International Journal of Information Management 42 (2018) 13–24

Contents lists available at ScienceDirect

International Journal of Information Management

journal homepage: www.elsevier.com/locate/ijinfomgt

Big data analytics for disaster response and recovery through sentiment analysis

J. Rexiline Ragini[a], P.M. Rubesh Anand[b,*], Vidhyacharan Bhaskar[c]

[a] Department of Computer Applications, Hindustan University, Chennai 603103, India
[b] Department of Electronics and Communication Engineering, Hindustan University, Chennai 603103, India
[c] Department of Electrical and Computer Engineering, San Francisco State University, 1600 Holloway Avenue, San Francisco, CA 94132, USA

ARTICLE INFO

Keywords:
Big data
Disaster management
Natural language processing
Sentiment analysis
Text classification
Social media analysis

ABSTRACT

Big data created by social media and mobile networks provide an exceptional opportunity to mine valuable insights from them. This information is harnessed by business entities to measure the level of customer satisfaction but its application in disaster response is still in its inflection point. Social networks are increasingly used for emergency communications and help related requests. During disaster situations, such emergency requests need to be mined from the pool of big data for providing timely help. Though government organizations and emergency responders work together through their respective national disaster response framework, the sentiment of the affected people during and after the disaster determines the success of the disaster response and recovery process. In this paper, we propose a big data driven approach for disaster response through sentiment analysis. The proposed model collects disaster data from social networks and categorize them according to the needs of the affected people. The categorized disaster data are classified through machine learning algorithm for analyzing the sentiment of the people. Various features like, parts of speech and lexicon are analyzed to identify the best classification strategy for disaster data. The results show that lexicon based approach is suitable for analyzing the needs of the people during disaster. The practical implication of the proposed methodology is the real-time categorization and classification of social media big data for disaster response and recovery. This analysis helps the emergency responders and rescue personnel to develop better strategies for effective information management of the rapidly changing disaster environment.

1. Introduction

Big data created from Twitter (Procter, Vis, & Voss, 2013; Gandomi & Haider, 2015) has made a prominent position in almost all the industries. The various applications of big data analytics include, smarter healthcare, multi-channel, finance, log analysis, homeland security, traffic control, telecommunications, manufacturing industries, trading analytics, retail marketing, crime analysis and prediction (Gerber, 2014; Yang, Lee, & Kuo, 2016; Lv, Chen, Zhang, Duan, & Li, 2017). Social media is used by people for sharing reviews and critiques about products and services (Gensler, Völckner, Liu-Thompkins, & Wiertz, 2013; Fang & Zhan, 2015). Social networks generate high volume of data every second and the major challenges are filtration and analysis of those big data for a specific query. While big data analytics has been successfully applied in many sectors, their application in disaster response is still at its early stages (Graham, Avery, & Park, 2015; Roshan, Warren, & Carr, 2016). The social network is rarely used for emergency help related requests during disaster situations. As crisis situations are more chaotic and disorganized, the analysis of the big data generated during such situation is the perfect fit for effective handling of the chaotic environment. In the event of disaster, it is important to make the right decision for helping the affected people with their needs. The disaster management team relies on incomplete or incorrect message[1] at most times due to the lack of direct communication from the affected people. In such situations, big data analytics and computational intelligence can help the rescue team to get the right information from a huge amount of data, analyze it and take the best course of action.

According to Qadir et al. (2016), the three major phases of disaster management are (i) preparedness and early warning, (ii) impact and response (iii) mitigation, risk and vulnerability modeling. In all the phases, the input data are of two types namely, user generated content such as Twitter, Flickr, Facebook and sensor generated data such as satellite images, drones. When these data are analyzed meticulously, the effects of the disaster situation can be handled effectively. Big data

* Corresponding author.
E-mail addresses: rexilineragini@gmail.com (J.R. Ragini), rubesh.anand@gmail.com (P.M.R. Anand), vcharan@gmail.com (V. Bhaskar).

[1] Text, Message and Tweet are used interchangeably throughout the paper which refers to information received from disaster affected area.

https://doi.org/10.1016/j.ijinfomgt.2018.05.004
Received 22 September 2017; Received in revised form 17 May 2018; Accepted 17 May 2018
Available online 23 May 2018

analytics provide solutions to handle these data in an operative way, such that all the three phases of the disaster are managed properly. Though disasters are big, messy and devastating, they bring the people together by creating a philanthropic community where people help one another to fight against the ongoing calamity. It is the intuitive nature of humans to express the opinions and feelings that surround them. It is important to analyze the emotional load of the messages to understand the true meaning of the text. Such analysis was first carried out on the text related to Haiti earthquake (Gurman & Ellenberger, 2015). It was the first incident which brought the people together where big data was effectively used to help the affected people. During the time of Haiti earthquake, the digital humanitarian was first introduced. Digital humanitarian is the process of employing techniques like, crowdsourcing to produce crisis maps (Tapia, Moore, & Johnson, 2013). After the incident of Haiti earthquake, the usage of digital technology for crisis response has become a practice. Though there are various studies that analyze the emotions of the people during disaster (Zielinski, Middleton, Tokarchuk, & Wang, 2013; Torkildson, Starbird, & Aragon, 2014; Mohammad & Kiritchenko, 2015), they are ineffective in analyzing the sentiment towards the needs of people during any crisis.

In this paper, we propose a method to identify the sentiment towards the philanthropic aids received by the people during and after a disaster. Though government and other rescue personnel try to help the people during disaster, people seldom get the full benefits as there are no proper means to understand the exact needs at that point of time. This research classifies the tweets during disaster and helps in building a sentiment model on the various needs of the people. The proposed model helps the rescue personnel to understand the disaster situation and act accordingly. The main contribution of this research is in three fold. First, we analyze and categorize the various needs of the people during and after the disaster. Secondly, various features like, bag of words, parts of speech based features and various lexicon based features are analyzed and the best performing algorithm for each of the category is identified. Lastly, a method to visualize the sentiment on the basic needs is proposed which would help the emergency responders to serve in a better way.

Further, the rest of the paper is organized in seven sections. The review of earlier works in the field of text analysis is discussed in Section 2. The case study and the dataset description is explained in Section 3. The proposed methodology is discussed in Section 4. The experimental results with comparative analysis is presented in Section 5. Discussions are included in the Section 6. Section 7 summarizes and provides the conclusion.

2. Review of literature

Big data from social media can be used in crisis response for various purposes like, communicating with public during disaster response and recovery, detect early warning messages, general community engagement services, communicate with other organizations involved in disaster management, monitoring the messages send by other humanitarian organization and general public.

Graham et al. (2015) had studied the pattern of usage of social media during the crisis situation. The results of their analysis revealed that social media could extensively be used during crisis but the available number of tools using social media to monitor the crisis situation were relatively less. Leong, Pan, Ractham, and Kaewkitipong (2015) had studied the effect of Information and Communication Technology (ICT) on the 2011 Thailand flooding. The study concentrated on analyzing the ways through which social media empowered the community from three dimensions namely, psychological, structural and resource empowerment. It also revealed the role of social media in empowering communications during crisis response. Abbasi and Kumar (2012) analyzed the use of social media during a simulated crisis response and a training platform was created to understand the ways of usage of social media during crisis which helped the first responders. The role of social media during the Tohoku earthquake was investigated by Umihara & Nishikitani (2013). In their work, Twitter users were divided into two groups as users affected by disaster and the users who were not affected. The psychological effect of the users affected by earthquake was analyzed. The psychological effect varied based on the gender of the people. Also, people affected by earthquake tend to tweet more at the time of disaster. Toriumi et al. (2013) identified the information sharing pattern and retweet pattern on Twitter messages during the great eastern earthquake. The results revealed that retweets during disaster were not only for information sharing by general public but they were mainly used as relying information on mass media. Twitter was used as an early warning system for detecting the earthquake shakes. Yates & Paquette (2011) investigated the use of social media information sharing pattern and the ways in which social media was used for decision making at the critical situation during the Haiti earthquake. A tweet-frequency time series with keyword earthquake was constructed which showed the large peaks correlation during the origin of any earthquake (Earle, Bowden, & Guy, 2011). Another earthquake detector system for Australia and New Zeeland was developed which sent notification to the joint Australian Tsunami warning system based on the tweets. The proposed algorithm was able to detect 28 real events which were minor out of 31 alerts (Robinson, Power, & Cameron, 2013).

