Analytics and Evolving Landscape of Machine Learning for Emergency Response

Minsung Hong · Rajendra Akerkar

Received: date / Accepted: date

Abstract The advances in information technology have had a profound impact on emergency management by making unprecedented volumes of data available to the decision makers. This has resulted in new challenges related to the effective management of large volumes of data. In this regard, the role of machine learning in mass emergency and humanitarian crises is constantly evolving and gaining traction. As a branch of artificial intelligence, machine learning technologies have the huge advantages of self-learning, self-organization, and self-adaptation, along with simpleness, generality and robustness. Although these technologies do not perfectly solve issues in emergency management. They have greatly improved the capability and effectiveness of emergency management. In this paper, we review the use of machine learning techniques to support the decision-making processes for the emergency management and discuss their challenges. Additionally, we discuss the challenges and opportunities of the machine learning approaches and intelligent data analysis to distinct phases of emergency management. Based on the literature review, we observe a trend to move from narrow in scope, problem-specific applications of machine learning to solutions that address a wider spectrum of problems, such as situational awareness and real-time threat assessment using diverse streams of data. This chapter also focuses on crowd-sourcing approaches with machine learning to achieve better understanding and decision support during an emergency.

Keywords emergency management · crisis analytics · data analysis · data mining · machine learning · decision making · situational awareness · real-time assessment · deep learning · data streams

Minsung Hong
Vestlandsforskning, Box 163, NO-6851 Sogndal, Norway
Tel.: +47-908-03-233
E-mail: msh@vestforsk.no

Rajendra Akerkar
Vestlandsforskning, Box 163, NO-6851 Sogndal, Norway
Tel.: +47-916-85-607
E-mail: rak@vestforsk.no

1 Introduction

In the contemporary society, a variety of emergencies take place more and more frequently. A considerable number of emergency incidents have threatened to human life, environmental protection, social stability and even political relationship of all countries around the world [29]. In this regard, sociologists have been working to define emergency for decades. There is a broad consensus that emergencies are social phenomena, characterized by a disruption of routine and of social structure, norms, and/or values. It implies that the severity of an emergency is more related to the extent of the disruption of social life as aspects of governments, business and individuals, than the measurable magnitude of the hazard [25]. Therefore, the negative effects of emergencies emphasize the need to improve the emergency management capability and strengthen the security worldwide.

1.1 Emergency Management

Unlike other, more structured disciplines, emergency management has expanded and contracted in response to events, congressional desires and leadership styles [41]. Some representative definitions in the literature are as follows:

- According to definition of the Federal Emergency Management Agency (FEMA) in USA, the process of emergency management consists of preparing for, mitigating, responding to and recovering from an emergency when a disaster arises [2].
- More modern emergency management involves processes to apply modern technologies and management methods to effectively and efficiently monitoring, response to, control and process events, by integrating various social resources and analysing scientifically the cause [28].
- Another definition for emergency management is “a discipline that deals with risk and risk avoidance.” Risk represents a wide range of issues and the range of situations that might possibly involve emergency management or the emergency management system is vast. This supports the premise that emergency management is essential to the security of everyone’s daily lives and should be integrated into daily decisions and not just called on during times of disasters [41].

In essence, emergency management is a complex and multifaceted task that involves a variety of activities, of managers and stakeholders, when emergency is not only arising but also before and after of emergency, so as to prevent the occurrence of unexpected events, to reduce the social damages and to mitigate the impacts. Based on the definitions of the emergency management, the evolution of an emergency can be distinguished as three stages, namely pre-emergency, in-emergency and post-emergency, as shown in figure 1 [3]. Chen et al. described emergency management as a ‘4R’ process, namely reduction, readiness, response and recovery. Reduction is referred to the pre-emergency phase, readiness and response belong to the in-emergency phase, and recovery is referred to the post-emergency phase. In each phase, the outcome of decision-making impacts substantially the evolution of events and the effectiveness of emergency management [28].

Fig. 1 The lifecycle of emergency management

As aforementioned, emergency management is a multifaceted process to prevent, reduce, respond to and recover from the impact of the emergency on the society. Because of the scale of events, emergency response requires the participation and cooperation of multiple organizations (e.g., government, public and private). This emphasizes the need for efficient and effective decision support systems, as it is practically impossible for a human decision maker to understand and manage the complexity of the situation. Instead, problems such as situational awareness [144] and building a common operating picture, shared among multiple actors who often have only partial view of the situation, are becoming some of the most urgent needs of emergency management [140].

However, the emergency data used to these decision support systems arises from the problems of delivering repetitive information and information overload [4]. Therefore, to improve the capability and effectiveness of emergency management, machine learning techniques have been utilised.

1.2 Machine Learning

Emergency management is concerned not only with predicting the course and consequences of emergencies, but also mitigating those undesired consequences. This process is undoubtedly a challenging task by the unprecedented volumes of data (e.g., forecast, news, web pages, data of social network service, and sensing data) and the pressure of time [140]. Machine learning techniques have been proven to successfully support the decision making processes in managing many complex problems. In that sense, emergency management is no exception; however, it presents various challenge to machine learning techniques for the emergency response. In this section, we briefly introduce category of machine learning.

Machine learning has progressed dramatically over the past two decades, from laboratory curiosity to a practical technology in widespread commercial use [58]. Machine learning and data mining often use the same methods and overlap significantly, but while machine learning focuses on prediction, data mining concentrates on the discovery of (previously) unknown properties in the data. According to depending on whether there is labeled instance which is consists of label and data, Machine learning are typically classified into three categories as follows:

- **Supervised learning:** Supervised machine learning makes predictions about future instances using externally supplied instances that consist of values and a label. Its goal is to build a concise model of the distribution of class labels, and then a classifier based on the model is used to assign class labels to the testing instances [65].
- **Unsupervised learning:** Unsupervised learning is inferring directly the properties of this probability density without the help of externally provided instances providing correct label or degree-of-error for each observation [43]. There are representative algorithms like as Apriori algorithm, K-means and so on.
- **Reinforcement learning:** Reinforcement learning deduces labels of instances with a dynamic environment. There are two main strategies. The first is to search in the space of behaviours in order to find one that performs well in the environment, such as genetic algorithms, and the second is to use statistical techniques and dynamic programming methods to estimate the utility of taking actions in states of the world [60].

1.3 Scope and Organizations

This chapter focuses on the application of machine learning techniques to support the decision-making processes for the emergency management. We start with the data-driven methodologies within the frameworks of machine learning and their roles and challenges in supporting different phases of emergency management. We then discuss the characteristics of disaster data akin to 5 Vs of big data and summarize various applications cases of big data analysis. Next, with respect to emphasizing the advance of the social media, we focus on reviewing the crowd-sourcing approaches with machine learning in emergency management, and issues of the approaches are discussed in terms of the data analysis. Finally, some examples of the tweet related to emergency are discussed to more deeply contemplate the challenges and opportunities.

Existing survey papers for machine learning techniques in emergency management have reviewed according to categories of the machine learning, or have considered only some part of emergency tasks. Whereas, in this chapter, we review the approaches of machine learning, have been discussed from 2010 to current, along each task of emergencies. Therefore, we believe that readers can easily find topics related to their interests and compare with existing approaches to little more concretely grasp potentiality of their methods.

2 Applications of Machine Learning in Emergency Response

As discussed earlier, within the present-day emergency management, the immediate and accurate decision making more and more relies on the capability of data analysis and processing. Therefore, there is an urgent need to enhance the machine learning functionality of emergency management, such as, to develop scalable and real-time algorithms for time-sensitive decisions, to integrate structured, unstructured and semi-structured data [29]. In this section, we attempt to introduce the tasks of machine learning in each phase of emergency management and review the

challenges and benefits of various machine learning techniques for the emergency management.

2.1 Machine Learning Techniques for Emergency Management Cycles

Successful emergency management requires a variety of tasks based on various technologies of machine learning within across the board three phases mentioned in Section 1.1. Figure 2 shows the tasks related to machine learning for each phase of the emergency management as follows: (1) predicting the occurrence of potential events and discovering the early warning signs; (2) during the emergency, detecting the events occurred and tracking change of the incidents, and recognizing situations of people, supply, and so on; (3) evaluating the loss caused by incidents and the execution of response, and simultaneously adjusting volunteer efforts based on crowdsourcing to recover from an emergency.

Fig. 2 Tasks related to machine learning in emergency management

- **Event Prediction:** it is forecasting emergency using technology and interpretation methods and can be achieved by extracting features or pattern. Although there is no prediction method with perfect accuracy, early detection of natural disasters reduces hazards in nearby locations [87].
- **Warning Systems:** to detect impending emergency can give that information to people at risk and enable those in danger to make decisions and early take action [123]. These systems have improved drastically in recent years but they are not perfect yet.
- **Event Detection & Tracking:** Most systems based on machine learning during crises start with detecting and tracking events. The events are mainly

associated with a specific time and location [21]. However, due to the online nature of collected data, events may or may not be necessarily associated with physical locations.

- **Situational Awareness:** it provides more deep recognition of events in emergency using social media data related to specific information (e.g., caution, advice, donations, casualties and damage) and smart-phones which typically mount various sensors such as camera, GPS and accelerometer [140].
- **Emergency Evaluation:** it is one of critical and complex tasks in emergency management [29]. In post-emergency, the activity outcomes (e.g., loss of resources, recoverability, performance and social influence) for current emergency should be measured to suppress the deterioration of next emergency.
- **Crowdsourcing:** this task is a sourcing model in which organizations use predominantly advanced Internet technologies to harness the efforts of a virtual crowd to perform specific organizational tasks [117]. It may allow to immediately collecting the statuses and requirements of people after an emergency and analysed and categorized the data collected to support relief operations.

Similarly, there are various tasks for emergencies, and machine learning techniques have been applied to each task to improve the effective and efficient emergency management. In next subsections, the approaches of machine learning will be reviewed for each task.

2.2 Event Prediction

Many researchers used Neural Networks techniques for predicting emergency. Shah and Ghazali proposed Improved Artificial Bee Colony (IABC) algorithm to improving the training process of Multilayer Perceptron (MLP) in order to overcome local minimal and slow convergence of ordinary backpropagation [118]. The IACB-MLP has showed that it is outperforms than the conventional backpropagation for forecasting earthquake magnitude with time series data in California. Also, to predict magnitudes of earthquake and the impending event following the occurrence of pre-seismic signals, Moustira et al. evaluated the performances of Artificial Neural Networks (ANNs) with various types of input data from the region of Greece [82]. In their study, a feed-forward MLP type Neural Network was implemented using the backpropagation learning algorithm for training. A feed-forward backpropagation algorithm has been also applied into development of a time-dependent surrogate model of storm surge [63]. As experimental results, storm surge was predicted by the 92 trained networks for approaching hurricane climatological and track parameters in a few seconds. Other approach based on ANN has been studied to forecast probabilities of occurrence and re-occurrences of earthquake in the region of Chile by Reyes et al., and they used input values (e.g., the b-value, the Bath's law and the Omori-Utsu's law) which are strongly correlated with seismicity [109]. The occurrences have been judged by threshold values which are adjusted for obtaining as few as false positives as possible. In addition, a combination an ANN and Genetic Algorithm (GA) has been proposed to predict 1-day-ahead Monsoon flood by Sahay and Scrivastava [111]. Four wavelet transform-genetic algorithm-neural network models (WAGANN) have been developed and evaluated for forecasting flows in two Indian rivers such as the Kosi and Gandak. In their

experiments, WAGANN models predicted relatively reasonable estimates for the extreme flows and showed little bias for under-prediction or over-prediction.

A variety of clustering methods have been also applied to the prediction task for emergency management. An approach for the prediction of the seasonal tropical cyclone activity over the western North Pacific has been developed to provide useful probabilistic information on the seasonal characteristics of the tropical cyclone tracks and vulnerable areas [62]. In a developed model, the fuzzy c-means clustering has been used to forecast tropical cyclone tracks and density over the entire basin. From experiments, seven patterns were founded to draw a map of the seasonal track density of tropical cyclone. Moreover, the k-means clustering technique has been combined with the statistical regression techniques for the inducement the weather phenomenon in forecasting the cloudburst [94]. The approach clusters atmospheric pressure according to areas of strong relative humidity for discovering weather patterns. To predict wildfire risk using weather data, Context-Based Fire Risk (CBFR) model has been developed based on clustering and ensemble learning techniques by considering the inherent challenge arising due to the temporal dynamicity of weather data [115]. These two machine learning techniques are used to anomaly detection. A particle swarm optimization algorithm-based clustering method with abnormally high-dimensional data has been also proposed to forecast earthquake [142]. A model analyses relationships between earthquake precursor data and earthquake magnitude, and an average distance between clusters is set as the evaluation function of the particle swarm optimization clustering algorithm. Experimental results indicate that this model can effectively and validly predict the earthquake magnitude in accordance with the earthquake precursor data than k-means algorithm model. Additionally, an ant-colony clustering algorithm has been introduced in earthquake prediction by Shao et al [119]. Measure parameters include spatial entropy, mean-fit and un-similar for clustering analysis. As their experiments, it showed that their algorithm could achieve better results than the traditional k-means algorithm to forecast of earthquake like the swarm optimization algorithm-based clustering method.

Decision trees, have often fast and accurate performance in machine learning, have been combined with other techniques to predict emergencies. To predict disaster before it has occurrence, there are many studies which combine the decision tree with various machine learning techniques such as regression [145], hidden Markov model [125], association rule learning [79,34] and fuzzy logic and particle swarm optimization [106]. In particular, decision tree techniques have been also applied into prediction of surroundings (e.g., flood susceptible areas [127] and landslide susceptible areas [30]) for emergency situation.

2.3 Warning Systems

Large magnitude emergency such as earthquake and flood kill and injure tens to thousands of people, inflicting lasting societal and economic disasters. Early warning could provide seconds to minutes of warning, allowing people to move to safe zones and prepare activities like automated slowdown and shutdown of transit and other machinery [64].

