



Sentiment analysis during Hurricane Sandy in emergency response



Venkata K. Neppalli^{a,*}, Cornelia Caragea^a, Anna Squicciarini^b, Andrea Tapia^b, Sam Stehle^b

^a University of North Texas, Denton, TX, 76201, USA

^b Pennsylvania State University, University Park, PA, 16801, USA

ARTICLE INFO

Keywords:

Sentiment classification
Disaster-related tweets
Hurricane Sandy
Emotional divergence
Retweetability

ABSTRACT

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 paper, 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 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.

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 [39]. 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 [12]. Emergency decision makers see the data produced through crowdsourcing as ubiquitous, rapid and accessible, with the potential to contribute to situational awareness [41]. Starbird et al. [35] 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 [39]. 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.” Through this research, 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 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 re-tweeted, 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 paper is organized as follows: Section 2 discusses the related work. Section 3 describes the sentiment classification followed

* Corresponding author.

E-mail address: kishoreneppalli@my.unt.edu (V.K. Neppalli).

by a geo-tagged sentiment analysis of tweets posted during Hurricane Sandy, in which sentiment classification of tweets is an important component. Section 4 describes an analysis on how the divergence of sentiments in a tweet is affecting tweets' retweetability. Section 5 concludes the paper with a summary and discussion.

2. Related work

2.1. The use of microblogged data in disaster response

Researchers have demonstrated the power of microblogging on the diffusion of news-related information [16,17]. Microblogging has been under the lens of researchers with regards to its use in disasters and other high profile events [14]. However, using microblogged feeds as information sources during a large-scale event is highly problematic for reasons including deception, focus of attention, quantification of performance, reliability, the inability to verify either the person or the information that the person posts [11,24,38]. Still, researchers are optimistic about the value of using microblogging data in disaster response. For example, several research groups have demonstrated that emergency managers and responders understand the value of social media for crisis communication [12]. 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 [7,13,8,37]. 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 [10,36]. Moreover, machine learning and natural language processing have made great leaps in extracting, processing and classifying Twitter feeds. Sakaki et al. [31] used machine learning techniques to detect earthquakes in Japan using Twitter data. Mendoza et al. [24] studied the propagation of rumors and misinformation from the Chilean earthquake using only a small set of cases. Castillo et al. [6] analyzed information credibility in Twitter. Specifically, the authors developed automatic methods to assess the credibility of tweets related to specific topics or events (although not restricted to disaster events), using features extracted from users' posting behavior and tweets' social context. Caragea et al. [3] used text classification approaches to build models for the classification of short text messages from the Haiti earthquake into classes representing people's most urgent needs so that NGOs, relief workers, people in Haiti, and their friends and families can easily access them. Li et al. [18] 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 and showed that source data can be useful for classifying target data. Similarly, Imran et al. [15] explored domain adaptation for identifying information nuggets using conditional random fields and data from two disasters, Joplin 2011 tornado (as source) and Hurricane Sandy (as target). Ashktorab et al. [1] used a combination of classification, clustering, and extraction methods to extract actionable information for disaster responders. Caragea et al. [5] presented an approach based on Convolutional Neural Networks for identifying informative messages in social media streams during disaster events. Neppalli et al. [27] focused on the task of automatically predicting the retweetability of a tweet in the context of the Hurricane Sandy, using features extracted from the tweets' content and user account information. In contrast to the above works, we use machine learning techniques to perform a geo-mapped sentiment analysis of tweets during the Hurricane Sandy.

2.2. Geo-mapped microblogged data in disaster response

Mapping crowd-sourced information in disaster response gained wide-scale media attention during the 2010 Haiti earthquake [34], with several challenges involved in mapping crowd-sourced communications, including the extraction of accurate location information, and the

application of useful and usable cartographic representations to visually support situational awareness in crises [22]. This occurs due to the need to display large volumes of data, while avoiding information overload [22], which is complicated further by the fact that potential users of crisis maps will have different expectations influenced by their social and physical relation to the crisis event [19]. According to McClendon & Robinson [22], "Mapping social media content provides a way to gather and visualize information from what can arguably be considered the true first responders - the affected citizens who are the first to assess the situation and request assistance through social media... Future research must focus on applications that go beyond basic crowd-sourcing to develop information collections, analytical tools, coordination of communications, and mapping visualization to support all phases of disaster management."

2.3. Sentiment analysis for disaster events

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 [25] focused on sentiment detection in Twitter during the San Bruno, California gas explosion and fires from 09/2010. They used SentiWordNet together with dictionaries of emoticons and out of vocabulary words, and a sentiment-based dictionary to identify the basic sentiment of a tweet. Schulz et al. [33] 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 [28] word list and SentiWordNet [2]. Schulz et al. [33] evaluated their models on tweets related to the Hurricane Sandy from October 2012. Mandel et al. [21] performed a demographic sentiment analysis using Twitter data during Hurricane Irene. Pfitzner et al. [30] 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. Their dataset contains tweets from a variety of popular events related to sports, entertainment, politics and technology. Building directly on these works on sentiment analysis, we take the next logical step and focus on geo-mapped 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. Our approach, originally introduced in Caragea et al. [4], can help increase situational awareness and can "visually" inform emergency response organizations about the geographical regions that are most affected by a disaster. In this 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.

