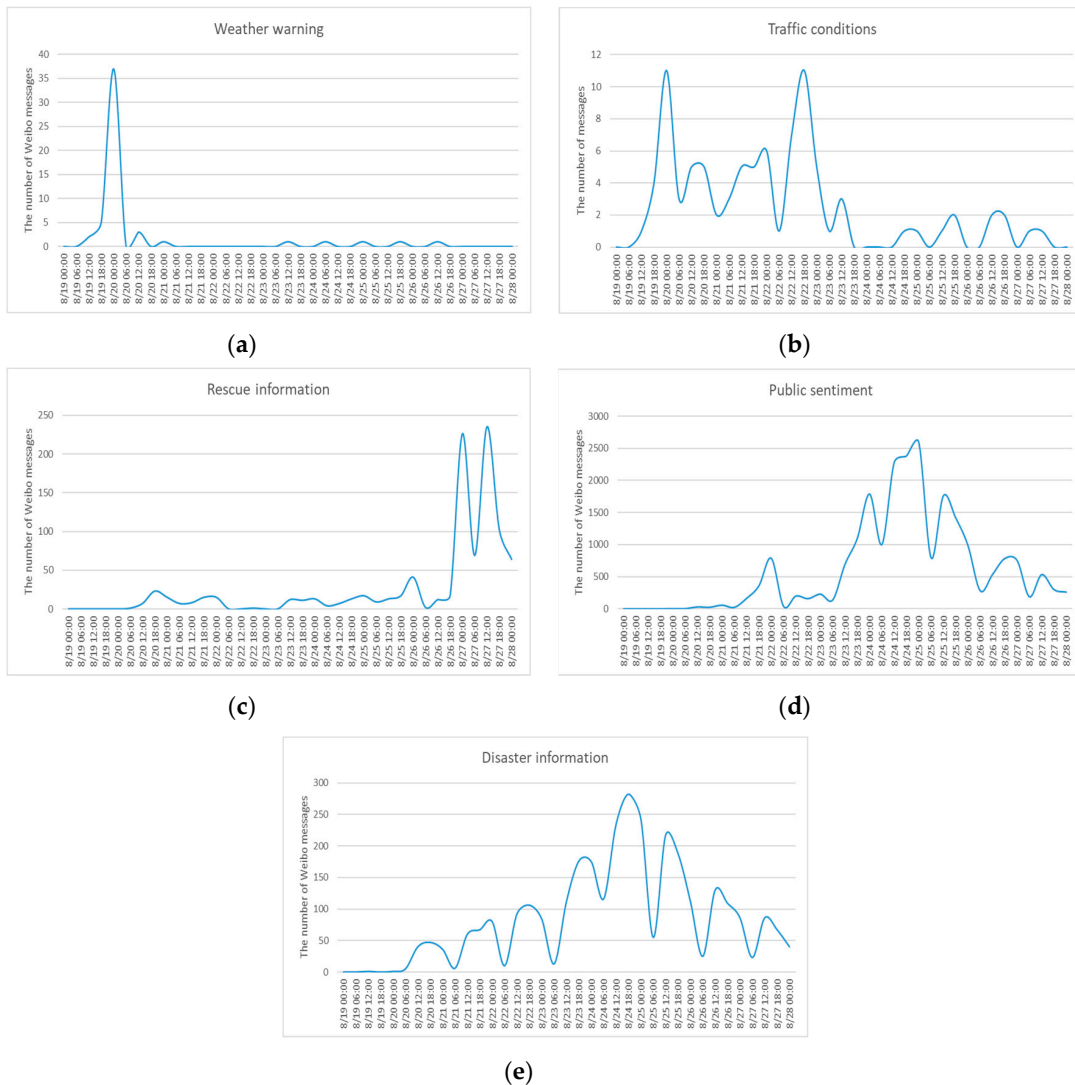

Figure 6. The trend of Weibo expressing different topics over time. (**a**) weather warning, (**b**) traffic conditions, (**c**) rescue information, (**d**) public sentiment, (**e**) disaster information.

