Mapping Moods: Geo-Mapped Sentiment Analysis During Hurricane Sandy

Cornelia Caragea[1], Anna Squicciarini[2], Sam Stehle[2], Kishore Neppalli[1], Andrea Tapia[2]
[1]Computer Science and Engineering, University of North Texas, Denton, TX-76203
[2]Information Sciences and Technology, Pennsylvania State University, University Park, PA-16801
ccaragea@unt.edu, asquicciarini@ist.psu.edu, samstehle@gmail.com,
KishoreNeppalli@my.unt.edu, atapia@ist.psu.edu

ABSTRACT

Sentiment analysis has been widely researched in the domain of online review sites with the aim of generating summarized opinions of product users about different aspects of the products. However, there has been little work focusing on identifying the polarity of sentiments expressed by users during disaster events. Identifying sentiments expressed by users in an online social networking site can help understand the dynamics of the network, e.g., the main users' concerns, panics, and the emotional impacts of interactions among members. Data produced through social networking sites is seen as ubiquitous, rapid and accessible, and it is believed to empower average citizens to become more situationally aware during disasters and coordinate to help themselves. In this work, we perform sentiment classification of user posts in Twitter during the Hurricane Sandy and visualize these sentiments on a geographical map centered around the hurricane. We show how users' sentiments change according not only to users' locations, but also based on the distance from the disaster.

Keywords

Sentiment classification, disaster-related geo-tagged tweets, Hurricane Sandy.

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 (Tapia et al., 2013). This research can be seen as grouped into three areas: (1) extracting and categorizing raw data into useful information, (2) visualizing the product of this useful information, and (3) integrating both the information and its visual display with responders' needs. According to MacEachren et al. (2011), research has focused more on the challenges of extracting action items and location information from social media and less on the utility of the extracted information and the effectiveness of associated crisis maps to support emergency response. We examine the integration of the first two areas, extraction and visualization, with the intent of leading to the third, utility to responders.

According to Starbird, Munzy and Palen (2012), 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 is derivative in nature: information in the form of reposts or pointers to information available elsewhere (Starbird et al., 2010). This derivative data is abundant, as a form of noise that must be filtered out to arrive at the signal of good data (Anderson & Schram, 2011). A small subset of the data comes from local affected populations in the form of citizen reports (Starbird et al., 2010). Starbird et al. 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." (2010)

Responders are hesitant to use social media data for several reasons (Tapia et al. 2011; 2013). 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. In this paper, 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.*

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 (Palen, Vieweg, Liu & Hughes, 2009; Sutton, Palen & Shklovski, 2008; Terpstra, 2012). The utility of this information has been enhanced by the creation of crisis maps based on location data extracted from social media communications (Liu & Palen, 2010; MacEachren et al., 2011).

Mapping crowd-sourced information in disaster response gained wide-scale media attention during the 2010 Haiti earthquake (Starbird, 2011), 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 (McClendon & Robinson, 2012). This occurs due to the need to display large volumes of data, while avoiding information overload (McClendon & Robinson, 2012), 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 (Liu & Palen, 2010). According to McClendon & Robinson (2012), "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."

The online sharing of information, opinions and sentiments has resulted in the generation of a huge amount of user-generated data during disaster-related events. Analyzing these data can help understand the dynamics of the network, e.g., the main users' concerns and panics, and the emotional impacts of interactions among users. In addition, this analysis can help obtain a holistic view about the general mood and the situation on the ground. Despite the evidence of its strong value to those experiencing the disaster and those seeking information concerning the disaster, there has not been much uptake of message data by large-scale, disaster response organizations (Tapia et al. 2011). Real-time message data being contributed by those affected by a disaster has not been incorporated into established mechanisms for organizational decision-making (Tapia et al. 2011). Through this research, we seek to find mechanisms to *automatically* classify the sentiment of users' Twitter posts during the Hurricane Sandy. Specifically, we formulate the problem as a classification problem and use supervised machine learning approaches to classify a post (or tweet) into one of the following classes: *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. These tweets are annotated as positive, negative, and neutral. The user names are replaced with a generic name.