3. Geo-mapped sentiment analysis of tweets from hurricane sandy

Through this research, we seek to find mechanisms to automatically classify the sentiment of tweets posted during the Hurricane Sandy. We formulate the problem as a classification problem and use supervised learning approaches to classify a tweet as positive, negative or neutral, based on the polarity of the emotion expressed in the tweet. Table 1 shows examples of tweets extracted from our Hurricane Sandy dataset.

The sentiment classification of tweets faces many challenges including dealing with very short texts as well as unstructured text and noisy user input, e.g., tweets contain many misspellings, "ole" instead of "old", or acronyms, "smh" (as can be seen from Table 1,

Table 1

Examples of tweets from Hurricane Sandy, labeled as positive, negative & neutral.

Tweet	Sentiment
1. "RT @User1: During this hurricane we are all going to reunite on Xbox like the good ole days"	Positive
2. "RT @User2: It doesnt look like a hurricane is coming."	Neutral
3. "User3: I got a feeling that #Sandy is about to screw up my work schedule for the week: (smh)"	Negative

examples 1 and 3, respectively). To detect the sentiment of tweets, we propose to use a combination of bag of word and sentiment features, e.g., emoticons and polarity clues, as the feature representation provided as input to machine learning algorithms.

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

3.1. Hurricane sandy data

The data used in our experiments is collected from Twitter during the disastrous Hurricane Sandy, using the Twitter Streaming API.¹ Examples of searched keywords used for data collection are: "hurricanesandy," "sandy," "hurricane," "sandyhurricane" and "hurricane east coast." The dataset contains 12,933,053 tweets crawled between 10–26-2012 and 11–12-2012, among which 4,818,318 have links to external sources, 6,095,524 are retweets, 622,664 contain emoticons, and 74,708 are geo-tagged. We selected this time-frame based on the number of tweets that we were able to collect per day. Fig. 1 shows the distribution per day of the 12.9 M collected tweets in our dataset. We can observe a burst after two days from the beginning of the hurricane. The delay can be explained by the hurricane progressive nature, i.e., it was forecasted a few days prior to the strike and the pace in postings picked up as it hit the coast from the Atlantic Ocean. As can be seen from the figure, the number of tweets per day decreases as time elapses.

We randomly sampled a subset of 700 tweets from all of our collected data and asked three annotators (volunteers from our research labs) to label them as positive, negative and neutral. The agreement between the annotators was 86%. From the set of 700 tweets, we removed the ones on which the annotators disagreed and were left with 602 tweets. In total, we obtained 249 positive examples, 216 negative examples and 137 neutral examples. These annotated tweets are used for the evaluation of our sentiment classifiers.

3.2. 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 $(\mathbf{x}_i, y_i)_{i=1, \dots, n}$, $\mathbf{x}_i \in \mathcal{X}$ 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

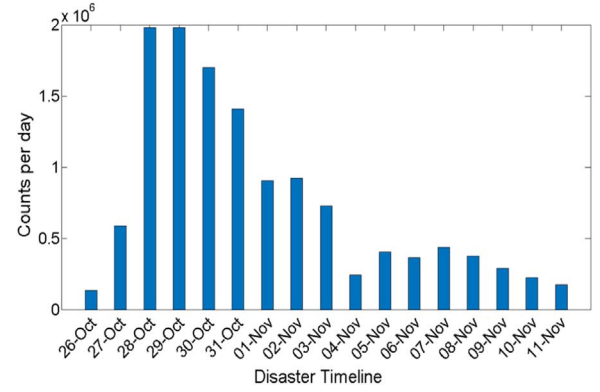


Fig. 1. The distribution of total posts per day in our collection during Hurricane Sandy.

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 [23,29]. 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, stop-word 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 $(\text{PosDensity}+1)/(\text{NegDensity}+1)$. We used a list of positive and negative words created by Hu & Liu [9]. 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.² If a match of the emoticon pattern is found, then the value for this feature is 1. Otherwise, the feature value is 0. In addition, we developed two more features which depend on the sentiment - positive emoticon presence and negative emoticon presence. For this, we divided the emoticons dictionary into positive and negative dictionaries, and the matched emoticon is checked against these dictionaries. For positive emoticon presence, we assign the feature value 1 if found in the positive emoticon dictionary, otherwise, we assign 0. Similarly, for negative emoticon presence, the feature value 1 is assigned if found in the negative emoticon dictionary, otherwise, 0 is assigned.

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

¹ <https://dev.twitter.com/streaming/overview>

² https://en.wikipedia.org/wiki/List_of_emoticons

Table 2
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

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.³ 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.

