Target Oriented Tweets Monitoring System during Natural Disasters

Si Si Mar Win
University of Computer Studies, Mandalay
Mandalay, Myanmar
sisimarwin@gmail.com

Than Nwe Aung
University of Computer Studies, Mandalay
Mandalay, Myanmar
mdytna@gmail.com

***Abstract*—Twitter, Social Networking Site, becomes most popular microblogging service and people have started publishing data on the use of it in natural disasters. Twitter has also created the opportunities for first responders to know the critical information and work effective reactions for impacted communities. This paper introduces the tweet monitoring system to identify the messages that people updated during natural disasters into a set of information categories and provide user desired target information type automatically. In this system, classification is done at tweet level with three labels by using LibLinear classifier. This system intended to extract the small number of informational and actionable tweets from large amounts of raw tweets on Twitter using machine learning and natural language processing (NLP). Feature extraction of this work exploited only linguistic features, sentiment lexicon based features and especially disaster lexicon based features. The annotation system also creates disaster related corpus with new tweets collected from Twitter API and annotation is done on real time manner. The performance of this system is evaluated based on four publicly available annotated datasets. The experiments showed the classification accuracy on the proposed features set is higher than the classifier based on neural word embeddings and standard bag-of-words models. This system automatically annotated the Myanmar_Earthquake_2016 dataset at 75% accuracy on average.**

Keywords— Twitter; NLP; LibLinear

I. INTRODUCTION

Due to their ease of use and simplicity, social media platforms can provide efficient delivery of information that can give better situational awareness for emergency response. Unfortunately, this vast amount of information can be useless or even dangerous, since its reliability is often unclear and any uncertainties can result in chaos [15].

Twitter allows its subscribers to express and share short text messages, called tweets, of up to 140 characters. These tweets are used to broadcast relevant information and report news of emergency situations. According to the March 2016 report by SOCIAL PILOT, total 500 million tweets are sent per day [18]. The rapid growth of online information services, social media and other digital format documents means that large amounts of information are becoming immediately available and readily accessible to numerous end-users. However, human ability to organize and understand a large numbers of social media text is limited. Methods requiring extensive human attention and interpretation do not scale robustly to large social media data sets.

Moreover, most of these numerous posts on twitter are not useful in providing information about the disasters. Due to the analysis of Gupta et al., only 17% of the total tweets posted about the event contained situational awareness information that was credible [1]. There is still needed to develop the technologies for filtering and retrieving the informational or credible tweets automatically during disasters. There are three problems involved in developing a system that can classify tweets as originating in social media applications into specific information categories. The first problem is to deal with the massive amount of tweets arriving per minute. The second is effective features extraction for noisy and not curated short text messages [17]. Tweets are highly vary in terms of subject and content and the influx of tweets particularly in the event of a disaster may be overwhelming. It is impractical to automatically classify these varied tweets in order to extract needed information. Tweets classification is therefore a third challenge.

Analyzing and extracting informational tweets from Twitter during disasters is one of the text mining researches in recent years. In this paper, a system for the monitoring of actionable social media post into corresponding predefined types is introduced that addresses the above mentioned three challenges. This system considers the problems as feature extraction, two layer classification and required disaster corpus creation problem. Each new data item uploaded to the system is either classified into one of three types. In summary, the contribution of this work is three folds: 1. Create different disasters corpus of tweets annotated with three labels: Related and Informative as Informative, Related and Not Informative as Other Information, and Not Related as Not Informative for future use. 2. Competitive, easily implementable feature extraction method that act as a benchmark for automated classification approaches for natural disaster related datasets by using natural disaster lexicon. 3. Creation of extended natural disasters lexicon based on publicly available annotated datasets and newly annotated corpus.

The rest of the paper is organized as follow: Section II presents the closely related work to this paper. Section III explains the methodology that we used in collecting, preprocessing, feature extraction, disaster lexicon creation, and classification scheme used for annotating the tweets. Section IV describes the architecture of the proposed system. Section V

978-1-5090-5507-4/17/$31.00 ©2017 IEEE
ICIS 2017, May 24-26, 2017, Wuhan, China

expresses the datasets details, experiments and analysis performed. And it also discusses the ranking based feature selection performed on the extracted features of the proposed system. Section VI summarizes the results from our analysis and highlights the implications of our results. This section also describes the future work of the proposed system.

