Analysing tweets describing during natural disasters in Europe and Asia

Kiran Zahra
Department of Geography, University of Zurich
Winterthurerstr. 190, 8057, Zurich, Switzerland
kiran.zahra@geo.uzh.ch

Ross Purves
Department of Geography, University of Zurich
Winterthurerstr. 190, 8057, Zurich, Switzerland
ross.purves@geo.uzh.ch

Abstract

Twitter is a widely known platform for speedy diffusion of views, ideas and information during different events. It has widely been used during disasters to communicate evacuation plans, help calls, and damage assessment. Reliability of information accessed during mass emergencies from social media for decision making is very important. In this research we reveal different aspects of credibility, granularity of geographic information reported in tweets and use of Naïve Bayes for tweet classification from the users of Europe and Asia. We used user-based features to assess credibility. Toponyms from tweet text are extracted with its frequency to reveal geographic feature granularity in the tweets. Naïve Bayes is used to classify tweets which is trained on one geographic location and tested for the event from another geographic region. Our results show that credibility assessment shows a complex picture for Italy and Myanmar based on user-based features. So-called fine geographic granularity has been reported from the users of Myanmar. And Naïve Bayes performs with high accuracy even if a new training data is not prepared for different geographic region.

Keywords: Naïve Bayes, Twitter, geographic feature granularity, credibility.

1 Introduction

The growth of social media over the last decade, and its possible use as a source of information about a wide variety of topics including events, news, personal opinions and many more (Hossmann *et al.*, 2011) ; (Terpstra *et al.*, 2012) is unquestionable. One widely studied investigated potential use is real-time monitoring of events (Middleton, Middleton & Modafferi, 2014). In particular, where events take the form of natural disasters any additional information with respect to casualties, damage, situational updates and evacuation plans is extremely valuable (Verma *et al.*, 2011).

However, not everything shared on social media can be considered as useful "information" with respect to natural disasters, since people also share spam, personal opinions and material which may explicitly harass other users (Senaratne *et al.*, 2016). Even if we collect tweets based on particular keywords related to a specific theme, content may still not be relevant since many words are polysemous and may also be used as synonyms or metaphors (Sakaki, Okazaki & Matsuo, 2013). Thus, I may "quake" in fear, results may be returned "like an avalanche" and we may be "flooded" with information. This makes the adoption of methods which can analysis the semantics behind particular terms very important if we wish to categorize information harvested from social media as relevant or irrelevant pieces of information.

Twitter offers free, real-time data in the form of tweets through its streaming Application Programming Interface (API). This API requires certain parameters to capture tweets such as particular keywords, tweets sent from particular users, or tweets originating from a particular region. For our project, an application was designed in R to capture near real-time tweets, based on disaster-related keywords such as earthquake, flood, hurricane etc. During the data collection phase of our project we observed a sudden rise in the number of tweets contemporaneously with events such as earthquakes or storms. A normal day tweet count over 24 hours period is around 50,000 tweets which rises to as high as 488,000 tweets in case of a disaster. It appears that users connect to Twitter even to verify a small earthquake experienced by themselves (example tweet text: "Was that #earthquake in Cali, or someone was rocking my chair?") or to know about damages and casualties caused by a major earthquake. This behavior is well known, and multiple studies have used Twitter to detect events such as earthquakes and attempt to, for example determine their geographical extent or magnitude (Sakaki, Okazaki & Matsuo, 2010). However, little attention appears to have been paid to either issues relating to the semantics of Tweets or the quality of information specifically with respect to such information, as opposed to many more general studies on the quality of Twitter and VGI more generally.

In the case-study reported on in this paper we therefore selected two natural disasters which occurred on the same date in two different geographic regions of the world to explore the nature of information and data (tweets) quality on analysis. The first disaster was an earthquake which occurred in Italy on 24

Aug, 2016 at 0336 local time and second was an earthquake in Myanmar on the same date at 1704 local time. The two earthquakes were both of strong magnitudes, (Italy 6.2 and Myanmar 6.8 on the Richter scale).

Since Tweets consist of free text, Twitter users report disasters in many different ways. One critical feature in terms of information while reporting a disaster is the granularity of geographic location reported in a tweet. Granularity with respect to a Tweet refers to the precision of the area described in a tweet – thus a tweet reporting on an event in Italy is of coarse granularity, and of limited information use, while one reporting on an event near the commune of Accumoli in the Province of Rieti in Italy has a fine granularity.

