Social Media for Disaster Situations

Methods, Opportunities and Challenges

Alivelu Mukkamala
BusinessIT Department
IT University of Copenhagen
Copenhagen, Denmark
ammu@itu.dk

Roman Beck
BusinessIT Department
IT University of Copenhagen
Copenhagen, Denmark
romb@itu.dk

***Abstract*— During disasters people start using social media as a platform to share and disseminate real-time disaster information to wider audiences. In order to understand the type of information that is being shared during disasters and how communities are using the technologies to respond to the disasters, we analyze two-different case studies on natural disasters using Twitter as a platform for gathering and sharing information. In both the case studies, we applied different content analysis methods, both manual and automated, to analyze the valuable information from the user-generated content produced during disaster situations. Based on our findings, we argue that social media platforms are facilitating collective level situation awareness among people and valuable information for disaster management agencies. However, in order to integrate social media in organizational work routines and processes, understanding the opportunities along with challenges is a key.**

Keywords—Disasters; Social Media; Twitter

I. Introduction

Humankind is prone to different types of disasters such as earthquakes, floods, epidemics, etc., and had to deal with devastating consequences of disasters in our history. Disasters are events that disrupt a communities' normal functioning and have an impact on people's lives, economy and environment. When compared to epidemic diseases or economic crises, both natural and man-made disasters occur suddenly and require fast relief activities [1]. To mitigate the damage, disasters typically are managed through four disaster management phases such as mitigation, preparedness, response and recover/reconstruction [2, 3].

In recent times, we have seen a noteworthy rise in the number of natural disasters throughout the world [4, 5]. Along with natural disasters, man-made disasters during the recent decades are also disrupting and making communities vulnerable. During disasters, the availability of real-time information is essential for both emergency management agencies as well as humanitarian organizations for effective decision making and coordinating their immediate response activities. The quicker the organizations react and respond; the more effective are the relief efforts. The rescue teams can save the affected population and assess the infrastructure damage, to estimate the economic loss and estimate the fatalities. Gauging who needs help, where to reach them, and what kind of help is needed, is limited, in reality, by the constraints of the difficulty in tracing and tracking the information needs of people [6]. However, compared to the 1990's, nowadays the time to get the information from the disaster zone has decreased tremendously because of technological developments such as Internet, mobile communication and most importantly social media [7].

Social media is playing an important role both in our personal and professional lives. The new digital technologies are facilitating the two-way communication as people become both consumers as well as producers of the information [8]. As discussed previously, disasters often disrupt traditional communication systems as infrastructure might get damaged. In these critical situations people often look for other means to share and receive the information. Hence social media become a privileged platform to share and disseminate real-time disaster information to an even wider audience [9] . In addition to individuals, government organizations, media and NGOs are also sharing information during different crisis events. For example, during Alberta floods, Boston bombing, Oklahoma tornado, West Texas explosion, affected people and others switched to social media like Twitter to gather information about a crisis situation [10-12]. The user-generated and unstructured data contains a large amount of valuable information. This information comprises, for example, rescue requests, situational awareness information, or coordinating relief activities. Moreover, affected communities are using social media to share information, to mobilize and self-organize to coordinate activities among themselves [13, 14]. Hence, it is very important for disaster management agencies to listen to the voices of affected communities to organize their relief activities because of the benefit of real time information [15]. However, the volume and the velocity of social media data makes it challenging to monitor and extract valuable information. It is important to develop methodologies and techniques that will automate the process to figure out the rescue requests and the urgencies of the needs of the affected people, which include shortages of food, clothing and shelter. In this regard, by taking example of two case studies that we have explored, we discuss the applied methods and their opportunities and challenges regarding analyzing disaster social media data to extract important information that will be an asset to the disaster management agencies. The two case studies are hurricane Sandy that happened in 2012 in US and Chennai floods that happened in India during 2015. Both these case studies belong to two different geographical locations with different cultural norms, but however, it is interesting to observe the commonalities between two different contexts. Moreover, the first disaster took place in a developed country where infrastructures and support/relief systems are at an advance level and the other case study is from the developing country where infrastructures as well as relief systems are at a minimal level. Our motivation for using these case studies is also to explore how social media is being used by the people in crisis situations, in different contexts, to share and receive the information.

II. Related work

The new trend of using technology to share disaster relevant information by the people during disasters, has gained the attention of research community way back in 2008 [16-18]. People used available technologies of their reach such as web-forums, blogs, Internet to locate their loved ones and to gain or disseminate information about disasters [16]. Social networking sites such as Twitter and Facebook penetrated into the personal lives of people and then also during times of disasters. People started using these networking sites tremendously to share the real-time information, which has changed the traditional disaster communication practices [16, 19]. Among the other social networking sites, research focused more on disaster related publicly available Twitter data because of its' message characteristics, public nature, and the speed with which the information reaches the wider audiences [20-22]. Since people are sharing and retweeting real-time information from the affected areas, significant research has focused on identifying the type and nature of information [23-28]. In addition to the opinions and emotion related messages, the social media messages also contain important situation updates, which can help in assessing the situation and creates situational awareness [13, 28].

