Research Paper

JIS

Journal of Information Science
1–12

DOI: 10.1177/0165551519828620
journals.sagepub.com/home/jis
SAGE

Twitter speaks: A case of national disaster situational awareness

Amir Karami
School of Library and Information Science, University of South Carolina, USA

Vishal Shah
Computer Science & Engineering Department, University of South Carolina, USA

Reza Vaezi
Coles College of Business, Kennesaw State University, USA

Amit Bansal
Computer Science & Engineering Department, University of South Carolina, USA

Abstract
In recent years, we have been faced with a series of natural disasters causing a tremendous amount of financial, environmental and human losses. The unpredictable nature of natural disasters behaviour makes it hard to have a comprehensive situational awareness (SA) to support disaster management. Using opinion surveys is a traditional approach to analyse public concerns during natural disasters; however, this approach is limited, expensive and time-consuming. Luckily, the advent of social media has provided scholars with an alternative means of analysing public concerns. Social media enable users (people) to freely communicate their opinions and disperse information regarding current events including natural disasters. This research emphasises the value of social media analysis and proposes an analytical framework: Twitter Situational Awareness (TwiSA). This framework uses text mining methods including sentiment analysis and topic modelling to create a better SA for disaster preparedness, response and recovery. TwiSA has also effectively deployed on a large number of tweets and tracks the negative concerns of people during the 2015 South Carolina flood.

Keywords
Flood; natural disaster; sentiment analysis; situational awareness; text mining; topic model; Twitter

1. Introduction

The growth of social media has provided an opportunity to track people's concerns in a new way. Twitter has 316 million users [1], provides capability of real-time feedback and utilises time stamp to provide conversation updates to users. Therefore, Twitter's potential as a reliable and relevant as a data source is evident and provides a unique opportunity to understand users' concerns [2,3]. In contrast to surveys and surveillance networks that can take weeks or even years to collect data, Twitter is publicly accessible with no waiting time.

Situational awareness (SA) is 'all knowledge that is accessible and can be integrated into a coherent picture, when required, to assess and cope with a situation' [4]. Obtaining reliable and accurate information and providing access to knowledge derived from the collected data is the main concern in SA. Thus, SA and its inherent processes play an essential role in helping people facing a natural disaster. During a natural disaster, people immediately try to collect or share

Corresponding author:
Amir Karami, School of Library and Information Science, University of South Carolina, 1501 Greene St., Columbia, SC 29208, USA.
Email: karami@sc.edu

information to keep them away from possible dangerous situations [5]. Developing high level of SA by analysing real-time big data helps to become more effective in disaster management [6,7].

In October 2015, hurricane Joaquin led to an unusually high amount of rainfall in the state of South Carolina (SC). The precipitation ranged from 11 to 26 in in different areas and caused server flooding in the state. This flood caused more than US$12 billion damage to roads, homes and infrastructures with 19 dead [8,9]. Twitter users expressed their feelings and opinions during the 2015 SC flood through their Tweets. This valuable time-sensitive data can improve SA development and give all involved parties a more accurate picture of the situation during the flood.

Crowdsourced data such as social media data are rapid and accessible source for SA development [10]. In other words, understanding users' concerns and their emotions can help disaster managers develop a better real-time SA and decision-making plans [11,12]. During the height of the 2015 SC flood, thousands of people were twitting about the flood on an average of ~3000 tweets per hour. One can imagine that manually analysing this many tweets is not easily possible; therefore, new customised approaches are necessary to disclose hidden sentiment and sentiment features from social media. This study proposes an analytical framework, Twitter Situational Awareness (TwiSA), to analyse unstructured flood-related tweets to better understand temporal concerns of people affected by this natural disaster. The proposed framework employs both sentiment analysis and topic modelling to uncover temporal patterns of public concerns to provide a better SA. We apply TwiSA on a large number of tweets and track the negative concerns of people during the 2015 SC flood.

2. Related work

Twitter data have been used for a broad range of applications such as business [13, 93, 94, 95, 96], politics [14, 77, 78, 79, 80, 81] and health [15, 82, 83, 84, 85, 86]. In this section, we review the applications of Twitter data in natural disasters including fire, flood, earthquake, hurricane and typhoon, and volcano eruption.

Abel et al. [16] developed a tool, called Twitcident, to collect fire-related tweets and search tweets using queries. Sinnappan et al. [17] collected tweets and used qualitative approach to find meaningful categorisation of fire-related tweets.

Earthquake-related studies proposed frameworks to improve disaster management process. These included using data mining and natural language processing for damage detection and assessment of earthquakes [18], proposing a probabilistic spatiotemporal model for reporting earthquake-related events [19], developing a detection algorithm based on the frequency of tweets to detect earthquake [20], applying classifier methods on tweets to detect earthquake [21,22], using qualitative approach to analyse people's behaviour after an earthquake [23], applying a keyword level analysis to track social attitudes during and after an earthquake [24] and analysing dynamic of rumour mill in tweets [25].

Literature also contains methods for disaster management during hurricanes and typhoon. This research area analysed online public communications by police and fire departments in Twitter and Facebook [26], used classification methods to detect fake images during the hurricane [27] and analysed external factors such as geolocation and internal factors (e.g. geolocation and stakeholders power) in a small sample of tweets to find patterns of information dissemination [28]. Recently, a study mapped online users' sentiments to track their changes during a hurricane [11]. Twitter data were also used to help disaster management using qualitative content analysis to categorise information inside the Twitter data during a volcano eruption [29].