We consider any tweet containing locational information about the earthquake to be information. Since the two earthquakes reported on here happened in different continents, more or less simultaneously, our first research question took advantage of this difference in geographic location, and asked:

RQ1: How does the detail with which an event is reported in terms of its geographic feature granularity vary in two different continents?

Since tweets are user-generated data, produced for many different reasons, they are also associated with varying quality with respect to particular contexts, since different users may produce geographic information and, to use this information for decision making, data quality must be considered (Senaratne *et al.*, 2016). One important aspect of data quality is the credibility of a Tweet, that is to say how likely is it that the content is for example, accurate, authoritive, objective and current (Gupta & Kumaraguru, 2012). In our second research question we therefore explore data quality of tweets describing the two earthquake events in terms of credibility:

RQ2: What is the quality of tweets in terms of credibility for a natural disaster in Europe and Asia?

Our dataset in this study was based on disaster related tweets, and we needed a simple, repeatable methods to classify tweets as disaster related and containing useful information. We used a common approach in text classification, the supervised machine learning algorithm Naïve Bayes. Using a supervised machine learning algorithm efficiently is in turn always critically dependent on the training dataset used. During a real disaster, time is of the essence, and building a new training dataset for every event would result in a significant delay in classification. Artificial Intelligence for Disaster Response (AIDR) proposed crowdsourcing for timely preparation of training data for a particular disaster, which can be volunteered with no or limited quality assurance or may also be generated as a paid task with associated costs (Imran *et al.*, 2014). Some researchers claim, classifiers trained for one disaster work well for another disaster of the same nature (Verma *et al.*, 2011), though others have shown that classification of specifically geographic information is a challenging task, often requiring local knowledge (Ostermann, Tomko & Purves, 2013). In our case we used data related to two disasters of the same nature in two different continents. To explore the need to prepare new training data for every disaster we formulated following research question:

RQ3: How well does Naïve Bayes perform with respect to text classification of informational content for another event of the same nature, when training data for the classifier is trained using an event of a similar nature in a different location?

2 Methods

In the following we firstly explain how our datasets were collected, before describing our methods for exploring geographic granularity, Tweet credibility and classification of information in turn.

2.1 Data collection

We collected Twitter data based on disaster related keywords from the Twitter Streaming API. This API allows downloading near real-time Tweets. The streaming API provides access to some 1 - 40% of Tweets. We chose keywords to query the Twitter streaming API on general words used in English to refer to a hazard which can cause disaster.

Query keywords used in the API are space sensitive but not case sensitive. The full set of keywords we used is illustrated in Table 1.

Table 1: Set of keywords used to query Twitter API

Tsunami	flood	earthquake
Landslide	earth quake	fore shock
fore-shock	after shock	after-shock
landslide	land slide	avalanche
rockfall	rock fall	mud slide
mudslide	earth slip	earth-slip
cloudburst	cloud burst	heavy rainfall
extensive rainfall	heavy rain	extensive rain
rain storm	forest fire	inundation
overflow	flash-flood	-

We aimed to collect only tweets written in English, with no spatial restrictions, for the following reasons:

- English is the most frequently learned and spoken foreign language all over the world.
- Many researchers have used English tweets in their research.
- We were not familiar with regional languages spoken in earthquake hit areas to efficiently analyze tweet content in the native language.

Table 2: Dataset details

Size	Tweets	Start time	End time
2.54 GB	488175	Wed Aug 24 2016 8:57	Thu Aug 25 2016 8:57

2.2 Geographic granularity of tweets

We analyzed tweet text to assess how users in different regions of the world (Asia and Europe) report an earthquake with its location. We selected 500 (Verma *et al.*, 2011) tweets through stratified sampling for each earthquake and manually analysed the content and identified every geographic location reported in the tweet text with number of times it appeared in the sample dataset. These geographic locations were searched later on Geonames gazetteer and we added feature class of every location as per gazetteer on the list (Table 3 and Table 4).