Based on the prior research, for example, hurricane sandy [29], red river valley floods and wild fires [30], and floods in Pakistan [31] it is evident that disaster related social media data of all types of disasters contain valuable information. However, with a view to extract situational awareness information from the huge amounts of social media data, classifiers were built using different machine learning algorithms with the help of labelled data as training sets [32]. In order to understand the practical relevance and the importance of usage of social media by the emergency management agencies, interviews were conducted to understand officials' opinions and concluded that social media does provide invaluable information, however the organizations are still struggling to integrate it into their day to day practices [33]. However, there are certain success stories where disaster management agencies started using social media successfully. For example, during the Queensland floods, emergency officials used social media to communicate flood updates and information, yet they have ended up mostly with one-way communication [34]. Moreover, applications and tools have been developed to identify the disaster or to analyze the disaster related social media data, to name a few, SensePlace2 [35] ,Twitcident [36], Tweedr [37] CrisisTracker [38], or Artificial Intelligence for Disaster Response (AIDR) [39]. These are the few examples which shed light on the way social media has been exploited during disasters and for disaster management.

III. CASE STUDY-HURRICANE SANDY

In our first case study [29], we wanted to understand what people discussed about during the storm and what type of information is being shared by the people. We especially wanted to understand if people share their personal experiences and observations (original source) or retweet others' tweets or media information and weather reports (secondary source) [30]. For this purpose, we analyzed 11 million tweets that were posted on Twitter when hurricane Sandy hit the east coast of US in 2012.

A. Hurricane Sandy:

Hurricane Sandy initially developed in the Caribbean waters on October 22, 2012. While gaining forward momentum and slowly intensified and developed into a superstorm affecting 60 million people across 24 states in the U.S. It made landfall on 29th of October and devastated New York city with power outages, flooded subway systems, disrupted communication systems, and also led to the shortages of gasoline, commodities and food [40]. However, people used social media extensively to share information about hurricane, for example, @tweetuser (2012-11-01 03:28:59)

> *"What's amazing is how Twitter and Facebook are more current and up to date with events on #Sandy then the actual news on tv...#media #fail"*

B. Data collection and pre-processing

In this case study, we analyzed tweets that originated when the U.S. east coast was hit by hurricane Sandy in October 2012. Hurricane Sandy Twitter dataset contains 15 million tweet IDs that are made available publicly [41] and out of the total dataset, we were only able to retrieve 11 million tweets between May and June 2015, using the Twitter Rest APIs. The number of tweets generated during the disaster situation indicates the interest among people including the non-effected, on hurricane Sandy, however, the tweets posted by the people in and around the hurricane impacted area are most useful and contains valuable situational awareness information. Hence our aim was to extract tweets that are produced and generated by the people in the affected areas. So, in our work, to extract and visualize the tweets from the affected area of hurricane Sandy, we used different tools that can process and filter the tweets that contain geo-location information.

Firstly, with the help of Tableau [42], a business analytics and visualization tool, first we visualized and identified the geographical spread of 11 million tweets. In general, if the twitter data contains longitude and latitude coordinates, it indicate the location of a Twitter user at the time of sharing the information [43]. With the help of Cosmos software [44], which can process geo-located information, we filtered and excluded tweets that do not contain geo-location information. Hence, we identified that out of 11 million tweets only 115,800 tweets, or 1.07% of tweets contain geo-location information. The next step was to extract tweets that originated from the hurricane Sandy path, that affected the coast-line, in addition to this we had to filter and extract tweets that were written in the English language. For this purpose, we used CartoDB [45] to identify the English language tweets and to narrow down the tweets that are originated from the hurricane Sandy path based on the geo-location information of the tweet as shown in Fig. 1. Finally, we extracted 68,800 tweets from the hurricane Sandy affected area that were produced between 25th October and 5th November, 2012, in the east coast of the U.S. The overall descriptive statistics of the twitter dataset that was extracted during hurricane Sandy is presented in Table I.

TABLE I. DESCRIPTIVE STATISTICS OF HURRICANE SANDY DATASET

Twitter messages (tweets)	**Absolute numbers**	**Percentage**
Total tweets	11,658,279	100%
Original tweets	5,369,520	49.50%
Retweeted tweets	5,478,562	50.50%
Tweets with geo-location	115,800	1.07%
English tweets with geo-location	100,700	0.93%

C. Methodology

To explore whether valuable and relevant information exists in the final extracted dataset containing 68,800 tweets, one of the authors manually read through all the tweets for the important information and identified 677 important tweets. Further to analyze the resulted tweets, we developed our own coding scheme based on two different prior coding schemes: original/secondary sources [26] and nature of messages [30]. The coding scheme developed by us helped us to analyze *a tweet along the two dimensions: information source and nature of message*. With the help of information source a tweet can be classified as 1) original source or 2) secondary source. Original source refers to an individuals' personal observations and experiences, and secondary source refers to media information and its' links, retweets, and other online sources. The coding scheme under nature of messages consists of 1) informational messages, 2) action-related, 3) opinion-related and 4) emotion-related. Hence along with information source dimension, a tweet produced by a user from the hurricane sandy affected area can be classified with a label from one of the 4 types i.e. informational, or action-related, opinion-related or emotion-related. Based on our new coding scheme, further both the authors conducted the content analysis of 677 tweets. In between the coded tweets were discussed and cross validated and discrepancies were sorted out. The results from the analysis were later checked again by another member from our research group.

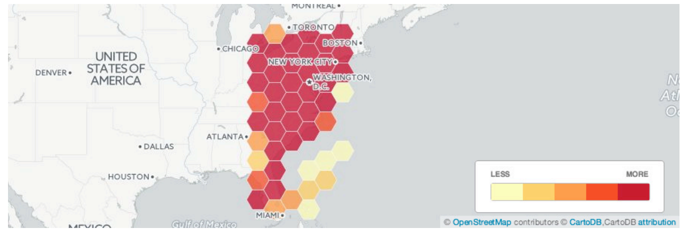

Fig. 1 Sandy affected area (East coast of the U.S.)