Sentiment analysis mainly deals with classifying the texts into positive and negative. Pang, Lee, and Vaithyanathan (2002) was the first to work on sentiment analysis by classifying the movie review data into positive and negative using machine learning approaches. The study was concluded by analyzing the challenges in sentiment analysis. Kim, Howland, and Park (2005) had studied the dimensionality reduction using Support Vector Machine (SVM) algorithm. The authors highlighted SVM as the best algorithm for any text classification task. Fang and Zhan (2015) carried out the sentiment analysis on product reviews through the data collected from Amazon. The experiment gave promising results for both sentence level and review level sentiment classification. Kapukaranov & Nakov (2015) had worked in a fine grained sentiment analysis for Bulgarian movie reviews. The authors added few contextual information features in the form of meta-data. The results revealed that adding the contextual features improved the classification accuracy. Jeong, Yoon, and Lee (2017) analyzed the usage of sentiment analysis in business for product opportunity exploration. Social media mining approach was utilized in topic modeling and sentiment analysis for identifying the changing customer needs. Sentiment analysis was also carried on various domains like pizza industry (He, Zha, & Li, 2013) and hotel reviews (Hu & Chen, 2016) for analyzing the customer satisfaction. Wu, Zheng, and Olson (2014) had used the Twitter messages to predict the Chinese stock market. The authors analyzed both lexicon based approach and machine learning approach for the data from Sina Finance web portal. The authors concluded that machine learning approach had higher classification accuracy than semantic approach. Aramaki, Maskawa, and Morita (2011) had categorized the tweets related to influenza disease outbreak using SVM. The study revealed that using Natural Language Processing (NLP) techniques significantly improved the classification accuracy. Paul, Dredze, and Broniatowski (2014) had also used tweets to forecast the outbreak of influenza. Their study revealed that tweets could forecast the disease outbreak with 30% improved accuracy than historical methods. Gitto and Mancuso (2017) had investigated the use of data collected from web in improving the airport services.

The main task of sentiment analysis relies in detecting the hidden subjective expression in the text. In order to detect the subjective content, various features are analyzed by different researchers. Subrahmanian and Reforgiato (2008) had used AVA (Adjective verb Adverb) framework for classifying the subjective sentence. Their work revealed that in any type of document, the adjectives and adverbs play an important role in calculating the sentiment. Cho, Kim, Lee, and Lee (2014) analyzed the use of various lexicons as features for sentiment analysis. The analysis used the merge, switch and remove methods to construct a new lexicon based on the domain of study and the method was tested with product reviews of smart phones, movies and books. Hogenboom, Heerschop, Frasincar, Kaymak, and De Jong (2014) had performed a lexicon based multi-lingual sentiment analysis to map the sentiment from English to Dutch language and the results revealed that sentiment related to the meaning of the word also tend to have language specific dimensions as well. According to Mudinas, Zhang, and Levene (2012), combining lexicon and learning based approaches improved the classification accuracy. The results were tested on two real world data set namely, the CNet software reviews and IMDB movie reviews which confirmed that combining these two yield better performance. According to Bravo-Marquez, Mendoza, and Poblete (2014), sentiment consist of two folds namely, polarity classification and subjective classification. The authors analyzed parts of speech tagging and lexicon based features. Parts of speech tagging performed well for subjectivity classification and lexicon based approach had good performance for polarity classification.

In recent days, researchers have focused on applying the sentiment analysis techniques in crisis domain. Tweets were used as an early warning system for detecting the earthquake. Sakaki, Okazaki, and Matsuo (2013) used features such as, keywords and number of words to detect the target event. Kalman filtering and particle filtering were used along with the identified feature to estimate the center of trajectory. Ragini and Anand (2016) had used machine learning algorithms to classify the crisis related tweets. Albuquerque, Herfort, Brenning, and Zipf (2015) studied the usage of social media along with the authoritative data for identifying the required information in managing the disaster situation. The messages that originated around 10 km distance from the affected area were relevant to the incident. Various emotions like calm, unpleasantness, sadness, anxiety, fear and relief were studied by Vo and Collier (2013) during the Japan earthquake. A sentiment analysis model was built by Sen, Rudra, and Ghosh (2015) which automatically detected tweets related to crisis and classified them into various categories like, personal or impersonal style, subjectivity, formal and informal linguistic text. Caragea, Squicciarini, Stehle, Neppalli, and Tapia (2014) had performed a sentiment classification of the tweets during hurricane sandy. The various sentiment of people like, panic and user's concerns were visualized in a map. The results revealed that the sentiments of the people changed according to users' locations and also depending on the distance from the disaster. Ragini, Anand, and Bhaskar (2017) had proposed a hybrid method to classify and segregate the crisis related tweets from the people who were trapped and struggling for survival during disaster situations.

Though there are various studies that measure the emotion and sentiment of the people (anger, fear, disgust, panic) during disaster, they are ineffective in identifying the sentiment of the people towards the philanthropic aid they receive. The summarized problem definition is that an automated text classification and analysis system in real-time is highly necessary for identifying the needs of the people during the times of disaster. The categorization of disaster data for the different needs of the people and the sentiment on each category of the need has to be analyzed. The main objective of this research is to develop a conceptual model or a framework for disaster response and recovery by identifying the best feature that classify the disaster data with highest accuracy. Effective categorization of the data will help the responders in building trust, courage, and confidence among the people during the events of the disaster.

3. Case study and data set description

3.1. Floods in South Asian countries

The disasters that are considered in this research include, India-Pakistan floods in September 2014, a severe cyclonic storm named HUDHUD in October 2014 and another severe cyclonic storm named Nilofar. In September 2014, heavy floods and landslides happened due to torrential rain in the border of India-Pakistan causing severe damages. During the floods in Kashmir, 2600 villages were affected out of which 390 villages were completely submerged in water. In a tweet, "*People of Pakistan have unfortunately suffered from destructive floods now for the fifth consecutive year.*" indicates that the flood occurs almost every year in these areas. During the Cyclonic storm named HUDHUD, the city of Visakhapatnam which is located in South India experienced a huge loss of life and severe damages. Over 2 million families were affected by Hudhud and the avalanche that happened as an effect of the cyclone. The third incident is Nilofar cyclonic storm. The cyclone was formed in the North Indian Ocean and it was about to hit Gujarat state in India. Later, the cyclone got weakened and did not cause much effect. But there was great anxiety among people, which created a lot of hits in social media.

3.2. Data set description

The disaster related data are collected from Twitter for the aforementioned disasters. The Twitter data collected for the text analysis contained 70,817 tweets. The corpus for India-Pakistan floods, Kashmir floods contained 30,817 tweets, HUDHUD contained 30,000 and Nilofar contained 10,000. A part of these tweets are collected using the Streaming Twitter API. As Twitter allows to collect only the past seven days' data using the Streaming Twitter API, there is no provision to collect the historical data using it. The rest of the data are collected using a third party vendor 'Followthehashtag' (Twitter Historical data recovery tool, 2017). The keywords that are used to collect these data are HUDHUD, Vizag flood, Nilofar, Kashmir floods, India-Pakistan floods, Pakistan floods. The data is collected in the specified date range for each of the disaster HUDHUD from 05/09/14 to 15/11/14, Nilofar from 05/09/14 to 15/11/14 and Kashmir floods from 01/09/14 to 15/10/14.

4. Conceptual model for sentiment classification

The proposed sentiment model for disaster response and recovery consists of five phases namely, data collection phase, data storage phase, data preprocessing phase, learning & classification phase and presentation phase.

The main contribution of this research is the learning and classification phase. The classification phase is domain dependent which need in-depth analysis of the data. The presentation phase helps the disaster responders with the visual sentiment analytics model. The learning & classification phase includes data preprocessing, text categorization, subjective sentence categorization, feature vector generation and machine learning algorithm as shown in Fig. 1.

4.1. Text filtering stage

4.1.1. Data preprocessing

In the text filtering stage, the disaster data is preprocessed so that the machine learning algorithm in the next stage can understand the data. In case of Twitter messages, the input data contains Uniform Resource Locators (URLs), numbers, foreign language words, abbreviations, symbols and emoticons. Data preprocessing is preformed to remove all these unnecessary characters from the input data. Once data preprocessing is performed, the input data is categorized according to the various needs of the people.

4.1.2. Categorization of the data

The disaster related tweets are categorized using keyword filtering technique (Vieweg, Hughes, Starbird, & Palen, 2010) which is a common practice in Twitter analysis. The keywords are coined for each category of the identified needs. The keywords are selected by identifying the words that are found more than five times and also relevant to

