Social Media: New Perspectives to Improve Remote Sensing for Emergency Response

This paper provides a detailed overview of strategies for the integration of social media and remote sensing data in time-critical applications. Several practical case studies and examples are presented in the context of applications focused on emergency response.

By Jun Li, Senior Member IEEE, Zhi He, Javier Plaza, Senior Member IEEE, Shutao Li, Senior Member IEEE, Jinfen Chen, Henglin Wu, Yandong Wang, and Yu Liu

ABSTRACT | Remote sensing is a powerful technology for Earth observation (EO), and it plays an essential role in many applications, including environmental monitoring, precision agriculture, resource managing, urban characterization, disaster and emergency response, etc. However, due to limitations in the spectral, spatial, and temporal resolution of EO sensors, there are many situations in which remote sensing data cannot be fully exploited, particularly in the context of emergency response (i.e., applications in which real/near-real-time response is needed). Recently, with the rapid development and availability of social media data, new opportunities have become available to complement and fill the gaps in remote sensing data for emergency response. In this paper, we provide an overview on the integration of social media and remote sensing in time-critical applications. First, we revisit the most recent advances in the integration of social media and remote sensing data. Then, we describe several practical case studies and examples addressing the use of social media data to improve remote sensing data and/or techniques for emergency response.

KEYWORDS | Deep learning; emergency response; remote sensing; social media

I. INTRODUCTION

Remote sensing [1], [2] involves the use of systems and algorithms to record information about the surface of the Earth from a remote location. Although reliable as a data source, remote sensing data may not be always available and can be complemented by other sources of data, such as geographic information systems (GIS) or social media [3], in order to address time-critical applications. For instance, relating publicly available social media information with remote sensing or GIS data can lead to a more efficient management of emergency response (which, in our context, refers to applications in which real/near-real-time response is needed). Social media [4]–[6] (generated from many individuals) is now playing a more relevant role in our daily lives and provides a unique opportunity to gain valuable insight on information flow and social networking within the society. Through data collection and analysis of its content, it supports a more detailed mapping and understanding of the evolving human landscape. As a result, there has been a growing interest in using social media to complement the information available from remote sensing and GIS systems.
In the following, we provide an overview of available strategies for emergency response grouped by the main source of information used in the process. First, we focus on techniques that mainly use remote sensing data for this purpose. Then, we describe techniques that are based mainly on social media data. Next, we describe strategies that are mainly based on GIS data. Finally, we provide a description of techniques that exploit both remote sensing and social media data in synergistic fashion.

A. Emergency Response Using Remote Sensing Data

There have been many techniques mainly based on remote sensing data (but also exploiting other sources of information) for emergency response. For instance, in [7], a new framework is developed to harvest the ambient geospatial information from social media data, leading to improved support situational awareness as related to human activities. In [8], a single-band density slicing technique and the maximum-likelihood (ML) algorithm are used to detect water bodies from remote sensing images acquired by the Landsat multispectral instrument. The results exhibit a significant accuracy in the characterization task. In addition to optical remote sensing images, radar images have also been used for emergency response purposes. For instance, in [9], an automatic and near-real-time flood water level extraction method is developed for the analysis of high-resolution synthetic aperture radar (SAR) images. This algorithm is shown to be successful in the task of improving estimates of an hydraulic model parameter, taking assimilation into consideration. Another effort based on SAR images is presented in [10], in which coarse resolution SAR images are analyzed in near real time for timely flood management. The experiments reported show that flood inundation models can be verified and recalibrated in a time shorter than the flood travel time. A more detailed review on the use of remote sensing data to provide information into flood inundation models is given in [11]. A more specific, application-oriented study is the one presented in [12], in which Dhakal et al. devise a methodology able to detect the flood and erosion areas affected by heavy rainfall using multitemporal Landsat data. In [13], Carrara et al. rely on GIS and remote sensing technologies in order to map landslide hazards. Despite the potential of such technological advancements, landslide hazard mapping remains a major and largely unsolved task. In addition to optical data, radar data have also been used for this purpose. For instance, in [14], Chen et al. develop a new flood mapping method for SAR images, and further evaluate the obtained result by a detailed comparison with multispectral images collected by the SPOT satellite. The effectiveness of using SAR data for flood monitoring is clearly demonstrated in this study. On the other hand, the study described in [15] is focused on the development of a new method to obtain the instantaneous profiles of flood waves using orbital remote sensing and topographic data. This approach is also shown to be useful to measure peak discharges and verify hydraulic models. In [16], Wang illustrates different strategies for the use of remote sensing in flood applications, including: 1) precipitation data sets derived from in situ observations; 2) airborne scanners or detectors; and 3) unmanned aircrafts and GIS-based flood analysis. In [17], André et al. show the potentials of flood mapping using radar/optical imagery and digital elevation models. Another relevant study is the one presented in [18], in which a simple, efficient, and economical method integrating the advantages of Landsat imagery and digital elevation models is presented for mapping flood extent in coastal floodplains.

