The Role of Social Media in Measuring Human Response to Urban Flash Flooding

By
Fei Xu

Bachelor of Landscape Architecture
China Agriculture University, 2008

Master of Geography (Landscape Architecture)
Peking University, 2010

Submitted to the Department of Urban Studies and Planning
in Partial fulfillment of the requirement for the degree of

Master in City Planning

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2015

Author________________________ Signature redacted
Department of Urban Studies and Planning
May 21, 2015

Certified by________________________ Signature redacted
Professor Joseph Ferreira
Department of Urban Studies and Planning
Thesis Supervisor

Accepted by ________________________ Signature redacted
Professor Dennis Frenchman
Chair, MCP Committee
Department of Urban Studies and Planning

The Role of Social Media in Measuring Human Response to Urban Flash Flooding

Members of the public...leverage their own social networks to find and provide information outside the official response effort, and to make critical decisions about, for example, heeding warning and making plans to evacuate (Mileti, et al., 2006).

Keywords: social media, disaster, sentiment analysis, sentiment map

Abstract

This thesis explores the role of social media in urban flooding. The author analyzes the activity of weibos (Chinese tweets) related to Beijing's "7.21" flash flood in the Sina-weibo system (the most popular social media open platform in China) and characterizes these weibos from the 37 hours following this disaster.

In order to understand the response of the public to urban flash flooding, multiple methods are used, including trend analysis, content analysis for high frequent terms and co-occurrence words, and lexicon-based sentiment analysis. In particular, weibos with geo-location information were extracted to draw different sentiment maps for the city. Sentiment maps show the public emotion (polarity, intensity and type) geographically. Through these analyses, I set out to construct a framework to process massive amount of data generated by social media, and proposed a methodology for converting the data into actionable knowledge.

This work explored extracting emotions from weibos, distilling crowd-wisdom by using filters and algorithms to smooth out the noise in the massive amount of data, and determining that human emotions closely correlated with severe natural disaster. By tracking human emotions, it was possible to track the progress of the disaster, and more importantly whether relief was provided to mitigate the disaster. Moreover, by projecting the emotional polarity, intensity and emotional type onto maps, the visualization can provide reasonably clear and timely picture of when and where the strongest emotions occurred. The methodology developed in this thesis could facilitate innovative approaches in the field of urban disaster planning.

Supervisor: Joseph Ferreira

Reader: Sarah Williams

Acknowledgements

Can't imagine I am going to complete this journey.
Cannot believe I have done so much. I do not know why I choose the hard level at the beginning, but the good news was that everything was absolutely new for me. Every step was the first step. I should be proud of myself.

I have a lot of thank-you s, for helping me come to this.
Thank you, Joe and Sarah. You taught me a lot and always encouraged me.
Thank you MIT, DUSP and CDD. You are the part of my life now. Thanks for giving me opportunities to join you and I appreciate every moment with you.

Thank you, Li lulu, you are amazing. You are the angle and make my June brighter.
Thanks to my great friends, Meng Ren, Zelin, Luxi, Qiuying. You are wonderful girls, and always support me. I would love to be your elder sister forever.
I would like to thank Kate, Callida and Thierno who help editing my thesis.

Last but not least, I want to thank my mom and dad. You are not perfect parents, but good enough. No words can fully express my gratitude. Love you forever.

Table of Content

List of Figures

List of Tables

Chapter 01 Introduction

1.1 Background

1.1.1 Global Flooding Challenge

Flooding is a common phenomenon in the world. Figure 1.1 shows that the occurrence of flooding and storms are the two most frequent types of natural disasters. From 1950 to 2015, floods hit the African continent 882 times, the Americas 1001 times, Asia 1801 times, Europe 540 times, and Oceania 132 times. 178 million people were affected by floods in 2010; the total financial impact of losses exceeded $49 billion (World Bank 2012). In 2012, 64 million people were affected and 3549 died due to floods; the total financial impact of losses exceeded $ 25 billion (Data from EM-DAT/CRED). Figure 1.2 shows the trend of flood events in the world. Unfortunately, one can see that the frequency of flooding is increasing. Floods can be divided into four categories: riverine flood, flash flood, coastal flood, and mixed causes flood.

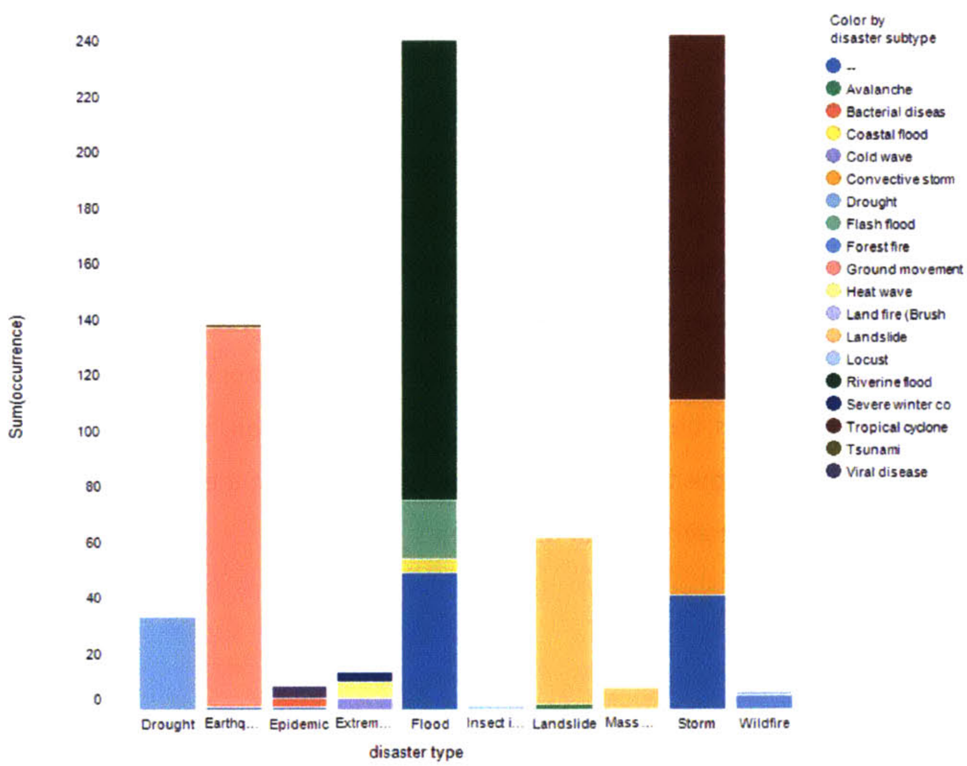


Figure 1.1 Histogram of all natural disasters (1950-2014) Data from EM-DAT/CRED

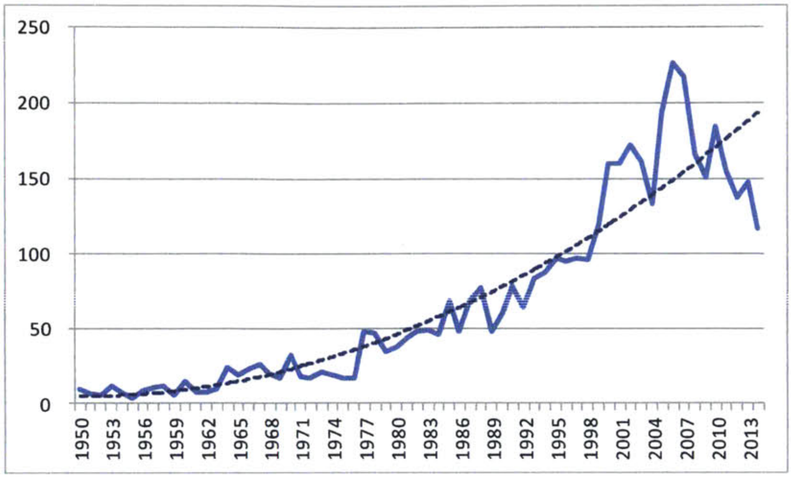


Figure 1.2 Numbers of reported flood events (1950-2014) Data from EM-DAT/CRED

Floods are not alike. Some floods develop slowly over a period of days. But flash floods are rapid flooding, and develop quickly. They may be caused by heavy rain or "without any visible signs of rain" (Federal Emergency Management Agency). Recently, flash floods have emerged as a serious and growing development problem in the world (World Bank, 2012). In rural areas, flash floods often surpass existing rivers, carrying rocks, mud or other debris and sweeping away everything in its path. In the city, flash floods are also destructive and can produce a significant negative influence on unprepared citizens. Since 1990, occurrences of flash floods have increased dramatically which brings about even tougher challenges (Data from EM-DAT/CRED).

In China, flash floods have increased considerably in recent years and have caused huge economic losses. Previously, those floods were more commonly seen in southwest and south of China, such as Sichuan, Guizhou, Hunan, and Guangdong Province. Not only have those places seen more frequent flash floods, but in light of recent rapid urbanization, more and more big cities are also experiencing serious flash floods. Beijing is one such city; the "7.21" Beijing flood was one of the most serious catastrophes in recent years.

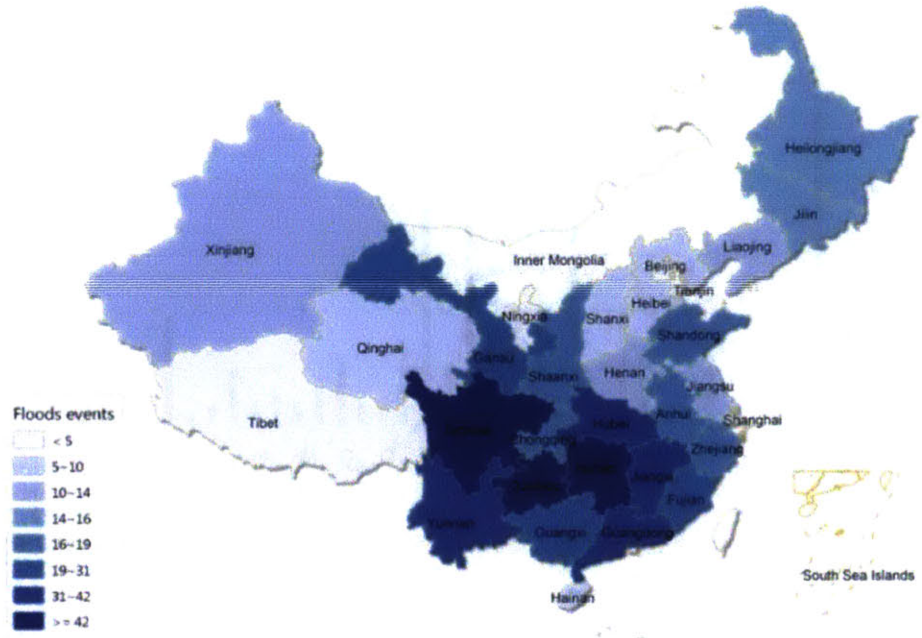


Figure 1.3 Numbers of flood events in China (1980-2014). Data from EM-DAT/CRED

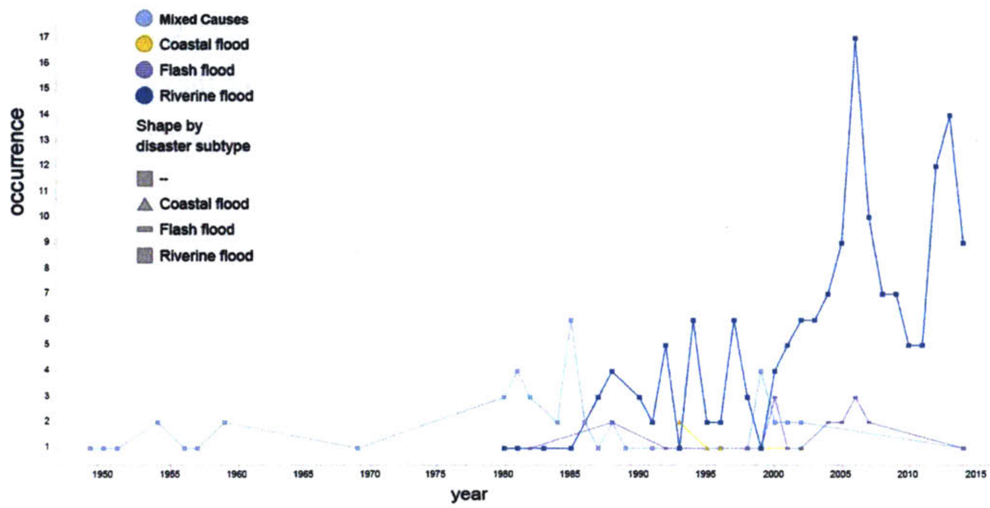


Figure 1.4 Numbers of reported flood events in China (1950-2014). Data from EM-DAT/CRED

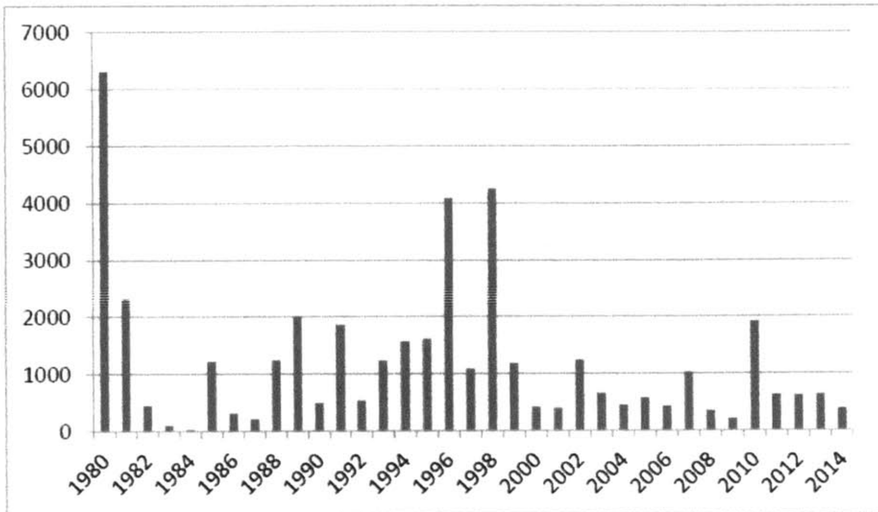


Figure 1.5 Numbers of death caused by flood in China (1980-2014) Data from EM-DAT/CRED

1.1.2 The Role of the City Planner in Urban Floods

How do flash floods influence the city? What are the responsibilities of city planners? Urban flash floods change urban physical surfaces and also disable urban functions; roads become rivers, residential areas become islands, and cars become boats if they can. City planners are responsible for preventing the development of flash floods and ensuring the city works well in the rain. Unfortunately, the high frequency of flash floods has generated public frustration and annoyance. More and more citizens blame city planners as part of the destruction. Although city planners are improving disaster response and resiliency, understanding public emotions may help city planners to manage the city.

The generally accepted classification of disasters and disaster response divides it into four phases: preparedness, response, recovery, and mitigation (Drabek, 1986). Two areas of city planning correspond to these four phases: one is Urban Storm Water Management (USWM) and the other is Urban Flood Management (UFM). The former focuses on how to prepare for the flash flood, and the latter emphasizes disaster relief.

Current USWM system can be summarized into three aspects, which also shows how this field has developed. The first aspect focuses on the drainage system, which relies on engineering techniques to enhance the system's capacity to combat disasters. The second one is water sensitive urban design. Many city planners, especially urban designers, are trained to develop suitable urban environments that create healthy ecosystems by integrating water sensitive design into the entire water cycle. The third one is related to "Green infrastructure" theory. This is an up-to-date method to manage flash floods in which vegetation, soils, and natural processes are used to prevent flooding. With this theory, more elements of landscape architecture are involved in city construction and can help support creating healthier urban environments. On the other hand, flood management also involve other tasks, such as informing the public through early warning notifications and situation updates, responding to and investigating floods, managing rescue and evacuation operations etc.

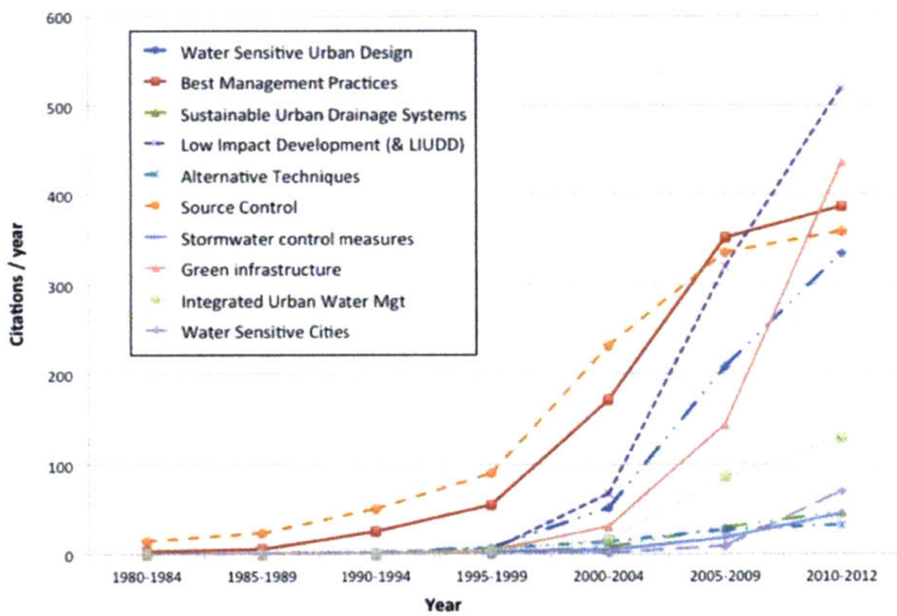


Figure 1.6 Evolution of new urban drainage terminology in the 32 years from 1980 to 2012. Retrieved from "SUDS, LID, BMPs, WSUD and more – The evolution and application of terminology surrounding urban drainage", Fletcher et, all, 2014. The data were extracted from Google Scholar on 23/09/2012.

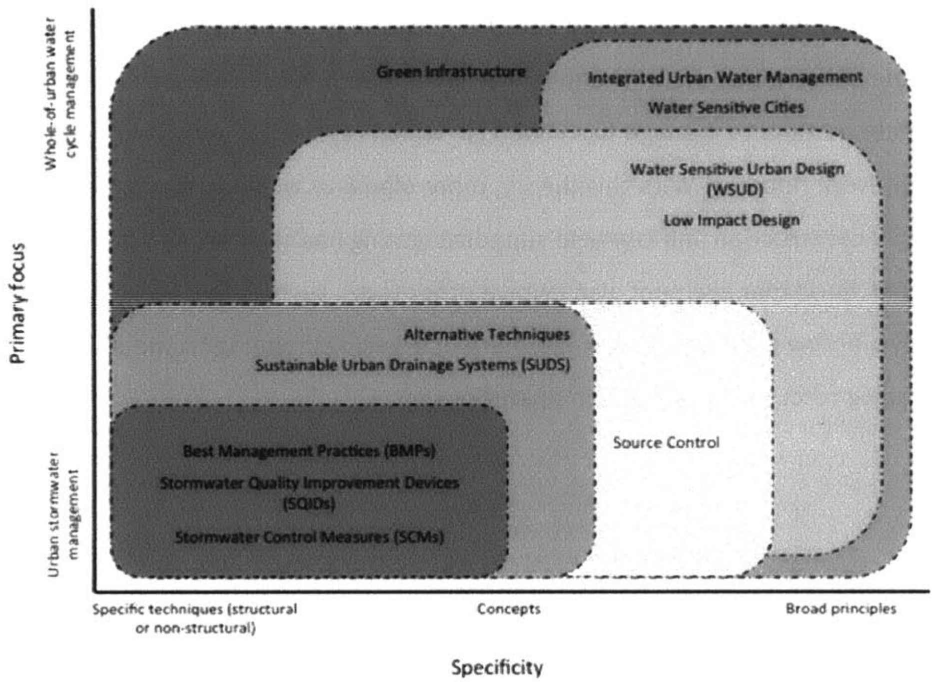


Figure 1.7 One possible classification of urban drainage terminology according to their specificity and their primary focus

Note: Retrieved from "SUDS, LID, BMPs, WSUD and more – The evolution and application of terminology surrounding urban drainage", Fletcher et. all, 2014

Previously, as a landscape architect and urban designer, I learned a great deal from USWM, but have no ways to verify these theories. Meanwhile, as a city planner I feel vulnerable and helpless when floods occur. Urban development should be guided by feedback; community participation should be encouraged and brought into the planning process. This is especially important in China, where urban development is generally managed in a top-down method. City planners need an effective approach to understand people; in this way, crowdsourcing [1] can be really powerful. Social media can create such platforms for community participation via web 2.0 techniques. "Big data" from social media platforms provide an opportunity for planners to integrate bottom-up and top-down processes and facilitate a more disaster-resilient city construction.

1.1.3 Social Media and Disasters

[1] Crowdsourcing is a way to obtain needed information or request services by soliciting contributions from a large group of people via web 2.0 techniques.

Social media is an emerging industry drawing on 'big data', which has the potential to discover unprecedented knowledge (Taylor Shelton, 2014). Social media depends on mobile and web-based technologies to develop platforms that emphasize user-generated content, usability, and interoperability (H. Kietzmann et.all 2011). It is exploding worldwide. Twitter is the most popular micro-blogging system in the world. It's also one of fastest growing social media platforms in the world, with more than 500 million users (Paul et al., 2015). Today, social media is still growing rapidly and maturing with time.

Social media in China started in 2009, but China has become an area with the most active users of social media in the world. "It has had the world's most social media active users in 2012" (McKinsey, 2012). The 2012 and 2014 'CIC China Social Media Landscape' maps show hundreds of current social media platforms in China and their prototypes, including 'Answers', 'Wikipedia', 'Twitter', 'Facebook', 'YouTube', 'eBay' etc.

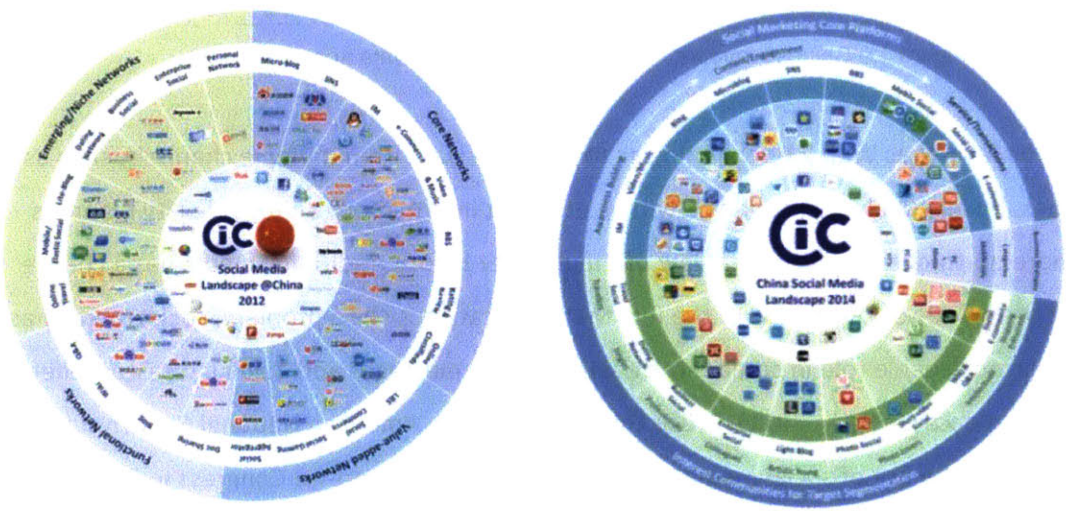


Figure 1.8 CIC China Social Media Landscape 2012 and 2014, Retrieved from http://www.ciccorporate.com

Sina-weibo is the top open social media platform in China. According to the McKinsey survey, Chinese consumers ranked the following social-media sites: Q-zone (starting from 2005), which 44% of the surveyed chose as their most frequently used social media tool; Sina-weibo (starting from 2009) and Renren (starting from 2005), each favored by 19% of respondents; Tencent Weibo, and Kaixin were chosen by 8 percent and by 7 percent, respectively (McKinsey, 'China's social-media boom', 2012). Q-zone is a private platform where only friends can share information. The second one, Sina-Weibo, is an open platform which can be described as

Chinese Twitter. Sina-weibo is also one of the fastest growing social media platforms in china. Until December of 2012, its registered total users reached 358 million, and it has roughly 36.5 million active users daily, who post more than 100 million weibos daily. Although Twitter had a total user base of about 500 million, they comes from it is all over the world worldwide, whereas Weibo is almost entirely Chinese, with no complete English interface.

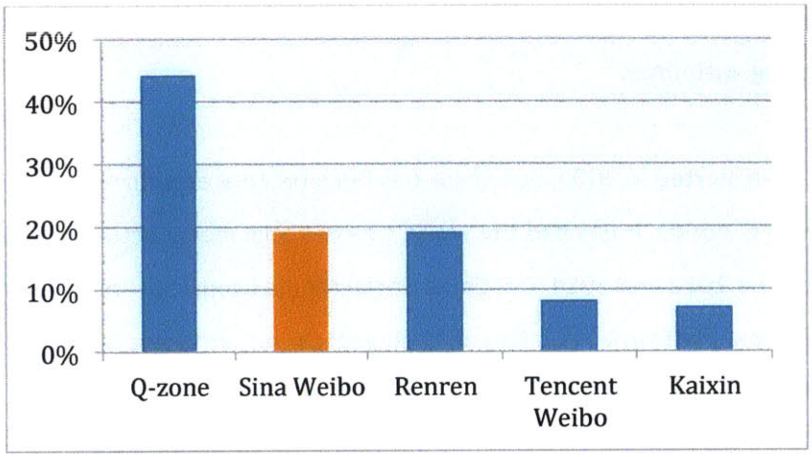


Figure 1.9 Chinese favorite social-media sites rank (Data from McKinsey)

Social media leverages Web 2.0 technologies to integrate data from multiple channels, such as cell phones, emails, Web applications, Twitter, etc. That information can provide city planners with a more up-to-date and broader view of the disaster before traditional media has a chance to report on the news. Regardless of the fact that verification is still a problem, social media has proved its irreplaceable role in spreading disaster-related messages and guiding disaster relief. (Huiji Gao, Rebecca Goolsby, 2011). Bruce Lindsay, Analyst in American National Government has encouraged the use of social media in disasters in the CRS report for Congress in September 2011.

1.2 Objects and Implication

Objectives

This thesis intends to explore how we can use social media data to understand crisis in the city and examine the feasibility of applying crowd-sourced data towards disaster management. It also looks to provide a preliminary framework with extensive analysis using real data from a real natural event in Beijing, China, and interpret the social media data into actionable knowledge for city planners.

Implication

This thesis' research has multiple important implications. One implication is the potential development of an efficient and low cost approach to understand the public response to disasters. In addition to a deficient drainage system and inadequate storm-water storage capacity, the incompleteness of knowledge about how the public responds to a flash flood is problematic. Social media is not a one-way communication, but flows in multiple directions. Thus, it can increase information channels dramatically. Besides the dissemination of information, it can support more functions including establishing situational awareness, receiving victim requests, and uploading images to estimate damage. Social media can help understand this knowledge and build an emergency response system.

Another important aspect would be the addition of a sentiment indicator to the current disaster defense system, improving its performance. Sentiment can measure public response to a disaster and filter the disaster-related information. This measurement can supplement a traditional warning system by reflecting affected areas and demographics. Real-time sentiment monitoring can help trace the path of the event, and measure the severity of the disaster, and can guide the after-flood reconstruction and improving flood-control projects.

Crowd-sourcing wisdom also has important implications for guiding city planning professionals. Using crowdsourcing power is possibly a smart way to address disasters and reduces the heavy burdens of the government. Compared to the cost to fix the entire city infrastructure, a more efficient resource allocation system is more affordable and more realistic. Moreover, it is always a good strategy to incorporate suggestions from the public into city planning practices.

Mapping the data and tracking the mood swings of the public may provide a totally different and more useful perspective for planners as well. Crowd sourced data with geo-tags produce more powerful guidance for disaster mitigation. Visualization can improve the accessibility of big data for the public. In addition, mapping moods and tracking them may provide an absolutely new perspective to capture response trends before and after disasters. The spatial patterns of emotions also help planners understand the relationship between public emotion and urban environment.

1.3 Questions and Methodology

Questions:

1. What approaches can be used to extract information and emotion from social media?
2. What measurements can be used to capture people's responses to disasters?
3. How is data translated into actionable knowledge?
4. Besides locating disaster areas, what else can we learn from geo-location messages?

Methodology:

I used qualitative analysis, quantitative analysis, and visualization. Qualitative analysis includes literature review and case studies. I also take some weibos as examples to explain what kind of messages was posted before, during, and after disasters. I applied several quantitative analysis methods. I relied on time series analysis in the trend analysis. As part of the content analysis, I extracted opinions based on high-frequency terms and co-occurrence words. The sentiment analysis relied on a Chinese-defined lexicon-based method to extract subjectivity and polarity. As part of the geo-location information analysis, semantic orientation refers to the polarity, emotion type, and strength of text. I also visualized the data and mapped the public mood reflecting the disasters as part of this analysis.

Chapter 02 Literature Review

2.1 Use of Social Media in Disaster Response

2.1.1 Individual: Share Information and Request for Help

A lot of literature recorded how Information and Communication Technologies (ICTs) were adopted in citizen-driven emergency response methods (Jaeger et al.). Residents affected by the 2004 and 2005 Gulf Coast hurricanes used public libraries to contact friends and families, find missing people, and obtain latest news about disaster (Hager,2015). A UK local farmer community used a grass roots computer network to share information about the foot and mouth disease crisis (Torrey et al., 2007). A Chinese online forum was developed in the 2008 Sichuan earthquake. (Qu et al., 2009) Researchers from Maryland university developed community response grids (CRGs), aiming to" provide local information and emergency instructions geared toward the immediate needs of residents in response to an emergency" (Shneiderman & Preece, 2007).

Micro-blogging systems, where registered users can read and send 140-character messages, have drawn increasing attention from research communities recently. Compared with other ICTs, micro-blogging systems (like Twitter, Sina-weibo) have shown some advantages such as speed, easy access, easy editing, and interactivity. There has been an increasing amount of literature on how Twitter was used in crises, such as in grassfires (Corvey et al., 2010) and river flooding (Vieweg et al., 2010). Vieweg et al. (2010) divided tweets into three categories: "namely, repeated information, links to information on external websites, and retweets (RTs)". Starbird et al. summarized four types of information evolution: "generative, derivative, synthetic, and innovative activities"(Starbird, K., Palen, 2010). They also explored what kinds of messages tend to be reposted.

At the Individual level, a micro level, Twitter use focuses on information sharing. Besides revealing the role of social media in situation update, other discussions in the literature mainly focus on two questions: what kind of information can attract people in different phrase of the disaster, and how the disaster-related information spread in the tweeting and retweeting process?

2.1.2 Organization: Update situation and Coordinator

Researchers have demonstrated the power of crowdsourcing to disseminate news-related information (Kwak et al., 2010). Media crowd-sourcing has been under the lens of researchers with regards to its use in disasters and other high profile events (Hui et al., 2012). Moreover, official sources tend to be too lagging, vague, and inaccurate for many reasons. Social media data may supplement these issues. "The medium is also seen as a place for "harvesting" information during a crisis event to determine what is happening on the ground (Palen et al.)

They focused more on the challenges of extracting action items and locating information. Compared to data collection, "a crowd-sourced filter" seems to be "a big thing" (Erik Herdsman, 2010). Organizations have to deal with these millions of messages first before using and publishing them. There are numerous challenges in extracting information such as issues of "verification, quantification of performance, translation and other technical barriers" (Hughes et al., 2009; Mendoza et al., 2010; Starbird et al., 2010; Tapia et al., 2011). Mendoza et al.(2010) examined the true and false information that posted on Twitter after the 2010 Chile earthquake. The findings revealed that the rumor propagation in social media is different from traditional news. The former is more likely to be doubted and denied by the Twitter users. Mendoza et al. also developed an approach to determine which tweets were factual based on the way the information was spread, with 70%-80% accuracy (Mendoze et al. 2011).

In case of the Haiti earthquake, thousands of volunteers met a data access problem. Outside data, especially massive social media data outputs, was not allowed to be incorporated into policies and procedures. They also needed technical support to mapping the disaster, coordinating rescue efforts, and filtering messages (Portsea, 2011). Volunteers had to extract factual messages manually. Although the verification issues were solved (Palen et al., 2009; Starbird et al., 2010), it still fell short in bringing adequate rescue, relief, and recovery efforts to affected people(Walton et al., 2011).

Sudha et al. (2011) built a classifier to extract tweets that could contribute to situation awareness. They collected tweets from three different crises, including the 2009 Oklahoma grass fires, the 2010 Red River flood, and the 2010 Haiti earthquake. They used both "hand annotated and automatically-extracted linguistic features" to extract related messages (Sudha et al.). Their system was over 80% accurate. The study drew attention to the problems faced by Twitter users who struggled to identify what information was important during the disaster; however, it was concluded that because of the varying length of entities, annotators cannot be assigned simple rules."

2.1.3 Disaster maps: Visualize and Geo-locate

Most people have difficulty dealing with large amounts of information. Visualizing texts can help them obtain useful information quickly and accurately. Additionally, responders can accurately locate specific requests for help on the map.

Case of "petajakarta.org/banjir" Flood platform

"petajakarta.org" is the first collaboration program in the world between a government disaster management, a university and Twitter. This project works on both model and real-time disaster-related responses. (Mark Gillis, 2014) Through post a message @ petajkt #banjir, people can report a flood-related tweet via twitter. If they open the geo-location function, they can flag the flood on the map. The public can use these community flood maps to check the situation and avoid danger while BPBD Jakarta's incident control room can use them to monitor real-time disaster response.