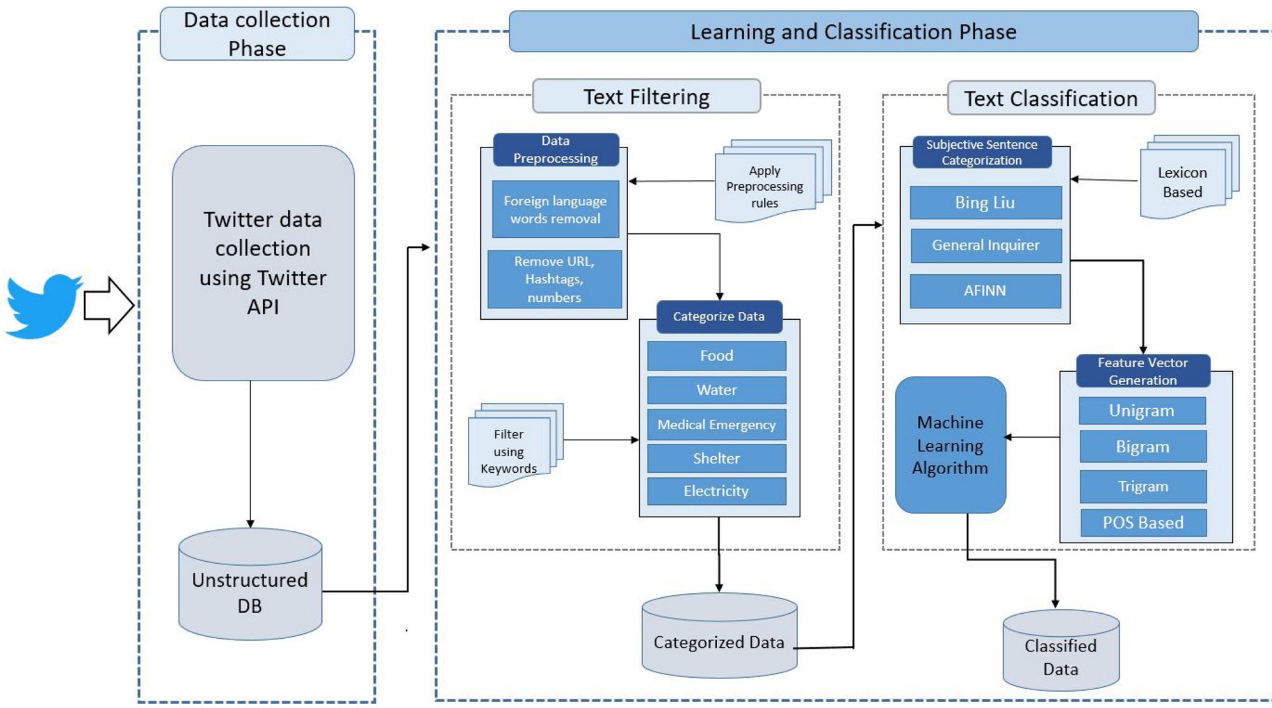


Fig. 1. Proposed model for sentiment categorization and classification.

Table 1
Summary of the Category List and Its Associated Keywords.

Category	Keywords
Water	Water, drink, drinking, thirsty, thirst, dehydration
Food	Food, starve, hungry, milk, bread, formula, eat, foodstuff
Shelter	Shelter, house, living place, sleep, rest, accommodation
Medical emergency	Medicine, clinic, hospital, medicine, doctor, nurse, syrup, first aid, tonic.
Electricity	Electricity, power, electricity, light, fan, energy, current, charge

each of the category (Albuquerque et al., 2015). The identified keywords are utilized to filter the required data from a large set of text. Table 1 shows the coined keywords pertaining to each category of need.

The considered keywords are applied to the collected dataset and the resulting data consisted of 6842 tweets. In the list of categories, collapsed structure and people trapped category are not considered since a different type of analysis (cost estimation analysis for collapsed structure and emergency rescue plan for people trapped) needs to be carried out to help the affected people. Rest of the categories include food, water, shelter, medical emergency, and electricity are the most demanding needs of the people affected by the disaster. The distribution of the Tweets in various categories is shown in Fig. 2.

4.2. Text classification stage

The data categorized according to the needs through text filtering is given as input to the text classification stage. The text classification consists of two steps namely, subjective sentences segregation and feature vector generation. The subjective sentences are the one that has sentiment related information. Segregation of such sentences is important to analyze the needs of the people. Features like, unigram, bigram and trigram are used to convert the subjective sentences into feature vectors. These feature vectors are given as input to the machine learning algorithm.

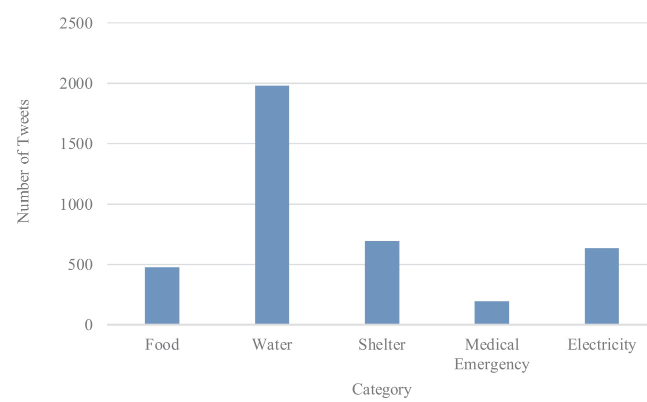


Fig. 2. Number of disaster related tweets in various categories.

4.2.1. *Subjective sentence categorization*

The major task involved in sentiment analysis is segregating the text into subjective and objective sentence by considering the usage of words. The subjective sentences have the information about the sentiment of the people. The objective sentences do not contribute in identifying the sentiment of the people affected by the disaster. Table 2 shows the examples of subjective and objective sentences in disaster data.

The manual segregation of subjective sentence and objective

Table 2
Categorization of Subjective and Objective Sentences.

TWEET	CATEGORY
hudhud power back in some more places including near siripuram ncbn salute to our cm and pm Hudhud	Subjective
Four days on trot I waited fr GVMC to organise water to my aptmt L wise waiting No help so far Hudhud AP CMO ncbn naralokesh	Subjective
Cyclone Hudhud's fall out brings rains in Rajasthan – Zee News http://t.co/ZVcaqn0nxZ	Objective
The storm should be called BadBad or and not HudHud. That would have given us an early indication of its severity.	Objective

Table 3
Comparative Analysis of Various Lexicons for Text Analysis.

Lexicon Name	Description	Domain	Mode of creation	Language	Reference
Bing Liu	Contains a list of positive and negative words	General	Manual	English	Liu (2012)
Sentiment 140	Method label (Positive, Negative and Neutral)	Tweets	Automatic: Created from tweets with emoticons.	English	Zhu, Kiritchenko, and Mohammad (2014)
General Inquirer	Contains about 26 categories pertaining to the sentiment	General	Manual	English	Stone et al. (1962)
NRC Word emotion lexicon	Contains list of positive and negative words	General	Manual	English	Mohammad and Turney (2013)
SentiFul	Contains list of positive and Negative words	General	Automatic	English	Neviarouskaya, Prendinger, and Ishizuka (2009)
SentiStrength	Method label (Positive, Negative and Neutral)	General	Automatic	English	Khan, Qamar, and Bashir (2016)
AFINN	Contains list of positive and negative words with score	Twitter	Manual	English	Deng and Wiebe (2015)
ANEW	Contains list of positive and negative words with score	General	Manual	English	Nielsen (2011)
Sentiwordnet	Sum of the positive and negative scores	Wordnet	Automatic	English	Bradley and Lang (1999)
NRC−Hashtag	Sum of the positive and negative scores	Twitter	Automatic	English	Baccianella, Esuli, and Sebastiani (2010)
SenticNet	Contains scores for pleasantness, attention, sensitivity, aptitude	General	Automatic	English	Cambria, Olsher, and Rajagopal (2014)
NRC Word-Colour Association Lexicon	Contains list of colors	General	Manual	English	Mohammad (2011)

sentence is a tiring task. In order to automate the process, publicly available sentiment based lexicons are utilized to categorize the subjective sentences. These lexicons are the dictionary of words that contribute to identify the polarity of the text. Table 3 summarizes the categorization of the various lexicons that are available for sentiment analysis.

In order to categorize the disaster data, the proposed model considers three manually created lexicons, one from Twitter domain (AFINN) and the other two lexicons which contains slang words, misspelled words, morphological variants (Bing Liu, General Inquirer). Manual lexicons are considered in this analysis as automatic lexicons are with noise which also reduces the classification accuracy (Bravo-Marquez et al., 2014). Moreover, the considered lexicons contain the complete list of slang and misspelled words which is appropriate for Twitter text classification.

4.2.2. *General Inquirer lexicon*

The General Inquirer (GI) lexicon provides content analysis for English language using the “Harvard” and “Lasswell” dictionaries (Stone, Bales, Namenwirth, & Ogilvie, 1962). This lexicon has 26 categories of emotion and sentiment to fit the various types of content analysis depending upon the need. In the proposed method, the first category of the GI lexicon “Two large valence category” is utilized which has about 1915 positive words and 2291 negative words.

4.2.3. *Bing Liu opinion lexicon*

Bing Liu opinion lexicon is maintained and freely distributed by Liu (2012). This lexicon has all the combinations of words which includes, misspelled, slang words and morphological variants of any word. This is a polarity based lexicon and has about 2006 positive words and 4683 negative words.

4.2.4. *AFINN lexicon*

Affective Norms for English words (ANEW) is a lexicon that includes emotional ratings of the words. The rating is based on the person’s psychological reaction to any word. ANEW was released prior to the rise of microblogging sites. After the rise of microblogging sites, there was a need that ANEW should be extended to include all the slang words and other text types related to microblogging. AFINN (Deng & Wiebe, 2015) is the extension of ANEW lexicon which focuses on the words used in social media. AFINN lexicon includes, slang words, web jargons, obscene and acronym words which focuses on calculating the strength of the text. The AFINN lexicon has about 2477 words. These words are categorized through scores as positive and negative, positive word score ranges from 1 to 5 and the negative score ranges from -1 to -5.