Flood-related studies developed numerous frameworks to improve disaster management during floods. These studies tracked users' behaviour during a flood [30], combined georeferenced social media messages, and geographic features of flood phenomena [31], used machine learning techniques to find tweets reporting damage [32], mapped geo-tagged tweets to directly support geotechnical experts for reconnaissance purposes [33], analysed crisis relevant information from Twitter using related keywords and geo-tagged tweets [34], investigated the correlation between location and the types of URLs inside tweets [35], explored the discursive aspects of Twitter communication with qualitative analysis [36], analysed the related tweets with qualitative data coding to find useful information [10] and used keyword analysis to determine the types of information in the flood-related tweets [37].

Although it is evident from the literature that past research on the use of social media during a natural disaster provides useful insight into disaster management, unpredictable nature of natural disasters provides a great motivation for disaster responders to constantly improve disaster management by exploring new perspectives. Thus, we believe we can improve disaster management through better integration of SA into the management practices specially when it comes to people concerns and negative feelings. To our knowledge, no other research studied the concerns of people behind their negative feelings for having a better SA. This study proposes an analytical framework to explore a large number of tweets using sentiment analysis and topic modelling for tracking temporal patterns of public negative concerns during the 2015 SC flood.

3. Methodology and results

This article proposes a framework, TwiSA, with four components: data collection, sentiment analysis, temporal topic discovery and topic content analysis.

3.1. Data collection

There are two methods to collect a large number of tweets in TwiSA. One method is to retrieve data using Twitter [38] Application Programming Interfaces (APIs), and the other method is to ask an independent service to provide the data. The first method, despite being free, is limited in that one can only retrieve a small portion of relevant tweets. Therefore, we asked an independent company, GNIP [39], to extract 100% of tweets related to our research. Similar to Palen et al. [30], we ordered the queries in Table 1 to collect tweets for the days having flood danger or flood side effect. One million tweets were filtered based on 13 days in October 2015. We chose this time frame because it represents a period that starts with the highest amount of rainfall (3 October 2015) and ends with the last day for boil advisory (15 October 2015) [8]. It is worth mentioning that we did not remove retweets and the tweets containing URLs from our dataset but the word 'rt' and URLs were removed from the collected tweets. Table 2 shows a sample of tweets that are related to insurance and damage issues using the #scflood query.

Table 1. Queries for Twitter data collection.

#floodsc **OR** #scflood2015 **OR** #SCFloodRelief **OR** #southcarolinastrong **OR** #prayforsc **OR** #scflood **OR** #scflooding **OR** #FloodGSSCMMwithlove **OR** #floodingsc **OR** #flood **OR** flood

Table 2. A sample of tweets.

Tweet 1	*How many South Carolinians have flood insurance? Few. They'll be looking to federal gov't or their tight-fisted governor for help. #SCFlood*
Tweet 2	*Damage inside Student Activities room at Westwood. #SCFlood*

3.2. Sentiment analysis

Sentiment analysis discloses the overall feelings inside text data [40]. Learning-based and lexicon-based methods are two main approaches for sentiment analysis [41]. The first approach uses machine learning classifiers when there is prior knowledge about data categories. In this case, a sample of the data is first labelled by human raters such as assigning spam and non-spam labels to a sample of emails [40]. The second approach, a cost-effective one, finds the frequency of a predefined dictionary of positive and negative terms to disclose sentiment in the data when there is not any prior knowledge about its categories [42]. We did not have any prior knowledge about the categories of the tweets in this research; therefore, we applied the second approach to find positive, negative and neutral tweets.

Linguistic Inquiry and Word Count (LIWC) [43] is a linguistics analysis tool that reveals thoughts, feelings, personality and motivations in a corpus based on lexicon-based approach. This software assumes that the frequency of words can reveal overall tone in the corpus [97–99]. This tool applies a simple word-count process in text documents and maps each word in the documents on the already developed internal LIWC dictionaries in different categories such as negative emotion [44].

LIWC has good sensitivity value, specificity value and English proficiency measure [45]. Comparing to competitors such as deep learning, this tool can be used alone or be combined with other methods [46]. Applying LIWC on the collected tweets shows that there are 217,074 negative tweets, 529,150 neutral tweets and 217,183 positive tweets. When people are exposed to unpredictable events, they show unpleasant negative feelings that come from the presence of threats with undesirable outcomes [47,48]. We used the negative tweets for the next step to identify unpleasant elements of the SC flood through temporal topic discovery analysis. Figure 1 shows that the number of negative tweets per day has changed from ~12,700 tweets to ~30,000 tweets with an average ~16,700 tweets per day.

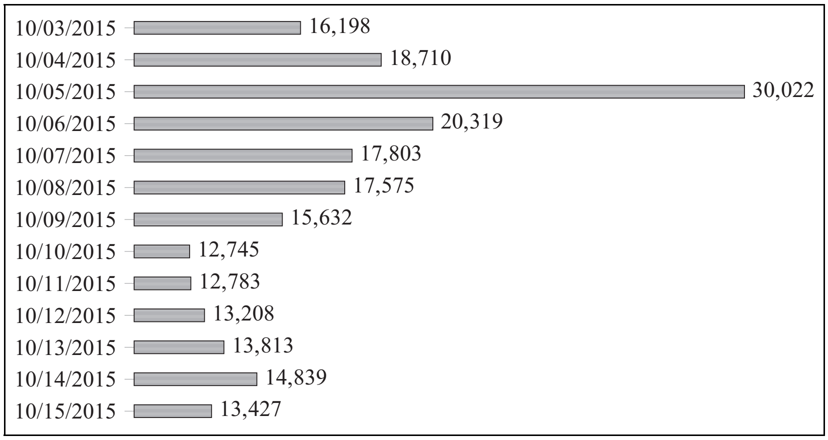

Figure 1. Number of negative tweets per day.

