



Target Oriented Tweets Monitoring System during Natural Disasters

Si Si Mar Win¹, Than Nwe Aung²

¹ *University of Computer Studies, Mandalay,
Patheingyi
Mandalay 1001, Myanmar
sisimarwin@gmail.com*

² *University of Computer Studies, Mandalay,
Patheingyi
Mandalay 1001, Myanmar
mdytina@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 is 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 monitoring 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 (BOW) models. This system automatically annotated the Myanmar_Earthquake_2016 dataset at 75% accuracy on average.

Keywords: Twitter; NLP; LibLinear, BOW.

1. 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 chaos²².

Among the popular social media platforms, Twitter becomes an important social media to get access to

breaking news and timely information during natural disasters. 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 day²⁸.

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 credible¹. There is still needed to develop the technologies for filtering and retrieving the informational or credible tweets automatically during disasters.

There are at least three problems involved in developing a tweets monitoring 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 messages²⁵. 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. We consider the problems as feature extraction, two layers 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:

- (i) 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.
- (ii) 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.

- (iii) 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 2 presents the closely related work to this paper. Section 3 explains the methodology that we used in collecting, preprocessing, feature extraction, disaster lexicon creation, and classification scheme used for annotating the tweets. Section 4 describes the architecture of the proposed system. Section 5 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 6 summarizes the results from our analysis and highlights the implications of our results. In this section, we also describe the future work of the proposed system.

2. Related Work

Since much of the data from the social media sites will be text -based data, the process of preparing and analyzing the data will focus on principles of preparing text data for analysis²³. The analysis of social media data is heavily rely on the ability to analyze text data. However, there are some unique considerations in the analysis of social media data that make it different than a normal text mining analysis.

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 emotion analysis, sentiment analysis, information diffusion, information extraction, topic detection, credibility prediction, event identification and other types of social media irrelevant content analysis. Feature extraction in text mining and some previous approaches to critical tweets classification are closely related to the proposed work.

In text mining, feature extraction is the most important task for text categorization. There are various feature extraction models in content analysis such as the vector space model, the probability model, Term Frequency Inverse Document Frequency (TFIDF) and Latent Dirichlet Allocation (LDA)²¹. Bag of words based on TF model is a well-known method to evaluate how important is a word in a document or tweet content. It

does not take any syntax and semantic features such as POS tag and lexical based features.

In recent years, a neural network implementation of distributed word vectors in deep learning becomes powerful semantic feature extraction tool and it has been used in many applications especially word prediction and translation. Zharmagambetov and Pak applied Word2vec to extract main features in their sentiment analysis⁴.

Sun et al. applied Feedforward Neural Network Language Model (NNLM) in deep learning to extract semantic code of each word for emotion analysis²⁷.

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 tweets⁹. In their research, they focused on the level of credibility of every trend on Twitter. Gupta et al.¹ used six types of features such as 3 tweet meta-data features, 10 tweet content features, at least 3 user based features, 4 network features, 5 linguistic features and 2 external resource features for credibility analysis. Besides, hashtags have been effectively utilized as critical features for various tasks of text or social media analysis, including tweet classification⁵.

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 Ref. 9. To predict the credibility levels on Twitter events, they used J48 decision tree classifier.

Moreover, the information detection and extraction system for microblog posts was described by Imran et al.¹⁸. 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 timeline².

According to the literatures, supervised machine learning algorithms have been applied by most of the researchers to detect and classify the content in online social media. Naive Bayes (NB) and Support Vector Machine (SVM) are used for tweets classification in Ref. 6.

In contrast, the proposed system uses a Library for Large Linear Classification (LibLinear) classifier to annotate the tweets from Twitter during natural disasters. 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 the specific disaster related lexicon creation based on the top natural disasters types that often cause around the world.

3. Methodology

Tweet contents analysis in social media mining 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. Although 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 covers all possible disaster situations. Data annotation has always been carried out by human annotator or crowd source workers in social media analysis. However, manual annotation of entire dataset in real time is impractical to huge amount of social media data.

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 four main functions of the proposed system: Data Collection, Feature Extraction, Lexicon Creation and Ground Truth Labeling or classification for annotation learning. The proposed system focuses on attaching only contents of the tweets.

3.1. 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.

3.2. 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 stem⁷. 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.