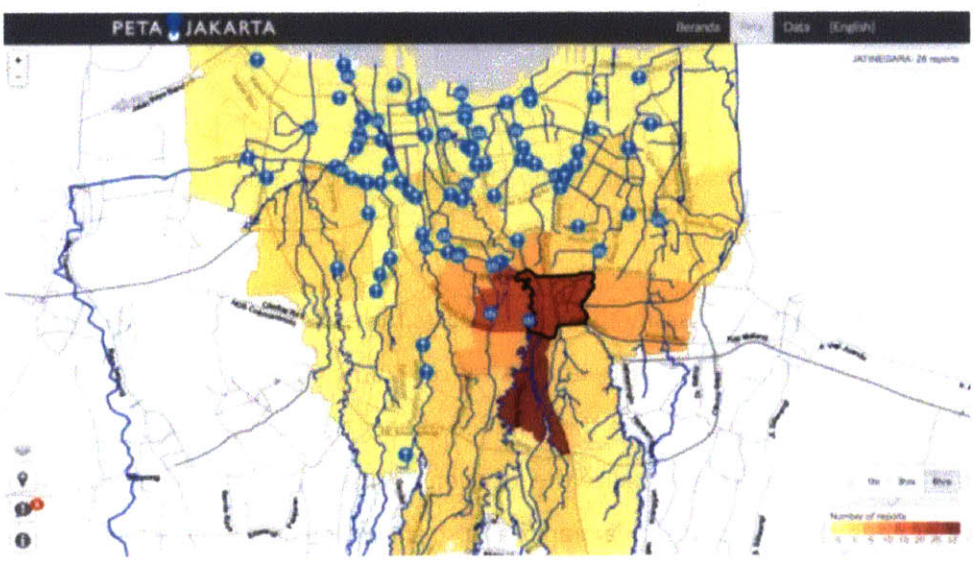


Figure 2.1 "petajakarta" Flood Map, Retrieved from http://petajakarta.org/banjir/in/map/

Case of Haiti Earthquake

The 2010 Haiti earthquake was a catastrophic magnitude 7.0 Mw earthquake, in which more than 222,000 people died and 300,000 were injured (Data from Haitian government). People published numerous texts, photos, and videos of their experiences on twitter, flicker, facebook and YouTube after it happened. Ushahidi, an open source crisis map platform, was created in the first two hours after the earthquake struck. It captured, organized, and shared important disaster-related information coming from social media. 4636 Volunteers manually filtered numerous messages and created 3,596 reports that enough relevant information to be mapped. Those crisis maps identified "clusters of incidents and urgent needs, helping responders target their response efforts." Moreover, Red Cross also confirmed that they received US$8 million through Ushahidi.

Figure 2.2 Ushahidi Disaster Map, Retrieved from http://voices.nationalgeographic.com

Case of Hurricane Sandy

In Cornelia, et al.'s work, they classify the Twitter posts the occurred during Hurricane Sandy based on the user's sentiment. Using Twitter data from Hurricane Sandy they identify the sentiment of tweets and then measure the distance of each categorized tweet form the epicenter of the hurricane. They visualize the location of different user's messages , categorized by positive or negative sentiments on a map that also shows the proximity to the hurricane. Cornelia, et a., show how users' sentiments change according to users' locations and also on the distance users are from the hurricane.

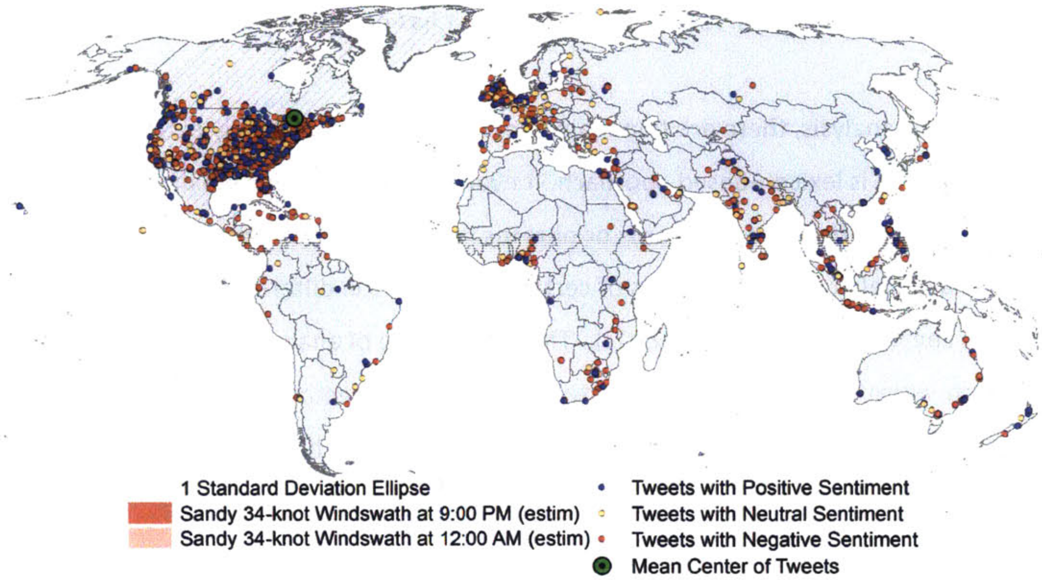


Figure 2.3 Hurricane Sandy Mood Map

Note: Retrieved from "Mapping Moods: Geo-Mapped Sentiment Analysis During Hurricane Sandy" Cornelia et.all 2014

2.2 Semantic orientation

One way to explore human sentiment is to utilize semantic orientation. Semantic orientation (SO) usually captures an evaluative factor (positive or negative) and potency or strength (degree to which the word, phrase, sentence, or document in question is positive or negative) towards a subject topic, person, or idea (Osggood, Suci, and Tannenbaum 1957). It is viewed as a tool for measuring of subjectivity and opinion.

Modern semantic orientation refers to Natural Language Processing (NLP) to identify and extract subjective information in source materials .NLP algorithms are based on machine learning, especially statistical machine learning. The analysis of semantic orientation can be found under different umbrella terms: point of view (Wiebe 1994; Scheibman 2002), evidentiality (Chafe and Nichols 1986),sentiment analysis (Pang and Lee 2008), subjectivity (Lyons 1981; Langacker 1985), opinion mining (Pang and Lee 2008), analysis of stance (Biber and Finegan 1988; Conrad and Biber 2000), appraisal (Martin and White 2005), and a few others, without expanding into additional disciplines and into the study of emotion (Ketal 1975; Ortony, Clore, and Collins 1988) and affect (Batson, Shaw, and Oleson 1992). In marketing, SO has been widely used to interpreted crowd opinions automatically and measure popularity, success, or failure, popularity and success, and compiling reviews. However, in planning it has not drawn much attention.

In terms of sentiment analysis, there are two main approaches to automatically extracting sentiment. The first one is lexicon-based approach. It evaluates orientation for a document by calculating the semantic orientation of words or phrases in the document (Turney 2002). Text classification can also help to label instances of certain sentiments in different texts or sentences (Pang, Lee, and Vaithyanathan 2002). The second type of analysis is a machine-learning approach, a statistical analysis, using Bayesian or Support vector machines models. I followed the lexicon-based approach to do sentiment analysis, in which I use dictionaries of words annotated with subjective and degree.

2.3 Related Works in Chinese Context

2.3.1 Social Media Related

Relevant literatures on social media in Chinese context focus on qualitative analysis. Regarding weibo (Chinese twitter), many studies suggest it is an efficient way for the government to monitor public sentiment. In terms of quantitative analysis, studies on that measure how information spreads attracted more attention due to its potential marketing value. In Wan Weiguo's doctoral dissertation, he discussed "User Behavior Analysis and Network Evolution in Microblogging Networks" (Wan Weiguo, 2014). What did he say about it?

In terms of how social media has been used in disasters, Yan Qu et al. (2011) analyzed the use of Sina-Weibo in response to the 2010 Yushu earthquake. This work "supplements existing works with an exploration of a non-Western socio-cultural system: how Chinese Internet users used micro-blogging in disaster response."(Yan Qu et all, 2011) They still focus on distribution problems in the tweeting and retweeting process. Yan Qu et all categorized Sina-weibo messages into five types: opinion-related (33%), situation update (25%), general earthquake related (18%), emotion-related (16%), and action-related (4%). There has been no research working to identify the polarity of sentiments expressed by users during specific disaster events, let alone research trying to map these emotions.

2.3.2 Semantic Orientation in Chinese

Semantic orientation in the English context has not been well developed. In the Chinese context, this field just started and only has limited relevant literature on the topic. According to search report from National Knowledge Infrastructure (CNKI, http://en.cnki.com.cn/), 116 papers/thesis were found in China on Integrated Knowledge Resources Database. All of them were developed in the domain of computer science. Some researchers in computer science tried to extract emotions via lexicon-based method or machine learning approaches. They compared results based on different models and work for model optimization. However, the results are much worse than English due to the complexity of the Chinese language. Since there is no space between Chinese characters, words segmentation should be conducted first. Moreover, Weibo is a kind of folk language, a mixture of informal writing language, cyberspeak and oral expression which makes it more difficult to analyze.

There are two evaluation systems in Chinese polarity judgment: National Institute of Informatics (NTCIR) and Chinese opinion analysis evaluation (COAE5) (Lixing Xie,2013). I cited the optimal evaluation results from these two systems:

Table2.1 NTCIR-8 Best Judgment Result

	Subjective/Objective		Positive/Negative	
	Traditional Chinese	**Simplified Chinese**	**Traditional Chinese**	**Simplified Chinese**
Precision	56.37%	41.34%	76.48%	67.39%
Recall	85.71%	83.35%	53.03%	52.90%
F-Measure	68.01%	55.27%	62.63%	59.27%

Data from http://research.nii.ac.jp/ntcir/

Table2.2 COAE-09 Best Judgment Result

Judge	P@1000	Precision	Recall	F1	Accuracy-1000
1	0.662	0.662	0.158033	0.255155	0.158033
2	0.612	0.612	0.158268	0.245143	0.153268
3	0.544	0.544	0.149986	0.235142	0.149986

Data from http://www.ir-china.org.cn/Information.html

Chapter 03 "7.21" Flood Disaster

"Let's go to the forbidden city to see the sea" "go to Baishiqiao subway station to see waterfall" "we do not need to wait until '2012'" "enjoy Beijing National Water Park"(Sina-weibo, July ,2012)……

On the 21st July 2012, the largest rainstorm in the last 61 years hit Beijing. The city received an average precipitation of 190.3mm that day, which caused a flash flood, paralyzed commuters, destroyed at least 8,200 homes and led to more than 23 billion [yen?] loss. 56,933 people were evacuated, 79 people died from the flood. At Beijing Capital International Airport, the flood resulted in the cancellation of over 500 flights, stranding 80,000 travelers. The depth of standing water in railways reached 6 meters. Reservoirs were created under bridges. Interchanges became graves.

Table I. Some consequences of the "7.21" rainstorm in Beijing, on 21st of July, 2012 (The information was summarized based on: [1] Yang, Bai 2012; [2] Liu 2012a; [3] Liu 2012b; [4] Wang 2012; [5] Wang, Yu 2012; [6] Wang, Liu 2012).

Disaster	Number	Notes
Fangshan District in Beijing [1]		
Rainfall	Average 301 mm; Maximum 541 mm	RP is 500 years
Mountain torrents, rock-mud flow	Happened	
Flood flow (the Jvma River)	Maximum 2 570 $m^3 s^{-1}$	The largest since 1963
The Beijing Capital Airport [2]		
Cancelled flights	571	On the 21^{th} of July
Delayed flights	701	
People stranding in the airport	80 000	
Beijing subway and expressway		
Collapsed roadbed of Beijing subway line No. 6. [3]	1	
Flooded section of the Beijing-Hong-Macao expressway [2]	Length of about 900 m, water depth of 4 m average and 6 m maximum	
Water-logging under the Nangangwa railway bridge [2]	> 200 000 m^3	
Effects on human livelihood		
Human deaths	79	By the 6^{th} of August [4]
People affected	1.9 million	By the 23^{th} of July [5]
Houses collapse	10 660	
Crop damage	238 688 hectares	
Immovable cultural relics damage [6]	163	
	About 210 000 m^2	
	About 850 million CNY	
Total economic loss [1]	More than 140 billion CNY	

Figure 3.1 Some Consequences of "7.21" flood. Note: Retrieved from "Eco-hydrology and good urban design for urban storm water-logging in Beijing, China", Congying Li, 2012

Heavy rains are commonly seen as responsible for weather disasters in China. Many scholars call "7.21" flood disaster as "7.21"catastrophic natural disaster. However, I believe the disaster is a consequence of natural causes and human causes. We can get more insights from the following analysis.

3.1 Mega-city of Beijing

3.1.1 Urbanization of Beijing

- Beijing urban sprawl has created a mega-city in the world.

Beijing is the capital of the People's Republic of China and the second most populous city in the world (from Cities proper by population Wikipedia). It located in the Northern China (39°N, 116°E) with 16808 km^2 area.

According to Professor Lu Dadao, president of the Geographical Society of China (GSC), China's urbanization took 22 years to increase to 39.1% from 17.9%. It took Britain 120 years, the US, 80 years, and Japan more than 30 years to accomplish this same urbanization ratio. Beijing is definitely one of the fastest growing cities in China.

In the past 30 years, the population of Beijing has doubled and urban areas have expanded over 700% extending the city 32.07 km^2 per year (Wang Sisi, Kongjian Yu, 2011). Until 2012, the urbanization rate of Beijing has reached 86.2 %.(Capital urbanization Report,2013) Currently, Beijing has a total population of 21.148 million within the municipality, of which 18.251 million reside in urban districts or suburbs and 2.897 million lived in rural areas.(Beijing Bureau of Statistics)

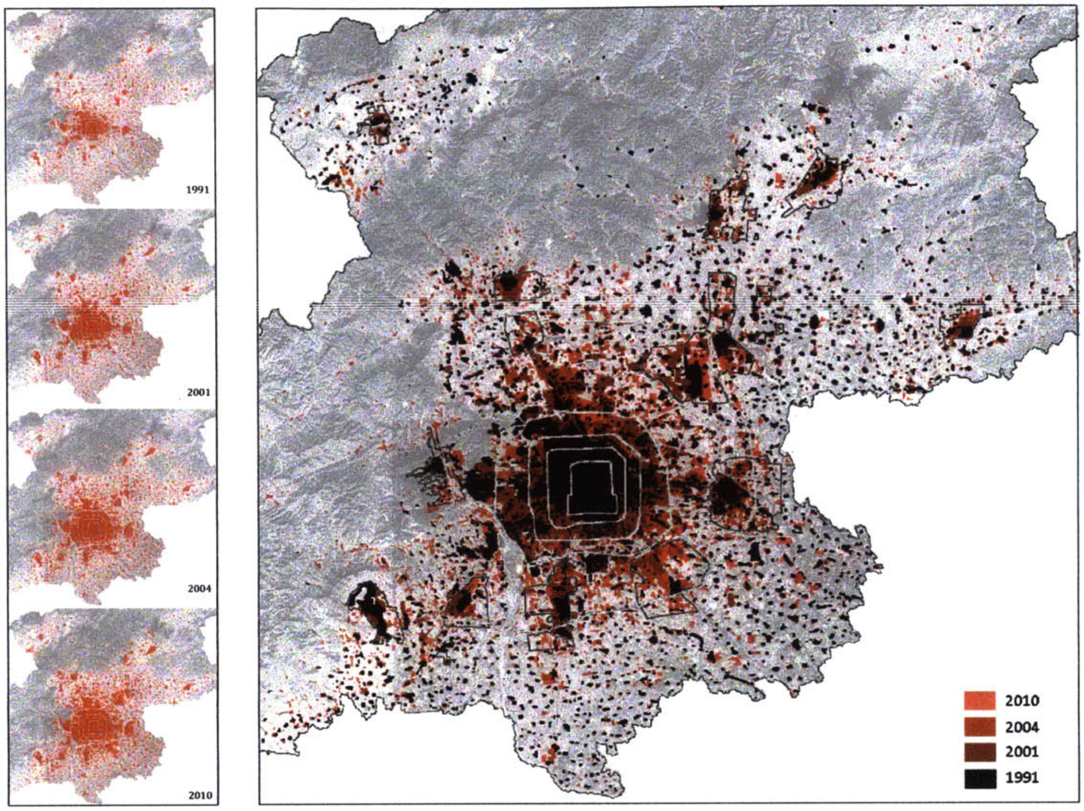


Figure 3.2 Urbanization of Beijing in 1991, 2001, 2004 and 2010

- Beijing can be divided into old city, inner suburbs, outer suburbs and rural areas. Urban floods are concentrated in the former three areas.

Beijing has 14 districts and 2 counties which can be divided into four categories based on the location and developed time. The Old city has historically been enclosed by city walls where much traditional architecture is located. The Old city is inside the Second Ring Road. The inner suburbs are located between the Second and Fifth Ring Road. The Outer suburbs linked by the Sixth Ring Road. The Old city, inner suburbs and outer suburbs are urban areas where citizens suffered heavy storm and flash floods. However, the Fangshan district is the hardest-hit areas of riverine flood and flash flood.

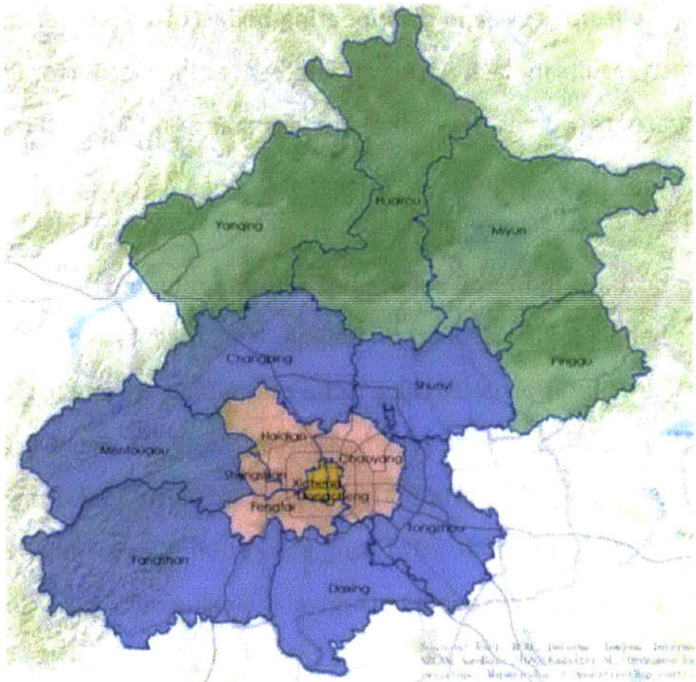


Figure 3.3 Districts and counties of Beijing

Table3.1 Basic information of districts and counties of Beijing

Data from Wikipedia and Baidu

Location	District / County	Population(2010)	Area(km²)	Density(per km²)
Old city	Dongcheng District	919,000	40.6	22,635
Old city	Xicheng District	1,243,000	46.5	26,731
Inner suburbs	Chaoyang District	3,545,000	470.8	7,530
Inner suburbs	Haidian District	3,281,000	426	7,702
Inner suburbs	Fengtai District	2,112,000	304.2	6,943
Inner suburbs	Shijingshan District	616,000	89.8	6,860
Outer suburbs	Tongzhou District	1,184,000	870	1,361
Outer suburbs	Shunyi District	877,000	980	895
Outer suburbs	Changping District	1,661,000	1,430.00	1,162
Outer suburbs	Daxing District	1,365,000	1,012.00	1,349
Outer suburbs	Mentougou District	290,000	1,331.30	218
Outer suburbs	Fangshan District	945,000	1,866.70	506
Rural areas	Pinggu District	416,000	1,075.00	387
Rural areas	Huairou District	373,000	2,557.30	146
Rural areas	Miyun County	468,000	2,335.60	200
Rural areas	Yanqing County	317,000	1,980.00	160

3.1.2 Water environment of Beijing

- Beijing is a water scarcity city.

Since 1990s to 2005, 80 acres (around 14%) of the urban bodies of water disappeared and another 10% bodies of water are threatened by urbanization till 2050. decreasing groundwater volumes. The growing built-up areas and disappearing bodies of water mean an increasing runoff coefficient which results in flash floods and decreases the water storage capacity. Currently, the fresh water resources per capital are less than 300m3 which is only 1/30 of the word's average and 1/8 of the national average (Cai,Ji 2010.)

- Beijing suffers from both drought and severe flash flood problems.

Since 1999, there has been a 9-year drought in Beijing. The precipitation each year was lower than the mean annual precipitation (Ji 2010). On the other hand, Beijing has experienced 20 heavy rainstorms within the most recent decade, such as the "7.10" event in 2004, the "7.31" in 2006, the "6.13" in 2008, the "7.13" in 2009, the "6.1" in 2010, the "6.23" in 2011, the "8.12" in 2013 and "9.02" in the 2014 (Chen et al. 2011; Lan, Yang 2009; Wang, Ma 2011, NetEase) .Besides, 22 urban flooding events occurred in Beijing during the flood seasons (June, July and August) from 2007 to 2014.(You et al. 2011, NetEase)

3.2 Climate and weather

3.2.1 Typical moderate continental climate

Beijing has a rather dry, monsoon-influenced humid continental climate, characterized by hot, humid summers due to the East Asian monsoon, and generally cold, windy, dry winters that reflect the influence of the vast Siberian anticyclone. The average daily temperature in January is −3.7 °C (25.3 °F), while in July it is 26.2 °C (79.2 °F). Precipitation averages around 570 mm (22.4 in) annually, with close to three-fourths of that total falling from June to August.

3.2.2 Weather Change

The average precipitation is 370mm per year from 1941 to 2008. The overall trend is a slight decrease, but it increased from 2003. The trends of rainfall and frequency for short-duration extreme events also showed a gentle wave pattern before 2003 and a sharp increase after 2013. In 2008, the extreme rainfall occurrence was 11 times. According to Beijing climate center, till 2050, the average annual rainfall of Beijing will increased 10.5% by the average rainfall from 1980 to 1990(Xiaoxin Zhang,2014).

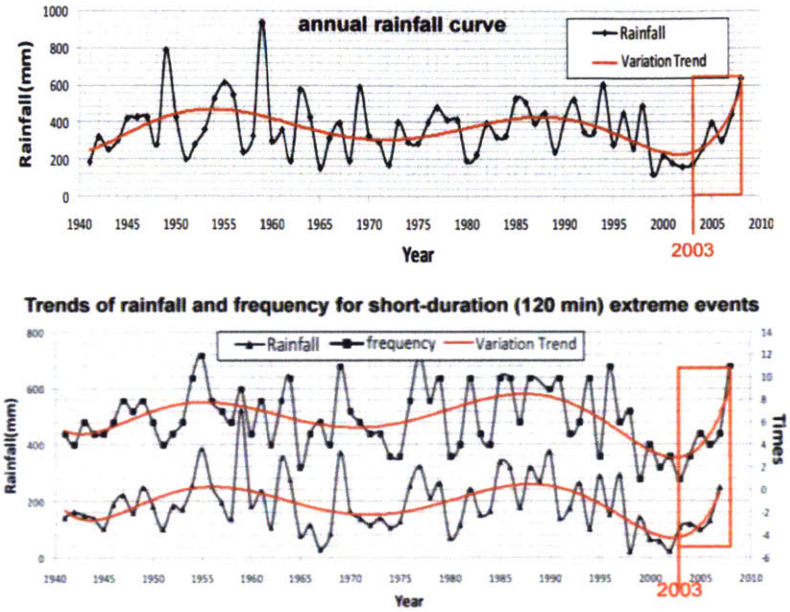


Figure 3.4 Annual rainfall trend and frequency for short-duration extreme events. Data from Beijing climate research center;

Note: Retrieved from Xiaoxin Zhang's report in 6th international conference on flood management (ICFM6)

3.2.3 Extreme Weather in "7.21" Flood Disaster

On 21 to 22 July 2012, the heaviest storm, characterized by great rainfall amount and intensity, wide range, and high impact, in over 61 years hit Beijing (Xiaoman Jiang et all, 2014). China Meteorological Data Service System (http://cdc.nmic.cn/) provides "hourly merged precipitation product with 0.1◦ × 0.1◦ resolution based on AWS (automatic weather station) observations in China and CMORPH (Climate Prediction Center MORPHing technique) satellite data" (Xiaoman Jiang et all, 2014). However, only the Chinese mainland authorized accounts can download the data. This product was considered the best data to describe the rainfall of this disaster. Although CMORPH showed a lower rainfall extreme intensity (389mm) than the observed (460mm), "CMORPH merged data have relatively high quality which shows an accurate location of precipitation". (Jiang Xiaoman, Yuan Huiling, 2014)

Researchers working outside the mainland of China cannot access to this data. I have to cite previous research using the same data to demonstrate the weather. "As shown in Figs. 2a1–a4, the precipitation area is quite concentrated with prominent mesoscale [2]feature. The rain starts to intensify at 14:00 and gradually increases to reach extreme at 20:00 21 July, presenting a southwest-northeast (SW–NE) oriented mesoscale, quasi-linear rain belt (Fig. 2a3). Accordingly, the rainstorm moves from southwest to northeast, gradually evolving into a convective line, leading to sustained downpour over Beijing. After 20:00, the system heads eastward and moves out of Beijing (Fig. 2a4)." (Jiang Xiaoman, Yuan Huiling, 2014)

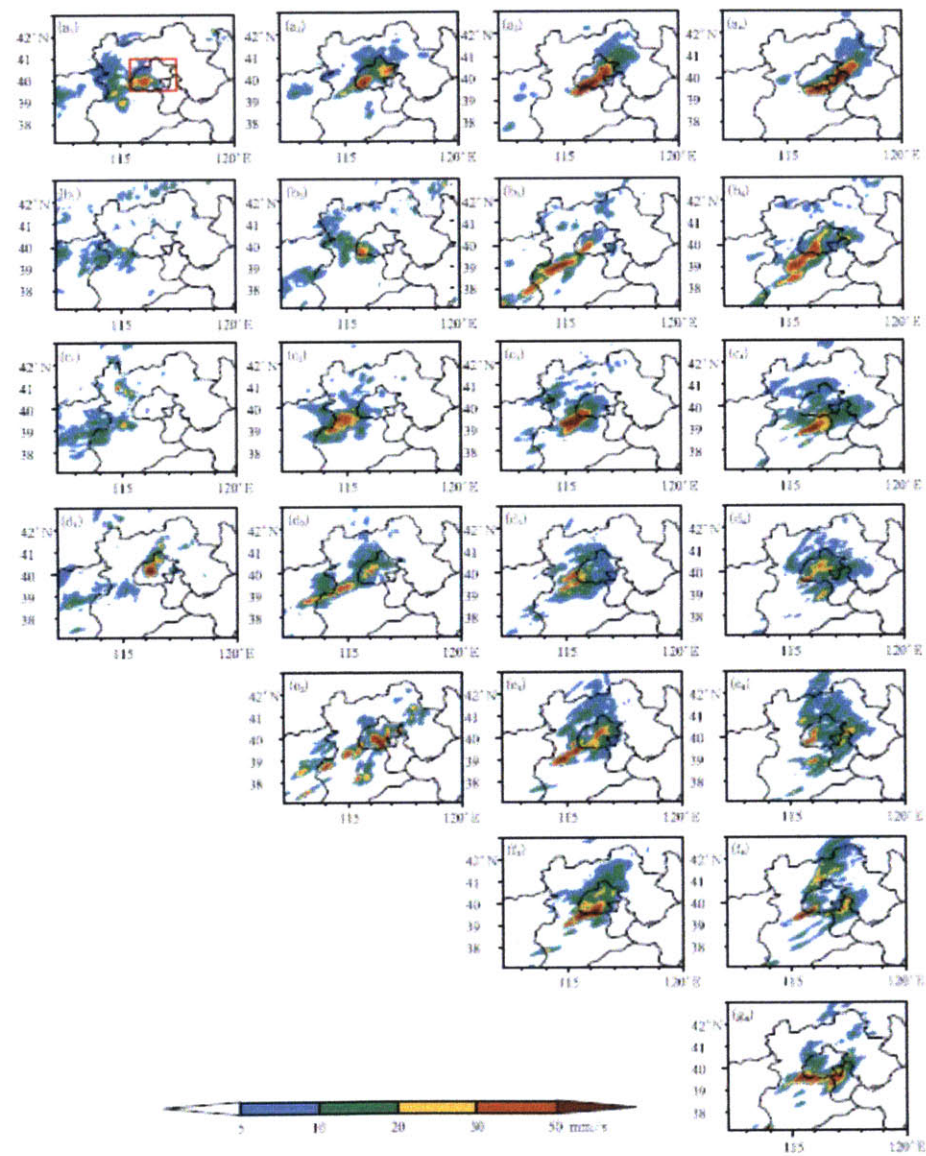


Figure 3.5 Hourly precipitation(mm h-1)at 0600, 0900, 1200, and 1500 UTC 21 July 2012 from(a1-a4)CMORPH merged precipitation and forecast precipitation initiated at(b1-b4) 2100 UTC 20 July, (c1-

[2] "Mesoscale meteorology is the study of weather systems smaller than synoptic scale systems but larger than microscale and storm-scale cumulus systems."(wikipedia)

c4)0000 UTC, (d1-d4)0300UTC, (e2-e4)0600 UTC, (f3, f4)0900 UTC, and (g4)1200 UTC 21 July 2012. Valid at(b1-d1)0600 UTC, (b2-e2)0900UTC, (b3-f3)1200 UTC, and (b4-g4)1500 UTC 21 July", Note: Retrieved from Jiang Xiaoman,2013

Weather Underground provides hourly weather history. The database includes Time CST, Temperature C, Dew Point C, Humidity, Sea Level Pressure hPa, Visibility Km, Wind Direction, Wind Speed Km/h, Gust , Speed Km/h, Precipitation mm, Events, Conditions, Wind Dir Degrees, DateUTC, observations. Observation provides information about the weather patterns which is useful and is lacking in other weather data.

Six intensities were defined based on activities records in every 30 minutes. I scored the rainfall intensity according to the description of observation. On a scale from 1 to 6, '1' is very light and '6' is extremely severe. I assigned the score to the different rain condition. Although weather underground can only provides observation, it can reflect the situation accurately. Fig? showed that observation matches to CMORPH data.

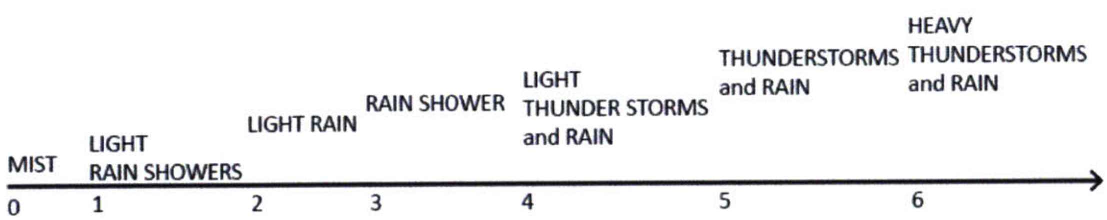


Figure 3.6 Storm description and intensity score

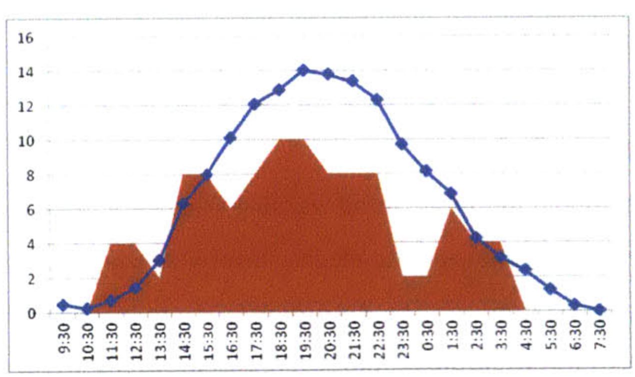


Figure 3.7 Rainfall intensity based on observation and averaged CMORPH merged precipitation (mm)

3.3 Flood Disaster Management

Beijing flood defense system includes Storm Water Management and flood management.

3.3.1 Existing Storm Water Management

Storm water management (SWM) in Beijing focuses on collecting storm water. The Yongding River and North Canal are two regional rivers that are the main storm water storage in Beijing. The other four main rivers and 120 channels (total length is about 580km) are the second storage body. Beijing's drainage system is a mystery. The best estimate is that this 3000-years ancient city has more than 6000 km pipes (Beijing municipal institute of city plan). The flooding frequency is unknown. There is a lot of ecological planning shaping the water reserve problem including the Beijing Ecological Infrastructure, urban flood control planning, greenbelt planning etc. However, little research mentions how these planning projects are implemented and whether they contribute to the storm management. At the beginning of this year, the Ministry of Housing and Urban-Rural Development of the People's Republic of China (MOHURD) published the, "Sponge City Guideline" which aims to change Beijing to a "sponge city", collecting water and using it effectively.

The Beijing Water Authority and related institutes is in charge of the Beijing Flood Management, including flood forecasting and flood operations. The Beijing flood risk model can simulate 5-year flood (Xiaotao Chen, 2014). Weather and water level alerts will be delivered by Beijing Meteorological Bureau, Beijing Water Authority and other related organization. First responder organizations include government agencies, police, firemen, medical and public health organizations and military responders.

3.3.2"7.21" Flood Disaster Defense

On July 21, warning signals and mountain flood warning level were delivered by radio, TV, mobile network, fixed network, internet and electric display device. However, it seems that the early warning system lags and is inaccurate. According to the following table, no red warning was published. The signals the government delivered also showed a less severe alert than the actual condition of the flood.

Table 3.2 "7.21" Flood disaster early warning

Delivery Time	Warning Signal/Level	Department
7/21 9:30am	1 rainstorm blue warning	Beijing Meteorological Bureau(BMB)
7/21 9:30am – 7/22 1:00am	3 rainstorm yellow warnings, 2 rainstorm orange warnings and 1 rainstorm blue and 1 thunder yellow warning	BMB
7/21 10:30am	1 geological disaster yellow warning	BMB + Beijing Municipal Bureau of Land and Resources (BMBLR)
7/21 11:00am-23:00pm	Mountain flood warning third level, Riverine flood risk third level	BMB
7/22 3:50am	Rainstorm blue warning lifted	BMB

Table 3.3 Storm warning signals and guidance Data from China Meteorological News Press

Signal	Meaning	Preventive Measures
	In the next 12 hours, the rainfall will be up to above 50 mm, or is likely to continue with a basis of 50 mm.	1. Government and related departments should take preventive measures for rainstorm. 2. Schools and kindergartens should take proper measures to ensure the safety of students and kids. 3. Drivers should pay attention to waterlogging and traffic jam. 4. Check the drainage system in city, farmland and fish pond and made preparation for drainage.
	In the next 6 hours, the rainfall will be up to above 50 mm, or is likely to continue with a basis of 50 mm.	1. Government and related departments should prevent against rainstorm. 2. Transport management departments should take traffic control in case of heavy rainfall and guide traffic in waterlogging section. 3. Cut off the dangerous outdoor electronic power in low-lying areas , pause the outdoor operation in open areas, and transfer people in dangerous areas and buildings. 4. Check the drainage system in city, farmland and fish pond system and take necessary measure for drainage.
	In the next 3 hours, the rainfall will be up to above 50 mm, or is likely to continue with a basis of 50 mm.	1. Government and related departments should take emergency response for rainstorm. 2. Cur off the dangerous outdoor electronic power and pause outdoor business. 3. Units in dangerous areas should stop activities and the students, children and other personnel in schools should be protected. 4. Carry out drainage in cities and farmland and prevent against flash flood, landslide, mud-rock flow and other disasters induced by heavy rainfall.
	In the next 3 hours, the rainfall will be up to above 100 mm, or is likely to continue with a basis of 100 mm.	1. Government and related departments should do a good job of emergency response and disaster rescue for rainstorm. 2. Stop assembly, classes, and business expect special industry. 3. Do a good job of prevention against and rescue of flash floods, landslide, mud-rock flow and other disasters.

http://www.cma.gov.cn/en/WeatherWarnings/WarningSignals/201203/t20120320_166872.html

Government published a report about "7.21" flood disaster emergency rescue. According that, 13,096 soldiers from Beijing Fire Forces, Beijing Emergency Medical Center, Beijing Traffic Administration Bureau, Beijing Drainage Group and Beijing Garrison took part in rescue missions on that night. The Beijing Municipal Public Security Bureau received 134,000 calls. The mission of flood defense is a many-sided, complex and difficult undertaking.