The sentiment classification of tweets faces many challenges including dealing with very short texts, e.g., a tweet is at most 140 characters in length, and dealing with 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, examples 1 and 3, respectively). In this paper, to detect the sentiment of tweets, we propose to use a combination of *bag of word* and *sentiment features* such as emoticons, acronyms, and polarity clues, as the feature representation provided as input to machine learning algorithms. Furthermore, in order to understand the general mood during the Hurricane Sandy, we perform a geo-mapped sentiment analysis: 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.

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

Table 1. Examples of tweets from the Hurricane Sandy labeled as positive, negative and neutral.

The rest of the paper is organized as follows: Section 2 discusses the related work, Section 3 describes our methodology and the features extraction; Section 4 provides experimental design and results, and Section 5 concludes the paper with a summary and discussion.

RELATED WORK

The Use of Micro-blogged Data in Disaster Response

Researchers have demonstrated the power of crowdsourcing on the diffusion of news-related information (Kwak et al., 2010; Lerman & Ghosh, 2010). Media crowd-sourcing has been under the lens of researchers with regards to its use in disasters and other high profile events (Hui et al., 2012). The American Red Cross and Dell have launched a new Digital Operations Center, the first social media-based operation devoted to humanitarian relief, demonstrating the growing importance of social media in emergency situations (http://content.dell.com/us/en/corp/d/secure/2012-03-07-dell-red-cross-digital-operations-center).

There are numerous challenges while considering the use of Twitter data, including issues of reliability, quantification of performance, deception, focus of attention, and translation of reported observations/inferences into a form that can be used to combine with other information. Using Twitter feeds as information sources during a large-scale event is highly problematic for several reasons, including the inability to verify either the person or the information that the person posts (Hughes et al., 2009; Mendoza et al., 2010; Starbird et al., 2010; Tapia et al., 2011). One problem became apparent during the Haiti earthquake when thousands of volunteers from around the world attempted to provide responders with mapping capabilities, translation services, people and resource allocation, all via Short Messaging Service (SMS) at a distance (Portsea, 2011). Despite the good will of field staff, their institutions' policies and procedures were never designed to incorporate data from outside their networks, especially at such an overwhelming flow. In addition, the organizations did not have the technical staff, or the analytical tools, to turn the flow of data into actionable knowledge (Portsea, 2011). Still, researchers are optimistic about the value of potential information provided, that issues surrounding trustworthiness can be reasonably resolved (Palen et al., 2009; Starbird et al., 2010). While not optimized to current expectations of speed, efficiency and knowledge, these mechanisms have been successful at bringing rescue, relief and recovery to millions (Walton et al., 2011).

Crowdsourcing Early Warning for Crises

The United Nations Office for Coordination of Humanitarian Affairs finds that slowly unfolding emergencies can be mitigated by early response (see http://reliefweb.int/sites/reliefweb.int/files/resources/report_36.pdf). If preparedness, early warning and early response systems are fully functioning, coordinated and integrated, the longer lead time means the humanitarian community can step in early enough to reduce human suffering and help prevent the downward spiral of increased vulnerability to future hazards. Hurricane warning time varies from weeks to hours. Public warnings that offer even seconds have the potential to save lives.

Several scholars have recently turned to social media, Twitter in particular, to test the potential for Twitter to act in some manner as an early warning system. An important common characteristic concerning Twitter as an early warning system is its real-time nature (Sakaki et al., 2010). Machine learning and natural language processing have made great leaps in extracting, processing and classifying micro-blogged feeds. For example, Sakaki et al. (2010) used machine-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. Castillo et al. (2011) analyzed information credibility in Twitter. Specifically, they 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. Castillo et al. tried to model whether end-users would believe the information reported in Twitter to be true or not, but they were not concerned with the detection of sentiment in tweets. Caragea et al. (2011) 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.

