Disentangling the Lexicons of Disaster Response in Twitter

Nathan O. Hodas
Pacific Northwest National Laboratory
Richland, WA 99352
nhodas@pnnl.gov

Greg Ver Steeg
Information Sciences Institute
University of Southern California
Marina Del Rey, CA 90292
gregv@isi.edu

Joshua Harrison
Pacific Northwest National Laboratory
Richland, WA 99352
Josh.Harrison@pnnl.gov

Satish Chikkagoudar
Pacific Northwest National Laboratory
Richland, WA 99352
Satish.Chikkagoudar@pnnl.gov

Eric Bell
Pacific Northwest National Laboratory
Richland, WA 99352
Eric.Bell@pnnl.gov

Courtney D. Corley
Pacific Northwest National Laboratory
Richland, WA 99352
court@pnnl.gov

ABSTRACT

People around the world use social media platforms such as Twitter to express their opinion and share activities about various aspects of daily life. In the same way social media changes communication in daily life, it also is transforming the way individuals communicate during disasters and emergencies. Because emergency officials have come to rely on social media to communicate alerts and updates, they must learn how users communicate disaster related content on social media. We used a novel information-theoretic unsupervised learning tool, CorEx, to extract and characterize highly relevant content used by the public on Twitter during known emergencies, such as fires, explosions, and hurricanes. Using the resulting analysis, authorities may be able to score social media content and prioritize their attention toward those messages most likely to be related to the disaster.

Categories and Subject Descriptors

J.4 [**Social and Behavioral Sciences**]: Sociology; H.3.3 [**Information Search and Retrieval**]: Information Filtering; K.4.1 [**Computers and Society**]: Public Policy Issues—*Human Safety*

Keywords

Mutual Information, Clustering, Disaster Response, Twitter, Lexicon, Risk

1. INTRODUCTION

The ability to clearly communicate between the public and those responsible for assessing, minimizing, and regulating risks is critical for successful resolution of a public emergency. Strong personal safety concerns cause the public to develop symptoms of emotional and behavioral distress that adversely affect the perception of risk during a crisis through the evocation of strong emotions such as fear, anxiety, distrust, anger, outrage, helplessness, and frustration[4]. Understanding the dynamics of risk perception during a crisis is crucial for successful emergency response because ultimately people act on the basis of what they believe to be true. Perceived risk is known to have a stronger impact on disaster recovery and preparedness than actual risk as communicated by emergency public information officers. For example, a recent study on risk communication shows households in America are more strongly motivated to prepare for terrorism and other hazards by observed preparations taken by others than they are by information received from preparedness information providers[8]. Although public officials depend on precision and clarity, properly tailoring their communications using derived lexicons is only now being investigated [11]. Herein we present an automated content analysis method to analyze social media in order to provide tools to study language used during emergencies. These tools could potentially be used to facilitate effective communication of risk and how to mitigate risk in crises.

Over the past few years, short messages have been used in various forms for disaster-related risk communication. The multistage developmental process for risk communication that is discussed by Fischhoff[5] is still relevant to short messages and the new generation of communication media. One notable instance was the use of social media to communicate warning messages during the 2008 terror attack in Mumbai, India. The consensus of review articles is that social media is not just a new means to carry out an old risk communication strategy[2, 3, 6, 10, 9]. However, language use varies widely, depending on proximity to the disaster, both geographic and experiential[7]. Here, we present a new strategy for analyzing tweets during an emergency to understand how language is being used and which words best help communicate latent factors. By analyzing the mutual information between words within a tweet and the extracted latent variables, we show that each type of disaster has a characteristic lexicon which is often surprisingly different from how words are used in typical tweets outside of emergencies.

WWW 2015 Companion, May 18–22, 2015, Florence, Italy.
ACM 978-1-4503-3473-0/15/05
http://dx.doi.org/10.1145/2740908.2741728 .

2. METHODS

2.1 Data

Social media is understood as an information propagation tool for reporting on and responding to natural disasters. Emergency management services use social media to issue alerts and warnings, look for reports of emergencies, and understand public response to emergencies. Social media is used to share information leading up to, during, and after the disasters[1]. For the purposes of this study, tweets around a set of well-known and documented disasters were gathered for examination. We sampled the public 1% streaming Twitter API for control tweets.

