EXTRACTING USEFUL INFORMATION FROM SOCIAL MEDIA DURING DISASTER EVENTS

Venkata Kishore Neppalli

Dissertation Prepared for the Degree of

DOCTOR OF PHILOSOPHY

UNIVERSITY OF NORTH TEXAS

May 2017

APPROVED:

Cornelia Caragea, Major Professor
Armin Mikler, Committee Member
Paul Tarau, Committee Member
Yan Huang, Committee Member
Barrett Bryant, Chair of the Department
of Computer Science and Engineering
Costas Tsatsoulis, Dean of the College of
Engineering
Victor Prybutok, Vice Provost of the
Toulouse Graduate School

Neppalli, Venkata Kishore. *Extracting Useful Information from Social Media during Disaster Events*. Doctor of Philosophy (Computer Science), May 2017, 82 pp., 18 tables, 14 figures, 74 numbered references.

In recent years, social media platforms such as Twitter and Facebook have emerged as effective tools for broadcasting messages worldwide during disaster events. With millions of messages posted through these services during such events, it has become imperative to identify valuable information that can help the emergency responders to develop effective relief efforts and aid victims. Many studies implied that the role of social media during disasters is invaluable and can be incorporated into emergency decision-making process. However, due to the "big data" nature of social media, it is very labor-intensive to employ human resources to sift through social media posts and categorize/classify them as useful information. Hence, there is a growing need for machine intelligence to automate the process of extracting useful information from the social media data during disaster events. This dissertation addresses the following questions: In a social media stream of messages, what is the useful information to be extracted that can help emergency response organizations to become more situationally aware during and following a disaster? What are the features (or patterns) that can contribute to automatically identifying messages that are useful during disasters? We explored a wide variety of features in conjunction with supervised learning algorithms to automatically identify messages that are useful during disaster events. The feature design includes sentiment features to extract the geo-mapped sentiment expressed in tweets, as well as tweet-content and user detail features to predict the likelihood of the information contained in a tweet to be quickly spread in the network. Further

experimentation is carried out to see how these features help in identifying the informative tweets and filter out those tweets that are conversational in nature.

ACKNOWLEDGMENTS

First of all, I am incredibly grateful to my advisor, Dr. Cornelia Caragea, for her visionary support and unwavering guidance throughout the course of this work. I could not have imagined a better advisor and mentor for my Ph.D. studies. Besides, from being a superb researcher, she has been a great mentor and guided me in the paths of righteousness throughout the years. Thank You so much, Dr. Caragea, for your support, encouragement, and guidance.

I express my sincere gratitude for the members of my Ph.D. committee, Dr. Armin Mikler, Dr. Paul Tarau, and Dr. Yan Huang. Their feedback was precious and helped me to improve my dissertation.

I would like to thank the National Science Foundation for the support from the grants IIS 1353418 and IIS 1526542 to Dr. Cornelia Caragea, as well as our collaborators, Dr. Andrea Tapia and Dr. Anna Squicciarini. The views and conclusions contained in this document should not be interpreted as necessarily representing the official policies, either express or implied, of the National Science Foundation.

I am also thankful to my colleagues for their various forms of support during my graduate studies. I would like to thank Dr. Bharath Dandala and many friends for their help during these years, their support and care helped me a lot during my Ph.D. I greatly appreciate their valuable friendship.

Finally and most importantly, none of this would have been possible without the love, understanding, patience and support of my family. I would like to express my heartfelt gratitude to my parents, Prasad Neppalli and Nagamani Neppalli, my wife Mounika Neppalli, and my brother Avinash Neppalli. I would like to add my gratitude for the incredible support from my uncles Krishna Rao Chappidi and Sudhakar Chappidi. This dissertation is dedicated to my family.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

In this chapter, we discuss the background and motivation of this field of research, our research goals and contributions.

1.1. Background and Motivation

In response to increased online public engagement through micro-blogging services and the emergence of digital volunteers, professional emergency responders have sought to better understand how they too can use social media to communicate with the public and collect intelligence [9, 28, 16, 63, 53]. Many emergency decision makers see the data produced through crowdsourcing as ubiquitous, rapid and accessible - with the potential to contribute to situational awareness [72]. As public social media use in crisis has increased, emergency responders have started to take notice of the way citizens engaged in social media and the information exchanges that took place there [19]. Consequently, responders began to consider if social media might be a useful tool for their practice. Research revealed that social media could be used to distribute information quickly to a wide-spread audience [26] and to engage more directly in a two-way conversation with members of the public [16, 28, 45]. According to Starbird et al. [56], social media data identified as coming from local bystanders during a disaster can be crucial to emergency responders. Most of the social media data surrounding a disaster are derivative in nature: information in the form of re-posts or pointers to information available elsewhere and only a small subset of the data comes from locally affected populations in the form of citizen reports [55]. Starbird et al. [55] assert that bystanders "on the ground are uniquely positioned to share information that may not yet be available elsewhere in the information space and may have knowledge about geographic or cultural features of the affected area that could be useful to those responding from outside the area."

Much has been written concerning the value of using messaging and micro-blogged data from crowds of non-professional participants during disasters. Often referred to as micro-blogging, the practice of average citizens reporting on activities “on-the-ground” during a disaster is seen as increasingly valuable [43, 61, 68]. The information that the public produced appeared to be useful, as researchers showed that it could contribute to situational awareness during a crisis event [4, 22]. Vieweg et al. [72] found that retweet ed tweets are likely to contain information that contributes to situational awareness and are actionable compared with non-retweeted tweets. In addition to the information which creates awareness to responders, people also post information related to relief efforts (such as offering shelters, donations, and food) during the disasters, for which the target consumers are the victims who need aid.

Despite the evidence of strong value to those experiencing the disaster and those seeking information concerning the disaster, there is still very little uptake of social media data by large-scale, disaster response organizations [67, 65]. Response organizations operate under conditions of extreme uncertainty. The uncertainty has many sources: the sporadic nature of emergencies, the lack of warning associated with some forms of emergencies, and the wide array of responders who may or may not respond to any one emergency. Moreover, along with informative messages, there are messages that do not convey any useful information (e.g., conversational in nature) and must be filtered out to arrive at the signal of “good data.” This increases the necessity for identifying appropriate information from the streaming data, which could make substantial improvements in the response process. These data are not yet seen as fully actionable because of the inability to sort and categorize the data into useful types. One strong reason for this is that the size of the messages streaming is tremendous. For example, in the survey by the Pew Research Center[1], it is stated that more than 20 million tweets were posted during the five-day interval of Hurricane Sandy. Due to this sheer amount of data, it is extremely difficult and time-consuming to manually sift through these data and categorize them into useful types. Therefore, there is growing need for machine intelligence to automate the process of categorizing the social media data into useful types.

[1] http://www.pewresearch.org/fact-tank/2013/10/28/twitter-served-as-a-lifeline-of-information-during-hurricane-sandy/

Machine learning [34] offers approaches to construct predictive models in the applications where training data are available.

However, the quality of the learned models highly depends on the choice of features that are extracted to capture the knowledge of distinctions in the given data. Through this research, we present a wide exploration of features towards solving the problem of categorizing social media data into useful types in the context of disasters. We specifically focus on designing features for three classification tasks based on these three aspects: sentiment (which offers insights about how people are feeling during the disaster), retweetability (which helps to inform on how to reach more people in a fastest way) and informativeness (which helps to keep people "on the ground" and emergency responders informed with valuable insights about the disaster).

1.2. Research Questions

Our research questions are: *In a social media stream of messages, what is the useful information to be extracted that can help emergency response organizations to become more situationally aware during or following a disaster?* and *What are the features (or patterns) that can help to automatically identify messages that are useful during disasters?*

A primary distinguishing factor between informational and conversational tweets is discrepancies that reveal formality. Aspects of formality include correct grammar, lack of slang, lack of swear words, etc. A formal tweet is likely informative since many credible sources of information will most likely structure their tweets to look as professional as possible. Moreover, tweets expressing sentiment are likely to be conversational tweets, because they express opinions/feelings of the user rather than conveying some useful information. For example, *"I hope this storm is TERRIBLEEEEEEE!!!!!! Lol forreal. Ill sleeep perfect."* This tweet has a positive sentiment expressing a happy feeling of an author and is not informational in nature. Another crucial factor that helps in identifying information is: retweetability of a tweet; it reveals the importance of the tweet with useful information. For example, "*# Sandy East coast, search for open shelters by texting: SHELTER + a zip code to 43362 (4FEMA)*", which has ≈ 3900 retweets. This tweet conveys information for those who seek shelters. A tweet with necessary information during disasters will typically get retweet ed. Along these lines, our research questions are refined as the following:

- *What are the features that could help identify the sentiment expressed in a tweet?*
- *What are the features that could help identify how likely a tweet can be retweeted?*
- *How could the features related to sentiment and retweetability of tweets help to model the task of identifying informative tweets?*

1.3. Dissertation Outline

In this section, we describe the brief outline of the dissertation. Each chapter presents the approaches to address respective questions raised in the above discussion. The goal of this research is to develop methods to automatically predict the messages that are useful for the emergency responders and the bystanders. We chose Twitter (one of the most popular social media) to carry out our research and the tweets posted during disasters. Most of the work in this dissertation has been published in conference proceedings and journals.

Chapter 2 presents our approach to understanding the general mood during Hurricane Sandy; we perform a geo-mapped sentiment analysis, where we first identify all geo-tagged tweets in our collection and label each of these tweets using our disaster-trained classifiers. We then associate the sentiments of tweets with their geo-locations. We show how users' sentiments change according not only to the locations of the users, but also based on the relative distance from the disaster. In this work, we offer a proof of concept. Using Twitter data from Hurricane Sandy, we identify the sentiment of tweets and then measure the distance of each categorized tweet from the epicenter of the hurricane. We find that extracting sentiments during a disaster may help responders develop stronger situational awareness of the disaster zone itself. We also performed a comparison with one of the previous works related to sentiment analysis and showed that our features are better compared with this work. Our approach can help increase situational awareness and can "visually" inform emergency response organizations about the geographical regions that are most affected by a disaster. We also study the effect of emotional divergence on retweetability during Hurricane Sandy and how the emotional divergence can affect information spread.

Chapter 3 addresses the problem of retweetability prediction and a set of analyses on retweet ed tweets during Hurricane Sandy and Hurricane Patricia to determine several aspects affecting retweetability. We explore a wide range of features including features extracted from the tweets' content and user account information and use them in conjunction with machine learning classifiers to predict the tweets' retweetability during the hurricane. We show that the classifiers trained on these features outperform those trained using the "bag of words" approach. We perform feature selection to understand what features are informative in identifying the retweetability of a tweet during disaster events.

Chapter 4 presents our approach with a wide range of feature exploration to identify informative messages (or tweets) from those that are not-informative in nature (i.e., pertaining to user feelings, informal communication or casual conversations). Our approach is based on a combination of "bag-of-words" features, which are typically used for text classification, and features that are extracted from tweets' content (e.g., URLs, hashtags, emoticons, slang), user details (such as number of friends, number of followers) and polarity clues (such as positive words, negative words). Furthermore, we study the extent to which our features can be generalized across different disaster types (e.g., natural and non-natural disasters) by developing models trained on one disaster type (such as natural disasters) and evaluating them on another disaster type (such as non-natural disasters).

Chapter 5 summarizes the findings of this research and provides a summary of contributions and possible future directions.

1.3.1. Published Work

- Chapter 2 on *sentiment analysis during hurricane Sandy in emergency response* has been published in *Proceedings of Information Systems for Crisis Response and Management (ISCRAM), 2014* [6]. This paper contains experiments on sentiment classification using Hurricane Sandy data and the association of sentiments predicted for tweets with geo-location on a geographical map. A journal article containing the work in this chapter has been published recently in *International Journal for*

Disaster Risk Reduction, 2017 in Vol. 21 (213-222) [39]. In the extended work, we augment our contributions to geo-mapped sentiment analysis in disaster events by studying the effect of emotional divergence on retweetability during Hurricane Sandy and how the emotional divergence can affect information spread during the disaster.

- Chapter 3 on *predicting retweetability during hurricane disasters* has been published in conference proceedings of *ISCRAM'16* [37], which contains experiments on retweetability prediction using the Hurricane Sandy dataset. A journal article containing the work in this chapter is published in the *International Journal of Information Systems for Crisis Response and Management (IJISCRAM), 2017* [40]. In the extended work, we performed the experiments on one more dataset, i.e., Hurricane Patricia; we modified user feature details with normalization, and compared with two previous works.
- Chapter 4 on *identifying informative tweets during disasters* is to be submitted. In this work, we experiment with a wide range of feature sets (namely user features, polarity features) and investigate the correlation among emotional divergence, informativeness and the retweet phenomenon.

1.3.2. Other published work (not included in this work)

- *Twitter mining for disaster response: A domain adaptation approach*, in *ISCRAM'15 (joint work with Doina Caragea, Nicolais Guevara, Hongmin Li, Nic Herndon, Andrea Tapia, and Anna Squicciarini)* [30]. Microblogging data such as Twitter data contains valuable information that has the potential to help improve the speed, quality, and efficiency of disaster response. Machine learning can help with this by prioritizing the tweets with respect to various classification criteria. However, supervised learning algorithms require labeled data to learn accurate classifiers. Unfortunately, for a new disaster, labeled tweets are not easily available, while they are usually available for previous disasters. Furthermore, unlabeled tweets from the current disaster are accumulating fast. We study the usefulness of labeled data from a prior source disaster, together with unlabeled data from the current target disaster to learn domain adaptation classifiers for the target. Experimental results suggest that, for some tasks, source data itself can be useful for classifying target data. However, for tasks specific to a particular disaster, domain adaptation approaches that use target unlabeled data in addition to source labeled data are superior.

- *MetaSeer.STEM: Towards Automating Meta-Analyses*, in *Innovative Applications in Artificial Intelligence (IAAI) 2016 (joint work with Robin Mayes, Kim Nimon, Fred Oswald)* [38]. Meta-analysis is a principled statistical approach for summarizing quantitative information reported across studies within a research domain of interest. Although the results of meta-analyses can be highly informative, the process of collecting and coding the data for a meta-analysis is often a labor-intensive effort fraught with the potential for human error and idiosyncrasy. This is due to the fact that researchers typically spend weeks poring over published journal articles, technical reports, book chapters and other materials to retrieve key data elements that are then manually coded for subsequent analyses (e.g., descriptive statistics, effect sizes, reliability estimates, demographics, and study conditions). In this paper, we propose a machine learning based system developed to support automated extraction of data pertinent to STEM education meta-analyses, including educational and human resource initiatives aimed at improving achievement, literacy and interest in the fields of science, technology, engineering, and mathematics.
- *Embracing Human Noise as Resilience Indicator: Twitter as Power Grid Correlate* in *Journal of Sustainable and Resilient Infrastructure, 2017, (joint work with Nick Lalone, Andrea H. Tapia, Christopher Zobel)* [27] The word resilience means many different things to many different disciplines and industries; measuring resilience is just as varied. Despite those differences, we find that there are typically two approaches to measuring resilience - technically dynamic, or the data produced by sensors attached to physical objects, and socially static, or those demographic indicators that represent a given geographic location. We find that this allows

resilience to represent before a disruption and examined post-disruption. During an event, there are few ways resilience, or even status of a population, can be accounted for. Through an analysis of tweets made during Hurricane Sandy and power outage data obtained after the event, we find that tweets that mention power, utility, or electricity were correlated. We believe this offers a proof of concept that social media data can be used to interpret critical infrastructure as well as now casting the events of an area. Despite this proof of concept, we conclude with a discussion of barriers to fully realizing this concept's potential.

1.4. Appendix

In this section, we describe the Twitter's data crawling APIs that we used to crawl the tweets for this work. Twitter provides access to most of the tweets through its Streaming and REST (REpresentational State Transfer) APIs. Streaming APIs offers continuous access to the real-time data that is flowing through Twitter and the REST API's gives us a programmatic access (such as adding new tweets, search a tweet). Streaming APIs are offered in three types, namely Public Streams (streams of public data that are posted rapidly on Twitter), User Streams (which confines the data collection to a specified user), and Site Streams (this is multi-user version of the User Streams and is a beta version). REST API provides a lot of methods to attain the data depending on tweet specific or user specific information (e.g., we can get the user details by querying the REST API with *user id*). I described the details of each API and the parameters used for our tasks.

Twitter Streaming API[2]: This API provides methods and parameters to access the publicly available tweets that are streaming. We particularly used the streaming URL "https://stream.twitter.com/1.1/statuses/filter.json" with "track" parameter set to our search terms to download the tweets from the stream. The "track" parameter facilitates the filtering of unnecessary tweets (not related to the search terms) from the stream. According to the information available from Twitter developers' website[3], it is mentioned

[2] https://dev.twitter.com/streaming/overview

[3] https://dev.twitter.com/streaming/overview/connecting

that as long as the connection between the client and the Streaming API is established and live, one can download as many tweets as possible from the public stream until the connection is lost. However, on the website, it is stated that too many connections in a short period limit the access to the stream, although the exact number of attempts that limit the access is not disclosed.

Twitter Search API[4]: This is a REST (REpresentational State Transfer) API which is different from the Streaming API. It provides several methods to download tweets based on the parameters. For example, if tweet ids are available, the total information about those tweets can be pulled from the Twitter database. We used the REST endpoint "`https://api.twitter.com/1.1/statues/lookup.json`" with "id" parameter. The "id" parameter can take up to 100 comma-separated ids. The rate limits[5] for Search API are - 900 per 15-min window for user level authentication and 300 per 15-min window for app-level authentication.

[4] https://dev.twitter.com/rest/public/search

[5] https://dev.twitter.com/rest/public/rate-limits

CHAPTER 2

SENTIMENT ANALYSIS DURING HURRICANE SANDY IN EMERGENCY RESPONSE

Sentiment analysis has been widely researched in the domain of online review sites with the aim of generating summarized opinions of users about different aspects of products. However, there has been little work focusing on identifying the polarity of sentiments expressed by users during disaster events. Identifying such sentiments from online social networking sites can help emergency responders understand the dynamics of the network, e.g., the main users' concerns, panics, and the emotional impacts of interactions among members. In this chapter, we perform a sentiment analysis of tweets posted on Twitter during the disastrous Hurricane Sandy and visualize online users' sentiments on a geographical map centered around the hurricane. We show how users' sentiments change according to not only to their locations, but also based on the distance from the disaster. In addition, we study how the divergence of sentiments in a tweet posted during the hurricane affects the tweet retweetability. We find that extracting sentiments during a disaster may help emergency responders develop stronger situational awareness of the disaster zone itself.

2.1. Introduction

In the field of disaster response, making social media data useful to emergency responders has been the single strongest research focus for the past several years [66]. In response to increased online public engagement and the emergence of digital volunteers, professional emergency responders have sought to better understand how they can use social media to collect intelligence [16]. Emergency decision makers see the data produced through crowdsourcing as ubiquitous, rapid and accessible, with the potential to contribute to situational awareness [72]. Starbird et al. [55] assert that bystanders "on the ground are uniquely positioned to share information that may not yet be available elsewhere in the information space and may have knowledge about geographic or cultural features of the affected area that could be useful to those responding from outside the area."

Despite the strong value to those experiencing the emergency and those seeking information concerning the emergency, responders are still hesitant to use social media data for several reasons [66]. One strong reason is insecurity and apprehension concerning the connection between the location of the disaster event and those tweeting about the disaster. Because of the nature of social media, contributors do not have to be bystanders. Responders interested in the wellbeing of physical bystanders seek methods of finding and measuring the concerns of those directly affected by a disaster. Analyzing social media data and extracting users' geo-mapped opinions and sentiments during a disaster can help emergency responders understand the dynamics of the network, e.g., the main users' concerns and panics, the emotional impacts of interactions among users, and the geographical regions that are most affected by the disaster. In addition, analyzing social media data can help obtain a holistic view about the general mood and the situation "on the ground."

2.1.1. Contributions and Organization

In this chapter, we aim to design accurate approaches to geo-mapped sentiment analysis during disaster events. More precisely, using Twitter data from Hurricane Sandy as a case study, we first develop models to identify the sentiment of tweets and then measure the distance of each categorized tweet from the epicenter of the hurricane. We show that users' sentiments change according to not only to the locations of the users, but also based on the relative distance from the disaster. We find that extracting sentiments during a disaster may help emergency responders develop stronger situational awareness of the disaster zone itself.

