AAAI Technical Report WS-13-04
When the City Meets the Citizen 2

The Car that Hit The Burning House: Understanding Small Scale Incident Related Information in Microblogs

Axel Schulz[1,2] and **Petar Ristoski**[1]
[1]SAP Research, Darmstadt, Germany
[2]Telecooperation Lab, Technische Universität Darmstadt, Germany
{axel.schulz, petar.ristoski}@sap.com

Abstract

Microblogs are increasingly gaining attention as an important information source in emergency management. In this case, state-of-the-art has shown that many valuable situational information is shared by citizens and official sources. However, current approaches focus on information shared during large scale incidents, with high amount of publicly available information.

In contrast, in this paper, we conduct two studies on every day small scale incidents. First, we propose the first machine learning algorithm to detect three different types of small scale incidents with a precision of 82.2% and 82% recall. Second, we manually classify users contributing situational information about small scale incidents and show that a variety of individual users publish incident related information. Furthermore, we show that those users are reporting faster than official sources.

Introduction

Social media platforms are widely used by citizens for sharing information about incidents. Ushahidi, a social platform used for crowd-based filtering of information (Okolloh 2008), was heavily used during the Haitian earthquake for labeling crisis related information. Also useful situational information was shared on Twitter during incidents like the Oklahoma grass fires and the Red River floods in April 2009 (Vieweg et al. 2010) or the terrorist attacks on Mumbai (Goolsby 2009). All these examples show that regular citizens already act as observers in crisis situations and provide potentially valuable information on different social media platforms.

Current analyses for the use of social media in emergency management mostly focus on detecting large scale incidents in microblogs, e.g., the works of (Krstajic et al. 2012) or (Sakaki and Okazaki 2010). In the case of large scale incidents, the number of information shared on social media platforms is rather high, because many people might be affected. In contrast, the absolute amount of information on smaller scale incidents, like car crashes or fires, is comparably low. Thus, the automatic detection of incident related tweets is much more difficult with only a dozen of postings available.

Besides the detection of such small scale incidents, it has not been investigated yet who is sharing information about these types of incidents. This paper aims to provide an understanding of who is contributing tweets during small scale incidents. In this case, we do not want to focus on information credibility, but we want to answer if citizens tend to report tweets about small scale incidents or if only governmental agencies or news media spread this information.

In the first part of the paper, we cope with the question how to detect small scale incidents in the massive amount of user-generated content. In this case, we contribute the first classifier to identify tweets related to three different types of incidents. Our approach has 82.2% precision and 82% recall. Furthermore, we contribute an in-depth analysis who is tweeting incident related information. We show that a variety of individual users share this information and that those users are reporting faster than official sources.

Related Work

Work related to this paper arises from two areas: (1) research on the detection of small scale incidents and (2) analysis of users sharing information during events.

In the first area, only few state-of-the-art approaches focus on the detection and analysis of small scale incidents in microblogs. Agarwal et al. (2012) propose an approach to detect events related to a fire in a factory using standard NLP-based features. They report a precision of 80% using a Naïve Bayes classifier. Wanichayapong et al. (2011) focus on extracting traffic information in microblogs from Thailand. Compared to other approaches, they detect tweets related to traffic information. The evaluation of the approach shows an accuracy of 91.7%, precision of 91.39%, and recall of 87.53%. Though the results are quite promising, they restricted their initial test set to tweets containing manually defined traffic related keywords, thus, the number of relevant tweets is significantly higher than in a random stream. Li et al. (2012) introduce a system for searching and visualization of tweets related to small scale incidents, based on keyword, spatial, and temporal filtering. Compared to other approaches, they iteratively refine a keyword-based search for retrieving a higher number of incident related tweets. Their classifier has an accuracy of 80% for detecting incident re-

lated tweets, although they do not provide any information about their evaluation approach.

In the second area, only few works analyze the originators of incident related information in microblogs. Vieweg et al. (2010) examine communication on Twitter during two large scale incidents (Oklahoma Grassfires and the Red River Floods that occurred in March and April 2009). They describe that 49% to 56% of all tweets related to the incidents contain situational update information. It remains unclear, which types of users share this information. Starbird et al. (2010) also analyze tweets shared during the Red River Floods. They show that individuals comprise 37% of all tweets. On the other side, only about 3% are flood specific organizations, but they share more than 44% of the tweets. As a result they conclude that mostly governmental agencies share situational information, though, this information can further be enhanced with insights from citizens. Despite crisis related scenarios, Choudhury et al. (2012) conducted a detailed analysis of user behavior during rather common events. They show that individuals share a lot of information during local events. Organizations or journalists share more information during national events or breaking news. Summarized, related work shows that the behavior of users during events seems to differ depending on the type of event. Furthermore, the user behavior during small scale incidents has not been evaluated so far.

