Typhoon Risk Perception: A Case Study of Typhoon Lekima in China

Jiting Tang¹²⁴ · Saini Yang¹²³⁴ · Yimeng Liu¹²⁴ · Kezhen Yao¹²⁴ · Guofu Wang⁵

Abstract The typhoon is one major threat to human societies and natural ecosystems, and its risk perception is crucial for contextualizing and managing disaster risks in different social settings. Social media data are a new data source for studying risk perception, because such data are timely, widely distributed, and sensitive to emergencies. However, few studies have focused on crowd sensitivity variation in social media data-based typhoon risk perception. Based on the regional disaster system theory, a framework of analysis for crowd risk perception was established to explore the feasibility of using social media data for typhoon risk perception analysis and crowd sensitivity variation. The goal was to quantitatively analyze the impact of hazard intensity and social and geographical environments on risk perception and its variation among population groups. Taking the Sina Weibo data during Typhoon Lekima of 2019 as an example, we found that: (1) Typhoon Lekima-related Weibo public attention changed in accordance with the evolution of the typhoon track and the number of Weibo posts shows a significantly positive correlation with disaster losses, while socioeconomic factors, including population, gross domestic product, and land area, are not explanatory factors of the spatial distribution of disaster-related Weibo posts; (2) Females, nonlocals with travel plans, and people living in areas with high hazard intensity, low elevation, or near waterbodies affected by Lekima paid more attention to the typhoon disaster; and (3) Descriptions of rainfall intensity by females are closer to the meteorological observation data.

Keywords China · Crowd sensitivity · Regional disaster system theory · Risk perception · Social media · Typhoon Lekima

1 Introduction

Typhoons are one of the major natural hazards in China. In 2019, typhoons caused about CNY 58.9 billion in direct economic losses in China, accounting for 18% of the annual direct economic losses caused by natural hazards (UNDRR 2020). When a typhoon occurs, people in the disaster-stricken areas often share personal experiences and views on the disaster events or the relief processes on social media (Tang et al. 2021). The popularity and content of disaster-related topics on social media reflect public perception of the disaster risk. In recent years, scholars have begun to use social media data in disaster impact assessment (Butler 2013; Kryvasheyeu et al. 2016; Baranowski et al. 2020; Eyre et al. 2020) and risk perception (Marshall 2019; Rahmanti et al. 2021), and have pointed out that there exist variations in the disaster impact and risk perceptions of different individuals. Fan et al. (2020) found uneven spatial distribution of social media attention,
although the hazard intensity of Hurricane Harvey was similar in some areas of the United States. Zhao et al. (2019) proposed that victims tend to share information, while the nondisaster-affected are more likely to express their feelings and opinions. Xiao et al. (2018) found that Twitter information was concentrated in hard-hit areas, landmark public places, and areas with easy access to the Internet. Macias (2016) indicated that variations in environmental risk perception existed among different race and ethnic groups.

Understanding disaster risk perceptions in a disaster-related context is crucial, because attitudes and risk preference shape how individuals, groups, and public- and private-sector organizations adopt risk mitigation practices (Gotham et al. 2018), how they develop post-disaster rebuilding activities (Albrecht et al. 2021), and how they plan for future disasters (Harman et al. 2021). Inevitably, different groups vary in responses to disasters (Lee et al. 2015). Social media data can be a valuable source to research disaster risk perception for more accurate and effective disaster education and risk communication. However, a comprehensive framework for the application of social media data in typhoon risk perception research has not been formed.

In this study, we established a novel framework based on the regional disaster system theory (Shi 2005) to comprehensively assess social media data for the analysis of risk perception in which the hazard, the surrounding social and geographical environments, and the affected population groups jointly affect typhoon risk perception. We took the Sina Weibo posts during Super Typhoon Lekima as an example to verify our model. Sina Weibo is a popular Chinese microblog that is similar to Twitter (Yao et al. 2021). Super Typhoon Lekima is the ninth named storm of the 2019 Pacific tropic cyclone season (Bao et al. 2020). At about 1:45 on 10 August 2019, it landed in the coastal area of Wenling City, Zhejiang Province, and moved into the Yellow Sea via Jiangsu Province. At around 20:50 on 11 August, it landed again in Qingdao City, Shandong Province twice, and then moved into the Bohai Sea and weakened continuously thereafter. At around 8:00 on 13 August, the National Meteorological Center downgraded Lekima to a tropical depression. Lekima was the fifth strongest typhoon that has landed in China since 1949. It brought extreme rainfall and a wide range of strong, long-duration winds. In eastern China, particularly in Zhejiang, Anhui, Jiangsu, and Shandong Provinces, urban and rural waterlogging, river floods, flash floods, and landslides have occurred in varying degrees. Lekima caused CNY 51.53 billion in direct economic losses. According to incomplete statistics, there are nearly 300 Lekima-related hashtags on Weibo, read by over 10 billion people, and over 1.7 million comments (Zhang et al. 2020). Lekima is a typical typhoon event with a wide impact area, high disaster loss, and strong public concern.

We used multiple machine learning models to extract hidden key information including spatial location, crowd sentiment, and rainfall intensity from Weibo. Spatial-temporal comparisons of multiple heterogeneous data were conducted based on three aspects: hazards, the disaster-formative environment, and hazard-affected elements especially humans. Combined with the three-dimensional risk perception cube, we investigated the feasibility of using social media data for typhoon disaster risk perception analysis and crowd sensitivity variation. Our goal was to provide directions for planners and policymakers on how to develop more effective typhoon risk communication strategies and risk reduction policies.