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 (2012) 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. (2013) 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 seven 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 word list and SentiWordNet. Schulz et al. (2013) evaluated their models on tweets related to the Hurricane Sandy from October 2012. Mandel et al. (2012) performed a demographic sentiment analysis using Twitter data during Hurricane Irene.

In contrast to these works, we focus on the sentiment classification of tweets from the Hurricane Sandy into one of the three classes: positive, negative and neutral, and associate the sentiment of a tweet with its geo-location in order to obtain a holistic view of the general mood and the situation "on the ground" during the hurricane.

PROBLEM CHARACTERIZATION

The supervised learning problem can be formally defined as follows. Given: (i) an *independent and identically distributed* (*iid*) dataset $\mathcal{D}$ of labeled examples $(\mathbf{x}_i, y_i)_{i=1,\ldots,l}$, $\mathbf{x}_i \in \mathcal{X}$ and $y_i \in \mathcal{Y}$, where $\mathcal{X}$ denotes a vocabulary of words and $\mathcal{Y}$ is the set of all possible class labels; (ii) a hypothesis class $\mathcal{H}$ representing the set of all possible hypotheses that can be learned; and (iii) 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. During classification, the task of h is to 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.

Next, we describe our features used as input to machine learning algorithms. We divide these features into two types: unigrams and sentiment features (polarity clues, emoticons, Internet acronyms, punctuation, and SentiStrength).

Unigrams: This approach is widely used in sentiment classification tasks (McDonald et al., 2007; Pang et al., 2002). 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 (PosDensity+1)/(NegDensity+1). We used a list of positive and negative words created by Hu & Liu (2004). 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[1]. 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.

EXPERIMENTS AND RESULTS

Sandy Twitter data. The data used in our experiments is collected from Twitter during the disastrous Hurricane Sandy. 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

[1] SentiStrength is available at http://sentistrength.wlv.ac.uk/

emoticons. We randomly sampled a subset of 602 tweets from the crawled 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.

Experimental Design: 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., Naïve 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.

For SVM, given a set of labeled inputs $(\mathbf{x}_i, y_i)_{i=1,\ldots,l}$, $\mathbf{x}_i \in \mathbf{R}^d$ and $y_i \in \{-1,+1\}$, learning an SVM is equivalent to learning a decision function $f(\mathbf{x})$ whose sign represents the class assigned to an input $\mathbf{x}$. This can be achieved by solving a quadratic optimization problem. During classification, an unlabeled input $\mathbf{x}_{test}$ is classified based on the sign of the decision function, $sign(f(\mathbf{x}_{test}))$ (i.e., if $f(\mathbf{x}_{test})>0$, then $\mathbf{x}_{test}$ is assigned to the positive class; otherwise, $\mathbf{x}_{test}$ is assigned to the negative class). In experiments, we used SVM with a linear kernel. In the case of linear SVM, when the training data is separable, it is possible to find linear decision functions $f(\mathbf{x})$ that accurately discriminate between positive and negative labeled inputs. When the training data is non-separable, the linear SVM does not find a feasible solution. In this case, an extra cost for errors can be assigned by introducing a set of positive slack variables ξ_i, $i=1,\ldots,l$ in the constraints of the optimization problem, where ξ_i measure the extent to which constraints are violated. SVM selects the decision function that minimizes $||\mathbf{w}||^2/2 + C\cdot\sum\xi_i$, where $\mathbf{w}$ is the parameter vector and C is a user parameter. The larger the value of C, the higher the penalty assigned to errors. We experimented with different values for C = 0.1, 0.5, 0.75, 1.0.

The Baseline Approach: For the positive vs. negative classification task, we used the SentiStrength algorithm as a strong 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 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%.

Results: Table 2 shows the results of the comparison of different classifiers, Naïve Bayes and SVM (C= 0.1,0.5,0.75,1.0), 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.

Feature type	**Naïve Bayes**	**SVM C=0.1**	**SVM C=0.5**	**SVM C=0.75**	**SVM 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

Table 2. Performance of Naïve Bayes and SVM for sentiment classification using various feature types.