B. Emergency Response Using Social Media Data

The amount of techniques exploiting social media data for emergency response is quite numerous. An early survey is provided in [19], describing existing research on content-based retrieval for multimedia databases from spatial, temporal, and spatio–temporal relations. The work in [20] analyzes the advantages of combining visual analytics with big data techniques, concluding that visual analysis of social media data enables a wide range of promising new applications. A similar study focused on big data (including new challenges and opportunities) is presented in [21] and [22]. In [23], Abel et al. develop an automatic social web stream filter system called Twicident that allows users to search and analyze information about incidents available on Twitter. This system takes advantage of data semantics to profile incidents and continuously improve stream filters. Another relevant work in this direction is presented in [24], in which Tan et al. explore how to use personal ad hoc clouds comprising individuals in social networks to address big data processing challenges. In [25], a tutorial on models and algorithms for interactive sensing in social networks is presented and discussed. The focus is on the way individuals act as sensors, and on how the information exchange between individuals is exploited to optimize sensing. In this context, social learning is used to model the interaction between individuals that aims to estimate an underlying state of nature. In [26], Field and O’Brien introduce a new framework (called cartoblography) for mapping the spatial context of microblogging. Here, focus is on the possibility to use tweets to perform real-time mapping of different phenomena. Along the lines of this research direction, the work in [27] develops a Twitter-based event detection and analysis system (TEDAS) aimed at detecting new events, analyzing the spatial and temporal pattern of such events, and identifying their importance. On the other hand, the research presented in [28] is focused on modeling city dynamics by analyzing the check-in data of its residents from social media, which serves as a mechanism to study the structure and composition of the city on a large scale. Another relevant approach for processing massive social media data is described in [29], which transforms the resulting knowledge into suitable information to increase efficiency in disaster warning. A similar study based on Twitter data is recently
presented in [30], in which a technique for classification of data coming from this social network is developed for situation awareness based on semisupervised learning, developing online interactive maps of the vulnerable areas. In [31], an automatic filter is presented which can generate quantitative data derived from photos and eye-witnesses in social media posts to support emergency response for rapid inundation mapping, including information about inundation extent and water depth in floods. The research presented in [32] describes a methodology that also leverages social media content to support rapid inundation mapping. In [33], some insights on the role of social media in order to devise warning responses in extreme events are explored. The results obtained in the study indicate that local social media could facilitate such warning responses. The work in [34] presents a procedure to detect and identify earthquakes based on Twitter data, with a conclusion that it is greatly beneficial to exploit social media data for this purpose. Another research work focused on earthquakes is presented in [35], in which Acar and Muraki investigate the posted tweets during the Great Tohoku earthquake followed by a devastating tsunami. The authors particularly address four main aspects: 1) how to classify related tweets; 2) the difference between the tweets posted directly and indirectly; 3) the problems that users faced during the event; and 4) the communication aspects of Twitter. In similar fashion, the study in [36] presents a probabilistic framework for estimating the city-level location of Twitter users based only on the content of the tweets. The result is a list of possible locations for each user in descending order of confidence, showing promising accuracy. A partially related work is presented in [37], in which the socioeconomic properties of users of social networks are investigated by means of three popular online location-based platforms based on two randomized null models. The work in [38], on the other hand, analyzes the 2013 Colorado flood by focusing on how communities in and outside the media spotlight obtained information about the event from social media platforms. Another application-oriented work is presented in [39], in which a flood event in the Brisbane River is addressed. Here, Bunce et al. explore four categories of information experience: 1) monitoring information; 2) community and communication; 3) affirmation; and 4) awareness, all concerning the behavior of individuals using social media during the flooding. On the other hand, the work in [40] studies several floods in Australia during 2010–2011, applying social network analysis to identify active users and their effectiveness in disseminating critical information, and also identifying important online resources disseminated by online communities. In related fashion, the paper in [41] reviews the actions of administrators and users in public network platforms during the Queensland and Victorian floods, elaborating on the value of social media to assist emergency services. Another relevant effort is presented in [42], in which an event notification system is presented that can monitor and identify tweets to predict and detect events automatically by exploiting the real-time nature of Twitter.