3.3. Temporal topic discovery

Next we turned our attention to detect topics among the pool of the 217,074 negative tweets to showcase the proposed framework. Different approaches for detecting the topics in a corpus have been developed based on the neural network and statistical distributions. During the last decade, the neural network has been considered for improving data analysis methods such as the deep neural network (DNN) [49]. The models based on neural network are costly and time-consuming methods [50,51,52] that require a lot of training data along with having a chance of overfitting problem [53]. For example, one study needed to train millions of tweets for the pre-training step [54] that is also a barrier for having an efficient real-time streaming data [55]. In addition, the deep learning methods have not shown significant performance over other methods in small or medium size corpora [49,55].

According to the literature, Latent Dirichlet Allocation (LDA) [56] is a valid [57,58,87,88] and widely used [3,59] model for discovering topics in a corpus [89]. LDA assumes that there are multiple topics in a corpus and each topic is a distribution over corpus's vocabulary and can be inferred from word-document co-occurrences [56, 90, 91].

LDA assigns each word in a document to each of topics with a different degree of membership using a statistical distribution [92]. For example, in a given corpus, LDA assigns 'flood', 'earthquake' and 'hurricane' into a topic with 'natural disaster' theme. TwiSA uses LDA to identify topics in the negative tweets. This research uses Mallet implementation of LDA [60] that uses Gibbs sampling that is a specific form of Markov Chain Monte Carlo (MCMC) [61] with removing standard stopwords list and having 1000 iterations to detect top 25 topics per day (325 topics in total) in the negative tweets per each of the 13 days.

3.4. Topic content analysis

The last component of TwiSA analyzes extracted topics. This component labels and categorises 25 topics per day, analyses frequency of them and explains possible reasons behind the detected topics comparing with official and published reports.

We manually labelled the negative topics (concerns) for each of the 13 days between 3 October 2015 and 15 October 2015. For example, if 'road', 'damage', 'roadway', 'st' and 'exit' are in a topic, we labelled the topic 'Road Damage'. Table 3 shows some examples of the detected topics. It is worth mentioning that the quantity of topics/day was more than 11 topics but we assigned related topics to one category. For example, if some topics represent injured people and victims, we assigned just 'Victims' label to each of those topics.

The analysed topics were categorised into 11 unique ones with different frequencies (Table 4). This analysis indicates that damages and animal support are the most and the least discussed negative topics, respectively. Moreover, Figure 2 shows that the number of unique negative topics per day has changed from 2 to 8 during the time period considered for this research. The highest number of topic diversity is on the fourth, the fifth and the sixth days, and the lowest number of topic diversity is on the last day. This figure indicates growing rate of topic diversity between the first day and the sixth day, and declining rate of topic diversity between the seventh day and the last day. It seems that SC disaster responders have tried to reduce people's negative concerns between the sixth day and the thirteenth day.

Table 5 shows the topics' distribution for each of the 13 days. We compare our results with official and published reports to explain possible reasons behind the detected topics.

Table 3. Flood topics examples.

Victims	Damage and Costs	Drinking Water	Insurance	Road Damage	Roof Damage
scflood victims flooding disaster death	damage property loss roadway construction	water drink boil bottle clean	families insurance homeowners destroyed support	road damage roadway st exit	roof damage repair danger home
Bridge Damage	**Flood Report**	**Homelessness**	**Power Lost**	**Animal**	
disaster bridge flooded natural flood	alert warning effect flood remain	flood homeless lives woman boy	lost flood home power family	Animal Shelter dog poor find	

Table 4. Total frequency of topics.

Topic	Percentage
Victims	18.75%
Damage and Costs	15.62%
Drinking Water	9.37%
Insurance	9.37%
Homelessness	9.37%
Road Damage	9.37%
Roof Damage	9.37%
Bridge Damage	6.25%
Flood Report	6.25%
Power Lost	4.68%
Animal	1.6%

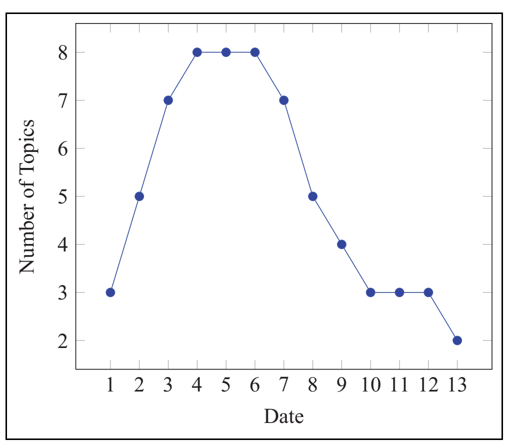

Figure 2. Number of negative topics per day.

'Victims' is the most frequent negative topic that is about the people who were harmed, injured or killed during the SC flood. South Carolina Emergency Management Department (SCEMD) develops and coordinates the South Carolina Emergency Management Program for effective response to disasters. SCEMD reported that 19 people were killed and more than 1500 were rescued from water [9].

'Damage and Costs' is the next high frequent topic that describes financial damage. SCEMD reported that 28,162 people received US$2.2 billion aid through Federal Emergency Management Agency (FEMA) and SCEMD. In addition, more than 3000 collisions [9] and more than 73,000 damaged structures [8] were recorded during the SC flood.

Table 5. Negative topics from 3 October 2015 to 15 October 2015.

