Tweet Factors Influencing Trust and Usefulness During Both Man-Made and Natural Disasters

Shane E. Halse
Pennsylvania State University
seh297@ist.psu.edu

Andrea Tapia
Pennsylvania State University
atapia@ist.psu.edu

Anna Squicciarini
Pennsylvania State University
asquicciarini@ist.psu.edu

Cornelia Caragea
University of North Texas
Cornelia.Caragea@unt.edu

ABSTRACT

To this date, research on crisis informatics has focused on the detection of trust in Twitter data through the use of message structure, sentiment, propagation and author. Little research has examined the usefulness of these messages in the crisis response domain. Toward detecting useful messages in case of crisis, in this paper, we characterize tweets, which are perceived useful or trustworthy, and determine their main features. Our analysis is carried out on two datasets (one natural and one man made) gathered from Twitter concerning hurricane Sandy in 2012 and the Boston Bombing 2013. The results indicate that there is a high correlation and similar factors (support for the victims, informational data, use of humor and type of emotion used) influencing trustworthiness and usefulness for both disaster types. This could have impacts on how messages from social media data are analyzed for use in crisis response.

Keywords

Twitter. Sandy. Hurricane. Boston Bombing. Trust. Usefulness.

INTRODUCTION

We believe that data directly contributed by citizens, and data scraped from bystanders witnessing a disaster, have a strongly positive potential to give responders more accurate and timely information than is possible with traditional information gathering methods. Many emergency decision makers see the data produced through crowd sourcing and social media as ubiquitous, rapid and accessible. In response to increased online public engagement and the emergence of digital volunteers, professional emergency responders have sought to better understand how they too can use online media to communicate with the public and collect intelligence (Latonero & Shklovski, 2011; Sutton et al., 2015; Vieweg et al. 2010).

Due to the perceived lack of authentication and validation of content posted in Twitter micro-blogs, large-scale responders have been reluctant to incorporate social media data into the process of assessing a disaster situation. Committing to the mobilization of valuable and time sensitive relief supplies and personnel, based on what may turn out be illegitimate claims, has been perceived to be too great a risk. Incorporating the products of digital volunteer activity into professional emergency practice has proved to be challenging due to issues with credibility, liability, training, and organizational process and procedure (Tapia, Bajpai, Jansen, Yen, & Giles, 2011; Tapia, Moore, & Johnson, 2013; Starbird & Paylen, 2013).

Our study resulted in a wealth of useful and informative insights. In particular, our analysis allows us to take a first step toward understanding whether it is possible to characterize the main features of potentially actionable tweets that should be perceived as trustworthy and/or useful. We focused on addressing the following research

questions: For two different disaster types, man-made and natural, are the perceived usefulness and trustworthiness related, and is there a difference in the factors that affect perceived usefulness and trustworthiness?

Our findings indicate that there is a significant correlation between the perceived usefulness and trustworthiness of a tweet across both man-made and natural crisis. In addition, the use of features found in previous research, such as geolocation data, are supported and reaffirmed.

RELEVANT LITERATURE

Using social media 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 (Mendoza, Poblete, & Castillo, 2010; Starbird, Palen, Hughes, & Vieweg, 2010; Tapia, Bajpai, Jansen, Yen, & Giles, 2011).

The research involving veracity in technologically mediated environments has had two distinct approaches. The first approach looks at the person supplying the information, while the second looks at the information itself. Identifying the reliability, credibility and position of who is providing information are extremely valuable factors to establish trustworthiness (Grabner-Kräuter, Kaluscha, & Fladnitzer, 2006). From the information side, information in a post may be considered verifiable when linked to a credible source (Starbird, Palen, Hughes, & Vieweg, 2010a), or when it is corroborated through multiple sources (Giacobe, Kim, & Faraz, 2010). Related to reputation, a micro-blogger who self-corrects information, or responds to criticism of information may be deemed credible and reliable (Shklovski, Palen, & Sutton, 2008).

