THE TYPHOON DISASTER ANALYSIS BASED ON WEIBO TOPIC HEAT

J. Yuan 1, 3, A. Gong 1, 2, 3, J. Wang 1, 2, 3, J. Li 1, 2, 3

1 Key Laboratory of Environmental Change and Natural Disaster, MOE, Beijing Normal University, Beijing, China - gado@bnu.edu.cn
2 Beijing Key Laboratory of Environmental Remote Sensing and Digital City, Beijing Normal University, Beijing, China - gado@bnu.edu.cn
3 Faculty of Geographical Science, Beijing Normal University, XinJieKouWai Street, Beijing, China - gado@bnu.edu.cn

Commission III, WG III/IV a

KEY WORDS: Typical cyclone disaster, Weibo data, Topic heat, Quantitative relationship, Damage assessment, Hazard management

ABSTRACT:

Could social media data be utilized in hazard evaluation? Typhoon disaster as one of the costly disasters has become devastating threats for human. Moreover, social media change the communication way of human and citizens can turn to this platform to express disaster-related information at real time. Therefore, social media improves situational awareness and widens the method of hazard information acquiring. With more and more studies investigating in relationship between social media response and degree of damage, the strong correlation has been proved. Weibo as one of the most popular social media in China can provide data with posted text, location, user identification and other additional information. Combining with 10 tropical cyclones and Weibo data in 2013, We perform a quantitative analysis between the grade of hazard situation and Weibo related topic heat in province scale. We provide a new model of Weibo topic heat to evaluate the Weibo activity in study area. Also we demonstrate the hazard assessing formula is $H = 1.8845 \ln (\alpha) + 15.636$ in tropical cyclone disaster. High level goodness of curve fitting also suggest that this equation can be used for rapid assessment of hazard caused by tropical cyclones.

1. INTRODUCTION

Typhoon disaster represents costly, which is the one of the most devastating threats for human nowadays (Hsiang et al., 2014; Smith et al., 2013). In the last 10 years, the direct economic loss of tropical cyclone around the world evenly reaches up to 50 billion US dollars every year, and the loss of China represents roughly 15% of the world (Chen et al., 2017). Due to the growing of population and assets continuously, there is increasing risking exposure to coastal areas of China (Pielke et al., 1998; Raghavan et al., 2003). Among such hazards, one of the challenges of emergency management is lack of the real-time status-information from affected areas.

Social media, including Twitter, Facebook, weibo and etc., has changed the way of communication for people (Kimanuka, 2015). Weibo, as one of the most famous social media platform in China, can provide real-time and round-the-clock situational awareness from people. In 2012, the daily active number of weibo users are up to 46.2 million, and the amount of posted messages expend 100 million (Tang et al., 2014). More over the number of Weibo users continues to be larger. Especially, as most texts are posted through mobile phone, weibo and other social media can offer data with location-based information and improve the hazard management research and practice (Roick et al., 2013).

Based on the advantages of social media data, more and more natural hazard studies make a good use of it in emergency situation. As for Chinese researchers, they are mainly attracted to dig the relationship between the disaster status presented from weibo text and the real. For instance, comparing the weibo data on the events Olypic Games and typhoon disaster was shown to strong correlation of the location distribution between affected areas and weibo texts, while Olympic event does not have (Chen et al., 2017). At abroad, existing researches on the use of Twitter in natural hazard is more manifold (Kryvashevyeu et al., 2016). Some researchers study its contribution to situational awareness (Vieweg et la., 2010; Power et al., 2014), and some pay attention to practical field of classifying disaster message, detecting events and identifying risking areas (Earle et al., 2011; Sakaki et al., 2010; Kumar et al., 2014; Licran et al., 2013; Caragea et al., 2011). Especially for tropical cyclone disaster, Kryvashevyeu and others presented a multiscale analysis of Twitter activity and Hurricane Sandy and found a strong relationship between Sandy’s path and topic related Twitter activity (Kryvashevyeu et al., 2016). Also, Yago’s research prove that Twitter peak occurred preparedness phase and decreased abruptly after the storm (Martin et al., 2017). However, most studies on the use of weibo data are qualitative research.