3.3. Experiments and results for sentiment classification

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 the combination of unigrams and sentiment-based features. We report the average classification accuracy obtained in 10-fold cross-validation experiments. In experiments, we used SVM with a linear kernel and 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 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 the absolute value of its positive sentiment score is greater than that of 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 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. We reported the accuracy of the classifiers. 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 less 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,

Table 3
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

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 the SVM classifier for C=0.1, i.e., 75.91%.

Comparison with Prior Work: Our work is the most similar with that of Nagy & Stamberger [25]. Hence, we performed a comparison with this work. The approach presented in Nagy & Stamberger [25] is to develop features based on emoticons and two sentiment dictionaries - AFINN [28] and SentiWordNet [2]. In addition to the features used in Nagy and Stamberger [25], 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 3. As can be seen from the table, the classifiers based on our features perform substantially better compared with the approach proposed by Nagy & Stamberger [25]. Specifically, the highest performance that our models achieve is 75.91% as compared with 64.30% achieved by models proposed in Nagy & Stamberger [25].

3.4. Geo-mapped tweets sentiment analysis

Labeling the Set of Unlabeled Geo-Tagged Tweets from Hurricane Sandy: 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 [40]. Temporal distance similarly impacts the tendency of a Twitter user to disseminate information about an emergency event [32]. In this section, we use the geographic representation and cluster measures to examine 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 Oct. 26 and Oct. 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