In recent work by (Gupta, Kumaraguru, & Castillo, 2014), a real-time system, called *TweetCred*, was developed to assign a credibility score to tweets in a user's timeline, however, not being focused on emergency-related tweets. (Castillo, Mendoza, & Poblete, 2011) developed automatic methods to assess the credibility of tweets related to specific topics or events using features extracted from users' posting behavior and the tweets' social context. Their work tried to model whether end-users would believe the information reported in tweets is true or not, but was not concerned with detecting whether the information in tweets was itself accurate or useful. (Dailey & Starbird, 2014) explored techniques such as *visible skepticism* to help control the spread of false rumors, but did not intend to automatically detect false rumors. Most research in disaster-related area has been performed post-hoc, and the most important aspect of any intelligence received, intelligence that is actionable and precisely geo-located, has not yet been achieved and is also complicated by translation and language understanding (McClendon & Robinson, 2012).

METHODS

In 2014 we started the complex process of creating gold-standard datasets of disaster-related data related to the Hurricane Sandy and Boston bombings and derived a sets of rules. Based on these derived rules, we used Mechanical Turk (MTurk) to manually label the tweets. In total about 2609 unique tweets were labeled (some had multiple labelers). Utilizing crowd sourced data such as Amazon Turk is representative of the volunteers that could be found during a crisis and is supported by the works by (Dailey & Starbird, 2014) and (Ann, Denis, & Hughes, 2012).

Two main dependent variables were investigated, the perceived trustworthiness and usefulness of each tweet. Each variable has 4 levels not including the "I don't know condition". The variables were coded as follows: DefT/VeryUse = 3, MaybeT/MaybeUse = 2, MaybeUnt/MaybeNotUse = 1, MostCUnt/NotUse = 0 and UnkT/IDKUseThe = NA. Participants were shown a tweet along with 9 questions asking about their perceptions of that tweet as seen in table 1. For the perceived trustworthiness users were asked "How trustworthy would you rate this tweet?" Likewise for the perceived usefulness users were asked "How useful would this tweet be to first responders (i.e. units trying to provide help)?"

In total there were 7 independent variables investigated through 6 questions in the format of "yes" (value = 1), "no" (value = 0), and "I don't know" (value = NA) responses. In addition there was a questions related to the emotion expressed within the tweet. This question was designed as an 8 point categorical scale with responses such as "anger", "disgust", "fear", "happiness", "sadness", "surprise", "neutral", and "irrelevant".

Table 1 - Independent Variable Coding Schema

Variable	Description	Response Codes
AboutDis	Is this Tweet about the disaster in question ?	IDKDis, NoDis, YesDis

Support	Does this tweet offer support for the victims of the disaster?	IDKSup, NoSup, YesSup
ShowEmo	Does this tweet express any emotion to the victims of the disaster?	IDKEmo, NoEmo, YesEmo
EmoType	Which emotions does this tweet express?	anger, disgust, fear, happiness, sadness, surprise, neutral, irrelevant
Humor	Does this tweet use humor in some way in relation to the disaster?")	IDKHum, NoHum, YesHum
Info	Does this tweet offer information about this disaster (not emotion, but facts)?")	IDKInfo, NoInfo, YesInfo
GeoData	Does this tweet contain information that could place the tweeter in a geographic location?	IDKgeo, Nogeo, Yesgeo

Data description

The first data set used in our experiment was 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 using the hashtag sandy and hurricane. We randomly sampled a subset of 1711 tweets from the crawled data for our labeling tasks. The second data set used in our experiment was collected from Twitter using the hashtag prayforboston, Boston, bomb, and bombs during the Boston Bombing incident between 04-15-2013 and 04-25-2013 and contained 23,642,905 tweets. We randomly sampled a subset of 898 tweets from the crawled data for our labeling tasks.

Labeling was done, and the inter-labeler agreement was addressed as follows: For yes/no/idk questions the average (yes = 1, no = 0, idk = NA) was taken. Those with greater than a 0.5 average were considered yes, those with less than 0.5 considered no. In the case of a ties (avg = 0.5) we used NA. For the usefulness and trustworthiness score and average value (coding described above) was assigned across all labelers. Lastly, for the Emotype the emotion with highest combined score was assigned, in the case of a tie, NA was used.

Overall for the Hurricane Sandy dataset, we employed 1702 workers, who labeled 1711 unique tweets, each worker labeled multiple tweets. Of these, we disregarded 149 (8.71 %) tweet messages and their corresponding labeler's data as these were perceived as not being related to Hurricane Sandy. For the Boston bombing dataset a total of 898 unique tweets were labeled. Of these, we disregarded 207 (23.05%) tweet messages and their corresponding labeler's data as these were perceived as not being related to the Boston Bombing. Note: as our dataset for the Boston bombing contained the filter "Boston" this resulted in a large number of tweets that were unrelated to the bombing. This allowed us to have tweets that were considered relevant to the Sandy Hurricane and the Boston Bombing by the labelers.