D. Data analysis and results

In our analysis, we identified a gradual increase in the number of tweets that began from 27th of October (which is before the landfall of hurricane Sandy), with a peak of 2.2 million tweets on the day of October 29th, when it actually made landfall. On October 30th 2012, this number rose to 2.7 million messages per day shared resulting in the highest peak. Fig. 1 shows a visualization of density map containing the 68,800 tweets, that were generated along the east coast, mainly from Florida, Connecticut, New Jersey, Massachusetts, New York. However, among a total of 68,800 tweets only 677 tweets contained valuable and relevant information pertaining to hurricane Sandy. The results revealed that during pre-disaster phase (27th of October) people shared hurricane relevant information from media sources, retweeted other's updates and shared weather reports, but just before the landfall of the hurricane, people started sharing real-time information.

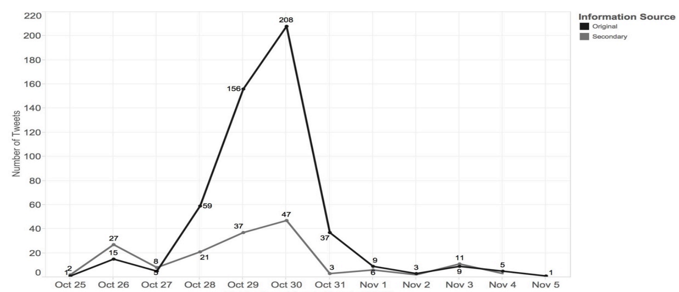

Fig. 2 Information Source

A significant amount of the information came from original sources, which means that people started sharing their valuable personal experiences and observations. The graph in Fig. 2 shows the representation of the 677 tweets that were coded among original and secondary sources. Along with information messages, people also shared their opinions, emotion-related and action related messages. However, information-related messages with situation updates are dominant among the dataset. Based on this case study, we argue that Twitter as a platform facilitates collective level situation awareness among people [29]. Situation awareness is important aspect during disasters either for government organizations, NGOs or relief organizations, to act quickly in disaster response and recovery phases.

IV. CASE STUDY- CHENNAI RAINS

In our second case study, from the theoretical perspective of social presence, we wanted to understand what drives people to feel for others and lend their helping hand during disasters by simply reading hastily written messages on social media. For this purpose, we analyzed 1.6 million tweets that were posted on Twitter when Chennai, a southern Indian city, was significantly affected by heavy rains and subsequent flooding in 2015.

A. *Chennai rains and floods-2015*

TABLE II. DESCRIPTIVE DATASET STATISTICS

Twitter messages (tweets)	**Absolute numbers**	**Percentage**
Total tweets	1,658,220	100%
Retweets	1,226,098	73.94%
Original tweets of Retweets	141,941	08.56%
Tweets never got retweeted	290,181	17.50%
Mean Retweet Ratio	8.64	
Total Unique Twitter Users	209,644	

Chennai received 34 times the normal daily amount of rain in the first week of December 2015. The downpour intensified on December 2nd, leading to massive flooding. Subsequently it affected homes, hospitals, roads, railway tracks and the city's airport. Three million people suffered due to the lack of access to food, and drinking water [46]. Social media played an important role in times of distress as people used it extensively to reach out to the affected people, coordinated their search and rescue activities, and also for food distribution [47]. Hence, we analyzed Twitter data of the Chennai floods to understand how social presence on social media helped affected people to participate in relief activities [48].

B. *Data collection*

In this empirical work, we analyze the tweets that were shared when Chennai was affected by heavy rains and subsequent flooding. We used Radian 6 tool to collect the Twitter messages by using hashtags #TNflood, #chennaiRains, #chennafloods #chennaiRainsHelp, #IndiaWithChennai and #chennaiMicro. We collected data from November 30th to December 16th, 2015. The total dataset consists of 1.65 million tweets with 209,644 unique users. Since the radian 6 does not provide some of the Twitter attributes like retweet status, retweet count, and original tweet Id for retweets, with the help of tweet Ids using open Twitter API we again downloaded the whole dataset and separated the original tweets from the retweets based on the retweet status information. Interestingly, 74% of total dataset consists of retweets. The descriptive statistics are presented in Table II.

C. *Methodology*

Fig. 3 represents the overall methodology of our research work in this case study. From the media related characteristics, social media falls into the medium category from social presence point of view.

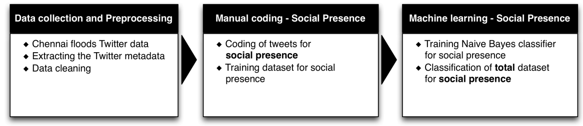

Fig. 3 Overall Methodology of Empirical Study

People perceive the presence of others on communicated medium [49]. In general, the concepts associated with social presence [50] are intimacy [51] and immediacy [52]. We operationalized the concepts *intimacy* and *immediacy* according to the disaster situations to conduct the content analysis. Our interest is to understand how these concepts were expressed in a tweet content.