Study 1: Small Scale Incident Detection

For analyzing incident related microblogs, we have to identify them with high accuracy in the stream of microblogs. For doing this, we trained a classifier for detecting incident related tweets.

Crawling and Dataset

For building a training dataset, we collected 6 million public tweets in English language using the Twitter Search API[1] from November 19th, 2012 to December 19th, 2012 in a 15km radius around the city centers of Seattle, WA and Memphis, TN. For labeling the tweets, we first extracted tweets containing incident related keywords and hyponyms of these keywords. The latter are extracted using WordNet[2]. We defined four classes in our training set: ”car crash”, ”fire”, ”shooting”, and ”not incident related”. 20.000 tweets were randomly selected from the initial set and manually labeled by scientific members of our departments. The final training set consists of 213 car accident related tweets, 212 fire incident related tweets, 231 shooting incident related tweets, and 219 not incident related tweets.

Preprocessing

Every collected tweet is preprocessed. First, we remove all retweets as these are just duplicates of other tweets and do not provide additional information. Second, @-mentions are removed as we assume that they are not relevant for detection of incident related tweets. Third, very frequent words like stop words are removed as they are not valuable as features for a machine learning algorithm. Fourth, abbreviations are resolved using a dictionary compiled from www.noslang.com. Furthermore, as tweets contain spelling errors, we apply the Google Spellchecking API[3] to identify and replace them if possible.

During our evaluations, we found out that around 18% of all incident related tweets contain temporal information. For identifying what the temporal relation of a tweet is, we adapted the HeidelTime (Strötgen and Gertz 2012) framework for temporal extraction. HeidelTime is a rule-based approach mainly using regular expressions for the extraction of temporal expressions in texts. As the system was developed for large text documents, we adapted it to work on microblogs. We use our adaptation to replace time mentions in microblogs with the annotations @DATE and @TIME to use temporal mentions as additional features.

Besides a temporal filtering, a spatial filtering is also applied. As only 1% of all tweets retrieved from the Twitter Search API are geotagged, location mentions in tweet messages or the user’s profile information have to be identified. For location extraction, we use a retrained model created with Stanford NER[4]. Thus, we are able to detect location and place mentions with a precision of 95.5% and 91.29% recall. We also use our adaptation to annotate the text with the annotations @LOC and @PLC so that a spatial mention can be used as an additional feature.

Before extracting features, we normalize the words using the Stanford NLP toolkit[5] for lemmatization and POS-tagging. We use the POS-tags to extract only nouns and proper nouns, because during our evaluation we found out that using only these word types for classification improve the accuracy of the classification.

Feature Extraction

For training a classifier, we use the following features. First, word unigrams are extracted to represent a tweet as a set of words. In this case, a vector with the frequency of words and a vector with the occurrence of words (as binary values) is used. Second, for every tweet we calculate an accumulated tf-idf score (Manning, Raghavan, and Schütze 2009). Third, we use syntactic features that might be related to tweets related to some incidents. In this case, we extract the following features: the number of “!” and “?” in a tweet and the number of capitalized characters. Fourth, as spatial and temporal mentions are replaced with corresponding annotations, they appear as word unigrams or character n-grams in our model and can therefore be regarded as additional features.

Our approach also incorporates Linked Open Data features. For this, we use the *FeGeLOD* (Paulheim and Fürnkranz 2012) framework which extracts features for a named entity from Linked Open Data. For example, for the instance `dbpedia:Ford_Mustang`, features that can be extracted include the instance’s direct types (such as

[1] https://dev.twitter.com/docs/api/1.1/get/search/tweets

[2] http://wordnet.princeton.edu

[3] https://code.google.com/p/google-api-spelling-java/

[4] http://nlp.stanford.edu/software/CRF-NER.shtml

[5] http://nlp.stanford.edu/software/corenlp.shtml

Method	Training Set			
	Acc	Prec	Rec	F
SVM	**81.99%**	**82.2%**	**82.0%**	**81.9%**
NB	80.74%	81.0%	80.7%	80.7%
JRip	75.08%	75.1%	75.1%	74.7%

Table 1: Classification results for small scale incident detection on 4-class problem (car crash, fire, shooting, no incident).

`Automobile`), as well as the categories of the corresponding Wikipedia page (such as `Road_transport`). We use the transitive closures of types and categories with respect to `rdfs:subClassOf` and `skos:broader`. This allows for a semantic abstraction from the concrete instances a tweet talks about. In order to generate features from tweets with FeGeLOD, we first preprocess the tweets using *DBpedia Spotlight* (Mendes et al. 2011) in order to identify the instances a tweet talks about. Unlike the other feature extractions, the extraction of Linked Open Data features is performed on the original tweet, not the preprocessed one, as we might lose valuable named entities.