Figure 7 presents the time series of all sentiments except "Other." From a general trends' perspective, the number of three sentiments, "questioning the government and media", "praying for the victims" and "seeking help" showed a similar variation tendency over time. The number of those four sentiments increased to a small peak on August 21, then improved quickly on August 23, and finally peaked between 6:00 and 12:00 on August 24. After the peak, the curve gradually fluctuated and decreased, stabilizing on August 26. In addition to this general trend, "thankful for the rescue" increased slowly from August 23 and peaked between 18:00 on August 26 and 0:00 on August 27. The discussion of "concerned about the disaster situation" appeared incrementally on August 21,

increased on August 23, and peaked between 6:00 and 12:00 on August 25. The sentiments of "making donations", "feeling sad about the disaster", and "worrying about vegetable prices" grew on August 23. However, the number of "worrying about vegetable prices" had two peaks on August 23 and August 24 whereas "feeling sad about the disaster" texts reached its maximum peak between 12:00 and 18:00 on August 26.
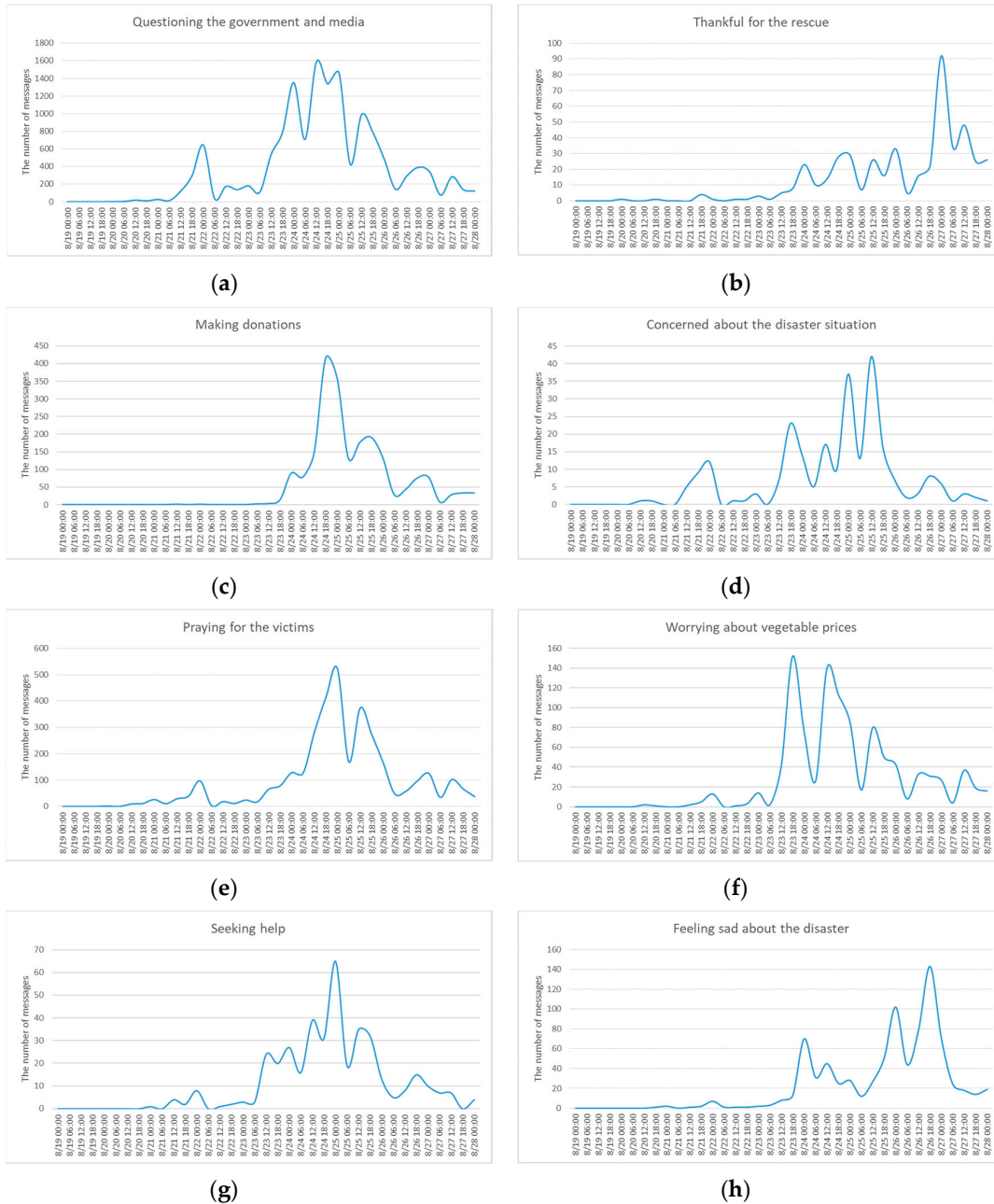

