

Assisting Coordination during Crisis: A Domain Ontology based Approach to Infer Resource Needs from Tweets

ABSTRACT

Ubiquitous social media during crises provides citizen reports on the situation, needs and supplies. Previous research extracts resource needs directly from the text (e.g. “Power cut to Coney Island and Brighton beach” indicates a power need). This approach assumes that citizens derive and write about specific needs from their observations, properly specified for the emergency response system, an assumption that is not consistent with general conversational behavior. In our study, Twitter¹ messages (tweets) from Hurricane Sandy in 2012 clearly indicate power blackouts, but not their probable implications (e.g. loss of power to hospital life support systems). We use a domain model to capture such interdependencies between resources and needs. Using semantic web technology, we represent these dependencies in an ontology that specifies the functional association between resources. Accurate interpretation of resource need/supply also depends on the location of a message. We show how inference based on a domain model combined with location detection and interpretation in the social data can enhance situational awareness, e.g., predicting a medical emergency before it is reported as critical.

Keywords

Crisis Response, Crisis Response Co-ordination, Domain Model, Semantic Inference, Social Media for Emergency Management (SMEM).

1. INTRODUCTION

The use of social media, especially Twitter, during crises (such as Hurricane Sandy in the U.S. in 2012 and Typhoon Haiyan in the Philippines in 2013) promises to revolutionize situational awareness in the emergency response community. Citizens serve as sensors [17], sharing information about the changing state of affairs and needs. Prior research applies information filtering and extraction techniques directly to the reported observation content to identify those that may assist the response activities [1,2,3,7]. However, explicit content is often incomplete due to unstated common ground among participants [11], and other constraints such as the 140 character limit for tweets. According to established conversational maxims [18], explicit content

appropriately provides only new information. For example, while many tweets identify power blackouts, citizens do not identify the implications of power blackouts for each hospital. Newspapers however, document the resulting hospital compromise during Hurricane Sandy: <http://j.mp/2Hospitals>. We focus on two related problems: 1) Recognizing resource compromise early (here to the medical system), and 2) Identifying geo-location of resources in messages.

Limitation of explicit tweet content is recognized and vocabulary tags are recommended in [4] for message creation but this can burden citizens who are likely unfamiliar with the professional work of emergency practice. Alternatively, the identification of relevant messages for human review can be improved substantially with computational approaches to filter and augment the Twitter data. Role of ontology is acknowledged in [4] but they do not mine the implications of resource inter-dependencies. We consider semantic knowledge bases, DBpedia [12] and crisis ontology (discussed in section 2), to include causal links (e.g. between power and medical resources) and infer the location of resource.

Location information can also assist in the screening of social media traffic, but like the rest of tweet content, requires interpretation with background knowledge. Geo-referenced data to anchor tweets to a map is used in [16], thereby providing the implicit context necessary for making the content actionable. However, their solution faces a severe limitation, as most tweets do not contain geo-location metadata. For example, out of 4 million tweets we collected during first week of Hurricane Sandy, only 20% of the tweets from the Hurricane Sandy data had geo-location information from tweet metadata (location from source device sensor or author profile). Consistent with [19] we found that location information may appear in text, e.g. “Manassas opens pet friendly shelter: City of Manassas VA will open a shelter for Hurricane Sandy victims”. However, the indication of location requires interpretation, which becomes particularly apparent for recipients unfamiliar with the area, e.g., references to street names or local stores, such as “The Home Depot in Manahawkin”. While [7] aims to locate the twitterer, we aim to locate resources. Geo-locations from tweet metadata are used in [1,5] but they do not consider location information from the tweet text. Location in text using machine learning techniques are discussed in [9], but without our emphasis on interpretation. We also focus on the text to identify the location but we exploit DBpedia [12] to identify the named locations in order to identify landmarks, hospitals, parks, etc.

Our main objective is to exploit and augment the natural patterns in social media crisis response, in order to infer compromise earlier than it is stated, and render the corpus more accessible to

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

¹Popular microblogging service: www.twitter.com

and informative for the emergency response community. The main contributions of this study are:

1. Demonstration of how can domain knowledge models of resources, during crisis response, capture the inter-dependencies between resources (section 2.1)
2. Identification and annotation of tweet metadata and text-based location information to assist domain-model based inference in crisis response (section 2.2)
3. The potential to identify resource scarcity or seeking behavior due to possible correlation with the inclusion of text-based location content. (section 3.2)

Next, we present our approach to model interdependent resources and location identification from tweet text.