Table 3: Italy earthquake granularity and occurrence of geographic locations

Class	Name	Occurrences
Independent political entity	Italy	314
	Vatican City	1
Capital of a political entity	Rome	28
First order administrative division	L'Aquila	2
	Umbria	3
	Perugia	23
Second order administrative division	-	-
Third order administrative division	Norcia	2
	Accumoli	12
	Marche	4
	Amatrice	42
Populated Place	-	-

Table 4: Myanmar earthquake granularity and occurrence of geographic locations

Class	Name	Occurrences
Independent political entity	Myanmar	95
	India	62
	Bangladesh	6
	Thailand	4
Capital of a political entity	Delhi	3
	Dhaka	4
	Bangkok	6
First order administrative division	Tripura	2
	Jharkhand	1
	Bihar	24
	Assam	36
	West Bengal	27
	Sikkim	1
	Kolkata	50
	Rajshahi	1
	Bhubaneswar	6
	Chittagong	3
	Patna	2
	Yangon	3
	Odisha	3
	Mizoram	1
	Agartala	1
	Ranchi	4
Second order administrative division	Deoghar	1
	Burdwan	1
	Jalpaiguri	1
	Khagaria	1
	Malda	1
Third order administrative division	-	-
Populated Place	Guwahati	12
	Balasore	1
	Noida	1
	Midnapur	1
	Gurgaon	1
	Durgapur	1

2.3 Credibility Assessment

To assess the quality of VGI where International Standard Organization (ISO) standards are not applicable, abstract quality indicators are used (Senaratne *et al.*, 2016). According to the Merriam Webster dictionary credibility is defined as "the quality of being believed or accepted as true, real, or honest"[1]. The credibility of Twitter has been studied by many researchers. A set of message, user, topic and propagation based features were highlighted as important features related to tweet credibility by (Poblete, Castillo & Mendoza, 2011). We adopt user-based features (Table 5) for this case-study to assess the credibility of tweets.

Table 5: User-based features for credibility assessment

User-based features	Description
Registration age	The time passed since the author registered their account
Statuses count	The number of tweet sent by the user
Followers count	Number of people following this user
Friends count	Number of people user is following
Verified	If the account has been verified
Has description	A non-empty bio
Has URL	A non-empty homepage URL

We selected user provided "location" field to filter tweets from our dataset. This field is entered by users at the time of creating their account, or may be added later, and is a free-text format field. For the Italian earthquake we filtered our dataset based on a query which selected all the records which contain Italy in location field. For Myanmar earthquake we used four countries India, Bangladesh, Myanmar and Thailand, because Myanmar earthquake was felt in these four countries. This query returned 4773 records for Italy earthquake and 16797 records for Myanmar earthquake. We selected 500 records by random sampling for each event to study credibility of tweets originating from these two regions.

We assume that credibility is a function of user-based features with the following form:

[1]https://www.merriam-webster.com/dictionary/credibility

C= f (FrC, SC , FoC , AG, U, D, V)

Where C is credibility, FrC, SC, FoC and AG are friends count, statuses count, followers count and account age (in years) respectively. Other features such as U represent whether users are associated with a Uniform Resource Locator (URL), D whether users have added a description or bio, and V if a user has a verified account. These three features are represented by Boolean values.

We compared the properties of each feature for our two areas, to test the hypothesis that credibility related attributes varied according to locations.

2.4 Classification rules

We defined two categories to classify our data into two classes: Information and Not information. These classes are defined as follow:

- **Information:** Tweet text about disaster event and its location.
- **Not Information:** Everything else falls in this category.

We then created training and test data by annotating 1000 Tweets from Italy, where we attempted to balance the number of Tweets in each class. A second smaller dataset was labelled for Myanmar, which was used exclusively as test data.

We used a state of the art supervised machine learning algorithm, Naïve Bayes, to classify tweets according to terms. In the case of the Italian earthquake, we used 70% of our initial annotated dataset as training data, and 30% to test performance. We ran the same classifier on our Myanmar data to explore the ability of the classifier to identify Tweets containing information when trained on annotated Tweets from a different region.

3 Results and interpretation

3.1 Geographic granularity

Figures 1 and 2 show places named and their frequencies in Myanmar and Italy. We attempted to use the hierarchy of administrative regions as used by Geonames to explore the granularity of the spatial information available (Tables 3 and 4). However, though Myanmar appears to contain information of finer granularities (populated places as opposed to the third order administrative regions in Italy) it is clear that the toponyms used in Italy cover a much more tightly defined region, while in Myanmar many tweets appear to be from the surrounding countries.