For early detection and warning of emergency in environments with wireless sensor networks (WSNs), Bahrepour et al. have tried to consolidate a general

decision tree with the reputation-based voting method [12,13]. In their works, early event warning of emergency are fulfilled with distributed event detection. As experimental results with wild and residential fire datasets, it was showed that their approach not only achieves a high detection rate but also has a low computational overhead and time complexity. In addition, a Random Forest (RF) based decision tree was applied into analysis of the potential factors affecting the satellite signal to announce the flood by Revilla-Romero et al [108]. They investigated various satellite data for 322 rivers in Africa, Asia, Europe, North America and South America. Their experiments shown that the mean discharge, climatic region, land cover and upstream catchment area are dominant variables to determine good or poor vulnerable of surrounding areas.

ANN techniques have been applied into warning emergency with various perspectives. Kong et al. used a smartphone-based seismic network which consists of smart device contains accelerometers [64]. The ANN was used to separate data collected from personal smartphone sensors into activities of the earthquake and human to warn earthquake. Additionally, to set an early warning threshold level of dam, a continuous monitoring of long-term static deformation based on three ANN approaches (i.e., the static neural network, the dynamic neural network and the auto-associate network) was proposed [61].

Krzhizhanovskaya et al. developed a flood early warning system to monitor sensor networks installed in flood defenses (e.g., dikes, dams and embankments), detect abnormalities in sensor signals, calculate dike failure probability, and simulate possible scenarios of dike breaching and flood propagation [66,98]. In the warning system, k-means clustering and Neural Clouds (NC) based classification has role to detect abnormality of sensor parameters in critical pre-failure conditions. Social media data that contains social concerns of people was used for early warning system by Avvenuti et al [11]. Their system applies classification techniques provided by Weka to distinguish Twitter messages into “useful” and “not useful”, and several machine learning techniques are utilized to temporal and spatial analysis of messages. There is also another study considers social media data. Persini et al. implemented a decision support system using machine learning and natural language processing to effectively detect and warn the earthquake [37]. On a real Twitter dataset, their system has shown outperformed results to identify messages related to the earthquakes and critical tremors.

2.4 Event Detection & Tracking

The Support vector machine (SVM) is a supervised learning model that defines a kernel function able to transform the data to a high dimensional feature space when the data can be separated by linear models. SVM is designed for binary classification in nature, but can also solve the multi-class classification problems through one-against-one or one-against-all strategy. SVM was found be effective in emergency rescue evacuation support system for detecting and tracking a sudden incident [81,45]. Pohl et al. took advantage of the Self-Organizing Map (SOM) and the Agglomerative Clustering (AC) for sub-event detection that operate on Flickr and YouTube data [102]. As they mention, multimedia data may be of particular importance to detect and track emergency event. Therefore, Vector Space Model (VSM) was also utilized to represent and annotate the media data before applying

clustering techniques. Their experiments showed that social multimedia data in the context of emergency is worth using for detecting sub-events. Similarly a method of cross-media analytic based on the clustering, sentiment analysis and keyword extraction was introduced to detect and track emergency events [134]. Moreover, the semantic expansion and sentiment analysis were adopted to quantify public sentiment time series.

Song et al. developed a model of human behaviour that takes into account several factors have been founded through empirical analysis between human mobility and emergency to detect and monitor human emergency behaviour and their mobility during large-scale emergency [122]. For the model, they used Hidden Markov Model (HMM) to model dependency between human behaviours in emergency. For the Great East Japan Earthquake and the Fukushima nuclear accident, the efficiency of the behaviour model was evaluated. HMM for speech recognition technique was applied into detecting earthquakes and tracking volcano activities [17]. To fit the model parameters in to earthquake detection, Beyreuther et al. introduced state clustering into their model to refine the intrinsically assumed time dependency. As experiments in during around four months, their earthquake detector of single station HMM showed that it can achieve similar detection rates as a common trigger in combination with coincidence sums over two stations. Akin to this, an approach, which used audio data, proposed to identify anthropogenic disasters by Ye et al [137]. In their approach, acoustic events are detected and learned using the dictionary learning and the spherical k-means clustering. And detected events then are classified into specific sounds (e.g., screaming, shouting, gun shout and explosion) by a clustering technique based on the hierarchical regularized logistic regression model. Experimental results with an audio dataset showed the effectiveness of the proposed hazard sound recognition method. Singh et al. investigated Twitter posts in a flood and proposed an algorithm to identify victims asking for help. To categorize the posts into high or low priority tweets, the SVM, the gradient boosting and the RF are applied. Furthermore, inferring user locations using the Markov model uses historical locations of users. In their experiments, the proposed algorithm worked with its classification accuracy of 81% and location prediction accuracy of 87% [121]. In addition, Caragea et al. developed an enhanced messaging for the emergency response sector, as a reusable information technology infrastructure, to detect and track emergency [24]. They focused on correct classifying messages during disasters by using the SVM classifier. Furthermore, the bag-of-words (BoW) approach, the feature abstraction, the feature selection and the Latent Dirichlet Allocation (LDA) were applied into feature representation as inputs for learning a classifier. Besides, various techniques of machine learning have been used for analysis of social media data to detect and track emergencies [112, 113, 128].

2.5 Situational Awareness

The recent advances of mobile devices, that are capable of wireless communications and have sufficient computing power, have caught attention of researchers and practitioners. The devices can serve as automated sensors (typically equipped with GPS, motion sensors, etc.) and are capable of relatively high quality imaging and video recording. Therefore, they can be used to enhance situational awareness

by gleaning various information to produce accurate results [132]. Furthermore, the advance of social networking services (e.g., Twitter, Instagram, Flickr and Facebook) allows people post their needs and gather the timely-relevant information. Tweets as one of these messages were also investigated to detect possible seismic events, to compare and contrast the people behaviour during emergency and to extract useful information using several extraction techniques [4]. Like this, social media is also a potential source for situational awareness in emergency management. Recent disasters, such as the Hurricane Sandy of 2012, the Typhoons Haiyan or Hagpuit in 2013-2014, or the Nepal earthquake in 2015 have shown that information provided by eye-witnesses through social networking services can greatly improve situational awareness.

Alam et al., for situational awareness during cyclone emergency, proposed a social media image processing pipeline, which includes a noise filter and a damage assessment classifier [5,88,6]. The Convolutional Neural Network (CNN) applied into filtering out irrelevant image content, and the perceptual hashing technique was employed for image de-duplication. Additionally, CNN technique was also used for situational awareness during heavy rainfall by Li et al [74]. They focused on using social remote sensing data for emergency response. The classification results obtained for the central parts of Wuhan and Shenzhen demonstrated the effectiveness of the CNN method considered for monitoring the heavy rainfall event that happened in both cities.

Shen et al. proposed a method to retrieve events based on event-specific hashtags preliminary to collected for situational awareness of emergencies [120]. In their experiments, the SVN showed best performance for extracting and classifying hashtags from data in Twitter. Then, the hashtags were used to collect relevant messages from not only Twitter but also other social media platforms. The SVM was also introduced to extract features and classify texts [104]. Raginia et al. proposed a hybrid method for segregating and classifying the texts obtained from people who are at risk in the affected region for situational awareness of emergency. The results showed that the text classification algorithm can help the emergency responders to locate the people at risk in real-time.

For understanding situations in disaster response, Li et al. proposed a domain adaptation approach, which learns classifiers from un-labeled target data, in addition to unlabeled data [73]. A Naive Bayes (NB) classifier and an iterative self-training strategy were adopted for tweet classification. Their experiment results showed that the domain adaptation classifiers are better as comparing with the supervised learning using only labeled data. Ramchurn et al. proposed an emergency management system called HAC-ER for situational awareness from large streams of reports posted by members of the public and trusted organizations [105]. They combined the independent Bayesian classifier combination with the Gaussian process to remove errors and to predict locations of events in affected areas by emergencies. Additionally, Imran et al. presented human-annotated Twitter corpora collected during 19 different crises and compared supervised learning techniques such as the SVM, the NB and the RF in terms of the utility of annotations [55]. Given the complexity of the multiclass classification of short messages, it was indicated that all three classifiers have decent results. To discover important topics from Twitter and provide useful information of situation awareness during emergency, Yin et al developed an online incremental clustering algorithm that automatically groups similar tweets into topic clusters [138]. They also adapted

optimization techniques (i.e., burst detection, text classification, online clustering and geotagging) to deal with real-time, high-volume text streams. It includes an early indicator identification of unexpected incidents, an impact exploration of events and an incidents evolution monitoring.

In general, to classify social media data may be tedious and time consuming task, since the collected data are not in the form of a labeled data. Therefore, Pandey and Natarajan utilized the semi-supervised machine learning approach to avoid the classification process and concurrently obtain useful information in situational awareness [95]. In addition, they also introduced an interactive map to grasp the vulnerable areas during an emergency. A reinforcement learning technique was also introduced to map dynamic situations in emergency. Sadhu et al. proposed a Multi-Agent Reinforcement Learning (MARL) framework implemented as a mobile application and a back-end server [110]. Via both simulations and real experiments, an evaluation of the framework in terms of effectiveness in tracking random dynamicity of the environment was performed.

2.6 Emergency Evaluation

Trekin et al. applied CNN into developing a method of change detection on remote sensing imagery to improve time efficiency of assessment of damaged buildings in disaster affected area [129]. Also, a deep learning-based framework for rapid regional tsunami damage recognition using post-event on synthetic radar imagery was proposed [15]. They applied the SqueezeNet network (as a CNN type) architecture into a selection algorithm, and a recognition algorithm with a modified wide residual network was developed to classify the damaged regions. Via experiments on Tohoku earthquake in 2011 and tsunami area, it was showed that the proposed framework is fast in model training and prediction calculations. The potential of CNN features was also explored for an online classification of satellite image to detect structural damages by Vetrivel et al [131]. A feature extraction and a classification process are carried out at an object level, where the objects are obtained by over-segmentation of satellite images. The proposed framework outperformed a batch classifier with lesser time and memory requirements. As other usage case of CNN for natural emergency evaluation, analysis of images posted on social media platforms using the CNN was proposed [89]. Experimental results indicated that the domain-specific fine-tuning of deep CNN outperforms Bag-of-Visual-Words (BoVW). In addition, high classification accuracy under both event-specific and cross-event test settings demonstrated that their approach can effectively adapt deep-CNN features to identify the severity of destruction from social media images taken after a disaster strikes. Additionally, Attari et al. introduced a Nazr-CNN, a deep learning pipeline for an object detection and fine-grained classification in aerial images for assessing and monitoring damage [10]. In their work, a hidden layer of a CNN was used to encode the popular BoVW of the segments generated from the first component in order to help discriminate between different levels of damage. Moreover, backpropagation neural network as a kind of multilayer feed-forward network was used in evaluation of city emergency management system for disaster event by Jiang and Li [57].

Cervone et al. proposed a methodology that leverages data harvested from social media for collecting remote-sensing imagery during disasters [26,27]. The

images are then fused with multiple sources for the damage assessment of transportation infrastructure. In this method, DT was used to classify entire scenes acquired. They also evaluated the proposed methodology with considering Colorado floods in 2013 [27]. Zhang and his colleagues proposed a machine learning framework to assess post-earthquake structural safety [143]. In this framework, Classification and Regression Tree (CART) and RF were implemented to map damage patterns to classified structural safety states. For assessment of sensitive area for landslide at the Pauri Garhwal in India, the RF and CART were also compared with Logistic Model Trees (LMT) and Best First Decision Trees (BFDt) [99]. The results showed that a RF model has the highest predictive capability followed by the LMT, BFDt and CART models, respectively. It was showed that although all four methods have shown good results, the performance of the RF method was the best for landslide spatial prediction.

An assessment model based on RF was adopted to evaluate regional flood hazard [133]. The risk assessment method was implemented in Dongjiang River Basin, China. In addition, SVM technique was used for risk assessment as a comparison, as well as an analysis of index importance degree. The spatial distributions of the RF and the SVM-based assessment maps showed a similar correlation coefficient, was indicated that the classification capacity of the two methods is similar in the majority of cases. Joshi et al. introduced a methodology for detection of damage post disasters by examining the textural features from high resolution aerial imagery [59]. The proposed technique considered DT, NB, SVM, RF, Voting Classifier and Adaptive Booster, and were compared to identify damaged regions from aerial images using only pre-event images as the input. As a result, the RF-based classifier comparatively had higher accuracy than other classifiers. Yoon and Jeong applied Cubist and RF techniques into assessment what vulnerability indicators are statistically associated with disaster damage in Korea and found twelve indicators to evaluate vulnerability of 230 local communities to disasters [139].

Zanini et al. proposed a procedure based on Fuzzy Logic (FL) for the evaluation of interactions between existing buildings and urban roadway networks after a seismic event [141]. The methodology was applied to the Municipality of Conegliano in Italy in the potential seismic damage scenario. Their experiments showed it is able to evaluate the network link functionality reductions caused by building damages, through the estimation of the residual road width, without the necessity of carrying out expensive and detailed surveys on the analysed area. Izadi et al. proposed a neuro-fuzzy approach based on GA and SVM for the semi-automatic detection and assessment of damaged roads in urban areas using pre-event vector map and both pre and post-earthquake QuickBird images [56]. Experimental results showed the efficiency and accuracy of the Neuro-Fuzzy systems for road damage assessment. Resch et al. introduced an approach based on analyse social media posts to assess footprints and the damage caused by natural disasters through combining LDA for semantic information extraction with spatial and temporal analysis for hot spot detection [107]. Furthermore, they provided a damage map that indicates where significant losses have occurred. Their experiments showed that earthquake footprints can be reliably and accurately identified in our use case. Nadi and Edrisi introduced a Markov decision process as a multi-agent assessment and a response system with reinforcement learning designed to ensure the integration of emergency response and relief assessment operations [84]. Experiments indicated that

the use of the proposed approach in assessing network conditions and true demand during search and rescue operations can decrease death tolls.

2.7 Crowdsourcing

Volunteers provide information and resources to the affected people and this process has been facilitated by social media in recent years [87]. In this regard, for some years now, both researchers and practitioners in the areas of emergency management have been exploring the role of crowdsourcing in collecting, processing and sharing information [101]. Although there are various roles of crowdsourcing in tasks of emergency management, in this section, we focus on its usages in tasks for post-emergency, especially related relief activities. The others will be investigated and discussed in Section 5. First we start with reviewing studies focused on crowdsourcing for post-emergency without considering machine learning techniques.