2. APPROACH

We use the Twitter Streaming API to collect tweets and further process them for resource detection and location detection. We annotate the tweet combining standard Natural Language Processing (NLP) techniques with a domain model and semantic web content.

2.1 Resource Detection

A resource detection module identifies resources (e.g. food, water, power), using a focused top-down approach. Alternatively, using a bottom-up approach, the data drive the extraction of an ontology using machine learning techniques (Imran et al., 2013). Using a broad top-down approach, domain experts provide a hierarchical classification scheme that includes class relationships e.g., the Management Of A Crisis (MOAC) vocabulary [14] and the Humanitarian Exchange Language (HXL) [13]. Though less vulnerable to bias in the data relative to the bottom-up approach, the top-down approach faces the daunting challenge of modeling the entire domain. Our crisis ontology is more focused, containing 43 concepts and 45 relationships². We spot the entities in tweet text for the entity set of concepts in the domain model. Tweets are annotated with the crisis ontology that includes the relationship between power, medical, and food/water resources.

2.2 Location Detection

Consider the resource identification in the tweet “*American Red Cross & City of Nashua Have Opened Shelter In Response to Hurricane Sandy*”. Understanding “Nashua” is critical to the utility of this tweet. We exploit two methods for the specification of actionable location: existing metadata with Tweet and location information embedded in tweet text. The Twitter API provides device sensor location (in form of geo-coordinates) as well as the location of an author profile in the metadata. We query author profile location with Google Maps API [20] to get co-ordinates of author location consistent with the techniques followed in [17]. However, the source identified in tweet metadata may not be in the affected crisis region. Furthermore, the geo coordinates in tweet metadata still require an interpretation in terms of place name in order to render the tweet actionable.

Our aim is to locate resources, a problem that is more directly related to assisting responders than identifying the location of a person, as in [7]. Responders need to know whether a named location is country, state, county, or a specific street in a city.

²<http://twitris.knoesis.org/images/datasets-and-models/Disaster-domain-model-v2.rdfxml.owl>

Moreover, detailed location information must be situated within a city, county, and state. While this is “common knowledge” for the local responder, non-local personnel often assist in large-scale crises.

First we use the Stanford NER [15] to detect location content from tweet text. Next we determine, via query, whether the identified location is part of the locations list in the DBpedia ontology. DBpedia includes both global locations as well as information for well-known hospitals, parks, buildings, etc., that people often mention while tweeting and that may not otherwise be covered by a geo-location ontology. Using this two-fold approach we ensure that the text identified by the Stanford NER is a true location. For example, the following tweet is annotated with two locations:

Evacuation in progress at Bellevue Hospital in New York City has begun. Their power generator has 1 hour left. #hurricane #sandy #nyc #nypd

http://dbpedia.org/resource/New_York_City,
http://dbpedia.org/resource/Bellevue_Hospital_Center

We traverse the DBpedia ontology to situate Bellevue Hospital. We annotate both text location and tweet metadata location. We use the term ‘affected region’ in the following discussion to consider the locations of crisis affected regional states, in this case, New York, Virginia, Delaware, New Jersey, Pennsylvania, Massachusetts, Connecticut, North Carolina, South Carolina.

2.3 Evaluation

Evaluation requires comparison against a standard, which is a challenge of this new area of crisis computing. To emulate this standard, we compared our text location methods to the ratings of three independent research assistants on a subset of 1000 tweets that contained either text or tweet metadata location indicators, or both, using the following prompts:

- *is_text_location_relevant* — Given the embedded text location identified by our technique, is that semantically relevant for the message context of the tweet (0 or 1).
- *is_seeker* — Whether a tweet mentions scarcity of resource (0 or 1).
- *is_supplier* — Whether a tweet mentions availability of a resource (0 or 1).

Table 1: Tweets related to specific resource and location types

Type	Total	Metadata locations	Metadata locations	Text w/o Metadata	Text and Metadata	Metadata w/o text
Power	103102	14969	24361	10974	3995	20366
Medical	16002	3243	6015	2057	1186	4829
Food/water	38952	5046	9152	3574	1472	7680
Power & Medical	948	231	377	134	97	280
Power & Food	2908	382	839	260	122	717
Power & Food/water	44	39	30	13	26	17

3. RESULTS AND DISCUSSION

Here we review some of the key findings for our full corpus of 4 million hurricane Sandy related tweets crawled during the period of 27th Oct to 4th Nov, 2012, addressing how the population naturally uses Twitter in the articulation of resources and location, and how we might exploit these patterns of usage. Frequency counts of respective resources and identified location types appear in Table 1.