3.4 Social media in "7.21" Flood Disaster

Sina-weibo is the most popular open platform of social media for the public to sharing disaster situation and disseminating rescue information. In the "7.21" flood disaster, millions of weibos were created and forwarded.

3.4.1 Hot topic with harsh tags

#北京暴雨(Beijing Storm)#become the top 2hot topic in July.(source??) Other hot topics including #"7.21"灾难"(7.21"disaster)#, #广渠门(Guangqumen)#, #暴雨 救援(Storm Rescue)#, #北京紧急救助(Beijing Emergency)#, #传递正能量(Convey positive energy)#, #祝福北京 (Blessing Beijing)# , #北京紧急(Beijing Emergency)# #北京突发(Beijing incidents)#are also become the popular harsh tags on July 21th and July 22th.

3.4.2 Information Distribution

One case: Death under Guangqumen overpass

On the individual level, weibos definitely distributed a lot of valuable information about situation awareness and rescue information. We can take reporting on a death under the Guangqumen overpass as an example. I extracted all weibos by the keyword of "Guangqumen" between 8:00 to 21:00 on July 21 before the death. I found 306 pieces of weibos and 60 of them have geo-tags. About 20% weibos in query "Guangqumen" attached geo-tags.

Query: Guangqumen, from 8am to 21pm, 07/21

13:00 the first weibo with geo-tag indicates Guangqumen started raining

13:24 #Beijing heavy storm# + photos showing the disaster situation of Guangqumen

13:26 #Guangqumen overpasses# we are ready to "watch sea" here

13:27 one weibo user complained she was trapped in Guangqumen because of the heavy rain

19:06 the storm caused flooding in Guangqumen

I also checked the top forwarded weibos and found that weibos created by verified person and verified organization were given a priority to repost. Verified organization and celebrity play a very important role (celebrity effect) in weibo distribution pattern.

Distribution pattern of one weibo from VIP[3] (Rank 1)

I selected the most retweeted weibo form VIP include keyword of "Guangqumen".

Table 3.4 The most retweeted weibo from VIP

User ID	Content	like	forward	comment	time
北京人不知道北京的事儿	广渠门桥下疑有人员被困！！！消防正在搜救！晚 21 时，一名淋得浑身透湿的女子……	20	10070	1477	20:26
Beijing local issues	Someone was trapped in Guangqumen!!! The firemen are attempting to rescue him! At 21pm, a woman was begging for help.....	20	10070	1477	20:26

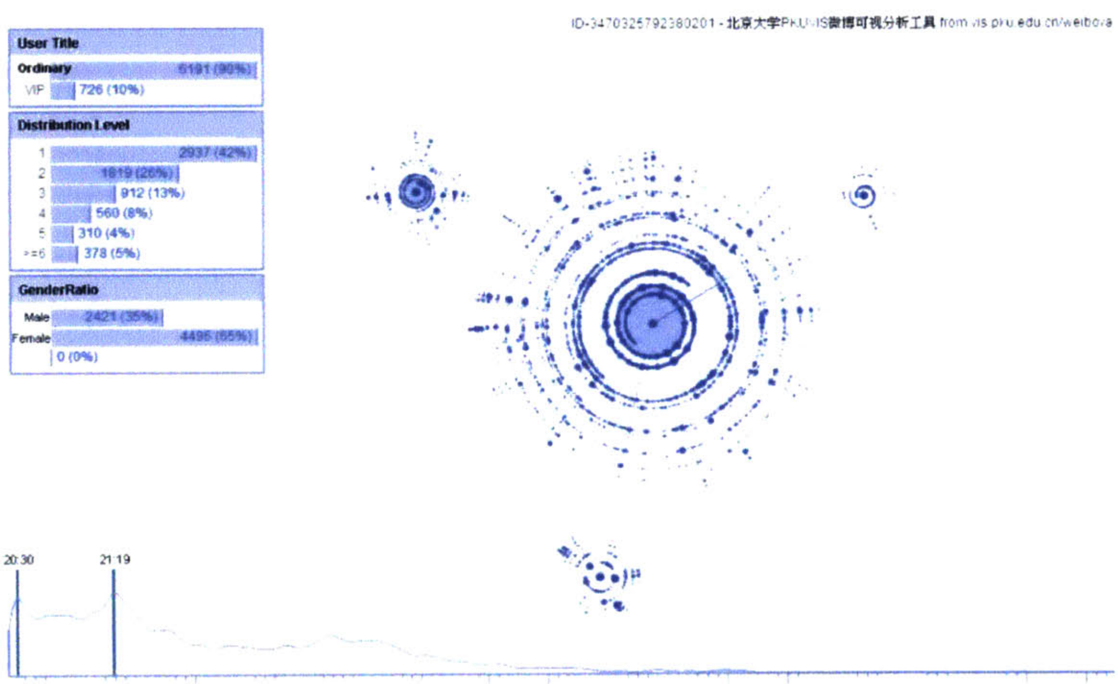


Figure 3.8 Distribution pattern of VIP via vis.pku.edu

[3] VIP: verified person and verified organization.

Distribution pattern of weibo from Non-VIP[4] (Rank 24)

Table 3.5 The most retweeted weibo from Ordinary User

User ID	Content	like	forward	comment	time
桃福子甜品店	两广广渠门，井盖全部被雨水冲起，形成了喷泉，地面也翘起@交通路况 @北京交通。	2	489	54	19:59
Taofuzi Dessert	All manhole covers were carried away by the flood, and created jet, the ground was lifted as well@Traffic condition @Beijing Traffic	2	489	54	19:59

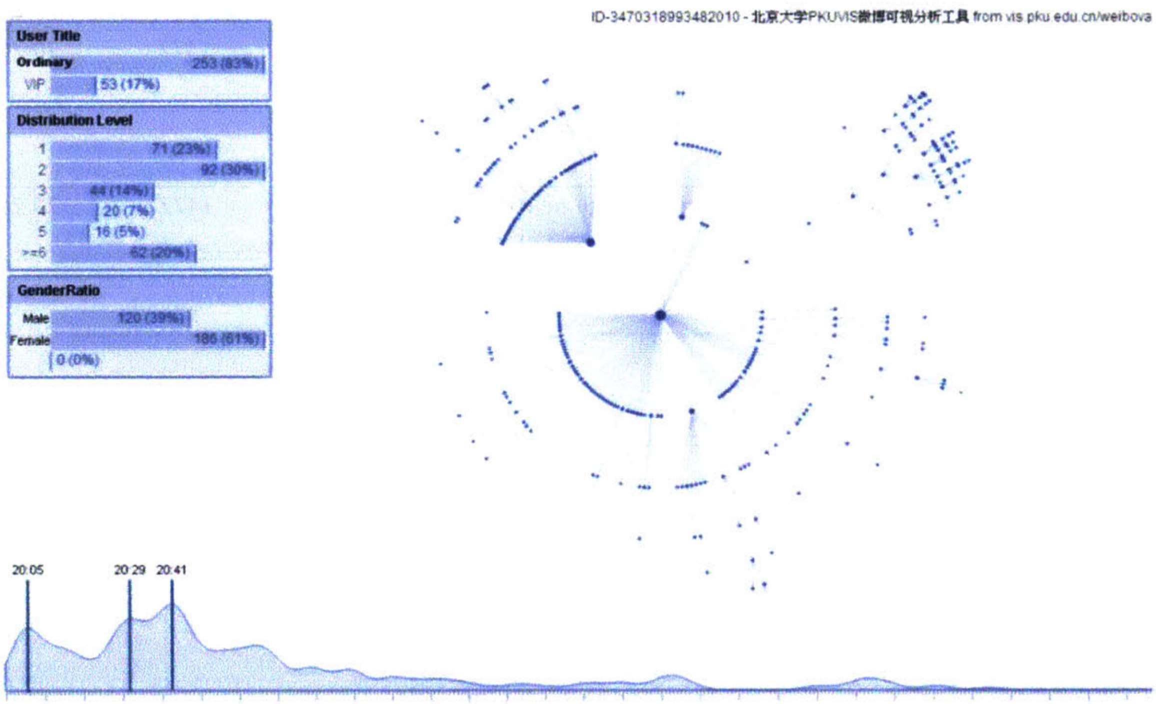


Figure 3.9 Distribution pattern of Non-VIP via vis.pku.edu

3.4.3 Disaster Maps

Netizens homemade disaster maps

Except sharing disaster related information, weibo netizens also made maps reflecting water-logging areas and traffic congestions caused by the flash flood. The following homemade map

---

4 Non-VIP: Ordinary user and master user

was the most popular one forwarded in Sina-weibo system. They reflect the disaster situation at the first time.

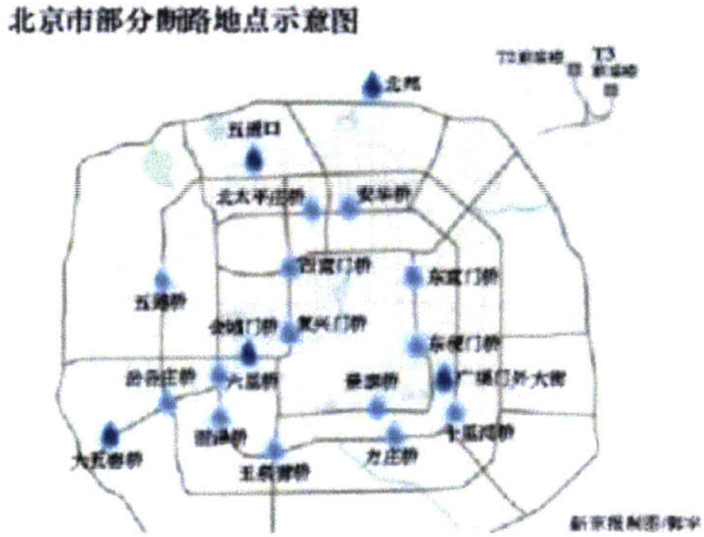


Figure 3.10 Homemade Water-logging Map. Retrieved from Guoyu's Sina-weibo

Official/authorized organization disaster maps

Dedicate websites were developed after the disaster occurs. Maps on these websites combined information from traditional news reports and weibos.

- Official water-logging map was developed by Beijing Water Authority (BWA) on July 24. 42 water-logging places were pointed out. However, it has been removed from the BWA website.
- The Beijing News collected related information and drawn a water-logging map with more points. Published on July 25 04:28
- Sogou combined all information from traditional media and social media and created Beijing water-logging website on July 25. This on-line map includes 114 water-logging points and invite edits from the public. The red points show the water-logging places and candle means the severe water-logging areas.

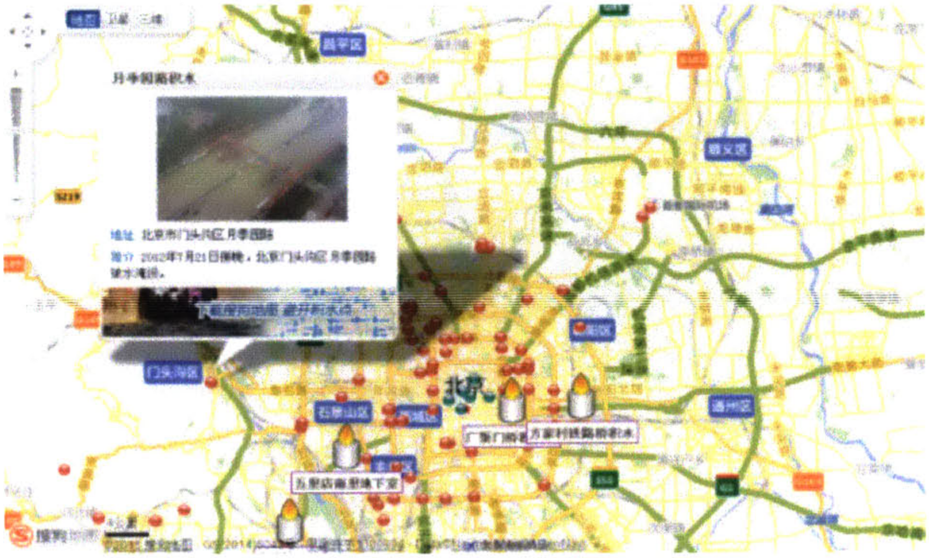


Figure 3.11 Sogou Water-logging Map

Retrieved from http://map.sogou.com/special/jishui/?IPLOC=CN1100

Tencent[5] created a "7.21" heavy storm website including latest news, pictures, videos, water-logging maps and self-rescue techniques and car insurance stuff. The water-logging map in this website was considered to have more authority because it integrated the official sources.

Figure 3.12 Tencent Water-logging Map

Retrieved from http://news.qq.com/zt2012/bjzyby/index.htm

---

[5] Tencent is one of China's largest and most used Internet service portal. It product, Q-zone, was he most favorite social media in China in 2012 (McKinsey, 2012).

Chapter 04 Data analysis

4.1 Data Collection

4.1.1 Data Sample and Method of Extraction

Weibo open platform (http://open.weibo.com/wiki/) has been the official platform of Sina-weibo for subscribing and downloading weibos. It provided Open APIs for developers to access the information related to weibos. However, this platform is not friendly to users due to the strict authentication process, encrypted filter principles and limitation on number of weibos that one could extract. Web application require fully use OAuth for user authentication. One had to pass the authentication process to access the advanced APIs, such as searching by topics. The advanced APIs even came with its own limitation. For instance, it could only return the latest 200 weibos regarding the searched topic. This advanced API was not available in the English version of the web interface.

Figure 4.1 Sample screen shot of two weibos on Sina-weibo web interface

Studying Sina-weibo communication was very challenging because Sina-weibo open platform did not provide access to historical database. The only available option was the search function provided by the web interface of the website. In its advanced search functions, users can search by key words, time range, and districts, and received 50 pages of weibos. The minimum time interval was one hour, thus the data collected in this study has the minimum resolution in time of one hour. In order to increase the sample size, this work combined weibos searched using one hour internal and two hours interval.

Figure 4.2 search function interface of Sina-weibo

To make up for the lack of API for historical data, this work used a web scraping approach to collect weibos from Sina-weibo web interface. By submitting the query "暴雨" ("storm") to Sina-weibo interface with the time set to one hour at a time, and district limited to Beijing, 38,528 weibos were retrieved related to "7.21" flood disaster for the time corresponding to during the storm and 20 hours after it. To be more specific, the data was collected from 07/21/2012 11:00:00 to 07/21/2012 23:59:59. Note that this was a subset of all the "7.21" flood disaster related weibos because the query was on the keyword "storm"(暴雨), while there could be relevant weibos that did not include the term explicitly.

"http://s.weibo.com/weibo/%25E6%259A%25B4%25E9%259B%25A8&region=custom:11:1000&typeall=1&suball=1&timescope=custom:2012-07-21-11:2012-07-21-11&nodup=1&page=2"

The above link was a sample that returned "storm" related weibos for 11am, July 21, 2012. The rules which composed this link could be broken down into the following:

-Fixed part: http://s.weibo.com/wb/
-Keyword (encode-in-two-phases): %25E6%259A%25B4%25E9%259B%25A8
-Region: custom: 11:1000
-Time scope: custom: 2012-07-21-11:2012-07-21-11
-Omit similar weibos: nodup=1
-Page: page=2

The process of web-scraping was: simulating a user login→ building URL of next hour→ building URL of next page → sending a request → Inputting identification code if you were suspected as a robot→ returning the results → writing to a file. The following flowchart showed the process f web scraping Sina-weibo.

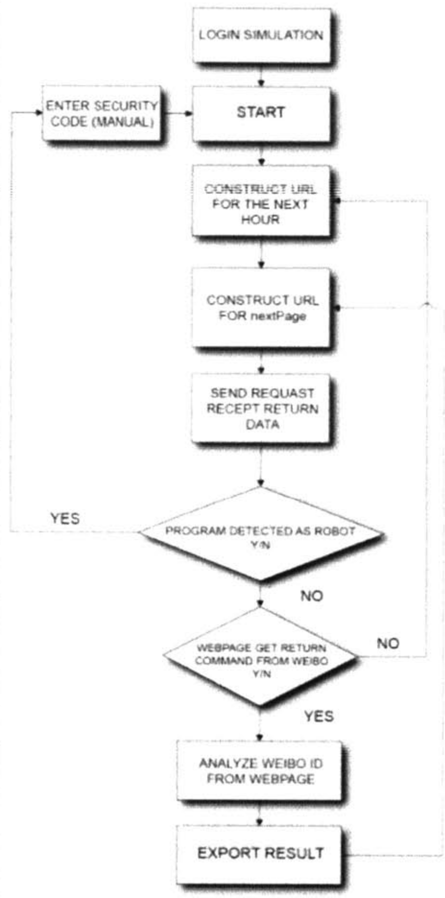


Figure 4.3 flowchart of web scraping Sina-weibo

4.1.2 Data Description

The weibos generated by web scraping were then exported into CSV files. For each of the retrieved weibo, basic information was extracted, including user's user name(ID), user's title, weibo type, weibo URL, weibo content (text, emoticons, short links etc.), number of likes, number of retweets, number of comments, published date and time.

For readers that might not be familiar with the Sina-weibo system, a user title referred to one of the following four titles: ordinary user, expert user, verified person, and verified organization. Weibo types included original posts and reposts. (Sina Weibo has an identification policy which can verify the identity of famous person, organization and so on. Ordinary user can bind the accounts with their phone numbers or post the numbers of weibos reach Sina's threshold in order to be expert user). In this study, only original posts of weibos were collected to avoid massive amount of duplicate information. However, if the user created part original weibo and repost other weibo as well, it also counted as the original weibo which would be collected.

In terms of weibo content, emoticons were converted into text, as would be discussed later in more details in Section 4.4.1. Weibo content could also include short links that would point to images, web pages, or a specific location on a map. As part of the preprocessing process that would be discussed in Section 5.1, these short links were expanded into normal links in this thesis.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	发布人	发布人类型	微博类	微博地址	转发原微博	微博内容		博文图片地	点赞数	转发数	评论数	发布日期	发布时间
2	沙漠里的山	微博达人	原创	http://weibo.com/2		北京今天大暴雨，只能呆在家里 //@缘分时			0	11	5	2012-7-21	11:29
3	秤子小米	微博达人	原创	http://weibo.com/1		我在:#故宫护城河# 故宫的人真多，排队买			0	0	4	2012-7-21	11:29
4	沙漠扬沙	普通用户	原创	http://weibo.com/2		闷热死了，天气预报说有大雨至暴雨，让雨			0	0	3	2012-7-21	11:29
5	宸諾NewLi	普通用户	原创	http://weibo.com/1		没有狂风预警，暴雨来的毫无征兆，雷电只			0	0	2	2012-7-21	11:29
6	想飞又懒惰	普通用户	原创	http://weibo.com/1		更喜微风！另外这暴雨啥时落到东边啊！			0	2	2	2012-7-21	11:29
7	二旦旦	普通用户	原创	http://weibo.com/1		大暴雨还没开始下啊。			0	0	2	2012-7-21	11:29
8	Ludwig_Bl	普通用户	原创	http://weibo.com/1		尼玛暴雨什么时候下？？？			0	2	2	2012-7-21	11:29
9	望穿尜	微博达人	原创	http://weibo.com/1		传说今日暴雨@丹丹的窝--丹小柔柔 @审判			0	0	1	2012-7-21	11:29
10	年轻的磊爷	微博达人	原创	http://weibo.com/1		据说午后有暴雨，俺们下午3点还得踢球呢,			0	0	1	2012-7-21	11:29
11	奥斯卡111	微博达人	原创	http://weibo.com/2		回复@丿亅丶于丿乚:北京下午是暴雨吗 //			0	0	0	2012-7-21	11:29
12	王春雷chx	微博达人	原创	http://weibo.com/1		大暴雨下吧！			0	0	0	2012-7-21	11:29
13	NN小样儿	微博达人	原创	http://weibo.com/1		这所谓的雷阵雨转大雨到暴雨在哪儿呐？@			0	0	0	2012-7-21	11:29
14	SabrinaGs	普通用户	原创	http://weibo.com/1		据说从中午开始下暴雨，所以逃了课回家。			0	0	9	2012-7-21	11:30
15	小小锣鼓咿	普通用户	原创	http://weibo.com/1		离凤凰山还有5公里，暴雨来袭，幸遇一采			0	0	8	2012-7-21	11:30
16	Double花	微博达人	原创	http://weibo.com/1		晚上会有暴雨吗哎			0	0	7	2012-7-21	11:30
17	奶罩大贝	微博个人认	原创	http://weibo.com/1		回家睡觉…还好一上午没白忙活…到家继续			1	0	7	2012-7-21	11:30
18	suzumax	普通用户	原创	http://weibo.com/1		我在太兴烧味餐厅(来福士店)：今天有暴雨			0	0	5	2012-7-21	11:30
19	mukaddass	微博机构认	原创	http://weibo.com/1		赞同 //@KASHGAR灿烂:喀什噶尔和北京同在			0	0	4	2012-7-21	11:30
20	幽_冥robe	微博机构认	原创	http://weibo.com/2		房山大到暴雨状态 //@霹雳球球:大雨来袭			0	0	2	2012-7-21	11:30
21	lwVampire	微博达人	原创	http://weibo.com/2		今天有特大暴雨啊~			0	0	2	2012-7-21	11:30
22	小饭的天空	微博达人	原创	http://weibo.com/2		今天有暴雨，赶紧回家吧！//@佟璐-小丸子			0	0	2	2012-7-21	11:30
23	小娜娜要养	微博达人	原创	http://weibo.com/1		说好的大暴雨呢？！这种温度32湿度100%的			0	0	1	2012-7-21	11:30
24	竹竹熊猫	微博达人	原创	http://weibo.com/1		下午还得出去咧~表暴雨哦			0	0	1	2012-7-21	11:30
25	吴小喵同学	微博达人	原创	http://weibo.com/1		等待今天的大暴雨~不知道会被拍在哪儿~哈			0	3	1	2012-7-21	11:30
26	摇摆的野马	普通用户	原创	http://weibo.com/1		说好的暴雨呢....			0	0	1	2012-7-21	11:30
27	李寂卓	微博个人认	原创	http://weibo.com/2		遭遇大暴雨，带着一家老小狂奔，白小C经			0	0	0	2012-7-21	11:30
28	Carrie小	微博达人	原创	http://weibo.com/1		说好的暴雨呢？？			0	0	0	2012-7-21	11:30
29	比尔盖饭_	微博达人	原创	http://weibo.com/1		千万别下了 晚上肯定变落汤鸡呀 工体顶棚			0	0	0	2012-7-21	11:30

Figure 4.4 Data Sample

The counts of weibos collected from each hour varied from 604 to 1413 (Gray bar in the Figure4.5). In order to increase the sample size, weibo searches of one-hour interval and two-hour interval were performed and data was combined. After performing the previous steps, 38528 pieces of weibos (original) form 252388 (original and retweets) dataset were collected for this thesis.

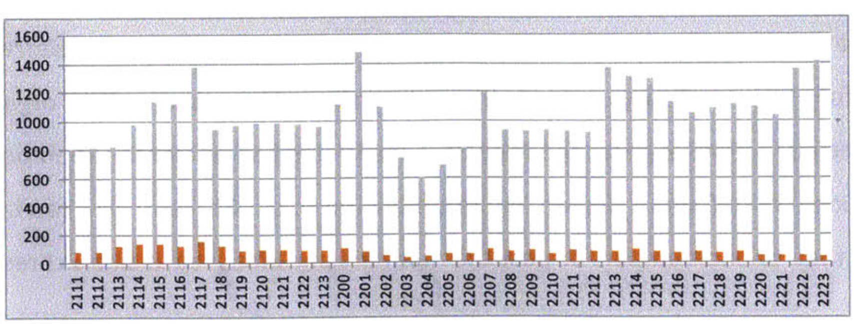


Figure 4.5 Weibo counts of each hour in the sample

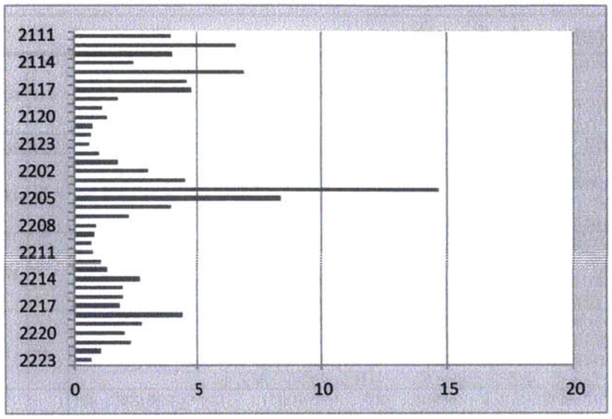


Figure 4.6 Percentage of weibos in the sample to all "storm" related weibos

In the sample, weibos created by ordinary users accounted for 50%, expert users' weibos accounted for 27%, verified personals' weibos accounted for 16%, and weibos from verified organization accounted for 7%.

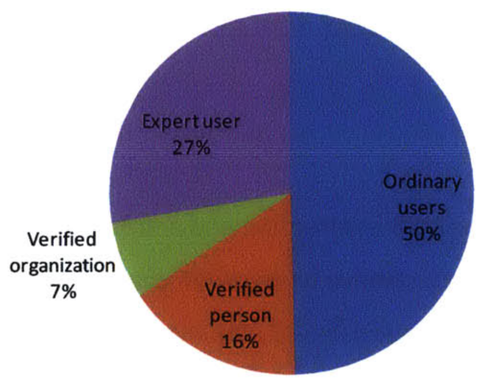


Figure 4.7 Percentage of weibos created by different users

For demonstration purpose, the top ten most retweeted weibos were listed in Table X. Among them, two came from ordinary users, three were written by verified personals, and the last half came from verified organizations. In particular, the most retweeted weibo was forwarded 62,500 times, and it consisted of very short narratives of eight different ordinary citizens' experience during the disaster, sharing the positive experience in a devastated disaster, and expressing the notion that ordinary citizens made history on a daily basis, which was a recurring theme on social media during this disaster. This weibo was posted by an ordinary user at 1:58am on July 22nd, 2012.

Table4.1 Top 10 most retweeted weibos

庞胡瑞	普通用户	#北京暴雨##加油北京#一个堵在车上临产的母亲；一个趴在水里疏通下水道的大爷；一个死在二环路上的兄弟；一个牺牲在救援前线的派出所长；一群挡在没了井盖的排水沟前的环卫工人，一群被大雨淋透的交警，一群搭载路人的好心朋友，一群招呼无法归家者留宿的爷们！加油北京！这一群小人物创造了历史！	247	62500	12469	2012-7-22	1:58
易鹏	微博个人认证	北京卫视新闻：北京市民高度肯定这次暴雨中北京政府的应对工作。	398	48258	316	2012-7-22	16:31
何恩培	微博个人认证	六十年一次的暴雨，让我误了明天的两个要约，却有幸体验‘’双闪志愿者‘’。凌晨3:30我出候机楼的门，迎面站着一个气质美女，背后一辆双闪奔驰：“去城里吗，免费送”，再看背后，一排双闪车，有别克有宝马有大众。。。，这样凝聚力的民族，必将崛起。 http://t.cn/zWXN1PN	88	14351	2432	2012-7-22	3:46
天气预报	微博机构认证	【北京全市大暴雨】北京市气象台发布10-20时雨量：全市平均117mm，平均大暴雨，城区更甚，平均166mm，石景山模式口277mm，八大处241mm，最大为门头沟的龙泉345mm。2011年北京623暴雨全市平均降雨量48mm，城区平均降雨量为72mm。今天这次降雨，超远去年623，至少为21世纪以来北京最强降雨，可载入史册。	23	13455	2816	2012-7-21	20:52
人民日报	微博机构认证	【北京暴雨灾情最新发布】记者魏薇6时从北京市防汛抗旱指挥部获悉，截至目前，北京受灾面积达1.6万平方公里，成灾面积1.4万平方公里，190万人受灾，其中房山区80万。北京市防汛抗旱指挥部副指挥、市水务局副局长潘安君透露，相关部门正在逐一核查伤亡人员，数据将会很沉痛，一旦核实清楚将立即公布。	86	11497	1358	2012-7-22	19:06
薛陈子	微博个人认证	【北京暴雨时 微博传递爱】一夜未眠，制作了一个小短片。汇总了天灾来临时，微博上的关爱、温暖、善良...转发就是力量，微博传递大爱。平安就好！！ http://t.cn/zWXpVrv	28	9015	1218	2012-7-22	4:33
天气预报	微博机构认证	【北京暴雨致多路段受阻 立汤路无法通行】京开高速出京方向新发地桥有事故，造成车辆行驶缓慢，目前队尾已经排过马家楼桥；万泉河快速路一亩园桥下路口有积水，过往车辆请注意小心驾驶；立汤路太平家园路口，受积水影响，各方向车辆无法通行，请过往车辆尽量避开天通苑地区。 @交通北京 @施 展	22	7321	1617	2012-7-21	18:45
电商妹子	普通用户	北京暴雨，鸭子上街，车全泡水，老外仰泳。以下照片全部来自手机拍摄哦。	33	7207	929	2012-7-21	17:45
微博小秘书	微博机构认证	#北京平安#突如其来的暴雨让我们措手不及，但正是每一个普通人默默的贡献携手让此刻的北京更安宁、更温暖。感谢每一个付出举手之劳的人，感谢每一个参与救助的人，让我们传递这份爱，让我们用行动祈祷#北京平安#，今晚，我们共同守护！ http://t.cn/zWXw4CB	28	7049	1968	2012-7-22	1:37
新浪评论	微博机构认证	【新华微评：暴雨中闪光的“北京精神” 】昨日暴雨袭击北京，突如其来暴雨不仅是对城市应急排险能力的考验，更是对人们精神上的一次洗礼。令人欣慰的是，政府职能部门积极行动，官民互助、人们守望相助。“北京精神”已不仅是根植于内心的道德操守，更是化为实践行动的闪光力量。http://t.cn/zWav1Sc	42	5563	2585	2012-7-22	23:16

4.2 Trend Analysis

The trend analysis reflected the shift in attention and activity during the different phases of the emergency responses. In order to remove background counts and isolate the pattern triggered by the heavy storm, the hour-by-hour data was smoothed using an ordinary weekend's weibos counts.

4.2.1 Overall Trend

To start, the number of weibos posted from 07/20-08/05 was plotted in Figure 4.8

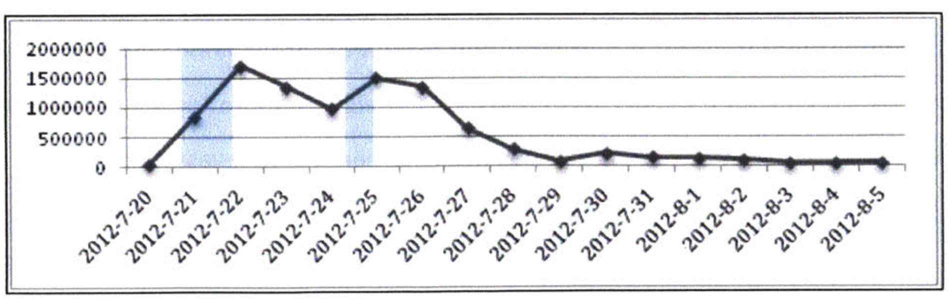


Figure 4.8 weibo counts from July 20th to August 5th, 2012

As illustrated in Figure 4.8, the number of weibos containing the key word "storm" peaked immediately after the heavy storm occurred and gradually decreased after July 22nd. The weibo volume surged again on July 25th due to another storm. The number of weibos continued to drop after July 25th and approached zero on July 29th.In the nine days after the storm, netizens have been continuously paying attention to the topic of heavy storm.

For comparison purpose, the number of weibos for August 10th-17th 2013 during the "8.1" storm in 2013, and weibos from August 27th – September 5th 2014 during the "9.1" storm in 2014 were extracted using the same methodology and illustrated below. A similar pattern could be observed from these three trends in that the weibo counts increased dramatically after the storm hit, and fairly rapidly drop back to pre-disaster count after the event.

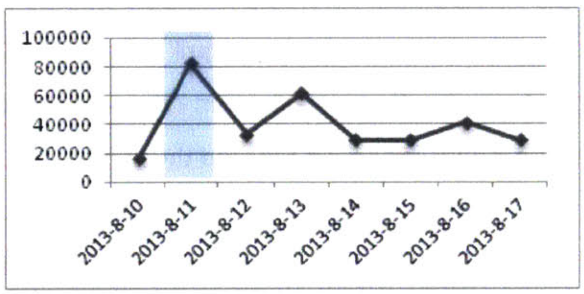


Figure 4.9 weibo counts from Aug 5th to 17th, 2013

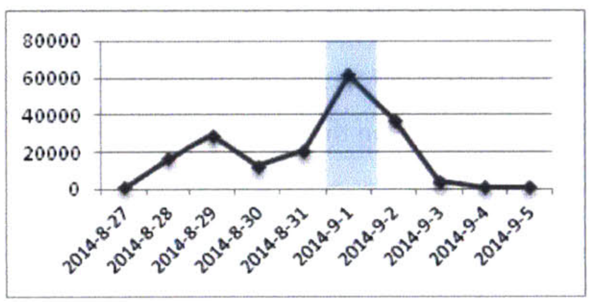


Figure 4.10 weibo counts from Aug 27th to Sep 5th, 2014

Secondly, I counted the number of weibos posted each hour from 07/21 11am to 07/23 11pm

Note: storm begins at 07/21 11:00am; ends at 07/22 3:30am-4:00am

Two peaks: one is 11pm on July 21 and the other one is 11pm on July 22.