Landwehr and Carley have reviewed how social media is used in disaster by individuals, first responders and disaster researchers. They have also introduced a variety of software tools that can be used by analysts to work with social media and have discussed of several different directions in which some of the research on social media usage in disaster is currently heading [69]. As aiming at efficiently harnessing crowdsourcing in remote assistance in real-time, Yang et al. designed and developed a crowdsourcing disaster support platform [136]. They considered three unique features as follows: selecting and notifying individual requests, providing collaborative working functionalities, improving answer credibility through “crowd voting.” In addition, Dubey et al attempted to develop a theoretical framework which can assist relief activities using valuable information derived using comprehensive crowdsourcing framework in environments with Internet of Things [35]. They have conducted extensive review of articles published in reputable journals, magazines and blogs by eminent practitioners and policy makers. Murali et al. proposed a multi-platform model to deal with disasters and support relief activities while handling the needs of victims, volunteers and government agencies [83]. Further, they used various techniques, ranging from Natural Language Processing (NLP) to crowdsourcing, for ensuring robustness and scalability of solution.

Various techniques of machine learning have been also combined with crowdsourcing to support relief activities in post-emergency. Most of the current systems allow volunteers to directly provide input to them [86]. Hence, many social media posts in the aftermath of disasters might contain useful information. There are several tries to extract the information from social media through a variety of machine learning techniques such as LDA and DT [86], RF and NB classifier [103]. In order to detect potential incidents implicated by victims negative emotions in the post-disaster situation, Bai et al. introduced a structured framework including three phases [14]. NB, RF, SVM and KNN techniques were compared each other in terms of to classify emergency-related messages; as a result, the RF and the SVM were outperformed than the others. Harris et al. proposed an approach of the post-phase situational awareness of an earthquake hit area to the rescue task [42]. To ensure the credibility of the crowdsourced data, their system considers the K-means clustering technique and maps coordinates of the calamity area through a short messaging service. In addition, experimentation was carried out to evaluate

the time taken to notify via SMS. Imran et al. also applied a clustering technique to the classification of crisis-related messages in microblog streams [52]. Social media messages are clustered together with the textual similarity, and human curators annotated larger clusters first to train the classifier. Liang et al. introduced a semi-supervised learning based cognitive framework to support emergency management through mapping crowdsourced data. The framework first divides the satellite or aerial image into patches leveraging a graph-based clustering approach [76]. The KNN classifier is then used to provide labels for a few patches. With over 50 participants working on three different tasks, their experiments showed that the crowdsourced variant performs well producing noise-tolerant flood maps.

3 Analysis of Emergency Data

Information exchange during and after the disaster periods can greatly reduce the losses caused by the disaster. This is because it allows people to make better use of the available resources and provides a channel through which reports on casualties and losses in each affected area can be delivered expeditiously [75]. Furthermore, the success of a disaster relief and response process is largely dependent on timely and accurate information regarding the status of the disaster, the surrounding environment and the affected people [87]. Therefore, understanding, analytic and utilization of data collected in emergency are vital. In this section, we describe data characteristics generated during disaster with considering the 5 Vs of big data and various application cases of Big Data Analysis (BDA) are looked over.

3.1 Big Data in Emergency Management

The role of data in emergency management has been evolving. Nowadays, scientists are facing one of the biggest challenges of managing large volumes of data generated at times of disasters. As a huge amount of emergency-related data is getting generated, traditional data storage and processing systems are facing challenges in fulfilling performance, scalability and availability [39]. Therefore, analytic methods to manage and process data in emergency management are particularly challenged due to the combination of its unique characteristics as follows:

- **Rapid increase of data** by many numbers of producers and consumers
- **The timely sensitivity** of detect and response
- **Combination of static and dynamic data** (e.g., maps and crowd emotion) [76,116]
- **Heterogeneous formats**, ranging from raw data (e.g., sensors) to structured data (e.g., metadata) and unstructured data (e.g., multimedia) [101]
- **Various levels of trustworthiness** of the data sources [126,47]
- **Possibility of extracting valuable information** like crowdsourcing, generated by people who are actually at the emergency scene for near real-time [75]

These characteristics similar to “5 Vs (i.e., volume, variety, velocity, veracity and value)” of big data. Besides, within the present-day emergency management systems, the immediate and accurate decision-making more and more relies on the

capability of data analysis and processing especially in the face of big data [29]. Therefore, there is an urgent need to enhance the BDA technologies of emergency management, such as, to develop scalable and real-time algorithms for time-sensitive decisions, to integrate structured, unstructured and semi-structured data, to deal with the imprecise and uncertain information, to extract dynamic patterns and outline the evolution of these patterns, to work in distributed environment and to present the multi-scale, multilevel and multi-dimensional patterns through various visualization approaches [8].

BDA was often defined as holistic process to process and analyse the 5 Vs in order to create actionable insights for sustained competitive advantages [36]. In this regard, Mehrotra et al. suggested that BDA can aid to create the next generation of emergency management technologies as it has the potential to mitigate the effects of disasters by enabling access to critical information in real-time [80]. Also, they emphasized that “accurate and timely analysis and assessment of the situation can empower decision makers during a crisis to make more informed decisions, take appropriate actions and better manage the response process and associated risks.” Thus, it is essential to reconsider how data on disasters should be properly and efficiently produced, organized, stored and analysed [44]. Here, we briefly discuss the BDA for emergency management including data collection, information extraction, data filtering and data integration.

3.2 Data Collection

Data collection is the process of inserting to a system data, which is coming from multiple heterogeneous sources. Scalability is an important issue in data collection, since the flow of information may be very high during the time of a critical event [49]. Traditionally, it was done in the form of paper reports or questionnaires. The development of IT allowed for use of word processors, spreadsheets and forms to enter data directly into the databases [140]. Moreover, sensing technologies are also currently undergoing rapid advances and are leading to might use of them significantly increases performance of situational awareness [19]. As another increasingly important source, social media is a new way of communication in the course of disasters. A major difference between social media and traditional sources is the possibility of receiving feedback from the affected people [87]. Additionally, for some years now, both researchers and practitioners in the areas of disaster and emergency management have been exploring the role of crowdsourcing in collecting, processing and sharing information across organizations and affected populations [101].

3.3 Information Extraction and Filtering

The goal of information extraction is to automatically extract structured information, i.e., categorized and contextually and semantically well-defined data from a certain domain, from unstructured machine-readable documents [49]. Whereas, according to Wikipedia, data filtering is that removes redundant or unwanted information from an data stream using (semi) automated or computerized methods prior to provide it. Its main goal is the management of the data overload

and increment of the semantic information. If all the disaster information were presented to the users, it would cause an overwhelming workload. Therefore, the disaster information should be filtered based on the specific purposes of the users [75]. Especially, in emergency management, the utilization of social media data can be extended for various tasks such as warning [11], detection & tracking [112, 102], situational awareness [86, 6], assessment [27, 107], etc. Therefore, it is more emphasized that the data extraction and filtering became the core of emergency management together with effects of social media data. However, the social media data contain texts, images, videos, tags and so on. Therefore, information is extracted from heterogeneous sources such as social media and monitoring devices. Typically, the disaster information from different sources varies greatly in structure or format. To support further analysis and processing, a specification of common format should be required for disaster information integration. Then, the integrated disaster information in this format can be organized and stored for further processing [75].

3.4 Data Integration

Emergency services sometimes should deal with the massive amount of data arriving through multiple channels such as existing records, sensors, satellite networks or social media [1]. Therefore, one of the biggest challenges in emergency management is to develop a data integration protocol. Data integration is the process of combining data residing at different sources and providing the user with a unified view of these data [72]. This process includes the following tasks: (1) to convert contents of different formats into a standard format; (2) to verify the credibility of various crowdsourcing data sources and attempt to leverage it to produce useful information for disaster decision-making; (3) to map images or texts with their corresponding geolocations to better capture the current situation; and (4) to process and analyse the data from different sources [75].

3.5 Applications for Data Analysis in Emergency

In this section, we introduce several applications and platforms comprehensively deal with data analysis for emergency managements. **Ushahidi** as the first large-scale crowdsourcing system developed to report Kenyan post-election violence in 2008 and since then has been applied into many major disasters such as Hurricane Sandy and Haiti Earthquake [91]. The Ushahidi is an open source and free systems which can either be deployed on external servers or on its hosting system CrowdMap. This system collects emergency-related data from several sources, web, Twitter, RSS feeds, emails, SMS and so on. Collected data is then visualized on the map. Last, The Ushahidi allows users to filter information based on several types (e.g. supplies or shelter).

Artificial Intelligence for Disaster Response (AIDR) is a free software platform which can be either run as a web application or created [53]. In the AIDR, tweets are collected according to pre-selected set of keywords. Prior labeled tweets will be used as the training set of a classifier which labels collected tweets based on the keywords. In the training process, n-grams of tweets are used as features

and therefore the classifier is retrained for every new category and disaster.

TweetTracker consists of tracking, analysing and understanding tweets related to a specific topic [67]. In the process of tracking the status of an event, data includes keywords, location and users can be collected using a set of criteria from Twitter, Facebook, YouTube, VK and Instagram. Fluctuations in the total number of posts or frequency of posts with specific words can be analysed for different time periods. Moreover, keywords, hashtags, links, images and videos with their frequencies are available to understand tweets by the user. For instance, to better understand the geographic distribution of posts, geo-tagged tweets are presented on a map.

DisasterMapper is a CyberGIS framework to ingest and archive massive amounts of social media data [50]. In this framework, to manage massive social media data, Apache Hive is used as the scalable storage solution. Furthermore, Hadoop platform is used as a scalable distributed computing environment to process social media data. Mahout is leveraged to support big data analytic. It can automatically synthesize multi-sourced data, such as social media and socioeconomic data, to track disaster events, to produce maps and to perform spatial and statistical analysis for emergency management.

IDDSS-Sensor is software for Geographic Information System (GIS) implemented to provide the functions of standard-based access, as well as on-the-fly harmonization, integration and usage of multi-agency sensor information [7]. The software has three layers, namely, the storage, service and presentation layer. In storage layer, PostgreSQL is used as an open-source object relational database of integration data models. For service layer, 520North and GeoServer was employed as SOS implementation and as spatial data service for serving static spatial data. As GUI of the system, third layer was developed by JavaScript and the Cesium, ExtJS3 and IDDSS were used for displaying the sensor data and visual indicators.

Table 1 lists approaches and applications in emergency management in terms of target types of emergency, belonged tasks, used technique of machine learning and data analysis.

4 Challenges and Opportunities of machine learning in Response

Now we discuss aforementioned approaches to draw challenges and opportunities of machine learning techniques in emergency management.

Information and Knowledge Learning. For this category, the following challenges are raised as further research issues: (1) Support sustainable annotation and classification on heterogeneous, streaming data from the multisource. To achieve this, effective and efficient learning algorithms must be studied considering the context of the emergency and the stakeholders belong to each task. (2) Facilitate real-time analysis and discover information across multiple streams through developing high-speed and flexible techniques. (3) Construct customized information extraction methods that can learn by the integration between domain experts and existing system.

Integration with Geographic Information System. GISs supports to integrate, store, share and display geographically referenced information. Users (e.g., stakeholders or decision maker) can understand overall situations and discover insight through GIS for emergency. In this regard, integration between a GIS and

Table 1 Approaches with Machine Learning for emergency management

Title	Emergency	Task	Machine learning technique	Data	Source	Data analysis
Prediction of Earthquake Magnitude by an Improved ABC-MLP [118]	Earthquake	Prediction	– Artificial bee colony – Multilayer perceptron – Backpropagation	– Earthquake parameter magnitude	– California earthquake data (2011) ^a	– Labeling
Artificial neural networks for earthquake prediction using time series magnitude data or Seismic Electric Signals [82]	Earthquake	Prediction	– Artificial Neural Network	– Time series magnitude data – Seismic electric signals	– VAN team in Greece Athens earthquake data (1981) – Seismological Institute National Observatory of Athens (SINOA)	– Labeling
A time-dependent surrogate model for storm surge prediction based on an artificial neural network using high-fidelity synthetic hurricane modeling [63]	Hurricane	Prediction	– Artificial neural network – Backpropagation	– Historical hurricane parameters	– Atlantic hurricane database, HURDAT2 ^b	– Labeling
Neural Networks to Predict Earthquakes in Chile [109]	Earthquake	Prediction	– Artificial neural network	– Earthquake magnitude	– Chile's National Seismological Service	– Labeling
Predicting Monsoon Floods in Rivers Embedding Wavelet Transform, Genetic Algorithm and Neural Network [111]	Flood	Prediction	– Artificial neural network – Genetic algorithm	– Flow data	– Monsoon flow (2001-07)	– Labeling
Track-pattern-based model for seasonal prediction of tropical cyclone activity in the western North Pacific [62]	Tropical cyclone	Prediction	– Fuzzy c-means clustering	– Tropical cyclone locations	– Regional Specialized Meteorological Centers Tokyo-Typhoon Center	– Labeling
The Application of Ant-Colony Clustering Algorithm to Earthquake Prediction [119]	Earthquake	Prediction	– Ant colony clustering algorithm	– Earthquake magnitude		– Labeling – Anomaly detection
Dynamic and robust wildfire risk prediction system: an unsupervised approach [115]	Wildfire	Prediction	– Clustering algorithms	– Weather data	– Fire & Rescue NSW from the Bureau of Meteorology (BoM)	– Labeling – Anomaly detection
The data mining technology of particle swarm optimization algorithm in earthquake prediction [142]	Earthquake	Prediction	– Particle swarm optimization clustering algorithm	– Earthquake magnitude		– Labeling – Anomaly detection
Spatial prediction of flood susceptible areas using rule based decision tree and a novel ensemble bivariate and multivariate statistical models in GIS [127]	Flood	Prediction	– Rule-based decision tree – Logistic regression statistical methods	– Rainfall data – Spatial database	– Landslide-hazard map [90]	– Labeling – Anomaly detection – Mapping flood susceptibilities of areas