2 Methodology

This section outlines the methodology that was employed in the study. The specific issues discussed include the regional disaster system theory, research data collection, and some algorithms and processes used for extracting hidden information in Weibo texts.

2.1 Regional Disaster System Theory

The concept of regional disaster system theory was first introduced in 1991 (Shi 1991) and elaborated further several times (Shi 1996, 2002, 2005, 2009). It is widely used to analyze natural hazard-related disasters (Jia et al. 2016) and technological disasters (Yang et al. 2020) in China. The regional disaster system theory considers that the intensity of hazards $H$, the stability or sensitivity of the disaster-formative environment $S$, and the vulnerability of hazard-affected elements $V$ together constitute the functional structure of a regional disaster system $D_t$. That is,

$$D_t = H \cap S \cap V$$  \hspace{1cm} (1)$$

According to this theory, we proposed a corresponding framework for the application of social media data in typhoon risk perception research (Fig. 1). The typhoon risk perception $R_p$ is a result of the combined effects of the hazard and disaster $H$ (hazard intensity and disaster damages), the surrounding social and geographical environment $S$ (socioeconomic stability, geographical environment sensitivity), and the affected population groups $V$ (gender variation, residency variation). That is,

$$R_p = H \cap S \cap V$$  \hspace{1cm} (2)$$

We included three adjustments according to the characteristics of risk perception: (1) For the general public, it is difficult to accurately distinguish between...
hazards and disasters, so we discussed both together; (2) We divided the environment into two parts: social environment and geographical environment; and (3) The public shares their perceived typhoon risk on social media platforms, so we regarded population groups as the most important hazard-affected elements and analyzed the gender and residency variations of those groups. Next, we explored the feasibility (red colored text in Fig. 1) and crowd sensitivity (purple colored text) based on risk perception-oriented regional disaster system theory. Here, the feasibility analysis of using social media data demonstrates the viability of the proposed framework, and the crowd sensitivity analysis highlights the variation of typhoon risk perception.

2.2 Research Data

We used Sina Weibo API to collect original Weibo texts and user attribute data, taking “Lekima (利奇马) | Typhoon (台风) | Storm (暴雨)” as the keywords, and 10 min as the minimum retrieval time unit. The period is from 8 August to 14 August 2019. As mentioned above, Typhoon Lekima landed on 10 August and stopped on 13 August. Before the typhoon landed, meteorological agencies issued early warnings, and the emergency management agencies communicated disaster prevention and preparedness news on Weibo. Therefore, the public paid attention to the typhoon trend in advance. With the approach of the typhoon, their concerns continued to rise. After the typhoon event, disaster relief in some areas was still in progress. The public was concerned about the disaster impact. Therefore, in order to show the whole process of change of the public attention to Typhoon Lekima on Weibo, we chose 8–14 August as our research time span that slightly exceeded the impact period of Typhoon Lekima in China. After eliminating duplication, 69,936 out of 342,984 Weibo texts were identified with location information, based on which we conducted further study. The data fields include the unique identifier of Weibo, location information, user registration position, user gender, and so on.

According to the tropical cyclone landfall records in China from the China Meteorological Administration (CMA) Tropical Cyclone Data Center,1 Guangdong (264), Hainan (165), Taiwan (149), Fujian (130), and Zhejiang (51) were the provinces where typhoons landed more than 50 times between 1949 and 2019. We designated these five provinces as typhoon-prone areas, and other regions of China as areas not prone to typhoons.

This study also used multisource data for model building and comparative analysis, which mainly include: Typhoon Lekima best track data2 (from CMA Tropical Cyclone Data Center); weather station observational data3 (from the National Climate Center); disaster data of some provinces and cities (from the National Disaster Reduction Center of China4); Digital Elevation Model (DEM) 90 m data of Zhejiang Province5 (from the Resource and Environment Science and Data Center); data of gross domestic product (GDP), population, and grain sown area of some provinces

---

1. https://tcdata.typhoon.org.cn/en/dlrdqx_zl.html.
2. https://tcdata.typhoon.org.cn/en/zjlyqj_zlhq.html.
3. http://data.cma.cn/.
4. http://www.ndrc.org.cn/.
5. https://www.resdc.cn/data.aspx?DATAID=284.
and cities (from the National Bureau of Statistics\(^6\) and Zhejiang Provincial Bureau of Statistics\(^7\)).

### 2.3 Hidden Information Extraction in Weibo Texts

Weibo texts contain abundant hidden information. In addition to the directly obtained Weibo data, we extracted the rainfall intensities, sentiments, spatial information, and keywords from the texts. We constructed several models including the text sentiment analysis model and the rainfall intensity classification model by using deep learning techniques. By comparing diverse data, we analyzed the feasibility of using social media data for typhoon risk perception analysis and crowd sensitivity variation based on the three aspects of the regional disaster system theory (Fig. 2). First, we compared the spatiotemporal distribution of Lekima-related Weibo posts with the hazard intensity data, disaster damage statistics, and socioeconomic data, which verified the feasibility of using the Weibo data for risk perception analysis. Then, the differences in crowd perception sensitivity were analyzed from the perspectives of gender, residency, and the environment.