We further analyze the impact of the divergence of sentiments in a tweet on the likelihood of the tweet to be retweeted, which affects the information spread in Twitter (also called as retweeting). Understanding how the retweet function inside Twitter works can potentially shed light into the type of information being spread during disasters in large microblogging communities. Identifying elements of a message that make it more likely to be retweeted during a disaster can better inform emergency managers on how to reach the widest audience in the fastest way.

The rest of the chapter is organized as follows: Section 2.2 describes the sentiment classification followed by a geo-tagged sentiment analysis of tweets posted during Hurricane Sandy, in which sentiment classification of tweets is an important component. Section 2.3 describes an analysis on how the divergence of sentiments in a tweet is affecting tweets' retweetability. Section 2.4 concludes the chapter with a summary and discussion including related work.

2.2. Geo-Tagged Sentiment Analysis of Tweets from Hurricane Sandy

In this section, we first present the feature extraction for sentiment classification. We then describe experiments and results of this classification task on tweets from the Hurricane Sandy and finally, analyze the set of geo-tagged tweets, which are automatically labeled with their sentiment polarity by our best sentiment classifier.

2.2.1. Dataset

The data used in these experiments is collected from Twitter during the disastrous Hurricane Sandy using Twitter Streaming API discussed in Section 1.4. Specifically, the dataset contains 12,933,053 tweets crawled between 10-26-2012 and 11-12-2012. Among these tweets, 4,818,318 have links to external sources, 6,095,524 are retweets and 622,664 contain emoticons. We randomly sampled a subset of 602 tweets from the collected data and asked three annotators (volunteers from our research labs) to label the 602 tweets as positive, negative and neutral. After the annotation process, we had 249 positive examples, 216 negative examples and 137 neutral examples. These annotated tweets are used for the evaluation of our sentiment classifiers.

2.2.2. Feature Extraction for Sentiment Classification

The supervised learning problem can be formally defined as follows. Given an *independent and identically distributed* (*iid*) data set $\mathcal{D}$ of labeled examples $(\mathbf{x}_i, y_i)_{i=1,\cdots,n}$, $\mathbf{x}_i \in \mathcal{X}$ and $y_i \in \mathcal{Y}$, where $\mathcal{X}$ denotes a vocabulary of words/features and $\mathcal{Y}$ denotes the set of all possible class labels; a hypothesis class $\mathcal{H}$ representing the set of all possible hypotheses that can be learned; and a performance criterion P (e.g., accuracy), a learning algorithm L outputs a hypothesis $h \in \mathcal{H}$ (i.e., a classifier) that optimizes P. The input $\mathbf{x}_i$ can represent natural text over a finite vocabulary of words $\mathcal{X}$, $\mathbf{x}_i \in \mathcal{X}^*$. During classification, the task of the classifier h is to accurately assign a new example $\mathbf{x}_{test}$ to a class label $y \in \mathcal{Y}$ In our case, examples are tweets posted during Hurricane Sandy. These tweets are labeled as positive, negative or neutral, based on the polarity of the emotion expressed in each tweet. In what follows, we describe our features used as input to machine learning algorithms. These features are divided into two types: unigrams and sentiment-based features (polarity clues, emoticons, Internet acronyms, punctuation, and SentiStrength).

- **Unigrams:** This approach is widely used in sentiment classification tasks [33, 44]. Each tweet is drawn from a multinomial distribution of words from a vocabulary, and the number of independent trials is equal to the length of the tweet. For unigrams, we consider frequency counts of words as features. We performed stemming, stopword removal, and punctuation removal.
- **Polarity Clues:** These are the words in a tweet that express the polarity of opinions/emotions. They are good indicators for calculating the sentiment of a given text. We extract three features: PosDensity, NegDensity and PosVsNegDensity from each tweet. PosDensity is the number of positive polarity clues (positive words) normalized by the number of words in the tweet. Similarly, we compute NegDensity for the negative polarity clues. PosVsNegDensity is the number of positive per negative polarity clues, calculated as (PosDensity+1)/(NegDensity+1). We used a list of positive and negative words created by Hu and Li [15]. We turned a negated positive word into a negative word and a negated negative word into a positive word.
- **Emoticons:** In online interactions, emoticons such as ":)" and ":(" are widely used to express emotional states. Each tweet is checked for emoticons by looking up an emoticon dictionary built from Wikipedia. If a match of the emoticon pattern is found, then the value for this feature is 1. Otherwise, the feature value is 0.
- **Internet Acronyms:** In Twitter, acronyms are fairly common since the length of a tweet is restricted to 140 characters. For example, "lol" is used for laughing

out loudly. We calculated positive and negative acronym counts by using positive and negative dictionaries and used them as features. We collected commonly used Internet acronyms and constructed positive and negative dictionaries.

- **Punctuation:** In online interactions, punctuation shows intensity of emotions. For example, “I hate this!” and “I hate this!!!!!!!!!!” represent different means of writing the same text, but with different intensities of emotion. Most commonly used punctuation marks are exclamation mark ‘!’ and question mark ‘?’. We extracted exclamation and question marks from tweets and used their counts as features.
- **SentiStrength:** The sentiment strength of a tweet is calculated with the SentiStrength algorithm. SentiStrength is a tool designed for short informal text in online social media. For a tweet, the algorithm computes a positive and a negative sentiment score. These scores are used as features in our model.

2.2.3. Experiments and Results

We treat our three-class classification problem as two binary classification problems as follows: first, we classify tweets as polar vs. neutral using the SentiStrength algorithm. The algorithm returns two sentiment scores for a given English short text: a positive score ranging from 1 to 5 and a negative score ranging from -5 to -1. A tweet with +1 and -1 scores is labeled as neutral; otherwise, it is labeled as polar. Second, we classify polar tweets as positive vs. negative using two machine-learning classifiers, i.e., Naive Bayes and Support Vector Machine (SVM) classifiers trained on three types of features: unigrams, sentiment-based features, and their combination. We report the average classification accuracy obtained in 10-fold cross-validation experiments. In the experiments, we used SVM with a linear kernel and with different values for C = 0.1, 0.5, 0.75, 1.0 (the value of C dictates the penalty assigned to errors).

For the positive vs. negative classification task, we used the SentiStrength algorithm as a baseline. For each of the 465 polar tweets in our labeled dataset, we generated positive and negative scores using SentiStrength, and used the two scores directly as rules for making

TABLE 2.1. Performance of Naive Bayes and SVM for sentiment classification using various features.

Feature type	**Naive Bayes**	SVM			
		C=0.1	**C=0.5**	**C=0.75**	**C=1**
Sentiment-based	68.60	67.95	67.52	67.09	67.09
Unigrams	71.82	72.25	72.04	70.10	68.60
Combination	73.33	**75.91**	73.54	72.47	71.61

inference about the sentiment of a tweet. Again, a score of +1 and −1 implies that the text is neutral. We say that a text is positive if its positive sentiment score is greater than its negative sentiment score. A similar rule is used for inferring negative sentiment. For example, a score of +3 and −2 implies positive polarity and a score of +2 and −3 implies negative polarity. If both scores are equal for a tweet (e.g., +4 and −4), we assigned the tweet to both classes. Applying this scheme on the 465 annotated tweets, we obtained an accuracy of 59.13%.

Table 2.1 shows the results of the comparison of different classifiers, Naive Bayes and SVM trained using three feature types: unigrams, sentiment-based features, and their combination. As can be seen from the table, all classifiers trained using the combination of unigrams and sentiment based features outperform classifiers trained using unigrams and sentiment-based features alone. This suggests that the two sets of features complement each other, e.g., the presence of emoticons boosts unigrams, and the presence of words not existent in the positive and negative dictionaries boosts sentiment-based features.

The performance of SVM keeps decreasing as we increase the value of the parameter C. This suggests that the higher the value of C, the fewer errors are allowed on the training set, which causes the models to overfit, and hence, to result in poor performance on the test set. SVM (C=0.1) achieves 75.91% accuracy using the combination of features as compared to 67.95% and 72.25% accuracy of SVM (C=0.1) using sentiment-based features and unigrams, respectively, and as compared with 59.13% accuracy achieved by SentiStrength. A naive approach that classifies all tweets in the majority class achieves 53.54% accuracy, which is much worse than that of SVM (C=0.1) i.e., 75.91%.

Comparison with Prior Work: Our work is the most similar with that of Nagy & Stamberger[35]. Hence, we performed a comparison with this work. The approach presented in Nagy & Stamberger [35] is to develop features based on emoticons and two sentiment dictionaries - AFINN [41] and SentiWordNet [3]. In addition to the features used in Nagy & Stamberger [35], i.e., emoticons and sentiment dictionary features, we used unigrams, punctuation, internet acronyms, and SentiStrength scores (our features are described in Subsection 3.2). The results of this comparison are shown in Table 2.2. As can be seen from the table, the classifiers based on our features perform substantially better compared with the approach proposed by Nagy & Stamberger [35]. Specifically, the highest performance that our models achieve is 75.91% as compared with 64.30% achieved by models proposed in Nagy & Stamberger [35].

Table 2.2. The comparison of our sentiment classification with Nagy and Stamberger (2012).

Feature type	**Naive Bayes**	SVM			
		C=0.1	**C=0.5**	**C=0.75**	**C=1**
Our features	73.33	**75.91**	73.54	72.47	71.61
Nagy and Stamberger (2012)'s approach	64.30	61.50	63.44	63.22	63.65

2.2.4. Geo-tagged Tweets Sentiment Analysis

In order to associate the sentiment of tweets with their geo-locations, we extracted the set of geo-tagged tweets from our Hurricane Sandy collection. In our data, there are 74,708 tweets with geo-location. We then used the SentiStrength to identify the neutral tweets (those for which SentiStrength returns +1 and −1 scores). Finally, we used our best performing classifier, i.e., SVM (C=0.1) with the combined features trained on Sandy data, to label the remaining tweets as positive and negative (i.e., the tweets with SentiStrength scores different from +1 and −1). In order to understand the general mood during the Hurricane Sandy, we performed a geo-mapped sentiment analysis. Although Hurricane Sandy had a physical impact that was regionally limited, the storm affected people in locations far away from the east coast of the United States. This is reflected in the global extent of geo-located tweets on the topic of Sandy. Regardless, in a disaster scenario of this magnitude, where the topic of the tweet is geographically specific and its physical impact isolated, spatial proximity to the event understandably has an impact on the credibility of the tweeted information [69]. Temporal distance similarly impacts the tendency of a Twitter user to disseminate information about an emergency event [50]. In this section, we use the geographic representation and cluster measures to examine the spatial and temporal variation of Twitter data with respect to Hurricane Sandy.

Given a dataset of tweets related to Sandy, we rely primarily on clustering methods to understand the spatial arrangement of geo-located tweets to avoid the stationarity of large population centers. Although tweets contain detailed temporal information, we aggregated them to the daily scale because of the effect of global time zones. We represent the spatial extent of Sandy using the National Oceanic and Atmospheric Association's (NOAA) National Hurricane Center 34-knot (NOAA's threshold for tropical storm classification) wind speed approximation between October 26 and October 29, 2012 the day the storm's threshold made landfall in New Jersey. After making landfall and dissipating in strength, we approximate the extent of the storm with buffers of decreasing diameter through October 31 around the best-track of the storm's center provided by NOAA [36]. Visual comparison of maps generated with this data and measures of the clustering tendency of tweets around Sandy's landfall point reinforce the hypothesis that Twitter users tweet about a developing disaster with greater proximity, reaching a peak of concentration during and at the location of the disasters impact [69].

We first visually examined the spatial arrangement of tweets. Observing the movement of the geographic mean center reveals the hemispheric shifts in Twitter use during the course of Sandy's development, landfall, and dissipation. The point at the mean center moves from a location more central to the area which Sandy impacted (the US east coast) to a more northern location following the onset of the storm as tweets around the globe pull the center of the cluster away. A one standard deviation ellipse surrounding the geographic center also shows a similar trend in the contraction and subsequent expansion of its diameter (Figure 2.1). This finding supports the use of social media in disaster management scenarios as individuals are much more likely to share information via Twitter about a disaster while and where it is occurring.

Following the visual analysis, we conducted a statistical measure of the clustering tendency of tweets based on their proximity to the point where Sandy made landfall. We evaluate the distance between each tweet and Sandy's landfall point then plot them based on the number of tweets that fall within predefined radii around that point. The positive skewness of the resulting histograms signify a minimal distance between tweets and the landfall point, and indicate an extreme tendency to cluster. Observing the histograms over time reinforces our visual analysis that Twitter users tweet about Hurricane Sandy with great proximity to it, increasing to a maximum during the storm's maximum impact, then quickly to a less clustered, global dispersion (Figures 2.2 and 2.3). Additionally, positive and negative sentiments expressed in tweets about Hurricane Sandy have unique patterns. Both positive and negative sentiment generally follow the trend of increasing clustering tendency to the point of Sandy's maximum impact and dispersion on the following days. However, negative sentiment tweets consistently cluster in closer proximity to Hurricane Sandy (Figure 2.4).

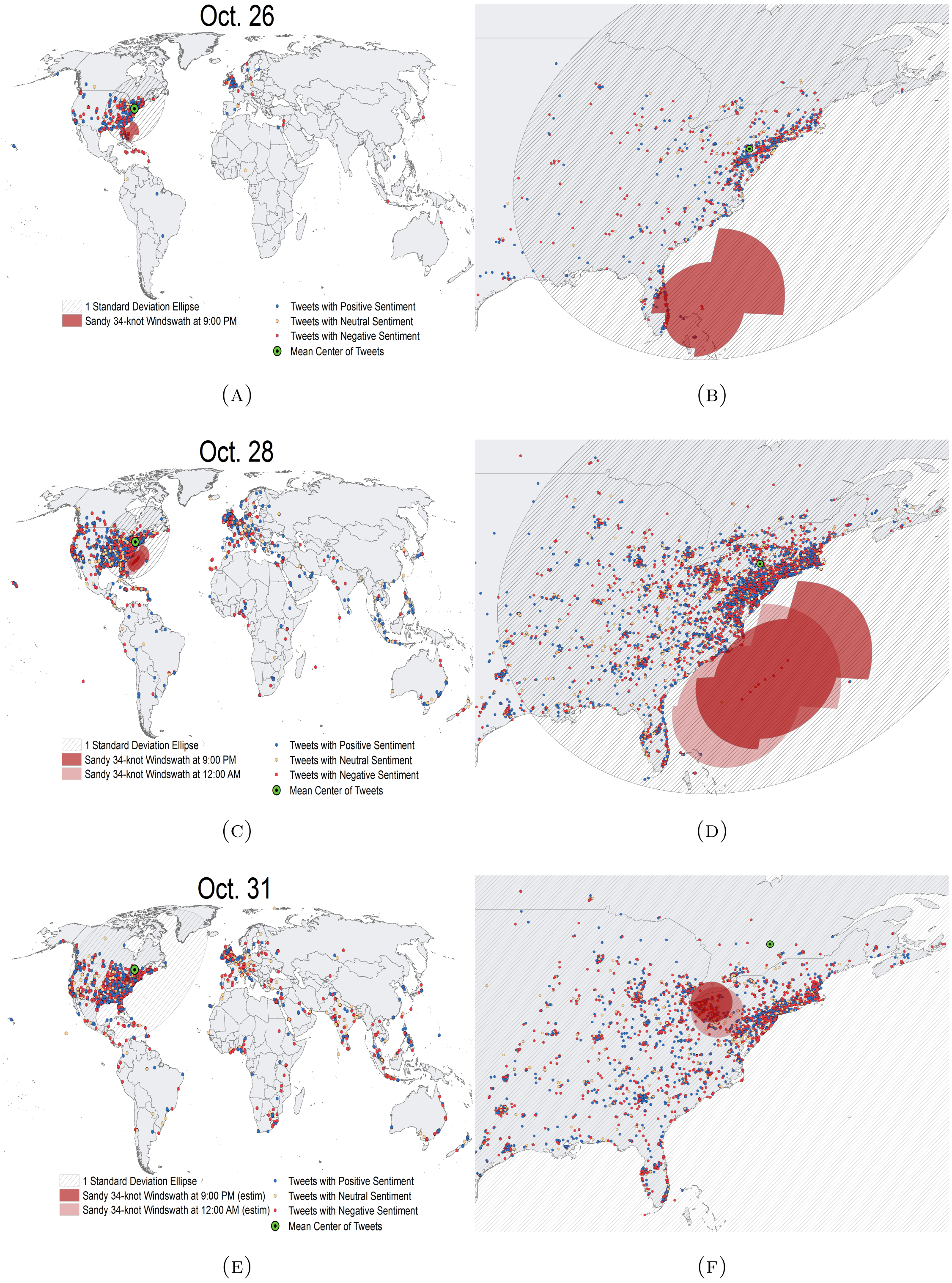
While sentiment alone cannot make social media information actionable for disaster responders, expressions of concern for others and notification of infrastructure failure, for example, present situations of negativity and potentially a cry for help. Furthermore, we have demonstrated that there is a spatial arrangement of positive/negative sentiment tweets. The arrangement indicates that sentimental expression is significant for the social and spatial environment of a disaster, and therefore for generating actionable information (such as negative sentiment clusters which intimates the people need help). In our perspective, we define actionable information as the data which helps in better decision making. Through our geo-tagged sentiment maps, emergency responders can interpret the emotional intensities on the ground and can plan the relief efforts more efficiently. They can have an overview of the how people are feeling about the disaster. For example, using our sentiment map emergency responders can focus their relief efforts on the clusters which emit negative sentiments that are closer to the proximity of the Hurricane landfall.

2.3. The Impact of Emotional Divergence on Retweetability of Tweets during Hurricane Sandy

In this section, we analyze the impact of emotional divergence on the retweetability of tweets during Hurricane Sandy. During natural disasters, identifying how likely a tweet is to be retweeted is very important since it can help promote the spread of "good" information in Twitter, as well as it can help stop the spread of misinformation, when corroborated with approaches that identify trustworthy information or misinformation, respectively.

We adopt the definition of "emotional divergence" (ED) from Pfitzner et al. [49], which is defined as *"the (normalized) absolute difference between the positive and the negative sentiment score delivered by SentiStrength"*. It is calculated using the formula $ED = \frac{p-n}{10}$, where p is a positive score and n is a negative score output by the SentiStrength algorithm. As mentioned in Section 3.2, SentiStrength algorithm outputs a positive score ranging from 1 and 5 and a negative score ranging from -1 and -5, hence $ED \in [0.2,\ 1]$. Emotional divergence in a given short text measures the spectrum of the emotions expressed in it, whereas emotional polarity (sentiment) capture the overall emotion from the text. For example, in the tweet "I hope this storm is TERRIBLEEEEEEE!!!!!! Lol forreal. Ill sleeep perfect", SentiStrength outputs a positive score (p) of 3 and a negative score (n) of -5, which makes the emotional polarity as negative (-2). We can see that the user is expecting the storm to be bad, but his intention to sleep well, is expressing a happy emotion. Even though the emotional polarity is negative, there is a high divergence in the emotions present in the tweet, which is precisely captured by the emotional divergence. In our example, ED is 0.8 showing a highly emotional contrast.

We first studied the emotional divergence using our geo-tagged tweets. In our data, there are 74,708 tweets with geo-location, which are also used in our geo-tagged sentiment analysis. From these geo-tagged tweets, we separated the tweets that are retweeted from


Oct. 26
1 Standard Deviation Ellipse
Sandy 34-knot Windswath at 9:00 PM
Tweets with Positive Sentiment
Tweets with Neutral Sentiment
Tweets with Negative Sentiment
Mean Center of Tweets
Oct. 28
1 Standard Deviation Ellipse
Sandy 34-knot Windswath at 9:00 PM
Sandy 34-knot Windswath at 12:00 AM
Tweets with Positive Sentiment
Tweets with Neutral Sentiment
Tweets with Negative Sentiment
Mean Center of Tweets
Oct. 31
1 Standard Deviation Ellipse
Sandy 34-knot Windswath at 9:00 PM (estim)
Sandy 34-knot Windswath at 12:00 AM (estim)
Tweets with Positive Sentiment
Tweets with Neutral Sentiment
Tweets with Negative Sentiment
Mean Center of Tweets

(A) (B)

(C) (D)

(E) (F)

FIGURE 2.1. Maps of Positive, Neutral, and Negative Tweets at global and regional scale. The maps are drawn using ArcGIS (www.esri.com/software/arcgis).

