AAAI Technical Report WS-13-03
Social Media Visualization 2

Visualizing Community Resilience Metrics from Twitter Data

Robert M. Patton, Chad A. Steed, Christopher G. Stahl
Oak Ridge National Laboratory, PO Box 2008, Oak Ridge, TN, USA, 37831
{pattonrm, steedca, stahlcg}@ornl.gov

Abstract

The recent explosive growth of smart phones and social media creates a unique opportunity to view events from various unique perspectives. Unfortunately, this relatively new form of communication lacks the structural integrity, accuracy, and reduced noise of other forms of communication. Nevertheless, social media increasingly plays a vita role in the observation of societal actions before, during, and after significant events. In October 2012, Hurricane Sandy making landfall on the northeastern coasts of the United States demonstrated this role. This work provides a preliminary view into how social media could be used to monitor and gauge community resilience to such natural disasters. We observe, evaluate, and visualize how Twitter data evolves over time before, during, and after a natural disaster such as Hurricane Sandy and what opportunities there may be to leverage social media for situational awareness and emergency response.

Introduction

Originally developed for entertainment purposes, social media systems have rapidly evolved to provide valuable benefits for such areas as business intelligence, national security, and disaster management. In combination with the development and adoption of smart phone technology, people have become mobile sensors providing “eyes” and “ears” as events unfold. Leveraging this capability provides significant advantages for situational awareness.

Unfortunately, this technology contains significant noise and error in the data as well as the inability to position the “sensors” in critical areas. Consequently, leveraging this capability can be quite challenging depending on the application purposes. Research into this issue has opened up several opportunities to overcome this problem. The work described here is a preliminary investigation into resolving the data noise issue as it relates to disaster management.

This work narrowly focuses on resolving a specific problem: identifying and characterizing the community resilience to natural disasters relating to physical infrastructure and social behavior using information obtained via Twitter. Community resilience is a measure of a community’s ability to prepare, respond, withstand, and recover from disasters both natural and man-made. In the unfortunate event of a disaster, communities that have low resilience will suffer significant loss of life and damage to critical infrastructure. In addition, communities with low resilience require longer periods of time to recover to pre-disaster operations and quality of life. Communities with higher resilience suffer fewer losses and return more quickly to pre-disaster operation and quality of life. Unfortunately, when a disaster occurs, the effects of any weakness in a community’s resiliency will be amplified. Any ability for disaster management personnel and leadership to observe in real-time the effects of a disaster on a community provides a significant advantage and opportunity to respond more quickly. Consequently, the work described here is focuses on utilizing social media, specifically the Twitter platform, as a means of providing a real-time view into the impacts of a disaster on the community.

Related Works

Investigating the impacts of natural disasters on society is not a new research area. There are many works relating to the identification of different impacts as well as different studies that have been performed (Alexander 1997) (Little 2003) (Kreimer, Arnold, and Carlin 2003). Many of these works and studies, however, have identified impacts that are difficult to measure quantitatively either before, during, or after the event. In addition, some works have relied on remote sensing techniques for assessing and monitoring disaster situations (Llinas 2002) (Chi, et. al. 2003) (Hussain, et. al. 2005). For our work, the goal is to develop an approach that provides a more quantitative view of the event impact as well as incorporate field reporting about the event via Twitter.

More recently, research has shifted into utilizing social media for disaster management. In (Chu, et. al. 2012), a system for crowdsourcing the collection of information for disaster management is described. Leveraging social networks, information is received and exchanged between volunteers in the field and disaster management command and control (DMCC). The DMCC can request the volunteers to visit specific locations and report on what is seen and heard, thereby making the DMCC much more adaptive to the circumstances. Unlike the work of (Chu, et. al. 2012) which relies on some structured data (e.g., automated geo-tagging of the volunteer's mobile device) to reduce the noise, the work of (Jie, et. al. 2012) uses natural language processing and data mining to extract situational awareness from Twitter data. Utilizing Twitter's location based search API, tweets are collected for specific areas and then processed and mined to provide a view into the current situation of a specific area. Results are then displayed as a combination of tag clouds and map-based user interfaces.