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, respectively, and as compared with 59.13% accuracy achieved by SentiStrength. A naïve approach that classifies all tweets in the majority class achieves 53.54% accuracy, which is much worse than that of SVM (C=0.1).

Labeling the Set of Unlabeled Geo-Tagged Tweets. In order to associate the sentiment of tweets with their geo-locations, we extracted the set of geo-tagged tweets from our collection. 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 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, which we discuss in the next section.

GEO-TAGGED TWEETS 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 (Thomson et al., 2012). Temporal distance similarly impacts the tendency of a Twitter user to disseminate information about an emergency event (Sakaki 2013). 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) windspeed 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 (National Weather Service 2013). 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 disaster's impact (Thomson et al., 2012).

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 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 and 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 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.

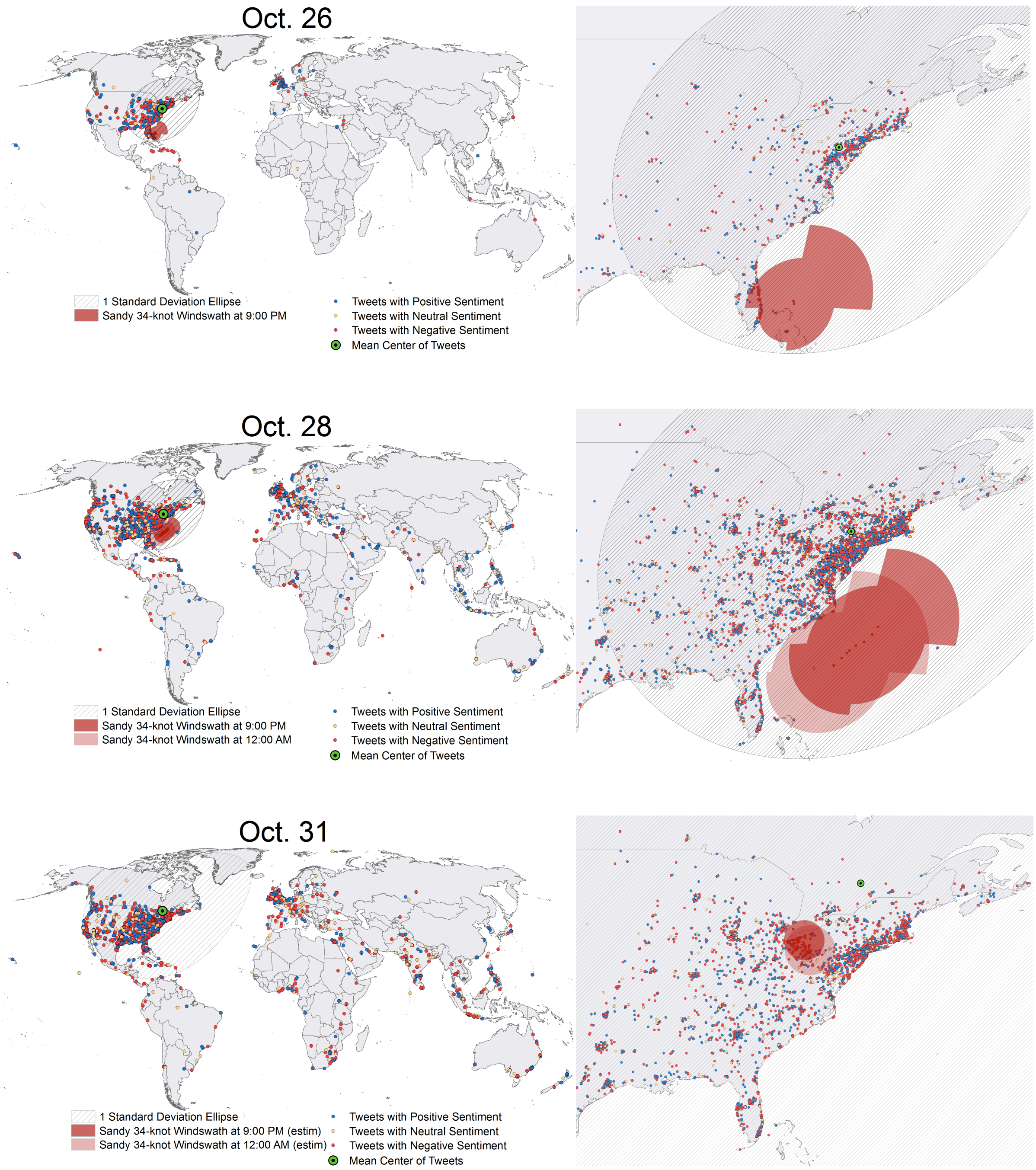

Figure 1. Maps of Positive, Neutral, and Negative Tweets at global and regional scale. The varying size and position of the standard deviational ellipse and mean center, respectively, are consistent with clustering measures. Maps are shown in the Robinson projection. All distance calculations are done in an azimuthal equidistant projection centered on Sandy's landfall point. The maps are drawn using ArcGIS (http://www.esri.com/software/arcgis).

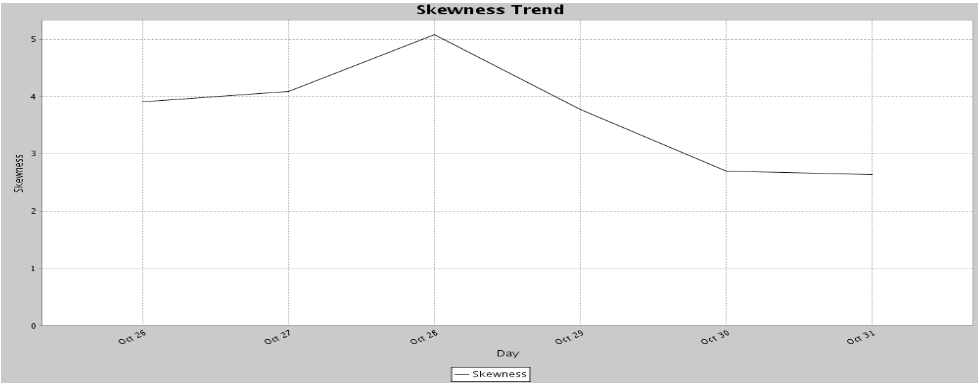

Figure 2. Skewness as a function of time.

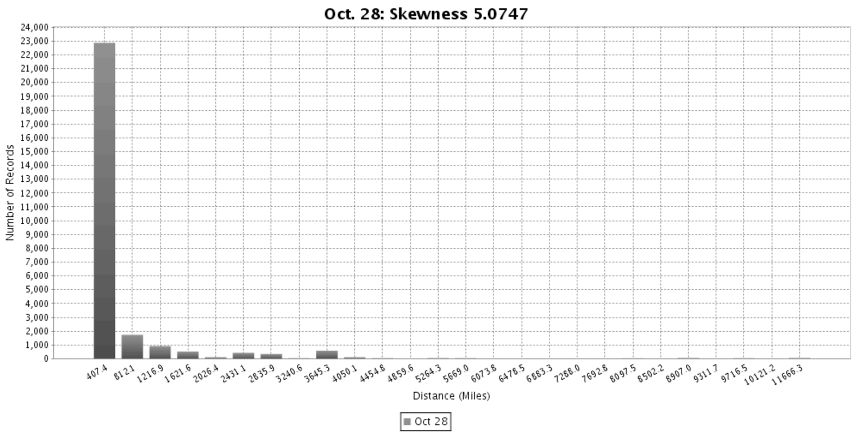

Figure 3. Histogram of October 28. The extreme positive skewness indicates short distances between each Tweet and the point where Hurricane Sandy made landfall.