Figure 7. The trend of Weibo conveying different sentiments over time. (**a**) questioning the government and media, (**b**) thankful for the rescue, (**c**) making donations, (**d**) concerned about the disaster situation, (**e**) Praying for the victims, (**f**) worrying about vegetable prices, (**g**) seeking help, (**h**) feeling sad about the disaster.

3.2.3. Spatial Distribution of Topics

As shown in Figure 8, we selected the Weibo texts with the topics of "public sentiment", "disaster information", and "rescue information", which had the highest proportions of public concerns, to analyze the sub-district level spatial distribution. The Shengcheng and Luocheng Sub-districts are

the spatial hot spots for the flood-related Weibo texts. Figure 8a shows that the sub-districts or towns along the Mihe river have higher numbers of Weibo texts related to "disaster information" than other areas. The Weibo texts on the topic of "rescue information" (Figure 8b) are primarily distributed in south-central Shouguang. The "public sentiment" topic (Figure 8c) covers all sub-districts or towns in Shouguang, but are largely concentrated in east-central Shouguang, including Shengcheng Sub-district, Luocheng Sub-district, and Daotian Town.
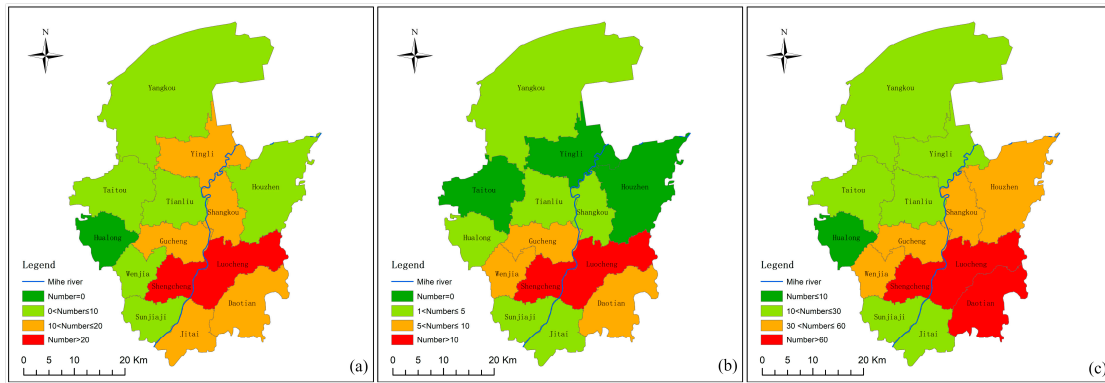

Figure 8. The spatial distribution of flood-related Weibo on different topics. (**a**) Disaster information; (**b**) rescue information; (**c**) public sentiment.