	Day 1: 3 Oct 15	Day 2: 4 Oct 15	Day 3: 5 Oct 15	Day 4: 6 Oct 15	Day 5: 7 Oct 15	Day 6: 8 Oct 15	Day 7: 9 Oct 15	Day 8: 10 Oct 15	Day 9: 11 Oct 15	Day 10: 12 Oct 15	Day 11: 13 Oct 15	Day 12: 14 Oct 15	Day 13: 15 Oct 15
Animal	0	1	0	0	0	0	0	0	0	0	0	0	0
Bridge Damage	0	0	1	0	0	1	1	0	0	0	0	1	0
Damage and Costs	1	1	1	1	1	0	1	1	1	0	0	1	1
Drinking Water	0	0	1	1	1	1	1	1	0	0	0	0	0
Flood Report	1	1	0	1	1	0	0	0	0	0	1	0	0
Homelessness	1	0	1	1	1	1	1	0	0	0	0	0	0
Insurance	0	0	0	1	1	1	0	0	1	1	1	0	0
Power Loss	0	1	0	0	0	1	0	1	0	0	0	0	0
Road Damage	0	0	1	1	1	1	1	0	0	1	0	0	0
Roof Damage	0	0	1	1	1	1	1	1	0	0	0	0	0
Victims	0	1	1	1	1	1	1	1	1	1	1	1	1

The next topic is 'Drinking Water'. During the 2015 SC flood, the drinking water system was collapsed in some area such as Columbia (South Carolina's capital city) [62] and thousands of SC residents did not have water for days [63]. In addition, boil water advisory was issued for some days in some parts of South Carolina [64]. With respect to 'Insurance' topic, most SC residents did not have flood insurance in 2015 and FEMA allocated more than US$89 million for individual assistance [65]. In addition, it was suggested to increase insurance penetration and accessibility in South Carolina [8].

The next topic is 'Homelessness'. Thousands of SC residents were displaced and hundreds of people were staying in shelters [63]. This topic covers two groups of people: first people who were forced to leave their home and the second people who did not have a permanent place to live [66]. This topic is an interesting one because SCEMD report did not mention this topic [9].

'Road', 'Roof' and 'Bridge' damages are the next three issues. South Carolina has the fourth largest state-owned highway system with 41,000 miles of road and 8400 bridges in United States [67]. More than 500 roads and bridges were damaged during the flood [9], and 8 out of the 19 deaths occurred on a flooded road [68]. The SC flood caused expensive roof damage. The estimates placed the repair cost close to US$137 million [8,69,70]. Hundreds of bridges were damaged and closed in October 2015 [9].

'Flood Report' topic shows the reports and warnings that were retweeted (rt). Forty-one warnings were issued during the flood, and different contents such as images from the floods, local storm reports and areas of major flooding were shared [68]. The next topic is 'Power Loss'. Thousands of people did not have power for some days during the flood [63]. Finally, 'Animals' was the last and the least discussed topic in the negative tweets. This topic shows main public concerns for animals. During the 2015 SC flood, hundreds of pets were rescued [71] and some of them lost their homes [72]. Supporting animals was one of the topics mentioned in tweets on the second day. This is the second interesting topic that was not addressed in the SCEMD's report. On the other side, we saw that food was the only topic that was in the SCEMD's [9] report but not in our analysis. More than 2 million meals were served and it seems that this topic was not among main concerns during the flood.

The findings show that this research helps SA in some directions. The first one is providing a big picture of damages and public concerns. For example, detecting damage reports by our framework shows show the large size of damages because LDA identifies topics discussed by a large number of users. That big picture can aid the state and federal agencies to have a better-cost estimation and resource allocation. This study can give the agencies a list of priorities for each day. For example, we think that the considering and addressing the insurance concern in the first 5 days were an excellent strategy to calm the people without an immediate cost.

Comparing to other studies, this study has several benefits. First, TwiSA is a fast, real-time and cost-effective framework. The second benefit is combining sentiment analysis and topic modelling methods. While other studies have used the two methods separately without a connection between them in SA analysis, this research provides a flexible framework that has a potential to embed other techniques such as utilising deep learning for sentiment analysis and LDA. The next benefit is focusing on disagreeable or negative experiences, instead of all experiences. The third advantage is introducing new issues such as insurance in a disaster, which were not considered in other studies and confirmed by official reports. The next asset is to have a dynamic perspective exploring the disaster issues during a time frame.

4. Conclusion

Accessing high quality and relevant real-time data during a natural disaster can increase safety and reduce damage and social effects [6]. Every user is a sensor and contributor in social media and can generate valuable and immediate real-time data to develop better SA. However, social media users generate a huge amount of data that need to be summarised to provide a big picture in a disastrous situation.

The opinion survey is a traditional method to analyse public opinion; however, this expensive method should be implemented after natural disasters. In addition, a low number of people participate in the data collection and data analysis steps take a considerable amount of time [73].

The growth of social media has provided a great opportunity to track public opinion. Big real-time social media data can help disaster managers to develop a better SA during a natural disaster. This research proposes an analytical framework to detect and track leading public concerns on social networks such as Twitter to provide a better picture of a disaster. TwiSA combines sentiment analysis and topic modelling to handle a hug number of tweets and to disclose negative public concerns.

This article found that 'victims', 'damages', 'drinking water', 'insurance', 'flood report', 'power loss', 'homelessness' and 'animals' are the topics that were discussed in negative tweets during the flood. We also found two topics, 'homelessness' and 'animals', were not mentioned in the official reports after the flood. It seems that human and non-human damages, drinking water and insurance were the most discussed topics during the SC flood. Even though TwiSA was applied to the 2015 SC flood twitter data, it is not limited to this context and can be applied to other natural disasters and emergency situations.