Here, we can, for the first time, present a quantitative analysis of disaster related to Weibo topic heat. We start at provincial level, and progressively use spatial information of Weibo data to analyze the relationship between different factors and topic heat. Finally, based on the Weibo topic heat (H) and association degree ($\alpha$), we find the equation to describe the quantitative correlation of $H$-$\alpha$.

In this paper, the studying hazards were triggered by the 10 hurricanes landing in China of the 2013 season. Also, the Weibo data, collected from crisis provinces in the hurricane phases, are used to analyze the topic heat.

2. DATA COLLECTION AND PREPROCESSING

2.1. Tropical cyclone Information

There are 10 tropical cyclones of the 2013 season selected from CMA Tropical Cyclone Database from ‘tcdata.typhoon.org.cn’
(Ying et al., 2014). Their names are Bebinca, Rumbia, Soulik, Cimaron, Jebi, Utor, Trami, Usagi, Fitow, and Haiyan, as shown in Figure 1. Also, we acquire more information of the tropical cyclones, including number, path, time, strength of wind. And all the selected made their landfall on the coastal area of China and caused different levels of disaster.

2.2. Damage Assessment: Tropical Cyclone

Because the tropical cyclone damage was mostly confined through different characteristics of hazard bearing body, to several provinces, we perform hazard analysis at a comprehensive perspective. Specifically, we examine it through the model for general grade division of typhoon disaster which is conducted by affected agricultural land, number of death, number of collapsed houses, and direct economic losses (Wang et al., 2014).

The primary data source contributes to the Yearbook of Meteorological Disaster in China (Song et al., 2014), including the needed characteristics of hazard bearing body for affected provinces of each tropical cyclone.

There are two steps for the above model applying. First, based on the graded standard of tropical cyclone damage in China, as shown in Table 1, we can establish the non-dimension function transform (Wang et al. 2014).

<table>
<thead>
<tr>
<th>Item</th>
<th>Very Sever</th>
<th>Severe</th>
<th>Moderate</th>
<th>Slight</th>
<th>Negligible</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affected Agricultural Land</td>
<td>(10^1, +∞)</td>
<td>(10^0,10^1)</td>
<td>(10^-1,10^0)</td>
<td>(10^-2,10^-1)</td>
<td>(10^-3,10^-2)</td>
</tr>
<tr>
<td>Number of Death</td>
<td>(10^1, +∞)</td>
<td>(10^0,10^1)</td>
<td>(10^-1,10^0)</td>
<td>(10^-2,10^-1)</td>
<td>(10^-3,10^-2)</td>
</tr>
<tr>
<td>Number of Collapsed Houses</td>
<td>(2*10^8, +∞)</td>
<td>(10^8,2*10^8)</td>
<td>(10^7,10^8)</td>
<td>(10^6,10^7)</td>
<td>(10^5,10^6)</td>
</tr>
<tr>
<td>Direct Economic Losses</td>
<td>(3*10^10, +∞)</td>
<td>(10^10,3*10^10)</td>
<td>(10^9,3*10^10)</td>
<td>(10^8,3*10^9)</td>
<td>(10^7,3*10^8)</td>
</tr>
</tbody>
</table>

Table 1. Graded standard of single disaster item in China

Second, by using the theory of grey association analysis (Yang, 1997; Wang et al. 2014; Fu, 1992), the association degree can be calculated and the general disaster level can be confined. The primary data and assessed results are summarized in Table 2. And the grading results covers all degree classes.