The spatial distributions of the nine sentiments are shown in Figure 9. The general spatial distribution of sentiments is concentrated in the central city along the Mihe river and extends outward, but the direction of the extension varies depending on the sentiment. The sentiments of "questioning the government and media", "seeking help", and "feeling sad about the disaster" are distributed mainly along the east of the Mihe river, a typical Weibo text accumulation area. "Praying for the victims" and "making donations" are concentrated in the south-central area, especially in the sub-districts of Shengcheng, Luocheng, Gucheng, and Wenjia.
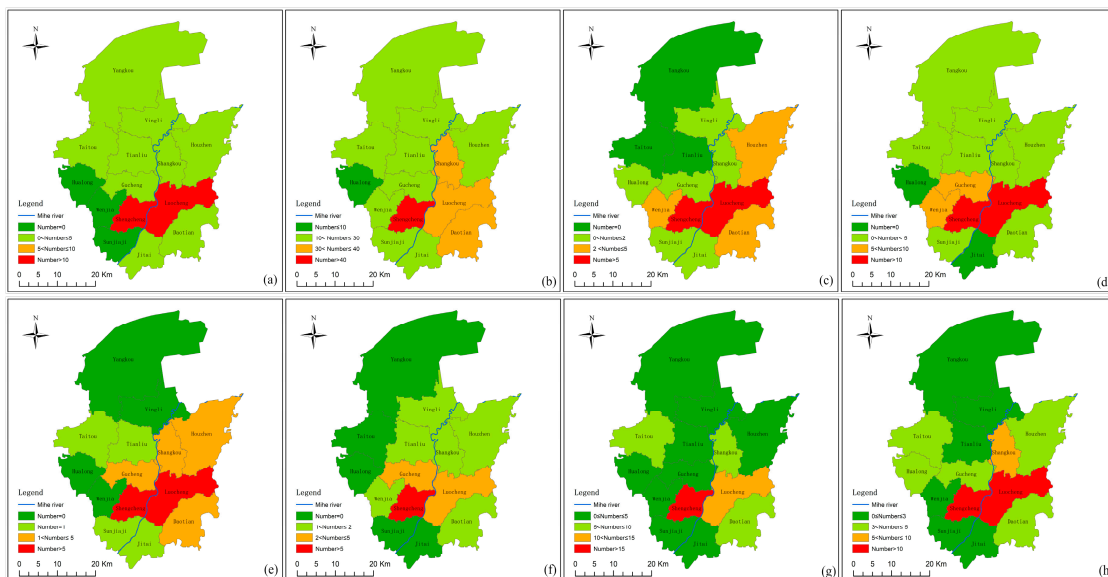

Figure 9. The spatial distributions of flood-related Weibo on different sentiments. (**a**) Concerned about the disaster situation; (**b**) questioning the government and media; (**c**) seeking help; (**d**) praying for the victims; (**e**) feeling sad about the disaster; (**f**) making donations; (**g**) thankful for the rescue; (**h**) worrying about vegetable prices.

4. Discussion

In this study, a topic extraction and classification model was built to hierarchically mine the topics and sentiments of flood-related Weibo texts. This is the first study to combine the LDA model with the RF algorithm to identify topics from Weibo texts and investigate public sentiments in depth. The model has been proven reliable and the topic classification results of flood-related Weibo are more accurate than previous studies. For example, Wang used the LDA model with the SVM algorithm to classify Weibo texts and provided an overall accuracy of 87.5% [22]. In this study, Weibo texts were classified with a precision of 89%, recall of 88%, and F1 of 88%. Moreover, our model effectively identified nine detailed sentiments from flood-related Weibo texts with a total precision of 78% (Table 3). The results of the sentiment identification quantitatively provided more qualitatively rich human-centric information to be used for disaster decision-making than previous studies. For example, Wang investigated seven topics by identifying and clustering important terms [20], whereas Zhu identified public sentiments from Weibo texts as positive, neutral, or negative [7]. In contrast, this study explicitly divided public sentiments during the flood into nine concrete sentiments. The results indicated that the secondary classification, based on natural language processing and machine learning, performed well in the analysis of public sentiments in the disaster-related Weibo.