The categorized data needs to be classified into positive and negative to train the machine learning system. In order to classify the text into positive and negative, these three lexicons namely, Bing Liu, General Inquirer and AFINN lexicons are used. The disaster data is filtered with the list of positive and negative word list from these lexicons. The text that is not filtered as either positive or negative is classified as neutral and is discarded. Table 4 shows some examples of the classified positive and negative tweets from the disaster data. The percentage of positive, negative and neutral tweets filtered from each categorized data with these lexicons are shown in Fig. 3.

4.3. *Data analysis and preparation*

The manually created lexicons from Twitter domain (AFINN) and the lexicons which contains slang words (Bing Liu, General Inquirer) are selected in order to have a better classification of the disaster data. In the tweets that are filtered with these lexicons, many words that are not utilized for sentiment classification in each category of disaster data is found in the positive and the negative word list. The positive and negative words obtained from filtering each category of disaster data

Table 4
Tweet Categorization.

TWEET	POLARITY
Can some body reading this plz plz plz please send some water and milk to Vizag kids r suffering a lot hudhud plz try to respond	Negative
No power No charging No networks No drinking water No proper food No proper shelter due to Hudhud cyclone I cannot explain now pls pray	Negative
kids happy food distribution cyclone hudhud vizag visakhapatnam relief hurricane	Positive
Cyclone Hudhud AP government airlifts vegetables from Delhi water supply restored in Vizag	Positive

with the considered lexica are shown as a word cloud in Fig. 4.

A word cloud is an image which contains the list of words used in a particular context, with the size of the word indicating the frequency of the word in the particular corpus. In the word cloud created for each of the category, it is evident that words available in the tweets categorized with the lexicon do not contribute much to indentify the polarity of the disaster related tweets. Since the filtered positive and negative tweets do not contribute much in identifying the polarity, subjective phrases are formed to understand the pattern in which each opinion word occurs in the tweet. In the process of subjective phrase identification, it is found that the number of positive subjective phrases are not equal to the number of positive sentences. This is due to the reason that a tweet can have more than one positive word or negative word. Table 5 shows the number of positive and negative subjective phrases in each category of the diaster data.

4.4. Feature vector generation

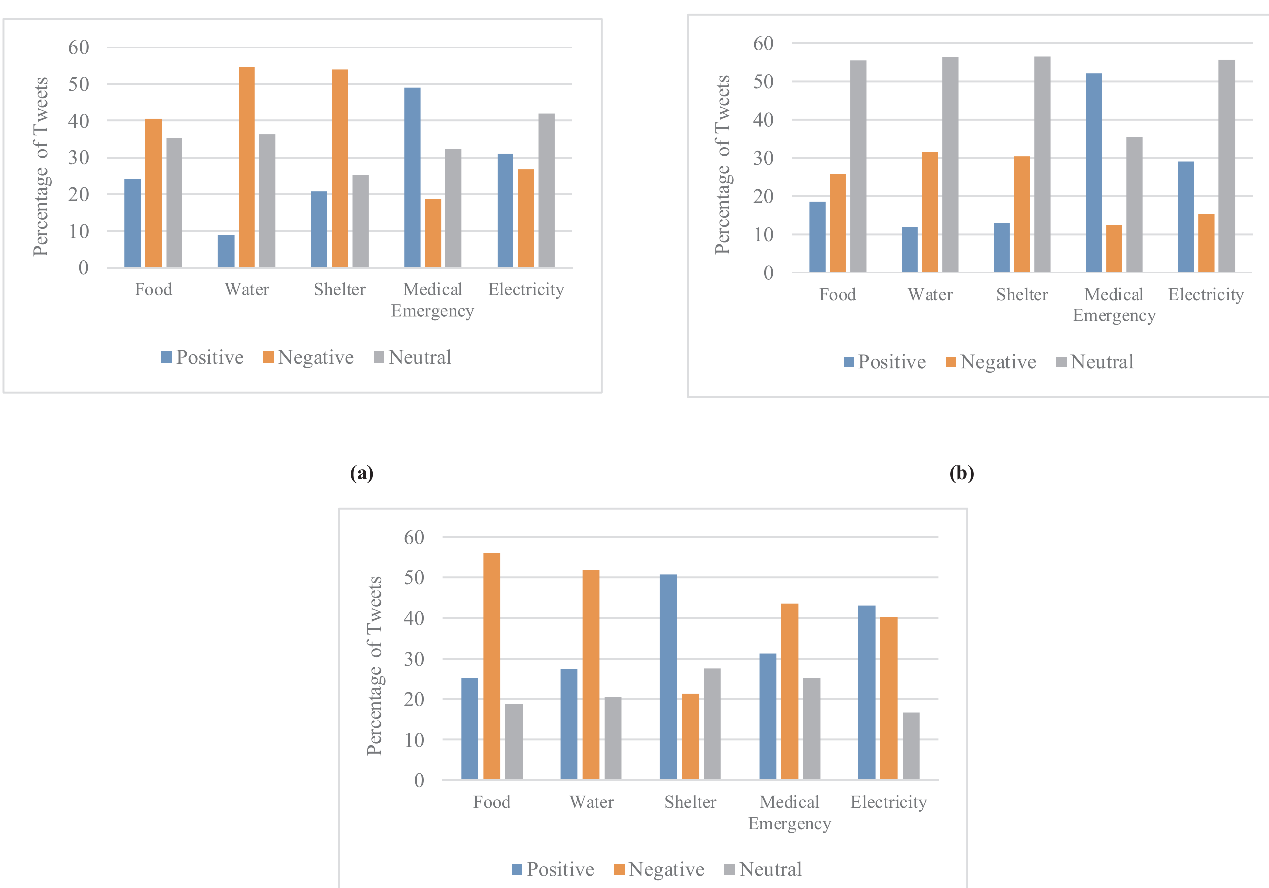
Features are also known as the sentiment tokens and they are vital in the sentiment categorization. The disaster data needs to be converted into feature vectors in order to train a classifier. The most commonly used features for sentiment classification through machine learning algorithm are bag of words, adjectives, adverbs, bigram, and trigram.

4.4.1. POS tagging based features

Adjectives, verbs and adverbs in any sentence contribute more in identifying the sentiment of the text. In order to filter the adjectives and adverbs from entire stream of tweets, it is necessary that Parts of Speech (POS) tagging needs to be carried out. Parts of speech tagging is the process of marking up the appropriate part of speech for each word in a sentence. The POS tagging is used to identify the syntactic role of a word in a sentence. There are eight parts of speech in English language out of which, adjectives and adverbs contribute mainly to identify the sentiment in a sentence. The Pen Tree Bank tagger (Marcus, Santorini, & Marcinkiewicz, 1993) has 46 tags. The syntactic roles of the sentences can be studied in detail with these 46 tags rather than just the eight tags in English. The Pen tree bank tag set from the Natural Language Tool Kit (NLTK) is used to tag each sentence. Table 6 shows the various forms of adverbs and adjectives used in the proposed framework.

4.4.2. Bag of words

In the bag of words (BOW) feature, the entire text is represented as a list of words. The occurrence of each word is used as a feature in training a classifier. Extensive preprocessing needs to be performed such that the bag of words feature gives good classification accuracy.



(c)

Fig. 3. Percentage of positive and negative tweets categorized by (a) Bing Liu, (b) General Inquirer and (c) AFINN.

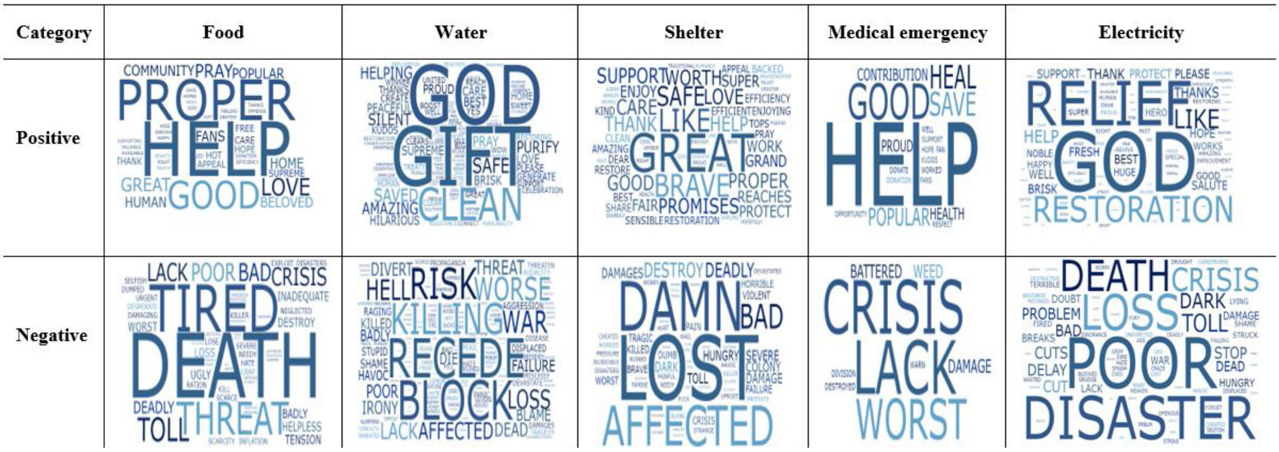


Fig. 4. Word cloud for various categories of disaster data.

