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

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