^a <http://www.data.scec.org/>^b <http://www.nhc.noaa.gov/data/#hurdat>

Table 1 continued

Title	Emergency	Task	Machine learning technique	Data	Source	Data analysis
Clustering technique to interpret Numerical Weather Prediction output products for forecast of Cloudburst [94]	Cloudburst	Prediction	– K-means clustering	– Flow pattern data – Temperature – Atmospheric pressure level	– European Center for Medium-range Weather Forecasting	– Labeling
GIS-based landslide susceptibility modelling: a comparative assessment of kernel logistic regression, Naïve-Bayes tree, and alternating decision tree models [30]	Landslide	Prediction	– Kernel logistic regression – Naïve Bayes tree – Alternating decision tree	– Landslide inventory map – Landslide conditioning factors – Landsat 8 OLI images – Precipitation data – Soil radon data	– Self-collection – NASA ^a – U.S Geological Survey (USGS) ^b	– Collecting landslide inventory map – Labeling – To manually ingrate early reports, aerial photographs and GPSs
Application of decision trees to the analysis of soil radon data for earthquake prediction [145]	Earthquake	Prediction	– Decision tree		– Soil radon data [146]	– Labeling
A new data mining model for hurricane intensity prediction [125]	Hurricane	Prediction	– Feature weight learning – Genetic algorithm	– Historical hurricane data	– Atlantic tropical cyclones (1982-2003) from National Center for Atmospheric Research NCAR	– Labeling
Computational Intelligence Techniques for Predicting Earthquakes [79]	Earthquake	Prediction	– Extensible Markov model – Association rules – Decision tree	– Location and magnitude of earthquakes	– Spanishs Geographical Institute (SGI)	– Labeling
Novel method for hurricane trajectory prediction based on data mining [34]	Hurricane	Prediction	– Association rules	– Hurricanes motion characteristics	– Atlantic weather Hurricane/Tropical Data from 1900 to 2008 ^c	– Labeling
Spatial Data Mining for Prediction of natural events and emergency management based on fuzzy logic using hybrid PSO [106]	Natural emergency	Prediction	– Fuzzy logic – Particle swarm optimization – K-means clustering – Naïve Bayes classifier	– Water level modifications data – Spatial-temporal data – Text data	– Bangladesh Country Almanac (BCA) dataset	– Labeling and detecting centroid – Filtering inconsistent and unnecessary data – Integrating spatial-temporal data and text
Use of wireless sensor networks for distributed event detection in emergency management applications[13]	Fire	Warning	– Decision tree	– Residential fire data (i.e., temperature, ionization, photo-electric and CO)	NIST ^d	– Labeling
Earthquake emergency management by social sensing [11]	Earthquake	Warning, Evaluation	– Bayesian statistics – Peak-detection algorithms – Corrected conditional entropy – Classification algorithms	– Social media data with keywords – multimedia data (i.e., video and photo)	– Using Streaming API of Twitter	– Collecting social media data – Labeling – Removing noise of different meaning and past event

^a <http://reverb.echo.nasa.gov/reverb/>^b <http://landsat.usgs.gov/landsat8.php>, <http://www.sxmb.gov.cn/index.php>^c <http://weather.unisys.com/hurricane/>^d <http://smokealarm.nist.gov/>

Table 1 continued

Title	Emergency	Task	Machine learning technique	Data	Source	Data analysis
Evaluation of the satellite-based Global Flood Detection System for measuring river discharge: influence of local factors [108]	Flood	Warning Detection & Tracking, Evaluation	Random forest	<ul style="list-style-type: none">Surface water extent and floodplains data from satelliteRiver widthPresence of floodplains flooded forest and wetlandsFlood extentLand coverLeaf area indexClimatic areasPresence of river iceDam location	<ul style="list-style-type: none">Global Flood Detection System, Global Runoff Data Centre and South African Water Affairs^a (1998 to 2010), SRTM Water Body Database and the HydroSHEDS [135]Global Lakes and Wetlands Database level 3 [70]Global Flood Hazard Map [96]Global Land Cover 2009 [20]SPOT-VGT^bKoppen-Geiger climate map of the world [97]Circum-Arctic map [22]Global Reservoir and Dam [71]Actual Hoogtebestand Nederland^c	<ul style="list-style-type: none">LabelingIntegrating data related to flood
Flood early warning system: design, implementation and computational modules [66]	Flood	Warning	K-means clustering	<ul style="list-style-type: none">Sensor stream dataDigital terrain model data	<ul style="list-style-type: none">Collecting sensor data in real-timeLabeling and detecting anomalyIntegrating map and sensor data	<ul style="list-style-type: none">Collecting sensor data in real-timeLabeling and detecting anomalyIntegrating map and sensor data
MyShake: A smartphone seismic network for earthquake early warning and beyond [64]	Earthquake	Warning	Artificial neural network	<ul style="list-style-type: none">Acceleration and GPS data from android phoneEarthquake data	<ul style="list-style-type: none">Self-collection, Northern and Southern California Earthquake Data Centers^dCenter for Engineering Strong Motion Data^eNational Research Institute for Earth Science and Disaster Prevention for Japanese^f	<ul style="list-style-type: none">Collecting smartphone sensorLabelingFiltering frequency bandIntegrating locations and acceleration according to time-stamps
Monitoring of long-term static deformation data of Fei-Tsu arch dam using artificial neural network-based approaches [61]	Flood	Warning	Artificial neural network (i.e., static neural network, dynamic neural network and auto-associate network)	<ul style="list-style-type: none">Residual deformation in Fei-Tsu arch dam (i.e., static deformation, water level and temperature distribution)	<ul style="list-style-type: none">Collecting static deformation, water level and temperature distribution	<ul style="list-style-type: none">Collecting static deformation, water level and temperature distributionLabeling

^a <http://www.dwa.gov.za/>^b <http://wdc.dlr.de/>^c <http://www.ahn.nl/>^d <http://www.ncedc.org/>^e <https://www.strongmotioncenter.org/>^f <http://www.bosai.go.jp/e/>

Table 1 continued

Title	Emergency	Task	Machine learning technique	Data	Source	Data analysis
Earthquake management: a decision support system based on natural language processing [37]	Earthquake	Warning	<ul style="list-style-type: none"> Bayesian model averaging Voting mechanism Classifiers (i.e., decision tree, support vector machine, naive Bayes, logistic regression and K-nearest neighbors) 	<ul style="list-style-type: none"> Social media data with keywords 	<ul style="list-style-type: none"> Using Streaming API of Twitter 	<ul style="list-style-type: none"> Collecting social media data Labeling
Development of ERESS in panic-type disasters: Disaster recognition algorithm by buffering-SVM [81] Disaster detection by statistics and SVM for emergency rescue evacuation support system [45]	Natural emergency	Detection & Tracking	<ul style="list-style-type: none"> Support vector machine 	<ul style="list-style-type: none"> Sensor data from smart phone 		<ul style="list-style-type: none"> Collecting sensor data from smart phone Extracting feature and labeling
Social media for crisis management: clustering approaches for sub-event detection [102]	Emergency	Detection & Tracking	<ul style="list-style-type: none"> Self-organizing map Agglomerative Clustering 	<ul style="list-style-type: none"> Social multimedia data (i.e., video and photo) 	<ul style="list-style-type: none"> Using APIs of Flickr^a and Youtube^b for 2011 Mississippi Flood, Oslo Bombing, UK Riots and Hurricane Irene 	<ul style="list-style-type: none"> Collecting social multimedia data Labeling Integrating spatial-temporal data, images and videos
Classifying text messages for the Haiti earthquake [24]	Earthquake	Detection & Tracking	<ul style="list-style-type: none"> Support vector machine Bag-of-words Feature abstraction Latent Dirichlet allocation 	<ul style="list-style-type: none"> Text messages in phone, e-mail, Twitter and web 	<ul style="list-style-type: none"> Ushahidi-Haiti^c 	<ul style="list-style-type: none"> Labeling Selecting feature to remove irrelevant and redundant feature
Prediction of human emergency behaviour and their mobility following large-scale disaster [122]	Natural emergency	Detection & Tracking	<ul style="list-style-type: none"> Hidden Markov model Bayesian filtering Maximum entropy inverse 	<ul style="list-style-type: none"> Human mobility database (GPS records) Disaster intensity data (i.e., seismic scale and damage level) Disaster reporting data 	<ul style="list-style-type: none"> The Great East Japan Earthquake and Fukushima nuclear from Japan Government statistical reports Government declarations Japanese Cabinet Secretariat^d News reports [46] 	<ul style="list-style-type: none"> Collecting GPSs of human Labeling GPSs Integrating map and human movements
Event classification and location prediction from tweets during disasters [121]	Emergency	Detection & Tracking	<ul style="list-style-type: none"> Markov chain Support vector machine Gradient boosting Random forest 	<ul style="list-style-type: none"> Social media data with geotags according to keywords 	<ul style="list-style-type: none"> Using Streaming API of Twitter 	<ul style="list-style-type: none"> Collecting social media data Labeling Filtering noise and redundancy

^a <http://flickrj.sourceforge.net/>^b <https://developers.google.com/gdata/>^c <http://haiti.ushahidi.com>^d <http://www.cas.go.jp/>

Table 1 continued

Title	Emergency	Task	Machine learning technique	Data	Source	Data analysis
Constructing a Hidden Markov Model based earthquake detector: application to induced seismicity [17]	Earthquake	Detection & Tracking	– Hidden Markov model	– Geothermal data for geothermal reservoir for volcano	– Geothermal plant in the municipality of Unterhaching in Germany (2008-2010)	– Labeling
Earthquake shakes Twitter users: real-time event detection by social sensors [112]	Earthquake	Prediction, Warning, Detection & Tracking	– Support vector machine	– Tweet feeds	– Using Streaming API of Twitter	– Filtering noise on sound recording
Tweet analysis for real-time event detection and earthquake reporting system development [113]	Earthquake		– Kalman filtering	– GPS and registered location of user		– Labeling
Leveraging cross-media analytics to detect events and mine opinions for emergency management [134]	Natural emergency	Detection & Tracking	– Particle filtering			– Mapping emergency event and tweet feeds
			– Clustering techniques	– Microblog data		– Collecting social media data
						– Labeling
						– Mapping emergency event and tweet feeds
						– Crawling microblog from online news and forums
						– Extracting relevance key-words
						– Filtering document by sentiment analysis
						– Integrating cross-media data
						– Labeling
Audio Data Mining for Anthropogenic Disaster Identification: An Automatic Taxonomy Approach [137]	Man-made Emergency	Detection & Tracking	– K-means dictionary learning	– Audio clips from four channel	– BBC Sound Effects Library	– Labeling
			– Clustering technique based on hierarchical regularized logistic regression model		– Urban-SoundSK datasets [114]	– Labeling
			– Convolutional neural network		– Sound Classification [100]	– Labeling
					– Sound effects from internet sources ^a	– Labeling
Social Media: New Perspectives to Improve Remote Sensing for Emergency Response [74]	Natural emergency	Situational awareness		– Social media data of Weibo with key-words (in Chinese)	– Using Sina Weibo open platform API for Weibo data in Wuhan (2016) and Shenzhen (2014)	– Collecting social media data
Disaster response aided by tweet classification with a domain adaptation approach [73]	Natural emergency	Situational awareness	– Naive Bayes classifier	– Social media data with hashtags	– CrisisLexT6 dataset [92]	– Labeling
Using social media to enhance emergency situation awareness [138]	Natural emergency	Situational awareness	– Iterative self-training strategy			
			– Incremental clustering algorithm	– Social media data	– Using Streaming API of Twitter (2010)	– Collecting social media data
How social media can contribute during disaster events? Case study of Chennai floods 2015 [95]	Natural emergency	Situational awareness	– Semi-supervised classification	– Social media data with hashtags and geotags according to keywords	– Using Streaming API of Twitter (2015)	– Labeling
				– Web Map service and satellite data		– Filtering stop words
				– Social media data with spatial-temporal data		– Collecting social media data
						– Labeling
						– Integrating map and tweet feeds
Post earthquake disaster awareness to emergency task force using crowd-sourced data [42]	Earthquake	Situational awareness, Crowd-sourcing	– K-means clustering		– Using Streaming API of Twitter	– Collecting social media data
						– Labeling
						– Mapping affected areas by disaster

^a <http://sound.natix.org/>

Table 1 continued

Title	Emergency	Task	Machine learning technique	Data	Source	Data analysis
Mining crisis information: A strategic approach for detection of people at risk through social media analysis [104]	Natural emergency	Situational awareness	Support vector machine	Social media data with hashtags and geotags	Using Streaming API of Twitter Follow hashtag ^a Test dataset [93]	Collecting social media data Labeling
Image4Act: Online Social Media Image Processing for Disaster Response [5]	Natural emergency	Situational awareness, Crowd-sourcing	Convolutional neural networks Perceptual Hashing	Social media data (i.e., image and geotags) according to keywords	Using Streaming API of Twitter	Collecting social media data Labeling Filtering irrelevant and duplicate images
Information retrieval of a disaster event from cross-platform social media [120]	Natural emergency	Situational awareness	Support vector machine	Social media data with hashtags	Using Streaming API of Twitter English language disaster datasets [51, 92, 93]	Collecting social media data Labeling Filtering noise Integrating cross-social media data
Twitter as a Lifeline: Human-annotated Twitter Corpora for NLP of Crisis-related Messages [55]	Natural emergency	Situational awareness	Naive Bayes Support vector machine Random forest	Social media data	Using Streaming API of Twitter for 19 different crises (2013-2015)	Collecting social media data Labeling Filtering stop words, URLs and user-mentions
Argus: Smartphone-enabled human cooperation via multi-agent reinforcement learning for disaster situational awareness [110]	Natural emergency	Situational awareness	Multi-agent reinforcement Learning (i.e., modified version of distributed Q-learning [78])	Multimedia data (i.e., audio and video)	Using a mobile application implemented by them-self	Collecting audio and video from user smartphone Labeling Generating 3D map Survey evaluation data Labeling
The evaluation of city emergency management system results based on BP neural network [57]	Natural emergency	Evaluation	Backpropagation neural network	Results evaluation of City emergency management system		
Damage assessment from social media imagery data during disasters [89]	Natural emergency	Evaluation	Convolutional neural network	Social media data with hashtags and keywords Google images data	Using AIDR [52] Typhoon Ruby/Hagupit (2014), Nepal Earthquake (2015), Ecuador Earthquake (2016) and Hurricane Matthew (2016)	Collecting images from web and social media Labeling
Nazr-CNN: Fine-Grained Classification of UAV Imagery for Damage Assessment [10]	Natural emergency	Evaluation	Nazr-convolutional neural network	Unmanned Aerial Vehicles (UAV) imagery	Google search Cyclone Pam in Vanuatu (2015) by the UAViators ^b	Collecting UAV imagery Labeling Mapping emergency damages
Automatic Image Filtering on Social Networks Using Deep Learning and Perceptual Hashing During Crises [88]	Natural emergency	Evaluation	Convolutional Neural Networks Perceptual hashing	Images from Twitter	Using AIDR [51] for Typhoon Ruby (2014), Nepal Earthquake (2015), Ecuador Earthquake (2016) and Hurricane Matthew (2016)	Collecting social media data Labeling Filtering duplicates and relevance
A machine learning framework for assessing post-earthquake structural safety [143]	Earthquake	Evaluation	Regression tree Random forest	Response data of engineering demand	Nonlinear response history analysis	Labeling