This research shows that social media are valuable data sources to explore people concerns during a natural disaster. TwiSA can help disaster management teams to find public's concerns in the fastest way to develop a better real-time crisis management plan. The proposed framework can be used not only during natural disasters but also after natural disasters as a post-event methodology review capability (PERC) for future disaster risk reduction. This research has applications for policymakers and disaster risk management experts in developing a better strategy for analysing social media contents and responding to it. It also helps social scientific research in developing Public Service Announcement (PSA) and unveiling public opinions [68].

Our future directions center on minimizing the impact of noises and misinformation, and analyzing positive and neutral tweets. A potential limitation of our approach is that LIWC tool does not usually detect emotions, misspellings, colloquialisms, foreign words, sarcasm and abbreviations [74]. Another limitation is related to the tweets containing wrong or inaccurate information and spams [75]. LDA detects major topics, not every single of topics, in a corpus. Therefore, we assumed that the massive number of tweets in a disaster buries noises. While we believe that the noises do not have a significant impact on the results, we will utilise some approaches to reduce the possible effects of the noises. The first approach is removing retweets and the tweets containing URL to avoid spams. The second approach is detecting unrelated or inaccurate topics and then eliminating the tweets whose primary topic is among the irrelevant or inaccurate topics. The primary topic is the topic that has the highest probability for a tweet. The next approach is using bot detection tools such as Botometer [76] to remove social bots spreading misinformation. We believe that the future plan will address these limitations to provide higher quality data and new insights to SA.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

This research is supported in part by the University of South Carolina Office of the Vice President for Research through the 2015 SC Floods Research Initiative. The authors are grateful for this support and all opinions, findings, conclusions and recommendations in this article are those of the authors and do not necessarily reflect the views of the funding agency.

ORCID iD

Amir Karami https://orcid.org/0000-0003-1936-7497

References

[1] Olanoff D. Twitter monthly active users crawl to 316M, Dorsey: 'We are not satisfied'. 2015, https://techcrunch.com/2015/07/28/twitter-monthly-active-users-crawl-to-316m-up-just-15-year-over-year/ (accessed 28 February 2017).
[2] Mejova Y, Weber I and Macy MW. *Twitter: a digital socioscope*. Cambridge: Cambridge University Press, 2015.
[3] Paul MJ and Dredze M. You are what you tweet: analyzing twitter for public health. *ICWSM* 2011; 20: 265–272.
[4] Sarter NB and Woods DD. Situation awareness: a critical but ill-defined phenomenon. *Int J Aviat Psychol* 1991; 1(1): 45–57.
[5] Castillo C. *Big crisis data*. Cambridge: Cambridge University Press, 2016.
[6] Kitchin R. The real-time city? Big data and smart urbanism. *GeoJ* 2014; 79(1): 1–14.
[7] Madey G, Barabási AL, Chawla N et al. Enhanced situational awareness: application of DDDAS concepts to emergency and disaster management. In: Shi Y, van Albada GD and Dongarra J (eds) *Computational science – ICCS 2007*. Berlin: Springer, 2007, pp. 1090–1097.
[8] ZURICH. What can be learned from the Columbia and Charleston floods 2015? https://www.zurich.com/_/media/dbe/corporate/docs/corporate-responsibility/risk-nexus-south-carolina-floods-2015.pdf?la=en&hash=CACE692203311CC2D883A5B0AA7B4954B9CA4742 (accessed 28 February 2017).
[9] Emergency Management Department (SCEMD). October, 2015 statewide flooding incident. 2016, https://www.scemd.org/news/october-2015-flooding-incident/ (accessed 28 February 2017).
[10] Vieweg S, Hughes AL, Starbird K et al. Microblogging during two natural hazards events: what twitter may contribute to situational awareness. In: *Proceedings of the SIGCHI conference on human factors in computing systems*, Atlanta, GA, 10–15 April 2010, pp. 1079–1088. New York: ACM.
[11] Neppalli VK, Caragea C, Squicciarini A et al. Sentiment analysis during hurricane sandy in emergency response. *Int J Dis Risk Reduct* 2017; 21: 213–222.