II. RELATED WORK

Most of the Social Media Analysis research have tried to address the problems of trust, scalability and credibility on Online Social Media using different techniques. There has been worked on the applications of online social media such as information diffusion, information extraction, topic detection, credibility prediction, event identification and other types of social media irrelevant content analysis. Feature extraction and some previous approaches to critical tweets classification are closely related to the proposed work.

Related with features extraction from tweets, some of the researchers have focused on what types of features are used. Castillo et al. utilized four types of features such as 21 message-based, 7 user-based, 35 topic-based, and 5 propagation-based to make a classifier for evaluating the credibility of tweets [8]. In their research, they focused on the level of credibility of every trend on Twitter. Gupta et al. used six types of features such as Tweet Meta-data Features, Tweet Content Features, User based Features, Network Features, Linguistic Features and External Resource Features for credibility analysis [1]. Besides, hashtags have been effectively utilized as critical features for various tasks of text or social media analysis, including tweet classification [3].

Unlike them, this system focuses on the content based features such as Linguistic Features, NRC Sentiment hashtags lexicons based Features and Disaster Lexicon based features only. Although, the rapid growth of social media, it continues to remain on the scalability issues of credibility prediction or tweet classification. Therefore, another area of related research is checking and classifying for informative messages in microblog platforms. The classification of tweets as Credible or Not Credible is presented in [8].

Moreover, the information detection and extraction system for microblog posts was described by Imran et al. [12]. In their work, Naïve Bayesian classifiers were used to classify a tweet into one of the types such as Caution and Advice, Informative source, Donation, and Causalities & damage. Gupta et al. also provided a SVM-rank based system, TweetCred to assign a credibility score to tweets in a user's timeline [2]. According to the literatures, supervised machine learning algorithms have been applied by most of the researchers to detect and classify the content in OSM. Naive Bayes (NB) and Support Vector Machine (SVM) are used for tweets classification in [4].

In contrast, the proposed system uses LibLinear classifier to annotate the tweets. This classifier is one of the most promising learning techniques for large sparse data with huge number of instances and features. This system performs not only the tweets annotation but also consider the creation of specific disaster related lexicon based on the top natural disasters types that often causes around the world.

III. METHODOLOGY

Tweet contents analysis can be applied to all kinds of text analysis but certain domains and modes of communication tend to have more expressions of very short text messages. Social media mining for disaster response and coordination has been receiving an increasing level of attention from the research community, no effort has been devoted to provide automatic data annotation from social media that cover all possible disaster situations. Data annotation has always been carried out by human annotator or crowd source workers.

Lack of well-defined values in choosing machine learning algorithms suitable for a given problem remains a major challenge. To address these problems, we analyze statuses updated on Twitter about natural disasters and perform automatic classification and annotation on these tweets. And then we also provide the annotated datasets for building appropriate model for credibility assessment.

The next section describes the three main functions of the proposed system: Data Collection, Feature Extraction, Ground Truth Labeling or classification for annotation learning. The proposed system focuses on attaching only contents of tweets.

A. Tweets Collection

This function works for new tweets collection. It collects messages from Twitter using the Twitter streaming API. The data collection process focuses on the exact matching of target keywords to acquire tweets and build the query using user defined keywords or hashtags.

Using the relevant keywords or hashtags for queries are the best way to extract the most relevant tweets during crisis or disasters. To get more tweets and to overcome the Twitter API limit, we use user desired target keywords or hashtags such as #MyanmarEarthquake hashtag is applied to acquire the news of earthquake that struck in Myanmar. And then we set the query time setting to the last six days that covers the some natural disasters which happen over a long period of time.

B. Tweets Preprocessing

This step preprocesses the tweet content before creating the numeric vector. Firstly, this task removes the tweets which already contains the same text in the previous preprocessed tweets to reduce the redundancy and noise.

Secondly, stop_words from collected tweets are used to reduce dimensionality of the dataset and thus terms left in the tweets can be identified more easily by the feature extraction process. Stop_words are common and high frequency words such as "a", "the", "of", "and", "an" "in" etc.

Finally, the stemming process converts all the inflected words present in the text into a root form called a stem [7]. For example, 'automatic,' 'automate,' and 'automation' are each converted into the stem 'automat'. For the purpose of stemming, this system uses a popular snowball stemmer.

C. Feature extraction

Feature extraction is the transformation of arbitrary data such as images or text into numerical features usable for

classification. In this work, it is concerned with altering tweet contents into a simple numeric vector representation. To do this, each tweet is tokenized into hashtags, user mention, URLs, special characters such as punctuation or emotion using ARK Tweet NLP [11].