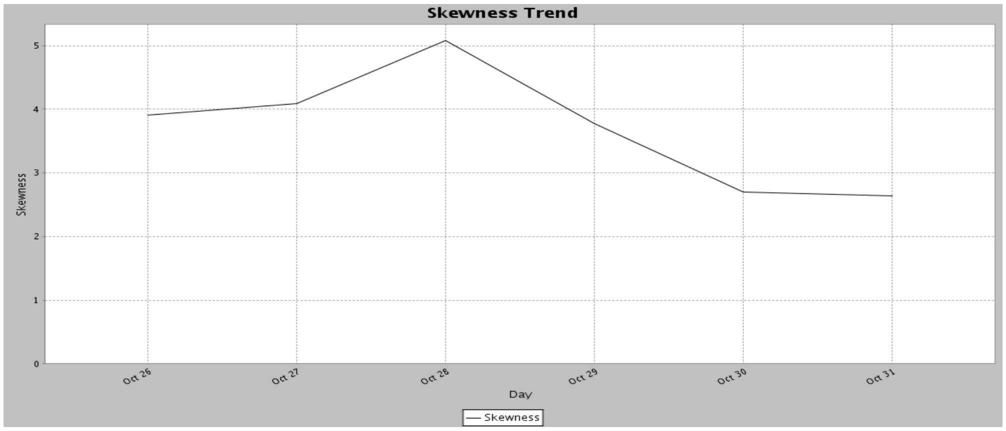


FIGURE 2.2. Skewness as a function of time.

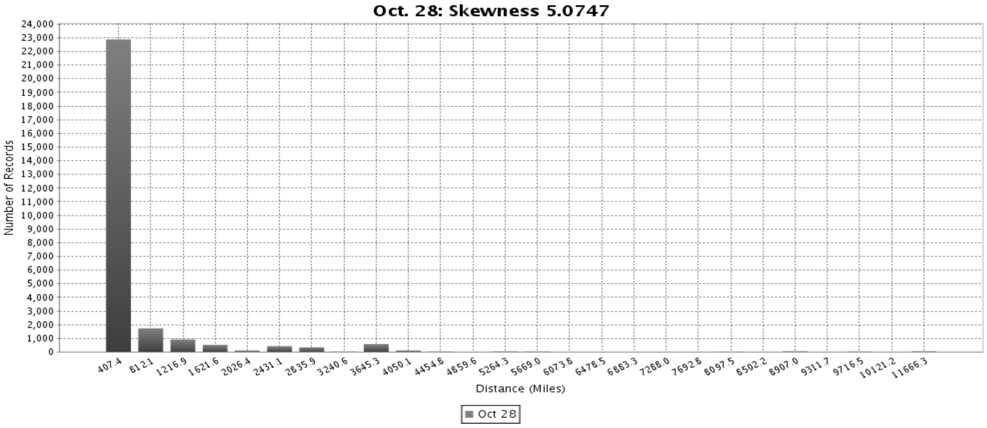


FIGURE 2.3. Histogram of October 28. The extreme positive skewness indicates short distances between each Tweet and the point where Hurricane Sandy made landfall.

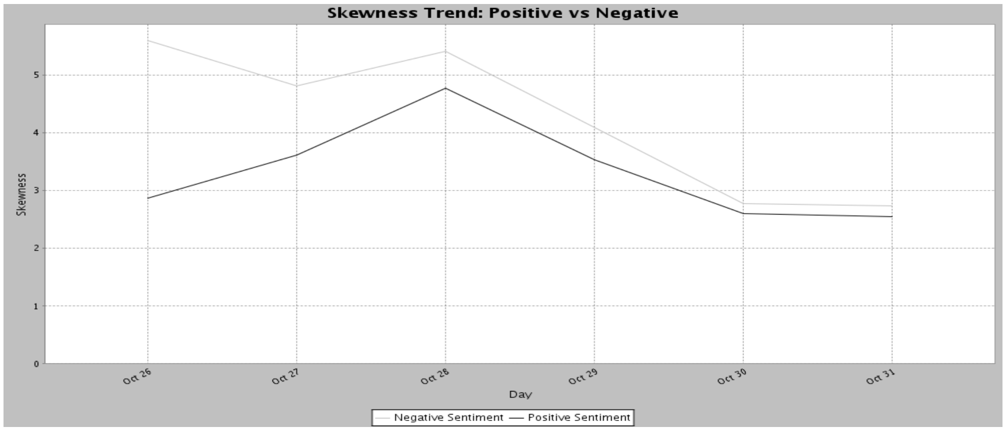
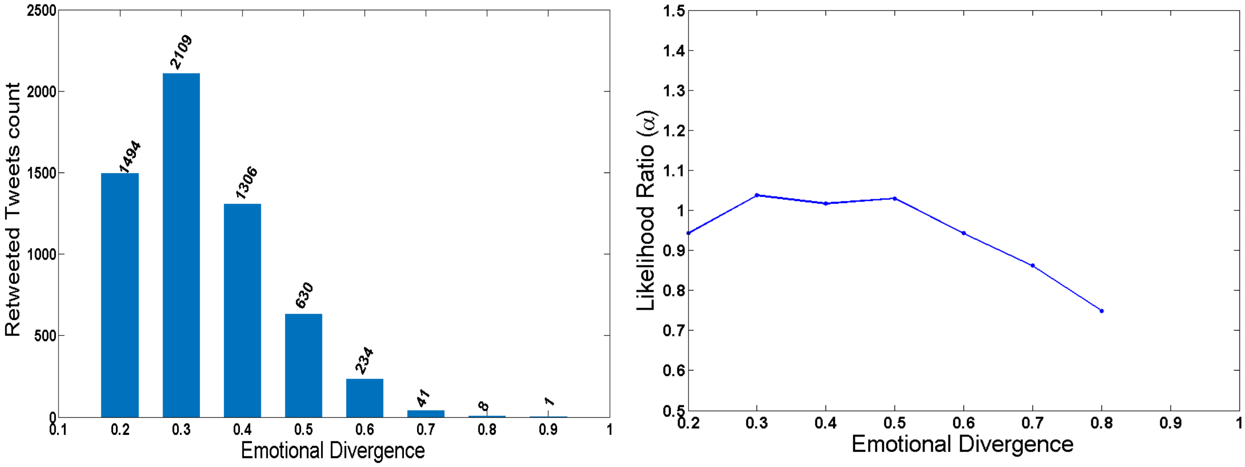


FIGURE 2.4. Positive vs. negative skewness as a function of time.

those that are not retweeted. We then analyzed the impact of emotional divergence on retweetability using this sample of geo-tagged tweets. The sample contains 5,823 retweeted tweets (only initial tweets) and 68,885 tweets that are not retweeted. We then calculated the emotional divergence value for each of geo-tagged tweets. In Figure 2.5(a), we plot the counts of the 5,823 retweeted tweets for each emotional divergence value, from 0.2 to 1. As can be seen from the figure, the number of retweeted tweets is decreasing with the increase in the emotional divergence. This implies that there is a good proportion of retweeted tweets for low emotional divergence values, and a tweet tends to have retweets if it is less emotionally divergent. However, the correlation between the counts and the emotional divergence for retweeted tweets does not elevate how likely a tweet can be retweeted. The likelihood of each type (i.e., retweeted (RT) and not-retweeted (T)) gives better insight about how emotional divergence affects retweetability.



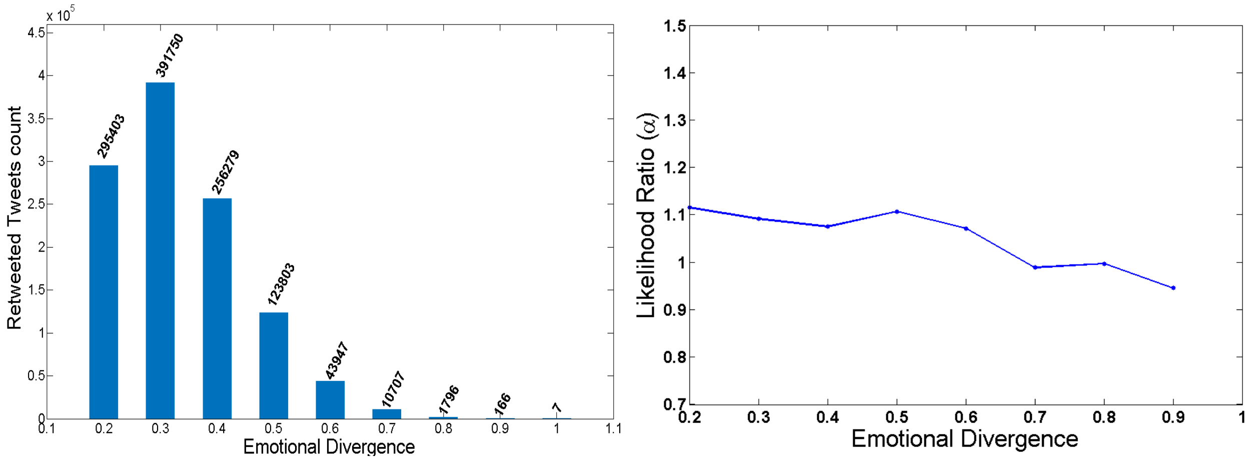
(A) ED vs. # retweeted tweets

(B) ED vs. Likelihood Ratio (α)

Figure 2.5. Impact of Emotional Divergence on Retweetability of geo-tweeets.

For each emotional divergence value (ED), we calculated $Pr(ED = x|T)$ for $x = 0.2, \ldots 1$, which is the probability of a not-retweeted tweet (T) to have emotional divergence x (i.e. the number of not-retweeted tweets having ED x, divided by the total number of not-retweeted tweets in the sample). Similarly, $Pr(ED = x|RT)$ is the likelihood of a retweeted tweet (RT) to have emotional divergence ED=x (i.e., the number of retweeted tweets having

ED=*x* divided by the total number of retweeted tweets in the sample). In Figure 2.5(b), we plotted the likelihood ratio with respective to the *ED*=*x* values. We observe a decreasing trend as the *ED* value is increasing. We noted that there was a sharp decrease in the ratio in the region $ED = 0.5$ to 0.8. This means that the chance of a tweet to be retweeted is higher for low emotional divergence values. In Pfitzner et al. [49], the trend in their plots show that the tweets with high emotional divergence value have a higher chance of being retweeted in the network, indicating that those tweets have more number of retweets. In contrast, we show an opposite trend, tweets that have high emotional divergence values have a low likelihood of being retweeted. We suspect that the variation is due to the data collected during different kinds of events. Our work is purely focused on disasters, whereas in Pfitzner et al. the dataset contains tweets from a variety of events related to Oscar ceremony, sports, and technological product launches, where highly emotional divergent tweets are more likely to be retweeted.



(A) ED vs. #retweeted tweets

(B) ED vs. Likelihood Ratio (α)

FIGURE 2.6. Impact of Emotional Divergence on Retweetability of tweets from the whole dataset

To further validate our findings, we performed the same analysis on the whole Sandy dataset. Out of 12.9 million tweets, 1.1 million tweets are retweeted during the disaster and around 7 million tweets are not retweeted. The analysis results in trends that are similar to what we found for the geo-tagged tweets sample. Figure 2.6(a), shows the correlation between emotional divergence and the corresponding retweeted tweets count. We observe that, among the retweeted tweets, the proportion of the retweeted tweets is decreasing as the emotional divergence value is increasing. We observe that the tweets having low emotional divergence values have many retweets. Similarly, Figure 2.6(b) shows the correlation between emotional divergence and the likelihood ratio. The chance of being retweeted is higher for the low *ED* values and decreased for high *ED* values. We observe that a tweet with $ED <$ 0.6 has a higher probability of being retweeted than the tweets with $ED > 0.6$.

Because in disaster events, tweets that convey information are more likely to be retweeted than those that are conversational in nature, we study the correlation between low emotional divergence (hence, high chance of retweetability) and the tweets' informativeness in the following section.

2.3.1. Emotional divergence vs. Informativeness

This experiment is to explore how the diversified emotions would affect the informativeness of the tweets. For each *ED* value, we ranked the tweets based on their retweets counts, meaning that a top-ranked tweet will have the highest number of retweets in the corresponding *ED* value. We then selected two sets of tweets, where one set contains tweets with *ED*=0.2, and the other contains tweets with *ED*=0.8. We noticed interesting patterns in these two sets. Tweets with *ED*=0.2 convey valuable information, which is useful for the disaster bystanders and emergency response organizations. Moreover, they express a neutral sentiment, are more objective (as opposed to being subjective), and are informative in nature. Table 2.3 shows these example tweets and their retweets count. Moreover, tweets with ED=0.8 are more conversational in nature. Table 2.4 shows the tweets with emotional divergence *ED*= 0.8 and their retweets count. They express personal opinion/feeling of users rather than conveying necessarily useful information. To validate this finding, we analyzed a set of tweets available at http://crisislex.org[1]. This sample was constructed by Olteanu et al. [42], which contains tweets annotated based on the tweet's informativeness, i.e, a tweet is labeled as informative if it conveys useful information and it is labeled as non-informative

[1] http://crisislex.org/data-collections.html

TABLE 2.3. Examples of retweeted tweets with low emotional divergence (ED = 0.2) and their retweets count.

Retweets Count	Tweet
251	# laguardia # lga flooded. Jet bridge is around 5 feet to the bottom where you enter the plane. @weatherchannel #sandy http://t.co/WSa2L9Ra
145	Norfolk continues to get hit hard by #Sandy #HRSandy http://t.co/MS9QAGE0
96	#Sandy power outages top 8.2 million http://t.co/gWYtG6Hx

TABLE 2.4. Example of retweeted tweets with high emotional divergence (ED = 0.8) and their retweets count.

Retweets Count	Tweet
6	Big Picture on Hurricane Sandy Carries far more impact than all the fakes. I quite liked num 6. The rest devastating. http://t.co/85h5t535
2	I hope this storm is TERRIBLEEEEEEE!!!!!! Lol forreal. Ill sleeep perfect.
1	Were just going to watch bad romantic comedies dance and maybe cry a little for the next two days. #sandy

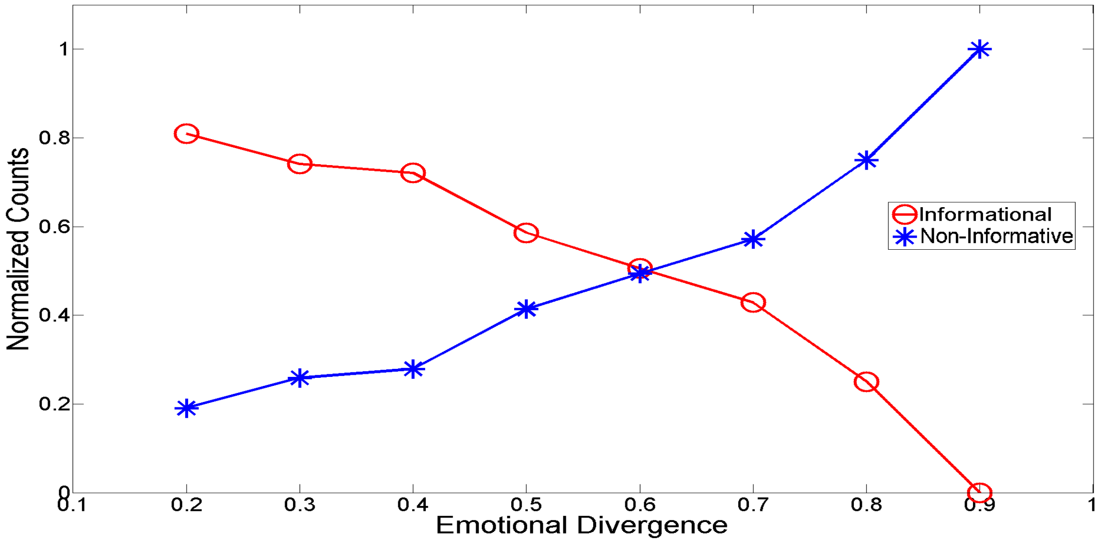


FIGURE 2.7. Emotional Divergence vs Informational/Conversational.

if it doesn't contain any useful information. The dataset contains tweets from 26 disasters happened during 2012 and 2013 all around the world. For each disaster there are around 1000 tweets coded with informational and conversational labels. For these tweets, we calculated ED values and recorded the total number of the tweets that are distributed in each ED value. For each ED value, we calculated the amount of tweets which are informational and conversational, and then normalized them with the total number of tweets distributed in each ED value.

Figure 2.7 shows the plot between emotional divergence and normalized counts of the tweets. In the figure, the red line represents the informational tweets and the blue is for conversational. As can be seen, the two curves show opposite trends. The curve for informational tweets shows a decreasing trend, as the emotional divergence is increasing the normalized counts is decreasing. Whereas, for conversational the trend is increasing, as ED value is increasing, the normalized count is also increasing. This implies that for at low ED values, the normalized counts value of informational tweets is higher than that of conversational tweets. Similarly, for high ED values, the normalized counts of conversational tweets is higher than that of informational tweets. This implies that the chance for the tweets with low ED values to be informational is more and the chance for the tweets to be conversational is higher at high ED values. In future, it would be interesting to see whether emotional divergence really impacts the prediction of informative tweets. This suggests that the chance for the tweets with low ED values to be informational is higher, whereas the chance for the tweets to be conversational is higher at high ED values.

2.4. Summary and Discussion

We performed a sentiment analysis of user posts in Twitter during Hurricane Sandy and visualized these sentiments on a geographical map centered around the hurricane. We show how users' sentiments change according not only to the locations of users, but also based on the relative distance from the disaster. In addition, we investigated the influence of *emotional divergence* on retweetability of a tweet and showed that the chance of retweeting a tweet decreases as the emotional divergence increases. Another interesting pattern that we discovered is that the content of tweets with low emotional divergence is generally informative in nature (see Table 2.3) whereas, the content for the tweets with high emotional divergence is more of personal opinions and do not necessarily convey any useful information. We supported this by using the tweets from CrisisLex datasets with informational and conversational labels. In this analysis, we found that the proportion of informative tweets is more than the conversational tweets at low ED values and the proportion of conversational tweets is more than informative tweets at high ED values.

2.4.1. Related Work

There have been very few works on identifying the polarity of sentiments expressed by users in social networking sites during disaster-related events. Nagy & Stamberger [35] focused on sentiment detection in Twitter during the San Bruno, California gas explosion and fires from 09/2010. They used SentiWordNet to identify the basic sentiment of a tweet, together with dictionaries of emoticons and out of vocabulary words, and a sentiment-based dictionary. Schulz et al. [52] proposed a fine-grained sentiment analysis to detect crisis related micro-posts and showed significant success in filtering out irrelevant information. The authors focused on the classification of human emotions into six classes: anger, disgust, fear, happiness, sadness, and surprise. As features, they used bag of words, part of speech tags, character n-grams (for n=3, 4), emoticons, and sentiment-based words compiled from the AFINN [41] word list and SentiWordNet[3]. They evaluated their models on tweets related to the Hurricane Sandy from October 2012. Mandel et al. [32] performed a demographic sentiment analysis using Twitter data during Hurricane Irene. Pfitzner et al. [49] introduced the concept of *emotional divergence* which measures the diversity of the emotions expressed in a text and analyzed how likely a tweet is to be retweeted with respect to its emotional divergence value.

In contrast to these works on sentiment analysis, we focus on geo-tagged sentiment analysis of tweets from Hurricane Sandy in order to obtain a holistic view of the general mood and the situation "on the ground" during the hurricane, which can help increase situational awareness and inform emergency response organizations. We also study the effect of emotional divergence on retweetability during Hurricane Sandy and how does the emotional divergence affect informativeness.