Statistical techniques

Multiple tests were completed in order to fully understand and describe the data. A significance level of $\rho < 0.05$ was used. The following two hypotheses were developed:

> H_1: There is a significant correlation between perceived usefulness and trustworthiness for both man-made and natural disasters
>
> H_2: There is no significant difference in the factors that affect perceived trustworthiness and usefulness across both man-made and natural disasters

For both datasets a five-way, 3 (Support) x 8 (EmoType) x 3 (Humor) x 3 (Info) x 3 (Geodata), factorial analysis of variance (ANOVA) was conducted to determine the influence of five independent variables on the perceived trustworthiness and usefulness of a tweet based on research by (Field, 2007). This allowed the researchers to examine which of the main effects and which of the interaction effects of the factors contributed significantly to the trustworthiness and usefulness scores and was consistent with those found in previous research.

FINDINGS

Correlation between trustworthiness and usefulness (Pearson's Correlation)

In order to investigate if the message's perceived usefulness and perceived trustworthiness were related a Pearson's correlation was run on each data set. The findings indicate that both the Sandy dataset ($r = 0.4574168$, $\rho < 2.2e\text{-}16$) and the Boston dataset ($r = 0.3228806$, $\rho < 2.2e\text{-}16$) had significant positive correlations between the two factors. This allowed us to accept the H_1 hypothesis. That is, it provides evidence that is a tweet is perceived as trustworthy then it would also be perceived useful.

Factors influencing Trustworthiness and Usefulness (ANOVA)

Sandy Dataset

For the trustworthiness score within the hurricane Sandy dataset (summarized in table 2) all main effects were found to be statistically significant ($\rho < 0.05$) and included; support, emotion type, humor, info and geodata. The interaction effects found to be significant was EmoType:Info indicating that when both emotion and information is present the trustworthiness is affected.

Table 2 - Sandy ANOVA Summary (DV: Trustworthiness, Only $\rho < 0.05$ shown)

Effects	F	Sig
Support	94.572	0.0000
EmoType	42.297	0.0000
Humor	143.066	0.0000
Info	124.035	0.0000
GeoData	40.072	0.0000
EmoType:Info	1.985	0.0155

For the usefulness score (summarized in table 3) all main effects were found to be statistically significant ($\rho < 0.05$) and included; support, emotion type used, humor, info and geodata. One of the more interesting findings in this dataset was that in terms of usefulness, a 5-way interaction effect was found to be significant between Support:EmoType:Humor:Info:GeoData.

Table 3 - Sandy ANOVA Summary (DV: Usefulness, Only $\rho < 0.05$ shown)

Effects	F	Sig
Support	463.723	0.0000
EmoType	27.611	0.0000
Humor	83.099	0.0000
Info	1121.28	0.0000
GeoData	90.02	0.0000
Support:EmoType	1.916	0.0206
Support:Info	6.121	0.0001
Humor:Info	3.204	0.0123
Support:GeoData	3.165	0.0132
EmoType:GeoData	2.113	0.0089
Humor:GeoData	3.194	0.0125
Support:EmoType:Humor	1.725	0.0210
Support:EmoType:GeoData	2.316	0.0021
Support:Info:GeoData	2.882	0.0053
EmoType:Info:GeoData	2.406	0.0008
Support:EmoType:Humor:Info	2.924	0.0076
Support:EmoType:Info:GeoData	2.43	0.0239

EmoType:Humor:Info:GeoData	3.59	0.0000
Support:EmoType:Humor:Info:GeoData	4.497	0.0340

Boston Dataset

For the trustworthiness score (summarized in table 4) all main effects were found to be statistically significant ($\rho < 0.05$) and included; support, emotion type used, humor, info and geodata. The interaction effects found to be significant included; Support:Humor. In addition a 3 way interaction effect was found between Support:Emotype:Humor which provides support for these factors influencing the trustworthiness.