This process receives the tweets from preprocessing step, it extracts the features by using ARK POS tagger and different lexicons. The features used in this work are only extracted from tweet contents. We do not consider source or user based features such as number of followers, number of friends, number of messages that user posted on, etc. These are described in TABLE I.

To extract neural word embeddings features for comparison to proposed feature set, this system used Word2vec model in Deep Learning4J. Word2Vec is the representations of words with the help of vectors in such manner that semantic relationships between words preserved as basic linear algebra operations [9]. The following parameters were used while training for Word2Vec: 100 dimensional space, 10 minimum words and 10 words in context. After transforming 100 dimension feature vector of each word in the corpus, this system used t-Distributed Stochastic Neighbor embedding (t-SNE) technique to reduce 100 dimensions of each feature vector to more relevant 10 dimensions feature vector.

In order to identify the optimal combination of features that provide good prediction results, this system reduced the features with little or no impact over the results to keep the number of features as small as possible. Using less features seemed to have affected the classifier. Therefore, after creating feature vector, all terms with occurrence less than 3 are removed from the feature space to exploit a reasonable amount of features.

TABLE I. FEATURES USED IN THE PROPOSED SYSTEM

Feature	Description
Word Ngram	Unigram and bigram are extracted for each word in tweet text after stop-word removing and stemming.
Word Cluster	1000 Brown clusters in Twitter Word Clusters made available by CMU ARK group
Emotions	Each emotion words contained in text
Lexicon Based Features	Lexical matching based on NRC Hashtag and CrisisLex with PMI Score
POS	List of part of speech tags that occur in the tweet generated by CMU ARK POS-Tagger
Hashtags	Number of Hashtags, each hashtags contained in tweets
URLs	Number of URLs, and each URL contained in tweets

D. Creating Disaster Lexicons from Annotated Tweets

The words used in Twitter include many abbreviations, acronyms, slang and misspelled words that are not observed in traditional media or covered by popular lexicons. However, we observed that different natural disaster related tweets may have composed of same terms such as need, pray, damage, death, destroy, survivor etc. According to this observation, we decided to apply the crisis lexicon for feature extraction process.

This system creates the disaster lexicon which contains specific natural disaster related terms with a point wise mutual information (PMI) based score and frequency distribution of these terms based on the set of annotated disaster datasets. This lexicon creation process follows the method of Olteanu et al. In this process, we exploit their natural disaster related datasets, publicly available natural disaster related datasets and newly annotated dataset such as Myanmar_Earthquake_2016 dataset collected by proposed system for lexicon expansion or keywords (disaster related terms) adaptation. This automatically created lexicon is used in feature extraction process of the proposed system.

E. Classification of Tweets

This function automatically classifies the information in tweets. To perform this task, the proposed system trained a LibLinear classifier operating on extracted features set. LibLinear solves large-scale classification problems in many applications such as text classification. It is very efficient for training large scale. It takes only several seconds to train more than 600,000 examples while LibSVM takes several hours for same task [16].

Given a set of features and a learning corpus (i.e. the annotated dataset), the classifier trains a statistical model using the feature statistics extracted from the corpus. This trained model is then employed in the classification of unknown tweets and, for each tweet, it assigns the probability of belonging to a class: Related and Informative, Related and Not Informative, and Not Related.

The annotated datasets required by the system can be obtained from using Artificial Intelligence for Disaster Response (AIDR) [14] and CrisisLexT26 [5]. This system uses datasets in English language only.

IV. THE ARCHITECTURE OF THE PROPOSED SYSTEM

The holy grail of text annotation is an automated system that accurately and reliably annotates very large numbers of cases using relatively small amounts of manually annotated training data [6]. In this system, annotation is restricted to tweets in English language. Non-English tweets are not considered.

The goal of this work is to provide a system that automatically creates the different disasters dataset with annotated tweets. The annotation at the tweet-level will be three types: Related and Informative, Related and Not Informative, and Not Related. This system also analyzed which features are important in the data to annotation. It applied the annotated corpus to train a classifier that automatically annotates the tweets.

The system, illustrated in Fig. 1, first collects the tweets from Twitter by using user desired query terms or target disaster related terms. After collecting the tweets, it removes the redundant tweets and then it also eliminates the stop_words.

In feature extraction, this system applies Linguistic features such as Word N-grams, POS features, Sentiment Lexicon features using NRC Hashtag Sentiment Lexicon [13] and Natural disaster lexicon which is extended from the

Olteanu et al. [4] and the other features such as Hashtags and URLs.