³ <http://sentistrength.wlv.ac.uk/>

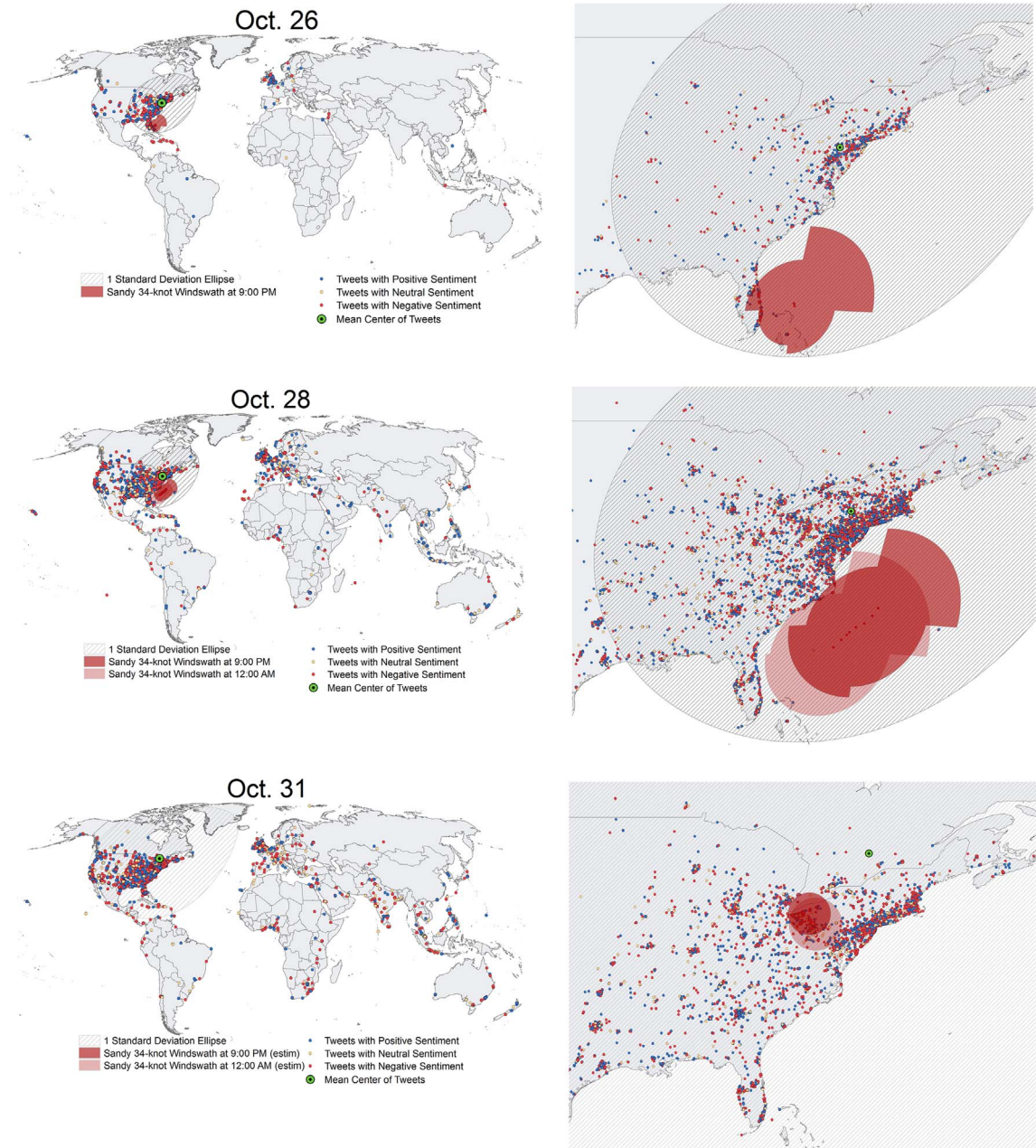


Fig. 2. Maps of Positive, Neutral, and Negative Tweets at global and regional scale. The varying size and position of the standard deviational ellipse and mean center, respectively, are consistent with clustering measures. Maps are shown in the Robinson projection. All distance calculations are done in an azimuthal equidistant projection centered on Sandy's landfall point. The maps are drawn using ArcGIS (<http://www.esri.com/software/arcgis>).

buffers of decreasing diameter through October 31 around the best-track of the storm's center provided by NOAA [26]. 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 [40].

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 (Fig. 2). 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

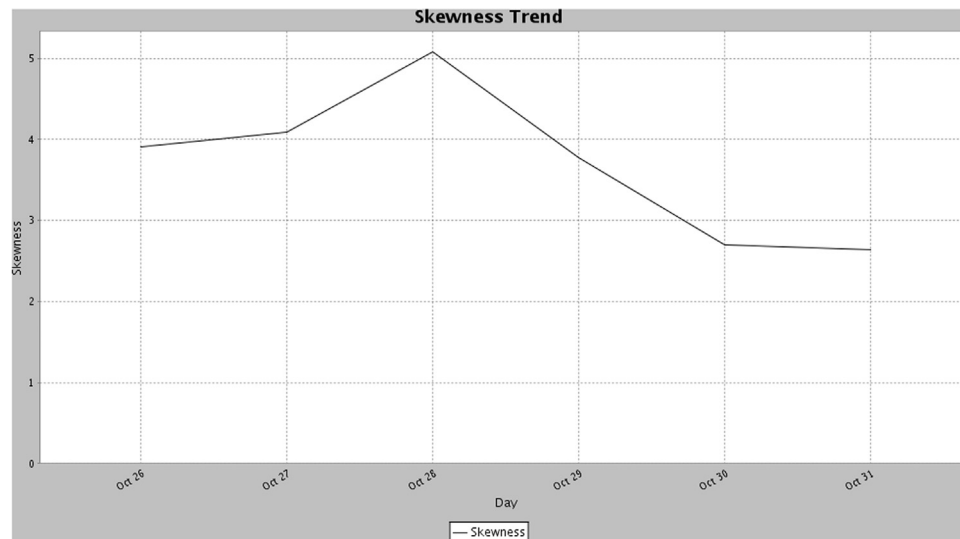


Fig. 3. Skewness as a function of time.

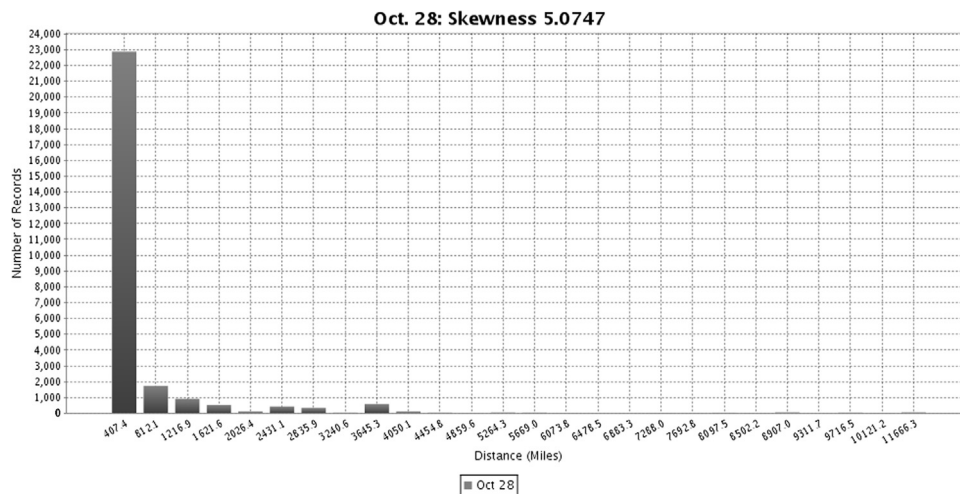


Fig. 4. Histogram of October 28. The extreme positive skewness indicates short distances between each Tweet and the point where Hurricane Sandy made landfall.

maximum impact, then quickly to a less clustered, global dispersion (Figs. 3 and 4). 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 (Fig. 5).

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 reflecting people in need of help). Through our geo-tagged sentiment maps, emergency responders can interpret the emotional intensities “on the ground” (positive and negative sentiment clusters) and can plan the relief efforts more efficiently. The sentiment maps can provide emergency responders with an overview of how people feel about the disaster. For example, using our sentiment map, emergency responders can focus their relief efforts on the clusters which emit high negative sentiments that are closer to the proximity of the Hurricane landfall.