TABLE III. OVERVIEW OF THE SOCIAL PRESENCE CONCEPTS AND THEIR OPERATIONALIZATION

categories	**Description**	**Tweet Content**
Intimacy	Feeling closeness: sharing road closure info, asking for help on behalf of others	Is there any way to provide any form of support monetary and supplies? #ChennaiRains
	Moral support: stand by people, providing hope for best	Hats off to the fighting spirit of Chennai! A salute all those volunteers who have been helping relentlessly! #staystrongchennai #chenna
Immediacy	Urgent action is needed: different types of rescue requests	any doctors in mudichur area? One pregnant lady in labour...no access to boat. Here is the Contact: 9940203871#chennairains
	Sharing information: to provide shelter, food and help	#food available at #tnagar gurdwara 9094790989#ChennaiRainsHelp #ChennaiMicro #ChennaiVolunteer https://t.co/JD1BAdWXSe

The operationalization of concepts for social presence is shown in Table III. Tweets displaying moral support, creating a feeling of closeness are placed into intimacy category. The tweet content that evokes a feeling of need for urgency and demands an immediate action placed into immediacy category. We used these concepts as our coding scheme in the subsequent step.

To automate the classification of the entire dataset, constituting 1.65 million tweets, along the intimacy and immediacy concepts, we have adopted the automated text classification approach using a supervised machine learning algorithm, the Naïve Bayes classifier. As mentioned above we used 1) Intimacy, 2) Immediacy and 3) None categories as coding scheme for the automated text classification and to prepare training dataset for social presence. The tweets that do not belong to either Intimacy or Immediacy are placed into the “None” category. Initially to ensure validity, authors discussed the concepts behind Intimacy and Immediacy extensively by taking examples, what constitutes the categories and what does not. Further authors independently coded randomly selected sample of approximately 500 tweets. Afterwards, the results were compared and both coders discussed to solve the discrepancies about the concepts. The inter coder agreement matrix for the text coded by the authors is presented in Table IV.

TABLE IV. INTER-CODER AGREEMENT MATRIX OF SOCIAL PRESENCE

		Coder 1			
		Intimacy	Immediacy	None	Marginal Totals
Coder 2	Intimacy	38 (0.06)	7 (0.01)	2 (0.01)	47 (0.08)
	Immediacy	7 (0.01)	83 (0.14)	6 (0.01)	96 (0.16)
	None	1 (0.01)	5 (0.01)	431 (0.74)	437 (0.76)
	Marginal Totals	46 (0.08)	95 (0.16)	439 (0.76)	580 (1.00)

In general, an inter-coder agreement of 0.40 to 0.80 is considered as a good indicator of valid agreement between coders [53]. In case of social presence, the proportion of agreement by chance = (0.08 * 0.08) + (0.16 * 0.16) + (0.76 * 0.76) = 0.62. and Cohen's Kappa value for social presence can be calculated by using the standard formula as: (0.95 – 0.62) / (1 – 0.62) = 0.868. In the final step, to get the trained dataset one of the authors conducted content analysis of randomly selected 5000 tweets for social presence concepts in the subsequent phase according to our coding scheme. Among the coded tweets, 80% of training set was used to train the classifier and the remaining 20% of the tweets were used to test the accuracy of the classifier.

TABLE V. PERFORMANCE MEASURES OF TEXT CLASSIFICATION USING NAÏVE BAYES CLASSIFIER

Model	Labels	Precision	Recall	F-Measure	Accuracy
Social Presence	Intimacy	0.158	0.466	0.236	0.805
	Immediacy	0.626	0.520	0.568	
	None	0.943	0.858	0.898	

D. Data analysis and results

In general, performance of a machine learning algorithm can be described by four measures: precision, recall, F-measure, and accuracy [54, 55]. The performance measures about text classification of tweet content using Naïve Bayes classifier are presented in Table V. The overall accuracy of the classifier is fairly high (i.e. around 80% of the prediction are accurate for correct predictions). In terms of precision and recall, immediacy received fairly good values in contrast to intimacy. In case of the F-measure, a value around 0.6-0.7 indicates a fairly better performance and the F-measure values for all labels/categories indicate reasonably good performance except for the categories: intimacy. The social presence categories in the total dataset are presented below in Table VI.

TABLE VI. SOCIAL PRESENCE CATEGORIES IN THE TOTAL DATASET

Social Presence	Total tweets	Percentage	Retweets	Percentage
Intimacy	285,292	17.20%	226,697	18.49%
Immediacy	335,877	20.26%	291,516	23.78%
None	1,037,051	62.54%	707,885	57.73%
Total	1,658,220	100%	1,226,098	100%

From the total dataset, 37% of tweets are classified as social presence categories. The remaining 63% of tweets belongs to the none category. The reason could be as mentioned in the previous studies [26, 28] people tend to post suggestions, comments, criticism and also discuss or vent their frustration about media or government. However, among the 37% classified tweets, 20% tweets belong to immediacy category and among the retweet category around 24% of retweets belongs to immediacy. Most importantly, 74% of the total dataset constitute retweets and the proportion of social presence retweets among the total retweets is higher with around 42% of retweets belonging to intimacy and immediacy, which indicates that social presence tweets are retweeted more than the tweets that belong to none category. Our results reveal that most of the immediacy tweets are conveying needs and urgencies of people. People perceive the online social presence through the messages sent by the online users and hence are drawn to Twitter to fulfill their social need for connections. Moreover, because of the perceived higher levels of social presence, people continue to use and interact more on Twitter [56].

V. DISCUSSION

In this section, we will discuss about the opportunities and challenges from both social media data point of view and the methods point of view. The following observations are beneficial to practitioners and academics who are interested in social media analytics in general. In particular, these insights might be useful to the emergency management agencies who wanted to integrate social media in their routines and processes.