^a <http://www.followthehashtag.com/features/twitter-historical-data-recover/>^b <http://uaviators.org/>

Table 1 continued

Title	Emergency	Task	Machine learning technique	Data	Source	Data analysis
Processing Social Media Images by Combining Human and Machine Computing during Crises [6]	Natural emergency	Evaluation	<ul style="list-style-type: none"> Convolutional neural networks Perceptual hashing Transfer learning 	<ul style="list-style-type: none"> Images from Twitter, Flickr, Instagram and so on. 	<ul style="list-style-type: none"> Using Streaming API of Twitter for Typhoon Ruby (2014), Nepal Earthquake (2015), Ecuador Earthquake (2016) and Hurricane Matthew (2016) 	<ul style="list-style-type: none"> Collecting social media data Labeling Filtering de-duplication and relevancy Integrating cross-social media data Collecting UAV imagery and tweet feeds Labeling Mapping emergency damages
Using Twitter for tasking remote-sensing data collection and damage assessment: 2013 Boulder flood case study [26]	Natural emergency	Evaluation	<ul style="list-style-type: none"> Decision tree 	<ul style="list-style-type: none"> Social media data in Twitter UAV imagery 	<ul style="list-style-type: none"> Using Streaming API of Twitter Falcon UAV 	<ul style="list-style-type: none"> Collecting UAV imagery and tweet feeds Labeling Mapping emergency damages
Damage identification and assessment using image processing on post-disaster satellite imagery [59]	Natural emergency	Evaluation	<ul style="list-style-type: none"> Decision tree Naive Bayes Support vector machine Random forest Voting classifier Adaptive booster 	<ul style="list-style-type: none"> Aerial images 	<ul style="list-style-type: none"> Using GeoEye1 for the Christchurch earthquake (2011) and Japan earthquake and tsunami (2011) 	<ul style="list-style-type: none"> Collecting images Labeling Filtering noise Mapping emergency damages
Application and comparison of decision tree-based machine learning methods in landslide susceptibility assessment at Pauri Garhwal Area, Uttarakhand, India [99]	Landslide	Evaluation	<ul style="list-style-type: none"> Random forest Logistic model tree Best first decision trees Regression tree 	<ul style="list-style-type: none"> Landslide and Non-landslide Data 	<ul style="list-style-type: none"> Interpretation of satellite and Google Earth images 	<ul style="list-style-type: none"> Collecting images Labeling Mapping emergency damages
A New Neuro-Fuzzy Approach for Post-earthquake Road Damage Assessment Using GA and SVM Classification from QuickBird Satellite Images [56]	Earthquake	Evaluation	<ul style="list-style-type: none"> Neurofuzzy inference Support vector machine Genetic algorithm 	<ul style="list-style-type: none"> QuickBird (high-resolution satellite images) 	<ul style="list-style-type: none"> Bam in Iran (2003) 	<ul style="list-style-type: none"> Collecting images Labeling mapping damages
Combining machine-learning topic models and spatio-temporal analysis of social media data for disaster footprint and damage assessment [107]	Earthquake	Evaluation	<ul style="list-style-type: none"> Latent Dirichlet allocation 	<ul style="list-style-type: none"> Social media data Earthquake footprints (Experiment) 	<ul style="list-style-type: none"> Using Streaming API of Twitter Napa earthquake (2014) Official earthquake footprint by USGS 	<ul style="list-style-type: none"> Collecting social media data for specific time and place Labeling Filtering stop words Unique words, stemming, etc. Mapping emergency damages
Satellite imagery analysis for operational damage assessment in Emergency situations [129]	Natural emergency	Evaluation	<ul style="list-style-type: none"> Convolutional neural network 	<ul style="list-style-type: none"> Two aerospace images for before and after emergency 	<ul style="list-style-type: none"> Two areas of Ventura and Santa Rosa counties(2017) from Digital-globe within Open Data Program^a 	<ul style="list-style-type: none"> Labeling Mapping emergency damages
A framework of rapid regional tsunami damage recognition from post-event TerraSAR-X imagery using deep neural networks [15]	Tsunami	Evaluation	<ul style="list-style-type: none"> Convolutional neural network 	<ul style="list-style-type: none"> TerraSAR-X data (i.e., high-resolution synthetic aperture radar) 	<ul style="list-style-type: none"> The Pacific coast of the Miyagi prefecture (2011) from the Pasco Corporation 	<ul style="list-style-type: none"> Collecting data in the StripMap mode Labeling Mapping damages

^a <https://www.digitalglobe.com/opendata>

Table 1 continued

Title	Emergency	Task	Machine learning technique	Data	Source	Data analysis
TweetTracker: An Analysis Tool for Humanitarian and Disaster Relief [67]	Natural emergency	Tracking, Situational awareness, Crowdsourcing	Various machine learning techniques	Various social data	Twitter, Facebook, YouTube, VK and Instagram	Collecting social media data Labeling Filtering noise Mapping emergency information
Emergency-relief coordination on social media: Automatically matching resource requests and offers [103]	Natural emergency	Crowdsourcing	Random forest Naive Bayes multinomial	Social media data with keywords and hashtags	Using Streaming API of Twitter	Collecting social media data for specific time and place Labeling
Finding requests in social media for disaster relief [86]	Natural emergency	Crowdsourcing	Latent Dirichlet allocation Decision tree	Social media data Request data (Experiment)	Using Streaming API of Twitter Request data [103]	Collecting social media data for specific time and place Labeling Removing duplicate tweets, re-tweets, etc. Mapping request tweet feeds
A Weibo-based approach to disaster informatics: incidents monitor in post-disaster situation via Weibo text negative sentiment analysis [14]	Natural emergency	Crowdsourcing	Random forest Naive Bayes Support vector machine	Social media data	Using Sina Weibo open platform API for Yushu earthquake (2010), Beijing rainstorm (2012) and Yuyao flood (2013)	Collecting social media data Labeling Filtering messages of positive and neutral emotion
Ushahidi, or testimony: Web 2.0 tools for crowd-sourcing crisis information [91]	Natural emergency	Warning, Tracking, Crowdsourcing	Various machine learning techniques	Web, Twitter, RSS feeds, emails, SMS, etc.	Using Streaming API of Twitter	Collecting social media data Labeling Filtering noise Mapping tweet feeds
DisasterMapper: A CyberGIS framework for emergency management using social media data [50]	Natural emergency	Crowdsourcing	Various machine learning techniques by Malout	Social media data	Using crawler based on Twitter4j ^a	Collecting social media data Labeling Mapping tweet feeds
Identifying valuable information from twitter during natural disasters [130]	Natural emergency	Situational awareness, Crowdsourcing	Naive Bayes classifiers	Social media data	Using Streaming API of Twitter	Collecting social media data Labeling Removing unnecessary data such as URLs, slang, etc.
Data mining Twitter during the UK floods: Investigating the potential use of social media in emergency management [124]	Flood	Situational awareness, Crowdsourcing	Naive Bayes classifiers	Social media data	Using Streaming API of Twitter	Collecting social media data Labeling Filtering noise
Crowdsourcing Incident Information for Disaster Response Using Twitter [68]	Natural emergency	Crowdsourcing	Naive Bayes classifiers Support vector machine	Social media data with specific keywords	Twitter dataset of Hurricane Sandy [147]	Labeling Mapping damages
#ChennaiFloods: Leveraging Human and Machine Learning for Crisis Mapping during Disasters Using Social Media [9]	Flood	Crowdsourcing	Smooth support vector machine	Social media data with specific hashtags	Using Streaming API of Twitter for a hashtag '#ChennaiFloods'	Collecting social media data Labeling Filtering stop words, stemming, etc. Mapping tweet feeds

^a <http://twitter4j.org/en/index.html>

other components is an important research. There are worth further investigation as follows: (1) automate or crowdsource the linkage construction between information/data and geo-map in real-time. For achieve this, as a prior work, automatic location extraction techniques from disaster data obtained or collected are essential for the real-time processing. (2) Intelligent alerts and location broadcasting when people enter a dangerous area. In this case, sensor-based or vision-based approaches can be integrated as the anomaly detection techniques.

Emergency Data Analysis. As aforementioned, data in emergency are in general generated from various sources and are heterogeneous in nature. Therefore, effective and efficient methods for data analysis in emergency management should consider discovering the inter-dependencies of data and extracting useful information and knowledge. An additional challenge is to integrate data with great diversity which may be caused from heterogeneous sources with different levels of redundancy, accuracy and uncertainty or may be due to different characteristics of data (e.g., structured/unstructured, real-time streams/static data). In this regards, some interesting research directions include: (1) a unified method for each specific algorithms to collect, extract, filter for heterogeneous, multisource disaster data; and (2) building an analysis method capable of real-time processing.

In addition, we discuss issues in crowdsourcing with the machine learning for emergency management in terms of data analysis such as data collection, information extraction, data filtering and data integration.

4.1 Data collection

Scalability issues. Large crises often generate an explosion of social media activity. In case of Twitter, although each message contains 140 characters, is around 4KB by considering the metadata. Furthermore, the significant amount of storage space may be required by attached multimedia objects such as images and videos. Data velocity is a more challenging issue, especially with considering frequent occurring drastic variations. The largest peak of tweets during a natural hazard was measured as 16,000 tweets per minute ¹. Finally, redundancy is in general cited as a scalability challenge. Repeated (e.g., retweets) messages are common in time-sensitive social media, even un-trusted tweet such as rumours and spam might gain more concern due to simply repeating more.

Content issues. Even although microtexts are brief and informal, to analysis this type of text is difficult work due to complexity with technological, cross-lingual and cross-cultural factors. This causes severe challenges to computational methods and can lead to poor and misleading results. Additionally, the texts are also highly heterogeneous with multiple sources and varying levels of quality. Quality itself is important question, encompassing many attributes including objectivity, clarity, timeliness, conciseness and so on.

¹ During Hurricane Sandy in 2012: <http://www.cbsnews.com/news/social-media-a-news-source-and-toolduring-superstorm-sandy/>.

4.2 Information extraction

Inadequate spatial information. As spatial and temporal information, are two components of an event, most systems encounter challenges to determine geographical information of social media that lack GPS information. In this case, additional information (e.g., geo-tag and locations in user profile) can be used.

Combining manual and automatic annotation. In a supervised learning setting, data labeled through manual works is necessary to training a model, but it may be costly to obtain. This is particularly problematic in emergency that attract concerns of a multilingual population, or for tasks that require domain knowledge, related to affected region or characteristics of emergency events. Also, labeled data are not always reliable and may not be available at the time of the emergency. In this case, a hybrid approach that combines human and automatic annotation can be used. An active learning with the selection of items to be labeled by humans can be applied to improve classification accuracy as new labels are received.

4.3 Data filtering

Mundane events. People post specific events as well as daily life on social media sites. These data as noise, which creates more challenges for event detection methods to overcome, should be separated of real-life big events like emergencies.

Rumours, spam and social bots. Filtering social media data in crowdsourcing is a necessary process before data usage in any stages of a disaster. Data which is overwhelmed with unwanted content (e.g., rumours, spam and content created by social bots) does not show real opinion of the crowd. To overcome challenge by rumours, methods have been proposed to automatically detect tweets by using their specific behaviours such as the difference of diffusion process of rumours and normal posts, the number of users and the depth related the diffusion, etc. In case of spam, characteristics of spammers can be used, such as posting numerous messages by one account and the few numbers of reciprocal connections. Specifically for bot, three major methods have been proposed through manual annotation, using the suspension mechanism of social media sites, or creating lure bots.

4.4 Data integration

Describing the events. Creating descriptions or labeling for a detected event is in general a challenge task. Although major keywords that are frequently posted during the event are presented as a description, it does not constitute a grammatically well-formed. In addition, with other useful data such as maps, images and video, even this issue will be more complex, but it may be practical and helpful to understand event on the whole.

Domain adaptation. Simply reusing an existing classifier trained on past data does not perform well in practice, as it yields a significant loss in accuracy even when emergencies have a lot of common elements. In machine learning, domain adaptation is a series of methods adapting it to continue to fulfil well on a dataset with different characteristics. Furthermore, this techniques may help to integrate a variety of data from different domain to supplement weaknesses each other. For

instance, briefly, sensor-based approach that provides more detail situation can be integrated with crowdsourcing-based methods, which support to detect event in broad area, for fast and accurate event detection for time-sensitive tasks in emergency.

From these challenges, several interest topics are able to be raised as opportunities such as:

- Annotation: What is the best way to collect and aggregate labels for unlabeled data from the crowd? How can we do the annotation in the most cost-efficient manner? What is the most effective way to collect probabilistic data from the crowd? How can we collect data requiring global domain knowledge via crowdsourcing?
- Time-sensitive and complex tasks: How can we design crowdsourcing systems to handle (near) real-time or time-sensitive tasks? How can we deal with work dependencies requiring more complexity?
- Data collection for specific domains: How can machine learning researchers apply the crowdsourcing principles for different domains where privacy and original characteristics are at play?
- Reliability, efficiency and scalability of a system: How can we deal with sparse, noisy and large number of label classes such as tagging images for Deep Learning based computer vision algorithms? How can we efficiently apply a series of useful methods (e.g., optimal budget allocation, label aggregation algorithms and active learning) into crowdsourcing disciplines?

5 Crowdsourcing in Emergency Management

In this section, combination crowdsourcing and machine learning for various tasks in emergencies are reviewed, and challenges and opportunities are raised in terms of data analysis. Moreover, we more deeply contemplate, with some examples (tweets in Twitter) for challenges and opportunities discussed.