#### 2.3.1 Rainfall Intensity Classification Based on BiLSTM Algorithm

This study used Bidirectional Long Short-Term Memory (BiLSTM) networks (Cai et al. 2021; Li et al. 2021) to build a rainfall intensity classification model to map the relationship between Weibo text and near-time local rainfall intensity (Fig. 3). For each Weibo text with location information, we matched the latest observation data from nearby weather stations, and converted the hourly precipitation into light rain, moderate rain, heavy rain, and torrential rain (and above). Then, a BiLSTM-based rainfall intensity classifier is based on BiLSTM in PaddlePaddle\(^8\) (PArallel Distributed DEeP Learning, Baidu’s open-source deep learning framework). PaddlePaddle outperforms other frameworks in Chinese NLP tasks with the help of powerful Baidu Chinese search database (Ma et al. 2019). The training dataset is the self-built dictionary mentioned above. The loss function is cross entropy. The optimizer is the adaptive gradient (Adagrad) algorithm. The learning rate is 0.002. The batch size is 128. Section 3.2.1 discusses whether the model error caused by the subjective description of netizens has a gender bias.

\(^6\) [http://www.stats.gov.cn/tdsj/tjjs/2019/indexeh.htm.](http://www.stats.gov.cn/tdsj/tjjs/2019/indexeh.htm)

\(^7\) [https://zjcssm.oss-cn-hangzhou-zwynet-d01-a.internet.cloud.zj.gov.cn/jcms_files/jcms1/web3077/site/flash/tj/Reports1/2020%E7%BB%9F%E8%AE%A1%E5%B9%B4%E9%89%B42000929/2019%E5%B9%B4%E7%BB%9F%E8%AE%A1%E5%B9%B4%E9%89%B4%E8%85%89%E7%9B%982000929/indexeh.htm](https://zjcssm.oss-cn-hangzhou-zwynet-d01-a.internet.cloud.zj.gov.cn/jcms_files/jcms1/web3077/site/flash/tj/Reports1/2020%E7%BB%9F%E8%AE%A1%E5%B9%B4%E9%89%B42000929/2019%E5%B9%B4%E7%BB%9F%E8%AE%A1%E5%B9%B4%E9%89%B4%E8%85%89%E7%9B%982000929/indexeh.htm)

\(^8\) [http://www.gb688.cn/bzk/gb/newGblInfo?hcno=B4A0044ABCF80F8C6A048C1D0121A97D](http://www.gb688.cn/bzk/gb/newGblInfo?hcno=B4A0044ABCF80F8C6A048C1D0121A97D).
2.3.2 Text Sentiment Analysis Based on BiLSTM Algorithm

Sentiment analysis is the process of interpreting the meaning or polarity of larger text units (sentences, paragraphs, articles) through the semantic composition of smaller elements. It is essentially a task of text classification or regression. Specifically, language is regarded as a digital signal, and words are represented by a word vector algorithm. Then, the traditional machine learning algorithms (such as Bayesian, random forest) or deep learning neural network algorithms are used to train vector features in the text, so as to realize the classification or regression of the text sentiment polarity (de Diego et al. 2018).

Our text sentiment analysis model was based on Senta, a Chinese sentiment analysis tool of PaddlePaddle. The model was trained based on the BiLSTM algorithm and the existing SentiCrop—a Chinese sentence-level sentiment classification dataset. Then the model outputs a sentiment polarity category of Weibo text and the corresponding sentiment intensity $x$. Here, the closer the $x$ value is to 0, the stronger the negative sentiment is. The model analyzes public attitudes towards disaster occurrence and relief processes. Section 3.2.1 discusses whether the sentiment of netizens has a gender bias.

2.3.3 Spatial Information Extraction

This study involved three kinds of spatial information: user registration positions, Weibo locations, and spatial information mentioned in the text. We got the registered province and city directly from the user attributes, and used the inverse geocoding interface of Baidu Map API to convert coordinates of Weibo location into corresponding administrative divisions. Extracting spatial information

---

10 https://www.paddlepaddle.org.cn/modelbasedetail/senta.
from text includes toponym recognition and toponym resolution (Hu et al. 2019).

Toponym recognition refers to extracting words representing place names from text. Due to the common improper grammar and expression, misspelling, and irregular alias abbreviations in social media text, toponym recognition is a major challenge in NLP tasks (Sagcan and Senkul 2015). The smaller the granularity of spatial information extraction, the greater the error. We identified the terms of provinces, cities, and districts in the Weibo text, and completed the geographical location information according to China’s three-level administrative divisions.

Toponym resolution is the conversion of the structured address nouns into coordinates. We then used the geocoding interface of Baidu Map API to achieve toponym resolution. Through the toponym recognition and resolution from Weibo texts, we established the spatial position in which netizens were interested.

2.3.4 Text Keyword Extraction Based on Latent Dirichlet Allocation (LDA) Algorithm

Text keyword extraction is to dismantle long text into several keywords for rough screening and clustering of disaster-related text. Common text keyword extraction methods include those based on statistical features, graphic models, and topic models (Onan et al. 2016). The statistical feature-based methods, such as the TF-IDF algorithm, rely heavily on a high-quality corpus related to the target sample set for training. The graphic model-based methods, such as the PageRank algorithm, have high computational complexity and are not widely used. The topic model-based methods, such as the Latent Dirichlet Allocation (LDA) algorithm (Jelodar et al. 2018), are statistical models that extract keywords from implicit semantic structures of documents in an unsupervised way. The LDA algorithm maps words into a semantic space by fitting the word-document-subject distribution based on the analysis of word cooccurrence information.