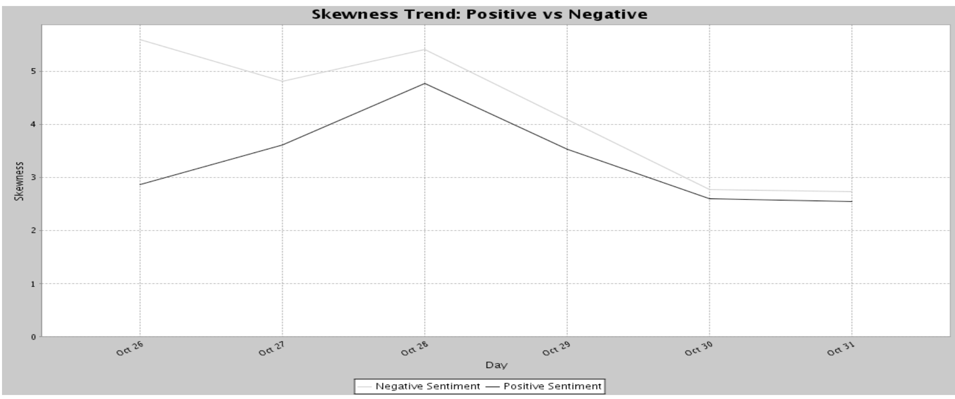

Figure 4: Positive vs. negative skewness as a function of time. Negative sentiment is expressed.

SUMMARY AND DISCUSSION

We performed sentiment classification of user posts in Twitter during the Hurricane Sandy and visualize 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.

Through previous research, we found that data gleaned from social media contributions have both significant value to emergency responders and are difficult to use. Responders seek an enhanced operational picture during any disaster, which grants them better situational awareness. Key to this is not only understanding the impact of the disaster on people and property, but also the actions and movements and response of the affected populations. Affected populations do not simply receive a disaster, they are often the first responders. They evacuate. They protect themselves and neighbors. They intervene in dangerous situations. They assist responders directly. Affected populations are dynamic elements of any response. To obtain a true awareness of a disaster, official responders must have an operational picture that includes the actions of the affected population.

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 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 form 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 are able to make phone calls. Our model 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. 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 (MacEachren et al., 2011). 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 presented a method by which the affected population's response to a disaster might be measured through a sentiment analysis and then mapped in relation to the disaster in space and time. This is one strong step along the path to providing official responders with truly actionable information in real time based on social media data.

ACKNOWLEDGMENTS

We wish to thank the National Science Foundation for support via EAGER: Collaborative Research: Establishing Trustworthy Citizen-Created Data for Disaster Response and Humanitarian Action. 2013-2014. 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.

REFERENCES

1. Anderson, K., & 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. 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.