The primary problem with the two previous works is that they are very granular in detail as it relates to situational awareness. In disaster management, there are various "layers" of detail that allow disaster response personnel to view the event at different levels of abstraction. The two previous works provide little to no layers of abstraction. In the work of (Birregah, et. al. 2012), Twitter data is processed in increasingly higher levels of abstraction such that individual tweets are eventually grouped into areas such as Health, Transport, Communication, and Lodging. While their work provides a critical connection of tweets to higher-level abstractions, what is lacking is a visualization or user interface to support a "top-down" view of the different layers of abstraction. The work described here attempts to address this deficit.

Finally, the work described here is an extension of our previous work in this area. In (Patton, et. al. 2008), news media was investigated in order to better understand various impacts of a disaster on society. In that work, the data source tended to be less noisy in terms of the language that was used. News media tend to be more factual and use more structured language as well as correct spellings. In contrast, the Twitter data uses abbreviated or incorrect syntax and spelling, and can be less factual and more biased to the individual needs and wants of a person. As addressed in (Birregah, et. al. 2012) work, a bottom up approach was developed to identify trends in a larger population of people based on the tweets that depict individual wants and needs. The approach described here supports visualizing the higher-level trends and enables the ability to dive into details for specific areas of concern.

Approach

The primary focus is to detect, analyze, and characterize community resiliency metrics from events impacting society as observed in Twitter data. Inspiration for this work partially originated from the concept of decomposing a time-series into sub-series as discussed in (Yu, et. al. 2001). Figure 1 shows the conceptual view of our approach.

The original data begins as a time-series of all of the word and phrase counts that are observed in the Twitter fire hose. Next, a "textual prism" is created that is comprised of a set of taxonomies that describe the words and phrases related to the topics of interest by the user. The taxonomies are defined by the user and applied to the original time-series in order to produce a component time-series. This component time-series shows how the words and phrases of the taxonomy change over time.

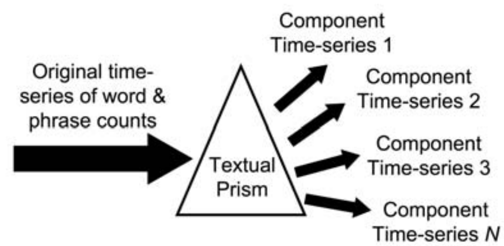

Figure 1. Concept of approach

As discussed previously, there have been investigations into the various impacts caused by natural disasters. For this work, the research performed by (Bruneau, et. al. 2003) provides the foundation for the impacts to be investigated. This work was chosen for its simplicity and extensibility. In (Bruneau, et. al. 2003), impacts can be observed in 4 areas: Technical, Organizational, Social, and Economic. Technical refers to the infrastructure (roads, bridges, power grid, water systems, etc.) of a society. Organizational refers to areas such as service crews that maintain or respond to the Technical aspects. Social refers to areas related to housing, shelters, provisions for human needs, etc. Economic refers to the impacts on the economy. This work focuses on the Technical and Social impacts of a natural disaster.

To begin, a set of taxonomies were developed for both the Technical and Social categories as shown in Table 1. These taxonomies were developed manually by analyzing news reports from the time period of August 25, 2005 to September 5, 2005. This is the time period when Hurricane Katrina made landfall in the Gulf Coast of the United States. This natural disaster created significant and widespread damage and resulted in extensive impacts on

U.S. society. The news reports during this time period were clustered. The clusters were then analyzed to determine the most popular words and phrases that were used to describe specific conditions. These words and phrases were then categorized according to the framework defined in (Bruneau, et. al. 2003) as shown in Table 1.

As a specific example, the Shelters taxonomy consists of word and phrases such as: taking cover, taking shelter, seek refuge, shelters. The Movement taxonomy consists of words and phrases such as: mandatory evacuation, dawn curfew, flights canceled, etc. The Power taxonomy consists of phrases such as: downed power lines, backup power, emergency generator, and emergency power. The Roads taxonomy consists of phrases such as: alternate route, traffic backups, evacuation route, major streets, and interstates.