LDA assumes that each document can be expressed as the probability distribution of latent topics, and that topic distribution in all documents shares a common Dirichlet prior. Each latent topic in the LDA model is also represented as a probability distribution over words. The word distributions of topics share a common Dirichlet prior as well. Given a corpus $D$ consisting of $M$ documents, document $d$ having $N_d$ words $(d \in 1, \ldots, M)$, the number of topics $T$, LDA models $D$ according to the following generation process (Blei et al. 2003): Choose a multinomial distribution $\phi_t$ for topic $t(t \in \{1, \ldots, T\})$ from a Dirichlet distribution with parameter $\beta$. Choose a multinomial distribution $\theta_d$ for document $d(d \in \{1, \ldots, M\})$ from a Dirichlet distribution with parameter $\alpha$. For a word $w_n(n \in \{1, \ldots, N_d\})$ in document $d$, select a topic $z_{dn}$ from $\theta_d$, and a word $w_n$ from $\phi_{z_{dn}}$.

In the above generative process, words in documents are the only observed variables while others are latent variables $(\phi, \theta)$ and hyper-parameters $(\alpha, \beta)$. The probability of observed data $D$ was calculated and obtained from a corpus as follows:

$$p(D|\alpha, \beta) = \prod_{d=1}^{M} \int p(\theta_d|\alpha) \left( \prod_{z=1}^{Z} \sum_{h=1}^{H} p(z_{dn}|\theta_d)p(w_{dn}|z_{dn}, \beta) \right) d\theta_d$$

(10)

LDA is a distinguished tool for latent topic distribution for a large corpus. To extract the hot topics from netizens, we used an LDA algorithm to explore the deep meaning of texts, that is, the topic of the texts. Specifically, we used lda_webpage$^{11}$—a Baidu self-built web page dataset—to create a dictionary and used an LDA algorithm to find the first (no more than) 10 keywords of each Weibo text and their corresponding similarity of lda_webpage dataset topics.

3 Results

This section presents the findings based on the risk perception-oriented regional disaster system theory, including the degree of hazard and disaster, the social and geographical environments, and the variation of population groups that jointly affect the typhoon risk perception.

3.1 The Degree of Hazard and Disaster

This section verifies the feasibility of using social media data for typhoon risk perception analysis according to the high correlation between Weibo posts and hazard intensity as well as elevated disaster damage.

3.1.1 Large Number of Weibo Posts in the Most Typhoon-Stricken Areas

Based on Weibo locations, we analyzed the spatial distribution (Fig. 4) and spatiotemporal evolution of Weibo locations (Fig. 5). The results suggest that there is reasonable consistency between public concern and hazard intensity.

From the overall spatial distribution of Weibo locations (Fig. 4), we found that Weibo locations are consistent with the spatial distribution of typhoon impact. From 8 to 13 August 2019, the overall point density distribution of

$^{11}$ https://github.com/PaddlePaddle/PaddleHub/tree/df850a187e53c2a1a9318ad1d3da5d716773619/modules/text/language_model/lda_webpage.
Fig. 4 Spatial distribution of typhoon Lekima-related *Weibo* posts

Fig. 5 Changes in daily *Weibo* amount rankings in the major affected areas of Lekima
Lekima-related Weibo posts indicates that most of the Weibo locations are concentrated in the southeast coastal area. The number of Weibo posts in typhoon-stricken areas is significantly larger than posts in other areas.

From the changes of Weibo posts’ amount rankings in some typhoon-stricken regions (Fig. 5), we found a high conformity between the spatiotemporal evolution of Weibo posts’ popularity and that of Typhoon Lekima. The netizens in Guangdong, Fujian, Zhejiang, Shanghai, Anhui, Jiangsu, Shandong, Hebei, and Liaoning paid high attention to this typhoon event. Before Lekima landed, netizens in Guangdong, Fujian, and Zhejiang paid close attention to the typhoon event, which may be related to local weather warnings and safety publicity. With the northward movement of the rain belt, the number of Weibo posts from the provinces and cities on the typhoon tracks also increased, and the heat of Lekima-related Weibo in the southeastern region remained high, especially in Zhejiang Province. Strongly typhoon-stricken areas have many Weibo posts before, during, and after the disaster.

3.1.2 High Spatial Matching between Weibo Posts and Severe Disaster Impacts

We analyzed the correlation between the number of Lekima-related Weibo posts and disaster damage statistics and socioeconomic data (Table 1). These data covered nine provinces (municipality) including Zhejiang, Shanghai, Jiangsu, Shandong, Anhui, Fujian, Hebei, Liaoning, and Jilin. The number of Weibo posts has the highest correlation with the affected population ($r = 0.8940$, $p = 0.0011$), followed by direct economic losses ($r = 0.8920$, $p = 0.0012$) and the number of houses moderately damaged ($r = 0.8791$, $p = 0.0018$). We concluded a high correlation between Weibo posts and disaster damage statistics.

3.2 Social and Geographical Environments

This section investigates the impact of social and geographical environments on public risk perception at the provincial and county level, and also verifies the feasibility of using social media data for risk perception analysis.