Volunteering is part of how community reacts to emergencies [31] and this process has been facilitated by social media in recent years. Volunteers provide information and resources to the affected people [87]. In this regard, for some years now, both researchers and practitioners, in the field of emergency management, have been exploring crowdsourcing for collecting, processing and sharing information between stakeholders. Jeff Howe in 2006 firstly defined the term as “the act of taking a job traditionally performed by a designated agent and outsourcing it to a generally large group of people in the form of an open call [48]”. Since Howes definition, an extended range of crowdsourcing researches have been carried out from a number of fields such as computer sciences, management, information systems and so on. Additionally, as a problem-solving method [32], crowdsourcing has caught the attention of emerging paradigms such as collective intelligence, human computation, or social computing [101]. With respect to this, Liu has analysed the distinct skills and expertise of different crowds typically involved in emergency management: (1) affected-populations, (2) social medias and (3) digital volunteer communities [77]. In this framework, affected populations generate local, timely and direct experiential information. And social media make available unexpected and fortuitous experience. Finally, digital volunteers offer their capabilities for

processing and managing emergency data. Also, Poblet et al. introduced four different crowd's roles such as sensors, social computers, reporters and micro-taskers. It considered four data types (i.e., raw data, unstructured data, semi-structured data and structured data) and two involvement types (i.e., active or passive) [101]. There have been recently efforts toward crowdsourcing such tasks in emergency management but it is still challenging. Social media posts come at a fast pace and immense volume. Moreover, it is challenging to collect all the posts which are related to a disaster due to the restrictions by social media services. The collected data contains daily information and is only in part insightful information. Another issue is malicious content such as spam and rumours which can cause panic and stress, especially when produced in large scale using bots [87]. In this regard, one of areas to potentially resolve these challenges is machine learning which has been applied to a variety of tasks in emergency management as mentioned in Section 2.

5.1 Crowdsourcing with Machine Learning for emergency management

Here, hybrid approach, crowdsourcing with machine learning, for various tasks in emergencies are reviewed, with exception of approaches for relief task, since the approaches are already reviewed in Section 2.7.

Pandey and Natarajan proposed a prototype solution to provide situation awareness during and post disaster event using semi-supervised machine learning technique based on SVM and creating interactive open street maps for crowdsourcing the user data providing threat and relief information [95]. Their model was evaluated with the data from Chennai flood 2015. Truong et al. developed a Bayesian approach to the classification of tweets during Hurricane Sandy in order to distinguish "informational" from "conversational" tweets [130]. They designed an effective set of features and used them as input to NB classifiers. The NB classifier was also introduced to reduce noise in crowdsourced data related to emergency management [124]. Their approach was assessed with the flood data of Cumbria 2015. As similarly, Imran et al. proposed automatic methods based on NB classifier for extracting information from microblog posts. They also focused on extracting valuable "information nuggets" relevant to disaster response [54]. Kurkcu et al. likewise used NB classifier with TF-IDF to identify keywords from tweets related to emergency [68]. Experimental result indicated that crowdsourced data refined could provide detailed location information of a specific incident according to its intensity and duration. Anbalagan and Valliyammai proposed a system that performs disaster tweet collection based on trending disaster hash tags [9]. They compared naive-Bayesian and Smooth Support Vector Machine (SSVM) classifications on collected tweets to identify the severity of the emergency. As a result, the SSVM outperformed the naive-Bayesian. Also, emergency geographic map was generated for the affected area through location to interpolation cluster proximity. Balena et al. evaluated supervised classification methods (AdaBoost, NB, RF, SVM and Neural Networks) to compare their effectiveness and potential for classifying message requests asking for/offering to help in emergency [16]. From their experiments, the RF and the Neural Network had better performance than the others. Nagy et al. introduced an evaluation of approaches to accurately and precisely identify crowd sentiment in social media data (Tweets) in emergency situation [85]. SVM was used to classify the lexicons linked to the seed. Their

technique performed better than Bayesian networks alone, and the combination with Bayesian networks improved the sentiment detection.

5.2 Example: Crowdsourcing and Machine Learning for Tracking Emergency

In this section, we more deeply contemplate, with several examples (tweet feeds in Twitter) for challenges and opportunities discussed in the previous section. Table 2 shows the tweet examples which will be discussed in this section.

Table 2 Tweet examples related to emergency

No	Tweet
E1	[11-03-2011T 05:49:01] BIG EARTHQUAKE!!!
E2	[11-03-2011T 05:50:00] Massive quake in Tokyo
E3	[29-10-2012T 03:16:51] #havetoacemysatexam this made me laugh, no idea why but omfg haha http://bit.ly/S6fbBt
E4	[15-11-2012T 13:03:15] ATTENTION ANYONE LOOKING FOR A JOB: FEMA needs assistance for South Jersey! \$1000.00/per l (904)797-5338!
E5	[15-11-2012T 05:50:13] #Sandy RUMOR CONTROL: The rumor that @fema is offering \$300 cash cards for food is FALSE. http://www.fema.gov/hurricane-sandy-rumor-control
E6	[12-05-2015T 05:50:45] Lets not forget the people in #NepalEarthquake let us all say a prayer for the people in Nepal MrAlMubarak
E7	[11-03-2011T 05:48:54] Huge earthquake in TK we are affected!
E8	[11-03-2011T 05:48:54] Earthquake!
E9	[11-03-2011T 05:48:08] Earthquake vertical shake!
E10	[11-03-2011T 05:48:14] Earthquake!!!!
E11	[11-03-2011T 05:52:23] I can see the tsunami coming!!!!

Discovering location of social media data. Geographical coordinate (known as geotagging) attached in message is useful for a number of tasks in disaster response [106]. For instance, it allows to search or verify information about a local event, by filtering the messages corresponding to a particular affected region; further geotagging can also be used for higher-level tasks, such as predict transmission of infectious disease. Unfortunately, only 2% of emergency-related messages includes GPS coordinates in practice, further, large portion of these messages may be made through the social bot [23]. For example, a place of the tweet E2 is able to be speculated from its content about Tohoku earthquake, while E1 doesn't contain explicit coordinates. In this regard, from user profiles, some information such as a home location (freeform text), preferred language and time zone may be considered for determining the location of the tweet [113].

Trustworthiness of crowdsourced data. Crowdsourced data reflects opinion of crowd; which sometimes contains more than the credible data. On the other hand, within crowdsourcing using social media service, rumour, spam and bots can be critical issues. Among the 8 million tweets according to Boston Marathon Bombing in 2013, around 29% were revealed to be rumours and 51% to be general opinions and comments [40]. This insight shows the perils of using keyword or hashtag-based topic definition. Additionally, as an example of spam (E3²), many

² <https://twitter.com/NevaKey9/status/262860388083843072>

tweets which generated in areas close to the occurrence of Hurricane Sandy had been related to promoting the material for Scholastic Assessment Test (SAT) together with *#HaveToAceMySATEexam* tag [87]. In case of E4³, it may be confirmed by other tweets which notice whether or not a rumor is true as E5⁴. For retrieving the relevant tweets, using the expert knowledge to compose high-precision queries has been emphasized as one possible solution [38].

Integration between different disciplines. Images and videos in tweets may open new opportunities to deeply understand situation and event. Information source such as website URLs, photos and videos in tweets related to emergency has been found to be around 18% [40]. Once images from the photos and videos are obtained, vision-based analysis can be used to find sources of the multimedia [18]. E6 is one example related to Nepal earthquake in 2015 and contains one image for broken apartments. The photo in tweet E6⁵ can be searched by Google Image Search to link with one YouTube video⁶ which contains a “Nepal” as a title.

Real-time analysis within sudden events that rapid increase data. In emergency such as Haiti earthquake in 2010 and Fukushima nuclear disaster in 2011, crowd power played an important role in emergency response [136]. However, it still difficult and complex to timely analyse the social data generated in emergencies because of a sharp and sudden rise in the number of tweets immediately. We note that the E1-2 and E7-11 were obtained from [33]. For instance, the E7-10 are tweets that have been generated in Tokyo after the Tohoku earthquake happened at the epicenter. In the E7-10 tweets are emerging rapidly with in seconds (even in milliseconds). Further E11 is the first tweet about a tsunami by an eyewitness at 6 minutes after the earthquake occurred; it indicates emergency may be dramatically changed from moment to moment. Therefore, it is essential to develop efficient and effective analysis algorithms that will be working in real-time; moreover, these algorithms should continuously adapt to existing situations and more correctly predict the next situation.

6 Conclusions

Some emergencies occurrence more and more frequently and are seriously affecting human life, society, environmental protection and even relationship between countries in the world. Therefore, stakeholders have been taking efforts to construct or study the emergency response for preventing and reducing damages, while researchers have been applying various techniques of machine learning to improve effect and efficiency of the emergency management. Although it was not perfect, many researches have proven successfully supporting tasks in the emergency phases.

The purpose of this chapter is to discuss a hybrid crowdsourcing and real-time machine learning approaches to rapidly process large volumes of data for emergency response in a time-sensitive manner. We then reviewed the application and the approach of machine learning techniques to support the emergency management for the each task, and the challenges and the opportunities were proposed.

³ <https://twitter.com/OccupySandyNJ/status/269183900113326082>

⁴ <https://twitter.com/FEMASandy/status/269180520993259520>

⁵ <https://twitter.com/safaher/status/598247877295644674>

⁶ <https://www.youtube.com/watch?v=aNZiLYEr6to>

We described characteristics of data being generated during disaster and discussed that the data characteristics akin to 5 Vs of big data. In addition, various applications cases of big data analysis were looked over. Moreover, we focused on crowdsourcing with machine learning in emergency management, and their challenges and opportunities were discussed in terms of the data analysis. Finally, example of the tweet feeds related to emergency was discussed to more deeply describe the challenges and opportunities.

Acknowledgements The work is funded from the Research Council of Norway (RCN) and the Norwegian Centre for International Cooperation in Education (SiU) grant through INT-PART programme.

References

1. Agarwal, R., Dhar, V.: Editorial - big data, data science, and analytics: The opportunity and challenge for IS research. *Information Systems Research* **25**(3), 443–448 (2014)
2. Agency, F.E.M.: Federal response plan (FRP). Tech. rep., Federal Emergency Management Agency (1999)
3. Akerkar, R.: Processing big data for emergency management. *Smart Technologies for Emergency Response and Disaster Management* p. 144 (2017)
4. Akter, S., Wamba, S.F.: Big data and disaster management: a systematic review and agenda for future research. *Annals of Operations Research* pp. 1–21 (2017)
5. Alam, F., Imran, M., Offi, F.: Image4act: Online social media image processing for disaster response. In: *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, pp. 601–604. ACM (2017)
6. Alam, F., Offi, F., Imran, M.: Processing social media images by combining human and machine computing during crises. *International Journal of Human–Computer Interaction* **34**(4), 311–327 (2018)
7. Alamdar, F., Kalantari, M., Rajabifard, A.: Towards multi-agency sensor information integration for disaster management. *Computers, Environment and Urban Systems* **56**, 68–85 (2016)
8. Amaye, A., Neville, K., Pope, A.: Bigpromises: using organisational mindfulness to integrate big data in emergency management decision making. *Journal of Decision Systems* **25**(sup1), 76–84 (2016)
9. Anbalagan, B., Valliyammai, C.: # chennai floods: Leveraging human and machine learning for crisis mapping during disasters using social media. In: *Proceedings of the IEEE 23rd International Conference on High Performance Computing Workshops (HiPCW)*, pp. 50–59. IEEE (2016)
10. Attari, N., Offi, F., Awad, M., Lucas, J., Chawla, S.: Nazr-cnn: Fine-grained classification of UAV imagery for damage assessment. In: *Proceedings of the 2017 IEEE International Conference on Data Science and Advanced Analytics, DSAA*, pp. 50–59. IEEE (2017)
11. Avvenuti, M., Cresci, S., Mariantonietta, P.N., Marchetti, A., Tesconi, M.: Earthquake emergency management by social sensing. In: *Proceedings of the 2014 IEEE international conference on Pervasive computing and communications workshops (PERCOM workshops)*, pp. 587–592. IEEE (2014)
12. Bahrepour, M., Meratnia, N., Poel, M., Taghikhaki, Z., Havinga, P.J.: Distributed event detection in wireless sensor networks for disaster management. In: *Proceedings of the 2nd International Conference on Intelligent Networking and Collaborative Systems (INCOS)*, pp. 507–512. IEEE (2010)
13. Bahrepour, M., Meratnia, N., Poel, M., Taghikhaki, Z., Havinga, P.J.: Use of wireless sensor networks for distributed event detection in disaster management applications. *International Journal of Space-Based and Situated Computing* **2**(1), 58–69 (2012)
14. Bai, H., Yu, G.: A weibo-based approach to disaster informatics: incidents monitor in post-disaster situation via weibo text negative sentiment analysis. *Natural Hazards* **83**(2), 1177–1196 (2016)