C. Emergency Response Using Geographical Information Systems

There has also been a significant amount of research on the use of GIS for emergency response purposes. In [43], Flanagan and Metzger explore the properties of a new system for volunteered geographic information (VGI), which refers to the harnessing of tools to create, assemble, and disseminate geographic data provided voluntarily by individuals. In [44], an assessment of flood risk analysis, management, and current GIS applications for flood damage modeling is presented and discussed, emphasizing the importance of GIS technology when used as a data management platform and decision support tool. In [45], the use of GIS in hydrology and water management is discussed. Hydrological GIS has become a very important aspect in the water and river management domain. In [46], a prototype of a dynamic and collaborative mapping system for flooding events based on VGI is presented, which is supported by citizens during the event. It uses a questionnaire result to evaluate this system in the city of Sao Paulo, Brazil. The work in [47] gives an overview on spatial interpolation methods aiming to provide guidelines and suggestions for different applications. The work compares a number of commonly applied methods and also provides a list of software tools for spatial interpolation. In [48], Poser and Dransch illustrate methods aiming to capture current research and future directions on VGIs used in preparedness and mitigation work. The work in [49] presents the design and prototypical implementation of a geospatial exploratory data mining web agent which reads webpage data and follows links to acquire knowledge in order to extract value on the geoinformation usable by a GIS. The agent creates a database from webpage text, mines it for location information, and then converts it to a proper geospatial data format. In [50], a case study is presented in which a VGI addresses the quality problem from innocent mistakes and intentional falsifications by aggregating input from many different people. Specifically, the work presents a technique to maintain a comprehensive list of points of interest for digital maps. In [51], Elwood et al. examine the content and characteristics of a VGI, including the technical and social processes through which it is produced, and appropriate methods for synthesizing and using these data in research. The work also explores emerging social and political concerns related to this new form of information. In [52], Sui and Goodchild consider GIS as social media. On the one hand, it is concluded that various users and contributors of online mapping sites have formed their own virtual community for exchanging information. On the other hand, it is concluded that interactions of online GIS users or neogeographers or neocartographers are not confined to cyberspace. The work in [53] describes social media geographic information, which comprises VGI from social media platforms used to explore novel methods and tools for analysis and knowledge construction. In [54], Yamamoto develops a social media GIS for disaster risk management in Japan, where the role of GIS
and social media are considered important for collection and transmission of disaster information. In [55], Croitoru et al. present a system prototype for harvesting, processing, modeling, and integrating heterogeneous social media feeds toward the generation of geosocial knowledge. Using a different perspective, the work in [56] describes the main contributions and challenges in applying GIS and remote sensing methods for disaster risk governance. This work analyzes emerging concepts and concludes that GIS and remote sensing technologies have an important impact in data processing and analysis. Finally, the work in [57] presents a case study of fast flood and waterlogging detection using GIS.