The temporal variability of the number of flood-related Weibo coincided with the developmental stages of the flood. The results of the temporal analysis showed that the number of flood-related Weibo texts slightly increased on August 21, rose substantially on August 23, and peaked at 4:00 on August 24, which corresponded to the different developmental stages of the flood. From the Shandong Disaster Reduction Centre, Shandong Province launched a level III emergency response to the flood on August 21 and the first batch of relief materials was sent to Shouguang that afternoon. On August 23, the Weifang Government held a press conference to inform the public about the flood situation, which produced widespread concern [27]. Thus, the amount of flood-related Weibo texts increased quickly from August 23. The distribution of topics and sentiments over time also matched the changes in the flood situation. The Weibo on the "public sentiment" topic showed a similar tendency with overall trends. The "Disaster information" topic appeared simultaneously with the flood on August 20 while the "weather warning" topic was shared during the pre-disaster phase. Both the "rescue information" topic and the "thankful for the rescue" sentiment were concentrated to August 26–27. According to the official report, rescue teams from all over the country began arriving in Shouguang in the early morning of August 26. Meanwhile, other scholars present similar findings. Wang observed a temporally concurrent evolution of wildfire and related tweets [20]. Wang found the developmental trend of Beijing rainstorm events and the trend of the number of Weibo texts were highly correlated [22].

The spatial distribution on a sub-district level of flood-related Weibo texts could be related to the degree of population aggregation, economic development, and disaster severity. Flood-related Weibo texts were mainly distributed in the central area of Shouguang, especially in the sub-districts of Shengcheng and Luocheng. As the political, economic, cultural, and information center of Shouguang, Shengcheng sub-district has a population of 220,000, which is one-fifth of the total population of Shouguang. According to news reports, the disaster-stricken area covered northern Shouguang (e.g., Shnangkou Town). The findings reflect a potential correlation between the spatial distribution of flood-related Weibo and population aggregation. A higher population density and a convenient network connection could have provided a better basis for the release of Weibo messages relative to other regions. The sub-district level spatial patterns of the content varied in topic and sentiment. "Public sentiments", "questioning the government and media", "seeking help", and "feeling sad about the disaster" had distributions similar to the overall spatial distribution of flood-related Weibo. "Praying for the victims" and "making donations were concentrated in the south-central region. Meanwhile, "disaster information" was distributed along the Mihe river. The phenomenon may be related to population density, economic situation, and disaster severity.

On a larger scale, the distance from the disaster area and the extent of the impact of the disaster may have influenced the spatial distribution of flood-related Weibo. As shown in Figure 5, the spatial distribution of flood-related Weibo in the cities around Shouguang indicated that as the distance from the disaster center increased, the number of flood-related Weibo decreased. This finding was also presented by Herfort [5]. Among the cities, Weifang, Qingdao, Jinan, and Beijing were the main centers for flood-related Weibo. Weifang is the prefectural city of Shouguang, and it had the largest number of flood-related Weibo. Qingdao and Jinan are the largest cities and the provincial capitals in Shandong Province, respectively. The people in Qingdao and Jinan may have been more interested in Shouguang's flood because of its geographic location and the flood's relevance to the government. However, the number of flood-related Weibo in Beijing showed an abnormal pattern. Albeit far from Shouguang, Beijing also showed an aggregation of flood-related Weibo. There are two possible explanations. First, as the capital of China, Beijing has a high level of economic development and abundant population distribution at all levels, which results in a large number of netizens. People in the "political center" pay more attention to government-related reports. Second, Shouguang is one of the main vegetable suppliers of Beijing and the public may have been concerned about whether or not the Shouguang flood would affect the vegetable prices.