3.2 Credibility assessment

We assessed the difference between a number of variables commonly associated with credibility for two events with the same number of Tweets and occurring at similar times. The count of friends, statuses, followers and account ages are illustrated in Table 6. We tested significance of differences using a Mann-Whitney U test, and found that the count of friends, followers and average account age were all significantly different ($p < 0.05$). However, these differences were asymmetric with accounts in Italy being associated with more friends and a greater account age, while those in Myanmar had more statuses (though not significantly) and more followers.

Table 6: Differences between credibility related attributes

Attribute	Italy	Myanmar
Friends count	1320 ± 3839	1073 ± 2369
Statuses count	31498 ± 79831	52067 ± 101762
Followers count	2082 ± 5479	3966 ± 22622
Account age	5.32 ± 2.25	3.45 ± 2.53

Finally, we found that users in Italy were more likely to have URLs associated with their accounts, while there was little difference in the number of users with descriptions between the two locations.

3.3 Classification results

We used our test data to evaluate our classifier's performance on data from Italy (Table 7). The precision of our classifier was very high 98% for tweets classified as containing information, suggesting that almost all tweets classified using this approach contain information, while the recall of 93% means that a small number of tweets were falsely discarded. When running the classifier on a different geographical region performance decreased somewhat but remained relatively high.

Table 7: Confusion Matrix for Italy

		Actual Class		
	Class	Information	Not Information	Precision
Predicted Class	Information	147	3	98%
	Not Information	11	139	92.68%
	Recall	93.038 %	97.887%	

We performed classification on Myanmar test data without preparing any training data from Myanmar earthquake event in our training dataset.

Table 8: Confusion Matrix for Myanmar

		Actual Class		
	Class	Information	Not Information	Precision
Predicted Class	Information	133	17	88.667%
	Not Information	12	138	92%
	Recall	91.724 %	89.032%	

4 Concluding Discussion

In this paper we set out to compare data related to natural hazard events that occurred more or less contemporaneously in two very different locations, Myanmar and Italy. When exploring the granularity of locations reported in tweets, an initial analysis based only on hierarchies derived from Geonames suggested that the toponyms used in Myanmar were of finer granularities. However, mapping the data clearly shows the more or less total absence of detailed data in Myanmar, as compared to the finer data in Italy. These results reinforce the importance of considering data divides (e.g. Graham *et al.* 2014) when analyzing such data, and also reflect the difficulties of using Volunteered Geographic Information (VGI) itself (here in the form of Geonames) to do so. Our results for credibility also suggest a complex picture. We expected a clear difference between Tweets related to Italy and Myanmar, but in fact observe that, at least for user attributes these perhaps better reflect different user characteristics (users reporting on events in Asia appear to tweet more often and have more followers, while those reporting on Europe seem to have older accounts and more friends. Finally, our approach to classifying tweets in terms of informational content did provide satisfying results, and suggested that, at least for tweets in English describing similar events, a machine learning approach not only performed well, but did so independently of geographic location.

References

Castillo, C., Mendoza, M. & Poblete, B., 2011, March. Information credibility on twitter. In *Proceedings of the 20th international conference on World wide web* (pp. 675-684). ACM.

Graham, M., Hogan, B., Straumann, R. K., & Medhat, A. (2014). Uneven geographies of user-generated information: patterns of increasing informational poverty. Annals of the Association of American Geographers, 104(4), 746-764.

Gupta, A. & Kumaraguru, P., 2012, April. Credibility ranking of tweets during high impact events. In *Proceedings of the 1st Workshop on Privacy and Security in Online Social Media* (p. 2). ACM. [Online] 2–8. Available from: http://dl.acm.org/citation.cfm?doid=2185354.2185356.

Hossmann, T., Carta, P., Schatzmann, D., Legendre, F., Gunningberg, P. & Rohner, C., 2011, December. Twitter in disaster mode: Security architecture. In *Proceedings of the Special Workshop on Internet and Disasters* (p. 7). ACM. [Online] Available from: doi:10.1145/2079360.2079367.