2.3. Raw Weibo Data and Filtering

The row data set for tropical cyclones in 2013 was collected through Weibo Application Programming Interface (API). And
the time interval of Weibo data was based on the duration of each tropical cyclone. Considering the post-evacuation process after the passage of storm, the time interval of each event extended three days after the ending of storm. Previous studies (Kryvasheyeu et al., 2016) suggested that there is a strong inverse relationship between distance to path hurricane and related social media activity, because the closer area can be directly hit by or close to. Therefore, the Weibo data gained from affected province can better contribute to the assessment of disaster situation.

All the Weibo data contain the text of the message, location of posting, and a range of additional significant information, such as time, user identification, follower counts, and so on. Due to repost messages in Weibo without location information, we only obtained the original to perform finer spatial analysis.

There has been study (Martin et al., 2017;) proved that tropical cyclone related social media activity can reach peak during pre-impact and preparedness phase and decrease after passage of storm. Moreover, previous study (Kryvasheyeu et al., 2016) also suggested that per-capita activity closely correlates with the per-capita economic damage. Therefore, we’d like to directly filter tropical cyclone related data to do further study in quantity relationship.

The Filtered data set should include some words which can convey the theme relevant to tropical cyclone event. There are Chinese keywords shown in Table 3, like “storm”, “super-storm”, which are selected from news, Weibo data samples and other text corpus. In total, for 10 tropical cyclones in 2013, we obtained 35,991 messages from unique 32,586 users.

<table>
<thead>
<tr>
<th>Name of Tropical Cyclone</th>
<th>Formation of Tropical Cyclone</th>
<th>Tropical Cyclone Degree</th>
<th>Warning Notice</th>
<th>Disaster Situation</th>
</tr>
</thead>
<tbody>
<tr>
<td>台风</td>
<td>台风</td>
<td>强热带风暴</td>
<td>暴雨</td>
<td>暴雨</td>
</tr>
<tr>
<td>洪灾</td>
<td>洪灾</td>
<td>灾害</td>
<td>灾害</td>
<td>灾害</td>
</tr>
<tr>
<td>资料</td>
<td>资料</td>
<td>资料</td>
<td>资料</td>
<td>资料</td>
</tr>
</tbody>
</table>

This is Table 3. The keywords for Weibo data filtering

2.4. Weibo Topic Heat

Weibo topic heat (H) here raised is one general comparison metrics to show Weibo activity. It can be conducted with amount of active users, number of text posted, number of active users’ followers, and the classification of verified users. In general, posted messages from users with more followers have greater influence, and gain more attention from public. Also, messages from verified users mostly convey real, valid, and noticeable information.

Relying on the location information of filtered Weibo data, we did a spatial analysis through thermodynamic diagram shown in Figure 2. This diagram directly shows the tropical cyclone path, the situation of landform, and the Weibo activity in each storm event through the density of posted messages. As we can see, the most active areas were not the location closest to the tropical cyclone path, but the spaces were, in affected provinces, with flat terrain, large cities and inferred great population. Therefore, such areas with high-density fortune and population likely suffer more losses and have plenty active Weibo users to post related messages. Taking all the factors into consideration, the situation of underlying surface, the area of affected place, and the degree of hazard can be partly reflected from the number of messages and users.

![Figure 2. The thermodynamic diagram of tropical cyclone related Weibo activity in affected provinces](image_url)
3. RESULTS

3.1. Weibo Topic Heat Response to the grade of damage situation

Weibo topic heat is put forward to quantitatively assess Weibo activity. For each tropical cyclone, we can calculate H results of each affected province and the synthesis based on Equation (1). Combining the data of H results above and the Association Degree (α) shown in Table 2, we analyze the quantitative relationship of \( H - \alpha \).

The result of analysis, presented in Figure 3, shows strong positive relationship between H and α. From a general view of the data points in Figure 3, Weibo topic heat of tropical cyclone grows gradually while association degree increases.

Relying on the scatter points, we create the fit curve to describe the quantitative relationship of H-α. In fact, Weibo topic heat cannot expend unlimitedly, as the amount of Weibo users in affected area is invariable. Therefore, logarithmic curve always stands out in all curve choices to present the changing trend.