3.2.1 No Clear Correlation between Weibo Posts and Socioeconomic Factors

Figure 4 shows that the Lekima-related Weibo posts were concentrated in the typhoon-affected areas. Table 1 shows that, at the provincial level, some socioeconomic factors, such as population, GDP, and grain sown area, are weakly correlated with the number of Weibo posts, which indicates that spatial heterogeneity of the population and economic factors may have little effect on the Lekima-related Weibo posts.

We took Zhejiang Province, one of the most severely affected provinces, as an example to compare the number of Weibo posts and economic indicators at the county level. The localized Weibo posted from Zhejiang were most intensive, with a total number of 20,946. The Moran’s I index of 0.76 for the Weibo post locations shows a clear spatial clustering phenomenon. We integrated the main socioeconomic indicators in 2019 at the city and county levels in Zhejiang Province, and obtained 62 effective data after data processing. Similarly and interestingly, we found that there was no significant relationship between the number of Weibo posts and these economic indicators, including land area, resident population, and per capita GDP. Thus, we speculated that the spatial distribution of disaster-related Weibo posts was not affected by socioeconomic factors.

3.2.2 Higher Geographical Environment Sensitivity, Higher Attention

The spatial distribution of Lekima-related Weibo posts in Zhejiang Province differed from the typhoon track (Fig. 6). It was noted that Weibo users living at lower altitudes or in areas with larger proportions of waterbodies were more concerned about typhoon disasters (shown as red dots in Fig. 6). The words “flooded (被淹, 积水)” and “flood (洪水)” appeared frequently, possibly because people in the low altitude areas or along the rivers generally paid more attention to flood risk. This indicates that public perception was consistent with the risk in their neighborhoods.

3.3 Variation among Population Groups

In this section, we take gender and residency as examples to illustrate the typhoon risk perception of different population groups.

3.3.1 Gender Variation

During Typhoon Lekima, the proportion of Weibo posts by male and female users was about 1:2 (24,140 by males, 45,343 by females, and 452 without user gender information). Compared with the data from Weibo User Development Report 2020, where the ratio of male and female users was 45.4, the difference in male and female ratio indicates that female users were more willing to voice their concerns about this typhoon disaster on Weibo than were male users. One possible reason is that females are more

---

12 https://data.weibo.com/report/reportDetail?id=456.
vulnerable to disasters (Ginige et al. 2009), and they tend to be the primary caregivers of families. They are possibly more sensitive about disasters also because they care more about the disaster event’s livelihood impact. This finding is consistent with previous studies in which females are more sensitive to negative events on social media than males (Zheng et al. 2019; Aak et al. 2020).

The overall accuracy of the rainfall intensity classification model fluctuated around 75%, and the cross entropy loss is around 0.9 (Fig. 7). Further, we tested the rainfall intensity classification model and the text sentiment analysis model for the two gender groups, respectively (Table 2). We found that males showed slightly higher negativity in their Weibo posts than females, while females’ hazard descriptions were slightly closer to the reported meteorological observation data. The univariate variance analysis results show that the negative sentiment expression \( F = 28.32794447, P = 1.02722E-07 \) and the hazard intensity description \( F = 488.4877355, P = 1.0528E-07 \) of the two genders were significantly different (passed the significance test of 1% level). This is a direction worth studying in the field of typhoon disaster risk perception.

### 3.3.2 Residency Variation

We obtained the user locations, user registration positions, and the text mentioned positions by spatial information extraction described in Sect. 2.2.3. Then we counted the number of nonempty data and the same-in-pair data (Table 3). We found that the proportion of provincial consistency is larger (all over 50%), and the proportion of municipal consistency is slightly lower.

Table 3 represents two statistical values of nonempty data and the same-in-pair data in the comparison among locations, registration positions, and text-mentioned positions. Of Weibo positioning provinces, 76.37% are the same as those mentioned in the text (16,390 out of 21,461), and this is 72.24% at the city level (9,589 out of 13,274). These two ratios illustrate that most users are concerned about the disaster information in their neighborhoods. For the Weibo texts whose locations and text-mentioned positions are inconsistent, most of the contents are related to the status of a specific extreme disaster-stricken area or travel destination.

We defined a user as a “resident” if the user registration city is the same as the user location. Lekima-related Weibo locations cover 324 cities, totaling 63,133 posts. Among them, 36,793 Weibo posts contain user registration
positions and 19,593 were from residents, which account for 53.25%. The number of Weibo posts in different regions varied greatly, so we took the natural logarithm of the number of Weibo posts as the Weibo popularity. The average popularity of residents’ Weibo posts was 2.27 and that of nonresidents’ Weibo posts was 2.04. We compared the difference of residents’ and nonresidents’ Lekima-related Weibo posts at the city level. Considering the three cases of all areas, typhoon-prone areas, and areas not prone to typhoons respectively, the univariate variance analysis results of the two residency groups failed the significance test. It shows that the difference in Weibo popularity between residents and nonresidents was not obvious in this event. Similar results appeared in some urban flood studies (Wang et al. 2020). Thus, we suggest that the distinction between locals and nonlocals may not be critical in the analysis of public responses towards typhoons on Weibo.