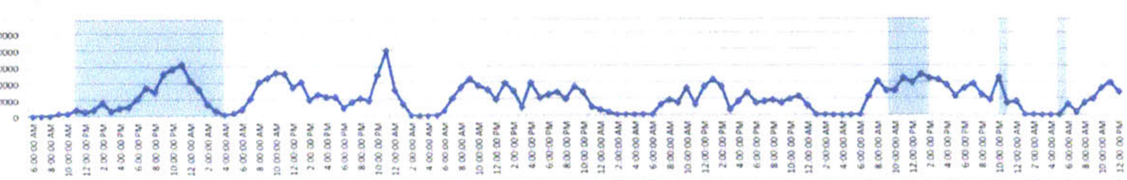


Figure 4.11 weibo counts by Hour, from 6am July 21st to 12pm July 25th, 2012

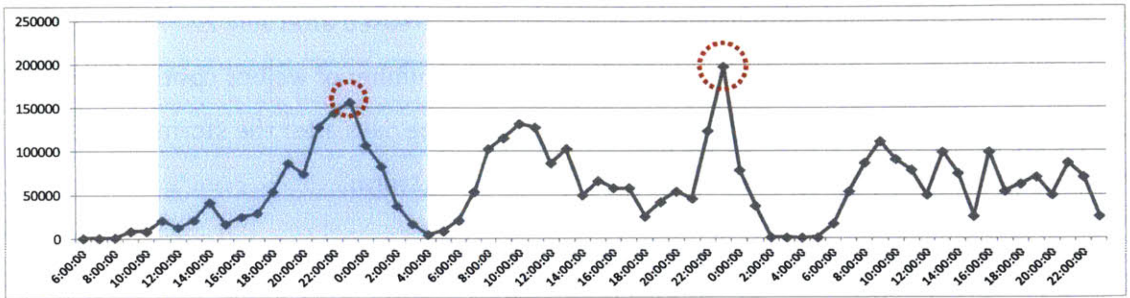


Figure 4.12 Two peaks of weibo counts following the 7.21 storm

4.2.2 Smoothed Trend

Weiguo Wan presented the posting pattern of weibo users in 24 hours in his PHD thesis (Wan, 2014). He collected real-time data of Sina-weibo from April 21st to April 27th in 2012 and drew the curves based on the numbers of weibos in each hour. The curves indicated that the number of weibos started to increase around 7-8am and reached the first peak at 11am. The curves gently changed between 12pm and 5pm and touched the bottom at 7pm. From 7pm it increased again and reached the second peak at 11pm. The whole week data showed a periodicity that the weibos posted reached the highest point at 11am and 11pm. Another research conducted in 2014 also proved that most weibo users like to post weibo at 11pm every day (Wenning Zhao, 2014). The biggest difference between these two researches was that the former one indicated that weibos posted on the weekend was less than weekdays, whereas the latter one did not reach this conclusion.

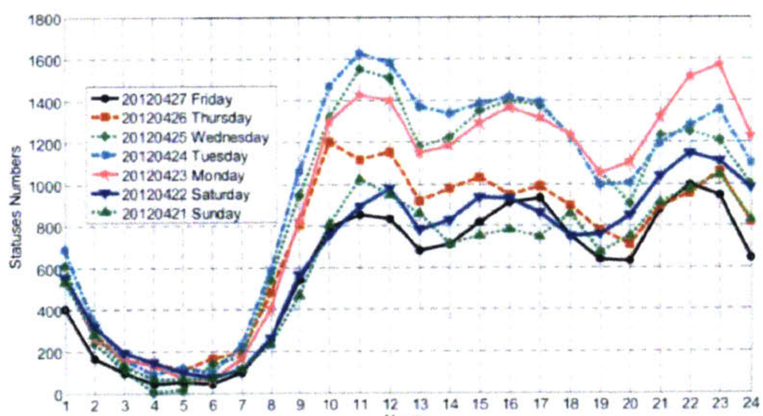


Figure 4.13 Posting pattern of weibo users in 24 hours

Note: Retrieved from Weiguo Wan, "Research on User Behavior Analysis and Network Evolution in Micro-blogging Networks", 2014

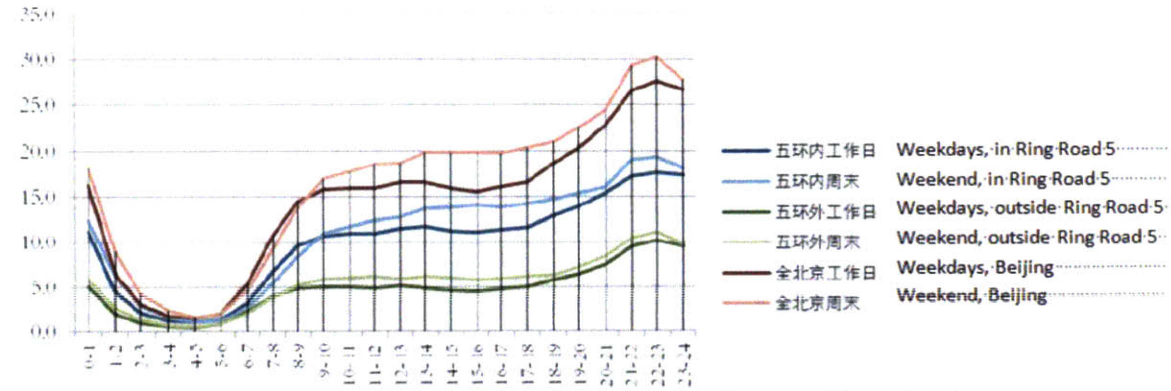


Figure 4.14 Posting pattern of weibo users in 24 hours in March,2014

Note: Retrieved from "Urban Spatial Analysis based on Sina-weibo", Wenning Zhao, 2014

In this work, the curve generated using the keyword "storm" also showed a similar pattern the numbers of weibos reached the highest point at 11pm. In order to identify the underlying pattern caused by the "storm," the count numbers were smoothed in order to remove background noise. Two approaches were applied for smoothing. The first method was to divide the data from the previous research. The resulting curve confirmed the periodicity for April. The second approach was using several ordinary words to estimate the periodicity. In particular, four words were selected for this analysis: dog, TV program, doctor and corruption. They came from four different domains and were popular words so that they would show the pattern. Since the disaster happened on a weekend, weibos related to these four words were collected three weekends before and after the event. The following figures illustrated the results for these four words.

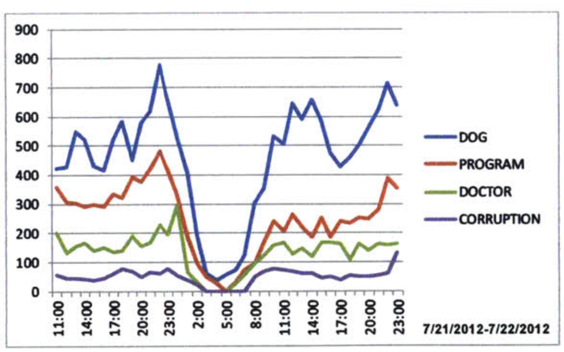


Figure 4.15 Weibo count s of four ordinary words by hour, at the weekend "7.21" storm occurred

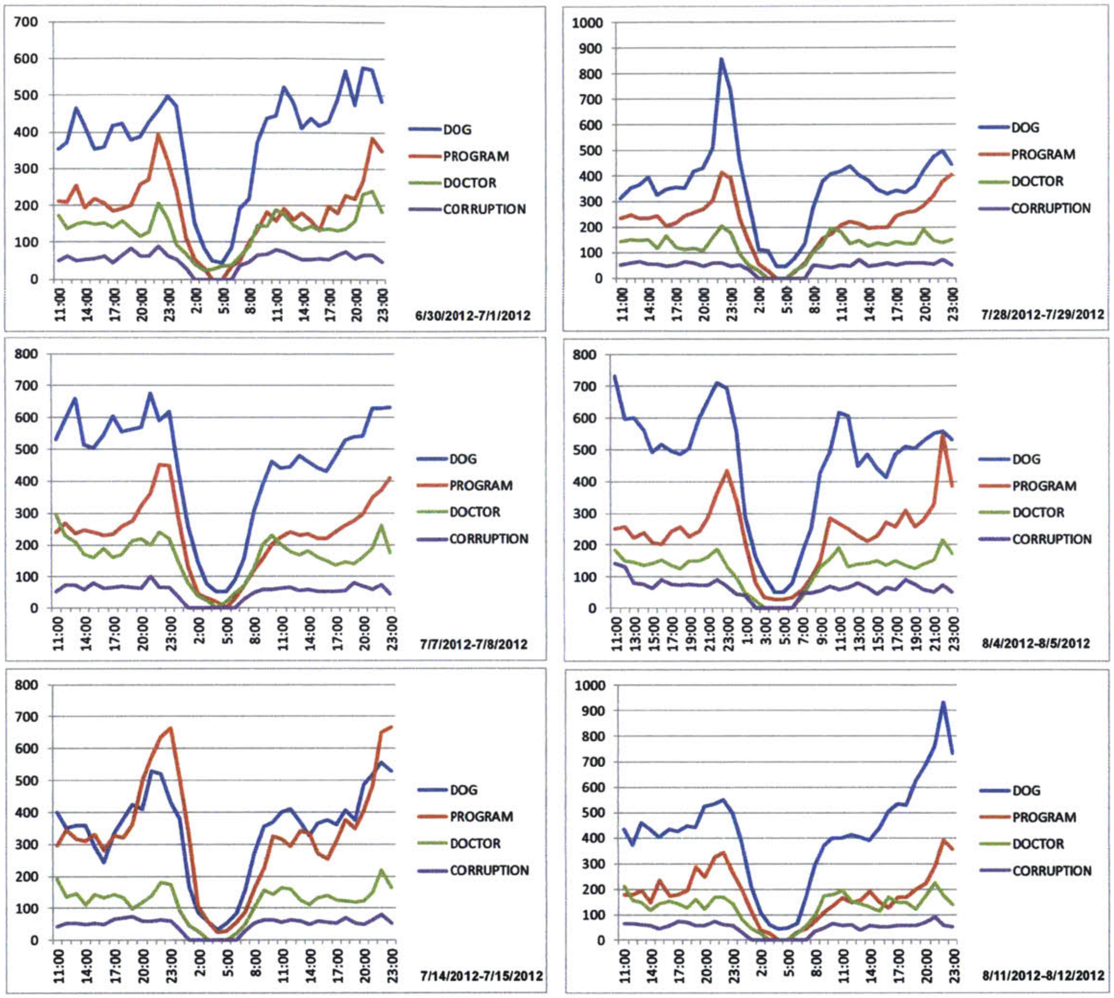


Figure 4.16. weibo count s of four ordinary words by hour, three weekends before and after "7.21" storm

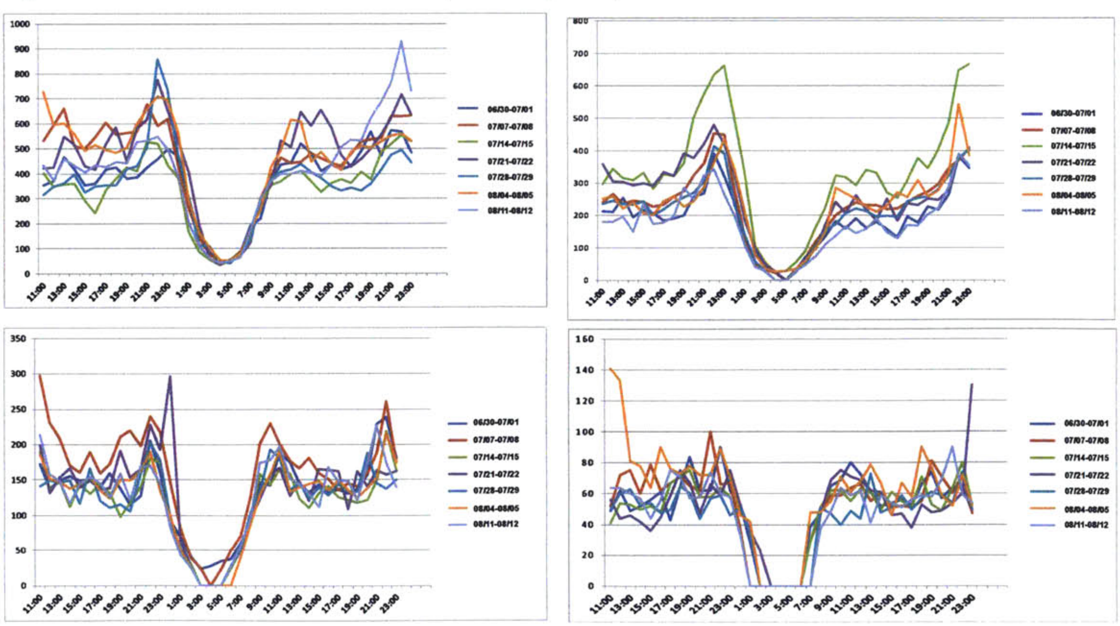


Figure 4.17 Weibo counts of each ordinary word by hour in the seven weekends

Although the scales differed from plot to plot, all of these words showed the periodicity clearly. Thus in this work, the storm weibo curve was divided by the average result of these four words. Figure 4.18 showed two curves representing the weibo posting pattern according to two smooth approaches.

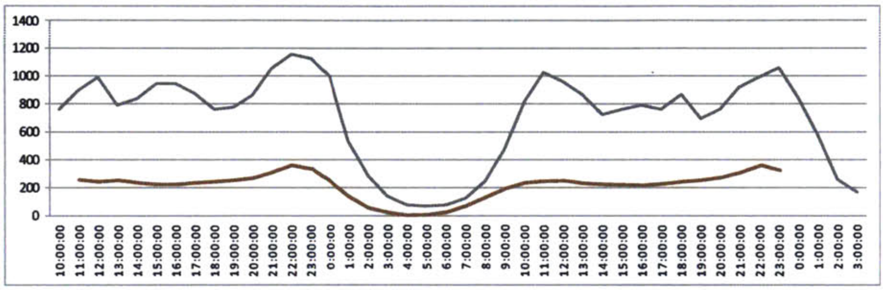


Figure 4.18 Two estimation of posting pattern of weibo users based on previous work and average counts of four words

The following figure showed the result of the normalized weibo count using the keyword "storm." The top one was the original curve, the middle was the curve by smoothed using the previous research, and the bottom was the result smoothed by the average counts of the four ordinary words.

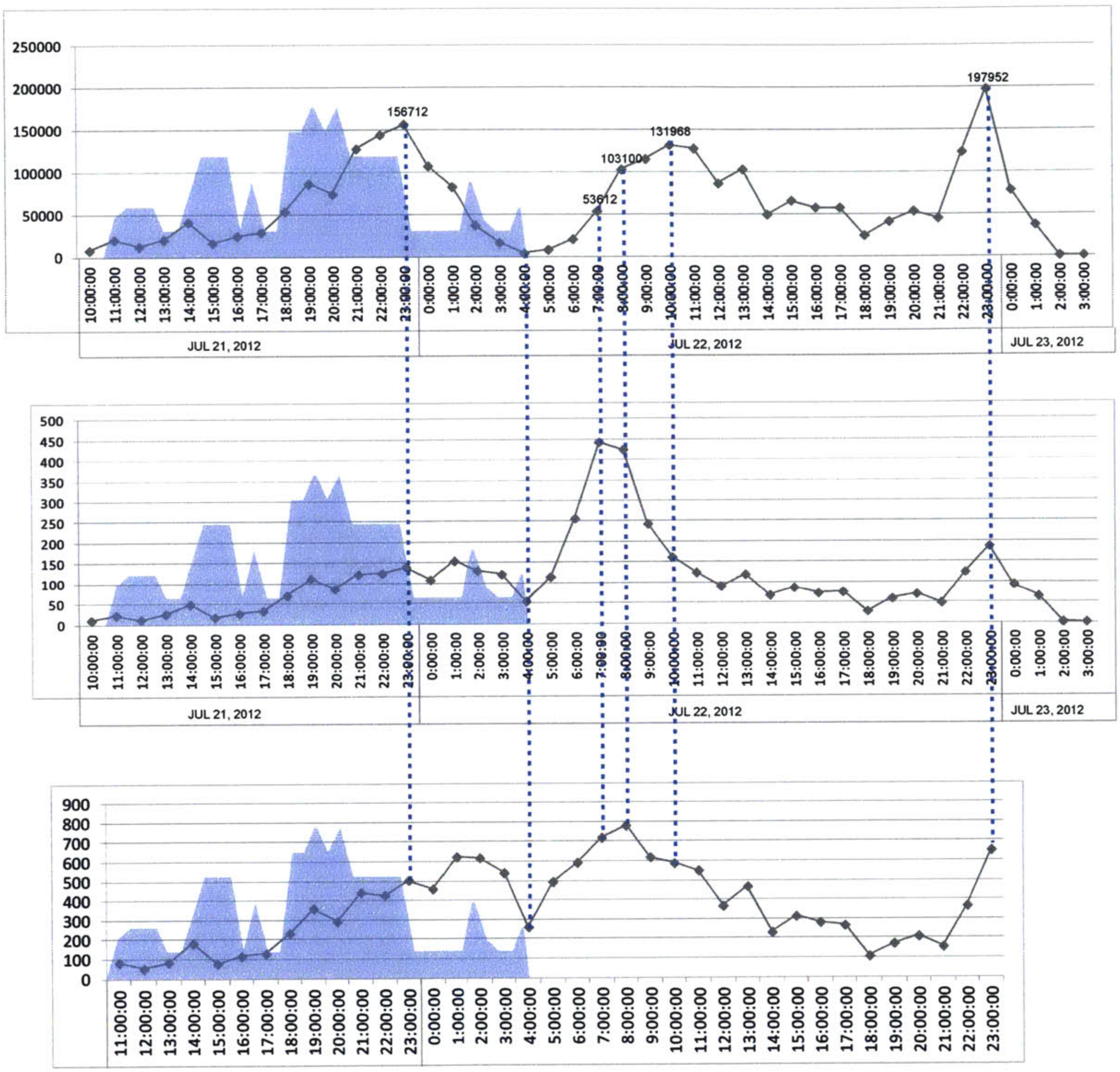


Figure 4.19 Original trend (top) and normalized trends (middle and bottom)

The normalized curve showed dramatically different patterns compared to the original one. Opposite to the original one, the smoothed curve dramatically increased after 4am when the rain stopped and reached the highest point at 7am, July 22nd, 2012. The result implicated that Beijing experienced an unusual night in the heavy storm. The curve reached the second smaller peak at 11pm on July 22nd which was the peak of the original weibo count curve.

The two smoothed curves provided the following insights:

1) The weibo volume closely followed the volume of rain fall with some time delay. The number of weibos posted increased from 11am (when the rain began) to 4 am (when the rain stopped)

which reflected a positive correlation with the storm. Comparing with the peak of the storm, weibo volume showed a little delay that was reasonable because it took time to respond to the event.

2) At 4am on July 22nd, the curve suddenly increased and showed an unusual pattern. Data between 4am-8am could be interpreted as self-filtered data, where because of the unusual hours (that most ordinary citizens would be at sleep), that the personals up and posting about the storm were the ones more involved with the storm. That is, the noise in the data was reduced because the weibos posted were more likely to be closely associated with the event.

3) At 8am, the curve started to decrease rapidly probably because a large portion of the population that used weibos started working around this time.

4) The amount of weibos keyed with "storm" decreased the day after the storm and continuously decreased until its final peak when reports regarding the death toll during the disaster were released.

4.3 Content Analysis

In this thesis, the frequency of words' occurrence was assumed to be correlated to the importance of the term. Hence key concepts were extracted from the high frequency terms, and their relations were explored with respect to time.

In addition, a preliminary classification framework was developed to examine the type of weibos that were likely to occur in response to the disaster by the hour.

4.3.1 High Frequent Term

One of the major challenges in parsing Chinese was that a character was not necessarily the smallest unit in Chinese. That is, depends on what character was in front of it or behind it, it might represent different meanings depending on the phrase. Hence a naive counting of the number of times each character occurred was not the appropriate method. Instead, word segmentation method has to be applied.

This work applied the ICTCLAS (Institute of Computing Technology, Chinese Lexical Analysis System) which was based on the multi-layer HMM to parse meaning from sentences in each weibo.

Next, characters that did not add much information for the purpose of this study were removed. This was done by using stop word lists. In particular three stop word lists were used, the first two were popular ones for Chinese language, and the third one was manually generated customized for this work.

Table 4.2 Stop word dictionary	
Harbin Institute of Technology	**1259**
Baidu stop word list	**842**
"Storm" weibos	**102**

The customized "Storm" weibo stop list included two types of words: 1)common words which were not related to the topic including grammatical words such as pronouns, prepositions and articles 2) words with meanings but such meanings were not relevant to this work, such as weibo users' names, and words like "repost" or "post weibo" etc. Any word in the weibos that matched these stop words was filtered out to reduce the noise in the data.

After performing word segmentation, the top 200 most high frequency words were extracted. Stop word lists removed 98 of them, leaving 102 words as shown in Figure 4.20 (left). In this list, many words expressed similar information. For example, heavy storm(暴雨), downpour(大雨), intense rainfall(强降雨) were near synonyms. Moreover, words like heaviest(最强), hit(袭击), 60 years-return-period(六十年一遇) all described the storm as well. Words with similar meanings were combined into several sets. Among these sets, series words of "storm" and "home" were important to convey the meaning of weibos. They were marked in the same color and were copied into the left portion of the figure Among the 61 terms on the left; terms with grey background were too general to indicate the particularity of the storm and the disaster. For instance, "finally", "news", "stuff" or "experience" were important in the context but too common.

1	大暴雨	heavy storm	52	家里	at home
2	回家	go home	53	房山	Fangshan
3	大雨	intense rainfall	54	去年	last year
4	积水	waterlogging	55	狂风暴雨	furious storm
5	城市	urban	56	中国	China
6	小时	hours	57	蓝天	blue sky
7	最大	heaviest	58	为什么	why
8	路上	on the road/way	59	电话	telephone
9	新闻	news	60	车辆	cars
10	平安	peace	61	情况	situation
11	希望	hope	62	阳光	sunshine
12	朋友	friend	63	工作	work
13	机场	airport	64	小区	residential area
14	排水	drainage	65	早上	morning
15	晚上	night	66	灾难	disater
16	过后	storm passed	67	人员	stuff
17	到家	arrive home	68	今晚	tonight
18	下午	afternoon	69	世界	world
19	遭遇	suffer	70	谢谢	thanks
20	出门	leave home	71	今日	today
21	雨后	after storm	72	雨水	rain water
22	终于	finally	73	报道	reports
23	救援	rescue	74	人民	the public
24	感动	move/moving	75	手机	cellphone
25	外面	outside/ourdoor	76	雨中	in the rain
26	回来	come back	77	致敬	respect
27	首都	capital	78	生活	living
28	死亡	death	79	不了	disenable
29	明天	tomorrow	80	地铁	subway
30	经历	experience	81	毫米	millimetre
31	预报	forecast	82	航班	flights
32	预警	early warnging	83	发布	publish
33	特大	disastrous (rain)	84	居然	unexpectly
34	系统	system	85	市民	citizens
35	感谢	appreciation	86	部门	department
36	看海	watch sea	87	司机	driver
37	感觉	feeling	88	影响	affect
38	交通	transportation	89	精神	spirit
39	赶上	catch	90	滞留	detention
40	年一遇	(a-fourty/sixty-)year return period	91	最强	strongest
41	游泳	swim/swimming	92	不知	not informed
42	生命	life	93	帮助	help
43	开车	drive a car	94	我家	my home
44	时间	time	95	一夜	one night
45	京城	capital city	96	温暖	warm
46	视频	video	97	汽车	motor vehicle
47	天空	sky	98	看不见	cannot see
48	政府	government	99	能量	energy
49	在家	stay at home	100	电视	televison
50	严重	serious	101	免费	free
51	昨晚	last night	102	孩子	children

4	积水	waterlogging	60	车辆	cars
5	城市	urban	61	情况	situation
9	新闻	news	63	工作	work
10	平安	peace	64	小区	residential area
11	希望	hope	66	灾难	disater
12	朋友	friend	67	人员	stuff
13	机场	airport	69	世界	world
14	排水	drainage	70	谢谢	thanks
22	终于	finally	73	报道	reports
23	救援	rescue	74	人民	the public
24	感动	move/moving	75	手机	cellphone
25	外面	outside/ourdoor	77	致敬	respect
28	死亡	death	78	生活	living
30	经历	experience	79	不了	disenable
31	预报	forecast	80	地铁	subway
32	预警	early warnging	82	航班	flights
34	系统	system	83	发布	publish
35	感谢	appreciation	84	居然	unexpectly
36	看海	watch sea	85	市民	citizens
37	感觉	feeling	86	部门	department
38	交通	transportation	87	司机	driver
39	赶上	catch	88	影响	affect
41	游泳	swim/swimming	89	精神	spirit
42	生命	life	90	滞留	detention
43	开车	drive a car	92	不知	not informed
46	视频	video	93	帮助	help
48	政府	government	96	温暖	warm
50	严重	serious	97	汽车	motor vehicle
53	房山	Fangshan	98	看不见	cannot see
56	中国	China	99	能量	energy
58	为什么	why	100	电视	televison
59	电话	telephone	101	免费	free
			102	孩子	children

Figure4.20 High frequency terms lists

4	积水	waterlogging	36	看海	watch sea	60	车辆	cars	87	司机	driver
10	平安	peace	38	交通	transportation	63	工作	work	89	精神	spirit
11	希望	hope	41	游泳	swim/swimming	64	小区	residential area	90	滞留	detention
13	机场	airport	42	生命	life	66	灾难	disater	93	帮助	help
14	排水	drainage	43	开车	drive a car	69	世界	world	96	温暖	warm
23	救援	rescue	48	政府	government	70	谢谢	thanks	97	汽车	motor vehicle
23	感动	move/moving	50	严重	serious	77	致敬	respect	99	能量	energy
28	死亡	death	53	房山	Fangshan	78	生活	living	101	免费	free
31	预报	forecast	56	中国	China	80	地铁	subway	102	孩子	children
32	预警	early warnging	58	为什么	why	82	航班	flights			
35	感谢	appreciation	59	电话	telephone	86	部门	department			

Figure 4.21 Key terms extracted by high frequency

After removing the words in gray background, 42 key words were generated. Notice that "storm" series and "home" series were both crucial. In additional, the top 30 frequent terms for each hour were also extracted and added terms such as paralysis, mountain torrents, license plate and insurance. The following charts showed the top five high frequent terms and several interesting high frequent terms in top 30 in each hour. Some of them were difficult to translate into English directly. For instance, "Watch sea" is an ironic humor which describing Beijing became an ocean when the storm occurred. Bejiyuan, Guangqumen, Lishui bridge and Tianjin were locations in Beijing Beiyuan was a residential area and Guangqumen and Lishui bridge were overpasses. Tianjin was another big city near Beijing, where storm occurred on July 22 and caused flash flood as well. From the chart, there was a football game at 7pm on July 21, and weibo users thought the city was paralyzed at 10pm on July 21. The first group of citizens had arrived home at 9pm on July 21. Thousands of people were still on their way home at 00am on July 22. In the first four hours of the storm, many people enjoyed it because it helped with relief some of the air population. Seven hours after the initial start of the rain, citizens realized that it was turning into a disaster. Rescue information was distributed quickly at 2am and positive emotion followed. From 7pm to midnight on July 22, critiques and suggestions occupied the sina-weibo, followed with strong negative emotion. At 23 pm on July 22, the topic of "death" became the hottest topic.

TOP 5													
1	heavy storm	heavy storm	heavy storm	heavy storm	heavy storm	heavy storm	heavy storm	heavy storm	heavy storm	heavy storm	heavy storm	heavy storm	heavy storm
2	Fangshan	forecast	go out	intense rainfall	go home	intense rainfall	Beiyuan	go home	water-logging	water-logging	go home	water-logging	go home
3	forecast	Fangshan	intense rainfall	go home	on the road/way	go home	swim/swimming	intense rainfall	go home	go home	on the road/way	on the road/way	arrive home
4	go out	go out	go home	go out	go out	go out	intense rainfall	water-logging	on the road/way	intense rainfall	intense rainfall	arrive home	water-logging
5	warning	news	watch sea	on the road/way	intense rainfall	Beiyuan	go home	come back	intense rainfall	on the road/way	arrive home	urban	on the road/way
	11:00	12:00	13:00	14:00	15:00	16:00	17:00	18:00	19:00	20:00	21:00	22:00	23:00
OTHER H-F TERM	blue 10	haha 12	downfall 16	mood 14	airport 10	cancel 22	foreigner 6	Guo'an 11	Guo'an 9	knee 10	drainage 9	Guangqumen 28	hope 8
	Haidian 21	sleep 14	like 18	ha-ha 23	Guo'an 17	delay 29	tragedy 17	dinner 19	football fans 22	residential area 13	40years 16	paralysis 30	rescue 24
	work overtime 29	waiting for (the rain) 20	work overtime 24	children 27	yellow warning 29		take a bath 19	transportation 21	serious 26	subway 19			
	swelter 30	happiness 24					Lishui bridge 24		heaviest 30	hope 22			

Figure 4.22-1 High frequency words in each hour

	0:00	1:00	2:00	3:00	4:00	5:00	6:00	7:00	8:00	9:00	10:00	11:00
TOP 5 1	go home	peace	peace	rescue	rescue	rescue	heavy storm	heavy storm	heaviest	water-logging	urban	heavy storm
2	heavy storm	go home	go home	energy	peace	heavy storm	peace	news	intense rainfall	urban	heavy storm	urban
3	arrive home	hope	friend	peace	friend	peace	heaviest	heaviest	heavy storm	intense rainfall	intense rainfall	water-logging
4	intense rainfall	heavy storm	hope	hope	hope	after rain	passed	urban	news	heaviest	water-logging	heaviest
5	heaviest	intense rainfall	rescue	friend	heavy storm	passed	after rain	drainage	water-logging	drainage	heaviest	intense rainfall
OTHER H-F TERM	peace 9	energy 14	love 14	ask for help 9	service 13	good morning 20	respect 6	death 27	sunshine 19			
	last year 18	respect 20	positive 16	hotel 20	summary information 14	blue sky 23	life 27	government 30				
	Guangqumen 30	moving 25	detention 24	distribution 21	disaster 29		free 29					
				phone number 28								

Figure 4.22-2 High frequency words in each hour

	12:00	13:00	14:00	15:00	16:00	17:00	18:00	19:00	20:00	21:00	22:00	23:00
TOP 5 1	heavy storm	heavy storm	heavy storm	after rain	heavy storm	heavy storm	Tianjing	heavy storm	heavy storm	heavy storm	death	death
2	airport	after rain	passed	heavy storm	water-logging	after rain	Ugly	after rain	after rain	intense rainfall	news	news
3	urban	passed	heaviest	passed	intense rainfall	intense rainfall	moving	moving	moving	heaviest	heavy storm	causing death
4	intense rainfall	urban	after rain	intense rainfall	after rain	water-logging	heavy storm	car license	peace	urban	urban	heavy storm
5	heaviest	water-logging	urban	urban	urban	heaviest	after rain	news	intense rainfall	drainage	catastrophic	urban
OTHER H-F TERM		sunny day30		pipes 30	citizens 21			ugly 9	department 9	free 10	spirit 22	life 9
								grief 30	life 28	world 23	loss 25	drowning 10
									sewer system 29	flood 25		collapsed 14
										Japanese 27		

Figure 4.22-3 High frequency words in each hour

From these words, it appeared that there were a couple of recurring themes among the type of messages people posted in response to the disaster. In order to examine what types of massages were posted, this work developed a classification system to categorize these high frequent words and weibos. The process of building the classification scheme was mixed. This work referred to the categorization schemes from existing literature as the top-down method (Yan Qu et.all, 2009; Vieweg et.all, 2010). A new classification scheme was developed for processing the weibos from this specific case using geo-tags.

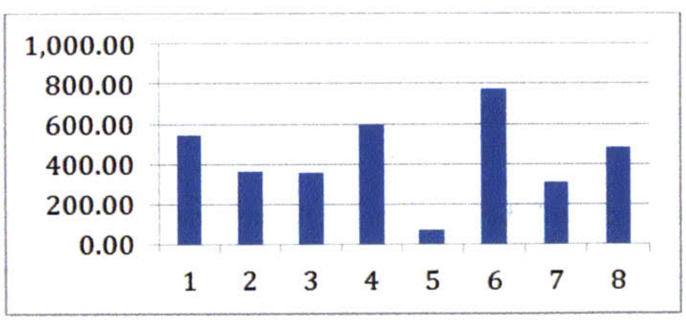
Table 4.3 Preliminary Classification		
1. INFORMATION RELATED		
Situation Awareness	Providing true disaster-related information describing what was happening	heavy storm, water-logging, mountain torrents,
Disaster Consequence	Providing factual information about the consequences the disaster caused	death, victim, life water, license plates, insurance
Critique &Suggestion	Criticizing disaster defense related departments; providing suggestions to mitigating the disaster	infrastructure, Japanese drainage system, urban water management
2. ACTION RELATED		
Personal Activity	Describing personal activities in the storm	have dinner, party, go home, visiting
Relief Action	Requesting help or disseminate rescue information	fire-fighters, seeking, missing people, vehicle-team, soldiers
3. EMOTION RELATED		
Personal Feeling	Expressing personal feeling to the storm	exciting, angry, sadness, anxiety, moving
Social Emotion	Expressing emotional support to the society	blessing, peace, love, mourning, positive energy
4. OTHERS:	Off-topic, irrelevant to the disaster	football games, restaurant, advertisement

After reviewing all the key concepts in the context, the following categorization scheme was developed. It was clear that the information-related was the largest category. In terms of subcategories, number of weibos describing situation awareness was higher than others. For future work, this categorization may help with sorting weibo messages and assigning priorities to different categories.