A. Social media-Provider of information

Social media has become a potential platform to gain and share valuable information during the different phases of a disaster. People are using different social media platforms, such as Twitter [57], Facebook [57, 58], Flickr [59] during disasters. Based on our first case study [29], it is evident that during disasters people share first-hand observations and experiences on social media. Moreover, people who are in the vicinity of an event or witnessing an unfolding event immediately share information on social media [15]. Hence social media has become a good source of information to recognize and tap into an unfolding or ongoing crisis/event. In one way, it complements the existing early warning systems because of rapid information dissemination [21] and in the other way, early warnings can be detected on social media by monitoring information [60].

For example, emergency situation awareness (ESA) system was developed to detect the earthquakes based on tweet burst detection method on Twitter [61]. Keywords play an important role in the burst detection in the twitter stream. Often, researchers with the help of keywords and hashtags extract the data using Twitter application programming interface (API) or tools like Radian 6 [29, 62]. Along with textual data, rich multimedia data is also available on social media platforms in the form of pictures and videos [59, 62]. Given the importance of information embedded in the images, micro mappers [63], with the help of crowds assessing the disaster damages by viewing the pictures. We believe that the disaster related social media data is useful in creating situation awareness and also in disseminating information during disasters. Situation awareness is important for disaster management agencies to make appropriate decisions in disaster response activities. Similarly, creating awareness among the people by disseminating disaster preparedness and warnings using social media is also quite important. However, we also noticed certain challenges in using disaster social media data.

- Data volume and velocity: These two aspects are major issues in handling and processing social media data in real-time as ongoing social media activity explodes during disasters. For example, 11 million tweets were produced just in a few days during hurricane Sandy.
- Data Relevance: From the huge volumes of social media data generated during disasters, identifying important useful information for disaster relief agencies is a challenge, as a major portion of the data contains unrelated information such as commercials and also repeated, re-posted messages (74% of the Chennai dataset consists of retweeted messages).
- Data Integration: During disasters, people use different social media platforms (e.g. Twitter and Facebook) to share information. Integrating and unifying information from these different platforms is a challenge due to different data formats. Integrating data generated from different sources (eyewitnesses, traditional media, outsiders) and different languages [12] is also a challenge.
- Rumors and Fake News: Validity and quality of information is also a challenge. Rumors and fake news create panic among the people. Most importantly, given the nature of social media, these rumors also spread very fast [64-66].

B. Geo-information

Disaster related social media data originates from different parts of the world, geographical coordinates [43] allows to filter and retrieve the information from particular disaster affected geographic region. Meta data of a tweet provides geographic information in the form of latitude and longitudinal coordinates. With the help of geo-location information one can select the tweets produced from one specific geographic location and visualize the information of an event on a map as presented in Fig. 1 as demonstrated in our first case study [29]. Given the importance of geo-location information, systems such as emergency situation awareness (ESA) [61] tag the geo information for event detection and systems like SensePlace2 [35] filter and extract geographic information from tweets to visualize the information in maps. Real time information from the disaster affected location is very important for emergency management officials to make tactical decisions. However, some of the challenges to consider:

- According to prior research and also based on our case study, only 1% of data consists of geographic information. An explanation could be that the end users are not revealing the location information due to privacy concerns.
- Applying different methods to infer the location information is considered as privacy intrusion.
- Some of the social media platforms do not provide geo-location information at all. For example Facebook offers a “check in” option [67], but does not provide geo-information in the metadata.

C. Digital emergent groups

During disasters, people in the affected area play a big role in taking part in immediate rescue operations. However, it is evident from our second case study, despite the lack of face to face interactions, people are perceiving the presence of others on the social media platforms. The feeling of connectedness in distress situations facilitating the emergence of a new type of volunteers who are not only sharing information and coordinating the rescue activities through social media in affected areas but are also helping in assessing the damage or collaboratively collecting the necessary information that is useful to the affected individuals. So far based on prior literature [33, 34, 68] we know that organizations are interested in social media to communicate disaster relevant information to people. However, new structures are evolving at the community level [58], hence it provides an opportunity for disaster management agencies while monitoring the information on social media can also engage with digital volunteers directly in conversations and in collaboration to perform relief operations [68]. However, the challenge is:

- These new types of volunteers only emerge during the times of disasters on an ad-hoc basis, hence it is a challenge to get in contact with them and include them in the disaster relief activities.

D. Methods to analyze the textual data

In general, in order to analyze the textual data from disaster related social media content, different methods have been applied by the research communities: manual content analysis and computational methods. As part of computational methods, two techniques: supervised and unsupervised machine learning techniques are in use. Classifiers fall into the supervised machine learning technique and topic modeling is considered as belonging to the unsupervised approach. However, in our case studies, we have applied both manual content analysis and classifiers. In our first case study, we analyzed the data manually by applying manuals content analysis, but in our second case study we applied supervised machine learning technique to classify the data for identifying social presence concepts in tweet content.

Manual content analysis is a very basic method, where human coders analyze the data and categorize the information. With the help of manual content analysis either one could derive categories from the data by analyzing it or one could analyze and extract the information based on pre-defined categories. One could also use crowd sourcing platforms, for example, Amazon Mechanic Turk or Crowd flower, to analyze the data with the help of crowds by providing instructions and coding schemes. However, this method is perfect and feasible when the dataset is rather "*small*", because analyzing millions of tweets is humanly not possible. However, there are systems like CrisisTracker [38] and Artificial Intelligence for Disaster Response (AIDR) [39] take the help of crowds to annotate the data. However, with the help of annotated data AIDR classifies the rest of the data automatically.