In annotation, to improve model performance, the best set of features were chosen by using Information gain theory based feature selection method. LibLinear classifier uses these selected features subset for tweets categorization to create annotated corpus and to provide informative tweets to the users.

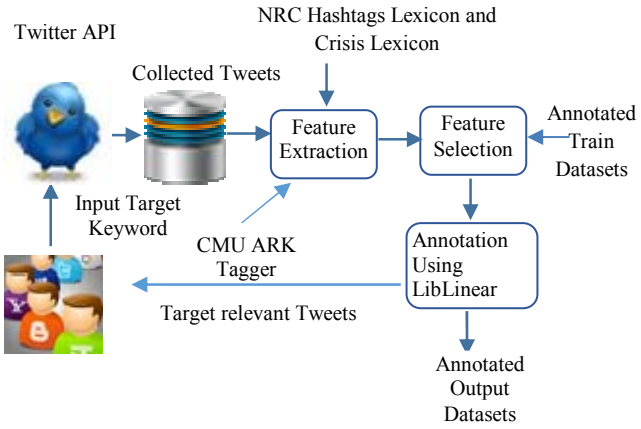

Fig. 1. Architecture of the Proposed Tweets Monitoring System

V. EXPERIMENTS

This system performs a set of preliminary experiments to evaluate the effectiveness of feature extraction, feature selection model and classifier model on the performance of the proposed approach. For feature extraction, the proposed system applied three models such as neural word embedding, BOW with Unigram and Bigram model and the proposed model.

The final experiment is done under the best development settings in order to evaluate the classifier model with the best feature set. This section presents experiments and results for classification of four annotated datasets. For each of the classification tasks, this system applies three different combinations of the following feature set:

1) Neural word embeddings and others features
2) Standard BOW model with Bigram and other features
3) Unigram, Bigram plus Crisis Lexicon based feature and the other features.

The third feature combination is used in this system. The other features represented the Word Cluster, Hashtags, Emoticons and URLs described in TABLE I. This system extracted the new lexical features such as natural disaster related terms using extended crisis lexicon or disaster related lexicon based on CrisisLex.

A. Datasets and Settings

In the experiment, this system uses people freely available 4 annotated natural disaster datasets from 2013 to 2015, occurring in different countries which affected up to several million. These datasets are already annotated with different information types. To annotate the dataset with ground truth label, this system normalized the four available datasets with only three classes. It annotated the data with all specific information type as Informative, Not related or irrelevant as Not Informative and Other News or related but not informative as Other Information. Detailed information of datasets is described in TABLE II.

TABLE II. DATASETS USED IN THE PROPOSED SYSTEM

Different Natural Disaster Related Datasets			
Name	***Informative***	***Not Informative***	***Other Information***
2013_Australia_bushfire	250	250	245
2013_Typhoon_Yolanda	765	105	175
2014_Iceland Volcano	165	250	-
2015_Nepal_ Earthquake	1500	2500	575

This system performed 10 fold cross validation to test the efficiency of the feature extraction and the model built during the training and testing phase. The results along with the experimentation of different datasets are described based on accuracy, precision, recall and F1 score of classifier model for feature extraction performance.

B. Effectiveness of Feature Extraction

This system assesses the effectiveness of extracted feature by comparing the combination of different features on classifier with cross validation mode. Accuracy, precision and recall values for classification tasks using different features on different datasets by LibLinear classifier are shown in TABLE III, TABLE IV and TABLE V.

We also tested the extracted feature set on four different classifiers such as Random Forest, SMO (fast training algorithm for SVM), Naïve Bayes and our LibLinear classifier that are well known in text classification process.

Due to the development experiments, the performance of Random Forest, Naïve Bayes and SMO was sensitive to the large number of features. Therefore, this system used Information Gain based feature selection method to get better performance and to reduce inconsistent features. This wok uses a well-known WEKA machine learning tools for implementation of Random Forests, Naïve Bayes, SMO, LibLinear and Information Gain based feature selection methods [10].