We extracted the keywords for Weibo posts located in Zhejiang Province. Then we found that nonresidents used more words such as “hotel (酒店),” “outage (停运),” “high-speed railway (高铁),” and “on the road (路上)” than residents, indicating that nonresidents pay more attention to travel. This typhoon greatly disrupted travel-related activities, so nonresidents with travel needs were more sensitive to its impact. This result is consistent with the behavior of the public, which shows the feasibility of using Weibo data in disaster assessment.

Fig. 6 Spatial distribution of Lekima-related Weibo posts in Zhejiang Province, China. DEM digital elevation model.
4 Discussion

This section summarizes the dynamic characteristics of social media-based typhoon risk perception, then discusses some implications of the results for typhoon risk communication and management, and gives future research directions based on the limitations of this work.

4.1 Dynamic Typhoon Risk Perception

We attempted to understand how people perceive typhoon risks based on regional disaster system theory. Our case study of Typhoon Lekima found that the public risk perception in typhoon disasters has dynamic characteristics in different times, regions, and populations.

(1) Public risk perception is closely related to the degree of hazard and disaster. The overall spatiotemporal distribution of Lekima-related Weibo posts is significantly clustered and highly correlated with typhoon processes (Fig. 4). The provinces and cities that Weibo users are located in are consistent with the daily typhoon-affected areas, and the number of Weibo posts generated in these areas is significantly higher than that in other areas (Fig. 5). Zhejiang and Shandong Provinces are where Lekima landed. These two provinces had the highest Weibo post activities. The number of Weibo posts has a high correlation to the disaster loss data (Table 1).

(2) The geographical, but not social, variables could better explain the spatial heterogeneity of Weibo interests. The population density, GDP, and other socioeconomic factors are not the main influencing factors of the spatial distribution of Lekima-related Weibo posts (Table 1). In previous studies, some researchers found that the number of social media users in a region is highly correlated with the level of socioeconomic development (Khare et al. 2020). Our results remind us that this conclusion may not be universal. Therefore, we suggest that before using social media data for disaster analysis, scholars should first examine the relationship between social media data and socioeconomic indicators. More importantly, our results show that the Weibo data exactly reflect the impact of typhoon disasters.

(3) The disaster-related social media data from different groups indicate biases. From the perspective of gender, female users’ overall response to disasters is more sensitive and objective. First, the number of Weibo posts shows that females are more active in online discussions related to typhoon disasters than males. Second, there is a gender bias in errors in the text sentiment analysis and rainfall intensity classification. Males have a slightly higher negative sentiment than females, and females’ descriptions of rainfall intensity are closer to observations (Table 2). From the perspective of geographical environment, people living in typhoon-affected areas at low altitudes or near waterbodies pay more attention to typhoon disasters (Fig. 6). It indicates that the perception of Weibo users is consistent with local risks, and further demonstrates the feasibility of retrieving disaster risk information from Weibo data.

From the perspective of residency, most people are concerned about their local disasters (Table 3), and there is no significant difference in being a resident or not for the focus on typhoon disasters, but nonresidents are generally more concerned about the travel impact caused by typhoon rains. It is worth noting that user registrations, positions mentioned in the text, and locations usually do not exactly match, but most of the existing social media-based natural hazard and disaster research only used one or two sets of the spatial information. In practical application, one local data source is used instead of the other because of data acquisition difficulty, such as regarding the user registration as a Weibo location. However, our study shows that these biases cannot be neglected.

This study revealed the crowd sensitivity variations of social media with regard to typhoon perception, which are reflected in three aspects: loss possibility, gender sensitivity, and travel needs. The confirmed feasibility of using social media data indicates the potential of this new data source in risk perception research.

| Table 2 Gender differences in text sentiment analysis and rainfall intensity classification models |
|-----------------------------------------------|
| User registration | Gender | Negative sentiment intensity | Accuracy of rainfall intensity classifier prediction |
|                  |        | Mean | Variance | Mean | Variance |
| Male             | 0.57171091 | 0.00816648 | 0.64456795 | 0.22910516 |
| Female           | 0.55216959 | 0.01055505 | 0.66464704 | 0.22290042 |
4.2 Implications

Social media platforms have a wide range of users, fast information production and dissemination, and are not subject to geographical constraints (Ilieva and McPhearson 2018). It is feasible to obtain disaster information and analyze public perception based on Weibo data. Social media conveys the public disaster risk perception and can reflect the occurrence and evolution of a disaster at multiple spatiotemporal scales. Based on the previous discussion, we verified the feasibility of using social media in typhoon disaster risk perception analysis, and emphasized that the crowd sensitivity variation cannot be ignored in risk communication. We have attempted to advise what risk management strategies governments can implement to improve risk perception and reduce the negative consequences of typhoons.

The typhoon disaster risk perception based on social media data indicates that disaster emergency management should be strengthened by an enriched understanding of differences in targeted public groups. Gender, residency, and geographical environment tend to be important sociodemographic predictors of risk perceptions. First, the higher sensitivity and accuracy of females’ typhoon risk perception suggest that scholars or policymakers should consider a more important role for females in disaster risk reduction, especially in early warning and evacuation. Social media platforms could also promote more targeted disaster information for male users to increase their disaster awareness. Second, there is no obvious difference in the attention to disasters caused by residency differences, but nonresidents have a lower cognition level of local vulnerability (Ushiyama et al. 2009), so they are a relatively higher risk group. Moreover, nonresidents with travel needs were more sensitive to typhoon impacts. Social media platforms could supplement and publicize more local environmental information and updated travel-related disaster news for nonresidents. Third, exposure affects public risk perception. People in areas with high exposure to typhoons are more concerned about the typhoon and secondary disasters such as floods. Disaster information dissemination for this group could be updated more intensively. We advocate individualizing disaster publicity and risk communications to improve the effectiveness of emergency management.