Nevertheless, this study has some limitations. First, the spatial information of the Weibo texts is place names (e.g., Shangkou Town, Kouzi Village) rather than latitude and longitude, causing inadequate analysis of the spatial distribution characteristics. Second, the specific reasons for the temporal-spatial distribution of the flood-related Weibo need further exploration. Third, the paper only analyzed the texts from social media while other content, such as pictures and videos, posted by social media users can also be informative to disaster management [8]. Fourth, the disaster-related Weibo texts were acquired based on the toponym keyword: "Shouguang". In other disaster scenarios, the toponym based method may collect unrelated information due to the ambiguity of place name. In addition, the related Weibo messages that do not contain the identified keywords may be ignored. This can be improved by extending the keyword lists and disambiguating toponyms in future studies. Despite the limitations, this study contributes to the existing research on uncovering emergency knowledge from social media by presenting a reliable approach to mining people's detailed opinions about a disaster. Most of the existing studies [20–22] used text mining methods (e.g., k-means clustering, the LDA model) to classify social media content into several categories. This study addresses the lack of a topic analysis that can identify detailed public sentiments during a disaster by providing an improved topic extraction and classification model.

5. Conclusions

This study comprehensively analyzed social media data related to the "2018 Shouguang flood" over time, space, and content and proposed a topic extraction and classification model focused on obtaining and analyzing public sentiments. The evaluation results show that the approach for topic extraction is accurate and viable for understanding public opinions during a disaster event. The findings indicate that the majority of the topics of the flood-related Weibo are "public sentiments" among which "questioning government and media" is the most commonly expressed. It suggests social media can help identify public perceptions of a disaster and potentially complement other methods for gauging public sentiment. The temporal changes in the disaster-related Weibo texts are synchronous with the development trend of the disaster event. The sub-district level spatial distribution of flood-related Weibo texts is largely concentrated in the south-central and eastern parts of Shouguang. Furthermore, at the city level, the flood-related Weibo are mainly distributed in Beijing and the cities of Shandong Province, centering in Weifang City. The spatial distribution of flood-relevant Weibo texts may be related to population density and distance from the disaster area. The study indicates that social media can help understand the public's attitudes towards a disaster, which can significantly enhance situational awareness and help to accelerate disaster response and support post-disaster management.

Further research is needed to improve the approach and further explore the deeper meanings behind the results. There is an urgent need to strengthen the acquisition of social media with spatial information, improve the identification of disaster-related information, and combine multi-source data to make the analysis more effective.

References

1. Klonner, C.; Marx, S.; Usón, T.; De Albuquerque, J.P.; Höfle, B. Volunteered Geographic Information in Natural Hazard Analysis: A Systematic Literature Review of Current Approaches with a Focus on Preparedness and Mitigation. *ISPRS Int. J. Geo-Inf.* **2016**, *5*, 103.

2. Kryvasheyeu, Y.; Chen, H.; Obradovich, N.; Moro, E.; Van Hentenryck, P.; Fowler, J.; Cebrian, M. Rapid assessment of disaster damage using social media activity. *Sci. Adv.* **2016**, *2*, e1500779. [CrossRef] [PubMed]

3. Shklovski, I.; Burke, M.; Kiesler, S.; Kraut, R. Technology Adoption and Use in the Aftermath of Hurricane Katrina in New Orleans. *Am. Behav. Sci.* **2010**, *53*, 1228–1246.

4. Sui, D.; Goodchild, M. The convergence of GIS and social media: challenges for GIScience. *Int. J. Geogr. Inf. Sci.* **2011**, *25*, 1737–1748. [CrossRef]

5. De Albuquerque, J.P.; Herfort, B.; Brenning, A.; Zipf, A. A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management. *Int. J. Geogr. Inf. Sci.* **2015**, *29*, 1–23. [CrossRef]

6. Li, Z.; Wang, C.; Emrich, C.T.; Guo, D. A novel approach to leveraging social media for rapid flood mapping: a case study of the 2015 South Carolina floods. *Cartogr. Geogr. Inf. Sci.* **2018**, *45*, 97–110. [CrossRef]

7. Zhu, R.; Lin, D.; Jendryke, M.; Zuo, C.; Ding, L.; Meng, L. Geo-Tagged Social Media Data-Based Analytical Approach for Perceiving Impacts of Social Events. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 15. [CrossRef]

8. Wang, Z.; Ye, X. Social media analytics for natural disaster management. *Int. J. Geogr. Inf. Sci.* **2018**, *32*, 49–72. [CrossRef]