Table 5
Number of Subjective Phrase for Different Lexicons.

Category	Lexicon					
	General Inquirer		Bing Liu		AFINN	
	Positive	Negative	Positive	Negative	Positive	Negative
Food	78	129	57	59	70	156
Water	233	232	269	650	546	1031
Medical Emergency	51	21	52	18	30	42
Shelter	80	21	45	98	149	63
Electricity	224	206	195	106	273	255

Table 6
Parts of Speech Tagging for Adjectives and Adverbs (Marcus et al., 1993).

Tag set	Definition
JJ	Adjective
JJR	Adjective, comparative
JJS	Adjective, superlative
RB	Adverb
RBR	Adverb, comparative
RBS	Adverb, superlative

The preprocessing step removes the unnecessary words from corpus so that the classifier omits all the unnecessary features during the learning phase. Combination of the words in a text is considered as a feature. The adjacent two words as the bigram feature, adjacent three words as the trigram feature and it is extended up to *n* numbers namely n-gram feature are used to train the classifier. As the order of the n-gram increases, the classification accuracy will decrease (Wu et al., 2014). Consequently, the input is limited to unigram (Bag of word) model while training the classifier.

4.4.3. *Lexicon based features*

Opinion words or phrases are the important part of any sentence which conveys the sentiment of the entire text. The same opinion word conveys different meaning in different occasions. When segregating the text with just the opinion words, the polarity of the text may result in incorrect identification. In order to understand the pattern in which the word occurs, subjective phrases are framed with a window size of three (the word, its predecessor and its successor). Linguistic features like, bigram and trigram are applied to the subjective phrases to learn the pattern of the text. Negation is an important aspect that needs to be addressed during the polarity classification. In order to remove negation from the text, the entire set of bigram and trigram is scanned. When a negation word like 'no' or 'not' is present in the positive polarity feature set, it is moved to the negative set of features.

4.5. *Machine learning method*

A machine learning algorithm has to be devised in order to map the extracted features to the object of interest. Though there are various machine learning algorithms, Support Vector Machine (SVM) is chosen for text classification as previous studies reveal that SVM performs better than other algorithms (Kim et al., 2005; Aramaki et al., 2011). A machine learning algorithm is required to map the extracted features to the text that needs to be classified. SVM is an artificial intelligence algorithm which works based on the decision boundaries. A decision plane separates the input into various classes based on the class membership. Though there are various supervised machine learning algorithms available, SVM is utilized as it is proven to perform better than any other algorithm in text classification problems. SVM is a supervised classification technique that performs regression and classification tasks by constructing nonlinear decision boundaries. For a given category $C = \{S_+, S_-\}$, where S_+ is the set of positive samples and S_- is the set of negative samples with S_+ is defined as $S_+ = \sum_{n}^{i=1} (d_i, +1)$and S_- is defined as $S_- = \sum_{n}^{i=1} (d_i, -1)$. The SVM calculates and constructs a hyper plane or set of hyper planes in the dimensional space that divides the data into sets with maximum margin. During preprocessing, the entire text is converted into vector X_i consisting of a set of features that represents the corresponding disaster related data. The SVM algorithm calculates and plots a hyperplane through supervised learning that divides the positive and negative texts with a maximum margin. The classification problem is defined as in which side of the hyperplane the test data lies. A kernel is the core of the learning algorithm which works based on the similarity function. There are three different types of kernels in SVM namely, linear, Radial Bias Function (RBF) and polynomial kernels. In the test trials of the disaster data with all the kernels, linear kernel is found to be best fit with the results outperforming the other kernels.

5. Experimental results and analysis

The prototype of the proposed big data approach is experimented using Twitter data collected for disaster events. The proposed method is implemented and tested using Apache Spark Big Data framework and Python programming language. The SVM based classification has two levels namely, training and classification level. In SVM, the training level builds a training model which is used to predict the input text as positive or negative in the classification level.

We have evaluated the learning and classification phase of the proposed model through two methods. The first method segerates the subjective sentence by filtering each category of the text related to disaster data with the considered Bing Liu, General Inquirer and AFINN lexicons. Parts of speech tagging is carried out for all the subjective sentence and the objective (Neutral) sentences are discarded. The subjective sentences are converted to feature vector by applying the adjectives, adverbs and bag of words features. The second method segerates the subjective phrases from the entire stream of disaster related texts. The subjective phrases are identified by scanning the entire text with Bing Liu, General Inquirer and AFINN lexicons. Once the positive or negative words are identified, the word is taken with a window size of three. These subjective phrases are converted into feature vector by applying bigram and trigram features. SVM algorithm is trained with the resulting feature vectors. The results of the machine learning algorithm is evaluated by calculating the precision, recall and F-measure. These are the three main metrics for measuring the performance of a classification system. In order to compute these measures, the class under investigation is considered as positive and all the other remaining classes as negative. Precision is a fraction of the classified text that are relevant. Recall is a fraction of the classified text that are retrieved. F-measure is the ratio of the combination of precision and recall. Tables 7 and 8 show the results of disaster text classification using different feature vectors by first method and second method respectively.

Although adjective and adverb in a sentence contribute more in analyzing the sentiment, this combination does not yield proper results in the crisis domain. The bag of words feature performs well in most of the text classification problems but, they are ineffective in the disaster domain data. As the features like, adjective, adverb, bag of words are not performing well, the second method is evaluated. The linguistic features like, bigram ($n = 2$) and trigram ($n = 3$) are extracted from these subjective phrases to train the machine learning algorithm for the efficient classification of disaster data.

In the disaster text classification using subjective phrases, the phrases formed with bigram performs better than phrases formed with trigram. The reason for the better performance of bigram is that the perfect fit is obtained when combination of two are considered in a sentences rather than more number of words. It is evident from the results that combining the subjective phrase and the machine learning algorithm yields better classification accuracy as in Table 8.

5.1. ROC analysis

Receiver Operating Characteristics (ROC) curve is plotted to further evaluate the performance of the disaster text classification system. ROC curve is an evaluation technique which is used to assess the ability of the algorithm to classify the data. The ROC curve is a graphical interface to measure the accuracy of the text classification through area under the curve. The graph is plotted by using the True Positive Rate (TPR) against the False Positive Rate (FPR) and varying the threshold settings. In the comparison of first method and second method from Tables 7 and 8, it is found that the second method performs better in this classification. In order to investigate further, ROC curve analysis is performed on the second method of text classification using bigram and trigram with Bing Liu, General Inquirer and AFINN lexicon as shown in Fig. 5. The ROC curve analysis using different lexicons for different category of disaster data shows that bigram performs better than trigram.

Table 8
Disaster text classification using subjective phrases.

Lexicon	Category	Bigram			Trigram		
		P	R	F_1	P	R	F_1
Bing-Liu	Food	0.92	0.91	0.90	0.89	0.87	0.85
	Water	0.93	0.92	0.92	0.89	0.88	0.86
	Shelter	0.87	0.85	0.82	0.86	0.83	0.79
	Medical Emergency	0.93	0.92	0.92	0.90	0.89	0.88
	Electricity	0.90	0.88	0.88	0.87	0.83	0.81
General Inquirer	Food	0.89	0.88	0.87	0.85	0.82	0.79
	Water	0.94	0.94	0.94	0.89	0.87	0.85
	Shelter	0.89	0.86	0.85	0.89	0.86	0.85
	Medical Emergency	0.91	0.89	0.89	0.84	0.79	0.77
	Electricity	0.89	0.86	0.86	0.82	0.73	0.71
AFINN	Food	0.95	0.94	0.94	0.91	0.90	0.89
	Water	0.93	0.93	0.92	0.91	0.89	0.89
	Shelter	0.84	0.79	0.72	0.59	0.77	0.67
	Medical Emergency	0.91	0.91	0.90	0.89	0.87	0.86
	Electricity	0.90	0.88	0.88	0.87	0.83	0.82

5.2. Sentiment strength estimation

Once the disaster data is classified by the SVM algorithm, the list of positive and negative words are used to identify the opinion orientation towards each category of data. The sentiment strength is calculated by assigning positive or negative values to the sentiment bearing words in each sentence. Each sentence can have more than one opinion bearing word. Semantic orientation score of *+1* is assigned to each positive word and a score of *–1* is assigned to every negative word. The overall positive score (P_C) for each category of disaster data is calculated as,

Table 7
Disaster Text Classification using BOW and POS Features.