Classification and Results

The different features are combined and evaluated using three classifiers. For classification, the machine learning library Weka (Witten and Frank 2005) is used. We compare a Naïve Bayes Binary Model (NBB), the Ripper rule learner (JRip), and a classifier based on a Support Vector Machine (SVM). Our classification results are calculated using stratified 10-fold cross validation on the training set. To measure the performance of the classification approaches, we report the accuracy (Acc), the averaged precision (Prec), the averaged recall (Rec), and the F-Measure (F).

In Table 1 the classification results are shown. With 82.2% precision and 82% recall on a 4-class problem, we outperform current state-of-the-art approaches for small scale incident detection. Thus, our study demonstrates that it is possible to detect potentially valuable information in the stream of microblogs with high precision and recall, even though the absolute amount of information related to incidents is low.

Study 2: Understanding User Behavior

For analyzing which types of users are contributing information about small scale incidents, we used the manually labeled dataset described in the previous section. We focus on the number of different users contributing to an incident type and the number of tweets sent by each type of user. Furthermore, as we are interested to understand who is first reporting about an incident, we had to aggregate all tweets that are describing the same incident. Hence, we applied the aggregation algorithm we presented in (Schulz, Ortmann, and Probst 2012) to determine all tweets that describe a particular incident. For that purpose, we applied the spatial and temporal extraction, which we also used for building the classifier. As a result, all 657 incident related tweets were associated with 347 incidents: 179 car crashes, 49 shootings, and 119 fires.

Evaluation Results

We were able to identify 246 unique users that are sharing incident related tweets. Using the description of the users' Twitter profile, we manually labeled all users with different categories. Following the approach described in Choudhury et al. (2012) we identified five user categories. Official organizations like the Seattle Fire Department are categorized as *emergency management organizations* (EMO). Organizations not related to emergencies, like magazines, are clustered as *other organizations* (ORG). Furthermore, we found specialized traffic reporters or journalists, which are represented as *journalists/bloggers focused on emergency management* (EMJ), in contrast to *other journalists/bloggers* (JOU). Citizens are categorized as *individual users* (I).

The first bar of each stacked cluster in Figure 1 shows the distribution of the number of users for each category according to the different types of incidents. We can notice that a variety of different individual users (196) are reporting about the three incident types. On the other side, only few emergency management organizations (11) and focused journalists (2) are publishing incident related microblogs.

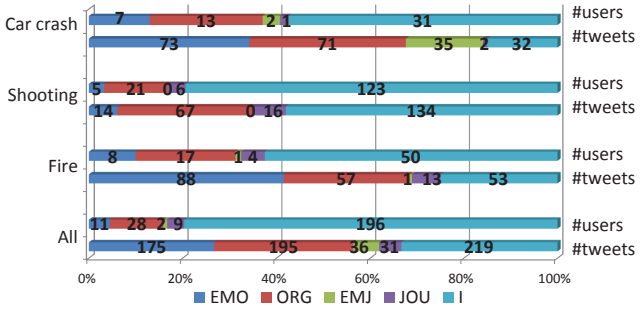

Figure 1: Distribution of number of users and number of tweets from different user categories by incident type.

The second bar of each stacked cluster in Figure 1 shows the overall number of tweets shared by each user category for the different types of incidents. We can notice that even though the number of the emergency management organizations and other organizations is significantly smaller than the number of individual users, most of the tweets are shared from users of these organizations (overall 56%). The individual users share 33.3% of the tweets, though, the number of tweets by individual users regarding the shootings is much higher. The reason for this might be that shootings are more of public interest compared to car crashes and fires. Furthermore, the results show that individual users contribute only one or at most two tweets regarding small scale incidents.

To understand who is first reporting about an incident, we compared the timestamps of the first incident reports sent by each user category. Figure 2 shows that 65% of all incident types are first reported by organizational users and only 23% of all incidents are first reported by individual users. But also

in this case, we can notice that 53% of the shootings are first reported by individual users, which might also be the case because of higher public attention.

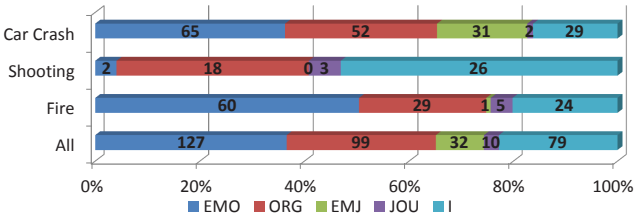

Figure 2: Distribution of first reports from different user categories by incident type.