4.3 Future Research Directions

Several extensions to the methodology are anticipated in future work. In this work, the social media data come only from Sina Weibo during Typhoon Lekima. Subsequent studies could consider data from other social media
platforms and disaster cases, and would make the model performance more robust.

We also recognize that the current rainfall intensity classification method does not directly map social media text to artificially labeled hazard or disaster intensity. We use discrete meteorological observation data as the classification label. There exist a few Chinese terminology thesauri, social media corpora, and sentiment analysis dictionaries in the field of typhoon disasters or emergency management. The accuracy of classification using existing datasets in other fields directly may not be high. Therefore, researchers mostly collected social media texts based on several event cases and labeled them manually, which was time-consuming and laborious. It is also susceptible to specific hazard vocabulary, such as typhoon names, which can lead to the over-fitting of models. Based on the principle of space-time proximity, our model directly matches Weibo text to meteorological observation data in batches without considering the influence of individual vocabulary weights in the text, which makes the classification efficient, and avoids subjective errors in manual labeling. But this method may not directly map the relationship between independent and dependent variables.

The proposed method is a comprehensive feasibility and crowd sensitivity variation analysis approach of social media in typhoon risk perception. The formation of crowd sensitivity on social media is a multidimensional issue. Crowd sensitivity varies among population groups. Due to the limited fields available in the Weibo API, this study excavates information from the fields of user gender, user registry, Weibo location, text content, and so on. In future research, we can further analyze the mechanism of Weibo information transmission, public opinion control measures on social media, and subjective error causes of different genders.

5 Conclusion

A new framework for the feasibility and crowd sensitivity variation analyses of using social media data in typhoon risk perception is proposed based on the regional disaster system theory. A comprehensive assessment result is accomplished with the availability of multisource data and machine learning techniques.

In this study, we obtained nearly 70,000 Weibo texts with the location and user attribute information during super typhoon Lekima by Sina Weibo API, extracted spatial information in each text by combining toponym recognition technology and Baidu Map API geocoding, and constructed several models for keyword extraction, rainfall intensity classification, and sentiment analysis based on LDA and BiLSTM algorithms. We used various machine learning techniques to extract hidden information from Weibo texts, breaking through the limitation of insufficient information channels.

Our study is the first attempt to understand the linkages among typhoon risk perceptions of Weibo users, gender differences, residency vulnerability, and geographical environment sensitivity. These findings enlighten us to the need to individualize disaster information dissemination rather than emphasize general broadcasting. Targeted disaster publicity and communications with localized risk information will be powerful tools to improve emergency management and implement disaster risk reduction strategies. This study provides evidence and technical support for the effective utilization of typhoon-related social media information in disaster risk perception research.

Acknowledgments This study was supported by the National Key Research and Development Program of China (No. 2018YFC1508903), the Science Technology Department of Zhejiang Province (No. 2022C03107), and the International Center for Collaborative Research on Disaster Risk Reduction.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Aak, A., A. Iar, and B. An. 2020. Gender-based approach for assessing risk perception in a multi-hazard environment: A study of high schools of Gilgit, Pakistan. International Journal of Disaster Risk Reduction 44: Article 101427.

Albrecht, R., J.B. Jarecki, D.S. Meier, and J. Rieskamp. 2021. Risk preferences and risk perception affect the acceptance of digital contact tracing. Humanities and Social Sciences Communications 8: Article 195.

Bao, X., L. Wu, S. Zhang, H. Yuan, and H. Wang. 2020. A comparison of convective raindrop size distributions in the eyewall and spiral rainbands of Typhoon Lekima (2019). Geophysical Research Letters 47(23): Article 2020GL090729.

Baranowski, D.B., M.K. Flatau, P.J. Flatau, D. Karnawati, K. Barabasz, M. Labuz, B. Latos, J.M. Schmidt, et al. 2020. Social-media and newspaper reports reveal large-scale meteorological drivers of floods on Sumatra. Nature Communications 11: Article 2503.

Blei, D.M., A.Y. Ng, and M.I. Jordan. 2003. Latent Dirichlet Allocation. Journal of Machine Learning Research 3: 993–1022.

Butler, D. 2013. Crowdsourcing goes mainstream in typhoon response. Nature 2013:14186: 2022. https://doi.org/10.1038/nature.2013.14186. Accessed 20 Feb 2022.
Cai, W., B. Liu, Z. Wei, M. Li, and J. Kan. 2021. TARDB-Net: Triple-attention guided residual dense and BiLSTM networks for hyperspectral image classification. Multimedia Tools and Applications 80(7): 11291–11312.

de Diego, I.M., A. Fernandez-Isabel, F. Ortega, and J.M. Moguerza. 2018. A visual framework for dynamic emotional web analysis. Knowledge-Based Systems 145: 264–273.

Eyre, R., F.D. Luca, and F. Simini. 2020. Social media usage reveals recovery of small businesses after natural hazard events. Nature Communications 11: Article 1629.

Fan, C., F. Wu, and A. Mostafavi. 2020. A hybrid machine learning pipeline for automated mapping of events and locations from social media in disasters. IEEE Access 8(8): 10478–10490.