Lexicon	Category	BOW			Adjectives			Adjectives + Adverbs		
		P	R	F_1	P	R	F_1	P	R	F_1
Bing-Liu	Food	0.84	0.78	0.76	0.60	0.58	0.58	0.62	0.59	0.59
	Water	0.96	0.96	0.95	0.78	0.60	0.66	0.78	0.60	0.66
	Shelter	0.83	0.77	0.69	0.62	0.54	0.57	0.62	0.54	0.57
	Medical Emergency	0.30	0.45	0.36	0.66	0.59	0.61	0.64	0.58	0.62
	Electricity	0.70	0.66	0.63	0.61	0.61	0.60	0.62	0.63	0.61
General Inquirer	Food	0.85	0.80	0.79	0.63	0.62	0.62	0.63	0.62	0.62
	Water	0.94	0.94	0.93	0.49	0.33	0.35	0.49	0.33	0.35
	Shelter	0.83	0.77	0.73	0.70	0.68	0.69	0.70	0.67	0.68
	Medical Emergency	0.97	0.96	0.97	0.93	0.92	0.93	0.93	0.92	0.93
	Electricity	0.79	0.77	0.69	0.62	0.54	0.57	0.62	0.54	0.57
AFINN	Food	0.58	0.69	0.63	0.63	0.44	0.50	0.60	0.55	0.57
	Water	0.77	0.78	0.77	0.56	0.50	0.51	0.55	0.53	0.51
	Shelter	0.70	0.72	0.71	0.62	0.57	0.58	0.65	0.63	0.61
	Medical Emergency	0.40	0.50	0.44	0.49	0.43	0.46	0.49	0.43	0.46
	Electricity	0.60	0.59	0.59	0.50	0.49	0.43	0.50	0.47	0.42

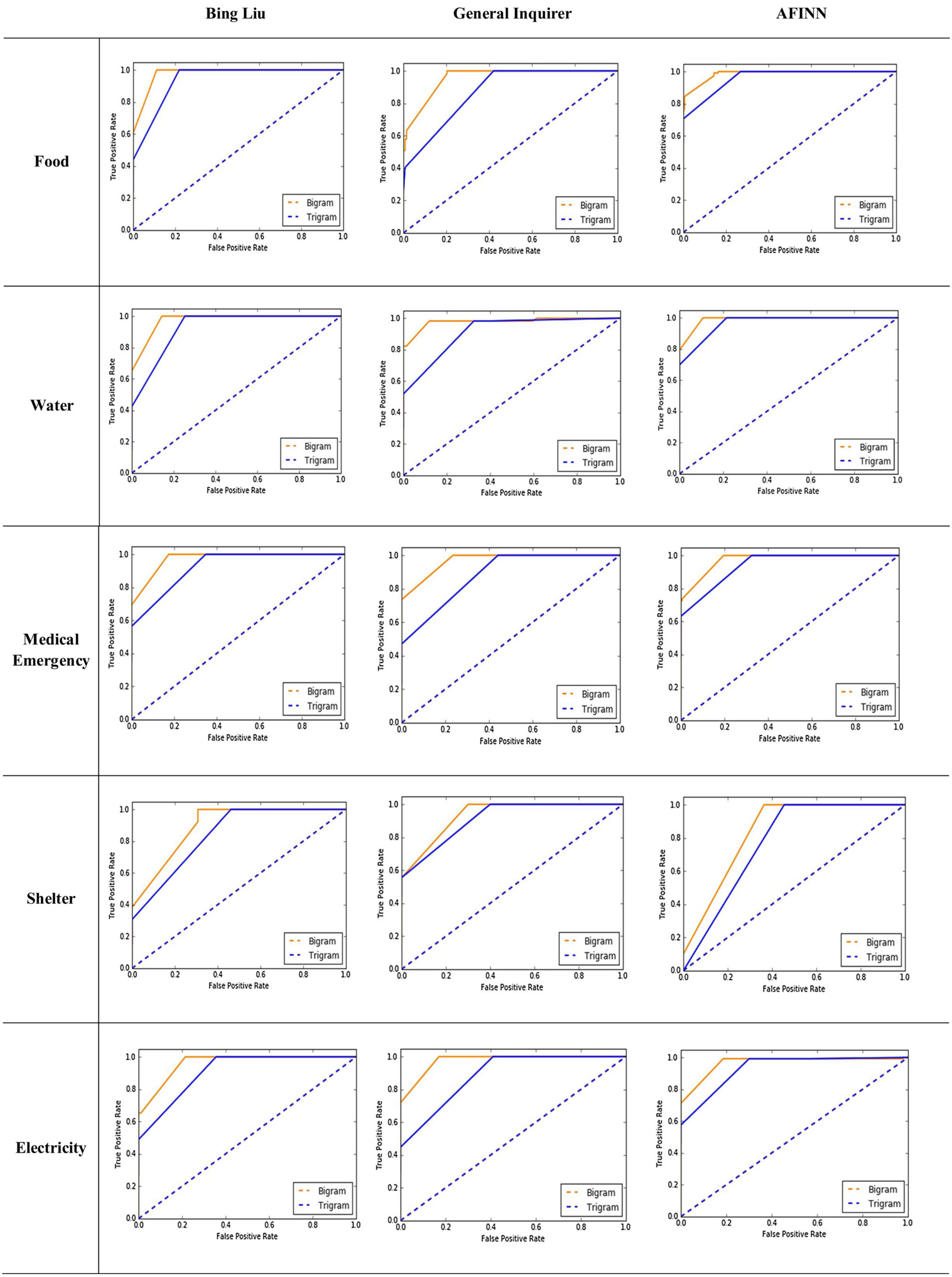


Fig. 5. ROC plot for bigram and trigram features using the subjective phrases.

$$P_C = \sum_{T_C=1}^{n} BL_P(T_C) + GI_P(T_C) + AFINN_P(T_C) \tag{1}$$

Where, n is the total number of tweets in each category, C is the category of disaster data, T_C is the tweets in a particular category, BL_P is the positive sentiment calculated by Bing Liu, GI_P is the positive sentiment calculated by General Inquirer and $AFINN_P$ is the positive sentiment calculated by AFINN lexicon. Similarly, the overall negative score (N_C) for each category of disaster data is calculated as,

$$N_C = \sum_{T_C=1}^{n} BL_N(T_C) + GI_N(T_C) + AFINN_N(T_C) \tag{2}$$

The overall sentiment score (SS_C) for each category of data like, water, food, shelter, medical emergency, electricity is calculated by using the three lexicon sets as,

$$SS_C = \frac{\left(\sum_{T_C=1}^{n} BL_P(T_C) + GI_P(T_C) + AFINN_P(T_C)\right) - \left(\sum_{T_C=1}^{n} BL_N(T_C) + GI_N(T_C) + AFINN_N(T_C)\right)}{\left(\sum_{T_C=1}^{n} BL_P(T_C) + GI_P(T_C) + AFINN_P(T_C)\right) + \left(\sum_{T_C=1}^{n} BL_N(T_C) + GI_N(T_C) + AFINN_N(T_C)\right)} \tag{3}$$

The sentiment analysis for each of the category of disaster data is calculated using Eqs. (1)–(3). Box plot are used to draw groups of sentiment scores. The distribution of the sentiment and the level of the sentiment score in each category are analyzed using box plot. Box plot of the sentiment score associated with each subjective word through the distribution of tweets in each category using AFINN is shown in Fig. 6.

6. Discussions

The paper presents an automated system to analyze the sentiment towards the various identified needs of the affected people during a disaster. The proposed text classification method is tested using the disaster dataset collected from Twitter domain. The research concentration is in identifying the right features that categorize the data with highest accuracy. In order to identify the right features, two methods are evaluated. The first method uses the bag words feature and POS tagging based features. In the second method, subjective phrases are extracted with a window size of three for the application of bigram and trigram features. A comparison of the F-score of all the results provide a clear idea on the feature that performs better in each category

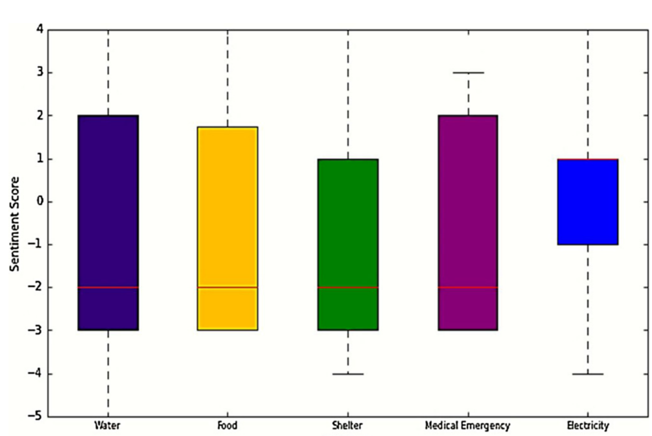


Fig. 6. Box plot of sentiment score for different category using AFINN.

of the disaster data. In the overall analysis, the combination of subjective phrase and machine learning algorithm with bigram feature yields better classification accuracy of disaster related data.