3.1 Inferring compromise from tweet content

The frequency of power related tweets is much greater than the frequency of medical tweets, or even food and water related tweets. Moreover, people rarely (less than 1% of the time) make the explicit link between compromise to the power grid and the consequences to medical, food, and water resources.

From a sample of 1.4 million tweets collected between the 29th and 30th of October 2012, we found 50 tweets that reported a blackout well before the 21 tweets that reported hospital evacuation. Table 2 illustrates the trend from general power-related comments to specific comments about the compromised hospital. Tweets confirm the power outage at midnight, yet the report of hospital compromise appears three hours later, at 3:00 a.m. the next day. Domain knowledge permits the inference about hospital compromise as soon as the power outage is confirmed, either due to direct consequences to the hospital power or the consequence of power outage to the surrounding community [21].

Table 2: Timeline of medical/power emergency tweets about event (<http://j.mp/2Hospitals>)

Time (2012)	Message text	Text Location identified
Oct. 29, 21:30:02	Lots of wind and some rain but still running and no power outage in Clinton Hill, Brooklyn. #Sandy #Hurricane	http://dbpedia.org/resource/Brooklyn
Oct. 29, 23:56:57	Power cut to coney island and Brighton beach #HurricaneSandy #NYC	http://dbpedia.org/resource/Coney_Island
Oct. 30, 01:30:36	Power may be cut off soon in south bklyn, Coney, Gravesend Sheedshed Bay etc #Sandy #Frankenstorm	http://dbpedia.org/resource/Coney_Island
Oct. 30, 03:15:08	@911BUFF: BREAKING CONEY ISLAND HOSPITAL ON FIRE. NYU HOSP. EVACUATED, BELLEVUE HOSPITAL ALSO LOSING BACKUP POWER #SANDY #NYC #frankenstorm	http://dbpedia.org/resource/Bellevue_Hospital_Center
Oct. 30, 20:20:42	SANDY: Bellevue Hospital is on backup power, trying to evacuate as much as possible. 2 young boys missing from SI since beginning of Hurricane	http://dbpedia.org/resource/Bellevue_Hospital_Center

3.2 Inferring location from tweet content

As shown in the Table 1, the ability to infer location from text increases location information over tweet metadata information by approximately 50%. To study this location dynamics we considered the sample of 1000 tweets having either metadata or text location from the above categories (food/water, medical,

Table 3: Comparison of affected region’s text and tweet metadata locations

Total tweets = 1000	Text indicator in affected region (T) = 340	Missing or misleading text indicator (NT) = 660
Metadata location in Affected Region (M) = 248	27	221
Missing or misleading Metadata location information (NM)= 752	313	439

power, and shelter).

Interpretation of rows and columns in the table 3 is as follows: M = AM, NM = WM + non-M, T = AT, and NT = WT + non-T; Where, each of the following mentions total number of tweets for- AM = Tweet metadata location falls in affected region,

WM = Tweet metadata location doesn’t fall in affected region,
 non-M = Tweet metadata locations were not available
 T = Text location falls in affected region,
 WT = Text location doesn’t fall in affected region
 non-T = Text locations were not identified.

We obtained 88% accuracy (agreement) for identifying location from text. That is, in 88 cases out of 100 human labelers considered the text location semantically relevant. The following example illustrates the potential for error: “Hurricane Sandy projected to slam into New England coast... binders of woman seek shelter”. This tweet was identified as having the location New_England, Australia. This happens because of location name ambiguity (New England exists in USA and Australia), which is not addressed in this paper.

Next we separated the 1000 tweets into two subsets: those inside and those outside the area of interest. We partitioned these subsets according to their consistency with the metadata, allowing us to examine the correlation between tweet metadata and text location. Table 3 illustrates the conditions for communicating location in text.

Tweets are more likely to contain location information when tweet metadata information is missing or indicates origination from outside the affected area.

Table 4: Text and tweet metadata location comparison

Location source	Total tweets	Total tweets with location in affected region	Total tweets with location not in affected region
Text	517	340 (66%)	177 (34%)
Metadata	313	238 (43%)	323 (57%)

Finally, we examined only those tweets that included location information (either text or tweet metadata or both) (see Table 4). This includes T along with a portion of the subset NT, removing any tweets that had no location identifiers. Table 4 shows that in almost 66% of cases, when text location was mentioned in a tweet, it was mentioning a location from the affected crisis region. However only 43% of time, were tweets with tweet metadata location consistent with the crisis affected region. This suggests that mention of text location is more reliable than the tweet metadata location to locate a resource need/supply in the affected region. This observation is somewhat unexpected, and emphasizes the need for text interpretation.