Ginghe, K., D.G. Amarantunga, and R. Haigh. 2009. Mainstreaming gender in disaster reduction: Why and how? Disaster Prevention and Management 18(1): 23–34.

Gotham, K.F., R. Campanella, K. Lauve-Moon, and B. Powers. 2018. Hazard experience, geophysical vulnerability, and flood risk perceptions in a postdisaster city, the case of New Orleans. Risk Analysis 38(2): 345–356.

Harman, J.L., J.M. Weinhardt, J.W. Beck, and I. Mai. 2021. Interpreting time-series COVID data: Reasoning biases, risk perception, and support for public health measures. Scientific Reports 11(1): Article 15585.

Hochreiter, S., and J. Schmidhuber. 1997. Long short-term memory. Neural Computation 9(8): 1735–1780.

Hu, Y., S. Gao, D. Lunga, W. Li, S. Newsam, and B. Bhaduri. 2019. GeoAI at ACM SIGSPATIAL: Progress, challenges, and future directions. ACM SIGspatial Special 11(2): 5–15.

Ilieva, R.T., and T. Mcpherson. 2018. Social-media data for urban sustainability. Nature Sustainability 1(10): 553–565.

Jelodar, H., Y. Wang, Y. Chi, F. Xia, X. Jiang, Y. Li, and L. Zhao. 2018. Latent Dirichlet Allocation (LDA) and topic modeling: Models, applications, a survey. Multimedia Tools and Applications 78(11): 15169–15211.

Jia, J.Y., L.Y. Han, Y.F. Liu, N. He, Q. Zhang, X. Wan, Y.F. Zhang, J. Ren, F. Huang, and W. Wang. 2021. Social media-based disaster communications: Public attention on the new normal during the COVID-19 pandemic in Indonesia. Computer Methods and Programs in Biomedicine 205: Article 106083.

Sagcan, M., and P. Senkul. 2015. Toponym recognition in social media for estimating the location of events. 2015 IEEE International Conference on Data Mining Workshop (ICDMW), 14–17 November 2015, Atlantic City, NJ, USA, 33–39.

Shi, P. 1991. On the theory of disaster research and its practice. Journal of Nanjing University Natural Science 11: 37–42 (in Chinese).

Shi, P. 1996. Theory and practice of disaster study. Journal of Natural Disasters 5(4): 6–17 (in Chinese).

Shi, P. 2002. Theory on disaster science and disaster dynamics. Journal of Natural Disasters 11(3): 1–9 (in Chinese).

Shi, P. 2005. Theory and practice on disaster system research on a fourth time. Journal of Natural Disasters 14(6): 1–7 (in Chinese).

Shi, P. 2009. Theory and practice on disaster system research on a fifth time. Journal of Natural Disasters 18(5): 1–9 (in Chinese).

Siami-Namini, S., N. Tavakoli, and A.S. Namin. 2019. The performance of LSTM and BiLSTM in forecasting time series. In Proceedings of 2019 IEEE International Conference on Big Data, 9–12 December 2019, Los Angeles, CA, USA, 3285–3292.

Tang, J., S. Yang, and W. Wang. 2021. Social media-based disaster research: Development, trends, and obstacles. International Journal of Disaster Risk Reduction 55: Article 102095.

UNDRR (United Nations Office for Disaster Risk Reduction). 2020. 2019 global natural disaster assessment report. https://www.preventionweb.net/files/73363_2019globalnaturaldisasterassessment.pdf. Accessed 10 Jan 2022.

Ushiyama, M., A. Yoshida, N. Kashiwagi, S. Sato, and Y. Sato. 2009. Analysis of the effects of disaster workshop for non-residents on the participants. Analysis of the Japan Society for Natural Disaster Science 27(4): 375–385.

Wang, B., B.P.Y. Loo, F. Zhen, and G. Xi. 2020. Urban resilience from the lens of social media data: Responses to urban flooding in Nanjing, China. Cities 106: Article 102884.

Xiao, Y., Q. Huang, and C. Zhang. 2018. Research on urban disaster analysis based on the big data mining of social media: Case study of Hurricane Sandy in New York. Urban Planning International 33(4): 84–92.

Yang, P., X. Huang, L. Peng, Z. Zheng, X. Wu, and C. Xing. 2020. Safety evaluation of major hazard installations based on regional disaster system theory. Journal of Loss Prevention in The Process Industries 69: Article 104346.

Yao, K., S. Yang, and J. Tang. 2021. Rapid assessment of seismic intensity based on Sina Weibo—A case study of the changing earthquake in Sichuan Province, China. International Journal of Disaster Risk Reduction 58: Article 102217.

Zhang, L., W. Ai, and X. Luo. 2020. Research on communication of network information under meteorological disaster: Case of typhoon. Journal of the Meteorological Sciences 40(6): 868–874.

Zhao, X., M.M. Zhan, and B.F. Liu. 2019. Understanding motivated publics during disasters: Examining message functions, frames, and styles of social media influentials and followers. Journal of Contingencies and Crisis Management 27(4): 387–399.

Zheng, S., J. Wang, C. Sun, X. Zhang, and E.K. Matthew. 2019. Air pollution lowers Chinese urbanites’ expressed happiness on social media. Nature Human Behaviour 3(3): 237–243.

Publisher’s Note

 Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.