3.3. Feature Extraction

Feature extraction is the transformation of arbitrary data such as images or text into numerical features usable for classification. In social media text analysis, features can be generally categorized into three main types according to their usage: network-based or propagation-based features which focus on the characteristics related to the retweets of a message such as number of retweets, number of mentions etc.^{9, 26}, user or source based features which are extracted based on the characteristics of users who post the tweets and content or message based features. Content based features are extracted based on only contents of the tweets. Feature extraction

process of this work is based on content based features only.

We can extract not only the term from the corpus of tweets but also many salient features from tweet content. Features extracted from tweet content are word NGram features, semantic features such as part of speech tag features, brown dictionary features, lexical features and linguistic features such as number of swear words, number of negation, number of positive emotion words, number of negative emotion words, number of elongated words, number of personal pronouns, number of hashtags, number of URLs etc.

In this work, we exploited the content based features only. 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. Therefore, feature extraction 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 NLP¹⁶.

This process receives the tweets from preprocessing step, it extracts the features by using ARK POS tagger and different lexicons. These are described in Table 1.

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 operations¹⁰.

Table 1. Features used in the proposed system

Features	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
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

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.

The proposed system also uses Bag of Words (BOW) model for feature vector creation and comparison purposes. Bag of Words model creates a numeric feature vector with word counts for each data instance (tweet). The count can be either absolute, binary (contains or does not contain) or sublinear (logarithm of the term frequency). Bag of words model can be used for predictive modelling.

In this system, we used BOW with unigram and bigram combination. The count of each word or term is considered in binary nature.

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.

3.4. 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 with equal number of Informative and Not-Informative tweets. This lexicon creation process follows the method of Ref. 3.

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.

3.5. Classification of Tweets

This function automatically classifies the information in tweets. To perform this task, the proposed system trained a LibLinear²⁴ 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 a Library for Support Vector Machines (LibSVM) takes several hours for same task.

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)¹⁹ and CrisisLexT26³. This system uses datasets in English language only.

4. 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 data⁸. 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

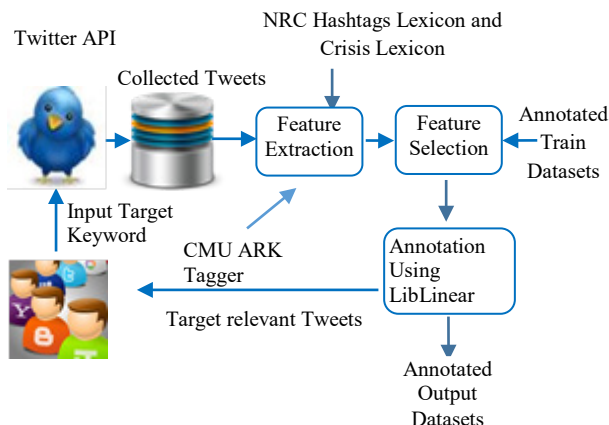


Fig. 1. Architecture of the Proposed Tweets Monitoring System.

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, Syntactic feature such as POS features, Lexical features such as Sentiment Lexicon features using NRC Hashtag Sentiment Lexicon²⁰ and Natural disaster lexicon which is extended from the Olteanu et al.³ and the other Twitter specific 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.

5. 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:

- (i) Neural word embeddings and others features
- (ii) Standard BOW model with Unigram, Bigram and other features
- (iii) Unigram, Bigram plus Hashtags and Crisis Lexicons 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 1. This system extracted the new lexical features such as natural disaster related terms using extended crisis lexicon or disaster related lexicon based on CrisisLex.

5.1. 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 2. 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.

Table 2. Datasets used in the proposed system

Different Natural Disaster Related Datasets			
Dataset	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



5.2. 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, F1-score and recall values for classification tasks using different features on different datasets by LibLinear classifier are shown in Table 3, Table 4 and Table 5.

We also tested the extracted feature set on four different classifiers such as Random Forest, which is an ensemble learning method, Sequential minimal optimization (SMO) which is a fast training algorithm for Support Vector Machine (SVM), Naïve Bayes which is a probabilistic estimator and our LibLinear classifier. These methods are well known in the text classification process.

Table 3. Classification Results using Neural Word Embeddings Model.

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 4. Classification Results using Standard Bag of Words Model

Standard 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 5. Classification Results using Proposed Feature Extraction Model

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%

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 work uses a well-known WEKA¹¹ machine learning tools for implementation of Random Forest, Naïve Bayes, SMO, LibLinear and Information Gain based feature selection methods.

5.3. Effectiveness of Feature Selection by Information Gain

With the presence of a large number of features in text categorization, a learning model tends to overfit and degrade in their performance¹⁴. Therefore, feature Selection is a key technology in text classification. It 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. Information Gain based method¹³ is the effective algorithm for text classification. Several research have been tried to improve this method for text classification^{12, 15, 17}.