[12] Callaghan CW. Disaster management, crowdsourced R&D and probabilistic innovation theory: toward real time disaster response capability. *Int J Dis Risk Reduct* 2016; 17: 238–250.
[13] Bollen J, Mao H and Zeng X. Twitter mood predicts the stock market. *J Comput Sci* 2011; 2(1): 1–8.
[14] Tumasjan A, Sprenger TO, Sandner PG et al. Predicting elections with twitter: what 140 characters reveal about political sentiment. In: *Proceedings of the ICWSM 10* Washington, DC, 17 May 23–26, 2010.
[15] Szomszor M, Kostkova P and De Quincey E. #Swineflu: twitter predicts swine flu outbreak in 2009. In: *Proceedings of the international conference on electronic healthcare*, Casablanca, Morocco, 13–15 December 2010, pp. 18–26. Berlin: Springer.
[16] Abel F, Hauff C, Houben GJ et al. Twitcident: fighting fire with information from social web streams. In: *Proceedings of the 21st international conference on World Wide Web*, Lyon, 16–20 April 2012, pp. 305–308. New York: ACM.
[17] Sinnappan S, Farrell C and Stewart E. Priceless tweets! A study on twitter messages posted during crisis: black saturday. In: *ACIS 2010 proceedings*, 2010, p. 39, https://aisel.aisnet.org/acis2010/39
[18] Avvenuti M, Cresci S, Marchetti A et al. Ears (earthquake alert and report system): a real time decision support system for earthquake crisis management. In: *Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining*, New York, 24–27 August 2014, pp. 1749–1758. New York: ACM.
[19] Sakaki T, Okazaki M and Matsuo Y. Tweet analysis for real-time event detection and earthquake reporting system development. *IEEE T Knowl Data Eng* 2013; 25(4): 919–931.
[20] Earle PS, Bowden DC and Guy M. Twitter earthquake detection: earthquake monitoring in a social world. *Ann Geophys* 2012; 54(6): 708–715.
[21] Sakaki T, Okazaki M and Matsuo Y. Earthquake shakes Twitter users: real-time event detection by social sensors. In: *Proceedings of the 19th international conference on World Wide Web*, Raleigh, NC, 26–30 April 2010, pp. 851–860. New York: ACM.
[22] Robinson B, Power R and Cameron M. A sensitive twitter earthquake detector. In: *Proceedings of the 22nd international conference on World Wide Web*, Rio de Janeiro, Brazil, 13–17 May 2013, pp. 999–1002. New York: ACM.
[23] Miyabe M, Miura A and Aramaki E. Use trend analysis of twitter after the great east japan earthquake. In: *Proceedings of the conference on computer supported cooperative work companion*, Seattle, WA, 11–15 February 2012, pp. 175–178. New York: ACM.
[24] Doan S, Vo BKH and Collier N. An analysis of twitter messages in the 2011 Tohoku earthquake. In: *Proceedings of the international conference on electronic healthcare*, Málaga, 21–23 November 2011, pp. 58–66. Berlin: Springer.
[25] Oh O, Kwon KH and Rao HR. An exploration of social media in extreme events: rumor theory and twitter during the Haiti earthquake 2010. In: *Proceedings of the international conference on information systems*.
[26] Hughes AL, St. Denis LA, Palen L et al. Online public communications by police & fire services during the 2012 Hurricane Sandy. In: *Proceedings of the 32nd annual ACM conference on human factors in computing systems*, Toronto, ON, 26 April–1 May 2014, pp. 1505–1514. New York: ACM.
[27] Gupta A, Lamba H, Kumaraguru P et al. Faking sandy: characterizing and identifying fake images on Twitter during Hurricane Sandy. In: *Proceedings of the 22nd international conference on World Wide Web*, Rio de Janeiro, Brazil, 13–17 May 2013, pp. 729–736. New York: ACM.
[28] Takahashi B, Tandoc EC and Carmichael C. Communicating on Twitter during a disaster: an analysis of tweets during Typhoon Haiyan in the Philippines. *Comput Hum Behav* 2015; 50: 392–398.
[29] Sreenivasan ND, Lee CS and Goh DHL. Tweet me home: exploring information use on twitter in crisis situations. In: *Proceedings of the international conference on online communities and social computing*, Orlando, FL, 9–14 July 2011, pp. 120–129. New York: ACM.
[30] Palen L, Starbird K, Vieweg S et al. Twitter-based information distribution during the 2009 Red River Valley flood threat. *Bull Am Soc Inform Sci Technol* 2010; 36(5): 13–17.
[31] De Albuquerque JP, Herfort B, Brenning A et al. A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management. *Int J Geograph Inform Sci* 2015; 29(4): 667–689.
[32] Ashktorab Z, Brown C, Nandi M et al. Tweedr: mining twitter to inform disaster response. In: *Proceedings of the 11th international conference on information systems for crisis response and management*, University Park, PA, 18–21 May 2014, pp. 77–80. New York: ACM.
[33] Dashti S, Palen L, Heris MP et al. Supporting disaster reconnaissance with social media data: a design-oriented case study of the 2013 Colorado floods. In: *Proceedings of the 11th international ISCRAM conference*, University Park, PA, 1 January 2014, pp. 18–21. New York: ACM.
[34] Herfort B, Schelhorn SJ, Albuquerque JPD et al. Does the spatiotemporal distribution of tweets match the spatiotemporal distribution of flood phenomena? A study about the River Elbe Flood in June 2013. In: *Proceedings of the international conference on information systems for crisis response and management*, University Park, PA, 18–21 May 2014.
[35] Murthy D and Longwell SA. Twitter and disasters: the uses of Twitter during the 2010 Pakistan floods. *Inform Commun Soc* 2013; 16(6): 837–855.
[36] Shaw F, Burgess J, Crawford K et al. Sharing news, making sense, saying thanks. *Aust J Commun* 2013; 40(1): 23–40.