Table 4.4 Key concepts/Domain dictionary in classification

INFORMATION RELATED			ACTION RELATED		EMOTION RELATED	
Situation Awareness	**Disaster Consequence**	**Critique and Suggestion**	**Personal Activity**	**Relief Action**	**Personal Feeling**	**Social Emotion**
waterlogging	death	drainage	drive a car	rescue	moving	peace
airport	life	government	work	help	exciting	hope
forecast	block	China	extra work	free	haha	blessing
early warnging	living	water management	"home" series	fire-fighers	anxiety	positive
watch sea	motor vehicle	world		seeking	angry	respect
transportation	tapwater	department		missing people	sadness	spirit
swim/swimming	licence plate	children		vehicle-team		warm
serious	victim	Japanese		soldiers		energy
Fangshan	insurance	agency		polices		
residential area	manhole cover	construction				
subway						
flights						
detention						
"storm" series						

Regarding to weibos with geo-tags, they were annotated by two volunteers and the result showed that personal feelings messages **occurred** more often than other categories.



1. Situation Awareness 2. Disaster Consequence 3. Critique &Suggestion 4. Personal Activity 5. Relief Action 6. Personal Feeling 7. Social Emotion 8. Others

Figure 4.23 Weibos with geo-tags classification

4.3.2 Key Concept in time series

The frequency of each key term was calculated by taking its number of occurrence and dividing by the number of total weibos in each hour.

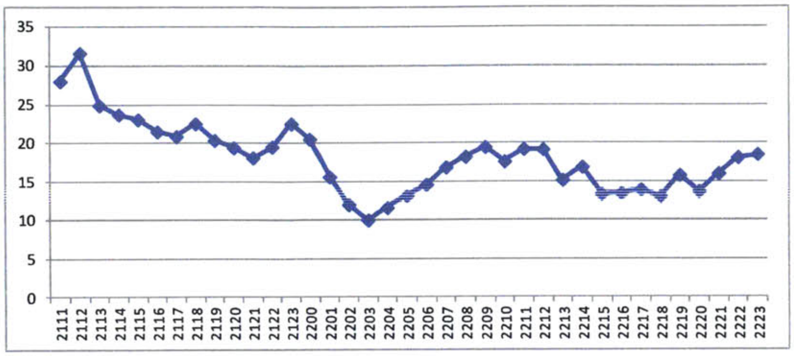


Figure 4.24 Trend of "Heavy storm"

Although the word "storm" was removed from the weibos, the term of "heavy storm" was still in the sample. The curve reached its nadir at 3-4am which coincided with the time when the rain stopped.

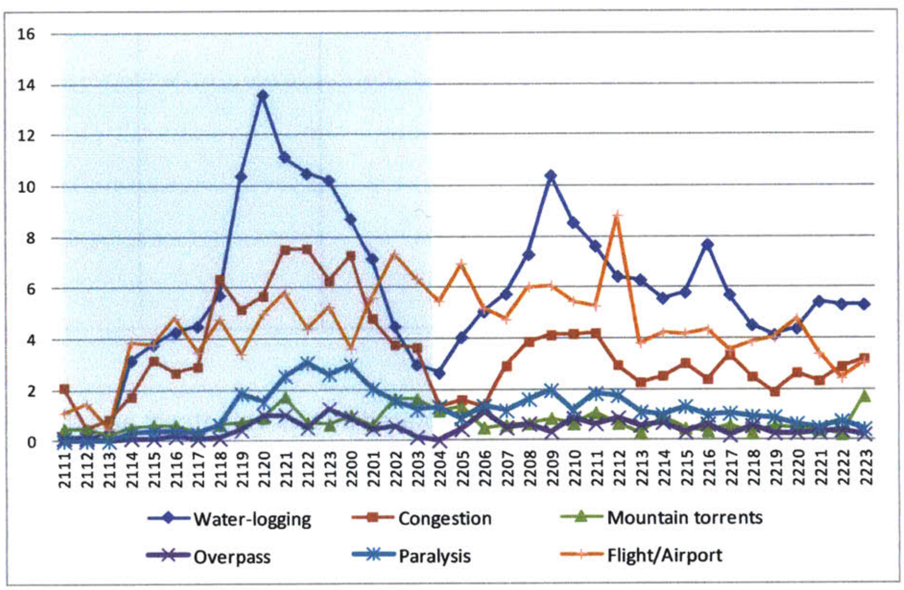


Figure 4.25 Trend of Information-related Keywords

The above figure displayed the occurrence probability of terms in the information-related category. An overall trend was that these terms occurred more frequently during the disaster than after. The curve of water-logging had multiple peaks, and it increased dramatically at 17pm on July 21 and reached the peak at 20pm on July 21. It reached two smaller peaks at 9am and 16am on July 22. The word of "paralysis" reflected the severity of the disaster to some degree. It reached the first high point at 19am 07/21 and reached the peak at 22pm 07/21.

Different from other terms, the terms "airport" and "flights" continued to appear frequently after storm and reached their peaks at 12pm 07/22.

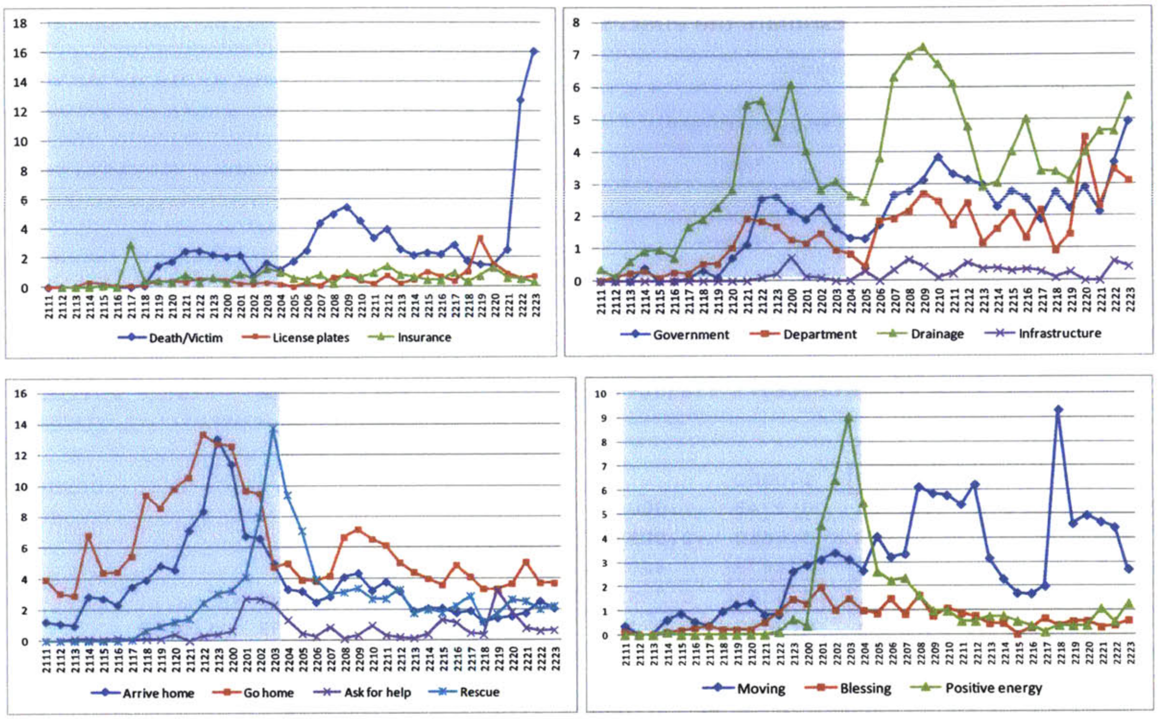


Figure 4.26 Trend of key concepts in other Sub-categories

4.3.2 Key Concepts' Relations

Several methods were used to identify the relation of key concepts. First, some key words from each sub-category were illustrated below for showing the relations between words.

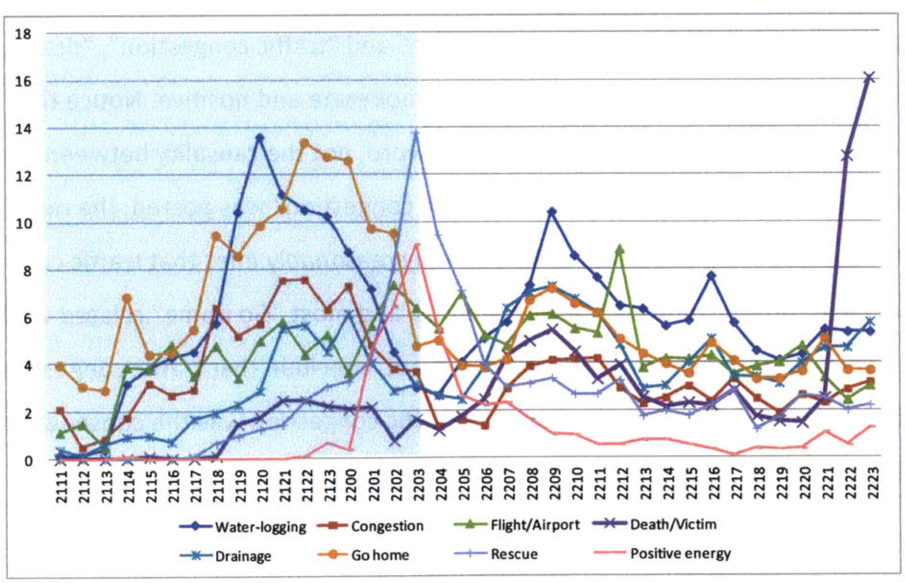


Figure 4.27 Trend of some key concepts from each sub-category

Secondly, Pearson's r (correlation coefficient) was proposed to identify the correlation between two key concepts. In statistics, the following rule of thumb chart was generally accepted, where the scale could be used to estimate the effect size.

Dancey and Reidy's (2004) categorisation:

Value of the Correlation Coefficient	Strength of Correlation
1	Perfect
0.7 - 0.9	Strong
0.4 - 0.6	Moderate
0.1 - 0.3	Weak
0	Zero

	Waterl~g	Conges~n	Flight~t	DeathV~m	Drainage	Gohome	Rescue	Positi~y
Waterlogging	1.0000							
Congestion	$\boxed{0.7841}$	1.0000						
FlightAirp~t	0.3924	0.2743	1.0000					
DeathVictim	0.2043	0.0610	-0.0086	1.0000				
Drainage	$\boxed{0.6256}$	$\boxed{0.4473}$	$\boxed{0.4200}$	$\boxed{0.5701}$	1.0000			
Gohome	$\boxed{0.6993}$	$\boxed{0.8888}$	0.2932	-0.1208	0.3039	1.0000		
Rescue	-0.0584	-0.0042	$\boxed{0.5727}$	0.0469	0.2130	0.0539	1.0000	
Positiveen~y	-0.1977	-0.0898	0.4696	-0.0037	0.0583	0.0104	$\boxed{0.9205}$	1.0000

Figure 4.28 Correlation coefficients of some key concepts

The correlation between "rescue" and "positive energy" was strong and positive. The more rescue activities, the more positive energy would be produced and disseminated in the society. The correlation between "go home" and "traffic congestion", "water-logging" and "traffic congestion", "water-logging" and "go home" were also strong and positive. The correlation between "Drainage" and "water-logging", "drainage" and "traffic congestion", "drainage" and "death/victim", "rescue" and "flight/airport" were moderate and positive. Notice that the result showed the correlation of word frequency in each word, not the causality between them. For example, one could conclude that the more "traffic congestion" was posted, the more "Go home" related weibos were posted as well. One may reasonably infer that traffic congestion can stimulate people to post "Go home" related weibos, and most "Go home" related weibos also mentioned traffic congestion. However, one could not conclude that traffic congestion leads to detention that many people cannot go home or traffic congestion was the consequence of too many people wanting to go home which caused traffic congestion.

Thirdly, co-occurrence words diagram indicates the probability of co-occurrence two specific terms. For example, Fig 4.29 shows that the co-occurrence of “water-logging” and “on the way” occurred 308 times in the same piece of weibo

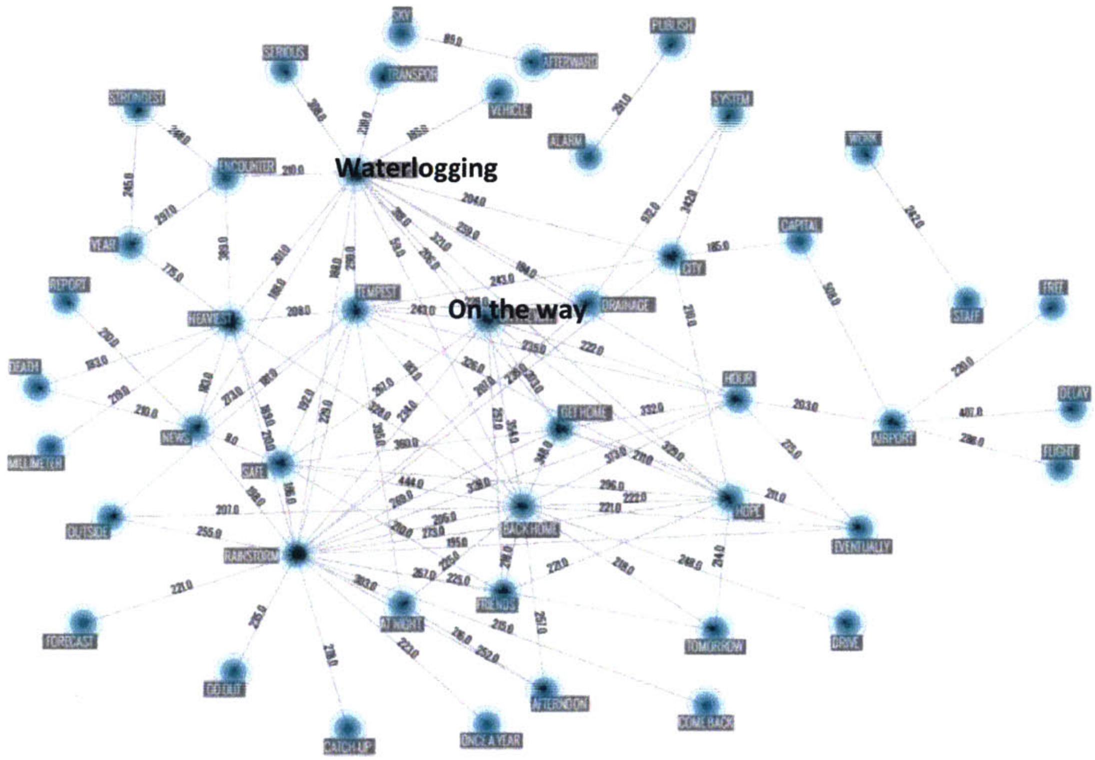


Figure 4.29 Co-occurrence words diagram

4.4 Sentiment Analysis

On a micro level, social media has expanded the information sources and propagated disaster-related messages more efficiently. However, on a macro level, there lacked an effective method to monitor and understand crowd responses. This work would introduce an indicator of sentiment for interpreting people’s responses in order to strengthen situation awareness and influence evaluation.

4.4.1 Data Preprocess

38,528 weibos related to “Storm”(暴雨) from Sina(新浪) were collected and all the contents, including phrases, sentences, and emoticons, to name a few, were translated. The following 249 emotional icons were considered. After translating these icons text into icon image and checking them in the original weibos, they were classified into three categories and seven sub-categories.

Each emotion icon was also identified with an emotional degree of one of the following: -3, -2, -1, -0.5, 0, 0.5,1,2,3.

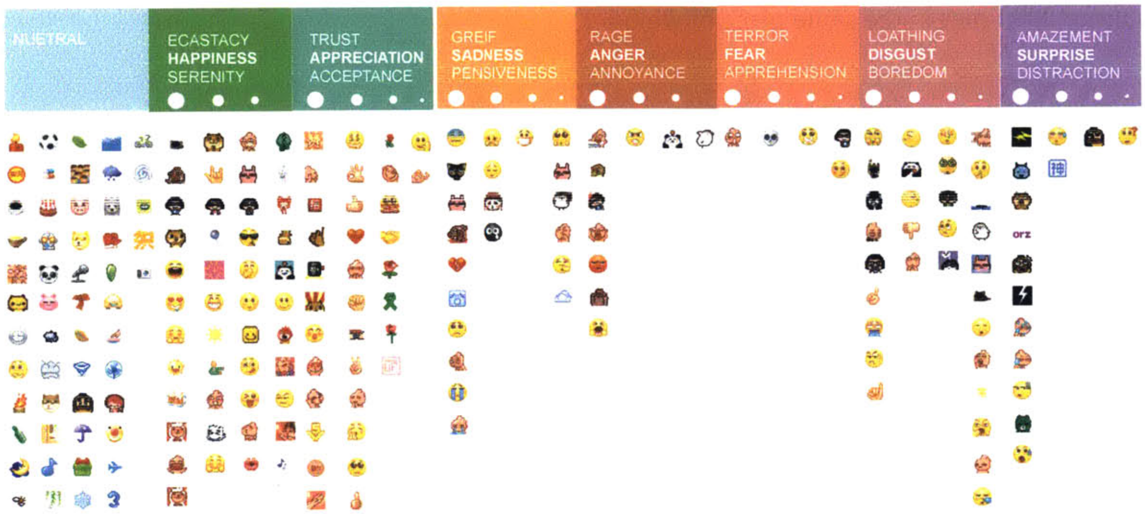


Figure 4.30 Emoticons classification

4.4.2 Sentiment Lexicons

There were few Chinese sentiment lexicons. Among them, the most popular one was the Hownet Chinese Vocabulary. In order to enlarge the positive and negative word lists, this work combined two authoritative dictionaries in Chinese context and this work's own domain dictionary based on the samples of weibos. The three lexicons are:

1. Hownet Chinese Vocabulary for Sentiment Analysis (VSA)(Beta version), published online 10/22/2007.

HowNet Chinese VSA contained 6 sub-files: plus feeling, minus feeling, plus sentiment, minus sentiment, opinion, and degree. It contained Chinese entries (unique forms) of 91,016 words with 4,566 positive words and 4,370 negative words.

2. NTUSD developed by Taiwan University which contained 2,810 positive words and 8,276 negative words.

3. Storm positive and negative words summarized from the 38,528 weibos.

After duplicated words were removed, the three dictionaries mentioned above were integrated into one which was the final lexicon dictionary for this work, and it contained 7,193 positive words and 12,673 negative words.

4.4.3 Sentiment Algorithm

In order to assign a positive, neutral or negative label to each weibo and obtain its opinion towards "heavy storm," the following algorithm was defined.

$$\text{Polarity}\begin{cases}\text{Positive (if posValue > negValue)}\\ \text{Negative (if posValue < negValue)}\\ \text{Neutral (if posValue = negValue)}\end{cases}$$

1. Positive/negative words

The value of a positive word was +1 and the value of a negative word was -1.

Table 4.5 Examples of words in Pos/Neg dictionary

Positive +1	Negative-1
Happy, admiration, consent, welcome, look forward to, good-looking, high-quality, effective, tranquility, safe and sound	Defy, disappointed, fear, criticize, regret, pull a long face, inferior, expensive, brutal, false, gawky, low
怀恋 ,怀念 ,怀想 ,欢 ,欢畅 ,,欢快 ,欢乐 ,欢闹 ,回礼 ,见爱 ,见义勇为 ,奖 ,奖励 ,奖赏,敬爱 ,敬奉 ,敬服 ,敬贺 ,敬礼 ,敬慕 ,敬佩	哀愁 ,哀怜 ,哀悯 ,哀戚 ,哀凄 ,哀切 ,哀伤 ,哀痛 ,恐怖 ,惊诧 ,惊呆 ,惊怪 ,惊骇 ,惊慌,怒, 怒叱 ,怕, 颓废,挖苦, 咒骂,咒诅,抓瞎,瀑布,游泳,海,船

2. Intensification

According to Quirk et al. (1985), intensifiers can amplify or dowtone the neighboring lexical word. They classified intensifiers into two major categories: amplifiers and downtoners. the former can increase the semantic intensity while downtowners decrease it (Quirk et al., 1985). This classification is also applied to Chinese. HowNet Chinese VSA also provides such intensifiers list which includes 219 words. I defined a modifier based on some previous researches. (Xiaodong Cheng, 2012;)

Table 4.6 Percentages for some intensifiers

Intensifier	Modifier (%)
most	+300

Very	+ 200
More	+100
-ish	+50
Insufficiently	-50
Over	-75

3. Negation words

In this algorithm, negation words can inverse the polarity of the sentimental words. Samples of negation words included: no(不，没有), never（未曾，难以）, and won't（不会）.

4. Question marks + question words

These two signs can also inverse the polarity of sentimental words.

5. Exclamation marks

Exclamation marks contribute to identify emotion intensity.

6. Sentiment Algorithm

```
for word in comment:
        i += 1
        if ((word in posdict) or (word in negdict)):
            from_last_emotion += 1
            for w in comment[a:i]:
                if w in mostdict:
                    from_last_emotion *= 4
                elif w in verydict:
                    from_last_emotion *= 3
                elif w in moredict:
                    from_last_emotion *= 2
                elif w in ishdict:
                    from_last_emotion /= 2
                elif w in insufficientdict:
                    from_last_emotion /= 4
            c = 0
            for w in comment[a:i]:
                if w in inversedict:
                    c += 1
                    from_last_emotion = 1
            if (((word in posdict) and (c % 2 == 0)) or
                ((word in negdict) and (c % 2 == 1))):
                poscount += from_last_emotion
            else:
                negcount += from_last_emotion
            save_from_last_emotion = from_last_emotion
            from_last_emotion = 0
            a = i + 1
        elif word == '!':
            for w in comment[::-1]:
                if w in posdict:
                    poscount += 2
                elif w in negdict:
                    negcount += 2
                break
        elif word == "?":
            j = i
            flag = True
            for w in comment[::-1]:
                if flag == True:
                    j -= 1
                    if w in questiondict:
                        for w2 in comment[j:i]:
                            if w2 in posdict:
                                poscount -= 2 * save_from_last_emotion
                                flag = False
                                break
                            elif w2 in negdict:
                                negcount -= 2 * save_from_last_emotion
                                flag = False
                                break
    if ((abs(poscount) > 20) or (abs(negcount) > 20)):
        out_lier[index] = comment
    return [poscount, negcount]
```

4.4.4 Test Set and Evaluation

Test set: 3119 weibos with geo-tags.

Experiment: 7 people annotate polarity by hand for each weibo independently and create 5 results sets.

Evaluation index: precision, recall and F-Measure

Precision: % of selected items that are correct

Recall: % of correct items that are selected

F-Measure: weighted harmonic mean which is very conservative average, people usually balanced F1 measure.

$$F = \frac{1}{\alpha \frac{1}{P} + (1-\alpha)\frac{1}{R}} = \frac{(\beta^2+1)PR}{\beta^2 P + R} \Rightarrow F = 2PR/(P+R)$$

To test the performance of the proposed algorithm, the geo-tagged weibo data set was used. This data set contained 3,119 weibos with polarity labels. The results were presented below.

Table 4.7 Evaluation results of text set

	POSITIVE	NEGATIVE
PRECISION	0.651862	0.731375
RECALL	0.462869	0.704225
F-Measure	0.541344	0.717543

4.4.5 Results

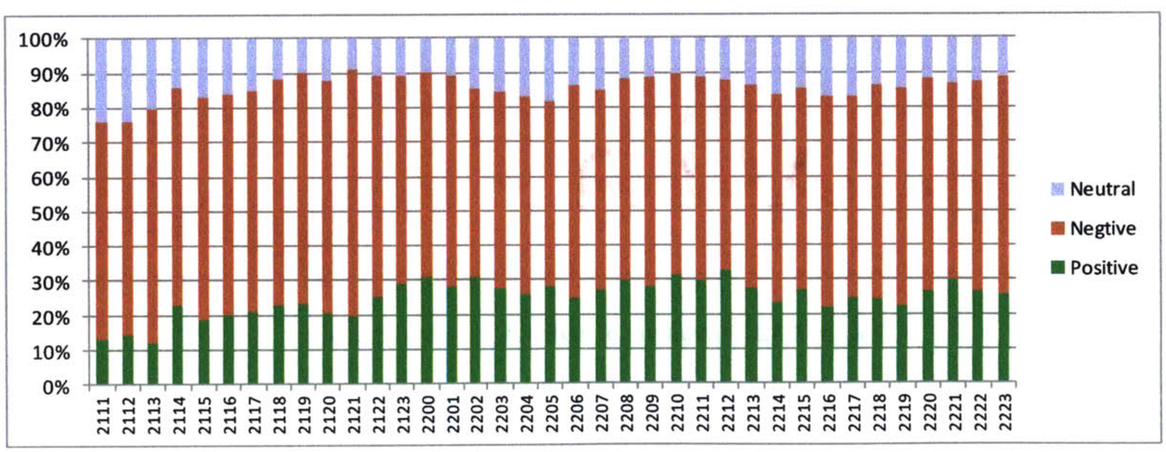


Figure 4.31 Percentage of negative, positive and neutral emotions

Table 4.8 Evaluation results of the sample

Variable	Mean	Std. Dev.	Min	Max
Positive	24.92211	4.948671	12.31884	32.60632
Negative	60.91488	3.777543	53.61272	71.16751
Difference	35.99277	8.044324	22.35551	54.95169

Over 70% of weibos had expressed some form of emotions. Negative emotion was much higher than positive emotion. The mean of negative emotion ratio is 69.9% and the mean of negative emotion ratio is 24.9. Positive emotion ratio was higher than the expectation. According to Fig.4.33, the difference of positive and negative emotion ratio is decreasing over the time. We could infer that the overall trend of negative emotion was decreasing.

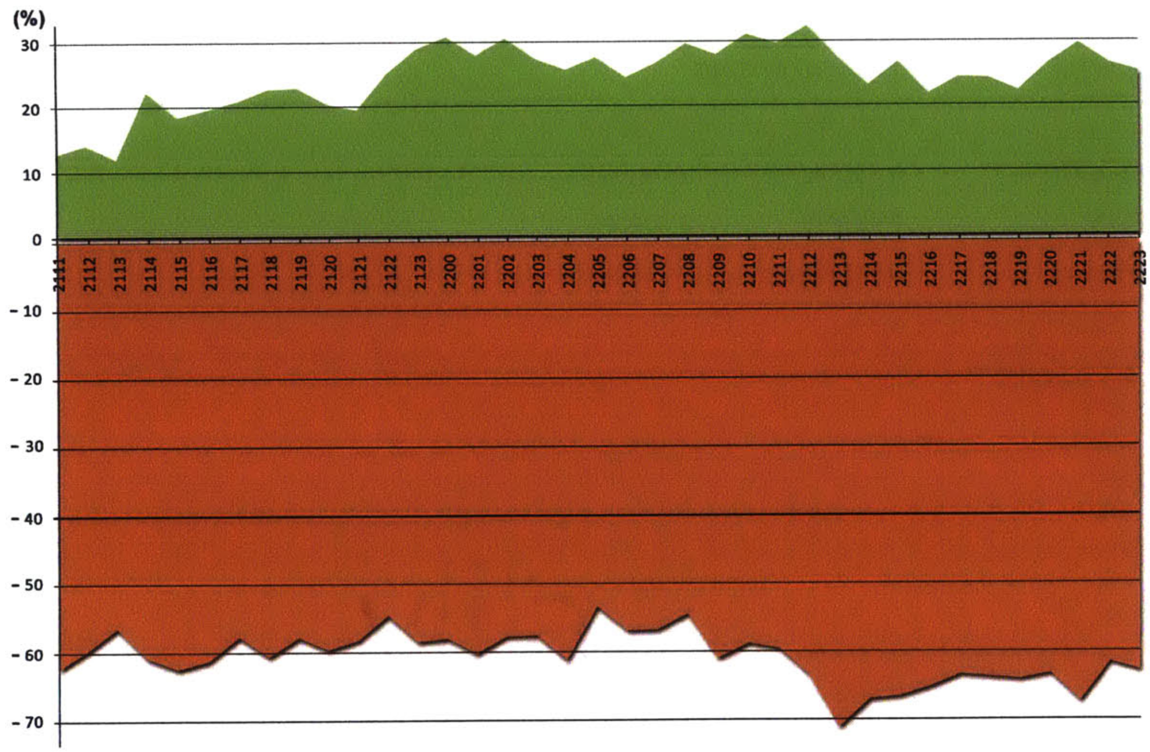


Figure 4.32 Percentage of negative, positive emotions

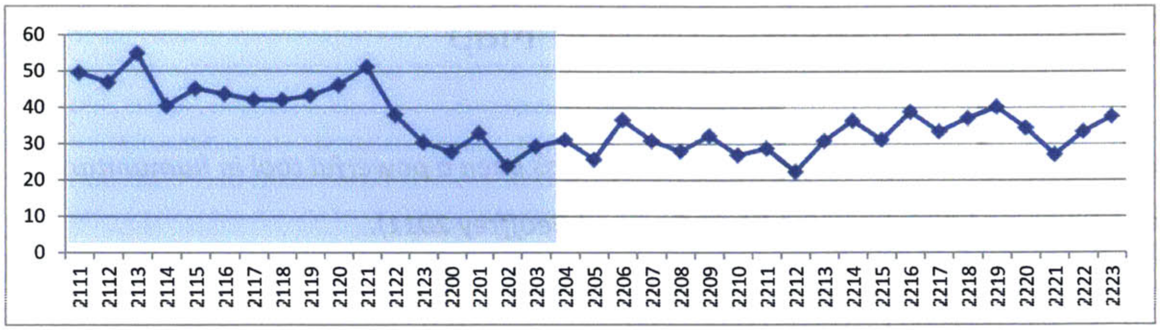


Figure 4.33 Difference of positive and negative emotion ratio in time series

The results provided the following insights:

1) The accuracy of the negative emotion was higher than that of the positive emotion. One of the possible reasons was that many messages did not express positive emotion directly, which made it difficult for the algorithm to correctly identify. For examples, many weibo users mentioned that it was not convenient for them to walk in the rain, and they were so excited to arrive home safely. Many people complained the rainfall was too heavy, and it cooled the air and made them happy.

2) The accuracy of the neutral emotion was very low. This problem was common in lexicon-based analysis approach which aimed to identify the positive words and negative words. The author remained consistent with previous practice in the literature. The low neutral accuracy showed a deficiency in dealing with values close to zero in the algorithm. In order to improve the accuracy of neutral, the definition can broaden to include 'close' instead of just exactly equal.

3) Expression icons, question marks, question words, exclamation marks could improve the accuracy. Among them, the expression icons were particular useful in identifying emotions.

4) Weibo messages tended to not follow established Chinese grammar. It more closely resembled spoken language and could use abbreviations and slangs. This feature added challenges for the algorithm proposed in this work to correctly identify the users' emotions. In addition, the volunteers' interpretations of the weibos could be different too. According to the results of test set, the probability of the identical judgment appearing in all five samples set was less than 50%. Compared to the disagreement in human interpretation, the algorithm-generated results appeared to be reasonable.

Chapter 05 Disaster Emotion Map

Crowd-sourcing integrated with crisis maps has been a powerful tool in humanitarian assistance and disaster relief (Huiji Gao and Geoffrey 2011).

5.1 Geo-location data preprocessing

5.1.1 Geo-data Acquisition

1. Extracting weibos with geo-tags

This work recognized geo-tags in the following forms, where <address> designated a specific address entered by individual users:

- I am here # <address>
- I am in # <address>
- Short links after text

2. Expand short links to standard links

This work expanded links that were previously shortened into their standard format so that the information can be extracted and compared. In particular, the following types of shortened links were preprocessed:

- Links with longitude and latitude
- Links with poiid
- Links with Jiepang geo-tags


```
{
    "urls": [
        {
            "url_short": "http://t.cn/h4DwT1",
            "url_long": "http://finance.sina.com.cn/",
            "type": 0,
            "result": "true"
        },
        {
            "url_short": "http://t.cn/h4DwT1",
            "url_long": "",
            "type": 0,
            "result": "false"
        },
        ...
    ]
}
```


A sample output form Sina API: https://api.weibo.com/2/short_url/expand.json

3. Converting poiid and Jiepang geo-tags to longitude and latitude

This work extracted the longitude and latitude from poiid and jiepang[6] geo-tags:


```
{
    "poiid": "B2094654D16CABFE419E",
    "title": "理想国际大厦",
    "address": "北四环西路58",
    "lon": "116.30987",
    "lat": "39.98437",
    "category": "46",
    "city": "0010",
    "province": "",
    "country": "",
    "url": "",
    "phone": "010-82625868",
    "postcode": "100000",
    "weibo_id": "0",
    "categorys": "44 46",
    "category_name": "楼宇",
    "icon": "http://ul.sinaimg.cn/upload/2012/03/23/1/lyjg.png",
    "checkin_num": 54484,
    "checkin_user_num": "55",
    "tip_num": 54420,
    "photo_num": 14,
    "todo_num": 9
}
```


A sample output form Sina API: https://api.weibo.com/2/place/pois/show.json

4. Converting Baidu map coordinates to Google map coordinates

Google map used iso-WGS84coorinates, while Baidu map used earth-BD-09 coordinates which encrypted mars-GCJ-02 coordinates.

This work converted the longitude and latitude from the Baidu map coordinates into the Google map coordinates, because the latter was more applicable to down-stream applications like the GIS tools or Google earth.

5. Description of Generated Data

After performing the data preprocessing described in the previous subsections, 38,528 weibos were extracted. Fig.5.1 illustrated the number of weibos accumulated every hour from 11am in July 21, 2012 to 11pm on July 22, 2012. The gray bars represent the total amount of weibos extracted from each hour, and the orange bars represent the ones with geo-tags. The dots represented the percentage of weibos that were geo-tagged, and the lines were linear curves connecting the dots. The shaded blue areas in the background illustrated the storm intensity during that period of time for reference.

---

[6] Jiepang is China's leading location based service (LBS) for the "check in", similar to Foursquare and was frequently called the Foursquare of China.

Over this illustrated domain of 37 hours, 38.528 weibos were sampled, of which 3,119 were tagged with geo-tags. That is, 8.1% of the weibos in this sample size was tagged with geographical information.