The proposed system used this method for our feature selection process. This allows us to select a desired number of superior features at very low cost. The original extracted features of the proposed system are 6727 for 2013_Australia_bushfire, 11048 for 2013_Typhoon_Yolanda, 4374 for 2014_Iceland_Volcano, and 37471 for 2015_Nepal_Earthquake respectively. The performance of feature selection method based on four different classifiers is shown in Table 6. In the feature selection process, this system chose the only top ranked 300 features for all datasets. Table 6, Table 7, Table 8 and Table 9 show the comparison results of reduced 300 feature subset with original feature set. According to the results, this work can eliminate at least 90% of original extracted features set on all datasets.

Table 6. Performance of the feature selection analysis on different datasets by SMO classifier.

Performance of Feature Selection by SMO				
Dataset	Accuracy (%)		Time (Seconds)	
	Original	Reduced	Original	Reduced
2013_Australia_bushfire	88.08%	88.28%	0.38	0.16
2013_Typhoon_Yolanda	91.92%	93.31%	0.43	0.13
2014_Iceland_Volcano	92.27%	93.31%	0.11	0.04
2015_Nepal_Earthquake	74.63%	76.43%	82.12	9.59



Table 7. Performance of the feature selection analysis on different datasets by Random Forest classifier.

Performance of Feature Selection by Random Forest				
Dataset	Accuracy (%)		Time (Seconds)	
	Original	Reduced	Original	Reduced
2013_Australia_bushfire	85.45	85.45%	0.62	0.15
2013_Typhoon_Yolanda	90.72	92.32%	1.39	0.19
2014_Iceland_Volcano	86.72%	92.32%	0.32	0.08
2015_Nepal_Earthquake	71.57%	72.13%	36.2	1.59

The classification 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 is decided to use the LibLinear as learning model for tweets annotation.

Table 8. Performance of the feature selection analysis on different datasets by LibLinear classifier.

Performance of Feature Selection by LibLinear				
Dataset	Accuracy (%)		Time (Seconds)	
	Original	Reduced	Original	Reduced
2013_Australia_bushfire	89.49	87.27	0.14	0.03
2013_Typhoon_Yolanda	92.02	92.71	0.1	0.05
2014_Iceland_Volcano	90.82	92.71	0.04	0.03
2015_Nepal_Earthquake	75.05	76.90	1.1	0.51

Table 9. Performance of the feature selection analysis on different datasets by Naïve Bayes classifier.

Performance of Feature Selection by Naïve Bayes				
Dataset	Accuracy (%)		Time (Seconds)	
	Original	Reduced	Original	Reduced
2013_Australia_bushfire	86.26	87.07	0.67	0.08
2013_Typhoon_Yolanda	83.93	83.83	2.18	0.07
2014_Iceland_Volcano	82.85	83.83	0.32	0.04
2015_Nepal_Earthquake	67.48	66.95	35.38	0.27

5.4. 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 such as #Myanmar, #Bagan, #earthquake and #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.

5.5. Finding and Discussion

As mentioned above, this system used four datasets for training and testing using 10 fold cross validation technique for evaluating four classifier models and feature selection model. 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 is chosen for further annotation process for categorizing the tweets into specific frequently found information type such as infrastructure damage, dead and injuries, etc.

The two most important factors in a real-time system are correctness and response time. Therefore, our evaluation is based on accuracy and time taken to build the model for annotation learning.

Based on the evaluation results, this system is considered to use the feature selection method to reduce irrelevant, redundant, and noisy features in text data and LibLinear to accurate classification. We chose the

information gain theory based method for feature selection and selected top 300 features based on their rank.