4. 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 [30], 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 sleep 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

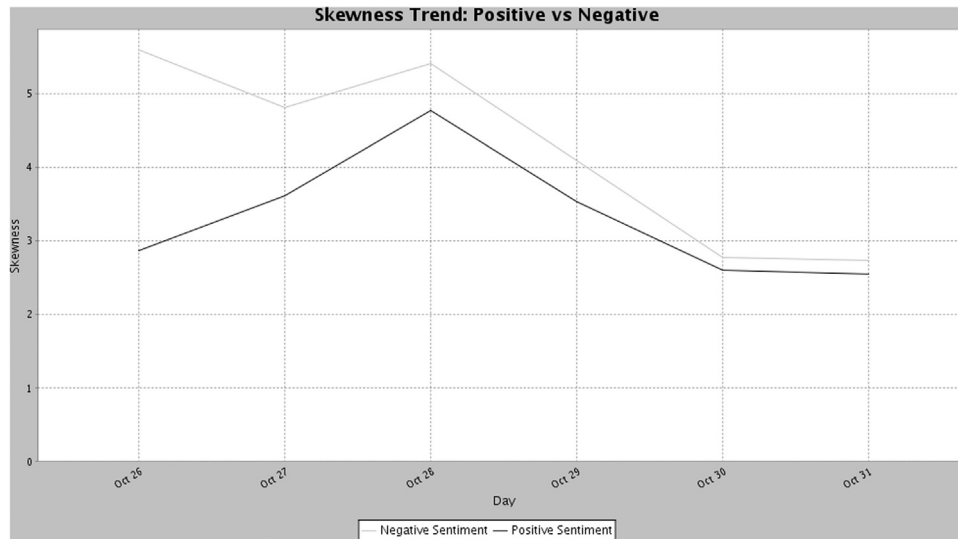


Fig. 5. Positive vs. negative skewness as a function of time.

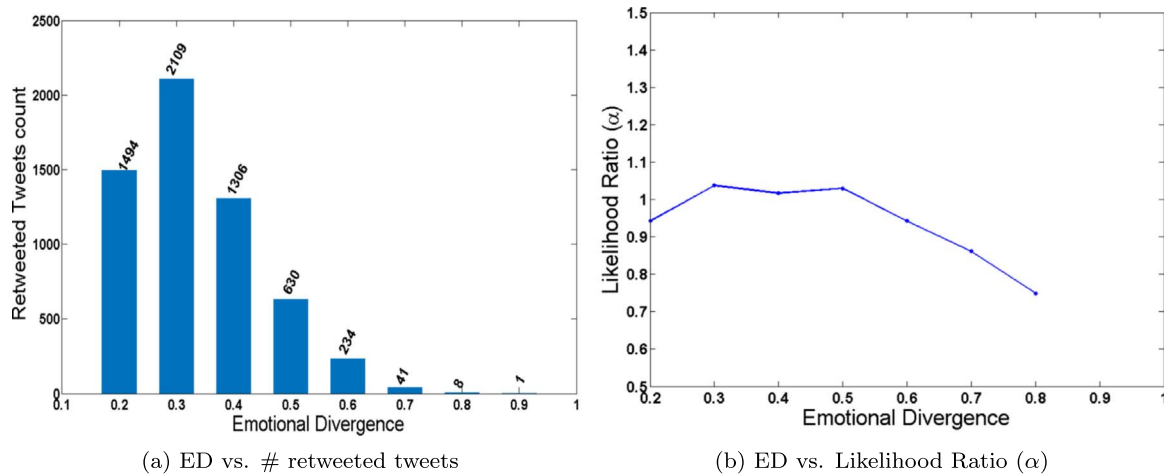


Fig. 6. Impact of Emotional Divergence on Retweetability of geo-tweets.

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 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 the geo-tagged tweets. In Fig. 6(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 decreases 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.

For each emotional divergence value (ED), we also calculate $Pr(ED = x|T)$ for $x = 0.2, \dots, 1$, 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 Fig. 6(b), we plot the likelihood ratio with respect 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 Pfizner et al. [30], 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 difference is due to different types of data. Our work is purely focused on disasters, whereas in Pfizner et al. [30] 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.