2.4.2. Discussion

There are probably many methods at gaining additional awareness of the affected population, some traditional and some using new techniques. In this paper, we offer one such new technique. We find that social media is a rich source of data surrounding a disaster event. Leading up to, during and after a disaster more and more people turn to social media to describe their experiences, express their needs, and communicate with other affected persons. This online discussion is a rich trove of information that could possibly inform responders, if made actionable. There are several reasons that this data is not yet seen as fully actionable including the sheer amount of data, the inability to sort and categorize the data into useful types, and the inability to fully trust data or unknown sources. One additional strong reason that the data is not currently used to its full potential is a lack of connection between the location of the disaster event and those tweeting about the disaster. Because of the nature of social media, contributors do not have to be bystanders. Responders interested in the wellbeing of physical bystanders seek methods of finding and measuring the concerns of those directly affected by a disaster.

The strongest contribution of this work is a proof of concept. Using Twitter data from Hurricane Sandy we identify the sentiment of tweets and then measure the distance of each categorized tweet from the epicenter of the hurricane. Currently, responders can track weather data to know where a hurricane hits an affected population, but they cannot know in real time the effect that disaster is having on the population. They often ask, “How bad is it out there?” Traditionally, they rely on either eyewitness accounts after the fact from survivors, or eyewitness information offered in real time by those who can make phone calls. In addition, through an analysis of the divergence of sentiments in tweets, we studied how likely a tweet is to be retweeted based on its emotional divergence. Our models can be integrated into systems that can help response organizations to have a real time map, which displays both the physical disaster and the spikes of intense emotional activity in proximity to the disaster. In time, such systems could pinpoint the joy of having survived a falling tree, the horror of a bridge washing out or the fear of looters in action. Responders might be able to use a future iteration of such a system to provide real-time alerts of the emotional status of the affected population. We find that mapping emotional intensity during a disaster may help responders develop stronger situational awareness of the disaster zone itself.

In their 2011 paper, MacEachren et al. [31] argue that extracting and categorizing social media data is where most researchers have focused their energy, and those efforts are not enough to change the data into actionable knowledge [31]. It is essential to refocus on the utility of the extracted information and the effectiveness of associated crisis maps to support emergency response. In our work, we presented a method by which the affected population's response to a disaster might be measured through a sentiment analysis and then mapped to the disaster in space and time. This is one big step along the path to providing official responders with truly actionable information in real time based on social media data.

CHAPTER 3

PREDICTING RETWEETABILITY DURING HURRICANE DISASTERS

During natural disasters, identifying how likely a tweet is to be retweeted is crucial since it can help promote the spread of useful information in a social network such as Twitter, as well as it can help stop the spread of misinformation when corroborated with approaches that identify misinformation. In this chapter, we present an analysis of retweeted tweets from two different hurricane disasters, to identify factors that affect retweetability. We then use these factors to extract features from tweets' content and user account information in order to develop models that automatically predict the retweetability of a tweet. The results of our experiments on Sandy and Patricia Hurricanes show the effectiveness of our features.

3.1. Introduction

Information diffusion is vital in the context of disasters to make the victims and responders aware about the situations surrounding them. Many emergency decision makers see the data produced through crowdsourcing as ubiquitous, rapid and accessible - with the potential to contribute to situational awareness [72]. As the use of public social media in crisis increased, emergency responders started to take notice of the way citizens engaged in social media and the information exchanges that took place there [19]. Consequently, responders began to consider if social media might be a useful tool for their practice. Research revealed that social media could be used to distribute information quickly to a wide-spread audience [26] and to engage more directly in a two-way conversation with members of the public [16]. The information that the public produced looked to be useful, as researchers showed that it could contribute to situational awareness during a crisis event [4, 22]. According to Starbird et al. [56], social media data that can be identified as coming from local bystanders to a disaster can be extremely important to emergency responders. Most of the social media data surrounding a disaster are derivative in nature: information in the form of reposts or pointers to information available elsewhere [55]. These derivative data are abundant, as a form of noise that must be filtered out to arrive at the signal of good data [1]. A small subset of the data comes from locally affected populations in the form of citizen reports [55]. Starbird et al. [55] assert that bystanders "on the ground are uniquely positioned to share information that may not yet be available elsewhere in the information space and may have knowledge about geographic or cultural features of the affected area that could be useful to those responding from outside the area."

Hughes and Palen [16] examined the role of social media in emergency management and found that emergency managers see the potential of social media as means of engaging the public quickly and widely during a crisis. Vieweg et al. [71] showed that retweeted tweets are likely to contain information that contributes to situational awareness and are likely to be actionable compared with non-retweeted tweets. In addition to the information that creates awareness to the responders, people also post information related to relief efforts during disasters (such as offering shelters, donations, and food), for which the target consumers are the victims who need aid. *We believe that understanding how likely a tweet is to be re-tweeted seconds after it is posted has the potential to help responders to influence the speed and spread of messages, which could make substantial improvements in the relief efforts and can positively impact people who are badly affected by a disaster.* However, the retweetability of a tweet is influenced by many factors including the aspects of a user who posted the information and the content present in it. In this paper, we focus on identifying factors that affect retweetability of a tweet during mass emergencies. These factors could be used in a real-time system to promote relevant tweets that convey useful information as well as to stop the spread of misinformation when corroborated with approaches that identify rumors and misinformation. Our research questions are: "*In a social media (e.g., Twitter) stream of messages, what features (or factors) affect the spreading (retweetability in our case) of a message? How well do these features help in automatically predicting a message retweetability during disasters?*" We specifically address these questions with our research agenda using Twitter datasets collected during Hurricane Sandy and Hurricane Patricia. Precisely, we present a supervised learning approach with a wide range of feature exploration to identify the retweetability of a tweet. Our approach, originally introduced in Neppalli et al. [39], can help increase situational awareness and can inform emergency response organizations on how to reach the widest audience in the fastest way during disaster events. In this extended work, we augment our contributions to retweetability analysis and prediction in disaster events with our findings from a larger spectrum of experiments using state-of-the-art classification approaches and two different hurricanes: Sandy and Patricia.

3.1.1. Contributions and Organiztion

The contributions of this chapter are as follows: (i) we present an analysis of retweeted tweets during Sandy and Patricia Hurricanes to determine several aspects affecting retweetability; (ii) we design features from tweets' content and user account information for learning models that predict the tweets' retweetability during disaster events, and study the predictive power of these features; (iii) we experimentally show that the models trained using the designed features perform better than strong baselines and previous approaches; and (iv) we find that the quantitative user features (e.g., #friends, #followers, #statuses) when normalized with the user age show better performance than the unnormalized user features.

The rest of the chapter is organized as follows: Definitions and Datasets sections describe the details of the datasets used for this work. Then, we present the data analysis performed on both disasters' tweets, followed by feature extraction. We then provide experimental design and results, and conclude the article with a summary, related work and future work.

3.2. Datasets

3.2.1. Our Definitions

: A post on Twitter (or a tweet) is a short message of up to 140 characters, posted by a user. A post may be direct or derivative. A direct post refers to a post that is published for the first time (by a user), whereas a derivative post refers to a re-post of a post from another user. In Twitter terminology, the former is called "tweeting" and the latter is called "retweeting." Retweeted messages have a common pattern as: "RT @A: message x," which specifies that the post is a retweet ("RT") or re-post of message x that was originally posted by user A ("@A"). A user A is called a follower of a user B if user A "follows" (or receives updates from) user B (but not vice-versa). If both users "follow" each other, then they are called friends. We believe that both relations "followers" and "friends" are important since they help pass the information in the network.

3.2.2. Hurricane Sandy

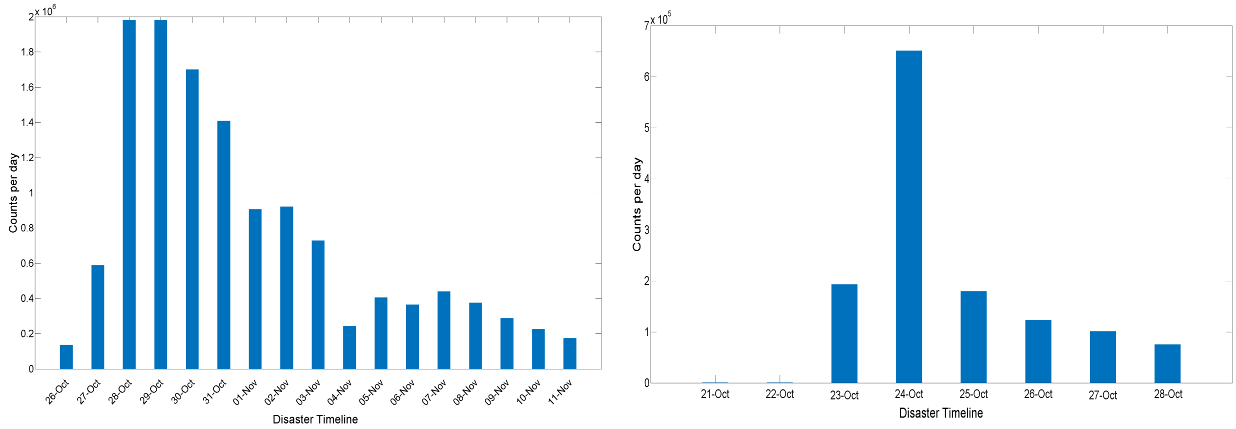
: We collected Twitter data posted during the Hurricane Sandy between October 26 and November 11, using the Twitter Streaming API (described in Section 1.4). We used the following keywords: "#hurricanesandy," "#sandy," "hurricane sandy," "#hurricane," "#sandyhurricane," and "hurricane east coast" to download the tweets during Hurricane Sandy. We used both search terms and the hashtags. Specifically, we collected 12.9 million (M) total tweets with 5.1M unique users. Out of the 12.9M tweets, 7.1M are initial tweets (or direct posts), and 5.8M are retweets (derivative posts). Out of the 7.1M initial tweets, only 1.1M tweets are retweeted, whereas the remaining 6M tweets are not retweeted.

3.2.3. Hurricane Patricia

: In addition to Hurricane Sandy, we used the tweets posted during Hurricane Patricia in October 2015. We used the tweet ids provided by Kate Starbird, Asst. Professor at the University of Washington. The search terms used by Starbird are: "#hurricane," "#patricia," "hurricane patricia," "#mexico," and "#hurricanepatricia." Due to Twitter terms and conditions, it is not allowed to directly share the tweets. Hence Starbird shared us the ids from their collection. Using the Twitter Search API (described in Section 1.4), we crawled 4.38M tweets, which comprise of tweets in different languages. Based on the "lang" attribute in the tweets' metadata, we extracted 1.38M tweets in English, 2.08M in Spanish and 59,559 tweets in other languages. We used the 1.38M ss English tweets to conduct our analysis and experiments. From these 1.38M tweets, we extracted 86,268 tweets, which are initial tweets that were retweeted to conduct our analysis and experiments.

3.3. Data Analysis

In this section, we present our analysis on the set of tweets crawled during Sandy and Patricia, and study how information is spread in the network via retweeting. Initially, we analyzed the distribution of the posts during the events to explore the tweets' posting activity. In Figure 3.1, we plot the per day distribution of all the tweets - 12.9M tweets of Sandy in the left plot and 1.38M tweets of Patricia in the right plot. We can observe a burst after two days from the beginning of both the events. The delay can be explained by the hurricanes' progressive nature, i.e., they were forecasted a few days before the strike and the pace in postings picked up as they hit the coast. As can be seen from Figure 3.1, the frequency of the tweets decreases as time elapses.

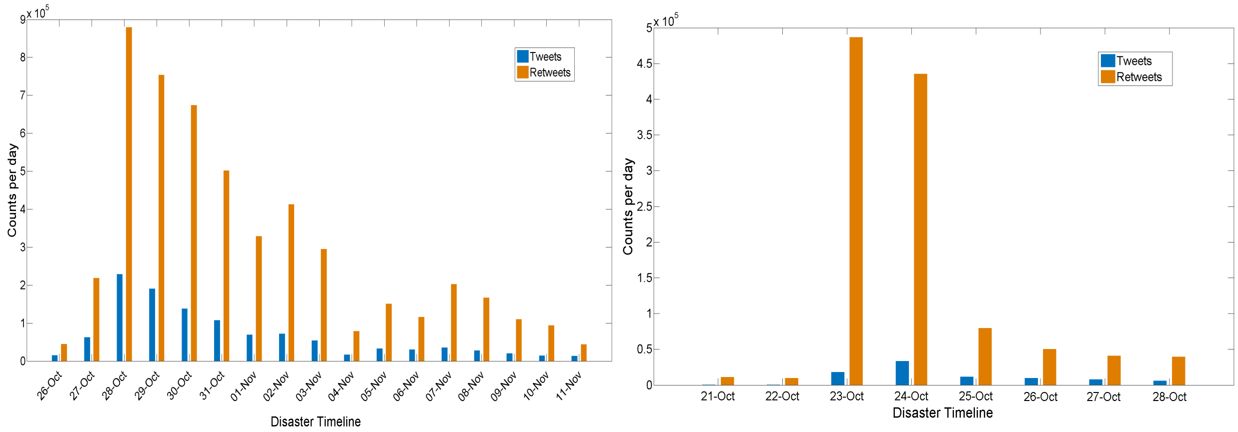


(A)

(B)

FIGURE 3.1. The distribution of total posts per day for Hurricane Sandy (A) and Hurricane Patricia (B).

An analysis of retweeted tweets - 1.1M of Sandy and 86K of Patricia, reveals a similar trend, as illustrated in Figure 3.2. Specifically, for each disaster, we remove the tweets that are not retweeted at all and keep only the tweets (direct posts) that are retweeted. In Figure 3.2, we show the number of tweets (direct posts) that are retweeted and their retweets count by day, where the left plot represents the data from Sandy and the right plot represents the data from Patricia. The trend is analogous to the plot in Figure 3.1, where we can observe a decrease in the number of tweets and retweets as the days elapse. We can also observe that the number of retweets is much higher than the number of tweets every day, showing that



(A)

(B)

Figure 3.2. The distribution of retweeted tweets and their retweets during Hurricane Sandy (A) and Hurricane Patricia (B).

the information is substantially spread in the network.

3.3.1. Retweetability vs. Number of Followers/Friends

Among the retweeted tweets (i.e., the direct tweets - 1.1M from Sandy and 86K from Patricia), we study how the number of followers or friends[1] of a user would affect the retweetability. Intuitively, we expect that a user with more followers or friends would have a better chance of having his/her tweets retweeted more often. For this analysis, we divided the direct tweets into five categories as shown in Table 3.1, based on their retweet count.

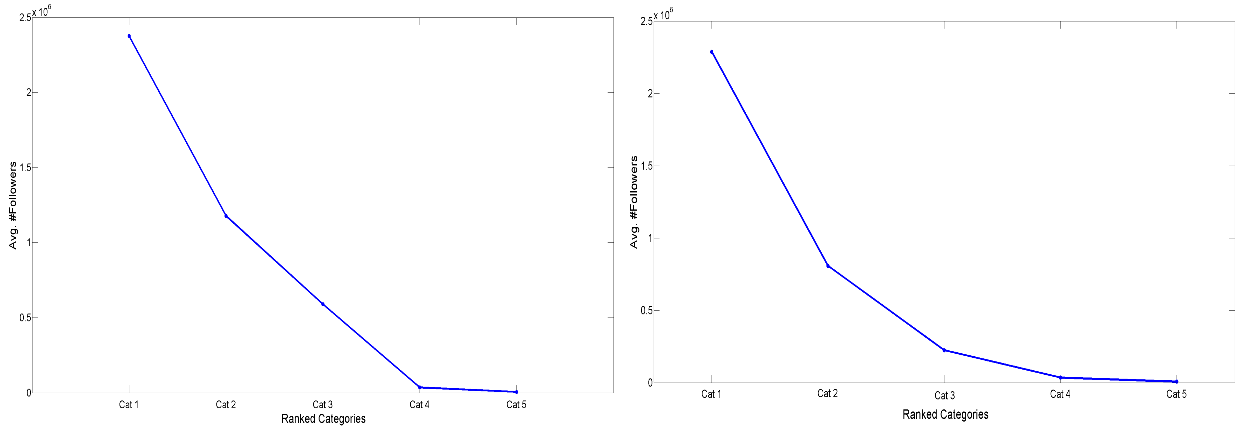
In Table 3.1 we show, for each category, from 1 to 5: the number of direct tweets split by category, the average retweets (in parenthesis next to # direct tweets) and the sum of the retweets count of all direct tweets (in a category) from Sandy and Patricia. We can see that the number of direct tweets that are retweeted only once (last row in the table) is significantly higher than the number of direct tweets that are retweeted more than 100 times (first row).

As we go from Category 1 to Category 5, the number of direct tweets keeps increasing, and the average retweets per tweet are decreasing. For Hurricane Sandy, the total number of

[1] Throughout this chapter, we refer to the followers (or friends) of a tweet as the followers (or friends) of the user who posted the tweet.

Table 3.1. The distribution of direct tweets and their retweets into five categories for Sandy and Patricia.

Retweet Category	**Hurricane Sandy**		**Hurricane Patricia**	
	# Direct Tweets (Avg.)	**# Retweets**	**# Direct Tweets (Avg.)**	**# Retweets**
1. Retweeted > 100 times (Category 1)	4,560 (398)	1,816,676	1,408 (532)	737,017
2. Retweeted > 50 & <= 100 times (Category 2)	5,226 (69)	362,107	1,268 (70)	89,681
3. Retweeted > 20 & <= 50 times (Category 3)	15,574 (30)	483,679	3,071 (31)	96,659
4. Retweeted > 1 & <= 20 times (Category 4)	434,654 (4)	1,740,840	37,997 (5)	185,505
5. Retweeted Only Once (Category 5)	665,437 (1)	665,437	42,524 (1)	42,524



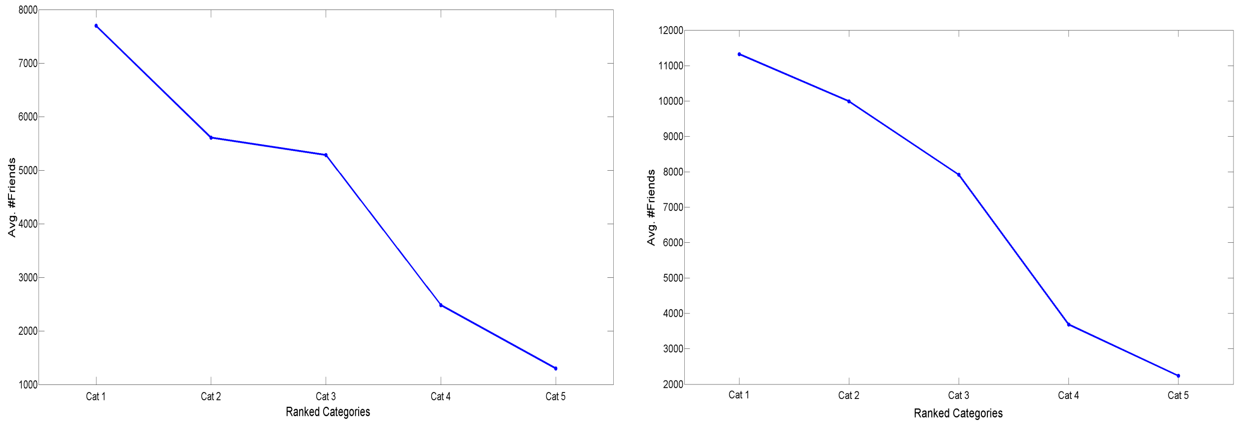
(A)

(B)

Figure 3.3. The distribution of average followers for the users of the tweets from the five categories in Table 3.1 for Hurricane Sandy (A) and Patricia (B).

retweets in Category 1 is very high compared with the other categories and has an average of ≈398 retweets per each tweet. Category 4 has the next highest number of retweets, but has a very low average of 4 retweets per each tweet. Similar patterns can be seen for Patricia as well (see Table 3.1).