Collection of data was limited to geotagged tweets (tweets containing latitude and longitude coordinates) within the specified timeframe to verify user proximity in both temporal and physical space to the disaster event. A primary goal of the study is characterizing the language around disaster events, as used by people likely to be impacted or otherwise directly involved. While the volume of geotagged tweets is low, we were still able to acquire workable sample volumes for each disaster.

To collect pertinent disaster-related tweets, we used 203 U.S. Federal Emergency Management Agency (FEMA) declared disasters in the United States from 2012 and 2013 (`http://www.fema.gov/disasters`). Historical Twitter data was obtained from Gnip, a provider of the Twitter firehose, using their historical data request API. Each query was composed of curated keyword lists, primarily named entities related to a particular disaster and informal language describing the nature of the event. Queries were further filtered both by geo-tagged location and date range. A date range for the historical query was selected by taking a range of ±5 days from the event date itself, except for the (non-weather-related) Alamo, California, gas leak, where a date range of +5 days and -1 day around the event date was used. Geographical filters for the query were established using the area of impact of the emergency declaration, such as a single 25-mile radius around a defined point, or entire regions were selected when more than a single point of impact exists. For example, Southern California flooding and wildfires searched within California, Hurricane Sandy covered multiple states, and the Alamo gas leak was a single point centered on Alamo, California. Each tweet was also labeled by the type of disaster from which it was querired (e.g., tornado, hurricane, fire, flood).

Upon acquisition of this data, it was ingested into Elasticsearch, a Lucene based search engine architecture. A sample of the queries are listed in Table 1, and the complete list is available on request. It is important to note that the nature of collecting only geo-tagged tweets means that we did not collect the subsequent retweets, but many people manually retweeted non-geocoded tweets and embedded their own geocode.

Table 1: Selected Query Definitions

Disaster Name	Date Range	Query Terms
Alamo, California, Gas Leak	July 23 to 29 2013	leak, gas, evacuation, pg&e, pge, pg+e, alamo, danville, shelter
El Reno, Oklahoma, Tornado	May 25 to June 6 2013	tornado, wind, shelter, evacuation, storm, chaser, funnel, EF, hail, moore, noise, warning, samaras, el reno, rotating, debris, disaster, twister, siren
Hurricane Sandy	October 26 to November 10 2012	storm, hurricane, sandy, frankenstorm, flood, danger
Yarnell Hill Fire, Arizona	June 29 to July 8 2013	fire, burn, Yarnell, hotshot

2.2 Clustering

The relevant geocoded tweets for each disaster were obtained as described in Section 2.1. We designed a lexicon to capture many of the potential ways people communicate about disasters. This includes categories of words including: units, interpersonal relationships, references to government and media, emotions, public directives, as well as descriptions of the disaster. We consulted with subject matter experts to grow the key term and phrase list, resulting in 292 regular expressions. We developed the regular expressions to capture lemmatized keywords with an intent to avoid ambiguity. For example, the words smolder and smoldering were represented as smold. Variants of "fire" can be expressed as `fire(?!(work|fi))`, to avoid tagging fireworks or firefighters as variant of fire, which were searched for separately. For the exact regular expressions, please contact the authors. The resulting crisis related lexicon does not capture how risk is communicated by the public. In order to obtain more relevant semantic clustering, we employed Correlation Explanation (CorEx)[12]. CorEx searches for latent variables that explain correlation between the usages of different terms.

A reduced subset of 50,000 tweets was randomly sampled for each disaster type for the final analysis. The sample size was chosen for computation purposes and to avoid over representation of any one particular disaster. Each tweet was converted into a vector, X, where X_i is the presence (or absence) of regular expression i in the tweet. For each type of disaster, we used CorEx to generate a tree of latent variables where each variable is constructed to maximally explain the correlations in its children. That is, we simultaneously search over latent variables, $Y_j, j = 1, \ldots, m$, and clusters of words G_j so that $\sum_j TC(X_{G_j}; Y_j)$ is maximized. TC represents the amount of correlation in a group of variables, X_{G_j}, that is explained by Y_j, and is specified by $TC(X_{G_j}; Y_j) = \sum_{i \in G_j} MI(X_i; Y_j) - MI(X_{G_j}; Y_j)$, where $MI(X; Y)$ is the mutual information between X and Y. For a group of uncorrelated X_i's, for instance, this expression would give zero, while it would be maximized if all the variables were identical copies. To construct a tree, we take the $Y^{(n-1)}$'s learned on one level and apply CorEx again to learn a representation, $Y^{(n)}$ [12].