To further validate our findings, we performed the same analysis on the entire 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. Fig. 7(a), shows the correlation between emotional divergence and the corresponding retweeted tweets

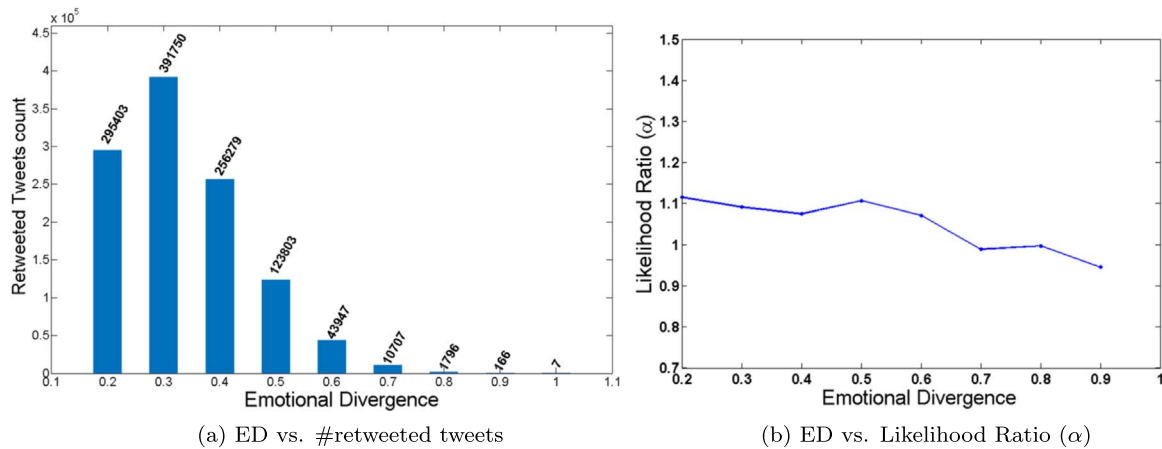


Fig. 7. Impact of Emotional Divergence on Retweetability of tweets from the entire dataset.

Table 4

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

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, Fig. 7(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 re-tweeted than those that are conversational in nature, we study the correlation between low emotional divergence (hence, high chance of re-tweetability) and the tweets' informativeness in the following section.

4.1. Emotional divergence vs. informativeness

In this section, we describe an experiment performed 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 manually inspected these two sets and noticed interesting patterns. 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 4 shows examples of retweeted tweets with low emotional divergence ($ED=0.2$) and their retweets count. Moreover, tweets with $ED=0.8$ are more conversational in nature.

Table 5 shows examples of retweeted tweets with high emotional divergence ($ED=0.8$) and their retweets count. They express personal opinion/feeling of users rather than conveying necessarily useful information. To better understand the connection between emotional divergence and informativeness, we analyze a set of tweets available at

Table 5

Examples 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 sleep perfect.
1	Were just going to watch bad romantic comedies dance and maybe cry a little for the next two days. #sandy

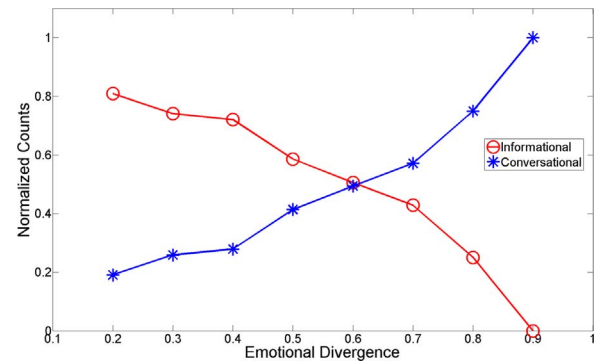


Fig. 8. Emotional Divergence vs. Informational/Conversational.

<http://crisislex.org>.⁴ This sample contains tweets coded with informational and conversational labels. The dataset contains tweets from 26 disasters that 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.

Fig. 8 shows the plot between emotional divergence and normalized counts of the tweets. In the figure, the red curve represents the informational tweets and the blue curve represents the conversational tweets. As can be seen, the two curves show opposite trends. The curve for informational tweets shows a decreasing trend: as the emotional

⁴ <http://crisislex.org/data-collections.html>

divergence is increasing, the normalized counts is decreasing, whereas for conversational, 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 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 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.

5. 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, whereas the content of 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.

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 people. This online discussion is a rich trove of information that could possibly inform responders, if made actionable. There are several reasons that these 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 these data are not currently used to their 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 paper 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. Through an analysis of the divergence of sentiments in tweets, we study how likely a tweet is to be retweeted based on its emotional divergence. Understanding elements of a message that make it more likely to be retweeted during a disaster can potentially shed light into the type of information being spread during disasters in large microblogging communities, and can better inform emergency managers on how to reach the widest audience in the fastest way. 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 close proximity to the disaster. 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 paper, MacEachren et al. [20] 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 [20]. It is essential to refocus on the utility of the extracted information and the effectiveness of associated crisis maps to support emergency response. In this paper, we present 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.

In future, it would be interesting to integrate more sophisticated models or design new ones for improved sentiment classification as well as fine-grained emotion detection. Another interesting future direction would be to understand how emotional divergence affects the automatic identification of informative tweets (i.e. tweets that convey information useful to emergency responders as well as people on the ground) in Twitter during a disaster event. Finally, understanding how the geo-mapped sentiment analysis works across different disasters is highly desirable.