3. Castillo, C., Mendoza, M., & Poblete, B. (2011). Information Credibility on Twitter. In: *WWW '11* (pp. 675–684). ACM. doi:10.1145/1963405.1963500
4. Hu, M. and Liu, B. "Mining and summarizing customer reviews," in SIGKDD '04, pp. 168–177.
5. Hughes, A. L., & Palen, L. (2009). Twitter Adoption and Use in Mass Convergence and Emergency Events. *International Journal of Emergency Management*, *6*(3/4), 248–260.
6. Hui, C., Tyshchuk, Y., Wallace, W., Goldberg, M., & Magdon-Ismail, M. (2012). Information cascades in social media in response to a crisis: a preliminary model and a case study. In: *WWW* (pp. 653–656).
7. 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. 591–600). ACM. doi:10.1145/1772690.1772751
8. Liu, S.B. & Palen, L. The new cartographers: crisis map mashups and the emergence of neogeographic practice, Cartography and Geographic Information Science, 37 (2010) 69-90.
9. MacEachren, A.M., Jaiswal, A., Robinson, A.C., Pezanowski, S., Savelyev, A., Mitra, P., Zhang, X., Blanford, J. SensePlace2: Geotwitter Analytics Support for Situation Awareness, in: IEEE Conference on Visual Analytics Science and Technology, Providence, RI, 2011.
10. Mandel, B., Culotta, A., Boulahanis, J., Stark, D., Lewis, B., and Rodrigue, J. 2012. A demographic analysis of online sentiment during hurricane Irene. In Workshop on Language in Social Media '12.
11. McClendon, S., & Robinson, A. C. (2012). Leveraging Geospatially-Oriented Social Media Communications in Disaster Response. In: *The International ISCRAM Conference 2012* (pp. 2–11).
12. Mendoza, M., Poblete, B., & Castillo, C. (2010). Twitter under crisis. *Proceedings of the First Workshop on Social Media Analytics SOMA 10* (pp. 71–79).
13. Nagy, A. and Stamberger, J (2012) Crowd Sentiment Detection during Disasters and Crises. In: Proceedings of the 9th International ISCRAM Conference, Vancouver, CA.
14. 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/.
15. Palen, L., S. Vieweg, S. Liu, and A.L. Hughes. (2009) Crisis in a Networked World: Features of Computer-Mediated Communication in 2007 Virginia Tech Event. *Social Science Computer Review*.
16. Pang, B., Lee, L. and Vaithyanathan, S. Thumbs up?: sentiment classification using machine learning techniques. In ACL-02, EMNLP '02, pages 79–86, Stroudsburg, PA, USA, 2002.
17. Portsea, L. J. (2011). *Disaster Relief 2.0: The Future of Information Sharing in Humanitarian Emergencies*. *Disasters*. pp. 1–8. UN Foundation & Vodafone Foundation Technology Partnership.
18. McDonald, R., Hannan, K., Neylon, T., Wells, M., and Reynar, J. Structured models for fine-to-coarse sentiment analysis. In ACL '07, volume 45, page 432, 2007.
19. Sakaki, T., Toriumi, F., Shinoda, K., Kazama, K., Kurihara, S., Noda, I. and Matsuo, Y. (2013) Regional Analysis of User Interactions on Social Media in Times of Disaster. In: *WWW 2013*, Brazil.
20. Sakaki, T., Okazaki, M., & Matsuo, Y. (2010). Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors. *WWW 2010, April 26-30* (pp. 851–860). Raleigh, North Carolina: ACM.
21. Schulz, A., Thanh, T.D., Paulheim, H., Schweizer, I. (2013) A Fine-Grained Sentiment Analysis Approach for Detecting Crisis Related Microposts. In: *ISCRAM 2013*.
22. Starbird, K. (2010). Pass It On? : Retweeting in Mass Emergency, (December 2004), 1–10.
23. Starbird, K. Digital Volunteerism During Disaster: Crowdsourcing Information Processing, in: CHI '11 Workshop on Crowdsourcing and Human Computation, Vancouver, BC, 2011.
24. 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. 241–250). New York.
25. Starbird, K., Munzy and Palen, L. (2012) Learning from the Crowd: Collaborative Filtering Techniques for Identifying On-The-Ground Twitterers during Mass Disruptions." In: ISCRAM 2012.
26. Sutton, J., Palen, L., and I. Shklovski. (2008) Backchannels on the Front Lines: Emergent Use of Social Media in the 2007 Southern California Fires. *ISCRAM*.
27. 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. 1–10).
28. Tapia, A. Moore, K. Johnson, N. "Beyond the Trustworthy Tweet: A Deeper Understanding of Micro-blogged Data Use by Disaster Response and Humanitarian Relief Organizations", In: ISCRAM 2013.
29. Terpstra, T. (2012). Towards a realtime Twitter analysis during crises for operational crisis management. *Proceedings of International ISCRAM Conference 2012* (pp. 1–9).
30. Thomson, R., N. Ito, H. Suda, F. Lin, Y. Liu, R. Hayasaka, R. Isochi, and Z. Wang. (2012) Trusting Tweets: The Fukushima Disaster and Information Source Credibility on Twitter. In: *ISCRAM 2012*.
31. Walton, R., Mays, R., & Haselkorn, M. (2011). Defining " Fast ": Factors Affecting the Experience of Speed in Humanitarian Logistics. In: *ISCRAM* (pp. 1–10). Lisbon, Portugal.