Table 1. Natural Disaster Impact Taxonomies

Technical	Social
Communications	Crime
Fuel	Death
Light	Health Hazards
Outage	Medical
Power	Movement
Rail	Shelter
Roads	Water (Health)
Structures	
Water (Infrastructure)	

After creating these taxonomies, they are then applied to the Twitter data as a form of "textual prism" to create component time-series as described in Figure 1. Every word and phrase observed from these categories is counted each day. For this particular investigation, the data was focused on the Hurricane Sandy event, and was collected using the Twitter location-based search to identify tweets in the geographic area of impact. When visualizing the volume of tweets for each taxonomy on the same chart, trends in numerically larger scales can obscure significant trends in smaller scales. To account for the volume of tweets for a particular taxonomy, the total counts of each day for each category was converted to be the percentage of total tweets for that taxonomy across the entire time that tweets were collected. For example, the Shelters taxonomy may have a total of 10,000 tweets over a 10-day period while the Medical taxonomy may have a total of 50 tweets over the same time period. On day 1, the Shelters total tweet count may have been 1,000 while the Medical may have been 5. For each taxonomy, the number of tweets on day 1 is 10% of the total number of tweets over the 10-day period, respectively. Thus, the data is normalized to a scale of 0 to 100.

Results & Discussion

The results from the approach described previously are shown in Figures 2 and 3. Figure 2 shows the trends of the Social taxonomies of Table 1. Figure 3 shows the trends of the Technical taxonomies.

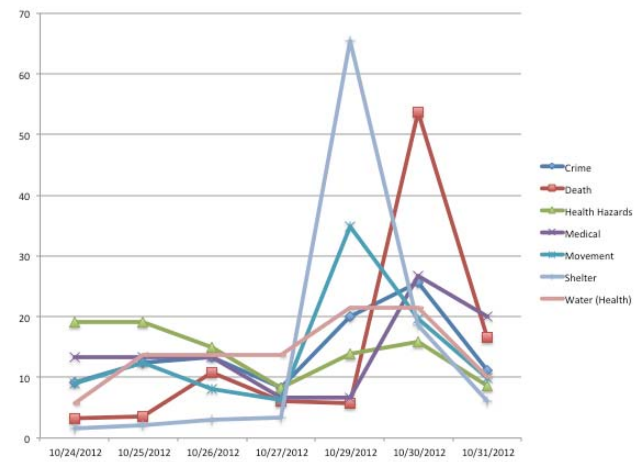

Figure 2. Social Taxonomies

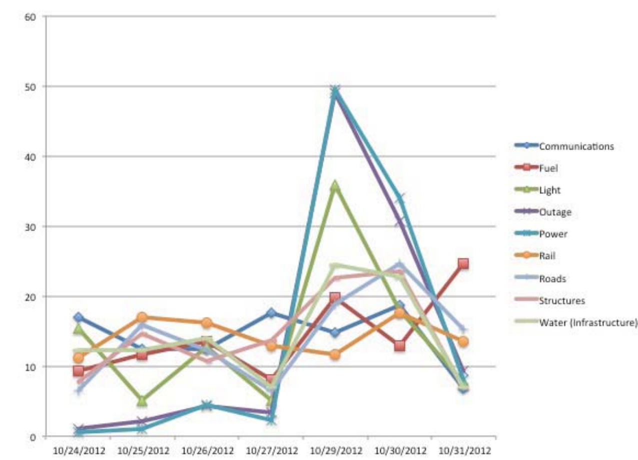

Figure 3. Technical Taxonomies

As Hurricane Sandy makes landfall in New Jersey around midnight of October 29th, the trends in nearly every area begin moving upward. Of particular note, the Shelter and Movement taxonomies begin trending upward considerably and prior to the upward trends Death and Medical taxonomies. Unfortunately, most of the tweets in the Shelter and Movement taxonomies are not particularly useful. Many of the tweets are either originating from news agencies, retweets from news agencies, or tweets that contain information that could more easily be obtained from news articles. Furthermore, many of the tweets are simply wishing or praying for people to stay safe and seek shelter.