[37] Kongthon A, Haruechaiyasak C, Pailai J et al. The role of Twitter during a natural disaster: case study of 2011 Thai flood. In: *Proceedings of the PICMET'12: technology management for emerging technologies*, Vancouver, BC, 29 July–2 August 2012, pp. 2227–2232. New YORK: IEEE.
[38] Twitter, https://dev.twitter.com/overview/documentation (accessed 28 February 2017).
[39] GNIP, https://gnip.com/ (accessed 28 February 2017).
[40] Taboada M, Brooke J, Tofiloski M et al. Lexicon-based methods for sentiment analysis. *Comput Linguist* 2011; 37(2): 267–307.
[41] Khan AZ, Atique M and Thakare V. Combining lexicon-based and learning-based methods for twitter sentiment analysis. *Int J Electron Commun Soft Comput Sci Eng* 2015; 4(4): 89–91.
[42] Medhat W, Hassan A and Korashy H. Sentiment analysis algorithms and applications: a survey. *Ain Shams Eng J* 2014; 5(4): 1093–1113.
[43] Linguistic Inquiry and Word Count (LIWC), http://liwc.wpengine.com/ (accessed 28 February 2017).
[44] Franklin E. Some theoretical considerations in off-the-shelf text analysis software. In: *Proceedings of the Student Research Workshop associated with RANLP 2015*, Hissar, 7–9 September 2015, pp. 8–15.
[45] Golder SA and Macy MW. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science* 2011; 333(6051): 1878–1881.
[46] Gonçalves P, Araújo M, Benevenuto F et al. Comparing and combining sentiment analysis methods. In: *Proceedings of the first ACM conference on online social networks*, Boston, MA, 7–8 October 2013, pp. 27–38. New York: ACM.
[47] Nolen-Hoeksema S and Morrow J. A prospective study of depression and posttraumatic stress symptoms after a natural disaster: the 1989 Loma Prieta earthquake. *J Pers Soc Psychol* 1991; 61(1): 115–121
[48] Yin D, Bond SD and Zhang H. Anxious or angry? Effects of discrete emotions on the perceived helpfulness of online reviews. *MIS Quart* 2014; 38(2): 539–560.
[49] Zhang D, Luo T and Wang D. Learning from LDA using deep neural networks. In: Lin CY, Xue N, Zhao D et al. (eds) *Natural language understanding and intelligent applications*. Berlin: Springer, 2016, pp. 657–664.
[50] Mohammadi M, Al-Fuqaha A, Sorour S et al. Deep learning for iot big data and streaming analytics: a survey. *IEEE Commun Surv Tut* 2018; 20: 2923–2960.
[51] Zhao W. Research on the deep learning of the small sample data based on transfer learning. *AIP Conf Proc* 2017; 1864: 020018.
[52] Chilimbi TM, Suzue Y, Apacible J et al. Project Adam: building an efficient and scalable deep learning training system. *OSDI* 2014; 14: 571–582.
[53] Erhan D, Bengio Y, Courville A et al. Why does unsupervised pre-training help deep learning? *J Mach Learn Res* 2010; 11: 625–660.
[54] Severyn A and Moschitti A. Twitter sentiment analysis with deep convolutional neural networks. In: *Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval*, Santiago, Chile, 9–13 August 2015, pp. 959–962. New York: ACM.
[55] Mahmud M, Kaiser MS, Hussain A et al. Applications of deep learning and reinforcement learning to biological data. *IEEE T Neur Net Learn Syst* 2018; 29(6): 2063–2079.
[56] Blei DM, Ng AY and Jordan MI. Latent dirichlet allocation. *J Mach Learn Res* 2003; 3: 993–1022.
[57] Blei DM and McAuliffe JD. Supervised topic models. In: *Proceedings of the advances in neural information processing systems*, https://papers.nips.cc/paper/3328-supervised-topic-models.pdf
[58] Hong L and Davison BD. Empirical study of topic modeling in Twitter. In: *Proceedings of the first workshop on social media analytics*, Washington, DC, 25–28 July 2010, pp. 80–88. New York: ACM.
[59] Lu Y, Mei Q and Zhai C. Investigating task performance of probabilistic topic models: an empirical study of PLSA and LDA. *Inform Retr* 2011; 14(2): 178–203.
[60] Mallet, http://mallet.cs.umass.edu/topics.php (accessed 28 February 2017).
[61] Steyvers M and Griffiths T. Probabilistic topic models. *Handb Latent Semant Anal* 2007; 427(7): 424–440.
[62] JournalStar, http://www.pjstar.com/article/20151008/NEWS/151009377 (accessed 28 February 2017).
[63] Senna A and Collins J. South Carolina flood: thousands without running water. 2015, http://www.thestate.com/news/nation-world/national/article37801893.html (accessed 28 February 2017).
[64] Wallace A. Curfew, boil water advisory issued as flooding continues. 2015, http://www.coladaily.com/2015/10/04/officials-issue-warnings-floods-sweep-columbia-area/ (accessed 28 February 2017).
[65] PR Newswire. Massive flooding in South Carolina; many consumers lack flood insurance. 2015, http://www.prnewswire.com/news-releases/massive-flooding-in-south-carolina-many-consumers-lack-flood-insurance-300154495.html (accessed 28 February 2017).
[66] Ohnesorge S. Homeless man loses 'home' in SC flooding. 2015, http://www.wbtv.com/story/30203344/homeless-man-loses-home-in-sc-flooding (accessed 28 February 2017).
[67] 2015 DOT annual report, http://dc.statelibrary.sc.gov/bitstream/handle/10827/21838/DOT_Annual_Report_2015-2016.pdf?sequence=1&isAllowed=y (accessed 28 February 2017).
[68] U.S. Department of Commerce. The historic South Carolina floods of October 1–5, 2015. 2015, http://www.weather.gov/media/publications/assessments/SCFlooding_072216_Signed_Final.pdf (accessed 28 February 2017).