Table 4 - Boston ANOVA Summary (DV: Trustworthiness, Only $\rho < 0.05$ shown)

Effects	F	Sig
Support	19.489	0.0000
EmoType	22.755	0.0000
Humor	106.352	0.0000
Info	59.963	0.0000
GeoData	10.948	0.0000
Support:Humor	3.176	0.0130
EmoType:Info	5.287	0.0000
Support:EmoType:Humor	2.604	0.0079

For the usefulness score (summarized in table 5) all main effects were found to be statistically significant ($\rho < 0.05$) and included; support, emotion type used, humor, info and geodata. There was a 4-way interaction effect found to be significant between EmoType:Humor:Info:GeoData.

Table 5- Boston ANOVA Summary (DV: Usefulness, Only $\rho < 0.05$ shown)

Effects	F	Sig
Support	11.138	0.0000
EmoType	20.515	0.0000
Humor	28.138	0.0000
Info	342.659	0.0000
GeoData	39.397	0.0000
EmoType:Info	4.01	0.0000
EmoType:GeoData	1.989	0.0183
Info:GeoData	3.122	0.0143
EmoType:Humor:GeoData	2.5	0.0206
EmoType:Info:GeoData	3.291	0.0002
EmoType:Humor:Info:GeoData	4.941	0.0072

Summary of ANOVA data

Below is a summary table that examines the significant factors that can influence both perceived usefulness and perceived trustworthiness

Table 6 - Summary table indicating significant main effects

Factor	Sandy Data (Trustworthiness)	Sandy Data (Usefulness)	Boston Data (Trustworthiness)	Boston Data (Usefulness)

Support	✓	✓	✓	✓
EmoType	✓	✓	✓	✓
Humor	✓	✓	✓	✓
Info	✓	✓	✓	✓
GeoData	✓	✓	✓	✓

From the table above we can see that the factors selected to measure both trustworthiness and usefulness across both main made and natural disasters yielded no significant difference. This allows us to accept the H_2 hypothesis.

DISCUSSION

Despite the barriers of social media data adoption being broad and numerous, the advantages and potential information that can be found within far outweigh the difficulty of obtaining it. (Tapia et al., 2011) have already described pockets of use of social media data and illustrate both the frustration and hope of one-day being able to use this data effectively. As seen in the data analysis section of this paper the correlation between perceived usefulness and trustworthiness contained significant positive correlations for both man-made and natural disasters. While measuring trust could yield fruitful results, we argue that this would not be enough to adequately reduce the amount of data found during a crisis. Instead by detecting the usefulness of a Tweet in conjunction with its trustworthiness the data would become much more relevant to crisis response and thus increase the adoption of its use.

In addition, this research allowed the investigation of two datasets and showed that despite the differences in the crisis type (man-made versus natural) the predictors of trustworthiness and usefulness were significantly similar. Not only does this research contribute additional factors for measuring trust, but it aligns with previous research. For example, the findings show that the use of Geolocational data play a significant role in the generation of trust (Vieweg, Hughes, Starbird, & Palen, 2010). Other significant factors that influence the perceived trustworthiness and usefulness included Tweets in which; provided support for the victims, provided informational data about the disaster, the use of humor in the tweet and type of emotion used within the tweet.

CONCLUSIONS

To fully develop an automated trust score for Twitter tweets is no easy task. This research however has shown another angle of attack that can be used to generate this score. By discovering the close correlations between the perceived trustworthiness and usefulness score, along with the factors that have significant effects on both of them, it could allow one to investigate perceived trust by investigating the perceived usefulness. Further as social media data provide a plethora of data, tools that would generate a trust assessment based on both man-made and natural disasters should use the key factors of; tweet usefulness, support for victims, informational data, geolocational data and humor usages.

Rather than having valuable resources devoted to sifting through and analyzing every tweet during a crisis; tools or volunteers could be used during this tedious process. In eliminating tweets that are neither useful nor trustworthy, only the tweets significant to the crisis responders would be presented. In providing both the usefulness and trustworthiness rankings these tools would be able provide more relevant information to first responders. For example, a message may have a large amount of factual data that would be considered truthful as it is verifiable, however if this data is not about the crisis it is irrelevant. Likewise, tweets that a tool may detect as untrustworthy, but still useful, could still provide valuable information that would otherwise be ignored and mark it for further vetting. This in turn would allow the people within crisis response to have a better situational awareness thus making them more effective and allowing them to make better decisions in time critical situations.