In regard to trends, each taxonomy experiences an increasing trend as Hurricane Sandy approaches landfall and through the initial hours of landfall, but then quickly start declining. One exception of particular notice is the Fuel taxonomy as shown in Figure 3. While the other taxonomies are declining, the Fuel taxonomy makes a significant increase on the last day of our collection. Analyzing the individual tweets from this taxonomy reveals that a number of the tweets are from news agencies that are reporting shortages at gas stations. Further, there are an increasing number of tweets by individuals expressing frustration at the gas shortages and long lines at gas stations. For example, one tweet stated "people are lining up at gas stations where there isn't even gas just waiting for gas to come this is scary." Another tweet stated, "what gas stations are open and are not packed with hundreds of people. i highly doubt any." In the days and weeks after Hurricane Sandy, this gas shortage becomes a significant problem resulting in near riots and citizens venting their anger in various ways. Consequently, there is an opportunity here to supplement this preliminary work with sentiment analysis as well as additional taxonomies to capture social unrest in order to further capture the dynamics of the community as an event transpires.

Summary & Future Work

This work is a preliminary investigation into the use of community resilience taxonomies applied to Twitter data. Visualizing these taxonomies over time provides insight into how communities respond to events in a variety of different ways. Our approach helps curb the effects of noise and errors in the Twitter data and provides a "top-down" visual analytics approach to leveraging this data for situational awareness purposes.

Additional work will incorporate sentiment analysis and additional taxonomies to capture more detail in the data as well as additional filters to further reduce noise and errors in the data. Furthermore, additional visualization involving parallel coordinates and a more sophisticated user interface to enable the user to thoroughly analyze the data will be developed.

Acknowledgements

Research sponsored by the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory, managed by UT-Battelle, LLC, for the U. S. Department of Energy. LDRD #6427.

Prepared by Oak Ridge National Laboratory, P.O. Box 2008, Oak Ridge, Tennessee 37831-6285; managed by UT-Battelle, LLC, for the U.S. Department of Energy under contract DE-AC05-00OR2225. This manuscript has been authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 for the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains non-exclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

References

Alexander, D. 1997. The Study of Natural Disasters, 1977-97: Some Reflections on a Changing Field of Knowledge, *Disasters*, Vol. 21, No. 4, pages 284–304.

Birregah, B.; Top, T.; Perez, C.; Chatelet, E.; Matta, N.; Lemercier, M.; Snoussi, H. 2012. Multi-layer Crisis Mapping: A Social Media-Based Approach. In Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), 2012 IEEE 21st International Workshop on , 379-384.

Bruneau, M., et. al. 2003. A Framework to Quantitatively Assess and Enhance the Seismic Resilience of Communities, *Earthquake Spectra*, Vol. 19, No. 4, pages 733-752.

Chi, T., et. al. 2003. Research on information system for natural disaster monitoring and assessment. In Proc of the 2003 IEEE International Geoscience and Remote Sensing Symposium.

Chu, E.T.; Yi-Lung Chen; Jyun-You Lin; Liu, J.W.-S. 2012. Crowdsourcing support system for disaster surveillance and response. In 2012 15th International Symposium on Wireless Personal Multimedia Communications (WPMC). pp.21-25.

Hussain, M., et. al. 2005. Emerging geo-information technologies (GIT) for natural disaster management in Pakistan: an overview. In Proc. of the 2nd International Conference on Recent Advances in Space Technologies.

Jie, et. al. 2012. Using Social Media to Enhance Emergency Situation Awareness. *Intelligent Systems* 27(6): 52-59.

Kreimer, A., Arnold, M., and Carlin, A. 2003. *Building Safer Cities: The Future of Disaster Risk*, World Bank Publications.

Little, R.G. 2003. Toward more robust infrastructure: observations on improving the resilience and reliability of critical systems. In Proc. of the 36th Annual Hawaii International Conference on System Sciences.

Llinas, J. 2002. Information fusion for natural and man-made disasters. In Proc. of the 5th International Conference on Information Fusion. pp.570-576.

Patton, R.M., et. al. 2008. Discovery, analysis, and characteristics of event impacts. In 2008 11th International Conference on Information Fusion. pp.1-8.

Yu, J. X., et. al. 2001. Patterns discovery based on time-series decomposition. In Proc. of 5th Pacific Asia Knowledge Discovery and Data Mining.