TABLE III. CLASSICICATION RESULTS USING NEURAL WORD EMBEDDINGS

Neural Word Embeddings by LibLinear				
Dataset	***Precision***	***Recall***	***F1***	***Accuracy***
2013_Australia_bushfire	0.718	0.709	0.713	70.85%
2013_Typhoon_Yolanda	0.747	0.735	0.74	73.45%
2014_Iceland Volcano	0.786	0.783	0.784	78.26%
2015_Nepal_ Earthquake	0.671	0.669	0.67	66.95%

TABLE IV. CLASSICICATION RESULTS USING STANDARD BAG OF WORDS MODEL

BOW by LibLinear				
Dataset	***Precision***	***Recall***	***F1***	***Accuracy***
2013_Australia _bushfire	0.787	0.792	0.787	79.22%
2013_Typhoon _Yolanda	0.782	0.797	0.777	79.74%
2014_Iceland Volcano	0.905	0.903	0.902	90.33%
2015_Nepal_ Earthquake	0.747	0,750	0.747	75.02%

TABLE V. CLASSICICATION RESULTS USING PROPOSED FEATURE SET

Proposed Features by LibLinear				
Dataset	***Precision***	***Recall***	***F1***	***Accuracy***
2013_Australia _bushfire	0.897	0.895	0.895	89.45%
2013_Typhoon _Yolanda	0,912	0,92	0,913	92,02%
2014_Iceland Volcano	0.908	0.908	0.908	90.82%
2015_Nepal_ Earthquake	0.748	0.751	0.749	75.05%

C. Effectiveness of Feature Selection by Information Gain

Feature selection is the process of selecting a subset of relevant and consistent features for use in model construction to reduce training time and to improve model performance. The performance of feature selection method by three different classifiers are shown in Table VI.

In the feature selection process, this system chose the only top ranked 300 features for all datasets. Table VI shows the results of feature reduction. According to the results, this work can eliminate at least 90% of original extracted features set.

TABLE VI. PERFORMANCE OF THE FEATURE SELECTION ANALYSIS ON DIFFERENT DATASETS

Dataset	**Feature Set**		**Accuracy of Classifiers**		
	Original Features	***Reduced Features***	***LibLinear (Accuracy)***	***SMO (Accuracy)***	***Random Forest (Accuracy)***
2013_Australia _bushfire	6727	300	87.28%	88.28%	85.45%
2013_Typhoon _Yolanda	11048	300	92.71%	93.31%	92.32%
2014_Iceland Volcano	4374	300	92.71%	93.31%	92.32%
2015_Nepal_ Earthquake	37471	300	76.90%	76.43%	72..13%

The accuracy of the reduced feature subset and the original feature set is almost equal or higher in most cases and can decrease up to 2% in only 2013_Australia_bushfire dataset.

Although the classification performance of the SMO classifier is as high as LibLinear, it takes so much longer time than the LibLinear for large dataset such as 2015_Nepal_ Earthquake. So this system used the LibLinear as learning model for tweets annotation.

D. Effectiveness of Annotation

In this section, the validation of this system on a real disaster study by classifying the data of Myanmar earthquake collected by Twitter API. The 6.8 magnitude earthquake that struck Myanmar on August 24, 2016 is among the strongest in recent Myanmar history. The earthquake was clearly perceived in all Central and Northern Myanmar and it caused 4 deaths and several damage to the Pagodas of the area of Bagan. This dataset is crawled for a three days period from August 24 to 26, 2016 by using the hashtags (#Myanmar, #Bagan, #earthquake, #Myanmarearthquake). And then it was randomly selected 1,800 tweets and was manually annotated based on the available news media in Myanmar such as Myanmar Times, The Global New Light of Myanmar and The Mirror.

Based on cross domain classification where we train the classifier on one dataset and test on another dataset, the experimental results using 2013_Typhoon Yolanda and 2015_Nepal Earthquake as training data and Myanmar Earthquake as test data confirmed the expected classification of this work. Myanmar_Earthquake_2016 was successfully annotated with predefined three labels at 75% accuracy on average which is pretty high.

E. Finding and Discussion

As mentioned above this system used four datasets for training and testing using 10 fold cross validation technique for evaluating classifier models and feature selection models based on different classifiers. Another new dataset for testing again for overall performance of the proposed system. According to the initial experimental results of classifiers, Naïve Bayes, SMO and Random Forest were very sensitive to large number of features and took longer time to build model. Among them, Naïve Bayes is the worst performance in accuracy. LibLinear is always faster than other classifiers.

The results of three feature extraction methods, the proposed method always outperforms the other two methods. Therefore, the proposed feature extraction model with LibLinear classifier was chosen for further annotation process for categorizing the tweets into specific frequently found information type such as infrastructure damage, dead and injuries, etc.