II. EMERGENCY RESPONSE THROUGH THE INTEGRATION OF SOCIAL MEDIA AND REMOTE SENSING

In recent years, there has been a significant increase in the number of applications and techniques that use remote sensing and social media jointly in the context of emergency response. For instance, in [58], Zhang et al. propose a transfer learning framework for urban waterlogging analysis, which takes advantage of social media and satellite data to perform multiview discriminant analysis. In [38], Denis et al. develop a comprehensive system integrating remote sensing and social media data for decision making and quick information broadcasting. This system is able to support evacuation strategies and real-time guidance during disasters (the authors specifically develop a tsunami risk map application in Padang, Indonesia). Zhang et al. [59] present a semantic method for social-sensor-based urban waterlogging monitoring. In [60], Backstrom et al. introduce a physical location prediction algorithm which takes into account social information such as population density. It produces measurable improvement in accuracy when compared to other related algorithms. In [61], Jones and Grandhi propose a people-to-people-to-geographical-places system, which aims at describing the design space for location-aware community systems. The study in [62] presents a scalable system for the contextual enrichment of satellite images by crawling and analyzing multimedia content from social media such as Twitter. This system benefits from an enhanced visualization system demonstrating different aspects of the event under consideration. In [63], Brivio et al. propose a new procedure that integrates topographic information with SAR images to overcome the constraints of temporal resolution in the exploitation of SAR images for flood mapping purposes. Goodchild and Glennon [64] focus on the latest developments of crowdsourcing VGI. In [65], Herfort et al. address the geographical features that can be used between social media data and flood phenomena and present a new approach to extract useful crisis-relevant information from social media platforms. A free and open-source web platform called CrisisTracker is built in [66] to extract distributed situation awareness from public tweets during large-scale events. In [67], the social media alert and response to threats to citizens (SMART-C) program developed by the U.S. Department of Homeland Security’s Science & Technology Directorate (DHS-S&T) is described. This program aims to develop citizen participatory sensing capabilities for decision support throughout the disaster life cycle via a multitude of devices and modalities. The authors provide an overview of the envisioned SMART-C system’s capabilities and discuss some of the most interesting and unique challenges that arise due to the combination of spatial computing and social media within the context of disaster management. A comprehensive survey [68] is presented on the processing of social media messages in mass emergencies. In [69], Middleton et al. develop a social media crisis mapping platform for natural disasters that uses locations from gazetteer, street map, and VGI sources for areas at risk of disaster, and matches them to geoparsed real-time tweet data streams, where statistical analysis is performed to generate real-time crisis maps. In [70], Yin et al. propose a comprehensive framework that combines hydrological modeling and GIS spatial analysis for small-scale risk assessments of urban waterlogging. The authors further derive the risk curve of rainstorm waterlogging hazards damage of a study area in Shanghai, China. A framework for disaster assessment using social media is developed in [71]. Geospatially oriented social media communications [72] are presented as a common information resource to support crisis management. In [73], Wang and Li propose a GIS and remote sensing-based urban waterlogging monitoring and warning system which takes into account hydrology, water dynamics, meteorology, and urban sewerage systems. In [74], a participatory sensing-based model for mining spatial information related to urban emergency events is introduced, where the Chan-hom typhoon is addressed as an example. Semantic analysis on microblog data is conducted, and high-frequency keywords in different provinces are extracted for different stages of the event. With the geo-tagged and time-tagged data, the collected microblog data can be classified into different categories. Correspondingly, some public opinion and requirements can be obtained from the spatial and temporal information to enhance situation awareness and help governments