The CorEx algorithm provides a tree of latent variables explaining correlation in the data, but it does not provide explicit labels for the latent variables. To label the latent variable nodes of the tree, we propagated up the tree the

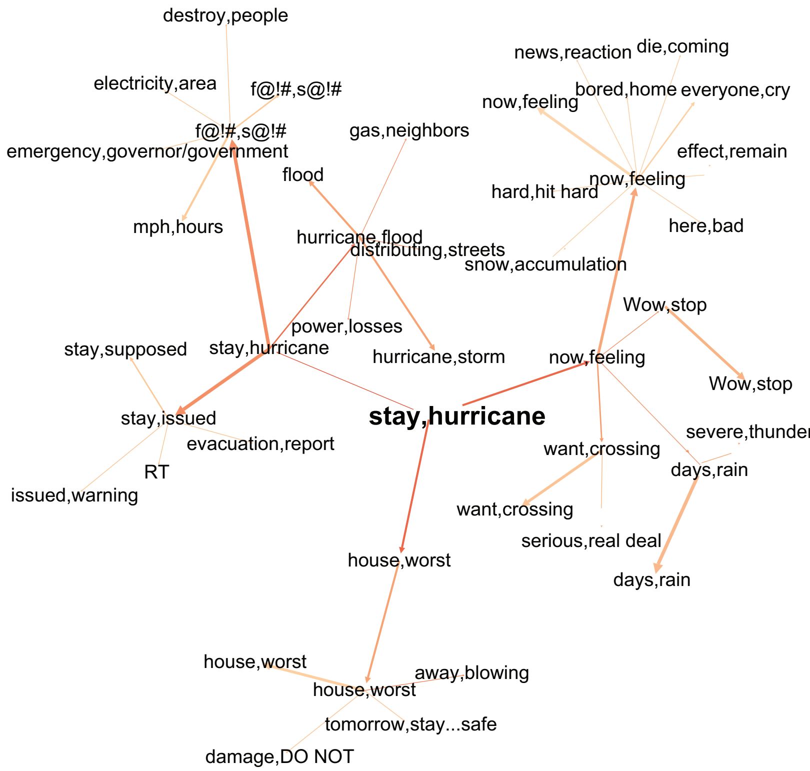

Figure 1: We assign labels to the latent variable tree by propagating highly informative words up the tree. Thus, some labels may appear more than once. The thickness of an arrow represents the mutual information between a node and its parent. "Stay, Hurricane" is the root node.

corpus entries with the highest weight according to the mutual information between the label and the latent variable to be labeled. Results for the Hurricane disaster are shown in Figure 1.

3. RESULTS AND DISCUSSION

The CorEx technique succeeds at identifying both disaster-specific themes as well as how word usage changes during emergencies. For the control tweets, the CorEx analysis produces very different clustering, implying different latent variables underlie the use of risk-related terms outside of emergence situations. For example, during fire events, the words "fire" and "firework" are closely associated, but this is not true outside of fire events. Similarly, outside of disaster events, the CorEx generates labels such as "hot/cold", "hours/minutes", "rain/sunny", and "char/destroy,". During disasters, CorEx tends to identify predictive n-grams, such as "power/loss", "evacuation/county", and "issued/until." Thus, we conclude that the resulting lexical analysis is specific to how users respond to disasters and not simply generic relations resulting from artifacts of corpus contruction.

Our results show that tweets tend to focus on announcing the emergency, advising others, or describing the damage. Other obvious clusters include expressions of anxiety, frustration, and expletives. The CorEx results enable us to show which words in the corpus are most predictive of the latent feature, and thus may communicate the intent of the latent feature most clearly.

For each type of disaster, we have produced a list of words that best communicate the latent variables. That is, for each feature X_i, there is a mutual information with each latent variable Y_j. High mutual information implies high explanatory power, and for each type of disasters we present a list of words that have the most explanatory power in each scenario. Because Y_j represent the latent variable which explains the co-occurence of words below it in the tree, having high $MI(X_i;Y_j)$ means that feature X_i also helps explain why other words are used in combination with it in that disaster scenario. We present a brief list of the most and least informative words for each type of disaster in Table 2.