15. Bai, Y., Gao, C., Singh, S., Koch, M., Adriano, B., Mas, E., Koshimura, S.: A framework of rapid regional tsunami damage recognition from post-event terrasars-x imagery using deep neural networks. *IEEE Geoscience and Remote Sensing Letters* **15**(1), 43–47 (2018)
16. Balena, P., Amoroso, N., Lucia, C.D.: Integrating supervised classification in social participation systems for disaster response. a pilot study. In: *International Conference on Computational Science and Its Applications*, pp. 675–686. Springer (2017)
17. Beyreuther, M., Hammer, C., Wassermann, J., Ohrnberger, M., Megies, T.: Constructing a hidden markov model based earthquake detector: application to induced seismicity. *Geophysical Journal International* **189**(1), 602–610 (2012)
18. Bica, M., Palen, L., Bopp, C.: Visual representations of disaster. In: *Proceedings of the 20th ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW 2017)*, pp. 1262–1276 (2017)
19. Boccardo, P., Tonolo, F.G.: Remote sensing role in emergency mapping for disaster response. In: *Engineering Geology for Society and Territory*, vol. 5, pp. 17–24. Springer (2015)
20. Bontemps, S., Defourny, P., Bogaert, E.V., Arino, O., Kalogirou, V., Perez, J.R.: *Globecover 2009-products description and validation report* (2011)
21. Brants, T., Chen, F., Farahat, A.: A system for new event detection. In: *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, pp. 330–337. ACM (2003)
22. Brown, J., Jr, O.F., Heginbottom, J., Melnikov, E.: *Circum-Arctic map of permafrost and ground-ice conditions*. US Geological Survey Reston (1997)
23. Burton, S.H., Tanner, K.W., Giraud-Carrier, C.G., West, J.H., Barnes, M.D.: "right time, right place" health communication on twitter: value and accuracy of location information. *Journal of medical Internet research* **14**(6) (2012)
24. Caragea, C., McNeese, N., Jaiswal, A., Traylor, G., Kim, H.W., Mitra, P., Wu, D., Tapia, A.H., Giles, L., Jansen, B.J., et al.: Classifying text messages for the haiti earthquake. In: *Proceedings of the 8th international conference on information systems for crisis response and management (ISCRAM2011)*. Citeseer (2011)
25. Castillo, C.: *Big Crisis Data: Social Media in Disasters and Time-Critical Situations*. Cambridge University Press (2016)
26. Cervone, G., Sava, E., Huang, Q., Schnebele, E., Harrison, J., Waters, N.: Using twitter for tasking remote-sensing data collection and damage assessment: 2013 boulder flood case study. *International Journal of Remote Sensing* **37**(1), 100–124 (2016)
27. Cervone, G., Schnebele, E., Waters, N., Moccaldi, M., Sicignano, R.: Using social media and satellite data for damage assessment in urban areas during emergencies. In: *Seeing Cities Through Big Data*, pp. 443–457. Springer (2017)
28. Chen, A., Chen, N., Ni, H., et al.: *Modern emergency management theory and method* (2009)
29. Chen, N., Wenjing, L., Ruizhen, B., Chen, A.: Application of computational intelligence technologies in emergency management: a literature review. *Artificial Intelligence Review* pp. 1–38 (2017)
30. Chen, W., Xie, X., Peng, J., Wang, J., Duan, Z., Hong, H.: Gis-based landslide susceptibility modelling: a comparative assessment of kernel logistic regression, naïve-bayes tree, and alternating decision tree models. *Geomatics, Natural Hazards and Risk* **8**(2), 950–973 (2017)
31. Dimitroff, R., Schmidt, L., Bond, T.: Organizational behavior and disaster. *Project Management Journal* **36**(1), 28–38 (2005)
32. Doan, A., Ramakrishnan, R., Halevy, A.Y.: Crowdsourcing systems on the world-wide web. *Communications of the ACM* **54**(4), 86–96 (2011)
33. Doan, S., Vo, B.K.H., Collier, N.: An analysis of twitter messages in the 2011 tohoku earthquake. In: *International Conference on Electronic Healthcare*, pp. 58–66. Springer (2011)
34. Dong, X., Pi, D.: Novel method for hurricane trajectory prediction based on data mining. *Natural Hazards and Earth System Sciences* **13**(12), 3211–3220 (2013)
35. Dubey, R., Luo, Z., Xu, M., Wamba, S.F.: Developing an integration framework for crowdsourcing and internet of things with applications for disaster response. In: *Proceedings of the IEEE International Conference on Data Science and Data Intensive Systems (DSDIS)*, pp. 520–524. IEEE (2015)
36. Fahimnia, B., Sarkis, J., Davarzani, H.: Green supply chain management: A review and bibliometric analysis. *International Journal of Production Economics* **162**, 101–114 (2015)

37. Fersini, E., Messina, E., Pozzi, F.A.: Earthquake management: a decision support system based on natural language processing. *Journal of Ambient Intelligence and Humanized Computing* **8**(1), 37–45 (2017)
38. Ghenai, A., Mejova, Y.: Catching zika fever: Application of crowdsourcing and machine learning for tracking health misinformation on twitter. *arXiv preprint arXiv:1707.03778* (2017)
39. Grolinger, K., Mezghani, E., Capretz, M., Exposito, E.: Knowledge as a service framework for collaborative data management in cloud environments-disaster domain. *Managing big data in cloud computing environments* pp. 183–209 (2016)
40. Gupta, A., Lamba, H., Kumaraguru, P.: \$1.00 per rt# bostonmarathon# prayforboston: Analyzing fake content on twitter. In: *eCrime Researchers Summit (eCRS)*, 2013, pp. 1–12. IEEE (2013)
41. Haddow, G.D., Bullock, J.A., Coppola, D.P.: *Introduction to Emergency Management*. Butterworth-Heinemann (2017)
42. Harris, P., Anitha, J.: Post earthquake disaster awareness to emergency task force using crowdsourced data. In: *Proceedings of the IEEE International Conference on Industrial and Information Systems (ICIIS)*, pp. 1–6. IEEE (2017)
43. Hastie, T., Tibshirani, R., Friedman, J.: *Unsupervised learning*. In: *The elements of statistical learning*, pp. 485–585. Springer (2009)
44. Hazen, B.T., Boone, C.A., Ezell, J.D., Jones-Farmer, L.A.: Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications. *International Journal of Production Economics* **154**, 72–80 (2014)
45. Higuchi, H., Fujimura, J., Nakamura, T., Kogo, K., Tsudaka, K., Wada, T., Okada, H., Ohtsuki, K.: Disaster detection by statistics and svm for emergency rescue evacuation support system. In: *Proceedings of the 43th International Conference on Parallel Processing Workshops (ICCPW)*, pp. 349–354. IEEE (2014)
46. Hoetzlein, R.C.: Visual communication in times of crisis: the fukushima nuclear accident. *Leonardo* **45**(2), 113–118 (2012)
47. Howard, C., Jones, D., Reece, S., Waldock, A.: Learning to trust the crowd: Validating crowdsources for improved situational awareness in disaster response. *Procedia engineering* **159**, 141–147 (2016)
48. Howe, J.: The rise of crowdsourcing. *Wired magazine* **14**(6), 1–4 (2006)
49. Hristidis, V., Chen, S.C., Li, T., Luis, S., Deng, Y.: Survey of data management and analysis in disaster situations. *Journal of Systems and Software* **83**(10), 1701–1714 (2010)
50. Huang, Q., Cervone, G., Jing, D., Chang, C.: Disastermapper: A cybergis framework for disaster management using social media data. In: V. Chandola, R.R. Vatsavai (eds.) *Proceedings of the 4th International ACM SIGSPATIAL Workshop on Analytics for Big Geospatial Data, BigSpatial@SIGSPATIAL 2015*, pp. 1–6. ACM (2015)
51. Imran, M.: *Aidr: Artificial intelligence for disaster*. Ph.D. thesis, Qatar Computing Research Institute (2014)
52. Imran, M., Castillo, C., Lucas, J., Meier, P., Rogstadius, J.: Coordinating human and machine intelligence to classify microblog communications in crises. In: *ISCRAM* (2014)
53. Imran, M., Castillo, C., Lucas, J., Meier, P., Vieweg, S.: *Aidr: Artificial intelligence for disaster response*. In: *Proceedings of the 23rd International Conference on World Wide Web*, pp. 159–162. ACM (2014)
54. Imran, M., Elbassuoni, S., Castillo, C., Diaz, F., Meier, P.: Extracting information nuggets from disaster-related messages in social media. In: *Isram* (2013)
55. Imran, M., Mitra, P., Castillo, C.: Twitter as a lifeline: Human-annotated twitter corpora for NLP of crisis-related messages. In: *Proceedings of the Tenth International Conference on Language Resources and Evaluation LREC 2016* (2016)
56. Izadi, M., Mohammadzadeh, A., Haghighattalab, A.: A new neuro-fuzzy approach for post-earthquake road damage assessment using ga and svm classification from quickbird satellite images. *Journal of the Indian Society of Remote Sensing* **45**(6), 965–977 (2017)
57. Jiang, D.M., Li, Z.B.: The evaluation of city emergency management system results based on bp neural network. In: *Applied Mechanics and Materials*, vol. 263, pp. 3288–3291. Trans Tech Publ (2013)
58. Jordan, M.I., Mitchell, T.M.: Machine learning: Trends, perspectives, and prospects. *Science* **349**(6245), 255–260 (2015)
59. Joshi, A.R., Tarte, I., Suresh, S., Koolagudi, S.G.: Damage identification and assessment using image processing on post-disaster satellite imagery. In: *IEEE transaction on Global Humanitarian Technology Conference (GHTC)*, pp. 1–7. IEEE (2017)

60. Kaelbling, L.P., Littman, M.L., Moore, A.W.: Reinforcement learning: A survey. *Journal of artificial intelligence research* **4**, 237–285 (1996)
61. Kao, C.Y., Loh, C.H.: Monitoring of long-term static deformation data of fei-tsui arch dam using artificial neural network-based approaches. *Structural Control and Health Monitoring* **20**(3), 282–303 (2013)
62. Kim, H.S., Ho, C.H., Kim, J.H., Chu, P.S.: Track-pattern-based model for seasonal prediction of tropical cyclone activity in the western north pacific. *Journal of Climate* **25**(13), 4660–4678 (2012)
63. Kim, S.W., Melby, J.A., Nadal-Caraballo, N.C., Ratcliff, J.: A time-dependent surrogate model for storm surge prediction based on an artificial neural network using high-fidelity synthetic hurricane modeling. *Natural Hazards* **76**(1), 565–585 (2015)
64. Kong, Q., Allen, R.M., Schreier, L., Kwon, Y.W.: Myshake: A smartphone seismic network for earthquake early warning and beyond. *Science advances* **2**(2), e1501,055 (2016)
65. Kotsiantis, S.B., Zaharakis, I., Pintelas, P.: Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering* **160**, 3–24 (2007)
66. Krzhizhanovskaya, V.V., Shirshov, G., Melnikova, N., Belleman, R.G., Rusadi, F., Broekhuijsen, B., Gouldby, B., Lhomme, J., Balis, B., Bubak, M., et al.: Flood early warning system: design, implementation and computational modules. *Procedia Computer Science* **4**, 106–115 (2011)
67. Kumar, S., Barbier, G., Abbasi, M.A., Liu, H.: Tweettracker: An analysis tool for humanitarian and disaster relief. In: L.A. Adamic, R.A. Baeza-Yates, S. Counts (eds.) *Proceedings of the Fifth International Conference on Weblogs and Social Media, Barcelona, Catalonia, Spain, July 17–21, 2011*. The AAAI Press (2011)
68. Kurcu, A., Zuo, F., Gao, J., Morgul, E.F., Ozbay, K.: Crowdsourcing incident information for disaster response using twitter. In: *Proceedings of the 65th Annual Meeting of Transportation Research Board* (2017)
69. Landwehr, P.M., Carley, K.M.: Social media in disaster relief. In: *Data mining and knowledge discovery for big data*, pp. 225–257. Springer (2014)
70. Lehner, B., Döll, P.: Development and validation of a global database of lakes, reservoirs and wetlands. *Journal of Hydrology* **296**(1–4), 1–22 (2004)
71. Lehner, B., Liermann, C.R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., Döll, P., Endejan, M., Frenken, K., Magome, J., et al.: High-resolution mapping of the world’s reservoirs and dams for sustainable river-flow management. *Frontiers in Ecology and the Environment* **9**(9), 494–502 (2011)
72. Lenzerini, M.: Data integration: A theoretical perspective. In: L. Popa, S. Abiteboul, P.G. Kolaitis (eds.) *Proceedings of the Twenty-first ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, pp. 233–246. ACM (2002)
73. Li, H., Caragea, D., Caragea, C., Herndon, N.: Disaster response aided by tweet classification with a domain adaptation approach. *Journal of Contingencies and Crisis Management* **26**(1), 16–27 (2018)
74. Li, J., He, Z., Plaza, J., Li, S., Chen, J., Wu, H., Wang, Y., Liu, Y.: Social media: New perspectives to improve remote sensing for emergency response. *Proceedings of the IEEE* **105**(10), 1900–1912 (2017)
75. Li, T., Xie, N., Zeng, C., Zhou, W., Zheng, L., Jiang, Y., Yang, Y., Ha, H., Xue, W., Huang, Y., Chen, S., Navlakha, J.K., Iyengar, S.S.: Data-driven techniques in disaster information management. *ACM Computing Surveys* **50**(1), 1:1–1:45 (2017)
76. Liang, J., Jacobs, P., Parthasarathy, S.: Human-guided flood mapping: From experts to the crowd (2018)
77. Liu, S.B.: Crisis crowdsourcing framework: designing strategic configurations of crowdsourcing for the emergency management domain. *Computer Supported Cooperative Work (CSCW)* **23**(4–6), 389–443 (2014)
78. Mariano, C., Morales, E.: A new distributed reinforcement learning algorithm for multiple objective optimization problems. In: *Advances in Artificial Intelligence*, pp. 290–299. Springer (2000)
79. Martínez-Álvarez, F., Lora, A.T., Morales-Esteban, A., Riquelme, J.C.: Computational intelligence techniques for predicting earthquakes. In: E. Corchado, M. Kurzynski, M. Wozniak (eds.) *Hybrid Artificial Intelligent Systems - 6th International Conference, HAIS 2011, Wroclaw, Poland, May 23–25, 2011, Proceedings, Part II, Lecture Notes in Computer Science*, vol. 6679, pp. 287–294. Springer (2011)