offer more effective assistance. Deng et al. [75] use social media data collected from Sina Weibo, which is a microblogging website like the hybrid of Twitter and Facebook, to analyze the public opinion on the spatial and temporal perspectives for emergency response, where the Chan-hom typhoon is again considered as an example. A crowdsourcing-based model is introduced in [76] for mining spatial information of urban emergency events. In [77], Chae et al. present a visual analytics approach that provides users with scalable and interactive social media data analysis and visualization including the exploration and examination of abnormal topics and events within various social media data sources, such as Twitter, Flickr, and YouTube. In [78], Earle et al. provide a quick review of Twitter and its capabilities and investigate the possibility of using tweets to detect seismic events. De Longueville et al. [79] study how Twitter can be used as a source of
spatio–temporal information. By focusing on a recent case of forest fire, the authors demonstrate the role of Twitter to support emergency planning, risk assessment, and damage assessment activities. In [80], Yamamoto and Fujita develop a social media GIS which is specially tailored to mash up the information that local residents and governments provide to support information utilization from normal times to disaster outbreak times, in order to promote disaster reduction. The newly developed social media GIS integrates a Web-GIS with Twitter, and includes a function for classifying the submitted information. A real-time situation awareness viewer is developed in [81] for monitoring disaster impacts using location-based social media messages in Twitter. In [82], Sakaki et al. develop an earthquake reporting system for use in Japan. Specifically, the authors investigate the real-time interaction of events such as earthquakes in Twitter and propose an algorithm to monitor tweets and to detect a target event. Carley et al. [83] examine the use of Twitter in the context of disaster management, with a focus on planning and early warnings. Landwehr et al. focus on the potential use of Twitter to support tsunami warning and response. In [84], Twitter is again presented as a social media that can be used to help mitigate disasters. Here, the authors describe the strengths and limitations of Twitter in this context, and identify the features needed in a Twitter system to support disaster planning, warning, and response. In [85], Xiao et al. examine the spatial heterogeneity in the generation of tweets after a major disaster. They develop a new model to explain the number of tweets by mass, material, access, and motivation. Empirical analysis of tweets after the Hurricane Sandy in New York City largely confirms the newly developed model. The authors also find that community socioeconomic factors are more important than population size and damage levels in predicting disaster-related tweets. An advanced system is presented in [86] for emergency management which fuses the potentiality offered by mobile social data and bottom-up communication with smart sensors. In [87], Yin et al. present a system that uses social media to enhance emergency situation awareness. The described system uses natural language processing and data mining techniques to extract situation awareness information from Twitter messages generated during various disasters and crises. Social media poses two types of challenges. The first is how to sift relevant information from the social media data, and the second is that the traditional natural language processing techniques are inapplicable to the user-generated content. To deal with these difficulties, the authors develop a coherent set of integrated components for extracting situation awareness by using various data mining techniques, including burst detection, text classification, online clustering, and geotagging. In [88], Crooks et al. analyze the spatial and temporal characteristics of the Twitter feed activity responding to an earthquake which occurred on the East Coast of the United States. They argue that these feeds represent a hybrid form of a sensor system that allows for the identification and localization of the impact area of the event. Schnebele and Cervone [89] propose a new pixel-based method to optimize and modify the contour of flood regions with VGI data obtained through Google news, videos and photos. In [90], Werts et al. build a Web-GIS framework which allows the public to upload photos and attributes of their sites and use these timely data in the context of soil and water conservation procedures.