6. 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 cross domain classification 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, in *Proc. 1st Workshop on Privacy and Security in Online Social Media*, ACM (2012) pp. 2.
2. A. Gupta, P. Kumaraguru, C. Castillo and P. Meier, TweetCred: Real time credibility assessment of content on Twitter, in *Proc. 6th Int. Conf. on Social Informatics (SocInfo)*. Springer, (Barcelona, Spain, 2014) pp. 228–243.
3. A. Olteanu, C. Castillo, F. Diaz and S. Vieweg, CrisisLex: A Lexicon for Collecting and Filtering Microblogged Communications in Crises, in *Proc. ICWSM*, (2014).
4. A. S. Zharmagambetov and A. A. Pak., Sentiment analysis of a document using deep learning approach and decision trees, in *Electronics Computer and Computation (ICECCO)*, 12th. Int. Conf. IEEE (2015), pp. 1–4.
5. A. Stavrianou, C. Brun, T. Silander and C. Roux, NLP-based Feature Extraction for Automated Tweet Classification, Interactions between Data Mining and Natural Language Processing: 145.
6. B. E. Parilla-Ferrer, L. Fernandez and J. T. Ballena, Automatic Classification of Disaster-Related Tweets, *International conference on Innovative Engineering Technologies (ICIET)*, (Bangkok, Thailand, 2014).
7. C. C. Aggarwal and C.-X. Zhai, Mining Text Data, (Springer, 2012).
8. C. Cardie and J. Wilkerson, Introduction: Text Annotation for Political Science Research, *Journal of Information Technology & Politics* (2008).
9. C. Castillo, M. Mendoza and B. Poblete, Information Credibility on Twitter, *International World Wide Web Conference Committee (IW3C2)*, (Hyderabad, India, 2011).
10. Deeplearning4j Development Team. Deeplearning4j: Open-source distributed deep learning for the JVM, Apache Software Foundation License 2.0. <http://deeplearning4j.org>.
11. E. Frank, M. A. Hall and I. H. Witten, *The WEKA Workbench. Data mining: Practical machine learning tools and techniques*, 4th edn. (Morgan Kaufmann, 2016).
12. G. Wu and J. Xu., Optimized approach of feature selection based on information gain., in *Computer Science and Mechanical Automation (CSMA)*, Int. Conf. on IEEE, (2015) pp. 157–161.
13. H. Peng, F. Long, and C. Ding, Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy, *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2005) pp. 1226–1238.
14. J. Tang, S. Alelyani, and H. Liu, Feature selection for classification: A review, *Data Classification: Algorithms and Applications* (2014) pp.37.
15. J. Xu, and H. Jiang, An improved information gain feature selection algorithm for SVM text classifier, in *Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*, International Conference on IEEE, (2015), pp. 273–276.
16. K. Gimpel, N. Schneider, B. O'Connor, D. Das, D. Mills, J. Eisenstein, M. Heilman, D. Yogatama, J. Flanigan, and N. A. Smith, Part-of-speech tagging for Twitter: Annotation, features, and experiments, in *Proceedings of 49th Annual Meeting of the Association for Computational Linguistics*, vol. 2, (Portland, June 2011) pp. 42–47.
17. L. Z. Feng, W. L. Zuo, and Y. W. Wang, Improved Comprehensive Measurement Feature Selection Method for Text Categorization, in *Network and Information Systems for Computers (ICNISC)*, 2015 International Conference on IEEE, (2015) pp. 125–128.
18. 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).
19. 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.
20. M. Mohammad, S. Kiritchenko, X. Zhu, NRC-Canada: Building the state-of-the-art in sentiment analysis of tweets, *National Research Council* (Canada, 2013).
21. M.S. Cha, S.Y. Kim, J.H. Ha, M.J. Lee, Y.J. Choi and K.A. Sohn, Topic Model based Approach for Improved Indexing in Content based Document Retrieval, *International Journal of Networked and Distributed Computing*, 4(1) (2016) pp. 55–64.
22. N. Antoniou and M. Ciaramicoli, Social media in the disaster cycle useful tools or mass distraction?, *International Astronautical Congress*, (Beijing, 2013).
23. R. C. Mosley Jr, Social media analytics: Data mining applied to insurance Twitter posts, *In Casualty Actuarial Society E-Forum*, vol. 2, (2012) pp. 1.
24. R. E. Fan, K.W Chang, C.J Hsieh, X.R Wang and C.J Lin, LIBLINEAR: A Library for Large Linear Classification, *J. Machine Learning Research* 9 (2008) pp. 1871–1874.
25. S. Kumar, Social Media Analytics for Crisis Response, Ph.D. thesis, (Arizona State University, 2015).
26. V. Qazvinian, E. Rosengren, D. R. Radev, and Q. Mei, Rumor has it: Identifying misinformation in microblogs, in *Proc. Conf. on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, (2011) pp. 1589–1599.
27. X. Sun, C. Sun, C. Quan, F. Ren, F. Tian and K. Wang, Fine-Grained Emotion Analysis Based on Mixed Model for Product Review, *International Journal of Networked and Distributed Computing*, 5(1) (2017) pp.1–11.
28. <https://socialpilot.co/blog/125-amazing-social-media-statistics-know-2016/>, (2 Feb 2017).