The strongest contribution of this paper is that it solves some of the challenges in using tweets for disaster response such as, the volume of information generated during disaster and the inability to segregate the data into different categories. The big data analytics help in managing such volume of data through distributed file system and parallel programing. The proposed method helps to categorize the data into various categories of needs to help the first responders. The existing literature identified that social media data is of significant importance in analyzing the sentiment of the people during disaster (Vo and Collier, 2013; Sen et al., 2015). Though these studies identify the emotions of the people like anger, unpleasantness, sadness, anxiety, fear and relief, the proposed method is unique from them since it analyzes the sentiment of the people towards the philanthropic needs during any disaster. According to Caragea et al. (2014), the affected people are often the first responders. They often involve in the rescue operations and they are well informed of the situation at the place of disaster. The segregation of the text is highly important and useful to the first responders for effective response during the disaster situation.

The main advantage of the proposed method is that it segregates tweets from the affected people about their needs and apply sentiment analysis on such tweets to identify whether people are served with their needs. Furthermore, the visual representation of the sentiment through the box plot helps the responders and emergency personnel to identify the sentiment towards the particular category of the need and act accordingly. The visual sentiment representation helps in monitoring and maintaining the required inventory to manage the needs of the people in disaster affected area. The geo annotation property of the tweet aids in identifying the location of the help request. This helps the government organizations to mobilize the required materials towards the identified location rather than dumping the relief materials where there is surplus supply. The visual sentiment analysis can be used by the government organizations and rescue personnel in the preparation phase for the upcoming disasters. The analysis helps the emergency responders to read the minds of the affected people and prepare themselves accordingly to soothe those affected people at the hour of their need. The analysis of the sentiment helps the emergency responders to build better strategies for relief operations.

Another advantage of the proposed method is that it tries to bridge the gap between common people who are willing to help the affected ones. During manual scrutiny of the disaster related tweets, few tweets like, "*Does anyone know where I can donate blankets, clothes food or just money to help out with the floods in Kashmir and parts of Pakistan*" reveal that though there are many volunteers who are willing to help the affected people, there is no provision to locate where help is needed. The results of the proposed method will be helpful in such situations to identify the exact area and type of needs of the affected people. Albuquerque et al. (2015) revealed that people in flood affected areas where water level is +0.75 m tend to tweet 54 times more on the topic than people who are far away from the affected region. In such situations, the geo tagged tweets are helpful in identifying the location of the affected people and reach them at the right time.

Despite of the various opportunities in utilizing social media for crisis response, there are few concerns in its usage. The challenges involved in using social media for disaster response include, difficulty in collecting the disaster related data to build a better sentiment model for disaster analysis, lack of standard crisis data set or disaster related lexicon for accurate evaluation of the needs of the people. In order to fulfill these challenges, the future work can focus on building an ontology on the needs of the people, and also create a lexicon with disaster related keywords. Also, it is highly necessary to collect the tweets and messages from various disasters for effective classification of the data and create a standard dataset for evaluation.

The major impact of the proposed model is the real time segregation of the texts according to the needs of the people. The output of the text classification system helps the emergency responders in providing the necessary services to the people during the worst time of their life. This indeed reduces the stress and number of causalities at the place of disaster. The proposed big data approach proves that tweets can be used as data for disaster text classification which helps in the effective response and recovery process. Once the disaster data is analyzed efficiently, they can be used as input to the preparedness phase for the upcoming disasters.

7. Conclusion

In this paper, a big data driven approach is proposed and the various phases involved in sentiment categorization are discussed. The main contribution of this paper is the detailed study of the learning and classification methodology to classify the needs of the people during the times of disaster. A method to visualize the sentiment is proposed for analyzing the sentiments about the various basic needs of the people affected in the disaster. The combination of subjective phrase and machine learning algorithm yields better classification accuracy for disaster data. The big data based research also analyzes about the various challenges that are involved in using Twitter data for disaster response and recovery. It is evident from the comparative analysis that the lexicons that are available for text analysis need to be extended to accommodate disaster related data. Also, building an ontology for the crisis data can improve the classification of the needs of the people during disaster.

References

Abbasi, M., & Kumar, S. (2012). Lessons learned in using social media for disaster relief-ASU crisis response game. *Social computing, behavioral – Cultural modeling and prediction (LNCS 7227),* 282–289. http://dx.doi.org/10.1007/978-3-642-29047-3_34.

Albuquerque, J. P., Herfort, B., Brenning, A., & Zipf, A. (2015). A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management. *International Journal of Geographical Information Science, 29*(4), 667–689. http://dx.doi.org/10.1080/13658816.2014.996567.

Aramaki, E., Maskawa, S., & Morita, M. (2011). Twitter catches the flu: Detecting influenza epidemics using twitter. *Proceedings of the 2011 conference on empirical methods in natural language processing,* 1568–1576.

Baccianella, S., Esuli, A., & Sebastiani, F. (2010). SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. *Proceedings of the seventh international conference on language resources and evaluation (LREC'10),* 2200–2204.

Bradley, M. M., & Lang, P. J. (1999). *Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings. Technical report C-1, the center for research in psychophysiology*. University of Florida1–45.

Bravo-Marquez, F., Mendoza, M., & Poblete, B. (2014). Meta-level sentiment models for big social data analysis. *Knowledge-Based Systems, 69*(1), 86–99. http://dx.doi.org/10.1016/j.knosys.2014.05.016.

Cambria, E., Olsher, D., & Rajagopal, D. (2014). SenticNet 3: A common and common-sense knowledge base for cognition-driven sentiment analysis. *Proceedings of the twenty-eighth AAAI conference on artificial intelligence,* 1515–1521.

Caragea, C., Squicciarini, A., Stehle, S., Neppalli, K., & Tapia, A. (2014). Mapping moods: Geo-mapped sentiment analysis during hurricane sandy. *Proceedings of the 11th international conference on information systems for crisis response and management – ISCRAM 2014,* 642–651.

Cho, H., Kim, S., Lee, J., & Lee, J. S. (2014). Data-driven integration of multiple sentiment dictionaries for lexicon-based sentiment classification of product reviews. *Knowledge-Based Systems, 71*, 61–71. http://dx.doi.org/10.1016/j.knosys.2014.06.001.

Deng, L., & Wiebe, J. (2015). MPQA 3.0: An entity/event-level sentiment corpus. *Proceedings of the 2015 conference of the North American chapter of the association for computational linguistics: Human language technologies,* 1323–1328.

Earle, P. S., Bowden, D. C., & Guy, M. (2011). Twitter earthquake detection: Earthquake monitoring in a social world. *Annals of Geophysics, 54*(6), 708–715. http://dx.doi.org/10.4401/ag-5364.

Fang, X., & Zhan, J. (2015). Sentiment analysis using product review data. *Journal of Big Data, 2*(5), 1–14. http://dx.doi.org/10.1186/s40537-015-0015-2.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management, 35*(2), 137–144. http://dx.doi.org/10.1016/j.ijinfomgt.2014.10.007.

Gensler, S., Völckner, F., Liu-Thompkins, Y., & Wiertz, C. (2013). Managing brands in the social media environment. *Journal of Interactive Marketing, 27*(4), 242–256. http://dx.doi.org/10.1016/j.intmar.2013.09.004.

Gerber, M. S. (2014). Predicting crime using Twitter and kernel density estimation. *Decision Support Systems, 61*(1), 115–125. http://dx.doi.org/10.1016/j.dss.2014.02.003.

Gitto, S., & Mancuso, P. (2017). Improving airport services using sentiment analysis of the websites. *Tourism Management Perspectives, 22*, 132–136. http://dx.doi.org/10.1016/j.tmp.2017.03.008.

Graham, M. W., Avery, E. J., & Park, S. (2015). The role of social media in local government crisis communications. *Public Relations Review, 41*(3), 386–394. http://dx.doi.org/10.1016/j.pubrev.2015.02.001.

Gurman, T. A., & Ellenberger, N. (2015). Reaching the global community during disasters: Findings from a content analysis of the organizational use of Twitter after the 2010 Haiti Earthquake. *Journal of Health Communication, 20*(6), 1–10. http://dx.doi.org/10.1080/10810730.2015.1018566.

He, W., Zha, S., & Li, L. (2013). Social media competitive analysis and text mining: A case study in the pizza industry. *International Journal of Information Management, 33*(3), 464–472. http://dx.doi.org/10.1016/j.ijinfomgt.2013.01.001.

Hogenboom, A., Heerschop, B., Frasincar, F., Kaymak, U., & De Jong, F. (2014). Multilingual support for lexicon-based sentiment analysis guided by semantics. *Decision Support Systems, 62*, 43–53. http://dx.doi.org/10.1016/j.dss.2014.03.004.

Hu, Y. H., & Chen, K. (2016). Predicting hotel review helpfulness: The impact of review visibility, and interaction between hotel stars and review ratings. *International Journal of Information Management, 36*(6), 929–944. http://dx.doi.org/10.1016/j.ijinfomgt.2016.06.003.