We note some additional patterns below in the specification of location information in tweets. We queried for tweets annotated with resource, specific text locations and stores like home depot, Walmart etc. We found tweets where people are offering a resource or reporting resource scarcity and location information suggesting the area of interest. Below we have considered an electric generator as the resource to be located.

Independent of the source of location information (tweet text, or device sensor or profile metadata) understanding the tweet still requires the background knowledge of a geo-ontology such as DBpedia. It implies that tweet metadata location will give us the geo-coordinates but we need to map the coordinates to an ontology such that city/state or country can be inferred. The

absence of useful specificity persists with text location. Consider the following Tweet and detected location:

I have a brand new 5500 watt generator for sale, local pickup in Monmouth county #sandy #ocean #generator #ocean #frankenstorm #nyc,

http://dbpedia.org/resource/Monmouth_County,_New_Jersey

The implication is that the generator is located at site of the tweet origins in the presence of nearby power outage, although the precise location still requires further exchange to establish a pickup location. In the following tweet, we note the persisting need to interpret “Home Depot in Manahawkin” in order to recover location information.

Picking up a gas can at Home Depot in Manahawkin and some guy told me there was a fist fight over a generator there yesterday! #hurricane

http://dbpedia.org/resource/Manahawkin,_New_Jersey

The above tweet is also interesting because it exemplifies a mismatch between information in the tweet metadata (i.e., user profile) and the episode in question. Users do not stay in one place. This episode did not likely occur in the location associated with the user’s profile, and we don’t know where he was when he wrote the tweet, or how this timing related to the timing of the episode. This elevates the importance of relying on text content for location inference.

The next example suggests a relationship between the availability of location information in the text and the articulation of resources and needs.

@terrence_j_r Turkey Hill on Harford St in #MilfordPA has a generator so they can pump gas and sell various sundries. #hurricane,
[http://dbpedia.org/resource/Turkey_Hill_%28Linthicum_Heights,_Maryland%29,](http://dbpedia.org/resource/Turkey_Hill_%28Linthicum_Heights,_Maryland%29)

This tweet suggests a potential relationship between very specific tweet location content and resource seekers or suppliers.

Table 5: Affected region text location with exploratory analysis of seeker/supplier behavior.

Total tweets with locations in text = 517	With affected region text location (340)	Without affected region text location (177)
With seeker supplier and with location = 89	64 (72%)	25 (28%)
Without seeker supplier but having location = 428	276 (64%)	152 (36%)

To explore the relationship between the articulation of resources and needs with location, Table 5 further analyzes the 517 tweets with text location as tabulated in Table 4. We required a majority (2) agreement among the three judges in order to classify a tweet as either a seeker or supplier. Though not statistically significant, the pattern suggests a promising relationship between the mention of seeker/supplier content and the inclusion of text that specifies the affected region with a larger sample size. This in turn suggests the potential of a simple location-based screening tool for the

identification of seeker/supplier content subject to subsequent computational analysis.

Conclusion and Future Work

We present an approach using a domain knowledge model for understanding contextually interdependent resource needs reported via social media during crisis response. We do not intend to modify user behavior, but rather exploit the natural communication patterns to situate the tweet in a broader context to facilitate response. Unlike prior approaches that focused on explicit content analysis for information filtering and extraction, we proposed the idea of implicit content analysis using relationships defined by ontological and domain knowledge, like power needs leading to medical needs. This may support the early identification of resource need, before it is articulated. Second, we relied on domain knowledge to situate the location information, whether or not the source was tweet metadata or text. This supports responder action, and is particularly critical for responders who are not local to the area affected by crisis. Third, we suggested deference to text indicators of location in the case of discrepancy with metadata, due to the dynamic nature of reporting. Finally, we noted a possible correlation between the presence of location information in the text and behavior of either seeking or supplying of resources. This may provide a simple screening tool for the identification of resources in social media. Future studies can leverage these simple model-based relationships to help manage resource demands in real-time. Our future work includes the following directions in order to:

- Exploit seeker-supplier behaviors in the tweets of the affected crisis region
- Exploit other relationships of interdependent resources
- Extend text location identification and verification methods to increase accuracy and eliminate the location ambiguity issue noted above.