Acknowledgements

We thank the National Science Foundation for support from the grants IIS 1526542 and IIS 1526678 to Cornelia Caragea, Andrea Tapia, and Anna Squicciarini. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either express or implied, of the National Science Foundation. We also wish to thank our anonymous reviewers for their constructive comments that helped improved the quality and readability of our paper.

References

- [1] Z. Ashktorab, C. Brown, M. Nandi, A. Culotta, Tweedr: Mining Twitter to Inform Disaster Response. ISCRAM'14 2014.
- [2] S. Baccianella, A. Esuli, F. Sebastiani, 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.
- [3] C. Caragea, N. McNeese, A. Jaiswal, G. Traylor, H.-W. Kim, P. Mitra, D. Wu, A. Tapia, L. Giles, B. Jansen, J. Yen, Classifying Text Messages for the Haiti Earthquake. ISCRAM'11, 2011.
- [4] C. Caragea, A. Squicciarini, S. Stehle, K. Neppalli, A. Tapia, Mapping Moods: Geo-mapped sentiment analysis during Hurricane Sandy, in: Proceedings of the 11th International ISCRAM Conference, 2014.
- [5] C. Caragea, A. Silvescu, A. Tapia, Identifying informative messages in disasters using Convolutional Neural Networks, in: Proceedings of the 13th International ISCRAM Conference, 2016.
- [6] C. Castillo, M. Mendoza, B. Poblete, Information credibility on twitter, WWW 11. ACM. 2011 pp. 675–684 (<http://dx.doi.org/10.1145/1963405.1963500>).
- [7] S. Denef, P.S. Bayerl, N.A. Kaptein, Social media and the police: Tweeting practices of british police forces during the august 2011 riots, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI '13, New York, NY, USA, 2013, pp. 3471–3480. ACM. (<http://dx.doi.org/10.1145/2470654.2466477>).
- [8] L.A.S. Denis, L. Palen, K.M. Anderson, Mastering Social Media: an Analysis of Jeffersons county's Communications during the 2013 Colorado Floods, ISCRAM, 2014.
- [9] M. Hu, B. Liu, Mining and Summarizing Customer Reviews, in: SIGKDD '04 KDD '04, 2004, pp. 168–177.
- [10] A.L. Hughes, A. Chauhan, Online Media as a Means to Affect Public Trust in Emergency Responders, ISCRAM'15, 2015.
- [11] A. L. Hughes, L. Palen, Twitter adoption and use in mass convergence and emergency events. International Journal of Emergency Management, vol. 6, 2009, pp. 248–260.
- [12] A.L. Hughes, L. Palen, The evolving role of the public information officer an examination of social media in emergency management, J. Homel. Secur. Emerg. Manag. 9 (1) (2012). <http://dx.doi.org/10.1515/1547-7355.1976>.
- [13] A.L. Hughes, L.A.A. St. Denis, L. Palen, K.M. Anderson, Online public communications by police and fire services during the 2012 hurricane sandy, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI '14, New York, NY, USA: ACM, 2014, pp. 1505–1514.
- [14] C. Hui, Y. Tyshchuk, W. Wallace, M. Goldberg, M. Magdon-Ismael, Information Cascades in Social Media in Response to a Crisis: a Preliminary Model and a Case Study. WWW, 2012, pp. 653–656.
- [15] M. Imran, S. Elbassuoni, C. Castillo, F. Diaz, P. Meier, Practical extraction of disaster-relevant information from social media, in: Proceedings of the 22nd International Conference on World Wide Web WWW '13 Companion, New York,