9. Zhang, S.; Feick, R.; Feick, R. Understanding Public Opinions from Geosocial Media. *ISPRS Int. J. Geo-Inf.* **2016**, *5*, 74. [CrossRef]

10. Innes, J.E.; Booher, D.E. Reframing public participation: strategies for the 21st century. *Plan. Theory Pract.* **2004**, *5*, 419–436. [CrossRef]

11. Balaji, S.; Gelernter, J. An algorithm for local geoparsing of microtext. *Geoinformatica* **2013**, *17*, 635–667.

12. Bakillah, M.; Li, R.Y.; Liang, S.H.L. Geo-located community detection in Twitter with enhanced fast-greedy optimization of modularity: the case study of typhoon Haiyan. *Int. J. Geogr. Inf. Sci.* **2015**, *29*, 258–279. [CrossRef]

13. Crooks, A.; Croitoru, A.; Stefanidis, A.; Radzikowski, J. # Earthquake: Twitter as a distributed sensor system. *Trans. GIS* **2013**, *17*, 124–147.

14. Blanford, J.I. Tweeting and tornadoes. In Proceedings of the 11th International ISCRAM Conference, State College, PA, USA, 19–21 May 2014; pp. 319–323.

15. Fohringer, J.; Dransch, D.; Kreibich, H.; Schröter, K. Social media as an information source for rapid flood inundation mapping. *Nat. Hazards Earth Syst. Sci. Discuss.* **2015**, *3*, 4231–4264. [CrossRef]

16. Rosser, J.F.; Leibovici, D.; Jackson, M.J. Rapid flood inundation mapping using social media, remote sensing and topographic data. *Nat. Hazards* **2017**, *87*, 103–120. [CrossRef]

17. Tom, B.; Dirk, E.; Arnejan, V.L. Probabilistic flood extent estimates from social media flood observations. *Nat. Hazard Earth Syst.* **2017**, *17*, 735–747.

18. Imran, M.; Mitra, P.; Castillo, C. Twitter as a Lifeline: Human-annotated Twitter Corpora for NLP of Crisis-related Messages. In Proceedings of the 10th Language Resources and Evaluation Conference (LREC), Portorož, Slovenia, 23–28 May 2016; pp. 1638–1643.

19. Imran, M.; Elbassuoni, S.; Castillo, C.; Diaz, F.; Meier, P. Practical extraction of disaster-relevant information from social media. In Proceedings of the 22nd International Conference on World Wide Web, Rio de Janeiro, Brazil, 13–17 May 2013; pp. 1021–1024.

20. Wang, Z.; Ye, X.; Tsou, M.-H. Spatial, temporal, and content analysis of Twitter for wildfire hazards. *Nat. Hazards* **2016**, *83*, 523–540. [CrossRef]

21. Ye, X.; Li, S.; Yang, X.; Qin, C. Use of Social Media for the Detection and Analysis of Infectious Diseases in China. *ISPRS Int. J. Geo-Inf.* **2016**, *5*, 156. [CrossRef]

22. Wang, Y.; Wang, T.; Ye, X.; Zhu, J.; Lee, J. Using Social Media for Emergency Response and Urban Sustainability: A Case Study of the 2012 Beijing Rainstorm. *Sustainability* **2015**, *8*, 25. [CrossRef]

23. Griffiths, T.L.; Steyvers, M. Finding scientific topics. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 5228–5235. [CrossRef] [PubMed]

24. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **2012**, *3*, 993–1022.

25. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

26. Thiago, S.; Marcos, G.; Victor, R. Improving Random Forests by Neighborhood Projection for Effective Text Classification. *Inf. Syst.* **2018**, *77*, 1–21.

27. Provincial Disaster Reduction Center Emergency Allocation of Relief Material to Guarantee the Resettlement of the Disaster-stricken Areas. Available online: http://www.sdjianzai.gov.cn/gzdt/6165.jhtml (accessed on 22 August 2018). (In Chinese)