4. REFERENCES

- [1] Varga, I.; Sano, M.; Torisawa, K.; Hashimoto, C.; Ohtake, K.; Kawai, T.; Jong-Hoon, O.; & De Saeger, S. (2013). Aid is out there: Looking for help from tweets during a large scale disaster. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics Vol. 1, pp. 1619-1629.
- [2] Purohit, H.; Castillo, C.; Diaz, F.; Sheth, A.; & Meier, P. (2013). Emergency-relief coordination on social media: Automatically matching resource requests and offers. First Monday, 19(1). doi:10.5210/fm.v19i1.4848
- [3] Imran, M.; Elbassuoni, S.; Castillo, C.; Diaz, F.; & Meier, P. 2013. Extracting information nuggets from disaster-related messages in social media,” In: T. Comes, F. Fiedrich, S. Fortier, J. Geldermann and T. Müller (editors). ISCRAM ’13: Proceedings of the 10th International ISCRAM Conference.
- [4] Starbird, K.; & Stamberger, J. 2010. Tweak the tweet: Leveraging microblogging proliferation with a prescriptive syntax to support citizen reporting. ISCRAM ’10: Proceedings of the 7th International ISCRAM Conference.
- [5] Cameron, M.A.; Power, R.; Robinson, B.; & Yin, J. 2012. Emergency situation awareness from twitter for crisis

- management. WWW '12 Companion: Proceedings of the 21st International Conference Companion on World Wide Web, pp. 695–698.
- [6] Vieweg, S.; Hughes, A.L.; Starbird, K.; & Palen, L. 2010. Microblogging during two natural hazards events: What Twitter may contribute to situational awareness. CHI '10: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1,079–1,088.
- [7] Starbird, K., Muzny, G., & Palen, L. (2012). Learning from the Crowd: Collaborative Filtering Techniques for Identifying On-the-Ground Twitterers during Mass Disruptions. In Proceedings of the Conference on Information Systems for Crisis Response and Management 2012. Leon Rothkrantz, Jozef Ristvej, & Zeno Franco (editors). Vancouver: Simon Fraser University.
- [8] Reuter, C.; Heger, O.; & Pipek, V. Combining Real and Virtual Volunteers through Social Media. Retrieved from: <http://www.iscramlive.org/ISCRAM2013/files/122.pdf>
- [9] MacEachren, A.M.; Robinson, A.C.; Jaiswal, A.; Pezanowski, S.; Savelyev, A.; Blanford, J.; & Mitra, P. 2011. Geo-twitter analytics: Applications in crisis management. In Proceedings, 25th International Cartographic Conference, Paris, France.
- [10] Jihan, S.H.; & Segev, A. Context Ontology for Humanitarian Assistance in Crisis Response. <http://www.iscramlive.org/ISCRAM2013/files/142.pdf>
- [11] Clark, Herbert H.; Brennan, Susan E. (1991), Resnick, L. B.; Levine, J. M., eds., Perspectives on socially shared cognition, American Psychological Association, ISBN 1-55798-376-3
- [12] Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., & Ives, Z. (2007). Dbpedia: A nucleus for a web of open data. The Semantic Web (pp. 722-735). Springer Berlin Heidelberg.
- [13] Keßler, C., Hendrix, C. J., & Limbu, M. (2013). Humanitarian eXchange Language (HXL) Situation and Response Standard. January 2013. <http://hxl.humanitarianresponse.info/ns/index.html>.
- [14] Limbu, M. Management Of A Crisis (MOAC). January 2012. <http://observedchange.com/moac/ns/>.
- [15] Marie-Catherine de Marneffe, Bill MacCartney and Christopher D. Manning. Generating Typed Dependency Parses from Phrase Structure Parses, LREC-2006
- [16] D. Pohl, A. Bouchachia, H. Hellwagner. Supporting Crisis Management via Detection of Sub-Events in Social Networks. International Journal of Information Systems for Crisis Response and Management. In Press (2013)
- [17] Nagarajan, M., Gomadam, K., Sheth, A. P., Ranabahu, A., Mutharaju, R., & Jadhav, A. (2009). Spatio-temporal-thematic analysis of citizen-sensor data - Challenges and experiences. Tenth International Conference on Web Information Systems Engineering, October 5-7, 2009, pp. 539-553.
- [18] Grice, H. P. (1989). Studies in the Way of Words. Harvard University Press.
- [19] Cheng, Z., Caverlee, J., & Lee, K. (2010). You are where you tweet: A Content-based approach to geo-locating Twitter users. In *Proceedings of the 19th ACM international conference on Information and knowledge management*. ACM, 2010.
- [20] Google Maps API, <https://developers.google.com/maps/documentation/>
- [21] Flach, J.M.; Steele-Johnson, D.; Shalin, V.L.; and Hamilton, G.C. 2012. Coordination and control in emergency response. To appear in *Handbook of Emergency Response*.