We record the average number of followers and the average number of friends of the unique users in each category. In Figure 3, we plot the distribution of the average number of followers on Y-axis to the ranked categories on X-axis. Similarly, in Figure 4, we plot the distribution of the average number of friends. In these figures, the left plots are for Sandy and the right plots are for Patricia. We observe that the trend in all the plots is



(A)

(B)

FIGURE 3.4. The distribution of average friends for the users of the tweets from the five categories in Table 3.1 for Hurricane Sandy (A) and Patricia (B).

in decreasing order. As can be seen from the figures, for both the disasters, Category 1 has the highest average number of followers and friends, whereas Category 5 has the lowest corresponding averages. This analysis indicates the importance of the followers and friends in the retweet phenomenon of a tweet. Despite the different datasets' sizes of Sandy and Patricia, we observe that the trend is the same for both the disasters.

3.3.2. Popularity Analysis among Users

We further analyze the popularity of users in terms of two measures: (1) the retweets count of their tweets during the event, and (2) the verified status of the user. Both of these

TABLE 3.2. Top 5 users during Sandy with total retweets in this disaster and verified status.

User Type	**User id**	**Retweets**	**isVerified**
Celeb	justinbieber	137,599	true
Politician	GovChristie	38,177	true
News Media	cnnbrk	34,359	true
News Media	HuffingtonPost	34,019	true
Parodic	FillWerrell	32,984	false

TABLE 3.3. Top 5 users during Patricia with total retweets in this disaster and verified status.

User Type	**User id**	**Retweets**	**isVerified**
Astronomer	StationCDRKelly	88084	true
Politician	POTUS	20023	true
News Media	publimetros	10874	true
Parodic	BillNyeTho	11355	true
News Media	ajplus	9607	false

measures are important for the fast spreading of information in a network. For this analysis, from all of our 1.1M retweeted tweets of Sandy and 86K retweeted tweets from Patricia, we extracted the unique users who posted these initial tweets that were retweeted. We found that the number of unique users in Sandy and Patricia are 487,026 and 55,459, respectively. We ranked these users based on the retweets count of their tweets and observed that most of the top ranked users are related to news media, celebrities (such as actors and musicians), politicians, and a small fraction is related to regular or anonymous users. The inspection of the top ranked users also revealed that, for these top ranked users, there is a significant number of other users who participated in retweeting their tweets. Tables 3.2 & 3.3 show the top five ranked users of Sandy and Patricia, respectively, along with their verifiability of the user account (last column of each table).

For users' credibility, we used the "verified account" attribute from Twitter, which helps in establishing the authenticity of a user. In this aspect, we found an interesting pattern. From the list of users ranked based on the number of retweets, we selected the top 1000 and the last 1000 users and discovered that the accounts of the users with more retweets are verified accounts. Figures 3.5 & 3.6 show the distribution of these 2000 users (on X-axis) with the verified status (on Y-axis), which takes 1 if an account is verified and 0 otherwise. The figures correspond to Sandy and Patricia, respectively.

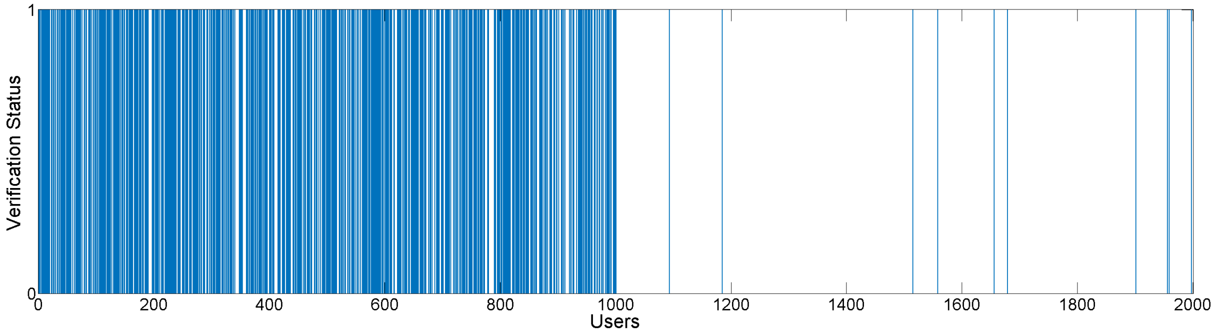


FIGURE 3.5. The distribution of users based on the verification status for top 1000 and last 1000 authors extracted for Hurricane Sandy.

In both the figures, as we descend to the users with less number of retweets, the verification status is faded off. The first half of the X-axis represents the users of the top

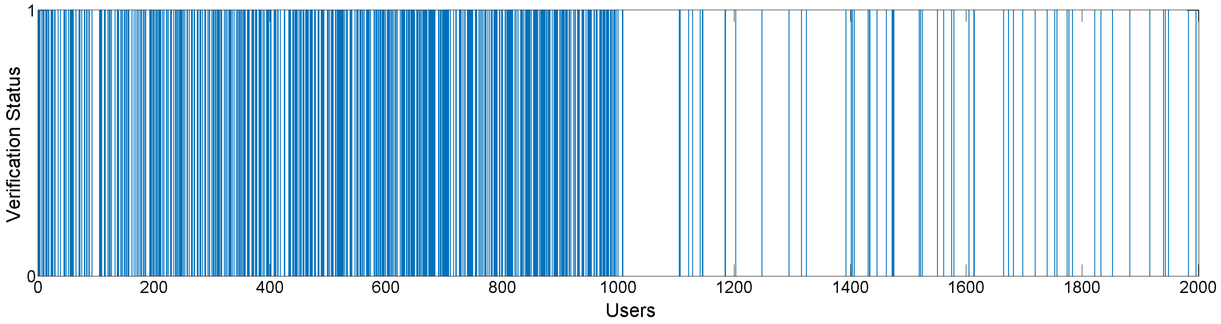


FIGURE 3.6. The distribution of users based on the verification status for top 1000 and last 1000 authors extracted for Hurricane Patricia.

1000 tweets with high retweets, where the density of blue bars is very high and the second half is for the users of the last 1000 tweets with only one retweet, where we see only a few blue bars. For Sandy, we found that there are around 584 verified users in the first 1000 users and only 11 verified users in the last 1000 users. Similarly, for Patricia, we found 506 verified users in the first 1000 users and 50 verified in the last 1000 users. Next, we present details of our retweetability classification task and show how factors related to user account information in the above analysis can affect the automatic prediction of tweets' retweetability.

3.4. Feature Engineering

In this section, we describe our features that we use as input to machine learning algorithms. We divide them into tweet content features (TC), user account features (U) and bag-of-words (BOW).

- *Tweet-content Features (TC)*: These features are extracted based on the tweet text. They are:
  - Contains Hashtag?: Hashtags are extremely relevant for the context of natural disasters because tweets from the same topic will likely contain the same hashtags. A user in search for information about a disaster may search for hashtags related to the particular disaster. We assign 1 if a hashtag is present in the

tweet, and 0 otherwise.

- Number of Hashtags: Not only it is important to verify the presence of a hashtag in a tweet, but also the number of hashtags may be important as well for retweetability. The more hashtags a tweet has, the more people can see it, thus increasing the chance of it being retweeted. The value of this feature is the number of hashtags.
- One-word sentences: We use OpenNLP[2] Java Libraries to check whether the tweet contains a one-word sentence. We assign a feature value of 1 for the presence of one-word sentences, otherwise 0.
- Multiple Sentences: Likewise one-word sentences, we also check for the presence of multiple sentences in the tweet and assign feature value 1 for its presence, otherwise 0.
- Presence of URL: URLs from news sources are important and are more likely to be shared because they provide a complete background about a natural disaster. The feature value is 1 for the URL presence, otherwise 0 is assigned.
- Is a Reply? A reply to a tweet usually indicates a conversation between users and is of more personal nature. We assign 1 if the tweet is a reply, otherwise 0.
- User mentions: We check if a tweet contains user mentions or not. We used regex

  s@.*

  s to check the pattern and assigned 1 as feature value if it is present, otherwise 0.
- Length of a tweet: We assign the string length as the feature value. If a tweet is very short, it might not contain useful information. In contrast, a longer tweet might contain useful information, and gets more retweets (Zhang et al. 2014).
- Phone Numbers: Tweets that contain phone numbers are likely to be shared by users because the phone number might be an emergency number or a donation

[2] https://opennlp.apache.org/

number. We assign feature value 1 for the presence and 0 otherwise.

- Measuring Units: In the case of natural disasters, there are many measurements involved, such as wind speed, flood depth. This information may be important for users participating in the disaster, and they may want to share it with other users.
- Date or Time: A disaster could last for days. Hence, dates may be important so that users can keep track of the progression of a disaster.

From the above features, we believe that the phone numbers, measuring units, date or time features are informative for disasters because they have the necessary vital information related to disasters, such as the magnitude of the disasters expressed in units, e.g., wind speed, water levels; phone numbers to inform or seek aid from the responders. In addition to the above feature, we extracted features that are derived based on certain word presences in the tweets. We manually parsed several random subsets of tweets in our set and went through several online resources to construct the dictionaries for each of these features as follows:

- Emoticons[3] : Emoticons are used in social networks to express emotions. We check for their presence and assign 1 for the presence or 0 otherwise.
- Cusswords: A tweet containing a cuss word indicates an informal way of expression, which may indicate no sign of useful information. We assign 1 for the presence of cuss words and 0 otherwise.
- Keywords: We manually went through several random subsets of tweets and selected a set of keywords such as "donate," "txt" and "pm" that exist in tweets. If a tweet contains a certain keyword, users will likely evaluate the tweet as useful for the natural disaster and will retweet this tweet.
- Abbreviations: Users commonly use abbreviations on Twitter, due to 140 characters limit. Abbreviations are a common way of expression, e.g., "LOL" which means "Laughing Out Loudly." We assign 1 for the presence of acronyms, and

[3] https://en.wikipedia.org/wiki/List_of_emoticons

0 otherwise. If a tweet contains abbreviations, it might be viewed as more informal. We selected most common abbreviations found in the tweets such as lol, lmfao, lmao, roflmao, etc., from slang lookup table available in the data folder of the SentiStrength project[4].

Using Stanford Named Entity Recognizer[5], we developed several other features. Given a text, Stanford NER outputs an inline XML string containing the labels for the entities. Named Entity is a reference to an entity or an object such as a person, organization or a location.

  - Contains Person Entity? We check for the presence of person entity and assign feature value 1 for the presence, otherwise 0.
  - Contains Location Entity? We check for the presence of location entity and assign feature value 1 for the presence, otherwise 0.
  - Contains Organization Entity? We check for the presence of organization entities such as NGO, NASA, and assigned feature value 1 for the presence, otherwise 0.

- *Normalized User Account Features*: These features are designed based on the attributes available in the user account information provided by Twitter. Generally, user account attributes listed below are informative for retweetability prediction, as shown in previous works [60, 74, 48]. However, these attributes do not provide information related to activeness of the users. Our intuition is to understand and quantify the activeness of a user, which plays a crucial role in the retweet phenomenon. For example, consider two users: A and B, whose Twitter account ages are 1000 days and 100 days, respectively, and each has posted 2000 statuses, and each has 1000 followers on their accounts. As a result, the rate at which user A posted is 2 tweets per day and for user B, it is 20 tweets per day. In this example, we can observe that user B is relatively more active than user A. Thus, tweets of user B

[4] http://sentistrength.wlv.ac.uk/

[5] http://nlp.stanford.edu/software/CRF-NER.shtml

will be more likely to get retweeted than the tweets of user A. Therefore, to extract the activeness, we normalized the following quantitative entities (except age and verification attribute) with the user account age:

  - Number of Friends: Friends are defined as all of the followers that a given user follows. The tweets of a user with more friends will likely be more retweeted.
  - Number of Followers: If a user has a big number of followers, his/her tweets will gain more visibility in the network. Since more people are visualizing the tweets, the probability of a tweet being retweeted increases.
  - Number of Favorites: A user with high favorites count indicates that other people like his/her tweets in general, so a tweet created by this user may be retweeted.
  - Number of lists a user belongs to: If a user is listed in multiple lists, he is connected and engaged with multiple communities. Consequently, the information that he posts is more likely to be seen by more people. Therefore, tweets made by this user may have a higher probability of being shared.
  - Verification: This is treated as a label of popularity. Users who are verified tend to post credible information, and this information is usually retweeted.
  - Status Count: This represents the number of statuses (tweets) posted by a user since the inception of the account. More statuses indicate active user, implying that there might be a chance of sharing his/her information.
  - Account age: If a user exists for a longer period in the network, he could potentially reach more people. Thus, information posted by this user would likely be retweeted. The feature value is the number of days since the creation of the account.

- *Bag of Words (BoW)*: A vocabulary is first constructed, which contains all unique words from the collection of training documents (tweets). Using this vocabulary, we use a binary representation to represent tweets as vectors. Specifically, we assign 1 for each word in the vocabulary if the word is found in the tweet, otherwise we assign 0.

3.5. Experiments and Results

In this section, we describe our experimental setup, and the results obtained using machine learning approaches for retweetability prediction. More precisely, our goal is to predict how likely a tweet is to be retweeted. We constructed several labeled datasets as follows: if a tweet was retweeted more than k times, then the tweet was labeled as positive. Otherwise, it was labeled as negative. For example, for a retweet threshold value k=1, we labeled a tweet as positive if the tweet was retweeted more than one time (the tweet has more than one retweet), and as negative otherwise. Since we are interested in predicting tweets that are likely to be highly retweeted on Twitter during disaster events, it is reasonable to label a tweet as negative, if the tweet is retweeted very few times. We performed experiments using various values of k, i.e., 0, 1, 2, 5, 20, 50, 75, 90 and 100, and show results for representative values of k = 0, 1, 5, and 20.

For evaluation, we show averaged results obtained using five disjoint train and test random splits, each containing 8000 tweets. Examples in each split are randomly sampled from all tweets in each disaster, depending on the retweet threshold. For example, to generate a training set for k=0, we randomly selected 4000 positive example from 1.1M retweeted tweets and 4000 negative examples from 6M non-retweeted tweets. Similarly, for the test set, we randomly sampled 2000 positive examples from 1.1M retweeted tweets (other than 4000 which are selected in the training set) and 6000 negative examples from the 6M non-retweeted tweets. Thus, the ratio of positive to negative class is 1:1 for the training set and 1:3 for the test set. For Patricia, we used the similar sampling method, but due to smaller dataset size, each of the train and test splits' size is reduced to 4000 tweets. This means each training set contains 4000 tweets (2000 positive examples and 2000 negative examples) and each test set contains 4000 tweets (1000 positive examples and 3000 negative examples). After the generation of the labeled datasets, we performed experiments with the following feature types: Tweet content features (TC), Normalized User account features (U) and Bag-of-words (BOW) (see Feature Extraction section for details). We report the average of the metrics: precision, recall, and F1-score. We used Naive Bayes and Support Vector Machines (SVM) classifiers to perform our experiments, and found that Naive Bayes performs better than SVM. Hence, we only show results for Naive Bayes.

We evaluated the quality of each feature type individually and then formulated several combinations from them. We found that not all feature combinations perform well on both classes, and hence, selected the following feature sets - TC, TC+U, TC+U+BOW, and Only BOW, which have good performance compared with the other combinations. In Tables 3.4 and 3.5, we show the results using these feature sets with several threshold values k for Sandy and Patricia, respectively. We discuss the results in terms of F1-score.

Table 3.4. The performance of Naive Bayes on Sandy data, for various retweet thresholds using TC, TC+U, TC+U+BOW and Only BOW. (P: Precision, R: Recall, F1: F1-score)

RT Threshold	**TC**			**TC+U**			**TC+U+BOW**			**Only BOW**		
	P	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**
Threshold_0	0.697	0.564	0.593	0.592	0.56	0.575	0.604	0.578	0.59	0.665	0.545	0.576
Threshold_1	0.657	0.564	0.592	0.672	0.731	0.68	0.68	0.729	0.69	0.659	0.597	0.62
Threshold_5	0.706	0.617	0.642	0.749	0.774	0.744	0.758	0.78	0.758	0.706	0.663	0.679
Threshold_20	0.729	0.64	0.663	0.787	0.802	0.783	**0.799**	**0.81**	**0.801**	0.742	0.695	0.711

3.5.1. Results Comparison on Feature Types

Among the results obtained for Sandy and Patricia, in Tables 3.4 and 3.5, the feature set TC+U+BOW gives the best performance compared with all other feature sets. Moreover, the best F1-score is achieved by TC+U+BOW for retweet threshold k=20. When we added normalized user features (U) to the tweet features (TC), the F1-score is increased from 0.663 to 0.783 (18% relative increase) and adding BOW to TC+U, boosted it to 0.801 (20% relative increase over TC). We can see that the conjunction of user account features and BOW with tweet content features improved the classifiers' performance for k=20. This suggests that the user account features, BOW and tweet content features are collaboratively assisting each other in boosting the classifier's performance.

Table 3.5. The performance of Naive Bayes on Patricia data, for various retweet thresholds using TC, TC+U, TC+U+BOW and Only BOW. (P: Precision, R: Recall, F1: F1-score)

RT Threshold	**TC**			**TC+U**			**TC+U+BOW**			**Only BOW**		
	P	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**
Threshold_0	0.698	0.608	0.633	0.742	0.77	0.732	0.752	0.772	0.755	0.708	0.588	0.615
Threshold_1	0.703	0.608	0.634	0.785	0.788	0.74	0.772	0.79	0.763	0.711	0.617	0.642
Threshold_5	0.719	0.612	0.638	0.821	0.808	0.768	0.825	0.821	0.791	0.727	0.652	0.673
Threshold_20	0.722	0.618	0.643	0.847	0.836	0.81	**0.854**	**0.847**	**0.827**	0.733	0.664	0.684

3.5.2. Results Comparison on Retweet Threshold (k)

As we can see from the tables, the performance of the classifiers is increasing with the increase in the retweet threshold value. In Tables 3.4 and 3.5, we reported the results for only 0, 1, 5 and 20, for which the performance is significantly improved when the threshold is increased. For example, the F1-score of TC+U feature set in Table 3.4 increases from 0.575 (for threshold k=0) to 0.744 (for threshold k=5) and increases further to 0.783 for k=20.

3.5.3. Comparison with previous works

We compared the results obtained using our best-performing feature set TC+U+BOW with the results of the features in Petrovic et al. [48] denoted by *Pt* and Suh et al. [60] denoted by *Sh*. These two works have the features extracted from the tweet text and the user account information. Both of these works have very similar tweet features and user account features. The tweet-based features are: URL (*Pt* & *Sh*), # hashtags (*Pt* & *Sh*), # mentions (*Pt* & *Sh*), is a reply? (*Pt*), # retweets (*Sh*) and BOW (*Pt*); and the social features are: # friends (*Pt* & *Sh*), # followers (*Pt* & *Sh*), # favorites (*Pt* & *Sh*), # listed (*Pt*), user account age (*Sh*), statuses (*Pt* & *Sh*), and verified status (*Pt*). In parenthesis, we show the work to which the feature belongs. In our work, we have additional features, which are described in the Feature Extraction section. We used the normalized values of the user account features instead of using the actual values.

In Table 3.6, we show the results comparison of our best performing feature set

Table 3.6. Performance comparison of our work with previous works.

Features	**Class**	**Hurricane Sandy**			**Hurricane Patricia**		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score
Our model	Positive	0.213	0.256	0.233	0.568	0.375	0.452
	Negative	0.734	0.686	0.709	0.813	0.905	0.856
Petrovic et al. (2011)	Positive	0.199	0.227	0.212	0.717	0.158	0.259
	Negative	0.729	0.695	0.712	0.777	0.979	0.867
Suh et al. (2010)	Positive	0.19	0.25	0.215	0.676	0.103	0.179
	Negative	0.72	0.645	0.68	0.767	0.984	0.862

TC+U+BOW with the feature sets of Petrovic et al. [48] and Suh et al. [60], and evaluate them using the train and test sets for retweet threshold 0 (since the data labeling in the two previous works is based on the retweet threshold 0). We report the metrics: precision, recall and F1-score for positive and negative classes in Table 3.6. For both the disasters, we observe that our feature set - TC+U+BOW performs better than the feature sets in Petrovic et al. and Suh et al. in terms of F1-score on the positive class, whereas our feature set performs on par or slightly worse compared with both previous works on the negative class. Hence, our model is more successful than previous works at predicting a retweetable tweet.

3.5.4. Effect of Normalized User Features

We investigate the effect of normalized user features on the classifier's performance. Our idea of normalizing the quantitative entities of user account information with user account age is to quantify the user activeness, which is a helpful factor in the retweet phenomenon. For our best-performing feature set -TC+U+BOW, two versions of the user account features (U) are used - one is normalized, and the other is without normalization.