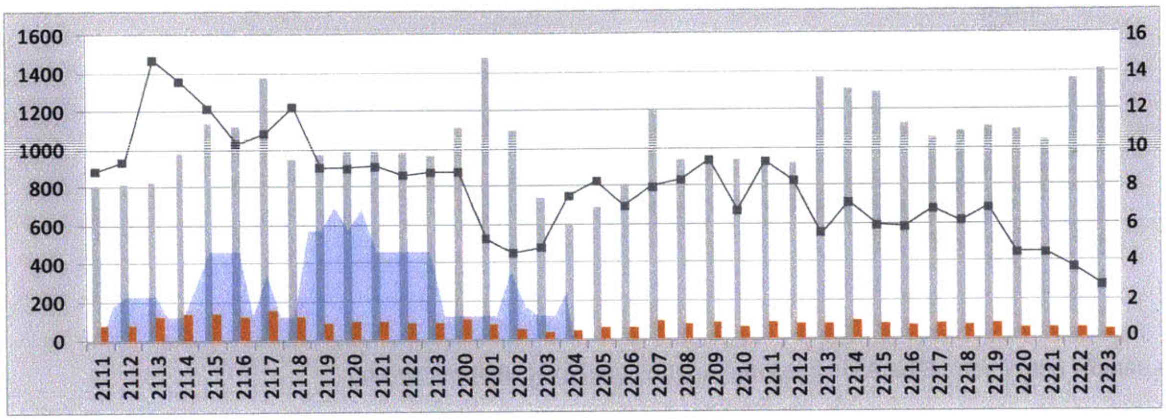


Figure 5.1 Counts and ratio of weibos with geo-tags by hour

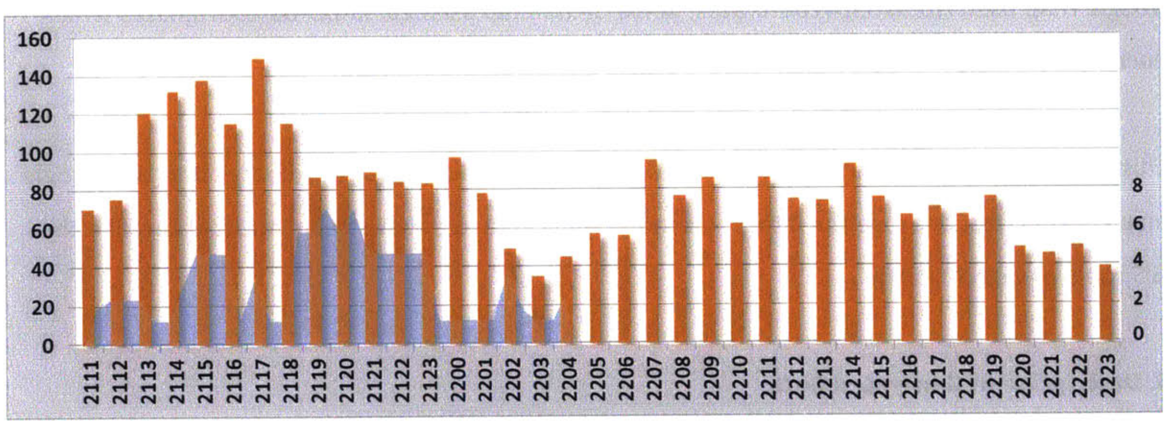


Figure 5.2 Numbers of weibos with geo-tags by hour

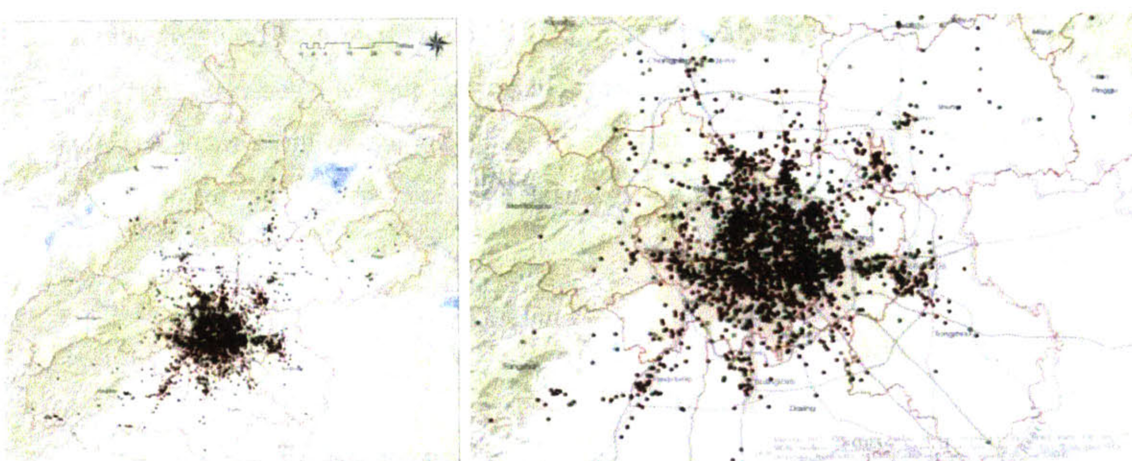


Figure 5.3 Spatial distributions of weibos with geo-tags

Data Distribution:

Fig.5.3 illustrated the spatial distribution of the geo-tagged weibos in the sample. The data were primarily concentrated within the five ring roads, corresponding to the most densely populated areas of Beijing. Outside the center of the City, weibos spread out following major highways. Visually, these figures demonstrate that a reasonably large amount of data was collected to provide enough information for the areas that most people reside in, as well as to capture some of the characteristics for the regions that did not have enough data previously without the help of crowd-sourced data.

5.5.2 Characterizing Emotion Polarity, Intensity and Type

Identifying seven emotion categories in Chinese:

Paul Ekman who ranked 59th out of the 100 as one of the most cited psychologists of the twentieth century, identified six basic types of emotions. Through an extended series of studies, Ekman found a high agreement across members of diverse Western and Eastern literate cultures on labeling facial expressions with emotion types (Ekman, 1999). He considered some expressions to be universal included those indicating anger, disgust, fear, happiness, sadness, and surprise. Previous research specific to the Chinese community further classified another category of emotions labeled as appreciation (Dalian Institute technology, Emotion Ontology).

This work focuses on the seven types of emotions listed in the table below. For each type, a couple of examples were given to further illustrate the concepts. Being able to distinguish one emotion from another, and characterizing the massive amount of weibos into these categories was crucial to this thesis because it provided detailed insights into how citizens of Beijing responded to one of the major disasters of this decade.

Table5.1: Seven Emotion in Chinese

Happiness 乐	Joy 快乐、Peace 安心、Serenity 平静;
Appreciation 好	Reverence 尊敬、Admiration 赞扬、Trust 相信、 Love 喜爱、Blessing 祝愿;
Sadness 哀	Greif 悲伤、Disappointment 失望、Guilt 疚、Miss 思;
Anger 怒	Annoyance 愤怒、Rage 狂暴;
Fear 惧	Panic 慌、Terror 恐惧、Timid 羞;
Disgust 恶	Anxiety 烦闷、Loathing 憎恶、Critique 贬责、 Jealousy 妒忌、Doubt 怀疑、Contempt 轻蔑;
Surprise 惊	Astonishment 惊奇、Surprise 惊讶

Labeling weibos with emotion types and intensities

For this analysis, six Chinese volunteers and I independently performed manual characterization of the weibos in terms of three qualities: emotion polarities, emotion intensities, and emotion types:

- Emotion Polarities: positive (+1), neutral (0), negative (-1);
- Emotion Intensities: 1,3,5,7,9;
- Emotion types: happiness, appreciation, sadness, anger, fear, disgust, surprise.

Among the seven persons performing independent analysis, three processed the whole set of 3,119 weibos with geo-tags, while the rest four each processed half of the dataset, making it in total five sets of analysis of the weibos.

These five sets of analysis were integrated manually, where for each specific weibo, the highest occurring characterization among the five was chosen as the final answer. The underlying assumption was that the emotions expressed in the weibos may not always be clear, and each individual person's response to the same message may vary often be different. To this end, I chose the answer with the highest number of occurrence, because the more the number of persons agreed with one characterization, the more likely it was to be representing the more general audience.

5.2 Flood Sentiment Map

Section 5.1 presents the weibo volume in the absolute terms. In these sections, disaster specific weibos would be discussed, and those data needed to be presented in a normalized fashion. This subsection presents two different methods for normalization.

5.2.1 Normalization by Districts

The top three districts and counties by volume of weibos with geo-tags were Chaoyang, Haidian and Changping. This ranking was generated by performing Sina-weibo searches by location, i.e., district and county. One note for this analysis was that the total weibo count by adding up the different districts and counties in Beijing was much less than the weibo count by searching by location "Beijing." However, the numbers can show a pattern of the weibo posted in each district and county which can be used for normalization. I got updated top three districts and counties where weibos attached with geo-tags. They are Shunyi, Fangshan, Miyun and Yanqing (they tied for the third places). The latter two are counties which are less developed areas and suffered riverine flooding. Fangshan is the hardest hit in this "7.21" Flood Disaster where the flash flood and the riverine flood killed 42 people.

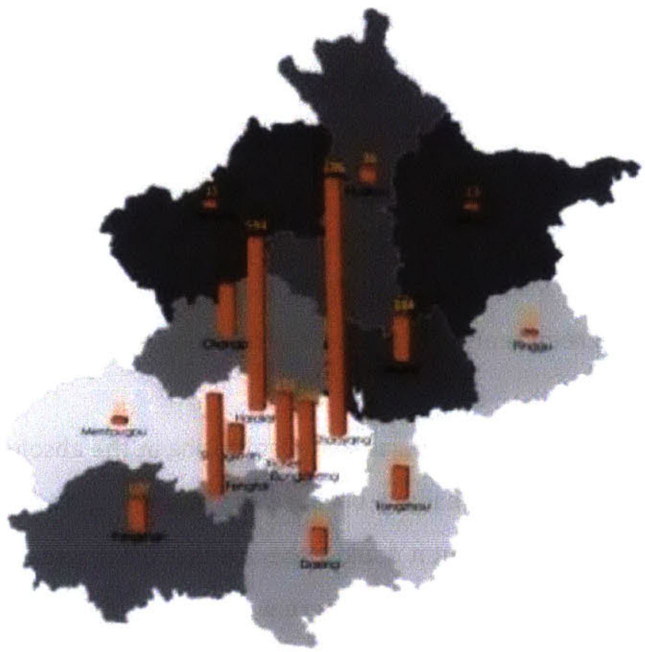


Figure 5.4 Weibo distribution normalized by weibo volume of district

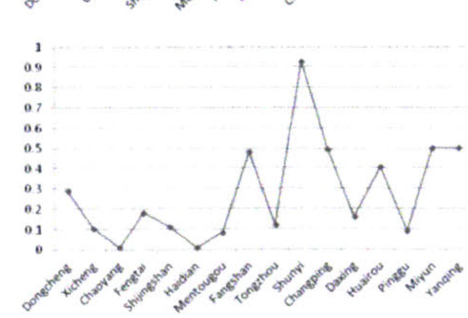


Figure 5.5 Numbers of weibos with geo-tags in district (top) and Percentage of weibos with geo-tags by district weibo volume (bottom)

5.2.2 Normalization by Raster

I have established methods to extract the weibos with geo-tags (3,119 for this sample), to access their geographical information so they can be visualized on maps, and to assign emotion types, intensities and polarities to them.

In this subsection I would present a method for combining the above information into one integrated map. First, I defined cell describing 1000m by 1000m in physical dimension. This dimension was chosen to be large enough to accumulate enough counts of weibo for generating useful information, but not too large such that the spatial resolution was lost. Next for each of the cell, I computed the percentage of weibos tagged with negative emotion over the total number of weibos. Figure 5.6 visualized these results, where red represented higher percentage of negative emotion, and green represented lower percentage of positive emotion.

The rationality behind this normalization was that just looking at the absolute number of weibos with negative emotions could be misleading, because the central region of Beijing had the highest number of negative weibos, but it could either be that there was actually higher concentration of negative emotions, or simply there were more population generating more volume of weibos in total.

This analysis showed that even if there was a higher volume of weibos with negative emotion in the central region of Beijing, the normalized results were actually lower in much of the central area. In comparison, the outer suburbs and rural areas had a significantly higher percentage of negative emotions, visualized by the red dots scattered in the figure.

We can see that the normalized sentiment map matched the locations of severe disaster. Figure 5.6 marked the locations of reported death and water-logging. The sentiment map successfully identified some of the affected regions in the inner city and inner suburb areas. It provided a better agreement for some of the outer suburb and rural areas.

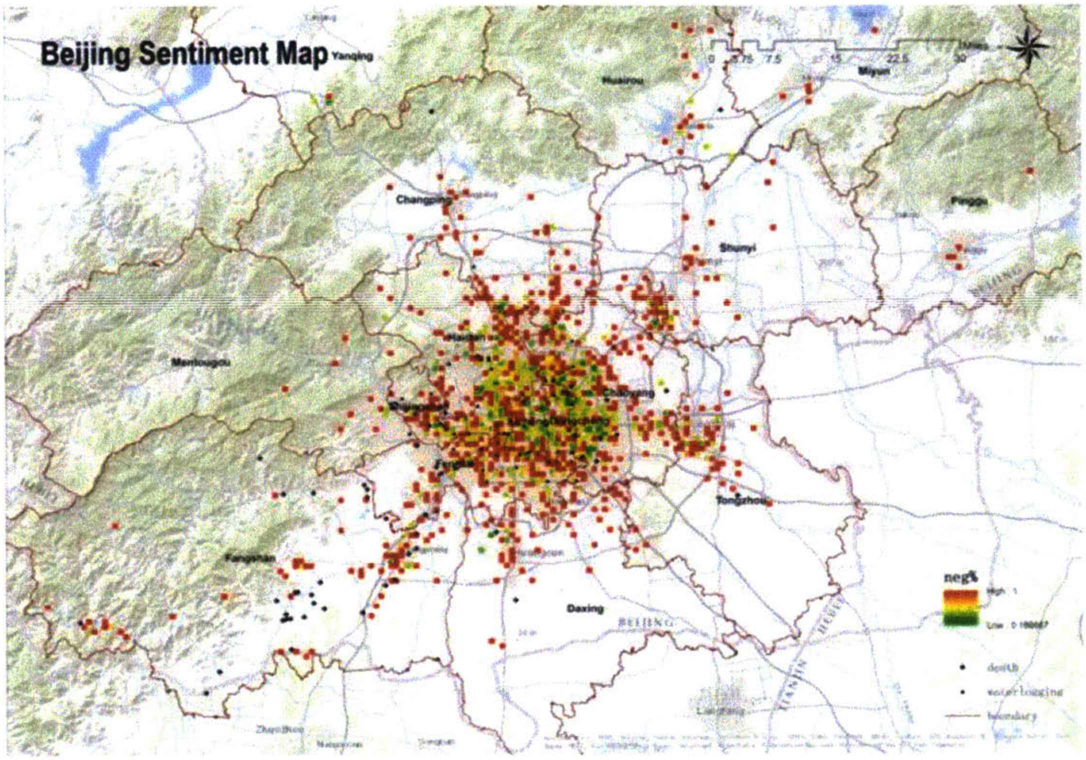


Figure 5.6 Sentiment distribution normalized by negative emotion ratio

The sentiment map could be a powerful tool in guiding crisis management and disaster response effort. For example, during the disasters when resources are limited, this method could process data during real-time and provide map isolating locations of the highest concentration of negative emotion. Based on the comparison in this subsection, these areas of highly concentrated negative emotions matched areas that had the most severe consequences. Hence this sentiment map could provide real-time guidance for disaster relief effort, especially in locations further away from population centers.

5.3 Flood Sentiment Intensity Map

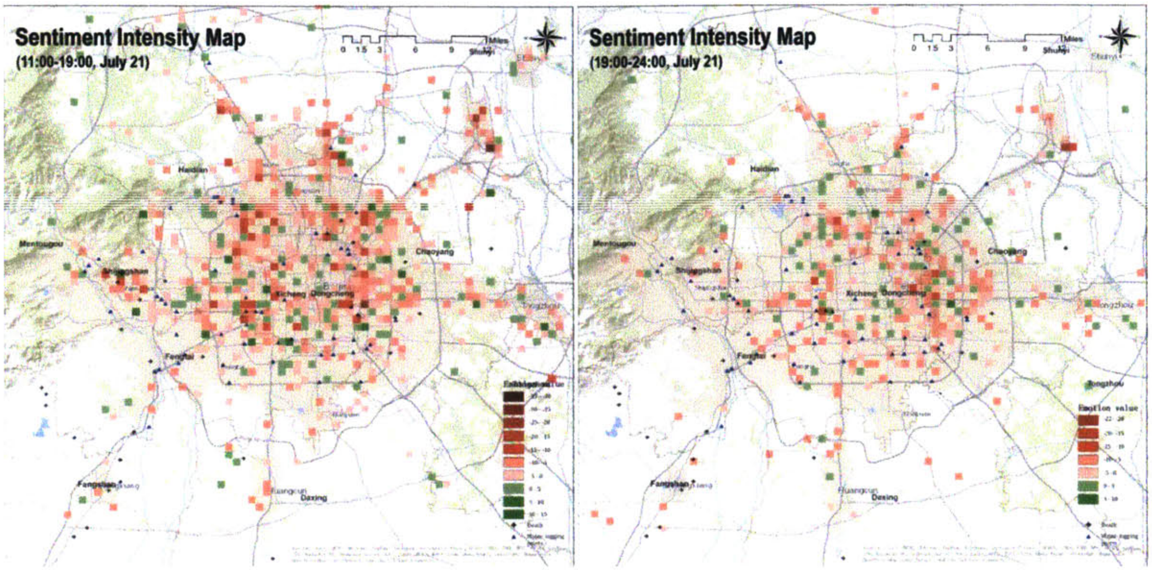


Figure 5.7-1 Sentiment intensity map

I divided the 37 hours period into six time period based on time and rainy and examined the spatial distribution of emotions for each of the time periods in the form of sentiment intensity map. The purpose of this analysis was to identify extreme emotions in both space and time. To be more specific, for each of the following time periods, the following districts generated the most negative emotions:

- From 11:00 to 19:00 on July 21, Capital airport, Tonghui River, Outer Guangquanmen overpass, Beijing Railway Station, Lotus overpass, Zhongguancun Business Area
- From 19:00 to 24:00 on July 21, Beijing Western Railway Station, Zhongguancun Business Area and South Street outside Chaoyangmen

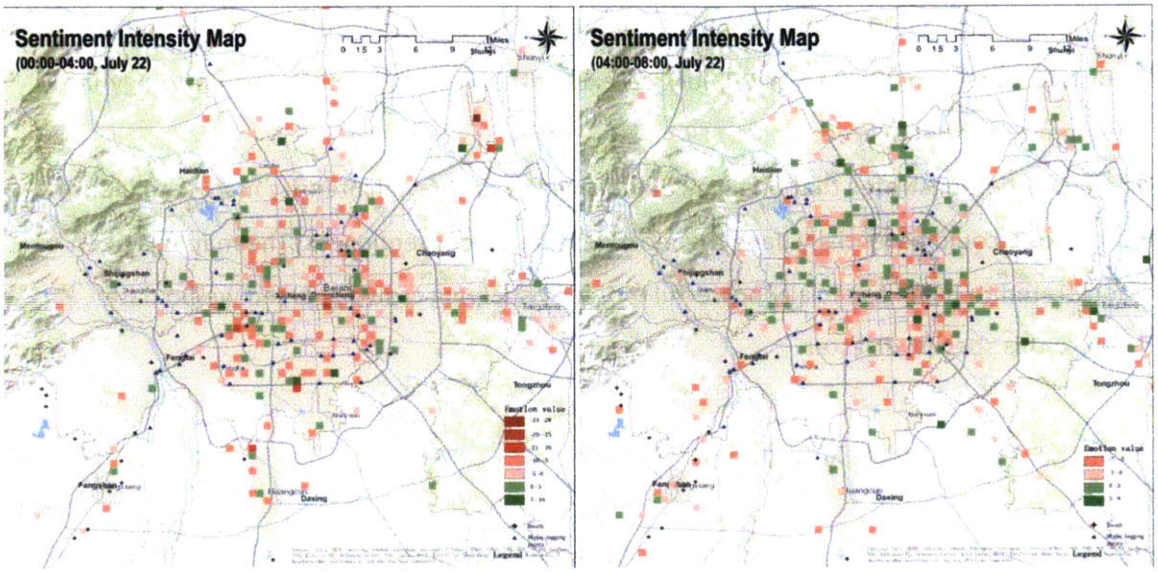


Figure 5.7-2 Sentiment intensity map

- From 00:00 to 04:00 on July 22: Beijing Municipal People's Government, Beijing-Shijiazhuang express way, Huaixin Garden neighborhood and Liuxiang community area.
- From 04:00 to 08:00 on July 22: the overall intensity decreased and some previous negative emotions changed to positive. For example, the dark red of Capital airport changed to light green in this period.

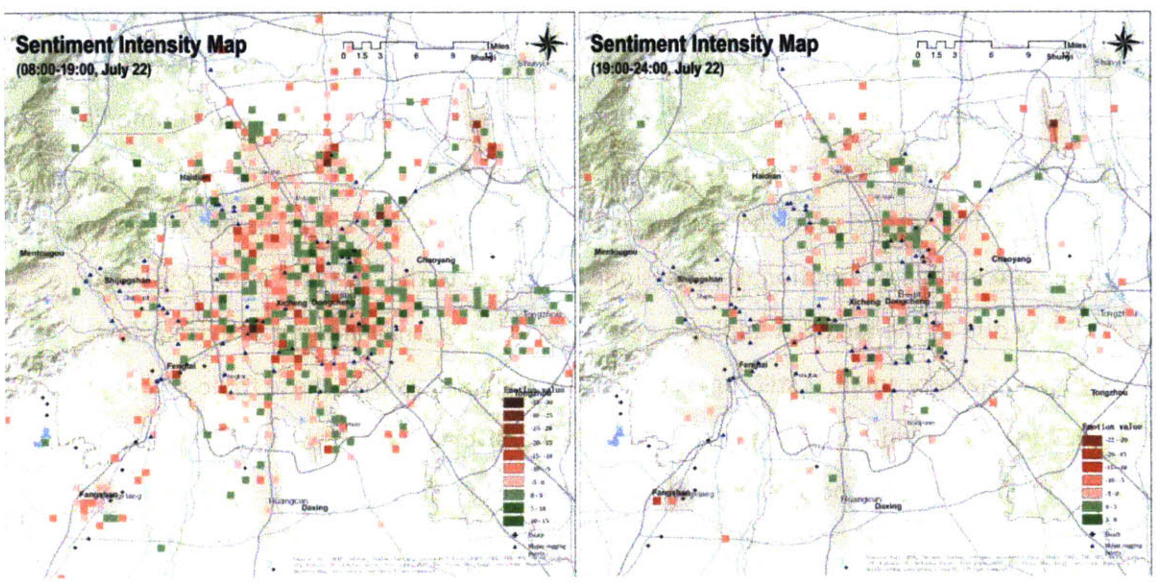


Figure 5.7-3 Sentiment intensity map

- From 08:00 to 19:00 on July 22: West of Wangjing, Tiantongyuan community area, Fudong, South of Lishui overpass, Beijing south station, Wanfang bridge and Lianhua Overpass.

- From 19:00 to 24:00 on July 22: the pattern of sentiment intensity was similar to the distribution from 08:00 to 19:00, July 22, except that negative emotion in Tiantongyuan community disappeared.

This analysis demonstrates that one way of utilizing the emotional intensity map, is to identify locations that could potentially have the most devastated consequences. The map has been demonstrated to be an especially powerful tool for visualizing crowd-sourced weibo data related to this specific disaster.

5.4 Flood Emotional Type Map

This section further examines the seven emotions identified previously, and examined how they change with respect to time and space.

5.4.1 Transitions of Emotion in Time

Figure 5.8 illustrated the percentage of each of the seven emotions at each hour. Positive emotions include happiness and appreciation and negative emotions include sadness, angry, fear, disgust, and surprise. Figure 5.9 is the result of sentiment analysis of all 38528 weibos (Ref 4.4.5). One interesting observation worth pointing out here was that the percentage of weibos with emotions was very high in this sample of weibos with geo-tags during this disaster. There were two possible contributing aspects. The first one was obvious that during and immediately after a severe disaster, there was likely a higher percentage of social media messages associated with emotions. The second one was an observation from this analysis that weibos with geo-tags had higher percentage of emotional expressions compared to the average of the total volume. This could be due to the fact that adding a geo-tag involved a small amount of extra effort. Although it was a very small, it still suggests that the weibo users that chose to do so also cared to do so. This can provide us with a way of filtering massive amount of weibo data, which could often time be massive and noisy as we have previously observed.

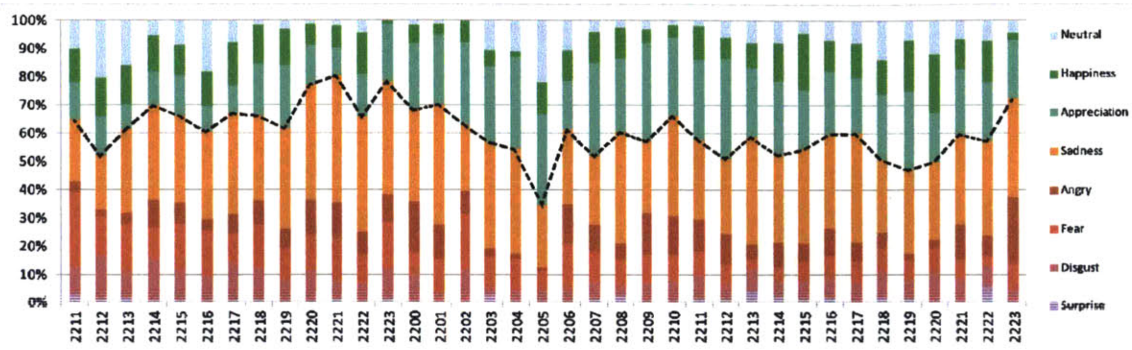


Figure 5.8 Seven types of emotions ratio in time series

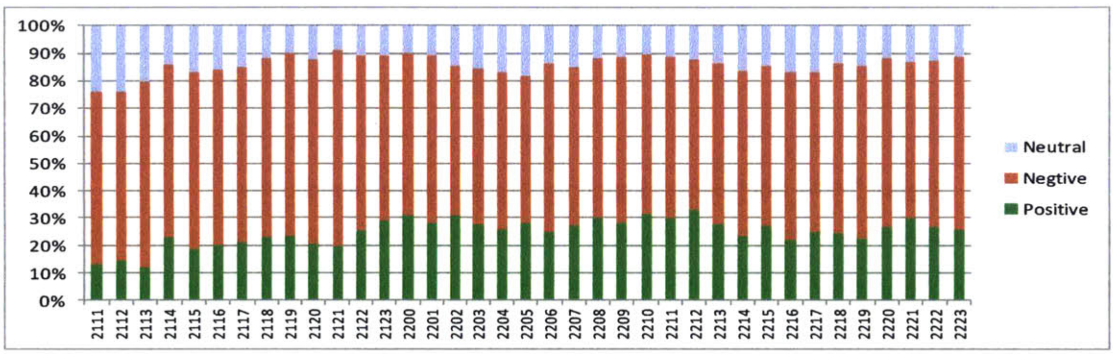


Figure 5.9 Positive and negative ratio in the sample

5.4.2 Hot Spot Map

This section visualizes the spatial distribution of each emotion type in the form of hot spot maps. More specifically, this analysis performed a kernel density estimation by calculating the density of the point features in a neighborhood around points, and producing smooth and continuous surfaces. Hot spot maps were generated to highlight areas with higher than average incidence of each type of emotion.

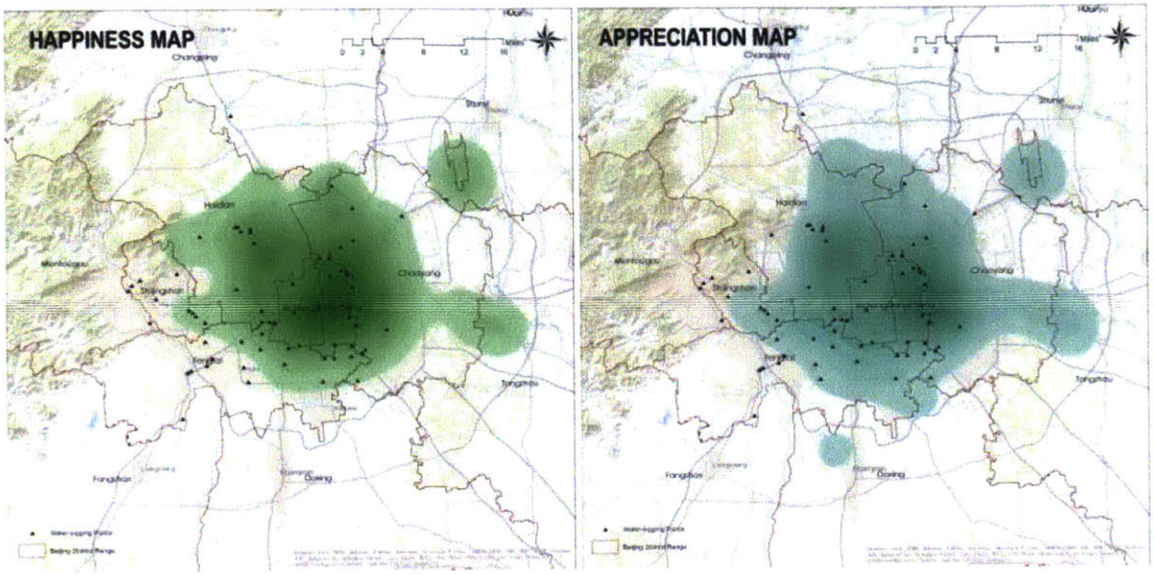


Figure 5.10-1 Sentiment hot spot map/Happiness and appreciation

The happiness map demonstrates that Dongchen was the happiest area during this flooding disaster. Part of Xicheng district including Tianmen Square and Forbidden City also expressed higher than average degree of emotions associated with happiness.

The appreciation map displayed a center on the border between Dongcheng and Chaoyang including the Worker Stadium, the Jianguomen Overpass and the Guomao Overpass. This was likely the result of a soccer game that was held in the Worker Stadium on the night of July 21, 2012.

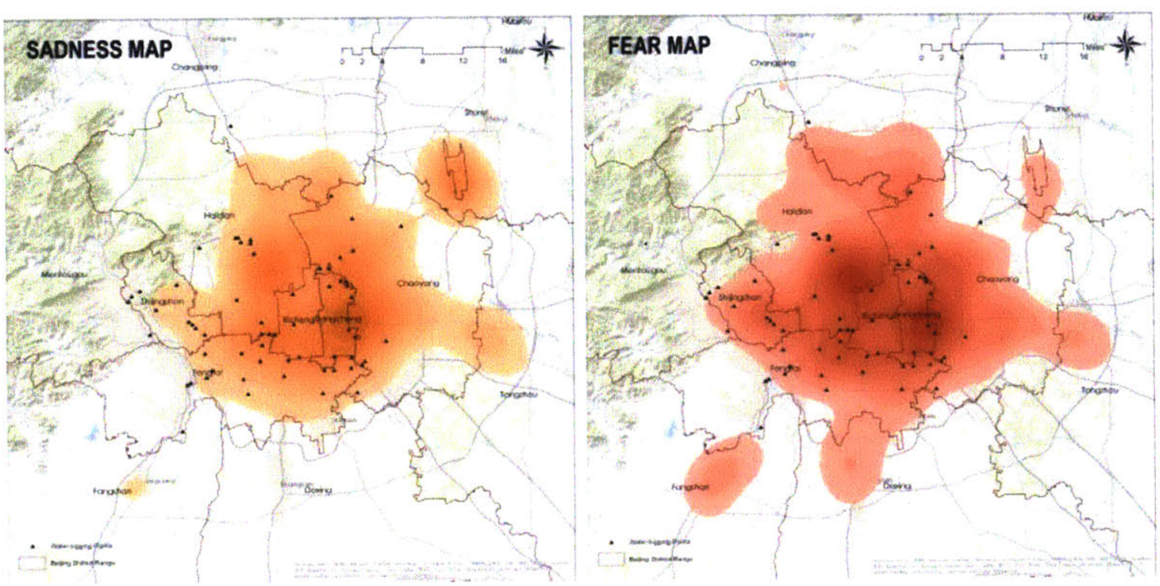


Figure 5.10-2 Sentiment hot spot map/Sadness and fear

Hot spot map of the sadness emotion showed two centers: one was on the border of Dongcheng and Chaoyang including Guangqumen overpass; the other was near Haidian Business Center where the best school districts of Beijing was located. Hot spot map of fear was geographically similar to the sadness map. It showed a clear center located in the south of Guanqumen overpass where the first victim drowned underneath.

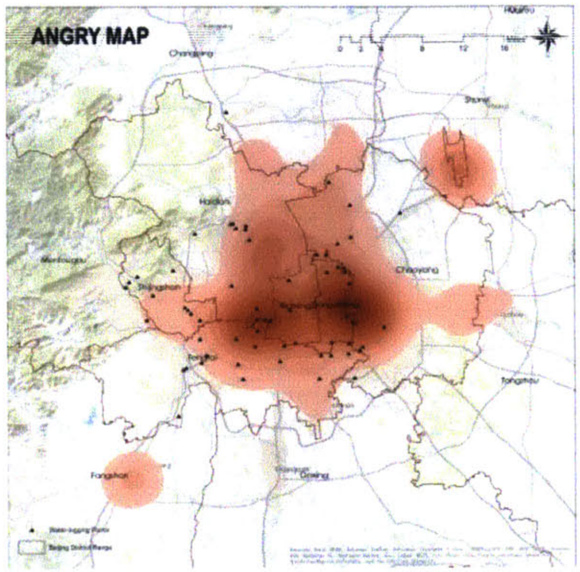


Figure 5.10-3 Sentiment hot spot map/Angry

The angry map highlighted Jianguomeng, the World Trade Center, Beijing Railway Station, Beijing West Railway Station and the Capital Airport.