III. CASE STUDY: HEAVY RAINFALL EVENT MONITORING USING DEEP LEARNING

In order to illustrate the potential of integrating remote sensing and social media data for emergency response, as a follow-up to the previous review section we now discuss several case studies based on the exploitation of Sina Weibo social media data sets from central Wuhan and Shenzhen cities. The goal of these case studies is to monitor heavy rainfall events recently happening in both cities. As shown in Fig. 1,
the regions marked in purple are the study areas in Wuhan and Shenzhen. Notably, the number of Sina Weibo messages published in the central part of a city is much larger than the number of messages generated in noncentral areas. As a result, we choose the urban centers of Wuhan and Shenzhen (i.e., Jiangan, Jianghan, Qiaokou, Hanyang, Wuchang, Qingshan, and Hongshan districts in Wuhan, Baohan, Nanshan, Futian, and Luohu districts in Shenzhen) rather than the whole cities as our target areas. In the following, we briefly review the existing emergency monitoring methodologies, followed by an introduction of our heavy rainfall monitoring framework. We show that, after analyzing the data collected from Sina Weibo websites, we can further explore the relationship between environmental phenomena and social media responses in the considered study areas.

A. Emergency Monitoring Methodologies

Monitoring of emergencies is fundamental for decreasing the potential impact of unexpected events, such as natural or human disasters. Making timely decisions and taking preventive actions saves human lives, reduces economic losses, and enhances the stability of the communities in the incident. Much work has been carried out to improve the monitoring performance in the literature. One of the most outstanding research aspects is based on machine learning algorithms, including the support vector machine (SVM) [91]–[94], multitask learning [95], probabilistic models [82], maximum entropy [96], logistic regression [92], etc. [97]–[100]. Recently, deep learning [101]–[110] has emerged as a state-of-the-art machine learning technique. It learns the representative and discriminative features in hierarchical architectures and shows successful results in a broad area of applications, such as image processing [111]–[115], computer vision [116], [117], and speech recognition [118], [119]. Many kinds of deep neural networks have been designed, e.g., deep belief networks (DBNs), convolutional neural networks (CNNs), deep stack networks (DSNs), and recurrent neural networks (RNNs). Specifically, deep learning has also been studied in the field of emergency monitoring. For example, a Chinese emergency event recognition model is proposed in [120] by applying the DBN. In our case study, CNN [121]–[123] is adopted to classify the clawed Sina Weibo data sets. The reasons for adopting CNN are that: 1) CNN is the first truly successful deep learning architecture benefited from the successful training of hierarchical layers; 2) CNN is an end-to-end learning process, from the raw data rather than hand-coded features to semantic labels, and thus no longer needs to manually devise the features; and 3) the shared weights within layers can reduce the free parameters and ease the burden of the computing. It is notable that this paper does not limit to the CNN but offers a general architecture (i.e., deep learning) for heavy rainfall event monitoring. As such, although CNN is adopted as a typical deep learning method in the case study, it is also allowed to use other state-of-the-art deep learning methods (e.g., DBN, DSN, or RNN).

The CNN, proposed by Lecun et al. [121] in 1998, is a biologically inspired variant of the multilayer perceptron (MLP). This type of hierarchical network is one of the most popular deep learning architectures consisting of various layers, including convolutional layers, pooling/subsampling layers, and fully connected layers. The layers can vary with how the data are sampled and trained. Favored by the development of computer hardware and software technologies, CNN has been of growing interest recently in many applications. In [124], three strategies are proposed to exploit the CNN for remote sensing scene classification. In [125], CNN models are directly trained to produce classification maps out of the input large-scale remote sensing images. In [126], predatory conversations are automatically identified by CNN-based method. In [127], opinion summarizations in Chinese microblogging systems are effectively mined by the CNN. Based upon the CNN, our newly developed framework for heavy rainfall monitoring is portrayed in Fig. 2. As shown in Fig. 2, the data sets with time and geospatial information are crawled by the Sina Weibo open platform API, and then a CNN model is designed to classify the Sina Weibo messages into positive and negative classes. Finally, the heavy rainfall emergence event can be detected according to the classification results.

![Fig. 2. Framework of the heavy rainfall monitoring methodology used for this case study.](image-url)
B. Data Collection

The experimental data sets used in this paper are Sina Weibo messages with specific keywords (in Chinese) like “雨,” “水,” “涝,” “淹,” “河,” and/or “海” from both central Wuhan and Shenzhen. For example, Fig. 3 displays some random Sina Weibo posts that contain the keyword “雨” (rain). Wuhan, which is known as a “Thoroughfare to Nine Provinces,” is the capital and the most populous city of central China’s Hubei Province. Wuhan’s climate is humid subtropical, with abundant rainfall. The summer season is the wettest period during which lots of disasters can happen, such as torrential rains and floods. Heavy rainfall can cause huge losses of life and property. For example, Wuhan saw 570 mm (about 22.44 in) of rainfall (see Fig. 4) during July 1–7, 2016, surpassing the record of the city in 1991. Red alerts for heavy rainfall were released on July 2 and 6, respectively. More than 27 people died and the economic losses reached ¥5.7 billion (about $850 million). On the other hand, Shenzhen is a major city in South China’s Guangdong Province. The city is located north of the Hong Kong Special Administrative Region and is the first special economic zone in China. Shenzhen has a humid subtropical climate. It is affected by monsoons, which leads to hot weather and typhoons, together with thunderstorms in the summer. Due to the heavy rainfall and floods, Shenzhen suffers severe water crises. For example, during May 8–11, 2014, as many as 1500 houses were destroyed and three people were killed by the torrential rain. The direct economic loss was more than ¥80 million (about $12.8 million). Figs. 4 and 5 show the influences of the heavy rainfall in Wuhan and Shenzhen. Moreover, it is notable that Sina Weibo is one of the most popular social media websites in China. It was launched by Sina Corporation in 2009 and as of March 2016 had more than 261 million monthly active users. It is able to convey information quickly since each message is limited to 140 characters.