In Table 3.7, we show the results (precision, recall and F1-Score) of positive and negative classes for TC+U+BOW. In both the disasters, we can see that the negative class is performing good and have a slight variation among the normalized and un-normalized user features. However, in this work, it is important to predict the retweetable tweet correctly. Hence we are interested in the performance of positive class, particularly. The performance

Table 3.7. The performance comparison between Un-normalized User features and Normalized User features. (P - Precision, R Recall, F1 F1-score)

RT Threshold	Class	Sandy						Patricia					
		Un-normalized			Normalized			Un-normalized			Normalized		
		P	R	F1	P	R	F1	P	R	F1	P	R	F1
Threshold_0	Positive	0.202	0.232	0.216	0.213	0.256	0.233	0.729	0.171	0.278	0.568	0.375	0.452
	Negative	0.731	0.694	0.712	0.734	0.686	0.709	0.78	0.979	0.868	0.813	0.905	0.856
Threshold_1	Positive	0.445	0.18	0.257	0.409	0.19	0.259	0.798	0.203	0.323	0.662	0.328	0.439
	Negative	0.772	0.925	0.842	0.771	0.908	0.834	0.787	0.983	0.874	0.808	0.944	0.871
Threshold_5	Positive	0.606	0.349	0.443	0.601	0.354	0.446	0.856	0.313	0.458	0.85	0.343	0.489
	Negative	0.81	0.925	0.863	0.811	0.922	0.863	0.811	0.982	0.888	0.817	0.98	0.891
Threshold_20	Positive	0.674	0.493	0.569	0.659	0.5	0.569	0.894	0.418	0.569	0.896	0.441	0.591
	Negative	0.845	0.921	0.881	0.846	0.914	0.879	0.835	0.983	0.903	0.841	0.983	0.906

of positive class in Sandy dataset shows minute improvements when normalized user features are used instead of un-normalized user features.

For Patricia, the positive class performance of normalized user features is better compared with the performance of un-normalized user features. These results prove that the normalization of the user attributes indeed helps in predicting the retweetability of a tweet. It is worth noting that for the Hurricane Sandy, some of the user details are missing, and hence, normalization may not help much, whereas for Hurricane Patricia, all user details are available.

3.5.5. Feature Selection

In order to find out which features are more informative in the model construction, we performed feature selection using the Weka toolkit[6]. The features are ranked according to their Information Gain. We performed this experiment on the data which gave us the best model, which is trained using TC+U+BOW with retweet threshold k=20. We present, in Table 3.8, the top 10 features of 3237 features in Sandy and of 1633 features in Patricia. We observe that the features belonging to user account feature type are top ranked. For

[6] http://www.cs.waikato.ac.nz/ml/weka/

both the disasters, the first four ranked features are the same, and after that, the ranking is different. From BOW, we observe that the words which are the disaster name hashtags: "#sandy" and "#patricia," are ranked among the top 10 ranked features, indicating the importance of the hashtags which are named after the events.

Table 3.8. Top 10 Ranked Features of TC+U+BOW for threshold 20.

Rank	**Sandy**	**Patricia**
1	Followers Count (U)	Followers Count (U)
2	Is Verified (U)	Is Verified (U)
3	Listed Count (U)	Listed Count (U)
4	Age (U)	Age (U)
5	Tweet Length (TC)	Status Count (U)
6	Contains Emoticon(TC)	Tweet Length (TC)
7	"#sandy" (BOW)	No. of hashtags (TC)
8	URL (TC)	Contains Hashtags (TC)
9	Status Count (U)	Friends Count (U)
10	No. of hashtags(TC)	"#patricia" (BOW)

This feature ranking shows that the user account features are more informative and important for achieving a promising model to predict the retweetability of a tweet. Now, we briefly discuss the first 3 user features. Followers count is ranked as the first most informative; this justifies the fact that followers are important means of retweeting a tweet. Listed Count specifies the number of groups or communities in which a user is listed in, which shows the active participation of a user, and hence, a higher chance of having his tweets highly retweeted. Verified accounts on Twitter show the authenticity of a user as a famous personality or as a well-known organization. As can be seen from the Figures 3.5 & 3.6 in Data Analysis section, the users of the highly retweeted tweets are mostly verified. Hence, verifiability is a highly informative feature for the classifier in predicting the retweetability of a tweet.

3.6. Summary and Discussion

In this chapter, we studied the problem of predicting the retweetability of a tweet in the context of disaster events. We used the tweets posted during the Hurricane Sandy in 2012 and Hurricane Patricia in 2015. The strongest contribution of this work is the design and exploration of features for training machine learning classifiers that can predict how likely a tweet is to be highly retweeted on Twitter. Unlike the features used in previous works (e.g., the number of retweets as in Suh et al. [60]), our features are not dependent on the retweet phenomenon. Our features are extracted from the tweet text and the user account information only. We developed models that automatically predict the retweetability of a tweet and found that classifiers trained on tweets' content features (TC), normalized user account features (U) and bag-of-words (BOW) together outperform those trained using solely the "bag of words."

The results of our experiments using different threshold values for labeling a tweet as being retweetable show improved performance for classifiers trained using TC+U+BOW over classifiers that are trained on each feature type independently. We then compared our best feature set TC+U+BOW with the features in Petrovic et al. [48] and Suh et al. [60], and the results indicate that our feature set performed better than the features in the other two works. We also found that normalizing user account attribute values with the user account age help in leveraging the activeness of a user, which is useful for predicting how likely a tweet can be retweeted.

3.6.1. Related Work

In Twitter, users can create tweets, i.e. posts that must not exceed 140 characters, and can share any public tweets in the network. This process of sharing a tweet is called "Retweeting." When users share tweets, all of the users' followers will be able to see it. Several research groups have demonstrated that emergency managers and responders understand the value of social media for crisis communication (see [16, 28, 13]). In addition, there have been several studies of emergency managers and responders who have used social media to get the word out during a crisis (see [9, 17, 53, 62]). More directly, there have been several research efforts to understand how emergency managers and responders have tried to influence the public's information or behavior via social media during crises (see [18, 63]). With the aforementioned research efforts, and with the limitation of only 140 characters per message, there is a strong agency for developing predefined terse messages to be used during a given crisis [63, 64].

Much work is done on how information is propagated through a network. For example, Sutton et al. [62] studied the effect of centrality on the dissemination information, and how this feature allows a certain organization to broker said information. Kwak et al. [25] conducted a quantitative study on Twitter data to find how information is diffused in the network. They suggested that the number of followers a user has and the number of times that a user's tweet is retweeted are different measures of popularity. Olteanu et al. [42] have studied the propagation of information in crisis situations using statistical analysis and have shown that different disasters contain similar tweets, and human-induced disasters are more analogous to each other than to natural disasters. Also, it was verified that tweets containing keywords related to a disaster and tweets by local media and emergency agencies are critical sources of information. Starbird and Palen [57] studied the information propagation in Twitter during Red River floods and Oklahoma Fires and found that people are more likely to use the retweet function to pass on crisis-related information than other types of information during a crisis event. Hochreiter et al. [14] have applied a genetic algorithm to improve message style and optimize a tweet composition for increasing the reach of a message. They found that the retweetability of the optimized tweet had increased significantly. Pervin et al. [47] studied the factors affecting the retweetability using Japan Earthquake Twitter data and observed that network features such as the type of user sharing a piece of information are very crucial for the propagation of the information.

Zaman et al. [74] developed a probabilistic model to predict a retweet given the tweet content, tweeter and retweeter. They found that features such as the name, number of retweet-followers and number of retweet following of the author of a given tweet were important. Suh et al. [60] found that the context of a tweet author (such as age, followers, and friends) influenced the retweetability. They also stated that tweets with URLs and hashtags were more likely to be retweeted. The authors developed a Generalized Linear Model to predict the retweetability. Petrovic et al. [48] addressed the problem of predicting retweetability in Twitter. They have shown that social features, i.e., the number of followers, friends, statuses, favorites, the number of times that a user was listed and verified status, play a major role in increasing the accuracy of the prediction. Uysal and Croft [70], have proposed methods to rank tweets using retweet behavior in order to bring more important tweets forward and also determined the audience of tweets by ranking users based on their likelihood of retweeting the tweets. Starbird and Palen [57], have performed statistical analysis on 2011 Egyptian uprising and showed that information diffusion is mostly due to the retweets. They found that during political events, the work of activism is accomplished by both local crowds through expressing social solidarity and through individual activists. Probably the work by Jenders et al. [24] is the most similar to ours. Instead of predicting a tweet which can be retweeted more than a threshold, we treated the problem as a binary classification problem and solved for multiple thresholds. We additionally have features that are designed based on the numbers in the tweet, such as phone numbers, measuring units, dates which are useful during disasters, and are not present in Jenders et al. [24] .

3.6.2. Discussion

Through this research, we explored a wide range of features and studied the predictive power of the features. We discovered that the features extracted from the tweet content (TC) and user details (U) are informative for predicting the retweetability. In particular, the user features have shown interesting results and suggest that user attributes are much informative in modeling the retweetability prediction, since the user network aspects (such as # followers, # friends, # listed) are important in spreading the information. Interesting directions for future work include predicting the number of retweets for a tweet. One more interesting future direction would be to integrate our approach with the approaches that identify trustworthy and misinformation in Twitter, would have the potential to help to promote useful information, which helps to flag those messages which are non-informative, but have a higher chance of being retweeted. Another direction would be to predict which of the retweetable tweet is actionable for the victims and responders.

CHAPTER 4

A FEATURE BASED APPROACH TOWARDS IDENDTIFYING INFORMATIVE TWEETS DURING DISASTERS

In recent years, micro-blogging services such as Twitter and Facebook have emerged as effective tools for broadcasting messages worldwide during disaster events. With millions of messages posted through these services during such events, it has become imperative to identify valuable information that can help the emergency responders to develop efficient relief efforts and aid victims. In this chapter, we focus on understanding what features are useful for identifying messages that convey information relevant to a disaster in the Twitter platform. In particular, we design various feature sets (based on tweet content, user details and polarity clues) and study the predictive power of these feature sets individually or in various combinations, using Naive Bayes classifiers. Moreover, we explore how our features generalize across different disaster types by developing models trained on one type (e.g., natural disasters) and evaluating them on another type (e.g., non-natural disasters). In addition, we perform an analysis to show how diversified emotions in a tweet affect the informativeness of the tweet. We find that tweets with high diversified emotions are more likely to be non-informative, whereas tweets with low diversified emotions are more likely to be informative.

4.1. Introduction

There is a growing body of research work on how to leverage micro-blogging information from crowds of non-professional participants during disasters. Data produced through micro-blogging is seen as ubiquitous, rapid and accessible [71], and it is believed to empower average citizens to become more situationally aware during disasters and coordinate to help themselves [46]. In disasters that have occurred recently in developed environments, average citizens offered ground-level information describing the local specifics of the disaster, keeping outsiders and emergency response organizations informed about the ground realities. In a survey performed by the Pew Research Center, it was found that Twitter served as a major source of information during Hurricane Sandy, with more than 30% of the analyzed Twitter data falling under the news and information category.

Despite the evidence of strong value to those experiencing the disaster and those seeking information concerning the disaster, there is still very little uptake of message data by large-scale, disaster response organizations[65, 67]. Response organizations operate under conditions of extreme uncertainty. The uncertainty has many sources: the sporadic nature of emergencies, the lack of warning associated with some forms of emergencies, and the wide array of responders who may or may not respond to any one emergency. This uncertainty increases the need for extracting appropriate information from streaming data, which could make substantial improvements in the response process. However, the amount of online streaming data is tremendous. For example, in the above-mentioned survey by the Pew Research Center, it is stated that more than 20 million tweets were posted during the five-day interval of Hurricane Sandy. Due to the sheer amount of data, it is extremely difficult and time-consuming to manually sift through these data and identify valuable informative messages. We believe that data directly contributed by citizens and data scraped from disaster bystanders have a huge potential to give responders more accurate and timely information than it is possible with traditional information gathering methods. Still, information quality and use in any area of disaster response remain as challenges.

Hence, one question that can be raised is: *In a social media stream of messages, what are the features that can help identify messages that convey information relevant to a disaster?* We specifically address this question with our research agenda in this chapter using the knowledge gained from the previous feature sets of the sentiment and retweetability, and with a Twitter dataset constructed by Olteanu et al. (2015).

4.1.1. Contributions and Organization

We present a supervised learning approach with a wide range of feature exploration to identify informative messages (or tweets) from those that are not-informative in nature (i.e., pertaining to user feelings, informal communication or casual conversations). We define "informative" tweets as any tweets, which would provide valuable, concrete information to anybody viewing the tweets. Our approach is based on a combination of "bag-of-words" features, which are typically used for text classification, and features that are extracted from tweets' content (e.g., URLs, hashtags, emoticons, slang), user details (such as number of friends, number of followers) and polarity clues (such as positive words, negative words). In summary, our contributions are:

- We propose an approach that combines "bag-of-words" features and features extracted from tweet content, user details, and polarity clues, for learning models that identify informative tweets during disaster events, and study the predictive power of our features.
- We show experimentally that models trained using the above combination of features perform better than models trained on each feature type independently (i.e., either "bag-of-words" or features extracted from the tweet content, user details, or polarity clues).
- We study how well our features generalize across different disaster types (e.g., natural and non-natural) by developing models trained on one disaster type (such as natural disasters) and evaluating them on another disaster type (such as non-natural disasters). We also investigate how the emotional divergence affects informativeness.

Our study provides evidence to the claim that proper analysis of streaming data may lead to applications able to help not only the disaster response providers to allocate resources more efficiently, but also the victims who are in need and seeking support, by filtering out necessary informative messages for immediate availability. The problem of finding relevant information as a disaster evolves faces many challenges since extracting informative tweets and filtering out those that are non-informative has to be done in real time and with high accuracy.

The remaining of this chapter is organized as follows: We first discuss the Twitter dataset and present the features used for identifying informative tweets from a stream of messages, and then we discuss our results and conclude this chapter.

Table 4.1. Summary of disasters used in the experiments.

S. No.	Non-natural Disasters	Natural Disasters
1.	Colorado Wildfires (2012)	Costa Rica Earthquake (2012)
2.	Australia Bushfires (2012)	Gautemala Earthquake (2012)
3.	Venezuela Refinery (2012) Italy	Earthquake (2012)
4.	Boston Marathon Bombings (2013)	Bohol Earthquake (2013)
5.	Brazil Night Club Fire (2013)	Typhoon Pablo (2012)
6.	Glasgow Helicopter Crash (2013)	Typhoon Yolanda (2013)
7.	LA Airport Shootings (2013)	Alberta Floods (2013)
8.	Lac Megantic Train Crash (2013)	Colorado Floods (2013)
9.	NY Train Crash (2013)	Philipinnes Floods (2012)
10.	Savar Building Collapse (2013)	Manila Floods (2013)
11.	Spain Train Crash (2013)	Queensland Floods (2013)
12.	Singapore Haze (2013)	Sardinia Floods (2013)
13.	West Texas Explosion (2013)	Russia Meteor (2013)

4.2. Dataset

We used a dataset constructed by Olteanu et al. [42]. This collection contains tweets from 26 disasters that occurred during 2012 and 2013, which are manually annotated by crowd-sourced workers with the labels: informative, not informative, and not related to the disaster. There are about 1000 tweets manually annotated in each of the 26 disasters. Among the 26 disasters, there are 13 non-natural disasters and 13 natural disasters. We consider a natural disaster as any catastrophic event caused by nature or natural process of the earth (e.g., cyclones, earthquakes, floods), whereas a non-natural disaster as an event caused by human actions which may be intentional (e.g., gun shootings, bombings) or indirect which leads to technological failures (e.g., industrial accidents, train crash). Table 4.1 shows the 26 disasters and their type. In these tweets, we found that some of the tweets were posted in languages other than English based on the language attribute available from Twitter. We removed the tweets which are not-related, non-english from the dataset and ended up with 7200 tweets related to non-natural disasters and 6480 tweets related to natural disasters.

4.3. Feature Engineering

Next, we describe our features that we use as input to Naive Bayes classifier available in Weka toolkit[1]. We divide them into four sets, namely bag-of-words, tweet content features, user details features and polarity features.

- *Bag-of-Words (BoW)*: A vocabulary is first constructed, which contains all unique words from the collection of training documents (tweets). Using this vocabulary, we use a binary representation to represent tweets as vectors. Specifically, we assign 1 for each word in the vocabulary if the word is found in the tweet, otherwise we assign 0.
- *Tweet Content Features (TC)*: We designed these features based on several aspects of the content of a tweet, as follows:
  - Presence of a URL: Checks whether URL is present or not.
  - Presence of Hashtags: Presence of “#hashtags” in a tweet.
  - Hashtag Count: Number of occurrences of hashtags in a tweet.
  - Emoticons: Presence of emoticons in a tweet, highly representative of conversational tweets.
  - Instructional keywords: Presence of instructions words such as “text,” “call,” and “donate.”
  - Phone Numbers: Presence of a phone number in the tweet, using regex to match phone number patterns (e.g., \d3-\d3-\d4 matches the phone numbers like xxx-xxx-xxxx).
  - Internet Slang: Presence of abbreviations and slang such as “OMG” (Oh My God), highly representative of informalities and mostly occurring in conversational tweets.
  - Is a Retweet (RT)?: Checks for the presence of the pattern “RT @” or “RT@” in the tweet.
  - Profanity - Checks for the presence of informal/cuss words.

[1] http://www.cs.waikato.ac.nz/ml/weka/downloading.html

  - Sentence Structure: We use OpenNLP Java Libraries to check whether the tweet contains a one-word sentence. We assign a feature value of 1 for the presence of one-word sentences, otherwise 0. Likewise, we also check for the presence of multiple sentences in the tweet and assign feature value 1 for multiple sentences, and otherwise 0.

- *User Details Features (UF)*: From the Twitter user attributes (namely, friends, followers, verified, lists, statuses, etc.), we use the attributes below as our features.
  - Followers count: Total number of followers of a user.
  - Friends count: Total number of friends of a user.
  - Favorites count: Total number of favorite tweets in the user account.
  - Listed Count: Total number of communities that the user is listed in.
  - Statuses Count: Total number of tweets posted by the user.
  - Verified account: Twitter follows a procedure to verify the authenticity of the users who are famous personalities, brands, etc. Each user account is assigned with a special emblem, if it meets the verification criteria of Twitter. The possible value for this feature is 1 (if verified account) and 0 (if not verified).

- *Polarity Features (PF)*: These features are formulated based on the polarity (positive or negative) of the words in a tweet. We identify the polarity words using lexicons of positive and negative words created by Hu and Liu (2004) and compute the following features:
  - Positive word count: Number of positive words in a tweet.
  - Negative word count: Number of negative words in a tweet.
  - Positive Score: Positive score returned by the SentiStrength algorithm.
  - Negative Score: Negative score returned by the SentiStrength algorithm.
  - Emotional Divergence: We adopt the definition of “emotional divergence” (ED) from Pfitzner et al. (2012). ED is defined as “the (normalized) absolute difference between the positive and the negative sentiment score delivered by SentiStrength. It is calculated using the formula ED =(p-n)/10, where p is a

positive score and n is a negative score output by the SentiStrength algorithm. This feature is an aggregate of the above four polarity features.

4.4. Experiments and Results

We evaluate the performance of models trained using the above features on both natural disasters (e.g., earthquakes and floods), and non-natural disasters (e.g., fire accidents and train crash), as well as across disaster types. Last, we show how emotional divergence in a tweet affects the informativeness of the tweet.

4.4.1. Comparison among Feature sets without BoW

We first contrast the feature sets described above in order to understand what feature set or feature combinations are most predictive in identifying messages that convey information relevant to a disaster. Thus, we compare the results of experiments obtained using different feature sets and feature combinations, using 10-fold cross-validation experiments. We experimented with Naive Bayes, Support Vector Machine, and Random Forest classifiers and the following features and feature combinations: individual feature types (namely TC, UF, and PF), bag-of-words (0/1 representation of tweets) and the combination of bag-of-words with various feature sets. For our experiments, we separated all tweets from the 26 disasters of CrisisLex based on the disaster type and formed two subsets: one consisting of 7200 tweets from non-natural disasters and another one consisting of 6480 tweets from natural disasters. In order to have an equal amount of data for training in each subset and remove the sample size bias, we randomly sampled a subset of 6400 tweets from each subset, of which 5000 are informative tweets and 1400 are non-informative.