With the help of classifiers, data can be classified automatically by using a machine learning algorithm. Using well defined training dataset and supervised algorithm, one can classify "*huge amounts*" of data or to extract the information easily. In our second case study, with the help of classifiers, we tested the social presence theory and extracted intimacy and immediacy tweets from the huge amounts of data. Previous research used classification approaches to extract situation awareness and informative messages from social media data [32, 69] and there are also a few systems developed based on text classification approach: Tweedr [37], ESA [61]. Some of the challenges associated with classifiers are:

- Well-defined training sets are necessary for building the classifiers, which might need resources such as a human coder to prepare them.
- Given the nature of unstructured data, getting the high accuracy of classifier is difficult.
- The classifier developed in one disaster scenario might not work well in another situation. For example, classifiers that are developed for earthquakes might not be suitable for floods.

VI. CONCLUSION

In this paper, based on the two case studies and the discussion above, we argue that emergency managers can use social media to disseminate disaster relevant information, gain situation awareness from the affected individuals, and can start a dialogue with people to reduce the risk [68]. Based on recent research [33, 70] emergency officials are still skeptical on fully integrating social media in their organizational routines. The reason could be the challenges that we mentioned previously regarding data and methods. However, in few cases, emergency organizations started using social media for disseminating information and can be coined as aspirational or early adopters of social media [71]. In order to reach to the next level, it is important for organizations to invest time, resources and personnel [68] to harvest the benefits of social media. Most importantly to gain value by using social media, organizations need to align their goals with the strategic initiatives.

However, based on our research we addressed certain challenges along with opportunities. Firstly, we argue that the various analysis techniques can be used to extract and understand the valuable information that is generated during the disaster situations. At the same time focusing on different case studies, we tried to understand how people used social media in different contexts during natural disasters such as floods and hurricanes. Based on case studies we also strongly argue that automated text analytical methods can be used to analyze and understand the user generated content. Most importantly, our current research helped us towards developing or applying other automated methods. For example, in future we would like to apply unsupervised topic modeling method to extract the hidden information from the user-generated texts without any human intervention. At the same time, we argue that the research community needs to focus on supervised and dictionary based approaches to classify the social media data, specifically focusing on extracting time relevant information with needs and urgencies that is embedded in social media data.

REFERENCES

[1] A. Mukkamala and R. Beck, "Disaster management and social media use for decision making by humanitarian organizations," in System Sciences (HICSS), 2016 49th Hawaii International Conference on, 2016, pp. 1379-1385.

[2] N.-S. Lee, S. Hirschmeier, S. Müller, and L. J. Luz, "Enablers in Crisis Information Management: A Literature Review," 2017.

[3] A. Cozzolino, Humanitarian logistics: cross-sector cooperation in disaster relief management: Springer Science & Business Media, 2012.

[4] R. Stiegler, S. Tilley, and T. Parveen, "Finding family and friends in the aftermath of a disaster using federated queries on social networks and websites," in Web systems evolution (WSE), 2011 13th IEEE international symposium on, 2011, pp. 21-26.

[5] P. H. Debarati Guha-Sapir , Regina Below, "Statistical Review 2015 The numbers and trends," Centre for Research on the Epidemiology of Disasters (CRED), Brussels, Belgium2016.

[6] L. K. Comfort, K. Ko, and A. Zagorecki, "Coordination in rapidly evolving disaster response systems: The role of information," American Behavioral Scientist, vol. 48, pp. 295-313, 2004.

[7] P. Meier. (2016). The Value of Timely Information During Disasters (Measured in Hours). Available: https://irevolutions.org/2016/04/06/the-value-of-timely-information-during-disasters-measured-in-hours/

[8] N. Cook, Enterprise 2.0: how social software will change the future of work: Gower Publishing, Ltd., 2008.

[9] T. Simon, A. Goldberg, and B. Adini, "Socializing in emergencies—A review of the use of social media in emergency situations," International Journal of Information Management, vol. 35, pp. 609-619, 2015.

[10] A. Olteanu, C. Castillo, F. Diaz, and S. Vieweg, "CrisisLex: A Lexicon for Collecting and Filtering Microblogged Communications in Crises," in ICWSM, 2014.

[11] A. Olteanu. (2017). CrisisLex.org. Available: http://crisislex.org/

[12] A. Olteanu, S. Vieweg, and C. Castillo, "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, 2015, pp. 994-1009.

[13] S. Vieweg, A. Hughes, K. Starbird, and L. Palen, "Supporting situational awareness during emergencies using microblogged information," SIGCHI 2010, 2010.

[14] K. Starbird and L. Palen, "Voluntweeters: Self-organizing by digital volunteers in times of crisis," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2011, pp. 1071-1080.

[15] M. Krstajic, C. Rohrdantz, M. Hund, and A. Weiler, "Getting there first: Real-time detection of real-world incidents on twitter," in 2nd Workshop on Interactive Visual Text Analytics : Task-Driven Analysis of Social Media Content with Visweek'12., 2012.

[16] J. Sutton, L. Palen, and I. Shklovski, "Backchannels on the front lines: Emergent uses of social media in the 2007 southern California wildfires," in In: Proceedings of the 5th International ISCRAM Conference, 2008, pp. 624-632.

[17] I. Shklovski, L. Palen, and J. Sutton, "Finding Community Through Information and Communication Technology During Disaster Events," in Proceedings of the 2008 ACM conference on Computer supported cooperative work, ed: ACM, 2008, pp. 1-10.

[18] L. Palen, "Online social media in crisis events," Educause Quarterly, vol. 31, pp. 76-78, 2008.