As described above, two data sets crawled from Sina Weibo are used to detect the heavy rainfall events in the experiments. The first is a collection of Weibo data from June 29, 2016 to July 10, 2016, located in the center of Wuhan, while the second data are from May 8, 2014 to May 11, 2014, located in the center of Shenzhen. We adopt the Sina Weibo open platform API to collect data from Weibo websites. To assess the API’s information, it is necessary to get the authentication of the Sina Weibo through the OAuth protocol. As depicted in Fig. 6, the geospatial Weibo data are collected by defining circular zones with different centers and fixed radius in a period to make the obtained data cover the whole city. In total, 8519 and 2442 significant Weibo messages are crawled in Wuhan and Shenzhen, respectively. Each Weibo message relates to a time and location, which includes the latitude and longitude coordinates.

1The keywords “雨,” “水,” “涝,” “淹,” “河,” and/or “海” refer to “rain,” “water,” “waterlogging,” “flood,” “river,” and/or “ocean” in English.
C. Classification of Sina Weibo Data by CNN

To perform the analysis of our considered heavy rainfall event using Sina Weibo data, it is important to identify useful information from the raw data. The crawled data with specific keywords usually include mention of the target event. For example, users may post Weibo messages such as “Torrential rain!,” “The rain just keeps pouring down,” or “Streets are all flooded.” However, users might also publish Weibo messages such as “Eventually, the rain stopped,” “Storms always give way to the sun,” or “I like to swim, especially in the sea.” Moreover, even though a message is associated with the heavy rainfall event, it may not be right to take the message as an event report. For example, users make Weibo messages such as “It was raining so hard yesterday” or “Three rains in two days.” Although the descriptions are truly related to the heavy rainfall event, they cannot reflect the real-time situation. In order to better conduct the analysis, it is worthwhile to distinguish whether a Weibo message truly reflects the actual occurrence of heavy rainfall.

Fig. 5. The heavy rainfall in Shenzhen. (a) Pedestrians walked in a flooded street (http://slide.news.sina.com.cn/weather/slide_1_29155_60229.html#p=6). (b) A car waded through the rainwater (http://slide.news.sina.com.cn/weather/slide_1_29155_60229.html#p=10).

Fig. 6. Circular zones with different centers and fixed radius are utilized to make the Sina Weibo data sets cover the research areas. (a) Circular zones of central Wuhan. (b) Circular zones of central Shenzhen.
events. Specifically, the true and real-time descriptions are denoted as positive class, whereas the others are labeled as negative class. In the case study, all of the Sina Weibo messages are manually labeled as positive or negative classes. As will be shown later, the classification performance of the CNN model can be evaluated by comparing the manual labels and the classified labels.

To classify a Sina Weibo message as belonging to the positive class or the negative class, we adopt the CNN model [122], [123], which has proven to be a powerful tool for machine learning tasks. Providing that a training set with positive and negative samples is available, the Sina Weibo data can be automatically classified into two classes by the CNN model. In the Wuhan data set, the samples used for training are 850, while the training set of Shenzhen dataset contains 100 samples.2