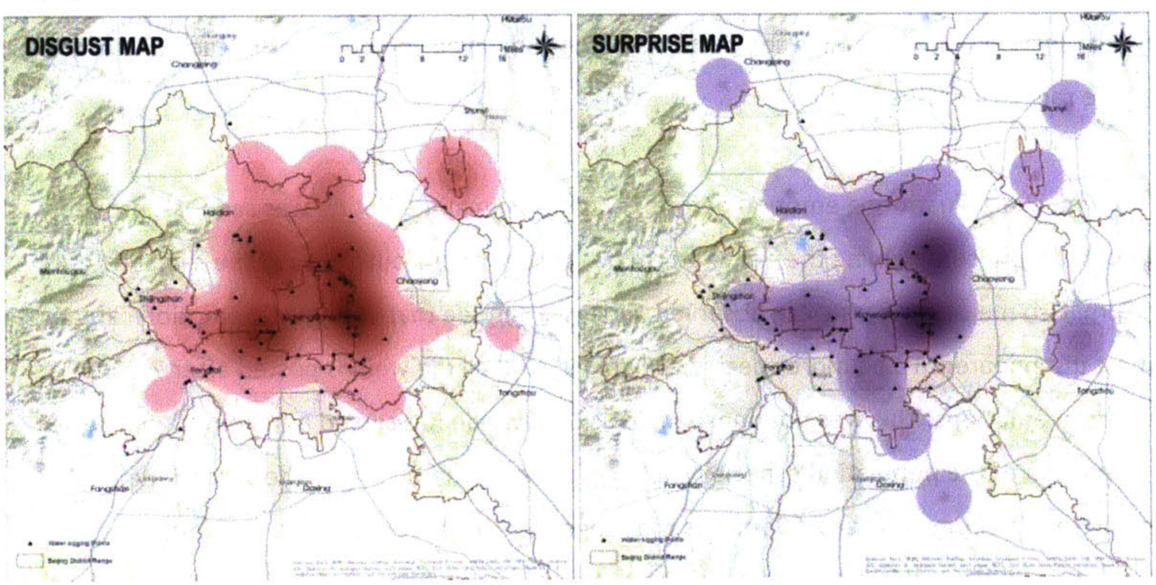


Figure 5.10-4 Sentiment hot spot map/Disgust and surprise

The disgust map showed multiple hot centers including Beijing West Railway Station, Zhongguancun Business Area, Sitongqiao Overpass and Beijing Railway Station

.

5.4.3 Emotional Type Pattern

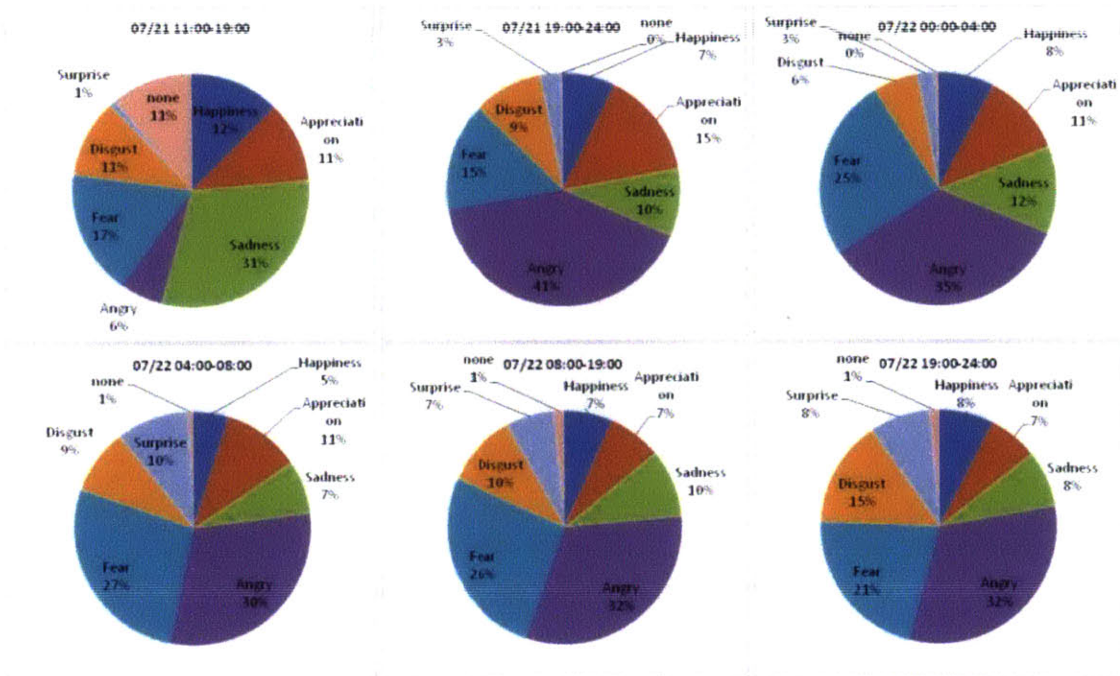


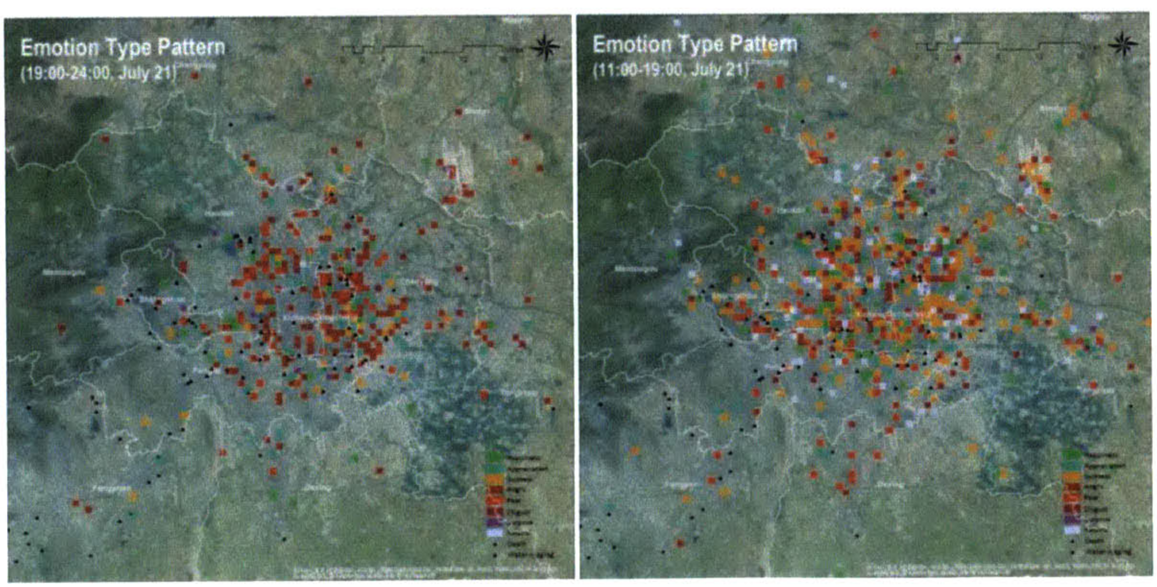
Figure 5.11 Seven types of emotions ratio in six time scope

This last section would examine the pattern in emotional type compositions in each one of the six time ranges identified earlier.

In the first chart illustrating emotion before 7:00pm, the major emotion was sadness, probably generated from complaining the heavy storm observing various outcomes of the disaster. After 7:00pm, two major changes were observed. First, we can see that the neutral emotion category almost disappeared. To be more specific, roughly 99% of all the weibos posted after 7:00pm expressed emotions. This was not unexpected as Sina-weibo has served as an emotional outlet for a portion of the population especially in large cities and urban areas. The second major shift from before 7:00pm to after 7:00pm was that anger replaced sadness to be the dominant emotion, as illustrated in the second pie chart. It remained to be one of the top emotions for the rest of the sampled period, accounting for more than 30% of the total emotion in each period of time. One possible explanation of the transition from sadness to anger was that during the earlier time period, public emotions were mostly centered on the power of natural disaster. As time continued, lack of satisfaction towards crisis management and disaster relief effort increased, and as a result angry emotions quickly surfaced during this period of time.

From July 22ndon, fear has also risen to be one of the stronger emotions, accounting for more than 30% of the total emotions for the duration of this analysis. This was likely to be correlated with that initial data summarizing the consequences of the disaster has started to be released, and reports on death caused by the disaster and coverage on the water-logging situation throughout the metropolitan area have started and would continue for the next hours and days to come.

In summary, these different ways of visualizing emotions in terms of emotional type, intensity and spatial distribution provided tools to understand the human response to devastated disasters. In particular, we could monitor the change in emotions from one district to another, and even isolate the specific locations that might need the most help using the methods presented in this work. We could analyze the shift in emotions in time, which could in turn guide us to identify the events that affect citizens' emotions the most, and understand how citizens' emotions were affected by such events. Such information and experience could provide more insights for a disaster relief team to help them understand what is most needed and what types of events are of most concern.



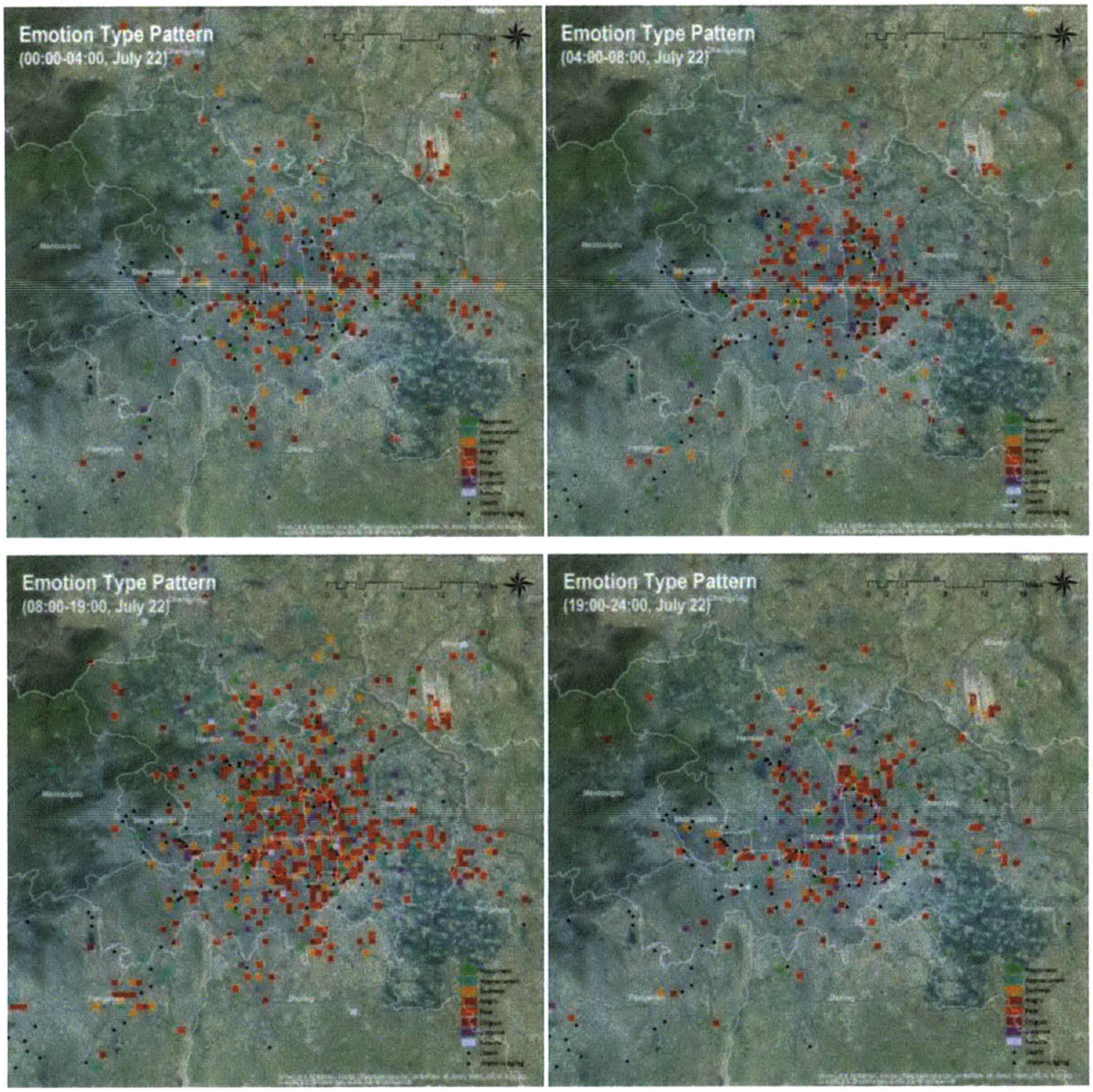


Figure 5.12 Emotion Type Pattern Maps

Chapter 06 Conclusion

This thesis focused on the heavy storm and flash flood disaster that happened in Beijing on July 21st, 2012. The thesis proposed to include data from the social media into the defense system to increase the efficiency in early warning and situation updates of disaster.

The"7.21" flood disaster revealed the deficiency in urban flood-control capability. In particular, the analysis in this work revealed two major issues in Beijing's urban planning and disaster mitigation system during this particular event. The first issue was that the alerting was too general. There was only one storm warning system covering the entire metropolitan area spanning 16,808 km^2, and the information it provided was not nearly sufficient in guiding citizens to mitigate the disaster. The second issue was that the warnings issued could be more time sensitive and more accurate. In this event, the storm warning level was blue (level I) before 9:30, most are yellow (level II) after the rain started, suggesting citizens to be aware of the situation and be prepared for potential disaster. It was not until 1am in the morning, after the storm rainfall intensity has significantly reduced to rain shower, then the red storm warning was issued. In addition to issues with the existing features of the disaster warning system, it also lacked the ability to reflect the damage caused by the storm, and could not provide information like flashflood, water-logging and traffic update which were all crucial to navigating citizens during and after the disaster. They still rely on eyewitness description or investigation to estimate consequences.

This thesis provided a framework for processing massive amount of data generated by social media, and proposed a methodology for converting the data into actionable knowledge. More specifically, this work was able to extract high-frequency terms at every time stage, and these terms reflect the important events that were occurring during that time span. For example, the frequency of the word "water-logging" rose significantly from 7pm of the day, signaling that water-logging was quickly becoming a major issue, and its multiple peaks suggested that efforts should be put into mitigating the situation. In addition, generating data from weibos provided advantages in time compared with conventional data collecting methods. For example, the death at Guangqumen overpass triggered a peak in this keyword on weibo within an hour of the incident, whereas the official channel did not release any death information until 7pm the next day, and the full list of victims was released on August 5th, fifty days after the incidents. The preliminary categorization method proposed in this work was able to filter weibos based on experience, and the categorized information could help the crisis mitigation agency to prioritize information and handle events from different categories with different methods.

This work also explored extracting emotions from the massive amount of weibos, distilling crowd-wisdom by using filters and algorithms to smooth out the noise in the massive amount of data, and determining that human emotions closely correlated with severe natural disaster. By tracking human emotions, it was possible to track the progress of the disaster, and more importantly whether relief was provided to mitigate the disaster. Moreover, by projecting the emotional polarity, intensity and emotional type onto maps, the visualization can provide reasonably clear and timely picture of when and where the strongest emotions occurred. By comparing with the actual locations where deaths and water-logging occurred during this particular event, this work demonstrated that the normalized negative emotion map was able to isolate the most affected areas, especially at areas slightly further away from the population center. This information could be extremely useful for guiding disaster relief and for allocating resources during a disaster. The emotional type map provided additional insights into understanding how human emotion shifted during the disaster, and could help us analyze the underlying causes of the emotions in each specific region for each specific demographic at each time. While the work performed in this thesis was a couple of years after the "7.21" disaster, the methods proposed could be applied real-time, thus providing real-time guidance and feedback for the disaster.

6.1 Research Results

One main contribution of this thesis is to provide a proof of concept for developing a framework to understanding and interpreting the social media data into actionable knowledge. Sentiment could be an effective measurement for situation update and disaster response.

This thesis explores multiple analytical methods to capture human responses towards urban flood disaster. Major findings of this research include four aspects: 1) weibo volume correlated strongly to disaster event in time series; 2) content analysis may provide a domain dictionary and preliminary classification for future application of this type; 3) human responses to disaster can be translated to emotion polarity which can help responders to develop stronger situation awareness; 4) emotion maps offered innovative perspectives to tracing disaster responses by streamlining data processing to update map as distinct types of sentiment maps. Next, each one of these aspects is expanded in more detail.

Trend analysis: trend correlation high

Weibo messages signaled the existence of the event in a clear and strong fashion. All of weibo volume trends by day in 2012, 2013 and 2014 demonstrated that "storm related "weibos were stimulated by the disaster event. The volume increased immediately after the disaster started, and decreased rapidly after the disaster. In 2012's trend, the peak of weibo counts reached 200,000 messages per hour. The strong signal proved that weibo can provide great insights into people's responses to certain disasters. Although the curve showed a slight delay in time, we observed that the trend matched closely to the intensity of the disaster. In terms of its fast speed in spreading information, concentration of information, and honest representation of human emotion, weibo played the indispensable role in sharing information and measuring the consequences. The trend analysis also presented that normalization techniques are effective in isolating information most pertinent to this event.

Content analysis: domain dictionary and preliminary classification

Authentication issue has always been a major challenge in crowdsourcing analysis. Historically, the best method to find the relevant information was to filter manually. In the case of Haiti, 4636 volunteers helped to extract the information. It was a labor intensive, fund intensive and time intensive process. In addition, each disaster event has its own characteristics. In this work, I proposed a domain dictionary based on word frequency. Frequent terms provided real time situation updates and could aid in directing government's attention to specific districts and urban functions that needed the most attention. After reviewing those words in the context of weibo, I categorized them into the preliminary classification I developed. The preliminary method for classifying words into different categories may automate the process of assigning disaster-related keywords and information into different priorities. In the future, responders can use this process to provide real time alerts and information to target recourses and assign needs.

Sentiment analysis: macroscopic monitoring

It has been verified in practice that social media plays an important and positive role in disasters, especially in information sharing and providing real-time information on the situation. However, there was a lack of public evidence supporting that Sina-weibo was ever used for situation awareness or disaster mitigation by organization or government in China. This work is the first attempt in applying sentiment analysis on real data from social media during a natural disaster. Although China's complexity and informal cyberspeak make it harder to perform sentiment analysis compared to the examples discussed in literature review section, sentiment analysis can still provide an effective measurement to monitor people responses to disaster on a macro level. Lessons learnt from this work could provide guidance for future analyses of this type.

Flood sentiment map: innovative method to visualize spatial pattern in time series

Four distinct sentiment maps were created: 1) sentiment map by negative ratio 2) emotional intensity map 3) emotional hot spot map 4) emotional type pattern map.

I found that sentiment map by negative ratio matches the disastrous consequences, especially in areas with less weibo volume. The emotional intensity map is a very effective method for isolating locations that resulted in the most devastated consequences. These two kinds of maps provide valuable information not only for situation awareness but also for allocating resources among different locations in Beijing. The emotional hot spot map and the emotional type pattern map can help us to develop a landscape of emotions: Where is the center of sadness? What kind of emotion was primarily expressed in the Forbidden City area during the disaster? Which areas generated the most criticism towards the city planners? It might be smart to give a priority to these areas.

6.2 Discussion and Implementation

Various parties could benefit from this real-time information.

NGOs/individuals

NGOs, academic institutions and the private sector could utilize the framework and methodology described in this thesis to develop an online platform or write a mobile application that would track disaster-related information real-time and present them through the app. Similar to the case of Ushahidi (Ref 2.1.3), an urban storm and flash flood platform could be created to publish filtered information and display disaster maps reflecting both the actual progress of the disaster itself and people's emotional responses to it. Users of the platform could get alerts automatically via email, SMS or social media within a given area near a disaster.

In the case of "petajakarta" (Ref 2.1.3), the university developed a community-led platform together with twitter to collect and disseminate information about the flooding and the critical water infrastructure in Jakarta. Users were able to tweet flooding related messages using @petajkt #banjir and then follow all related information. The methodology in this work could also support such implementation based on accessible data of Sina-weibo. Users could post weibos using @#urban flash flood# to share disaster-related information. In addition, the methodology could covert people's responses into shifts in their sentiment and present such information via various sentiment maps.

Another possible implementation would be to develop a cellphone application entirely independent from the use of Sina-weibo. Individuals could download the mobile application and use it to obtain warning on nearby locations such that they can mitigate the situation by avoiding heavily affected areas. They could also submit updates and corrections to the real-time available information through the app, trace sentiment maps and launch crowd-based self help. Either way could provide one more channel for the public to obtain situational updates in a timely and accurate fashion. Additionally, the presence of such disaster mitigation program written by NGOs could make governments more accountable to citizens.

City planners

According to the results of the content analysis, terms such as water-logging, transportation congestion and overpass etc were the hottest ones during the "7.21" disaster. These terms can help urban planners to identify the deficiencies in urban functions. Sentiment maps could further pinpoint the geographical locations of these weaknesses in the city for planners. The indicator of sentiment was very important in testing the urban drainage system and the rationality of the city planning. A database could be developed to record these variations in space and time via the volume, the emphasis and the sentiment of comments. In terms of the planning process, this work could help combining the top-down and the bottom up approaches and promoting a feed-back mechanism. It could provide a way for urban users to evaluate the effectiveness of urban planner's work. It could also provide a way for city planners to trace urban development and allocate limited resources.

Crisis response team

The framework descried in this thesis could be incorporated into the current disaster defense system in China. In addition, the data process and mapping results could help the government to assist in providing to the government a real time map, which reflects both the physical disaster and the emotion (polarity, intensity and type) in proximity to it.

Current early warning system for storms has the issue of lagging in time, and providing information that was not specific enough or even inaccurate (Ref 3.3). Beijing is a mega city and flash flood has been one of the main natural disaster challenges. This is because flash flood has the distinct characteristics of occurring frequently, developing rapidly upon occurring, and affecting a large area of citizens. Obviously, the existing weather alert was not enough. Crisis responders should be responsible for keeping the public informed on the disaster situation such as the water levels, locations with water-logging or locations of transportation congestions. Integrating the methodology in this work into current early warning system could improve the performance of early warning system. The integrated product would not only keep track of current updates but also guide resource allocation based on the location and disaster evaluation. This work would significantly improve the current crisis management system, which is an essential element in government's public images established through how it handles emergency situation affecting a massive amount of its citizens.

Other interested parties

Other organizations could utilize these techniques as well for further research. Commercial entities, such as insurance companies and architecture firms, could apply information generated by these methods based on real data toward their products, whether it would be auto insurance or water-sensitive designs.

6.3 Limitations and Future Work

Limitation

This thesis has aimed to break boundaries of various different disciplines. As one could see, most of the related work was developed in the domain of computer science. Furthermore, few city planners have experience in computer science, and as such, it is important and critical for more cooperation across different disciplines to be developed to take advantage of the research and knowledge accumulated in various disciplines that are contributing to this particular topic. Moreover, in order to obtain the optimal results, other fields besides computer science, such as statistics, linguistics, psychology, meteorology, civil engineering and geography should also be integrated into this multi-disciplinary framework. In terms of implementations, multi-sectoral cooperation would be greatly beneficial and advantageous.

The second limitation of this work is the data sample size. Sina-weibo heavily restricts the sample size of weibos one can collect. Monitoring real time data can increase the sample to some degree. In terms of historical data, the only solution to the author's knowledge is to use web scraping method to collect all weibos we can locate. Due to the limitation of sample size, the representation of the results and application of sentiment is maybe affected.

The third limitation is the lack of existing sentiment analysis and techniques for Chinese text. While there was a large collection of research into performing sentiment analysis in English, our understanding of performing sentiment analysis in Chinese is still somewhat limited. For example, Chinese characters are dramatically different from English text, and parsing the characters into information unit itself is a non-trivial job. Two popular models are Stanford Chinese word segmenter (CRF) and ICTCLAS. CRF "relies on a linear-chain conditional random filed model, which treats word segmentation as a binary decision"[7] (The Stanford Natural Language Processing Group). CEF was considered to be the best word segmenter, but still left much room for improvement. In addition, emotion lexicon is another serious problem. In both lexicon-based approach and machine learning approach, Lexicons are always the basis for evaluating the emotion. Moreover, the lexicon should fit to the results of word segmentation. On the other hand, weibo is kind of folk language, a mixture of informal writing language, cyberspeak and oral expression which make it more difficult to analyze.

---

[7] http://nlp.stanford.edu/software/segmenter.shtml

Future works

I would like to propose the following aspects for future research:

1) Update the framework to use real time data, and since flash floods occur frequently in China, monitoring social media information and data during a disaster is possible. We can increase the sample of disaster-related data and obtain more evenly distributed data within the restriction of Sina-weibo.

2) Meteorological data can help to show the relation between weibo and weather.

3) There are some additional ways to increase the sample size. By using a different query combined with machine learning, one can identify more weibos related to key phrases like "storm" and "flash flood". In this thesis, I only used one keyword. Interestingly, there are many weibos, which respond to the disaster, but are not included in the word of "storm".

4) Using lexicon-based combined machine learning approaches to perform the sentiment analysis, this combined method can increase the accuracy and relevance of the analysis. A control sample can be prepared to test the overall emotion trend of Sina-weibo. We can have several training sets, where weibos are selected randomly and labeled by different people.

5) We can get more weibos with geo-tags in proportion to the increased sample size, then increase cell size and time interval to get emotion spatial pattern each time.

Bibliography

1. Abhas K Jha, Robin Bloch, Jessica Lamond. Cities and Flooding: A Guide to Integrated Urban Flood Risk Management for the 21st Century. 2012. International Bank for Reconstruction and Development / International Development Association.
2. Anderson, K., & Schram, A. (2011). Design and Implementation of a Data Analytics Infrastructure In Support of Crisis Informatics Research. ICSE 2011, 21-28 May 2011, Honolulu, Hawaii.
3. Approach for Detecting Crisis Related Micro-posts. In: ISCRAM 2013.
4. Brandon Truong, Cornelia Caragea, Anna Squicciarini, Andrea H. Tapia. "Extracting Valuable Information from Twitter during Natural Disasters." In: The 2014 Annual Meeting of the Association for Information Science and Technology (ASIS&T 2014), Seattle, WA, USA. [pdf]
5. Bailey, N. T. J. The Mathematical Theory of Infectious Diseases and its Applications. Griffin, London, 2nd edition, (1975).
6. Bernardo A. Huberman, Daniel M. Romero, & Fang Wu. Social networks that matter: Twitter under the microscope. First Monday, 14, 1, (2009).
7. Blanford, J. SensePlace2: Geo-twitter Analytics Support for Situation Awareness, in: IEEE Conference
8. Boyd, D., Golder, S. & Lotan, G, Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter.Proc. HICSS, (2010).
9. Brooks, M., Kuksenok, K., Torkildson, M., Perry, D., Robinson, J., Scott, T., Anicello, O., Zukowski, O., Harris, P., Aragon, C. (2013). Statistical Affect Detection in Collaborative Chat. Proceedings of CSCW 2013, 317-328.
10. Congying Li, 2012. Eco-hydrology and good urban design for urban storm water-logging in Beijing, China. Ecohydrology & Hydrobiology, Volume 12, Issue 4, 2012, Pages 287–300
11. Cornelia Caragea1, Anna Squicciarini2, Sam Stehle2, Kishore Neppalli1, Andrea Tapia2. Mapping Moods: Geo-Mapped Sentiment Analysis During Hurricane Sandy. In: Proceedings of the 11th International Conference on Information Systems for Crisis Response and Management (ISCRAM 2014), University Park, Pennsylvania, USA, 2014.
12. Giles, L., Jansen, B., Yen, J. (2011). Classifying Text Messages for the Haiti Earthquake. In: ISCRAM 2011

13. Castillo, C., Mendoza, M., & Poblete, B. (2011). Information Credibility on Twitter. In: WWW '11 (pp.675–684). ACM. doi:10.1145/1963405.1963500
14. Cha, M., Haddadi, H., Benevenuto, F., & Gummadi, K.P. (2010). Measuring User Influence in Twitter: The MillonFollower Fallacy. Proc. ICWSM, (2010).
15. Collier, N., Doan, S.: Syndromic Classification of Twitter Messages. In: Kostkova, P.,Szomszor, M., Fowler, D. (eds.) eHealth 2011. LNICST, vol. 91, pp. 186–195. Springer,
16. Communications in Disaster Response. In: The International ISCRAM Conference 2012 (pp. 2–11).
17. Computer-Mediated Communication in 2007 Virginia Tech Event. Social Science Computer Review.
18. Microblogged Data Fit the Information Needs of Disaster Response and Humanitarian Relief
19. Data Use by Disaster Response and Humanitarian Relief Organizations", In: ISCRAM 2013.
20. Detection by Social Sensors. WWW 2010, April 26-30 (pp. 851–860). Raleigh, North Carolina: ACM.
21. Emergencies. Disasters. pp. 1–8. UN Foundation & Vodafone Foundation Technology Partnership.
22. Events. International Journal of Emergency Management, 6(3/4), 248–260.
23. Glaser, M. California Wildfire Coverage by Local Media, Blogs, Twitter, Maps and More. PBS MediaShift.http://www.pbs.org/mediashift/2007/10/the_listcalifornia_wildfire_co_1.html
24. Goldenberg, J., Libai, B., & Muller, E. Talk of the network: A complex systems look at the underlying process of word-of-mouth. Marketing Letters, 12, 3, (2001), 211–223
25. Gao, K., Zhou, E.-L., & Grover, S. (2014). Applied methods and techniques for modeling and control on micro-blog data crawler. Applied Methods and Techniques for Mechatronic Systems, 452, 171-188.
26. He, H. (2013). Sentiment analysis of Sina Weibo based on semantic sentiment space model. In Proceedings of management science and engineering (pp. 206-211).IEEE.
27. Huang, S., Peng, W., Li, J., & Lee, D. (2013). Sentiment and topic analysis on social media: a multi-task multi-label classification approach. In Proceedings of the fifth annual ACM web science conference (WebSci' 13) (pp. 172-181). ACM.
28. Hu, X., Tang, L., Tang, J., & Liu, H. (2013). Exploiting social relations for sentiment analysis in microblogging. In Proceedings of the sixth ACM international conference on web search and data mining (pp. 537-546). ACM.

29. Hagar, C. & Haythornthwaite, C. Crisis, Farming &Community. The Journal of Community Informatics, 1, 3,(2005), 41-52.
30. Honeycutt, C., & Herring, S. C., Beyond microblogging: Conversation and collaboration via Twitter. Proc. HICSS,(2009).
31. Hughes, A. L., & Palen, L. (2009). Twitter Adoption and Use in Mass Convergence and Emergency Events. International Journal of Emergency Management, 6(3/4), 248–260.
32. Hughes, A. L., & Palen, L., Twitter Adoption and Use in Mass Convergence and Emergency Events. Proc. ISCRAM, (2009).
33. Hui, C., Tyshchuk, Y., Wallace, W., Goldberg, M., & Magdon-Ismail, M. (2012). Information cascades in social media in response to a crisis: a preliminary model and a case study. In: WWW (pp. 653–656).
34. Jaeger, P. T., Langa, L. A., McClure, C. R. & Bertot, J.C., The 2004 and 2005 Gulf Coast Hurricanes: Evolving Roles and Lessons Learned for Public Libraries in Disaster Preparedness and Community Services. Public Library Quarterly, 25, 3/4, (2007), 199-214.
35. Jansen, B., Yen, J. (2011). Classifying Text Messages for the Haiti Earthquake. In: ISCRAM 2011.
36. Java, A., Song, X., Finin, T., & Tseng. B. Why we twitter: understanding microblogging usage and communities. Proc. WebKDD/SNAKDD, (2007).
37. Jiahao Chen, Jianhua Wu, Sentiment Analysis of Chinese Micro Blog using Machine Learning and an Improved Feature Selection Method. Mechanics and Materials Vols. 631-632 (2014) pp 1219-1223.
38. Workshop on Social Media Analytics SOMA 10 (pp. 71–79).
39. Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media? In: World Wide Web (Vol. 112, pp. 591–600). ACM. doi:10.1145/1772690.1772751 8. Liu, S.B. & Palen,L. The new cartographers: crisis map mashups and the emergence of eogeographic practice, Cartography and Geographic Information Science, 37 (2010) 69-90.
40. Lerman, K.; Ghosh, R., Information Contagion: An Empirical Study of the Spread of News on Digg and Twitter Social Networks. Proc. ICWSM, (2010).
41. Leskovec, J.; Adamic, L., & Huberman, B. 2006. The dynamics of viral marketing. Proc. EC, (2006), 228–237.
42. Li, Y., Li, X., Li, F., & Zhang, X. (2014). A lexicon-based multi-class semantic orientation analysis for microblogs. Web Technologies and Applications, 8709, 81-92.

43. Liben-Nowell, D., & Kleinberg, J., Tracing information flow on a global scale using internet chain-letter data. PNAS, 105, 12, 4633–4638. (2008)
44. Liu, S.B. & Palen, L. The new cartographers: crisis map mashups and the emergence of neo-geographic
45. MacEachren, A.M., Jaiswal, A., Robinson, A.C., Pezanowski, S., Savelyev, A., Mitra, P., Zhang, X., Blanford, J. SensePlace2: Geo-twitter Analytics Support for Situation Awareness, in: IEEE Conference on Visual Analytics Science and Technology, Providence, RI, 2011.
46. Mandel, B., Culotta, A., Boulahanis, J., Stark, D., Lewis, B., and Rodrigue, J. 2012. A demographic
47. Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, Manfred Stede.2011. Lexicon-BasedMethods for Sentiment Analysis. Computational Linguistics 37 (2): 267-307.
48. McClendon, S., & Robinson, A. C. (2012). Leveraging Geospatially-Oriented Social Media Communications in Disaster Response. In: The International ISCRAM Conference 2012 (pp. 2–11).
49. McDonald, R., Hannan, K., Neylon, T., Wells, M., and Reynar, J. Structured models for fine-to-coarse sentiment analysis. In ACL '07, volume 45, page 432, 2007.
50. Mendoza, M., Poblete, B., & Castillo, C. (2010). Twitter under crisis. Proceedings of the First
51. Microblogged Data Fit the Information Needs of Disaster Response and Humanitarian Relief
52. Moore, C., & Newman, M. E. J., Epidemics &percolation in small-world networks. Physical Review E,61, (2000), 5678–5682.
53. Pang, B., Lee, L. and Vaithyanathan, S. Thumbs up?: sentiment classification using machine learning techniques. In ACL-02, EMNLP '02, pages 79–86, Stroudsburg, PA, USA, 2002.
54. Pang Bo, Lee Lillian and Vaithyanathan Shivakumar. Thumbs up? Sentiment classification using machine learning techniques. In Proceedings of EMNLP, 2002. 79-86.
55. Portsea, L. J. (2011). Disaster Relief 2.0: The Future of Information Sharing in Humanitarian
56. Qu, Y., P.F. Wu, & X. Wang. Online Community Response to Major Disaster: A Study of Tianya Forum inthe 2008 Sichuan Earthquake. Proc. HICCS, (2009).
57. Regional Analysis of User Interactions on Social Media in Times of Disaster. In: WWW 2013, Brazil.
58. Reveals About the Social Life of Microblogged Information. In: CSCW '10 (pp. 241–250). New York.