REFERENCES

Ann, L., Denis, S., & Hughes, A. L. (2012). Trial by Fire : The Deployment of Trusted Digital Volunteers in the 2011 Shadow Lake Fire, (April), 1–10.

Castillo, C., Mendoza, M., & Poblete, B. (2011). Information credibility on twitter. *Proceedings of the 20th International Conference on World Wide Web - WWW '11*, 675. http://doi.org/10.1145/1963405.1963500

Dailey, D., & Starbird, K. (2014). Visibl e Skepticism : Community Vetting after Hurricane Irene, 775–779.

Field, C. A. P. (2007). *Analysis of Variance (ANOVA)*. (N. J. Salkind & K. Rasmussen, Eds.). Thousand Oaks: Sage Publications, Inc. http://doi.org/http://dx.doi.org/10.4135/9781412952644.n19

Giacobe, N. A., Kim, H.-W., & Faraz, A. (2010). Mining social media in extreme events : Lessons learned from the DARPA network challenge. In *2010 IEEE International Conference on Technologies for Homeland Security (HST)* (pp. 165–171). Waltham, MA: IEEE. http://doi.org/10.1109/THS.2010.5655067

Grabner-Kräuter, S., Kaluscha, E. A., & Fladnitzer, M. (2006). Perspectives of online trust and similar constructs. In *Proceedings of the ICEC '06* (p. 235). New York, New York, USA: ACM Press.

Gupta, A., Kumaraguru, P., & Castillo, C. (2014). TweetCred : Real-Time Credibility Assessment, 228–243.

Latonero, M., & Shklovski, I. (2011). Emergency Management , Twitter , and Social Media Evangelism, *3*(December), 1–16. http://doi.org/10.4018/jiscrm.2011100101

McClendon, S., & Robinson, A. C. (2012). Leveraging Geospatially-Oriented Social Media Communications in Disaster Response. *Iscram*, (April), 2–11. http://doi.org/10.4018/jiscrm.2013010102

Mendoza, M., Poblete, B., & Castillo, C. (2010). Twitter Under Crisis: Can we trust what we RT? *Workshop on Social Media Analytics*, 9. http://doi.org/10.1145/1964858.1964869

Munro, R. (2011). Subword and spatiotemporal models for identifying actionable information in Haitian Kreyol, (June), 68–77.

Shklovski, I., Palen, L., & Sutton, J. (2008). Finding community through information and communication technology in disaster response. *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work. ACM*, 127–136. http://doi.org/10.1145/1460563.1460584

Starbird, K., Palen, L., Hughes, A. L., & Vieweg, S. (2010a). Chatter on the Red: What Hazards Threat Reveals About the Social Life of Microblogged Information. In *CSCW '10* (pp. 241–250). New York, USA

Starbird, K., Palen, L., Hughes, A. L., & Vieweg, S. (2010b). Chatter on The Red: What Hazards Thret Reveals about the Social Life of Microblogges Information. *Proceedings of the CSCW*, 241–250.

Starbird, K., & Paylen, L. (2013). Working & Sustaining the Virtual " Disaster Desk ."

Sutton, J., Gibson, C. Ben, Spiro, E. S., League, C., Fitzhugh, S. M., & Butts, C. T. (2015). What it Takes to Get Passed On: Message Content, Style, and Structure as Predictors of Retransmission in the Boston Marathon Bombing Response. *Plos One*, *10*(8), e0134452. http://doi.org/10.1371/journal.pone.0134452

Tapia, A. H., Bajpai, K., Jansen, J., Yen, J., & Giles, L. (2011). Seeking the Trustworthy Tweet: Can Microblogged Data Fit the Information Needs of Disaster Response and Humanitarian Relief Organizations. *Proceedings of the 8th International ISCRAM Conference*, (May), 1–10.

Tapia, A. H., Moore, K. a, & Johnson, N. (2013). Beyond the Trustworthy Tweet: A Deeper Understanding of Microblogged Data Use by Disaster Response and Humanitarian Relief Organizations. *Proceedings of the 10th International ISCRAM Conference*, (May), 770–779.

Vieweg, S., Hughes, A. L., Starbird, K., & Palen, L. (2010). Microblogging during two natural hazards events. *Proceedings of the 28th International Conference on Human Factors in Computing Systems - CHI '10*, 1079. http://doi.org/10.1145/1753326.1753486