The structure of the CNN considered in our experiments is shown in Fig. 7. In the CNN, the first layer embeds the words of the input sentence into low-dimensional vectors. Word embeddings compute the distributed representations of the input words in the form of continuous vectors. It can alleviate the data sparsity problem by capturing meaningful semantic and syntactic regularities. Mapping the words to vectors of real numbers facilitates subsequent processing of the CNN model. The subsequent layer conducts convolutions on the embedded word vectors with multiple filter windows. Different sizes of filter windows can slide a different number of words at a time. Next, max-pooling is performed on the result of the convolutional layer. Finally, dropout regularization is added in the fully connected layer and the classification results are obtained by a softmax operation. Moreover, it is essential to choose proper hyperparameters in the CNN model. We set the filter windows as 1, 2, 3, 4, 5, 6, and 8, the learning rate as 0.0001, the dropout rate as 0.5, the $l_2$ constraint as 3 (i.e., the weight vectors are rescaled to $\|W\|_2 = 3$ whenever the $l_2$-norm of the weight vectors is larger than 3), the vector size of the word embedding as 50, and the batch size as 32. We initialize the word vectors by an unsupervised neural language model to obtain improved performance. The word embeddings are trained by the Skip-gram algorithm with default parameters in Word2Vec.3

Wikipedia dumps in Chinese are used in the word embeddings. Words that are not contained in the pretrained vectors are initialized randomly.

D. Monitoring Results of the Heavy Rainfall Events

The classification performance in our experiments is evaluated by precision, recall, and $F$-value. Precision refers to the fraction of true positives in total determined positives (i.e., the sum of true positives and false positives). Recall is the ratio of true positives to the sum of true positives and false negatives. $F$-value measures both the precision and recall synthetically. The definitions of precision, recall, and $F$-value are given by

\[
\text{Precision} = \frac{TP}{TP + FP} \quad (1)
\]
\[
\text{Recall} = \frac{TP}{TP + FN} \quad (2)
\]
\[
F\text{-value} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)
\]

where $TP$, $FP$, and $FN$ denote the number of true positives, false positives, and false negatives, respectively.

The classification results of both data sets are depicted in Fig. 8, from which we can observe that the precision, recall, and $F$-value of the center of Wuhan heavy rainfall event are lower than those of Shenzhen, respectively. In both data sets, the precision is not as high as the recall, and the $F$-value offers a compromise between precision and recall. For visualization purposes, we display the heat maps in Figs. 9 and 10, which show the location estimation of heavy rainfall events occurred in the center of Wuhan and Shenzhen, respectively. Without loss of generality, Fig. 9 presents the

---

2Although only a few training samples are used, the classification results of the CNN-based method are still satisfactory.

3https://code.google.com/p/word2vec/
rainfall monitoring results on June 30 and July 2 and 6, 2016, while Fig. 10 depicts the results on May 8, 9, and 11, 2014. It can be observed from Fig. 9 that the rainfall on July 2 and 6 is heavier than on June 30 in the center of Wuhan. This corresponds to the actual case. In reality, the rainfall started from June 30, and became heavy in the next weeks. Red rainstorm warnings were issued on July 2 and 6. Furthermore, as shown in Fig. 10, the rainfall on May 11 is much larger than on May 8 and 9 in the center of Shenzhen city. This observation is also consistent with reality. In a nutshell, the classification results obtained for the central areas of Wuhan and Shenzhen using Sina Weibo data sets demonstrate the effectiveness of the CNN-based method in monitoring the heavy rainfall event that occurred in those areas.

IV. CONCLUSION

In this paper, we have presented a comprehensive review of the state of the art in the use of social media data as a complement to remote sensing and GIS data in emergency response scenarios. First, we have focused on strategies that mainly use remote sensing data for emergency response. Then, we have described techniques that are based mainly on social media data. Next, we describe strategies that are mainly based on GIS data. The limitations of these techniques are pointed out by including a thorough review of techniques that exploit both remote sensing and social media data. In order to provide a realistic illustration of such techniques, a specific application case study is further discussed in detail. The case study addresses
the possibility of monitoring heavy rainfall events using a combination of remote sensing data and social media data. Specifically, we study heavy rainfall events in the cities of Wuhan and Shenzhen, China, by exploiting messages published on Sina Weibo. The analysis and classification is conducted using a CNN architecture. The classification results obtained for the central parts of Wuhan and Shenzhen (in which the Sina Weibo messages were more numerous) demonstrate the effectiveness of the considered CNN-based method in monitoring the heavy rainfall event that happened in both cities. This illustrates the potential of using social media data as a complement to remote sensing data sets for emergency response applications.

REFERENCES


Developments in Social Software