59. Roberts, K., Roach, M., Johnson, J., Guthrie, J., & Harabagiu, S. (2012). Empatweet: Annotating and detecting emotions on Twitter. Proceedings of the LREC, 3806-3813.
60. Romero, D. M., Galuba, W., Asur, S. & Huberman, B.A. (2010). Influence and Passivity in Social Media.http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1653135
61. Sakaki, T., Okazaki, M., & Matsuo, Y. (2010). Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors. WWW 2010, April 26-30 (pp. 851–860). Raleigh, North Carolina: ACM.
62. Schulz, A., Thanh, T., Paulheim, H., Schweizer, I. (2013). A Fine-Grained Sentiment Analysis Approach for Detecting Crisis Related Micro-posts. Proceedings of the 10th International ISCRAM Conference, 846-851.
63. Shneiderman, B. & Preece, J. 911.gov: Community Response Grids. Science, 315, (2007), 944.
64. Son Doan, Bao-Khanh Ho Vo, and Nigel Collier.(2012) An Analysis of Twitter Messages in the 2011 Tohoku Earthquake.
65. Starbird, K. Digital Volunteerism During Disaster: Crowdsourcing Information Processing, in: CHI '11
66. Starbird, K., Munzy and Palen, L. (2012) Learning from the Crowd: Collaborative Filtering Techniques for Identifying On-The-Ground Twitters during Mass Disruptions." In: ISCRAM
67. Starbird, K., Palen, L., Hughes, A. L., & Vieweg, S.,Chatter on The Red: What Hazards Threat Reveals aboutthe Social Life of Microblogged Information. Proc. CSCW (2010).
68. Social Media use During Disasters: A review of the knowledge base and gaps. National Consortium for the Study of Terrorism and Responses to Terrorism A Department of Homeland Security Science and Technology Center of Excellence Based at the University of Maryland.2011.
69. Suh, B., Hong, L., Pirolli, P., & Chi, Ed H. "Want tobe Retweeted? Large Scale Analytics on Factors ImpactingRetweet in Twitter Network", Proc. SocialCom (2010).
70. Support of Crisis Informatics Research. ICSE 2011, 21-28 May 2011, Honolulu, Hawaii.
71. Sutton, J., Palen, L., & Shklovski, I. Backchannels on the Front Lines: Emergent Use of Social Media in the 2007Southern California Fires. Proc. ISCRAM, (2008).
72. Tapia, A. Moore, K. Johnson, N. "Beyond the Trustworthy Tweet: A Deeper Understanding of Microblogged

73. Torrey, C., Burke, M., Lee, M., Dey, A., Fussell, S. &Kiesler, S. Connected Giving: Ordinary People Coordinating Disaster Relief on the Internet. Proc. HICSS,(2007).

74. Tweets: The Fukushima Disaster and Information Source Credibility on Twitter. In: ISCRAM 2012.

75. Vieweg, S., Hughes, A. L., Starbird, K., & Palen, L., Microblogging during two natural hazards events: what twitter may contribute to situational awareness. Proc. CHI,(2010), 1079-1088.

76. Visual Analytics Science and Technology, Providence, RI, 2011.

77. Walton, R., Mays, R., & Haselkorn, M. (2011). Defining " Fast ": Factors Affecting the Experience of Speed in Humanitarian Logistics. In: ISCRAM (pp. 1–10). Lisbon, Portugal.

78. Walton, R., Mays, R., & Haselkorn, M. (2011). Defining " Fast ": Factors Affecting the Experience of Speed in Humanitarian Logistics. In: ISCRAM (pp. 1–10). Lisbon, Portugal.

79. Workshop on Crowdsourcing and Human Computation, Vancouver, BC, 2011.

80. World wide web (Vol. 112, pp. 591–600). ACM. doi:10.1145/1772690.1772751

81. Wu, F.; Huberman, B.; Adamic, L.; & Tyler, J., Information flow in social groups. Physica A, 337, (2004),327-335

82. Yang, J. & Counts, S. (2010). Predicting the Speed, Scale, and Range of Information Diffusion in Twitter. Proc.ICWSM, (2010).

83. Yi, S. K. M., Steyvers, M., Lee, M. D. and Dry, M. J. (April 2012). "The Wisdom of the Crowd in Combinatorial Problems". Cognitive Science 36 (3). doi:10.1111/j.1551-6709.2011.01223.x.

84. Young, H. P., the Diffusion of Innovations in Social Networks. In The Economy as an Evolving Complex System, Vol. III, ed. Lawrence E. Blume & Steven N.Durlauf, (2003), 267–82. Oxford: Oxford University Press.

85. Yan Qu, Chen Huang, Pengyi Zhang,2011. Microblogging after a Major Disaster in China: A Case Study of the 2010 Yushu Earthquake.In: CSCW 2011, PP.25-34, ACM(2011)

86. Xiaodong Chen, 2012. 基于情感词典的中文微博情感倾向分析研究 [Research on Sentiment Dictionary based Emotional Tendency Analysis of Chinese MicroBlog]. Huazhong University of Science and Technology. Master Thesis [In Chinese].

87. Chen, X. 2012. 北京"7.21" 和深圳"6.13" 暴雨内涝成因对比与分析 [Cause comparison and analysis of water logging in Beijing "7.21" and Shenzhen "6.13"rainstorm events]. Water Resources Development Research 13, 39-43 [In Chinese].

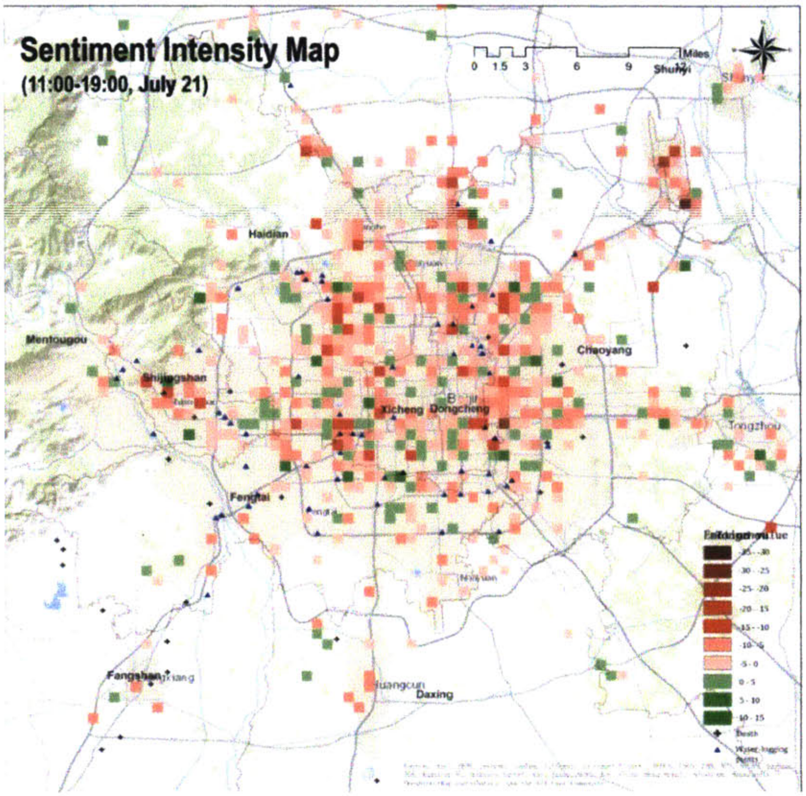
88. Chen, J., Wang, H., Zhang, S., Su, D., Zhao, F., Kong, G.2011. 北京市道路积滞水原因及案例

分析 [Causal analysis and case study of the road ponding in Beijing City]. Water & Wastewater Engineering 37, 37-40 [In Chinese].

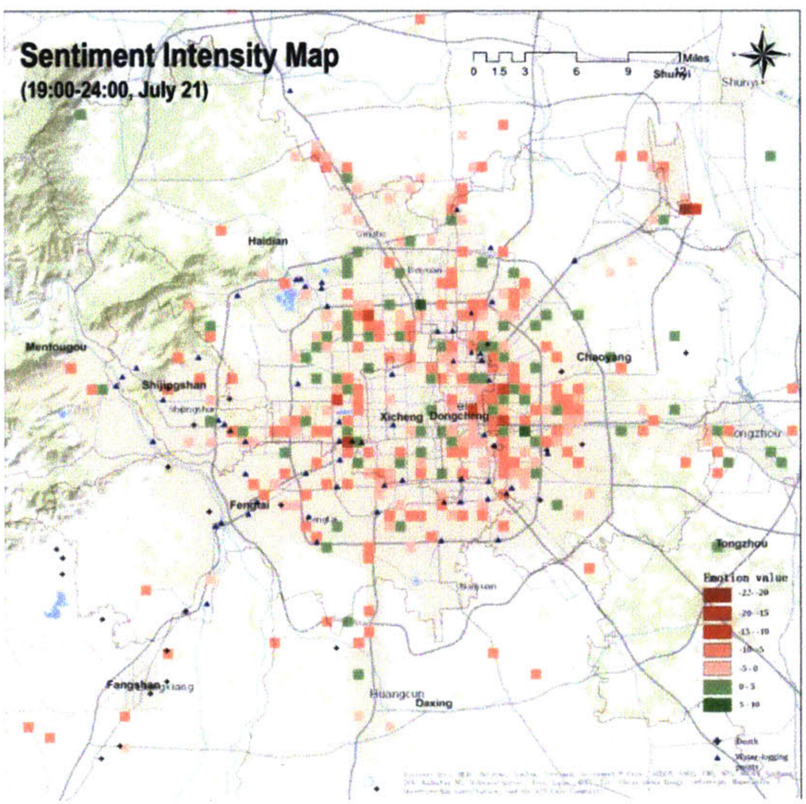
89. Nan Liu. 2013. 面向微博短文本的情感分析研究 [The Research of Microblogging Short Text Oriented Sentiment Analysis]. Wuhan University. Doctor Dissertation [In Chinese].

90. Weiguo Wan.2014. 微博用户行为分析和网络结构演化的研究 [Research on User Behavior Analysis and Network Evolution in Microblogging Networks]. Beijing Jiaotong University. Doctor Dissertation [In Chinese].

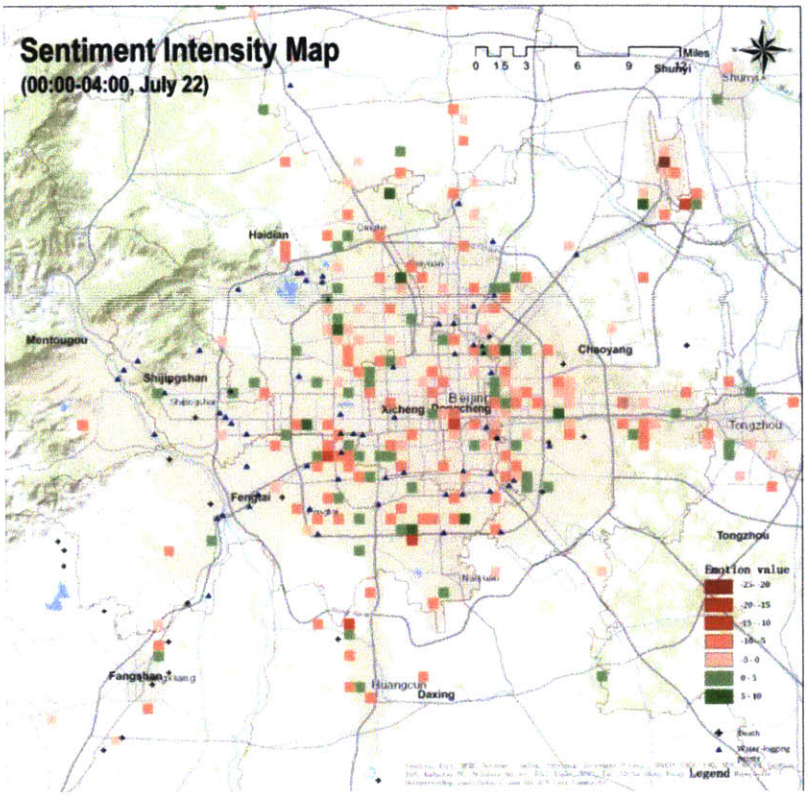
Appendix



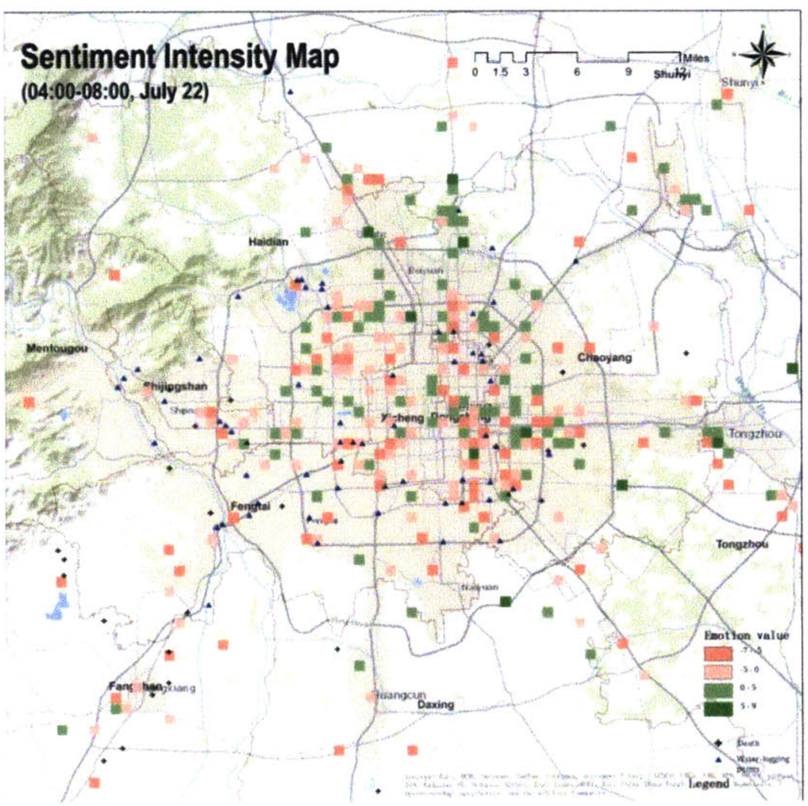
Sentiment Intensity Map
(11:00-19:00, July 21)
0 1.5 3 6 9 12
Miles
Shunyi
Haidian
Mentougou
Shijingshan
Chaoyang
Xicheng
Dongcheng
Tongzhou
Fengtai
Fangshan
Huangcun
Daxing
-35 - -30
-30 - -25
-25 - -20
-20 - -15
-15 - -10
-10 - -5
-5 - 0
0 - 5
5 - 10
10 - 15
Death
Water-logging Points



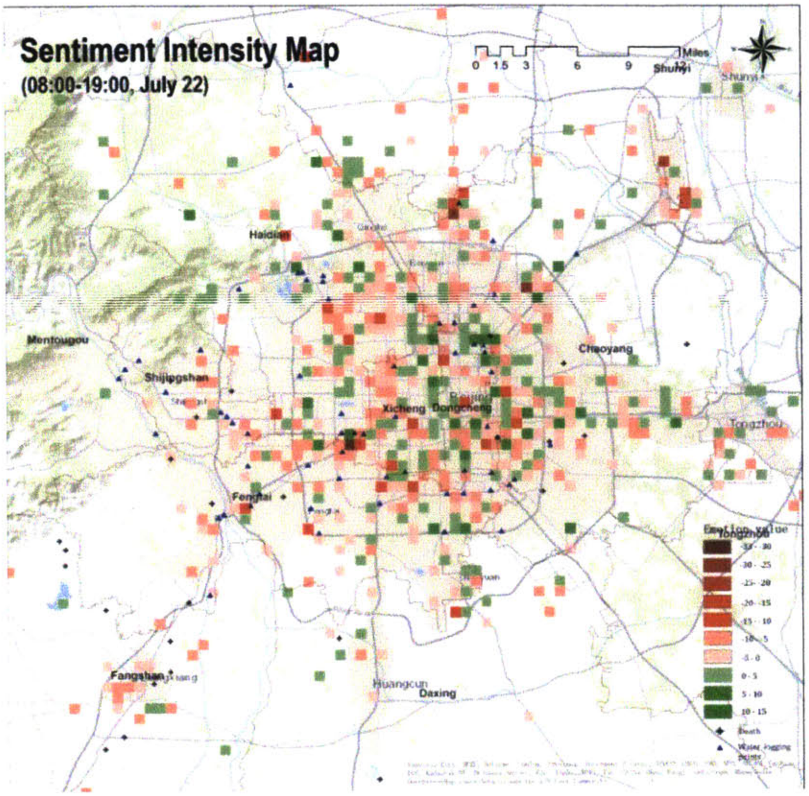
Sentiment Intensity Map
(19:00-24:00, July 21)
0 1.5 3 6 9 12
Miles
Shunyi
Haidian
Mentougou
Shijingshan
Chaoyang
Xicheng
Dongcheng
Fengtai
Tongzhou
Emotion value
-25 - -20
-20 - -15
-15 - -10
-10 - -5
-5 - 0
0 - 5
5 - 10
Death
Water-logging Points
Fangshan
Huangcun
Daxing

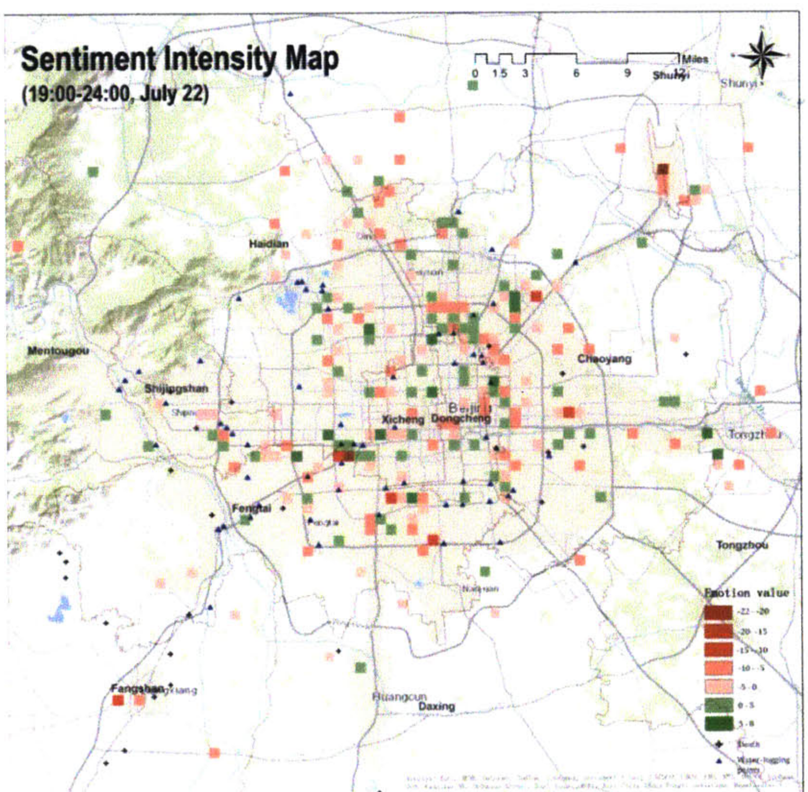


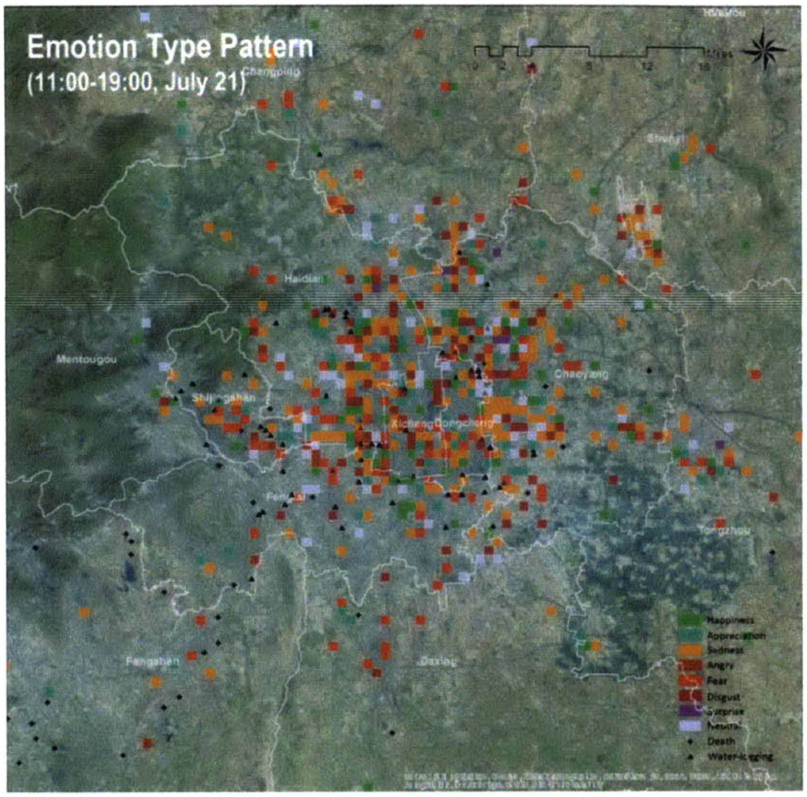
Sentiment Intensity Map
(00:00-04:00, July 22)
Haidian
Mentougou
Shijingshan
Chaoyang
Xicheng
Fengtai
Tongzhou
Fangshan
Daxing
Emotion value
Death
Legend

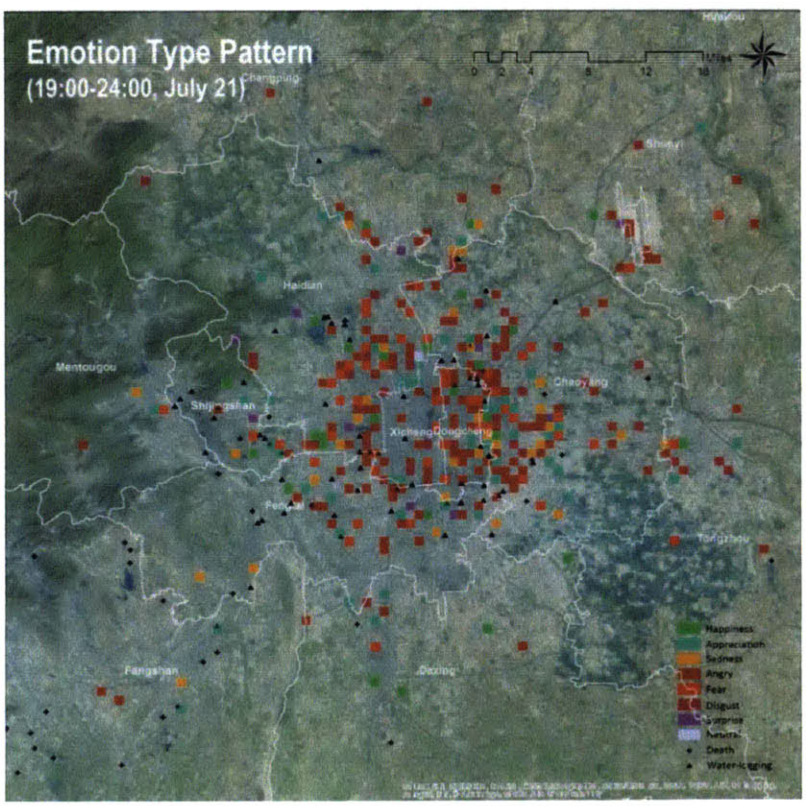


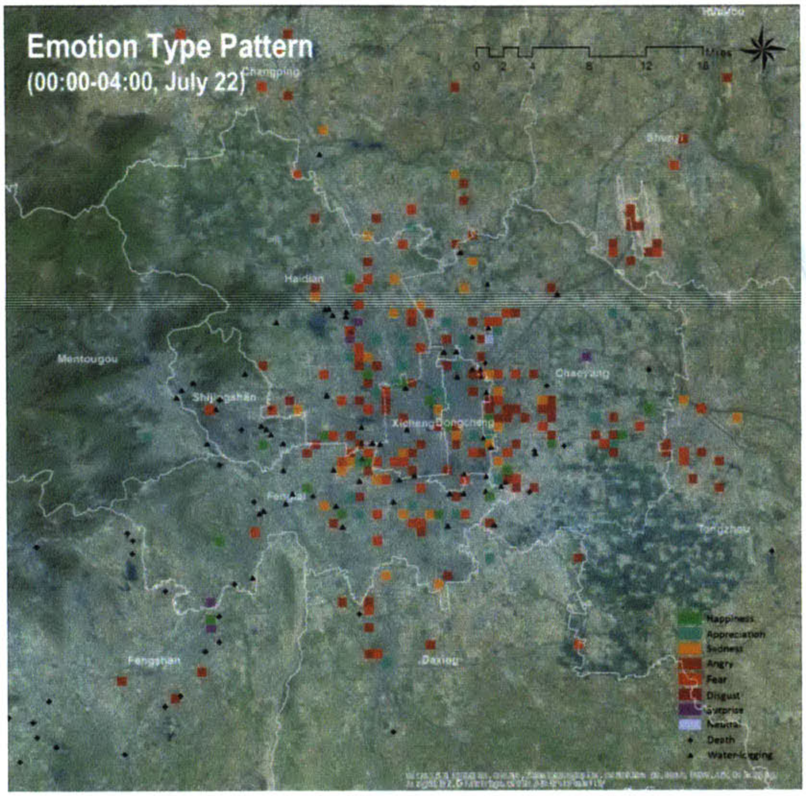
Sentiment Intensity Map
(04:00-08:00, July 22)
Haidian
Mentougou
Chaoyang
Fengtai
Tongzhou
Daxing
Emotion value
Death
Legend

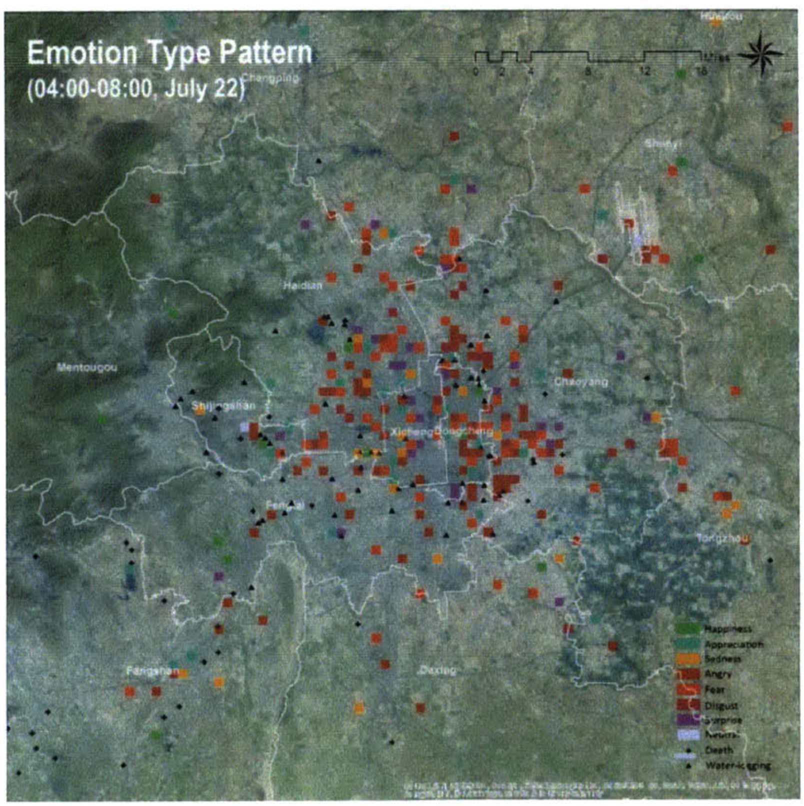

Sentiment Intensity Map
(08:00-19:00, July 22)
Miles
Haidian
Mentougou
Shijingshan
Chaoyang
Xicheng
Dongcheng
Fengtai
Tongzhou
Fangshan
Daxing
Emotion value
Death
Water-logging points

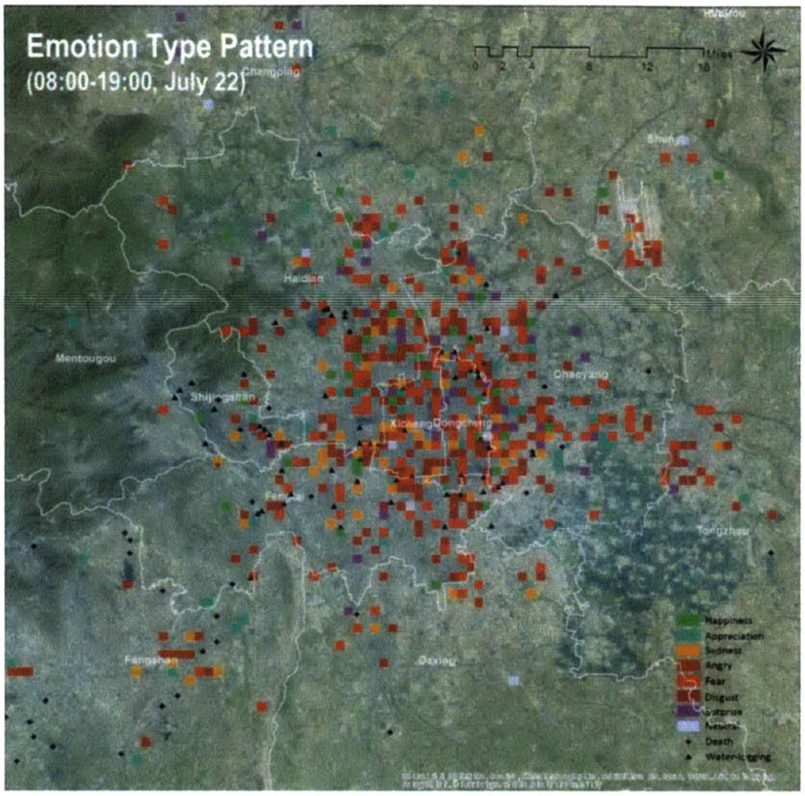

Sentiment Intensity Map
(19:00-24:00, July 22)
Miles
Haidian
Mentougou
Shijingshan
Chaoyang
Xicheng
Dongcheng
Fengtai
Tongzhou
Fangshan
Daxing
Emotion value
Death
Water-logging points

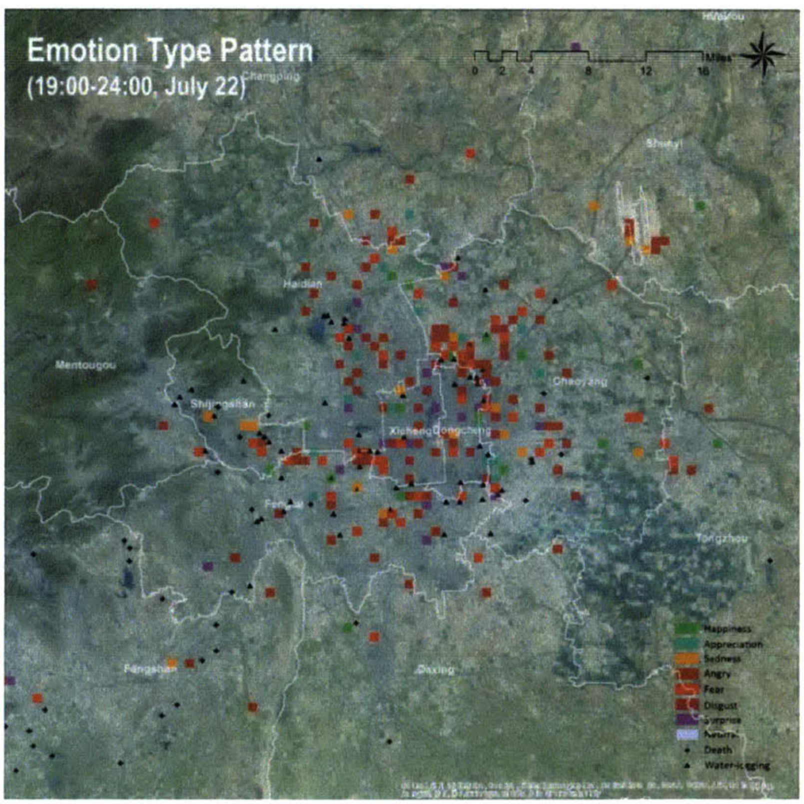

Emotion Type Pattern
(11:00-19:00, July 21)
Changping
Haidian
Mentougou
Shijingshan
Chaoyang
Xicheng
Dongcheng
Fengtai
Fangshan
Daxing
Tongzhou
Shunyi
Huairou
Happiness
Appreciation
Sadness
Angry
Fear
Disgust
Surprise
Neutral
Death
Water-logging


Emotion Type Pattern
(19:00-24:00, July 21)
Changping
Haidian
Mentougou
Shijingshan
Chaoyang
Xicheng
Dongcheng
Fengtai
Fangshan
Daxing
Tongzhou
Shunyi
Huairou
Happiness
Appreciation
Sadness
Angry
Fear
Disgust
Surprise
Neutral
Death
Water-logging


Emotion Type Pattern
(00:00-04:00, July 22)
Haidian
Mentougou
Shijingshan
Chaoyang
Xicheng
Dongcheng
Fengtai
Daxing
Tongzhou
Fangshan
Miles


Emotion Type Pattern
(04:00-08:00, July 22)
Haidian
Mentougou
Shijingshan
Chaoyang
Xicheng
Dongcheng
Fengtai
Daxing
Tongzhou
Fangshan
Miles


Emotion Type Pattern
(08:00-19:00, July 22)
Miles
0 2 4 8 12 16
Haidian
Mentougou
Shijingshan
Chaoyang
Xicheng
Dongcheng
Tongzhou
Fengtai
Fangshan
Daxing
Happiness
Appreciation
Sadness
Angry
Fear
Disgust
Surprise
Neutral
Death
Water-logging


Emotion Type Pattern
(19:00-24:00, July 22)
Miles
0 2 4 8 12 16
Haidian
Mentougou
Shijingshan
Chaoyang
Xicheng
Dongcheng
Tongzhou
Fangshan
Daxing
Happiness
Appreciation
Sadness
Angry
Fear
Disgust
Surprise
Neutral
Death
Water-logging