The CorEx technique enables us to analyze each tweet to determine how it is using words from the risk corpus. We score each tweet by summing its total correlation with respect to each latent variable for that disaster type. Low-scoring tweets utilize combinations of words in unexpected ways, often poorly communicating intent. These may be characterized as surprising, yet uninformative tweets. Conversely, words that are highly predictive of latent variables, and therefore other words, are found in tweets that can be considered unsurprising and informative. While the content of the tweet may refer to a surprising event, (tornadoes, explosions, etc.), the reader should be able to interpret how all the terms communicate the inferred latent variable. For example, disaster alerts and discussion of disaster repercussions and response tend to to highly informative and unsurprising, while passing comments and references tend to be uninformative and surprising from our current total-correlation perspective.

Table 2: Informative Words and Phrases. We exclude trivial disaster labels, e.g. "hurricane". (01234 indicates numbers)

Disaster Type	Most Informative	Least Informative
Control	love, 01234, want, now, day	mangled, fire dept, funnel, national guard, hoax
Hurricane	01234, house, power, flood, listen	false alarm, mangled, impassible, wind, fire fighters
Fire	01234, burn, smoke, police, firework	doozy, struggle, ice, blah, arson
Explosion	01234, scared, house, expletive, bomb	mph, temperature, anxious, remain inside, hoax
Tornado	01234, until, issued, mph, severe	accumulation, wind, impassible, go figure, remain inside

4. CONCLUSIONS

The CorEx analysis provides us a tool to extract useful ways to communicate with the public using a risk corpus, using language already being used on social networks today. By extracting latent variables, we reveal how words are used together to communicate coherent messages, and we quantify the latent content of each tweet. That is, each latent variable we identify helps to explain the mutual occurrence of words from the corpus in each tweet. In addition to being a useful clustering tool, the CorEx analysis provides us with a dimensionality reduction by mapping each tweet into a vector of probabilities for representing each of the latent variables. Although we applied the same risk corpus to analyze many types of disasters, words and phrases convey information very specific to each type of emergency event. Thus, CorEx and similar analyses can be used to characterize tweet composition, which may help for constructing emergency announcements. As a potential use case, CorEx can be trained on a selected set of disaster data. The resulting model can be used to evaluate new tweets, providing a filtered and ranked social media stream for use by emergency management personnel to react to rapidly changing and emerging conditions on the ground.

5. ACKNOWLEDGMENTS

The research at Pacific Northwest National Laboratory was supported through a contract with the U.S. Department of Homeland Security's Science and Technology Directorate First Responders Group. Pacific Northwest National Laboratory is operated for the U.S. Department of Energy by Battelle under contract DE-AC05-76RLO 1830.

6. REFERENCES

[1] J. P. Bagrow, D. Wang, and A.-L. Barabasi. Collective response of human populations to large-scale emergencies. *PloS one*, 6(3):e17680, 2011.

[2] Booz-Allen-Hamilton. Crisis communications social media round table special report. Report, 2009.

[3] W. J. Burns. Risk perception: a review. Report, USC, 2007.

[4] V. Covello and P. M. Sandman. Risk communication: evolution and revolution. *Solutions to an Environment in Peril*, pages 164–178, 2001.

[5] B. Fischhoff. Risk perception and communication unplugged: Twenty years of process1. *Risk analysis*, 15(2):137–145, 1995.

[6] R. E. Kasperson. Six propositions on public participation and their relevance for risk communication. *Risk analysis*, 6(3):275–281, 1986.

[7] Y. R. Lin. The ripples of fear, comfort and community identity during the boston bombings. In *iConference 2014 Proceedings*, pages 708–720, 2014.

[8] P. M. Sandman. Hazard versus outrage in the public perception of risk. *Effective risk communication*, pages 45–49, 1989.

[9] P. Slovic. *The feeling of risk: New perspectives on risk perception*. Routledge, 2010.

[10] P. Slovic, B. Fischhoff, and S. Lichtenstein. Why study risk perception? *Risk analysis*, 2(2):83–93, 1982.

[11] I. Temnikova, andrea Varga, and D. Biyikli. Building a crisis management term resource for social media: The case of floods and protests. In N. C. C. Chair), K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, and S. Piperidis, editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland, may 2014. European Language Resources Association (ELRA).

[12] G. Ver Steeg and A. Galstyan. Discovering structure in high-dimensional data through correlation explanation. *http://arxiv.org/abs/1406.1222*.