[19] L. Palen, S. Vieweg, S. B. Liu, and A. L. Hughes, "Crisis in a Networked World: Features of Computer-Mediated Communication in the April 16, 2007, Virginia Tech Event," Social Science Computer Review, vol. 27, pp. 467-480, 2009.

[20] A. T. Chatfield, H. J. Scholl, and U. Brajawidagda, "# Sandy Tweets: Citizens' Co-Production of Time-Critical Information during an Unfolding Catastrophe," in 2014 47th Hawaii International Conference on System Sciences, 2014, pp. 1947-1957.

[21] A. T. Chatfield and U. Brajawidagda, "Twitter early tsunami warning system: A case study in indonesia's natural disaster management," in System sciences (HICSS), 2013 46th Hawaii international conference on, 2013, pp. 2050-2060.

[22] A. Java, X. Song, T. Finin, and B. Tseng, "Why we twitter: understanding microblogging usage and communities," in In: Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis, 2007, pp. 56-65.

[23] A. Ahmed and J. Sargent, "Analysis of Post-Crisis Twitter Communication: A Study of the Iquique, Chile Earthquake," in In: Proceedings of the Australasian Conference on Information Systems 2014.

[24] T. Heverin and L. Zach, "Microblogging for Crisis Communication : Examination of Twitter Use in Response to a 2009 Violent Crisis in the Seattle-Tacoma , Washington Area," presented at the In: Proceedings of the 7th International ISCRAM Conference, 2010.

[25] S. Muralidharan, L. Rasmussen, D. Patterson, and J.-H. Shin, ""Hope for Haiti: An analysis of Facebook and Twitter usage during the earthquake relief efforts"," Public Relations Review, vol. 37, pp. 175-177, 2011.

[26] Y. Qu, C. Huang, P. Zhang, and J. Zhang, "Microblogging after a major disaster in China: a case study of the 2010 Yushu earthquake," in In: Proceedings of the ACM 2011 conference on Computer supported cooperative work, 2011, pp. 25-34.

[27] S. Vieweg, A. L. A. Hughes, K. Starbird, and L. Palen, "Microblogging during two natural hazards events: what twitter may contribute to situational awareness," in In: Proceedings of the 28th international conference on Human factors in computing systems - CHI '10, 2010, pp. 1079-1088.

[28] S. E. Vieweg, "Situational awareness in mass emergency: A behavioral and linguistic analysis of microblogged communications," PhD thesis, University of Colorado at Boulder, 2012.

[29] A. Mukkamala and R. Beck, "Enhancing Disaster Management Through Social Media Analytics To Develop Situation Awareness What Can Be Learned From Twitter Messages About Hurricane Sandy?," presented at the The Pacific Asia Conference on Information Systems (PACIS) 2016.

[30] K. Starbird, L. Palen, A. L. Hughes, and S. Vieweg, "Chatter on The Red: What Hazards Thret Reveals about the Social Life of Microblogges Information," in In: Proceedings of the Conference on Computer Supported Cooperative Work (CSCW), 2010, pp. 241-250.

[31] D. Murthy and S. A. Longwell, "Twitter and disasters: The uses of Twitter during the 2010 Pakistan floods," Information, Communication & Society, vol. 16, pp. 837-855, 2013.

[32] S. Verma, S. Vieweg, W. J. Corvey, L. Palen, J. H. Martin, M. Palmer, et al., "Natural Language Processing to the Rescue? Extracting" Situational Awareness" Tweets During Mass Emergency," in ICWSM, 2011.

[33] C. Reuter, T. Ludwig, M.-A. Kaufhold, and T. Spielhofer, "Emergency services′ attitudes towards social media: A quantitative and qualitative survey across Europe," International Journal of Human-Computer Studies, vol. 95, pp. 96-111, 2016.

[34] C. Ehnis and D. Bunker, "Social media in disaster response: Queensland Police Service-public engagement during the 2011 floods," in ACIS 2012: Location, location, location: Proceedings of the 23rd Australasian Conference on Information Systems 2012, 2012, pp. 1-10.

[35] A. M. MacEachren, A. Jaiswal, A. C. Robinson, S. Pezanowski, A. Savelyev, P. Mitra, et al., "Senseplace2: Geotwitter analytics support for situational awareness," in Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on, 2011, pp. 181-190.

[36] F. Abel, C. Hauff, G.-J. Houben, R. Stronkman, and K. Tao, "Twitcident: fighting fire with information from social web streams," in Proceedings of the 21st International Conference on World Wide Web, 2012, pp. 305-308.

[37] Z. Ashktorab, C. Brown, M. Nandi, and A. Culotta, "Tweedr: Mining twitter to inform disaster response," in ISCRAM, 2014.

[38] J. Rogstadius, M. Vukovic, C. Teixeira, V. Kostakos, E. Karapanos, and J. A. Laredo, "CrisisTracker: Crowdsourced social media curation for disaster awareness," IBM Journal of Research and Development, vol. 57, pp. 4: 1-4: 13, 2013.

[39] M. Imran, C. Castillo, J. Lucas, P. Meier, and S. Vieweg, "AIDR: Artificial intelligence for disaster response," in Proceedings of the 23rd International Conference on World Wide Web, 2014, pp. 159-162.

[40] Y. Neria and J. M. Shultz, "Mental health effects of Hurricane Sandy: Characteristics, potential aftermath, and response," Jama, vol. 308, pp. 2571-2572, 2012.