This system is considered to use the feature selection method to reduce irrelevant, redundant, and noisy features in text data. We chose the information gain theory based method for feature selection and selected top 300 features based on their rank.

VI. CONCLUSION

Social media mining for disaster response and coordination has been receiving an increasing level of attention from the research community. It is still necessary to develop automated mechanisms to find critical and actionable information on Social Media in real-time. The proposed system combines effective feature extraction using NLP and machine learning approach to obtain the annotated datasets to improve disaster response efforts. Expanded disaster lexicon is also used to

extract the relevant disaster related lexical features for annotation. The proposed feature extraction method significantly outperforms the standard bag of word model and neural word embedding model. By using LibLinear classifier based on the proposed method, this system successfully annotated the Myanmar Earthquake data at 75% accuracy on average. In future, we will investigate the specific variation of terms over different disasters to perform annotation on all disasters. We hope to formalize disaster lexicon in more detail to improve accuracy. And then we will continue to automatically annotate the informative tweets into more specific information types that are frequently found in natural disasters.

REFERENCES

[1] A. Gupta, P. Kumaraguru, "Credibility Ranking of Tweets during High Impact Events", PSOSM'12, 2012.

[2] A. Gupta, P. Kumaraguru, C. Castillo and P. Meier, "TweetCred: Real-time credibility assessment of content on Twitter", In Proc. Of SocInfo. Springer, 2014, pp. 228–243.

[3] A. Stavrianou, C. Brun, T. Silander, C. Roux, "NLP-based Feature Extraction for Automated Tweet Classification", Interactions between Data Mining and Natural Language Processing: 145.

[4] B. E. Parilla-Ferrer, L. Fernandez and J. T. Ballena, "Automatic Classification of Disaster-Related Tweets", International conference on Innovative Engineering Technologies (ICIET'2014), Bangkok (Thailand), 2014.

[5] A. Olteanu, C. Castillo, F. Diaz and S. Vieweg, "CrisisLex: A Lexicon for Collecting and Filtering Microblogged Communications in Crises", Association for the Advancement of Artificial Intelligence (www.aaai.org), 2014.

[6] C. Cardie and J. Wilkerson, "Introduction: Text Annotation for Political Science Research", Journal of Information Technology & Politics, 2008.

[7] C. C. Aggarwal and C.-X. Zhai, "Mining Text Data, Springer", 2012.

[8] C. Castillo, M. Mendoza and B. Poblete, "Information Credibility on Twitter", International World Wide Web Conference Committee (IW3C2), Hyderabad, India, 2011.

[9] Deeplearning4j Development Team. Deeplearning4j: Open-source distributed deep learning for the JVM, Apache Software Foundation License 2.0. http://deeplearning4j.org

[10] E. Frank, Mark A. Hall, and Ian H. Witten (2016). The WEKA Workbench. Online Appendix for "Data mining: Practical machine learning tools and techniques", Morgan Kaufmann, Fourth Edition, 2016.

[11] K. Gimpel et al., "Part-of-speech tagging for Twitter: Annotation, features, and experiments", In Proceedings of the Annual Meeting of the Association for Computational Linguistics, companion volume, Portland, June 2011.

[12] M. Imran, S. Elbassuoni, C. Castillo, F. Diaz and P. Meier, "Extracting Information Nuggets from Disaster Related Messages in Social Media", 10th International ISCRAM Conference–Baden-Baden, Germany, 2013.

[13] M. Mohammad, S. Kiritchenko, X. Zhu, "NRC-Canada: Building the state-of-the-art in sentiment analysis of tweets", National Research Council Canada, 2013.

[14] M. Imran, C. Castillo, J. Lucas, P. Meier and S. Vieweg, "AIDR: Artificial intelligence for disaster response", In Proc. of WWW (companion). IW3C2, 2014, pp. 159–162.

[15] N. Antoniou and M. Ciaramicoli, "Social media in the disaster cycle useful tools or mass distraction?", International Astronautical Congress. Beijing, 2013.

[16] R.E Fan, K.W Chang, C.J Hsieh, X.R Wang and C.J Lin, "LIBLINEAR: A Library for Large Linear Classification", Journal of Machine Learning Research 9, 2008, pp. 1871-1874.

[17] S. Kumar, "Social Media Analytics for Crisis Response", Ph.D. thesis, Arizona State University, 2015.

[18] https://socialpilot.co/blog/125-amazing-social-media-statistics-know-2016/, 2 Feb 2017.