- NY, USA: ACM, 2013, pp. 1021–1024. (<http://dx.doi.org/10.1145/2487788.2488109>)
- [16] H. Kwak, C. Lee, H. Park, S. Moon, What is twitter, a social network or a news media?, *World Wide Web* 112 (2010) 591–600. <http://dx.doi.org/10.1145/1772690.1772751>.
- [17] K. Lerman, R. Ghosh, Information contagion: an empirical study of spread of news on digg and twitter social networks, in: *Proceedings of the 4th International Conference on Weblogs and Social Media (ICWSM)*, 2010, (<http://dx.doi.org/10.1145/1772690.1772751>).
- [18] H. Li, N. Guevara, N. Herndon, D. Caragea, K. Neppalli, C. Caragea, A. Squicciarini, A. Tapia, Twitter mining for disaster response: A domain adaptation approach, in: *Proceedings of the 12th International ISCRAM Conference*, 2015.
- [19] S.B. Liu, L. Palen, The new cartographers: crisis map mashups and the emergence of neogeographic practice. *Cartography and Geographic Information Science*, vol. 37, 2010, pp. 69–90.
- [20] A. MacEachren, A. Jaiswal, A. Robinson, S. Pezanowski, A.S.P. Mitra, X. Zhang, J. Blanford, Senseplace2: Geotwitter analytics support for situation awareness, in: *Proceedings of the IEEE Conference on Visual Analytics Science and Technology* Providence RI, 2011.
- [21] B. Mandel, A. Culotta, J. Boulahanis, D. Stark, B. Lewis, J. Rodrigue, A demographic analysis of online sentiment during hurricane irene, in: *Proceedings of the Workshop on Language in Social Media*, vol. 12, 2013.
- [22] S. McClendon, A.C. Robinson, Leveraging geospatially-oriented social media communications in disaster response. *ISCRAM*, vol. 12, 2012, pp. 2–11.
- [23] R. McDonald, K. Hannan, T. Neylon, M. Wells, J. Reynar, Structured models for fine-to-coarse sentiment analysis. vol. 45, 2007, p. 432.
- [24] M. Mendoza, B. Poblete, C. Castillo, Twitter under crisis, in: *Proceedings of the First Workshop on Social Media Analytics SOMA 10*, vol. 6, 2010, pp. 71–79.
- [25] A. Nagy, J. Stamberger, Crowd sentiment detection during disasters and crises, in: *Proceedings of the 9th International ISCRAM Conference*, 2012.
- [26] National Weather Service 2013. Nhc data in gis formats.
- [27] K. Neppalli, M. C. Medeiros, C. Caragea, D. Caragea, A. Tapia, S. Halse. Retweetability Analysis and Prediction during Hurricane Sandy, in: *Proceedings of the 13th International ISCRAM Conference*, Rio de Janeiro, Brazil, 2016.
- [28] F.Å. Nielsen, A new ANEW: Evaluation of a Word List for Sentiment Analysis in Microblogs. *CoRR*, 2011. <http://abs/1103.2903abs/1103.2903>
- [29] B. Pang, L. Lee, S. Vaithyanathan, Thumbs up?: Sentiment Classification Using Machine Learning Techniques. *ACL-02, EMNLP 02*, 2002, pp. 79–86.
- [30] R. Pfitzner, A. Garas, F. Schewietzer, Emotional divergence influences information spreading in twitter, in: *Proceedings of the 6th International AAAI Conference on Weblogs and Social Media*, AAAI Press, 2012, pp. 543–546.
- [31] T. Sakaki, M. Okazaki, Y. Matsuo, Earthquake shakes twitter users: Real-time event detection by social sensors. *WWW*, Raleigh, North Carolina: ACM, 2010, pp. 851–860. 2010.
- [32] T. Sakaki, F. Toriumi, K. Shinoda, K. Kazama, S. Kurihara, I. Noda, Y. Matsuo, Regional Analysis of User Interactions on Social Media in Times Of Disaster, *WWW*, . Brazil, 2013.
- [33] A. Schulz, T. Thanh, H. Paulheim, I. Schweizer, A Fine-grained Sentiment Analysis Approach for Detecting Crisis Related Microposts, *ISCRAM*, 2013.
- [34] K. Starbird, Digital volunteerism during disaster: Crowdsourcing information processing, in: *Proceedings of the CHI '11 Workshop on Crowdsourcing and Human Computation*, Vancouver, BC, 2011.
- [35] K. Starbird, L. Palen, A.L. Hughes, S. Vieweg, Chatter on the Red: What Hazards Threat Reveals about the Social Life of Microblogged Information. *CSCW*, New York, vol. 10, 2010, pp. 241–250.
- [36] J. Sutton, E. Spiro, B. Johnson, S. Fitzhugh, C. Butts, Terse message amplification in the boston bombing response, in: *Proceedings of the ISCRAM 2014 Conference -11th International Conference on Information Systems for Crisis Response and Management*, 2014.
- [37] J. Sutton, E. Spiro, B. Johnson, S. Fitzhugh, M. Greczek, C. Butts, Connected communications: Network structures of official communications in a technological disaster, in: *Proceedings of the ISCRAM 2012 Conference 9th International Conference on Information Systems for Crisis Response and Management*. Simon Fraser University, 2012.
- [38] A.H. Tapia, K. Bajpai, B. Jansen, J. Yen, Seeking the Trustworthy Tweet: Can Microblogged Data Fit the Information Needs of Disaster Response and Humanitarian Relief Organizations, *ISCRAM '2011*, pp. 1–10.
- [39] A.H. Tapia, K. Moore, N. Johnson, Beyond tHe Trustworthy Tweet: A Deeper Understanding of Micro-blogged Data Use by Disaster Response and Humanitarian Relief Organizations, *ISCRAM 2013*.
- [40] R. Thomson, N. Ito, H. Suda, F. Lin, Y. Liu, R. Hayasaka, R. Isochi, Z. Wang, Trusting tweets: The Fukushima Disaster and Information Source Credibility on Twitter, *ISCRAM*, 2012.
- [41] S. Vieweg, A.L. Hughes, K. Starbird, L. Palen, Microblogging during two natural hazards events, in: *Proceedings of the 28th international conference on Human factors in computing systems - CHI New York, USA: ACM Press*, 2010, pp. 1079–1088.