![Figure 3. The fitting curve of Weibo Topic Heat and Association Degree](image)

The equation of fitting result is as follow:

\[
H = 1.8845\ln(\alpha) + 15.636
\]

The significance of this regression is 0.000 presented in Table 4 and the \( R^2 \) is 0.3813, which both prove the high level of goodness of curve fitting.

<table>
<thead>
<tr>
<th>Quadratic Term</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>32.812</td>
<td>24.037</td>
<td>0.000</td>
</tr>
<tr>
<td>Residual</td>
<td>38</td>
<td>1.365</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>46</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4. The ANOVA of H-α curve fitting

Arguably, the Weibo topic heat can be higher with the increasing degree of disaster. It also proves the result of previous study that there is strong correlation with per-capital social media activity and the per-capital economic damage (Kryvasheyeu et al., 2016). This phenomenon can be explained by the dissemination of emergency information in different forms through media. More severe disaster, more people affected or close to, and finally more related information posted in Weibo.

4. DISCUSSION

4.1. Finer spatial analysis

There is limitation to the approach in this paper. In fact, finer spatial analysis can better assist governments in hazard management. They eager to grasp real-time information in each affected county even finer scale to allocate resources.

However, when we pay attention to it, we not only need support in data source, but also should consider complex situation in real life. The density of related Weibo activity shown in Figure 2 has proved that there is no longer strong inverse relationship between the distance to path of cyclones and Weibo activity in county scale. There are a number of factors contribute to this result, including underlying surface, population, losses amount and others. For example, places with expanse of crop and large population can suffer more losses during one strong tropical cyclone. On the contrary, the losses at mountainous area with few people can be much less and negligible.

Further, we need to consider some situations in real-life, which may disturb the assessment by using Weibo topic heat. For example, people directly hit by very severe typhoon do not have mood and internet environment to post Weibo text, which may lead to lower Weibo heat in affected severely area. Moreover, there may be not great population to post Weibo in the countryside with badly affected agricultural land. Therefore, whether the conclusion of previous study (Kryvasheyeu et al., 2016), per-capita Twitter in terms of per-capita economic damage, can be adaptive finer scale in China is still need to be investigated.

4.2. The applicability of H-α equation

With time going, all factors can be changed. For example, the amount of Weibo users continues growing larger; the efficiency of hazard management becomes better; And the frequency of tropical cyclones may increase, which may lead more people pay attention to cyclones. Therefore, the applicability of H-α equation should be modified and verified by several-years data for further application.

5. CONCLUSION

This article did an analysis on the quantitative relationship between the grade of hazard situation and Weibo topic heat, which can be a new potential means to assess the situation in risk area at real time.

Considering the open source and tags of data, we finally utilize the Weibo API to acquire social media data to monitor social response to tropical cyclone events. Also, the hazard information and analysis here is based on the provincial scale.

The Weibo topic heat was conducted by several factors, including the number of posted messages, the amount of active users, the followers of user and whether the user being verified. Taking the relationships with each factor into consideration, we put forward to the Weibo topic heat equation.
Combining the results of Weibo heat and the association degree (α, utilized to describe the general degree of damage), we use logarithmic curve to fit it. Finally, we acquire one equation of H-α as followed H=1.8845ln(α) + 15.636, which got the high level of goodness of curve fitting. This result suggested that this equation can be made a good use in further rapid assessment application.

ACKNOWLEDGEMENTS

We sincerely thank Adu Gong as supervisor for constructive comments on an earlier version of the manuscript. And we also thank Dongyu Zhou of GeoHey for guidance on coding and which improved the efficiency of Weibo data acquiring.

REFERENCES


Martin, L., Cutter, 2017. Leveraging Twitter to gauge evacuation compliance: Spatiotemporal analysis of Hurricane Matthew, 12(7), e0181701.


Revised March 2018