We show only the results for Naive Bayes classifiers, which yield better results than SVM and Random Forest. We report F1-score of the classifiers on informative and non-informative classes and their average. Table 4.2 shows the results of the classifiers evaluated using 10-fold cross-validation with various feature set combinations for both disaster types, natural and non-natural. Among individual feature sets, the classifiers trained using only TC perform better than using only UF and only PF, on both disaster types. For example, the average F1-score of only TC for non-natural is 0.776, which is much better than 0.583 (for only UF) and 0.742 (for only PF). When user features (UF) are added to any feature set, the performance does not improve for any disaster type, which indicates that the user features are not informative and that the informativeness of a tweet is independent of the user. Overall, we observe the best performance for TC+PF among our feature set combinations. When polarity features are added to TC, we observe that the classifiers' performance improves, implying that these two feature sets are mutually assisting each other to boost classifier performance. For example, for the non-natural disaster type, the average F1-score is increased from 0.776 (for only TC) to 0.789 (for TC+PF), which is 1.68% increase.

Table 4.2. The performance (F1-score) of the classifiers with 10-fold cross-validation setting using various feature sets.

Disaster type	**Class**	**Only TC**	**Only UF**	**Only PF**	**UF+PF**	**TC+UF**	**TC+PF**	**TC+UF+PF**
Non-Natural	Info.	0.841	0.656	0.867	0.801	0.808	0.87	0.837
	Non-Info.	0.555	0.323	0.297	0.39	0.539	0.5	0.577
	Avg.	0.778	0.583	0.742	0.711	0.749	0.789	0.781
Natural	Info.	0.835	0.195	0.855	0.279	0.679	0.865	0.718
	Non-Info.	0.435	0.369	0.341	0.375	0.47	0.457	0.494
	Avg.	0.748	0.233	0.743	0.3	0.633	0.776	0.669

4.4.2. Comparison among the combinations in conjunction with BoW

In Table 4.3, we show the results of the comparison for the classifiers trained using best-performing feature set combinations (TC, TC+UF, TC+PF) in conjunction with BoW. When tweet content features are combined with BoW, the classifier performance is similar to Only BoW. Adding UF to BoW+TC, the classifier performance is degraded and when PF is added to BoW+TC, the performance is slightly improved in comparison to only BoW. Overall, from the results in Table 4.2 and Table 4.3, we find that among our feature sets, user features are not useful for identifying the informative tweets, and TC and PF are informative features for this classification task. Next, we investigate if the feature sets or their combination will result in models that will generalize well from one type of disaster to another (e.g., from natural to non-natural disasters). More precisely, we use one disaster type for training and the other disaster type for testing. Table 4.4 shows the results of these experiments with the feature sets and BoW.

Table 4.3. The performance (F1-Score) of the classifiers with 10-fold cross-validation setting using various feature sets and BoW.

Disaster Type	**Class**	**Only BoW**	**BoW+TC**	**BoW+TC+UF**	**BoW+TC+PF**
Non-natural	Info.	0.903	0.902	0.894	0.905
	Non-Info.	0.704	0.703	0.696	0.710
	Avg.	0.859	0.858	0.850	0.862
Natural	Info.	0.900	0.901	0.880	0.902
	Non-Info.	0.681	0.688	0.670	0.688
	Avg.	0.852	0.855	0.834	0.855

4.4.3. Domain Adaptation

We compare these results in Table 4.4 with the 10-fold CV of the test set, since in cross-validation all of the examples are used at least once in testing. As can be seen from the table, the classifier trained using all natural disasters data (denoted as N_{Train}) and tested on all non-natural disasters data (denoted as NN_{Test}), shows similar performance to the classifier evaluated using 10-fold cross-validation on all non-natural data. When trained on all non-natural disaster data (denoted as NN_{Train}) and tested on all natural disaster data (denoted as N_{Test}), the classifier performance is worse than that of all natural disaster 10-fold cross-validation. In terms of features, when trained and tested on opposite disaster types (e.g., NN_{Train}/N_{Test}), we observe that classifiers trained using feature sets in conjunction with BoW are performing better than the classifiers trained on only BoW, for example in $NN_{Train}/\ N_{Test}$, the average F1-score is 0.707 (for Only BoW) and is 0.725 (for BOW+TC+PF) which is a 2.55% increase.

Overall, we find that the features extracted from natural disasters can be used for developing models that could predict all disaster types (non-natural or natural), whereas non-natural disasters do not generalize well for natural disasters. We investigate why the latter

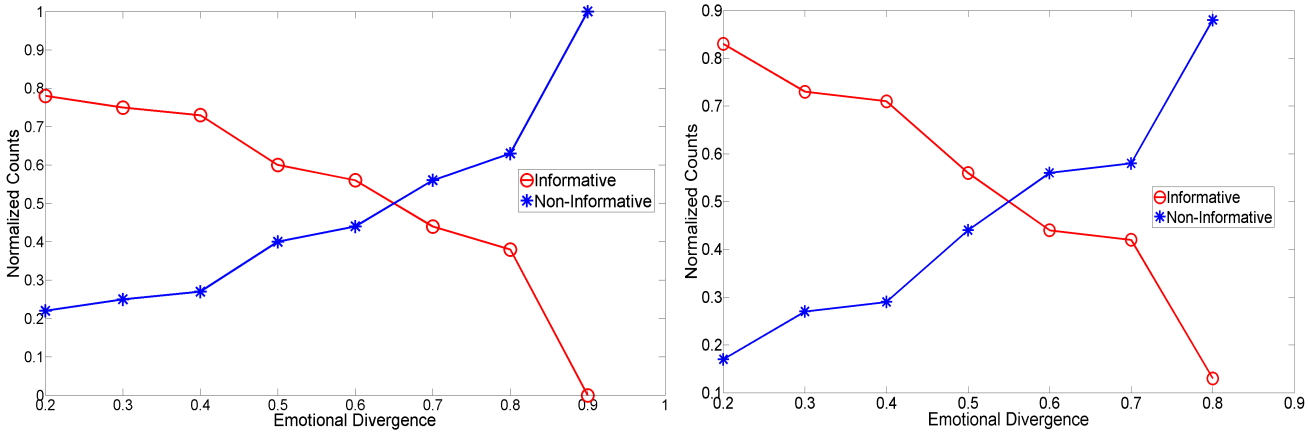
TABLE 4.4. Summary of results for cross-domain experiment with Train-on-all/Test-on-All strategy. (10 CV- 10 fold Cross Validation; F1-scores reported)

Disaster Type	**Class**	**Only BoW**	**Only TC**	**BoW+TC**	**BoW+TC+PF**
Non-natural 10 CV	Info.	0.903	0.841	0.902	0.905
	Non-Info.	0.704	0.555	0.703	0.710
	Avg.	0.859	0.778	0.858	0.862
N_{Train}/NN_{Test}	Info.	0.891	0.871	0.894	0.897
	Non-Info.	0.656	0.446	0.666	0.669
	Avg.	0.839	0.778	0.844	0.847
Natural	Info.	0.900	0.837	0.902	0.902
	Non-Info.	0.681	0.441	0.687	0.688
	Avg.	0.852	0.751	0.855	0.855
NN_{Train}/N_{Test}	Info.	0.753	0.817	0.768	0.773
	Non-Info.	0.540	0.497	0.546	0.555
	Avg.	0.707	0.747	0.719	0.725

scenario happened by comparing the examples which are misclassified in NN_{Train}/N_{Test}, but correctly classified in Natural 10-fold cross-validation. We noticed that the hashtags usage in natural disasters tweets is greater than the usage in non-natural disasters tweets. For example, in our natural disaster subset, there are 204 unique hashtags with occurrences accounted up to 6174, for non-natural disasters it is 154 unique hashtags with 3758 occurrences. One more observation we observed is that the non-natural disasters tweets have a lot of slang words and very fewer slang words usage seen in natural disaster tweets.

4.4.4. Emotional Divergence vs. Informativeness

In this section, we analyze the impact of emotional divergence on the informativeness of the tweets from CrisisLex. As can be seen from the experiments it is observed that the classifiers' performance spike when polarity features are added at some points. To understand the correlation between the polarity clues present in a tweet and the informativeness, we used the emotional divergence to capture the polarity clues present in the tweets and analyze the distribution with respect to the informative and non-informative tweets. We separated these tweets into non-natural and natural disasters. For these tweets, we calculated ED values and recorded the total number of the tweets that are distributed in each ED value. For each ED value, we calculated the number of tweets which are informative and non-informative, and then normalized them with the total number of tweets distributed in each ED value.



(A)

(B)

Figure 4.1. Emotional Divergence vs. Informative/Non-informative for Non-natural disasters (A) and Natural disasters (B).

Figure 4.1 shows the plot between emotional divergence and normalized counts of the tweets from non-natural (left) and natural (right) disasters. We observe a similar trend for both natural and non-natural disasters. In the figure, the red curve represents the informative tweets, and the blue curve represents the non-informative tweets. As can be seen, the two curves show opposite trends. The curve for informational tweets shows a decreasing trend: as the emotional divergence is increasing, the normalized count value is decreasing, whereas for non-informative, the trend is opposite: as the ED value is increasing, the normalized count is also increasing. This implies that for low ED values, the normalized counts value of informative tweets is higher than that of non-informative tweets. Similarly, for high ED values, the normalized counts of non-informative tweets is higher than that of informative tweets. This suggests that the chance for the tweets with low ED values to be informational is higher, whereas the chance for the tweets to be non-informative is higher at high ED values.

Table 4.5. Classifiers' performance using feature selection for both disaster types.

Deleted Feature	**Non-natural**	**Natural**
None Deleted	0.778	0.751
Contains hashtag	0.778	0.756
Hashtag count	0.778	0.753
Internet slang	0.778	0.75
One-word sentence	0.778	0.747
Multiple sentences	0.777	0.739
URL	**0.736**	**0.702**
Phone number	**0.738**	**0.721**
Emoticons	**0.721**	**0.703**
Is a retweet?	0.778	0.751
Keywords	0.777	0.726
Profanity	0.778	0.751

Table 4.6. Feature Rankings based on Information Gain for TC for both disaster types.

Rank	**Non-natural**	**Natural**
1	URL	URL
2	Emoticons	Emoticons
3	Phone number	Phone number
4	Multiple sentences	Hashtag count
5	Internet slang	Keywords
6	Hashtag count	Contains hashtag
7	Contains hashtag	Internet slang
8	One-word sentence	Multiple sentences
9	Profanity	One-word sentence
10	Keywords	Profanity
11	Is a retweet?	Is a retweet?

4.4.5. Feature Selection

From the above experiments, it is evident that the user features are not informative for this classification task and we analyzed the effect of polarity features on identifying informative tweets in the previous section. In this section, we study which features among the tweet content features are more informative in the model construction. First, we perform feature selection by training models with one feature removed at each iteration and record the F1-score. This gives the information about the effect of the features on classifier performance. Table 4.5 shows the results of feature selection on TC features for both disaster types. The first row is for none of the TC features deleted. From row two, onwards, the presented feature is removed and the remaining features are used for training. As can be seen from the table, for both disaster types, we observe that the classifiers' performance is severely dropped when these features are removed - URL, phone number, and emoticons. In addition, we performed feature rankings for TC using the Weka toolkit.

The features are ranked according to their Information Gain. We present the ranked features for both disaster types in Table 4.6. We observe that URL, emoticons and phone number features are ranked top 3. For both the disasters, the first three ranked features are the same, and after that, the ranking is different. From feature selection and ranking, we observe that URL, emoticons and phone number features are more informative features for identifying informative tweets despite the disaster type.

4.5. Summary

In this work, we designed several feature sets for identifying informative tweets using CrisisLex dataset. We experimented with BoW and various combinations of our designed feature sets TC, UF, and PF to obtain the best performing model. Among our feature set combinations, TC+PF is performing better than other feature set combinations and adding BoW to TC+PF shown improved classifier performance over TC+PF. We find out that using user features is not useful for this classification task, in general, the informativeness of a tweet is independent of user characteristics. We also explored how the models developed for natural disasters is useful for non-natural disasters and vice-versa. We found that the natural disasters are well generalized to all kinds of disasters, but non-natural disasters are not.

In addition, we performed an analysis to investigate why polarity features help in improving classifiers' performance. We used emotional divergence (ED) to capture and quantify the effect of having polarity clues in the tweets. We found that the tweets with high ED values, have a high chance of being non-informative and for the tweets with low ED values have high chance to be informative. We performed feature selection and found that among the TC features, we found URL, emoticons and phone number features are more informative features. This work will be very helpful when integrated into a real-time system and filter out necessary informative data during the disasters, for immediate availability to the emergency responders and reducing/minimizing the damage. In future, we can also develop models that automatically identify the tweets of the users who are posting for help, which will be an effortless task for the first responders to find those who are seeking help.

One more interesting direction would be to identify the tweets which convey actionable information.

4.5.1. Related Work

Sakaki et al. [51] used learning techniques to detect earthquakes in Japan using Twitter data. They designed a model to build an autonomous earthquake reporting system in Japan using Twitter users as sensors. Mendoza et al. (2010) studied the propagation of rumors and misinformation from the Chilean earthquake using only a small set of cases. Gupta et al. [12] focused on identifying fake images that were shared on Twitter, during Hurricane Sandy. Dailey & Starbird [8] explored techniques such as visible skepticism to help control the spread of false rumors during crisis events. Caragea et al. [5] built models for classifying short text messages from the Haiti earthquake into classes representing people's most urgent needs so that relief workers, people in Haiti, and their friends and families can easily access them. Ashktorab et al. [2] used a combination of classification, clustering, and extraction methods to extract actionable information for disaster responders. Li et al. [30] used a domain adaptation approach to study the usefulness of labeled data from a source disaster, together with unlabeled data from a target disaster to learn classifiers for the target. The authors showed that source data could be useful for classifying target data. Similarly, Imran et al. [20] explored domain adaptation for identifying information nuggets using conditional random fields and data from two natural disasters, Joplin 2011 tornado (as source) and Hurricane Sandy (as target). Moreover, Imran et al. [21] performed cross-language domain adaptation using a supervised learning approach with unigrams and bigrams features, and found that for similar languages (e.g., Italian and Spanish), the cross-language domain adaptation is useful, whereas, for dissimilar languages (e.g., Italian and English), it is not useful.

Starbird and Palen [57] studied information propagation in Twitter during mass emergencies through the re-tweet feature of Twitter, using North Dakota Red River floods and Oklahoma Wildfires. They focused on the retweet aspect and analyzed the percentage of the retweets among the collected tweets to show that retweeting plays a major role in information sharing. Neppalli et al. [37] focused on the task of automatically predicting the retweetability of a tweet during disasters. They studied features that are useful for predicting retweetability and found that features related to user details (e.g., friends and followers) are generally very useful. In a similar fashion, Fujio et al. [11] analyzed behavioral changes in users before and after the 2011 catastrophic great east Japan earthquake and showed that the retweets are the most used form of information sharing. Cresci et al. [7] developed a system that automatically detects the tweets that are useful for assessing the damage incurred during natural disasters with a focus on cross-event performance.

In contrast to the above works, we aim at understanding, in a social media stream of messages, what are the features that can help identify messages that convey information relevant to a disaster. More precisely, we contrast several sets of features extracted from the tweet content, user details and polarity clues, and study what are the best-performing ones. We evaluate the performance of models trained using these features on natural disasters (e.g., earthquakes and floods), as well as non-natural disasters (e.g., fire accidents and train crash), and we investigate if these features will result in models that will generalize well from one type of disaster to another (e.g., from natural to non-natural disasters). Last, we show how diversified emotions in a tweet affect the informativeness of the tweet.

CHAPTER 5

CONCLUSION AND FUTURE DIRECTIONS

In this chapter, we summarize the contributions of this work and present some interesting future directions.

5.1. Dissertation Summary

In recent years, micro-blogging services such as Twitter and Facebook have emerged as effective tools to broadcast information world-wide [59]. Scholars from the field of disaster management see hope in social media to extract useful information required during disasters. Many studies were conducted concerning the value of using social media for disaster response and their findings implied that the social media data could really help in improving accuracy in decision making, consequently helping them to focus the relief efforts effectively [4, 43, 61, 72]. Despite the strong value of using the social media data during disasters, it is not yet incorporated into the response processes due to several challenges which include the heterogeneity in the content of the posted messages and the big data nature of the social media streams. More precisely, along with the informative messages, there are messages as a kind of noise (which do not convey any useful information) and must be filtered out to arrive at the signal of good data. This increases the necessity for identifying appropriate information from the streaming data, which could make substantial improvements in the response process. Moreover, due to the big data nature of the social media, it is tough and time-consuming to assign human resources to sift through these data and categorize them into useful types.

Through this research, we presented several feature-based approaches to solving the problem of categorizing social media data into useful types in the context of the disasters. We specifically focused on classification based on these three categories: sentiment (which offers insights about how people are feeling during the disaster), retweetability (which helps to inform on how to reach more individuals in the fastest way) and informativeness (which helps to keep people informed with valuable insights about the disaster). The following are the summary of this dissertation:

- **Geo-mapped Sentiment Analysis during Disasters:** In this work, we presented our approach to understanding the general mood during Hurricane Sandy. We performed a geo-mapped sentiment analysis, where we first identified all geotagged tweets in our collection and labeled each of them with sentiment labels using our disaster-trained classifiers. We then associated the sentiments of tweets with their geo-locations. We showed how users' sentiments change according not only to the locations of the users, but also based on the relative distance from the disaster. Additionally, we investigated the impact of emotional divergence on retweetability. We also performed a comparison with one of the previous works related to sentiment analysis and showed that our features are better than the features in the other work.
- **Retweetability Analysis and Prediction during Hurricane Disasters:** In this work we presented a set of analyses on retweeted tweets during Hurricane Sandy and Hurricane Patricia to determine several aspects affecting the retweetability and addressed the problem of retweetability prediction. We explored a broad range of features including features extracted from tweets' content and user account information and use them in conjunction with machine learning classifiers to predict the tweets' retweetability during the hurricane. We also performed a comparison with some of the previous works related to this task and showed that our features are better.
- **Identifying Informative Tweets during Disaster Events:** We performed a broad range of feature exploration to identify informative messages (or tweets) from those that are not-informative in nature (i.e., about user feelings, informal communication or casual conversations). Our approach is based on a combination of bag-of-words features, which are typically used for text classification, and features that are extracted from tweets' content (e.g., URLs, hashtags, emoticons, slang), user details (such as number of friends, number of followers) and polarity clues (such as positive words, negative words). Furthermore, we studied how well our features

generalize across different disaster types (e.g., natural and non-natural disasters) by developing models trained on one disaster type (such as natural disasters) and evaluating them on another disaster type (such as non-natural disasters).

5.2. Summary of Contributions

This section presents the contributions of this dissertation and outlines some of the directions for further research.

- **Geo-mapped Sentiment Analysis during Disasters:** We performed the experiments using the tweets posted during Hurricane Sandy. We found that:
  - The performance of the classifiers trained using the combination of unigrams and sentiment features outperform the baseline and the classifiers trained using unigrams and sentiment-based features individually. This suggests that the two sets of features complement each other, e.g., the presence of emoticons boosts unigrams, and the presence of words not existent in the positive and negative dictionaries boosts sentiment-based features.
  - Through our geo-tagged sentiment maps, emergency responders can interpret the emotional intensities on the ground and can plan the relief efforts more efficiently. They can have an overview of the how people are feeling about the disaster. For example, using our sentiment map emergency responders can focus their relief efforts on the clusters which emit negative sentiments that are closer to the proximity of the Hurricane landfall.
  - The chance of a tweet to be retweeted is higher for low emotional divergence values using geo-tagged tweets (≈74,000). We further validated this finding using all the tweets from the Hurricane Sandy dataset (8.1 M), which confirms that our finding holds for both the samples (only geo-tagged tweets and the whole dataset). We observed that a tweet with $ED < 0.6$ has a higher probability of being retweeted than the tweets with $ED > 0.6$.
  - The content of tweets with low emotional divergence is generally informative in nature whereas, the content for the tweets with high emotional divergence

is more of personal opinions and do not necessarily convey any useful information. We supported this by using the tweets from CrisisLex datasets with informational and non-informative labels. In this analysis, we found that the proportion of informative tweets is more than the conversational tweets at low ED values and the percentage of conversational tweets is more than informative tweets at high ED values.