Imran, Muhammad, et al. "AIDR: Artificial intelligence for disaster response." *Proceedings of the 23rd International Conference on World Wide Web*. ACM, 2014. [Online]. 2014 pp. 159–162. Available from: doi:10.1145/2567948.2577034.

Middleton, S.E., Middleton, L. & Modafferi, S., 2014. Real-time crisis mapping of natural disasters using social media. *IEEE Intelligent Systems*, *29*(2), pp.9-17. [Online] 29 (2), 9–17. Available from: doi:10.1109/MIS.2013.126.

Ostermann, F.O., Tomko, M. & Purves, R., 2013. User evaluation of automatically generated keywords and toponyms for geo-referenced images. *Journal of the American Society for Information Science and Technology*, *64*(3), pp.480-499.

Senaratne, H., Mobasheri, A., Ali, A.L., Capineri, C. & Haklay, M., 2017. A review of volunteered geographic information quality assessment methods. *International Journal of Geographical Information Science*, *31*(1), pp.139-167.

Sakaki, T., Okazaki, M. and Matsuo, Y., 2010, April. Earthquake shakes Twitter users: real-time event detection by social sensors. In *Proceedings of the 19th international conference on World wide web* (pp. 851-860). ACM.. [Online] 851–860. Available from: doi:10.1145/1772690.1772777.

Sakaki, T., Okazaki, M. & Matsuo, Y., 2013. Tweet analysis for real-time event detection and earthquake reporting system development. *IEEE Transactions on Knowledge and Data Engineering*, *25*(4), pp.919-931. [Online] 25 (4), 919–931. Available from: doi:10.1109/TKDE.2012.29.

Terpstra, T., De Vries, A., Stronkman, R. & Paradies, G.L., 2012. *Towards a realtime Twitter analysis during crises for operational crisis management* (pp. 1-9). Simon Fraser University.

Verma, S., Vieweg, S., Corvey, W., Palen, L., Martin, J.H., Palmer, M., Schram, A. & Anderson, K.M., 2011. NLP to the Rescue?: Extracting" Situational Awareness" Tweets During Mass Emergency. In *Fifth International AAAI Conference on Weblogs and Social Media*.

Figure 1: Myanmar earthquake geographic feature granularity

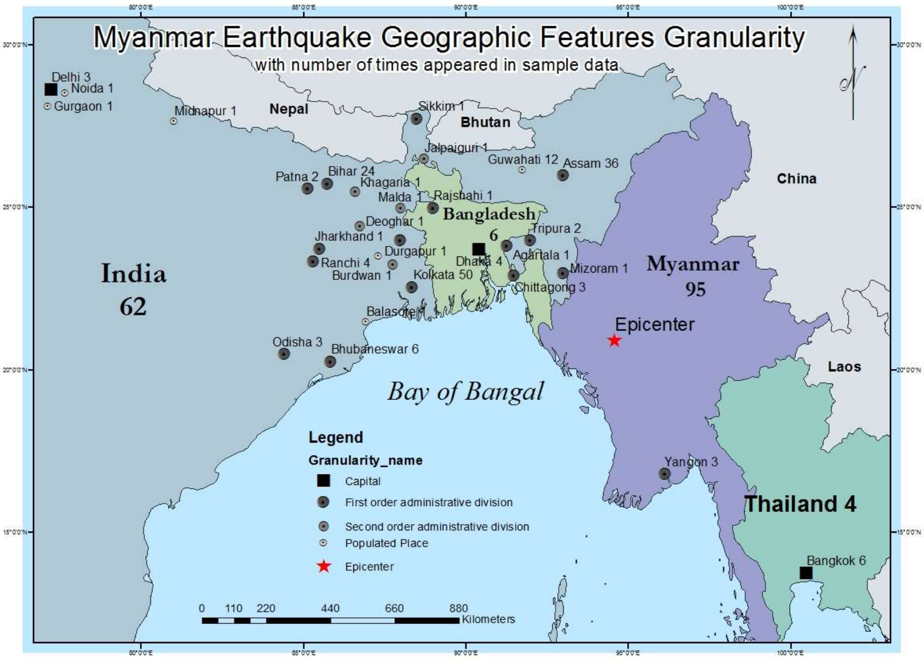

Figure 2: Italy earthquake geographic feature granularity