Jeong, B., Yoon, J., & Lee, J. M. (2017). Social media mining for product planning: A product opportunity mining approach based on topic modeling and sentiment analysis. *International Journal of Information Management*. http://dx.doi.org/10.1016/j.ijinfomgt.2017.09.009 (In Press).

Kapukaranov, B., & Nakov, P. (2015). Fine-grained sentiment analysis for movie reviews in Bulgarian. *Proceedings of recent advances in natural language processing,* 266–274.

Khan, F. H., Qamar, U., & Bashir, S. (2016). Senti-CS: Building a lexical resource for sentiment analysis using subjective feature selection and normalized Chi-Square-based feature weight generation. *Expert Systems, 33*(5), 489–500. http://dx.doi.org/10.1111/exsy.12161.

Kim, H., Howland, P., & Park, H. (2005). Dimension reduction in text classification with support vector machines. *Journal of Machine Learning Research, 6*, 37–53.

Leong, C. M. L., Pan, S. L., Ractham, P., & Kaewkitipong, L. (2015). ICT-enabled community empowerment in crisis response: Social media in Thailand flooding 2011. *Journal of the Association for Information Systems, 16*(3), 174–212.

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies,* 1–108. http://dx.doi.org/10.2200/S00416ED1V01Y201204HLT016.

Lv, Y., Chen, Y., Zhang, X., Duan, Y., & Li, N. L. (2017). Social media based transportation research: the state of the work and the networking. *IEEE/CAA Journal of Automatica Sinica, 4*(1), 19–26. http://dx.doi.org/10.1109/JAS.2017.7510316.

Marcus, M. P., Santorini, B., & Marcinkiewicz, M. A. (1993). Building a large annotated corpus of english: the penn treebank. *Computational Linguistics, 19*(2), 313–330.

Mohammad, S. M., & Kiritchenko, S. (2015). Using hashtags to capture fine emotion categories from tweets. *Computational Intelligence, 31*(2), 301–326. http://dx.doi.org/10.1111/coin.12024.

Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word-emotion association lexicon. *Computational Intelligence, 29*(3), 436–465. http://dx.doi.org/10.1111/j.1467-8640.2012.00460.x.

Mohammad, S. (2011). Colourful language: Measuring word –Colour associations. *Proceedings of the 2nd workshop on cognitive modeling and computational linguistics,* 97–106.

Mudinas, A., Zhang, D., & Levene, M. (2012). Combining lexicon and learning based approaches for concept-level sentiment analysis. *Proceedings of the first international workshop on issues of sentiment discovery and opinion mining – WISDOM '12,* 1–8. http://dx.doi.org/10.1145/2346676.2346681.

Neviarouskaya, A., Prendinger, H., & Ishizuka, M. (2009). SentiFul: Generating a reliable lexicon for sentiment analysis. *Proceedings of the 3rd international conference on affective computing and intelligent interaction and workshops,* 1–6. http://dx.doi.org/10.1109/ACII.2009.5349575.

Nielsen, F.Å. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *CEUR workshop proceedings,* 93–98.

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. *Proceedings of the ACL-02 conference on empirical methods in natural language processing,* 79–86. http://dx.doi.org/10.3115/1118693.1118704 10.

Paul, M. J., Dredze, M., & Broniatowski, D. (2014). Twitter improves influenza forecasting. *PLOS Currents Outbreaks, 1*, 1–13. http://dx.doi.org/10.1371/currents.outbreaks.90b9ed0f59bae4ccaa683a39865d9117.

Procter, R., Vis, F., & Voss, A. (2013). Reading the riots on Twitter: Methodological innovation for the analysis of big data. *International Journal of Social Research Methodology, 16*(3), 197–214. http://dx.doi.org/10.1080/13645579.2013.774172.

Qadir, J., Ali, A., ur Rasool, R., Zwitter, A., Sathiaseelan, A., & Crowcroft, J. (2016). Crisis analytics: Big data-driven crisis response. *Journal of International Humanitarian Action, 1*(1), 12–10.1186/s41018-016-0013-9.

Ragini, J. R., & Anand, P. M. R. (2016). An empirical analysis and classification of crisis related tweets. *Proceedings of the IEEE international conference on computational intelligence and computing research (ICCIC),* 1–4. http://dx.doi.org/10.1109/ICCIC.2016.7919608.

Ragini, J. R., Anand, P. M. R., & Bhaskar, V. (2017). Mining crisis information: A strategic approach for detection of people at risk through social media analysis. *International Journal of Disaster Risk Reduction, 27*(December), 556–566. http://dx.doi.org/10.1016/j.ijdrr.2017.12.002.

Robinson, B., Power, R., & Cameron, M. (2013). A sensitive twitter earthquake detector. *Proceedings of the 22nd international conference on world wide web,* 999–1002. http://dx.doi.org/10.1145/2487788.2488101.

Roshan, M., Warren, M., & Carr, R. (2016). Understanding the use of social media by organisations for crisis communication. *Computers in Human Behavior, 63*, 350–361. http://dx.doi.org/10.1016/j.chb.2016.05.016.

Sakaki, T., Okazaki, M., & Matsuo, Y. (2013). Tweet analysis for real-time event detection and earthquake reporting system development. *IEEE Transactions on Knowledge and Data Engineering, 25*(4), 919–931. http://dx.doi.org/10.1109/TKDE.2012.29.

Sen, A., Rudra, K., & Ghosh, S. (2015). Extracting situational awareness from microblogs during disaster events. *Proceedings of the 7th international conference on communication systems and networks ? COMSNETS 2015,* 1–6. http://dx.doi.org/10.1109/COMSNETS.2015.7098720.

Stone, P. J., Bales, R. F., Namenwirth, J. Z., & Ogilvie, D. M. (1962). The general inquirer: A computer system for content analysis and retrieval based on the sentence as unit of information. *Computers in Behavioral Science, 7*(4), 484–498.

Subrahmanian, V. S., & Reforgiato, D. (2008). AVA: adjective-verb-adverb combinations for sentiment analysis. *IEEE Intelligent Systems, 23*(4), 43–50. http://dx.doi.org/10.1109/MIS.2008.57.

Tapia, A. H., Moore, K. A., & Johnson, N. J. (2013). Beyond the trustworthy tweet: A deeper understanding of microblogged data use by disaster response and humanitarian relief organizations. *Proceedings of the 10th international ISCRAM conference,* 770–779.

Toriumi, F., Sakaki, T., Shinoda, K., Kazama, K., Kurihara, S., & Noda, I. (2013). Information sharing on Twitter during the 2011 catastrophic earthquake. *Proceedings of the 22nd international conference on world wide web,* 1025–1028.

Torkildson, M. K., Starbird, K., & Aragon, C. R. (2014). Analysis and visualization of sentiment and emotion on crisis tweets. *Lecture Notes in Computer Science, 8683*, 64–67. http://dx.doi.org/10.1007/978-3-319-10831-5_9.

Twitter Historical data recovery tool. http://www.followthehashtag.com/features/Twitter-historical-data-recover/ [Accessed 18.03.17].

Umihara, J., & Nishikitani, M. (2013). Emergent use of twitter in the 2011 Tohoku earthquake. *Prehospital and Disaster Medicine, 28*(5), 434–440. http://dx.doi.org/10.1017/S1049023X13008704.

Vieweg, S., Hughes, A. L., Starbird, K., & Palen, L. (2010). Microblogging during two natural hazards events: What twitter may contribute to situational awareness. *Proceedings of the SIGCHI conference on human factors in computing systems – CHI '10,* 1079–1088. http://dx.doi.org/10.1145/1753326.1753486.

Vo, B.-K. H., & Collier, N. (2013). Twitter emotion analysis in earthquake situations. *International Journal of Computational Linguistics and Applications, 4*(1), 159–173.

Wu, D. D., Zheng, L., & Olson, D. L. (2014). A decision support approach for online stock forum sentiment analysis. *IEEE Transactions on Systems, Man, and Cybernetics: Systems, 44*(8), 1077–1087. http://dx.doi.org/10.1109/TSMC.2013.2295353.

Yang, F.-C., Lee, A. J. T., & Kuo, S.-C. (2016). Mining health social media with sentiment analysis. *Journal of Medical Systems, 40*(11), 236. http://dx.doi.org/10.1007/s10916-016-0604-4.

Yates, D., & Paquette, S. (2011). Emergency knowledge management and social media technologies: A case study of the 2010 Haitian earthquake. *International Journal of Information Management, 31*(1), 6–13. http://dx.doi.org/10.1016/j.ijinfomgt.2010.10.001.

Zhu, X., Kiritchenko, S., & Mohammad, S. M. (2014). NRC-Canada-2014: Recent improvements in the sentiment analysis of tweets. *Proceedings of the eighth international workshop on semantic evaluation exercises (SemEval-2014),* 443–447.

Zielinski, A., Middleton, S. E., Tokarchuk, L., & Wang, X. (2013). Social media text mining and network analysis for decision support in natural crisis management. *Proceedings of the 10th international ISCRAM conference,* 840–845.