Moreover, we focused on the incidents that were first reported by individual users (79 in total). We calculated the time difference between the first tweet shared by individual users and the first tweet sent by users from other categories. The results show that the individual users are reporting 24.13 minutes in average before users of other categories. This underlines that if citizens are sharing incident related information, they share it promptly. Nevertheless, we did not evaluate if those tweets contain valuable situational information, which is subject to future work. We can thus summarize that a variety of individual users are sharing small incident related information, though, the absolute amount of tweets is comparably low. This is also contradicting the results of Choudhury et al. (2012), which might be because they focus on common events. Second, individual users are timely reporting information. Nevertheless, large amounts of incident related information is shared by official sources, specialized bloggers, or journalists.

Conclusion

In this work, we have made several contributions. We demonstrated how machine learning can be used to detect small scale incidents with high precision. In this case, we contribute the first classifier to detect three different types of small scale incidents with 82.2% precision and 82% recall. Furthermore, we examined the user behavior during small scale incidents and showed that a variety of individual users share small incident related information and that those users are reporting faster than official sources.

For future work, it might be interesting to do a qualitative analysis of the content shared by citizens. E.g., a differentiation which information really can contribute to increasing situational awareness is necessary. Furthermore, the analysis should be extended throughout different cities.

Acknowledgments

This work has been partly funded by EIT ICT Labs under the activities 12113 Emergent Social Mobility and 13064 City-CrowdSource of the Business Plan 2012.

References

Agarwal, P.; Vaithiyanathan, R.; Sharma, S.; and Shroff, G. 2012. Catching the long-tail: Extracting local news events from twitter. In *Proceedings of the Sixth International Conference on Weblogs and Social Media ICWSM 2012, Dublin, Ireland.*

De Choudhury, M.; Diakopoulos, N.; and Naaman, M. 2012. Unfolding the event landscape on twitter: classification and exploration of user categories. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, CSCW '12, 241–244.

Goolsby, R. 2009. Lifting Elephants: Twitter and Blogging in Global Perspective. In *Social Computing and Behavioral Modeling*. Springer, Berlin, Heidelberg. 1–6.

Krstajic, M.; Rohrdantz, C.; Hund, M.; and Weiler, A. 2012. Getting there first: Real-time detection of real-world incidents on twitter. In *Published at the 2nd IEEE Workshop on Interactive Visual Text Analytics*.

Li, R.; Lei, K. H.; Khadiwala, R.; and Chang, K. C.-C. 2012. Tedas: A twitter-based event detection and analysis system. In *Proceedings of the 11th International Conference on ITS Telecommunications (ITST)*, 1273–1276.

Manning, C. D.; Raghavan, P.; and Schütze, H. 2009. *An Introduction to Information Retrieval*. Cambridge University Press. 117–120.

Mendes, P. N.; Jakob, M.; García-Silva, A.; and Bizer, C. 2011. DBpedia Spotlight: Shedding Light on the Web of Documents. In *Proceedings of the 7th International Conference on Semantic Systems (I-Semantics), Graz, Austria*. ACM.

Okolloh, O. 2008. Ushahidi, or 'testimony': Web 2.0 tools for crowdsourcing crisis information. *Participatory Learning and Action* 59(January):65–70.

Paulheim, H., and Fürnkranz, J. 2012. Unsupervised Feature Generation from Linked Open Data. In *International Conference on Web Intelligence, Mining, and Semantics (WIMS'12)*.

Sakaki, T., and Okazaki, M. 2010. Earthquake shakes Twitter users: real-time event detection by social sensors. In *WWW '10 Proceedings of the 19th international conference on World Wide Web*, 851–860.

Schulz, A.; Ortmann, J.; and Probst, F. 2012. Getting user-generated content structured: Overcoming information overload in emergency management. In *Proceedings of 2012 IEEE Global Humanitarian Technology Conference (GHTC 2012)*, 1–10.

Starbird, K.; Palen, L.; Hughes, A. L.; and Vieweg, S. 2010. Chatter on the red: what hazards threat reveals about the social life of microblogged information. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, CSCW '10, 241–250.

Strötgen, J., and Gertz, M. 2012. Multilingual and cross-domain temporal tagging. *Language Resources and Evaluation*.

Vieweg, S.; Hughes, A. L.; Starbird, K.; and Palen, L. 2010. Microblogging during two natural hazards events: what twitter may contribute to situational awareness. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, 1079–1088. New York, NY, USA: ACM.

Wanichayapong, N.; Pruthipunyaskul, W.; Pattara-Atikom, W.; and Chaovalit, P. 2011. Social-based traffic information extraction and classification. In *11th International Conference on ITS Telecommunications (ITST)*, 107–112.

Witten, I. H., and Frank, E. 2005. *Data mining: practical machine learning tools and techniques*. Amsterdam, Netherlands: Elsevier.