[69] GreenvilleOnline. Storm system brings property damage, flooding, record rains. 2015, http://www.greenvilleonline.com/story/news/local/2015/06/03/storm-system-brings-property-damage-flooding-record-rains/28403225/ (accessed 28 February 2017).
[70] John McDermott. S.C. insurers brace for surge of flood claims. 2015, http://www.postandcourier.com/business/s-c-insurers-brace-for-surge-of-flood-claims/article_d67b6317-2145-5063-ba1b-61850ded7c0e.html (accessed 28 February 2017).
[71] Ferris R. South Carolina's rain and floods, by the numbers. 2015, http://www.cnbc.com/2015/10/08/south-carolinas-rain-and-floods-by-the-numbers.html (accessed 28 February 2017).
[72] Lu F. Dogs left homeless after S.C. flooding arrive in Mass. in search of new owners. 2015, http://www.masslive.com/news/index.ssf/2015/10/dogs_left_homeless_after_sc_fl.html (accessed 28 February 2017).
[73] Urcan I. Flood hazards perception: the result of an opinion survey made in the little towns from lower aries corridor. *J Catastrophe* 2012; 11: 202–210.
[74] Pennebaker JW, Boyd RL, Jordan K et al. *The development and psychometric properties of LIWC2015*. Technical report, 2015, https://repositories.lib.utexas.edu/bitstream/handle/2152/31333/LIWC2015_LanguageManual.pdf
[75] Najafabadi MM. A research agenda for distributed hashtag spoiling: tails of a survived trending hashtag. In: *Proceedings of the 18th annual international conference on digital government research*, Staten Island, NY, 7–9 June 2017, pp. 21–29. New York: ACM.
[76] Ferrara E, Varol O, Davis C et al. The rise of social bots. *Commun ACM* 2016; 59(7): 96–104.
[77] Karami A and Elkouri A. Political popularity analysis in social media. *Proceedings of the iConference*. Washington, DC, 31 March – 3 April 2019.
[78] Karami A, Bennett LS and He X. Mining public opinion about economic issues: twitter and the US Presidential election. *Int J Strat Deci Sci (IJSDS)* 2018; 9(1): 18–28.
[79] Alperin JP, Dumas C, Karami A, et al. Politicians & the public: The analysis of political communication in social media. *Proceedings of the Association for Information Science and Technology*, Vancouver, 10–14 November 2018; 55(1):646–9.
[80] Kitzie VL, Mohammadi E, Karami A. Life never matters in the DEMOCRATS MIND: examining strategies of retweeted social bots during a mass shooting event. *Proceedings of the Association for Information Science and Technology*, Vancouver, 10–14 November 2018; 55(1):254–63.
[81] Najafabadi MM and Domanski RJ. Hacktivism and distributed hashtag spoiling on Twitter: Tales of the# IranTalks. *First Monday* 2018; 23(4).
[82] Shaw G and Karami A. An Exploratory Study of (#) Exercise in the Twittersphere. *Proceedings of the iConference*, Washington, D.C, 31 March–3 April 2019.
[83] Webb F, Karami A and Kitzie V. Characterizing Diseases and disorders in Gay Users' tweets. *Proceedings of the Southern Association for Information Systems* (SAIS), Atlanta, GA, 23–24 March 2018.
[84] Karami A, Dahl AA, Turner-McGrievy G, et al. Characterizing diabetes, diet, exercise, and obesity comments on Twitter. *Int J Inform Manage* 2018; 38(1): 1–6.
[85] Karami A, Webb F and Kitzie VL. Characterizing transgender health issues in twitter. *Proceedings of the Association for Information Science and Technology*, Vancouver, 10–14 November 2018; 55(1): 207–215.
[86] Shaw Jr G and Karami A. Computational content analysis of negative tweets for obesity, diet, diabetes, and exercise. *Proceedings of the Association for Information Science and Technology*, Washington, DC, 27 October–1 November 2017; 54(1): 357–365.
[87] Karami A and Gangopadhyay A. Fftm: a fuzzy feature transformation method for medical documents. *Proceedings of BioNLP*, Baltimore, MD, 26 June-2 July 2014, pp.128–133.
[88] Karami A, Gangopadhyay A, Zhou B, et al. A fuzzy approach model for uncovering hidden latent semantic structure in medical text collections. *Proceedings of the iConference*, Newport Beach, CA, 24–27 March 2015.
[89] Karami A and Pendergraft NM. Computational analysis of insurance complaints: Geico case study. *International Conference on Social Computing, Behavioral-Cultural Modeling, & Prediction and Behavior Representation in Modeling and Simulation*, Washington DC, 10–13 July 2018.
[90] Karami A. ProQuest Dissertations Publishing, University of Maryland, Baltimore County, 2015. 3721828.
[91] Karami A, Aryya G, Bin Z, et al. Fuzzy approach topic discovery in health and medical corpora. *Int J Fuzzy Syst* 2018; 20: 1334–1345.
[92] Karami A, Gangopadhyay A, Zhou B, et al. Flatm: A fuzzy logic approach topic model for medical documents. In: *the 2015 Annual Conference of the North American Fuzzy Information Processing Society (NAFIPS) held jointly with 2015 5th World Conference on Soft Computing (WConSC)*, Redmond, WA, 17–19 August 2015. Piscataway, NJ: IEEE.
[93] Collins M and Karami A. Social Media Analysis For Organizations: Us Northeastern *Public And State Libraries Case Study. Proceedings of the Southern Association for Information Systems (SAIS)*, Atlanta, Georgia. March 23–24, 2018.
[94] Karami A and Collins M. What do the US West Coast public libraries post on Twitter? *Proceedings of the Association for Information Science and Technology. Vancouver*, Canada. 2018; 55(1): 216–225.
[95] Kordzadeh N. Exploring the use of twitter by leading medical centers in the United States. In: *Proceedings of the 52nd Hawaii International Conference on System Sciences*, Grand Wailea, Maui, Hawaii. 7–10 January 2019.
[96] Kordzadeh N and Chen S. Investigating The Use Of Twitter By Three Major Us Medical Centers. *Proceedings of the Southern Association for Information Systems (SAIS)*, Atlanta, Georgia, March 23 –24, 2018.

[97] Karami A and Zhou B. Online review spam detection by new linguistic features. *iConference 2015 Proceedings*, 15 March 2015.
[98] Karami A and Zhou L. Exploiting latent content based features for the detection of static SMS spams. *Proceedings of the American Society for Information Science and Technology*, Seattle, WA, 31 October-5 November 2014; 51(1): 1–4.
[99] Karami A and Zhou L. Improving static SMS spam detection by using new content-based features. *Proceedings of the 20th Americas Conference on Information Systems (AMCIS)*, Savannah, GA, 7–9 August 2014.