80. Mehrotra, S., Qiu, X., Cao, Z., Tate, A.: Technological challenges in emergency response. *IEEE Intelligent Systems* **28**(4), 5–8 (2013)
81. Mori, K., Nakamura, T., Fujimura, J., Tsudaka, K., Wada, T., Okada, H., Ohtsuki, K.: Development of ERESS in panic-type disasters: Disaster recognition algorithm by buffering-SVM. In: *Proceedings of the 13th International Conference on ITS Telecommunications (ITST)*, pp. 337–343. IEEE (2013)
82. Moustra, M., Avraamides, M., Christodoulou, C.: Artificial neural networks for earthquake prediction using time series magnitude data or seismic electric signals. *Expert systems with applications* **38**(12), 15,032–15,039 (2011)
83. Murali, S., Krishnapriya, V., Thomas, A.: Crowdsourcing for disaster relief: A multi-platform model. In: *Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER)*, IEEE, pp. 264–268. IEEE (2016)
84. Nadi, A., Edrisi, A.: Adaptive multi-agent relief assessment and emergency response. *International Journal of Disaster Risk Reduction* **24**, 12–23 (2017)
85. Nagy, A., Stamberger, J.: Crowd sentiment detection during disasters and crises. In: *Proceedings of the 9th International ISCRAM Conference*, pp. 1–9 (2012)
86. Nazer, T.H., Morstatter, F., Dani, H., Liu, H.: Finding requests in social media for disaster relief. In: *Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pp. 1410–1413. IEEE (2016)
87. Nazer, T.H., Xue, G., Ji, Y., Liu, H.: Intelligent disaster response via social media analysis a survey. *ACM SIGKDD Explorations Newsletter* **19**(1), 46–59 (2017)
88. Nguyen, D.T., Alam, F., Ofli, F., Imran, M.: Automatic image filtering on social networks using deep learning and perceptual hashing during crises. *Computing Research Repository* **abs/1704.02602** (2017)
89. Nguyen, D.T., Ofli, F., Imran, M., Mitra, P.: Damage assessment from social media imagery data during disasters. In: *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, pp. 569–576. ACM (2017)
90. Ohlmacher, G.C., Davis, J.C.: Using multiple logistic regression and gis technology to predict landslide hazard in northeast kansas, usa. *Engineering geology* **69**(3-4), 331–343 (2003)
91. Okolloh, O.: Ushahidi, or testimony: Web 2.0 tools for crowdsourcing crisis information. *Participatory learning and action* **59**(1), 65–70 (2009)
92. Olteanu, A., Castillo, C., Diaz, F., Vieweg, S.: Crisislex: A lexicon for collecting and filtering microblogged communications in crises. In: *Proceedings of the 8th International AAAI Conference on Weblogs and Social Media (ICWSM)* (2014)
93. Olteanu, A., Vieweg, S., Castillo, C.: What to expect when the unexpected happens: Social media communications across crises. In: *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pp. 994–1009. ACM (2015)
94. Pabreja, K.: Clustering technique to interpret numerical weather prediction output products for forecast of cloudburst. *International Journal of Computer Science and Information Technologies (IJCSIT)* **3**(1), 2996–2999 (2012)
95. Pandey, N., Natarajan, S.: How social media can contribute during disaster events? case study of chennai floods 2015. In: *Proceedings of the International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 1352–1356. IEEE (2016)
96. Pappenberger, F., Matgen, P., Beven, K.J., Henry, J.B., Pfister, L., et al.: Influence of uncertain boundary conditions and model structure on flood inundation predictions. *Advances in Water Resources* **29**(10), 1430–1449 (2006)
97. Peel, M.C., Finlayson, B.L., McMahon, T.A.: Updated world map of the köppen-geiger climate classification. *Hydrology and earth system sciences discussions* **4**(2), 439–473 (2007)
98. Pengel, B., Krzhizhanovskaya, V., Melnikova, N., Shirshov, G., Koelewijn, A., Pyayt, A., Mokhov, I., et al.: Flood early warning system: sensors and internet. *IAHS Red Book* **357**, 445–453 (2013)
99. Pham, B.T., Khosravi, K., Prakash, I.: Application and comparison of decision tree-based machine learning methods in landside susceptibility assessment at pauri garhwal area, uttarakhand, india. *Environmental Processes* **4**(3), 711–730 (2017)
100. Piczak, K.J.: ESC: Dataset for environmental sound classification. In: *Proceedings of the 23rd ACM international conference on Multimedia*, pp. 1015–1018. ACM (2015)

101. Poblet, M., García-Cuesta, E., Casanovas, P.: Crowdsourcing roles, methods and tools for data-intensive disaster management. *Information Systems Frontiers* pp. 1–17 (2017)
102. Pohl, D., Bouchachia, A., Hellwagner, H.: Social media for crisis management: clustering approaches for sub-event detection. *Multimedia Tools and Applications* **74**(11), 3901–3932 (2015)
103. Purohit, H., Castillo, C., Diaz, F., Sheth, A., Meier, P.: Emergency-relief coordination on social media: Automatically matching resource requests and offers. *First Monday* **19**(1) (2013)
104. Ragini, J.R., Anand, P.R., Bhaskar, V.: Mining crisis information: A strategic approach for detection of people at risk through social media analysis. *International Journal of Disaster Risk Reduction* **27**, 556–566 (2018)
105. Ramchurn, S.D., Huynh, T.D., Ikuno, Y., Flann, J., Wu, F., Moreau, L., Jennings, N.R., Fischer, J.E., Jiang, W., Rodden, T., et al.: Hac-er: A disaster response system based on human-agent collectives. In: *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*, pp. 533–541. International Foundation for Autonomous Agents and Multiagent Systems (2015)
106. Ravikumar, K., Kannan, A.R.: Spatial data mining for prediction of natural events and disaster management based on fuzzy logic using hybrid pso (2018)
107. Resch, B., Usländer, F., Havas, C.: Combining machine-learning topic models and spatiotemporal analysis of social media data for disaster footprint and damage assessment. *Cartography and Geographic Information Science* pp. 1–15 (2017)
108. Revilla-Romero, B., Thielen, J., Salamon, P., Groeve, T.D., Brakenridge, G.: Evaluation of the satellite-based global flood detection system for measuring river discharge: influence of local factors. *Hydrology and Earth System Sciences* **18**(11), 4467 (2014)
109. Reyes, J., Morales-Esteban, A., Martínez-Álvarez, F.: Neural networks to predict earthquakes in chile. *Applied Soft Computing* **13**(2), 1314–1328 (2013)
110. Sadhu, V., Salles-Loustau, G., Pompili, D., Zonouz, S., Sritapan, V.: Argus: Smartphone-enabled human cooperation via multi-agent reinforcement learning for disaster situational awareness. In: *Proceedings of the IEEE International Conference on Autonomic Computing (ICAC)*, pp. 251–256. IEEE (2016)
111. Sahay, R.R., Srivastava, A.: Predicting monsoon floods in rivers embedding wavelet transform, genetic algorithm and neural network. *Water resources management* **28**(2), 301–317 (2014)
112. Sakaki, T., Okazaki, M., Matsuo, Y.: Earthquake shakes twitter users: real-time event detection by social sensors. In: *Proceedings of the 19th international conference on World wide web*, pp. 851–860. ACM (2010)
113. Sakaki, T., Okazaki, M., Matsuo, Y.: Tweet analysis for real-time event detection and earthquake reporting system development. *IEEE Transactions on Knowledge and Data Engineering* **25**(4), 919–931 (2013)
114. Salamon, J., Jacoby, C., Bello, J.P.: A dataset and taxonomy for urban sound research. In: *Proceedings of the 22nd ACM international conference on Multimedia*, pp. 1041–1044. ACM (2014)
115. Salehi, M., Rusu, L.I., Lynar, T., Phan, A.: Dynamic and robust wildfire risk prediction system: an unsupervised approach. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 245–254. ACM (2016)
116. Salfinger, A., Girtelschmid, S., Pröll, B., Retschitzegger, W., Schwinger, W.: Crowdsensing meets situation awareness: A research roadmap for crisis management. In: *Proceedings of the 48th Hawaii International Conference on System Sciences, HICSS*, pp. 153–162. Kauai, Hawaii, USA (2015)
117. Saxton, G.D., Oh, O., Kishore, R.: Rules of crowdsourcing: Models, issues, and systems of control. *Information Systems Management* **30**(1), 2–20 (2013)
118. Shah, H., Ghazali, R.: Prediction of earthquake magnitude by an improved abc-mlp. In: *Developments in E-systems Engineering (DeSE)*, pp. 312–317. IEEE (2011)
119. Shao, X., Li, X., Li, L., Hu, X.: The application of ant-colony clustering algorithm to earthquake prediction. In: *Advances in Electronic Engineering, Communication and Management Vol. 2*, pp. 145–150. Springer (2012)
120. Shen, S., Murzintcev, N., Song, C., Cheng, C.: Information retrieval of a disaster event from cross-platform social media. *Information Discovery and Delivery* **45**(4), 220–226 (2017)

121. Singh, J.P., Dwivedi, Y.K., Rana, N.P., Kumar, A., Kapoor, K.K.: Event classification and location prediction from tweets during disasters. *Annals of Operations Research* pp. 1–21 (2017)
122. Song, X., Zhang, Q., Sekimoto, Y., Shibasaki, R.: Prediction of human emergency behavior and their mobility following large-scale disaster. In: *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 5–14. ACM (2014)
123. Sorensen, J.H.: Hazard warning systems: Review of 20years of progress. *Natural Hazards Review* **1**(2), 119–125 (2000)
124. Spielhofer, T., Greenlaw, R., Markham, D., Hahne, A.: Data mining twitter during the uk floods: Investigating the potential use of social media in emergency management. In: *Proceedings of the 3rd International Conference on Information and Communication Technologies for Disaster Management (ICT-DM)*, pp. 1–6. IEEE (2016)
125. Su, Y., Chelluboina, S., Hahsler, M., Dunham, M.H.: A new data mining model for hurricane intensity prediction. In: *Proceedings of the 2010 IEEE International Conference on Data Mining Workshops (ICDMW)*, pp. 98–105. IEEE (2010)
126. Tapia, A.H., Bajpai, K., Jansen, B.J., Yen, J., Giles, L.: Seeking the trustworthy tweet: Can microblogged data fit the information needs of disaster response and humanitarian relief organizations. In: *Proceedings of the 8th International ISCRAM Conference*, pp. 1–10 (2011)
127. Tehrany, M.S., Pradhan, B., Jebur, M.N.: Spatial prediction of flood susceptible areas using rule based decision tree (dt) and a novel ensemble bivariate and multivariate statistical models in gis. *Journal of Hydrology* **504**, 69–79 (2013)
128. Teodorescu, H.N.: Using analytics and social media for monitoring and mitigation of social disasters. *Procedia Engineering* **107**, 325–334 (2015)
129. Trekin, A., Novikov, G., Potapov, G., Ignatiev, V., Burnaev, E.: Satellite imagery analysis for operational damage assessment in emergency situations. *arXiv preprint arXiv:1803.00397* (2018)
130. Truong, B., Caragea, C., Squicciarini, A., Tapia, A.H.: Identifying valuable information from twitter during natural disasters. *Proceedings of the Association for Information Science and Technology* **51**(1), 1–4 (2014)
131. Vetrivel, A., Kerle, N., Gerke, M., Nex, F., Vosselman, G.: Towards automated satellite image segmentation and classification for assessing disaster damage using data-specific features with incremental learning (2016)
132. Vieweg, S., Hughes, A.L., Starbird, K., Palen, L.: Microblogging during two natural hazards events: what twitter may contribute to situational awareness. In: E.D. Mynatt, D. Schoner, G. Fitzpatrick, S.E. Hudson, W.K. Edwards, T. Rodden (eds.) *Proceedings of the 28th International Conference on Human Factors in Computing Systems, CHI*, pp. 1079–1088. ACM, Atlanta, Georgia (2010)
133. Wang, Z., Lai, C., Chen, X., Yang, B., Zhao, S., Bai, X.: Flood hazard risk assessment model based on random forest. *Journal of Hydrology* **527**, 1130–1141 (2015)
134. Xu, W., Liu, L., Shang, W.: Leveraging cross-media analytics to detect events and mine opinions for emergency management. *Online Information Review* **41**(4), 487–506 (2017)
135. Yamazaki, D., O’Loughlin, F., Trigg, M.A., Miller, Z.F., Pavelsky, T.M., Bates, P.D.: Development of the global width database for large rivers. *Water Resources Research* **50**(4), 3467–3480 (2014)
136. Yang, D., Zhang, D., Frank, K., Robertson, P., Jennings, E., Roddy, M., Lichtenstern, M.: Providing real-time assistance in disaster relief by leveraging crowdsourcing power. *Personal and Ubiquitous Computing* **18**(8), 2025–2034 (2014)
137. Ye, J., Kobayashi, T., Wang, X., Tsuda, H., Masahiro, M.: Audio data mining for anthropogenic disaster identification: An automatic taxonomy approach. *IEEE Transactions on Emerging Topics in Computing* (2017)
138. Yin, J., Lampert, A., Cameron, M., Robinson, B., Power, R.: Using social media to enhance emergency situation awareness. *IEEE Intelligent Systems* **27**(6), 52–59 (2012)
139. Yoon, D.K., Jeong, S.: Assessment of community vulnerability to natural disasters in korea by using gis and machine learning techniques. In: *Quantitative Regional Economic and Environmental Analysis for Sustainability in Korea*, pp. 123–140. Springer (2016)
140. Zagorecki, A.T., David, E.J., Ristvej, J.: Data mining and machine learning in the context of disaster and crisis management. *International Journal of Emergency Management* **9**(4), 351–365 (2013)

141. Zanini, M.A., Faleschini, F., Zampieri, P., Pellegrino, C., Gecchele, G., Gastaldi, M., Rossi, R.: Post-quake urban road network functionality assessment for seismic emergency management in historical centres. *Structure and Infrastructure Engineering* **13**(9), 1117–1129 (2017)
142. Zhang, X.Y., Li, X., Lin, X.: The data mining technology of particle swarm optimization algorithm in earthquake prediction. In: *Advanced Materials Research*, vol. 989, pp. 1570–1573. Trans Tech Publ (2014)
143. Zhang, Y., Burton, H.V., Sun, H., Shokrabadi, M.: A machine learning framework for assessing post-earthquake structural safety. *Structural Safety* **72**, 1–16 (2018)
144. Zheng, L., Shen, C., Tang, L., Li, T., Luis, S., Chen, S.: Applying data mining techniques to address disaster information management challenges on mobile devices. In: C. Apté, J. Ghosh, P. Smyth (eds.) *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 283–291. ACM (2011)
145. Zmazek, B., Todorovski, L., Džeroski, S., Vaupotič, J., Kobal, I.: Application of decision trees to the analysis of soil radon data for earthquake prediction. *Applied Radiation and Isotopes* **58**(6), 697–706 (2003)
146. Zmazek, B., Živčič, M., Vaupotič, J., Bidovec, M., Poljak, M., Kobal, I.: Soil radon monitoring in the krško basin, slovenia. *Applied Radiation and Isotopes* **56**(4), 649–657 (2002)
147. Zubiaga, A., Ji, H.: Tweet, but verify: epistemic study of information verification on twitter. *Social Network Analysis and Mining* **4**(1), 163 (2014)