[41] A. Zubiaga. (2015). Dataset: Hurricane Sandy tweets. Available: http://www.zubiaga.org/datasets/hurricane-sandy-tweets/

[42] Tableau. (2015). Available: http://www.tableau.com/

[43] M. Graham, S. A. Hale, and D. Gaffney, "Where in the world are you? Geolocation and language identification in Twitter," The Professional Geographer, vol. 66, pp. 568-578, 2014.

[44] P. Burnap, O. Rana, M. Williams, W. Housley, A. Edwards, J. Morgan, et al., "COSMOS: Towards an integrated and scalable service for analysing social media on demand," International Journal of Parallel, Emergent and Distributed Systems, vol. 30, pp. 80-100, 2015.

[45] CartoDB. (2015). Available: https://cartodb.com/

[46] T. Misra. (2015). Chennai Floods: A Perfect Storm of Climate Change and Failed Planning. Available: http://www.citylab.com/weather/2015/12/chennai-flood-climate-change-urban-planning-cop21-paris/419349/

[47] V. Pandey. (2015). How Chennai residents are helping flood affected via social media. Available: http://www.bbc.com/news/world-asia-india-34981360

[48] A. Mukkamala and R. Beck, "Presence of Social Presence during Disasters," in Pacific Asia Conference on Information Syatems, 2017.

[49] A. M. Kaplan and M. Haenlein, "Users of the world, unite! The challenges and opportunities of Social Media," Business horizons, vol. 53, pp. 59-68, 2010.

[50] J. Short, E. Williams, and B. Christie, The social psychology of telecommunications, 1976.

[51] M. Argyle and J. Dean, "Eye-contact, distance and affiliation," Sociometry, vol. 28, pp. 289-304, 1965.

[52] M. Wiener and A. Mehrabian, Language within language: Immediacy, a channel in verbal communication: Ardent Media, 1968.

[53] S. Stemler, "An overview of content analysis," Practical assessment, research & evaluation, vol. 7, pp. 137-146, 2001.

[54] D. M. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," School of Informatics and Engineering, Flinders University , Adelaide, Australia. 2229-3981, 2011.

[55] Y. Yang, "An evaluation of statistical approaches to text categorization," Information retrieval, vol. 1, pp. 69-90, 1999.

[56] [56] S. Han, J. Min, and H. Lee, "Antecedents of social presence and gratification of social connection needs in SNS: a study of Twitter users

and their mobile and non-mobile usage," International Journal of Information Management, vol. 35, pp. 459-471, 2015.

[57] K. Eismann, O. Posegga, and K. Fischbach, "COLLECTIVE BEHAVIOUR, SOCIAL MEDIA, AND DISASTERS: A SYSTEMATIC LITERATURE REVIEW," presented at the European Conference on Information Systems, 2016.

[58] L. Kaewkitipong, C. C. Chen, and P. Ractham, "A community-based approach to sharing knowledge before, during, and after crisis events: A case study from Thailand," Computers in Human Behavior, vol. 54, pp. 653-666, 2016.

[59] S. B. Liu, L. Palen, J. Sutton, A. Hughes, and S. Vieweg, "In search of the bigger picture: The emergent role of on-line photo sharing in times of disaster," in Proceedings of the information systems for crisis response and management conference (ISCRAM), 2008, pp. 969-980.

[60] T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake shakes Twitter users: real-time event detection by social sensors," in WWW '10: Proceedings of the 19th international conference on World wide web, ed, 2010, pp. 851-860.

[61] J. Yin, S. Karimi, B. Robinson, and M. Cameron, "ESA: Emergency Situation Awareness via Microbloggers," in Proceedings of the 21st ACM international conference on Information and knowledge management - CIKM '12, ed: ACM, 2012, pp. 2701-2703.

[62] M. Imran, C. Castillo, F. Diaz, and S. Vieweg, "Processing social media messages in mass emergency: A survey," ACM Computing Surveys (CSUR), vol. 47, p. 67, 2015.

[63] P. Meier. (2016-08-09). MicroMappers Digital Disaster Response. With a Single Click! Available: http://www.micromappers.com/

[64] M. Mendoza, B. Poblete, and C. Castillo, "Twitter under crisis: can we trust what we RT?," in Proceedings of the first workshop on social media analytics, 2010, pp. 71-79.

[65] O. Oh, K. H. Kwon, and H. R. Rao, "An Exploration of Social Media in Extreme Events: Rumor Theory and Twitter during the Haiti Earthquake 2010," in ICIS, 2010, p. 231.

[66] K. Starbird, J. Maddock, M. Orand, P. Achterman, and R. M. Mason, "Rumors, false flags, and digital vigilantes: Misinformation on twitter after the 2013 boston marathon bombing," iConference 2014 Proceedings, 2014.

[67] P. Hannay and G. Baatard, "GeoIntelligence: Data mining locational social media content for profiling and information gathering," 2011.

[68] C. Wukich, "Social media use in emergency management," Journal of Emergency Management, vol. 13, pp. 281-295, 2015.

[69] M. Imran, S. M. Elbassuoni, C. Castillo, F. Diaz, and P. Meier, "Extracting information nuggets from disaster-related messages in social media," Proc. of ISCRAM, Baden-Baden, Germany, 2013.

[70] A. Sheombar, C. Urquhart, T. Ndhlovu, and P. Ravesteijn, "Social Media in the Context of Development: A Case Study of Dutch NGOs," in ECIS, 2015.

[71] S. LaValle, E. Lesser, R. Shockley, M. S. Hopkins, and N. Kruschwitz, "Big data, analytics and the path from insights to value," MIT sloan management review, vol. 52, p. 21, 2011.