- **Retweetability Analysis and Prediction:** We used the tweets from Hurricane Sandy and Hurricane Patricia and performed the analysis and the classification experiments. We found that:
  - The retweetability of a tweet is highly influenced by the user account details (such as #followers, #friends, #statuses).
  - The users with more number of followers and friends have more retweeted tweets than the users with lesser followers and friends. This trend is same for both the disasters (Sandy and Patricia).
  - From the list of users ranked based on the number of retweets, we selected the top 1000 and the last 1000 users and discovered that the accounts of the users with more retweets are verified accounts.
  - The strongest contribution of this work is the design and exploration of features for training machine learning classifiers that can predict how likely a tweet is to be highly retweeted on Twitter. Unlike the features used in previous works (e.g., the number of retweets as in (Suh et al. 2010)), our features are not dependent on the retweet phenomenon. Our features are extracted from the tweet text and the user account information only.
  - The performance of the classifiers trained using the combination of tweet content features (TC), user features(U) and bag-of-words (BOW) outperformed the performance of the classifiers trained using TC, U and BOW features alone. This suggests that the user account features, BOW and tweet content features are collaboratively assisting each other in boosting the classifier's performance.

  - For both the disasters, we observe that our feature set - TC+U+BOW performs better than the feature sets in Petrovic et al. (2011) and Suh et al. (2010) in terms of F1-score on the positive class, whereas our feature set performs on par or slightly worse compared with both previous works on the negative class. Hence, our model is more successful than previous works at predicting a retweetable tweet.
  - The positive class performance using normalized user features is better compared with the performance of un-normalized user features. These results prove that the normalization of the user attributes indeed helps in predicting the retweetability of a tweet.
  - We performed feature ranking to find out the informative features using Information Gain algorithm and observed that the features belonging to user account feature type are top ranked.
- **Identifying Informative Tweets:** We designed several feature sets for identifying informative tweets using CrisisLex dataset. We experimented with BOW and various combinations of our designed feature sets TC, UF, and PF to obtain the best performing model. We found that:
  - Among our feature set combinations, TC+PF is performing better than other feature set combinations and adding BOW to TC+PF shown improved classifier performance over TC+PF.
  - User features are not useful for this classification task; in general, the informativeness of a tweet is necessarily dependent on the user characteristics.
  - Among the TC features, the features - URL, emoticons and phone number features are more informative for this task.
  - Domain adaptation results indicate that the natural disasters are well generalized to all kinds of disasters, but non-natural disasters are not.
  - The tweets with high emotional divergence (ED) values, have a high chance of being non-informative and for those with low ED values have high chance of

being informative.

5.3. Future Directions

Some directions for future research in the context of social media for emergency response may include:

- Exploration of features to automatically identify the tweets of the users who are posting for help, which will be an effortless task for the first responders to find those who are seeking help.
- Identifying the tweets which convey actionable information (we define actionable information as the data which helps in better decision making).
- Designing methods to identify the type of support offered (e.g., emotional support, offering food and shelter) through the tweets.
- Exploration of features and identifying the trustworthiness of a tweet posted during a disaster.

BIBLIOGRAPHY

[1] Anderson, K. M., & Schram, A. (2011). Design and Implementation of a Data Analytics Infrastructure In Support of Crisis Informatics Research. ICSE 2011, 21-28 May 2011, Honolulu, Hawaii.

[2] Ashktorab, Z., Brown, C., Nandi, M., & Culotta, A. (2014). Tweedr: Mining Twitter to Inform Disaster Response. In ISCRAM'14.

[3] Baccianella, S., Esuli, A. & Sebastiani, F. (2010). Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining, in: Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10). Valletta, Malta: European Language Resources Association (ELRA), 2010.

[4] Cameron, M. A., Power, R., Robinson, B., & Yin, J. (2012). Emergency situation awareness from twitter for crisis management. World Wide Web - WWW 12 Companion.

[5] Caragea, C., McNeese, N., Jaiswal, A., Traylor, G., Kim, H.-W., Mitra, P., Wu, D., Tapia, A., Giles, L., Jansen, B., Yen, J. (2011). Classifying Text Messages for the Haiti Earthquake. In: ISCRAM 2011.

[6] Caragea, C., Squicciarini, A., Stehle, S., Neppalli, K. & Tapia, A. (2016). Mapping Moods: Geo-Mapped Sentiment Analysis during Hurricane Sandy. In Proceedings of the 11th International ISCRAM Conference, PA, USA. 2014.

[7] Cresci, S., Tesconi, A., Cimino, A., & DellOrletta, F. (2015). A Linguistically-driven Approach to Cross-Event Damage Assessment of Natural Disasters from Social Media Messages. In International World Wide Web Conference Committee (IW3C2), Florence, Italy.

[8] Dailey, D., & Starbird, K. (2014). Visible Skepticism: Community Vetting after Hurricane Irene. In Proceedings of the 11th International ISCRAM Conference. University Park, Pennsylvania, USA. 777- 81.

[9] Denef, S., Bayerl, P.S & Kaptein, N. (2013). Social Media and the Police-Tweeting

Practices of British Police Forces during the August 2011 Riots. In CHI 2013, 34713480. New York, NY.

[10] Dandala, B. (2013). Multilingual Word Sense Disambiguation Using Wikipedia.(Order No. 3691039). Available from Dissertations & Theses @ University of North Texas; ProQuest Dissertations & Theses Global.

[11] Fujio, T., Takeshi, S., Kosuke, S., Kazuhiro, K., Satoshi, K., & Itsuki, N. (2013). Information sharing on twitter during the 2011 catastrophic earthquake. Proc. of the 22nd International Conf. on World Wide Web companion. International World Wide Web Conferences Steering Committee, 1025-1028.

[12] Gupta, A., Ponnurangam, K., Castillo, C., & Meier, P. (2014). TweetCred: Real-Time Credibility Assessment of Content on Twitter. Social Informatics. Springer, 8851, (pp. 228-243).

[13] Hiltz, S.R., Kushma, J.A. & Plotnick, L. (2014). Use of Social Media by U.S. Public Sector Emergency Managers: Barriers and Wish Lists. In: ISCRAM 2014. University Park, PA.

[14] Hochreiter, R. & Waldhauser, C. (2013). A Genetic Algorithm to Optimize a Tweet for Retweetability. In: I Proceeding of MENDEL 2013.

[15] Hu, M. and Liu, B. (2004). Mining and summarizing customer reviews. In SIGKDD 04, pp. 168177.

[16] Hughes, A.L. & Palen, L. (2012). The Evolving Role of the Public Information Officer: An Examination of Social Media in Emergency Management. Journal of Homeland Security and Emergency Management 9, no. 1.

[17] Hughes, A.L., St. Denis, L.A., Palen, L. & Anderson, K.M. (2014). Online Public Communications by Police & Fire Services during the 2012 Hurricane Sandy. In: CHI 2014, 15051514. New York.

[18] Hughes, A.L. & Chauhan, A. (2015). Online Media as a Means to Affect Public Trust in Emergency Responders. In: ISCRAM 2015.

[19] Hughes, A. L., & Tapia, A. (2015). Social Media in Crisis: When Professional Respon-

ders Meet Digital Volunteers. Journal of Homeland Security and Emergency Management.

[20] Imran, M., Elbassuoni, S., Castillo, C., Diaz, F., & Meier, P. (2013b). Practical extraction of disaster-relevant information from social media. In Proceedings of the 22nd international conference on World Wide Web companion (pp. 1021-1024). International World Wide Web Conferences Steering Committee.

[21] Imran, M., Mitra, P. & Srivastava, J. (2016). Cross-Language Domain Adaptation for Classifying Crisis-Related Short Messages. In ISCRAM16, Rio de Janeiro, Brazil.

[22] Ireson, N. (2009). Local community situational awareness during an emergency. In: 3rd IEEE Intl. Conference on Digital Ecosystems and Technologies, 4954.

[23] Java, A.; Song, X.; Finin, T.; & Tseng, B. (2007). Why we twitter: understanding microblogging usage and communities. Network, 43(1), 5665.

[24] Jenders, M., Kasneci, G. & Naumann, F. (2013). Analyzing and Predicting Viral Tweets. In: WWW 2013 Companion, (pp. 657- 664) May 13 17.

[25] Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media? In: World wide web (Vol. 112, pp. 591600). ACM. doi:10.1145/1772690.1772751

[26] Kodrich, K. & Laituri, M (2011). Making a Connection: Social Medias Key Role in the Haiti Earthquake. In: Journal of Communication and Computer, Vol. 8, No. 8, pp. 624-627.

[27] Lalone, N., Tapia, A., Zobel, C., Caragea, C., & Neppalli, V. K. (2017). Embracing Human Noise as Resilience Indicator: Twitter as Power Grid Correlate. In the Journal of Sustainable and Resilient Infrastructure, 2017.

[28] Latonero, M. & Shklovski, I. (2011). Emergency Management, Twitter, and Social Media Evangelism. In the International Journal of Information Systems for Crisis Response and Management 3, no. 4: 116.

[29] Lerman, K., & Ghosh, R. (2010). Information Contagion: an Empirical Study of the

Spread of News on Digg and Twitter Social Networks. Fourth International AAAI Conference on Weblogs and Social Media, 9097.

[30] Li, H., Guevara, N., Herndon, N., Caragea, D., Neppalli, K., Caragea, C., Squicciarini, A., & Tapia, A. (2015). Twitter Mining for Disaster Response: A Domain Adaptation Approach. In ISCRAM 2015.

[31] MacEachren, A.M., Jaiswal, A., Robinson, A.C., Pezanowski, S., Savelyev, A., Mitra, P., Zhang, X., & Blanford, J. (2011). SensePlace2: Geo-twitter Analytics Support for Situation Awareness, in: IEEE Conference on Visual Analytics Science and Technology, Providence, RI, 2011.

[32] Mandel, B., Culotta, A., Boulahanis, J., Stark, D., Lewis, B., & Rodrigue, J. 2012. A demographic analysis of online sentiment during hurricane Irene. In Workshop on Language in Social Media '12.

[33] McClendon, S., & Robinson, A. C. (2012). Leveraging Geospatially-Oriented Social Media Communications in Disaster Response. In: The International ISCRAM Conference 2012 (pp. 211).

[34] Mitchell, T. M. Machine Learning. McGraw-Hill, New York, 1997.

[35] Nagy, A.&Stamberger, J (2012) Crowd Sentiment Detection during Disasters and Crises. In: Proceedings of the 9th International ISCRAM Conference, Vancouver, CA.

[36] National Weather Service (2013, Oct 7 Published) NHC Data in GIS Formats. National Hurricane Center. Retrieved Oct 20, 2013 from www.nhc.noaa.gov/gis/.

[37] Neppalli, V. K., Medeiros, M. C., Caragea, C., Caragea, D., Tapia, A., & Halse, S. (2016). Retweetability Analysis and Prediction during Hurricane Sandy. In Proceedings of the 13th International ISCRAM Conference, Rio de Janeiro, Brazil, 2016.

[38] Neppalli, V. K., Caragea, C., Mayes, R., Nimon, K. & Oswald, F. (2016). MetaSeer. STEM: Towards Automating Meta-Analyses. In the Conference of Innovative Applications in Artificial Intelligence, Phoenix, AZ, USA, 2016.

[39] Neppalli, V. K., Caragea, C., Squicciarini, A., Tapia, A., & Stehle, S. (2017). Sentiment

Analysis during Hurricane Sandy in Emergency Response. In the International Journal of Disaster Risk Reduction, Vol. 21 (213-222), 2017.

[40] Neppalli, K., Caragea, C., Caragea, D., Medeiros, M. C., Tapia, A. & Halse, S. (2017). Predicting Tweet Retweetability during Hurricane Disasters. In the International Journal of Information Systems for Crisis Response and Management, Vol. 8 (32-50), IGI Global, 2017.

[41] Nielsen, F. A. A new ANEW: Evaluation of a Word List for Sentiment Analysis in Microblogs. CoRR, 2011. http://abs/1103.2903abs/1103.2903

[42] Olteanu, A., Vieweg, S., & Castillo, C. (2015). What to Expect When the Unexpected Happens: Social Media Communications across Crises. In: Proc. of 18th ACM Conference on Computer Supportive Cooperative Work & Social Computing. (pp. 9941009). DOI: 10.1145/2675133.2675242

[43] Palen, L., S. Vieweg, S. Liu,&A.L. Hughes. (2009) Crisis in a Networked World: Features of Computer-Mediated Communication in 2007 Virginia Tech Event. Social Science Computer Review.

[44] Pang, B., Lee, L.&Vaithyanathan, S. Thumbs up?: sentiment classification using machine learning techniques. In ACL-02, EMNLP 02, pages 7986, Stroudsburg, PA, USA, 2002.

[45] Palen, L. & Liu, S.B. (2007). Citizen communications in crisis: Anticipating a future of ICT-supported participation. In: Proc. of the CHI Conference, San Jose, CA (pp. 727-736). New York: ACM Press

[46] Palen, L., Vieweg, S. & Anderson, K. M. (2010). Supporting "Everyday Analysts" in Safety- and Time-Critical Situations. In: The Information Society. Vol 27 pp. 52-62.

[47] Pervin, N., Takeda, H., & Toriumi, F. (2014). Factor Effecting Retweetability: An event-Centric Analysis on Twitter. In: 35th International Conference on Information Systems, Auckland 2014

[48] Petrovic, S., Osborne, M., & Lavrenko, V. (2011). RT to Win! Predicting Message

Propagation in Twitter. In 5th International AAAI Conference on Weblogs and Social Media.

[49] Pfitzner, R., Garas, A. & Scheweitzer, F. (2012). Emotional divergence influences information spreading in twitter, in: Proceedings of the 6th International AAAI Conference on Weblogs and Social Media, AAAI Press, 2012, pp. 543546.

[50] Sakaki, T., Toriumi, F., Shinoda, K., Kazama, K., Kurihara, S., Noda, I.&Matsuo, Y. (2013) Regional Analysis of User Interactions on Social Media in Times of Disaster. In: WWW 2013, Brazil.

[51] Sakaki, T., Okazaki, M., & Matsuo, Y. (2010). Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors. WWW 2010, April 26-30 (pp. 851860). Raleigh, North Carolina: ACM.

[52] Schulz, A., Thanh, T.D., Paulheim, H., & Schweizer, I. (2013) A Fine-Grained Sentiment Analysis Approach for Detecting Crisis Related Microposts. In: ISCRAM 2013.

[53] St. Denis, L. A., Palen, L.&Anderson, K. M. (2014). Mastering Social Media: An Analysis of Jefferson Countys Communications during the 2013 Colorado Floods. In: Proc. of the Information Systems for Crisis Response and Management Conference (ISCRAM 2014).

[54] Starbird, K. Digital Volunteerism During Disaster: Crowdsourcing Information Processing, in: CHI '11 Workshop on Crowdsourcing and Human Computation, Vancouver, BC, 2011.

[55] Starbird, K., Palen, L., Hughes, A. L., & Vieweg, S. (2010). Chatter on the Red: What Hazards Threat Reveals About the Social Life of Microblogged Information. In: CSCW 10 (pp. 241250). New York.

[56] Starbird, K., Munzy&Palen, L. (2012) Learning from the Crowd: Collaborative Filtering Techniques for Identifying On-The-Ground Twitterers during Mass Disruptions. In: ISCRAM 2012.

[57] Starbird, K. & Palen, L. (2010). Pass It On? : Retweeting in Mass Emergency. In: Proc. of 7th ISCRAM Conference, Seattle, USA (May 2010).110.

[58] Starbird, K., & Palen, L. (2012). (How) will the revolution be retweeted?: information diffusion and the 2011 Egyptian uprising. In: Proceedings of the Conference on Computer-Supported Cooperative Work (CSCW). Bellevue, WA, USA: ACM Press. Retrieved from http://dl.acm.org/citation.cfm?id=2145212

[59] Stefanidis, A.; Crooks, A.T.; Radzikowski, J.; Croitoru, A. & Rice, M. (2014), Social Media and the Emergence of Open-Source Geospatial Intelligence, in Murdock, D.G., Tomes, R. & Tucker, C. (eds.), Human Geography: Socio-Cultural Dynamics and Global Security, US Geospatial Intelligence Foundation, Herndon, VA, pp. 109-123.

[60] Suh, B., Hong, L., Pirolli, P., & Chi, E. H. (2010). Want to be retweeted? Large scale analytics on factors impacting retweet in twitter network. In: SocialCom 2010, (pp. 177184).

[61] Sutton, J., Palen, L.,&I. Shklovski. (2008) Backchannels on the Front Lines: Emergent Use of Social Media in the 2007 Southern California Fires. ISCRAM.

[62] Sutton, J. N., Spiro E. S., Johnson, B., Fitzhugh, S.M., Greczek, M. & Butts, C.T. (2012). Connected Communications: Network Structures of Official Communications in a Technological Disaster. In ISCRAM 2012). Vancouver, BC.

[63] Sutton, J. N., Spiro E. S., Fitzhugh, S.M., Johnson, B., Greczek, M. & Butts, C.T. (2014). Terse Message Amplification in the Boston Bombing Response. In: ISCRAM 2014.

[64] Sutton, J., Gibson, C. Ben, Spiro, E. S., League, C., Fitzhugh, S. M., & Butts, C. T. (2015). What it takes to Get Passed On: Message Content, Style, and Structure as Predictors of Retransmission in the Boston Marathon Bombing Response. Plos One, 10(8).

[65] Tapia, A. H., Bajpai, K., Jansen, B. J., & Yen, J. (2011). Seeking the Trustworthy Tweet: Can Microblogged Data Fit the Information Needs of Disaster Response and Humanitarian Relief Organizations. In: ISCRAM 2011 (pp. 110).

[66] Tapia, A. Moore, K. Johnson, N. (2013). Beyond the Trustworthy Tweet: A Deeper

Understanding of Microblogged Data Use by Disaster Response and Humanitarian Relief Organizations. In: ISCRAM 13.

[67] Tapia, A. & Moore, K. (2014). Good Enough is Good Enough: Overcoming Disaster Response Organizations Slow Social Media Data Adoption Special Issue on Technologies for Disaster Response. In: Journal of Computer Supported Cooperative Work. 2014.

[68] Terpstra, T. (2012). Towards a real-time Twitter analysis during crises for operational crisis management. Proceedings of International ISCRAM Conference 2012 (pp. 19).

[69] Thomson, R., Ito, N., Suda, H., Lin, F., Liu, Y., Hayasaka, R., Isochi, R.&Z. Wang. (2012) Trusting Tweets: The Fukushima Disaster and Information Source Credibility on Twitter. In: ISCRAM 2012.

[70] Uysal, I., & Croft, W. B. (2011). User Oriented Tweet Ranking: A Filtering Approach to Microblogs. In Proceedings of the International conference on Information and knowledge management (CIKM) (pp. 22612264). Glasgow, Great Britain.

[71] Vieweg, S. (2010). Microblogged Contributions to the Emergency Arena: Discovery, Interpretation, and Implications. CSCW 2010, February 6-10 (pp. 515516). Savanah, GA: ACM. Retrieved from http://www.citeulike.org/user/ChaTo/article/6761693

[72] Vieweg, S., Hughes, A.L., Starbird, K. & Palen, L. (2010) Supporting Situational Awareness During Emergencies Using Microblogged Information. In: CHI 2010.

[73] Yang, F. (2012). Automatic Detection of Rumor on Sina Weibo Categories and Subject Descriptors. MDS12 (Vol. 2). Beijing, China.

[74] Zaman, T. R., Herbrich, R., Van Gael, J., & Stern, D. (2010). Predicting information spreading in twitter. In: Workshop on Computational Social Science and the Wisdom of Crowds, NIPS.