

Geographic variability of Twitter usage characteristics during disaster events

Kiran Zahra, Frank O. Ostermann & Ross S. Purves

To cite this article: Kiran Zahra, Frank O. Ostermann & Ross S. Purves (2017) Geographic variability of Twitter usage characteristics during disaster events, Geo-spatial Information Science, 20:3, 231-240, DOI: [10.1080/10095020.2017.1371903](https://doi.org/10.1080/10095020.2017.1371903)

To link to this article: <https://doi.org/10.1080/10095020.2017.1371903>



© 2017 Wuhan University. Published by Taylor & Francis Group



Published online: 07 Sep 2017.



Submit your article to this journal [↗](#)



Article views: 546



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 1 View citing articles [↗](#)

Geographic variability of Twitter usage characteristics during disaster events

Kiran Zahra^a, Frank O. Ostermann^b and Ross S. Purves^a

^aDepartment of Geography, University of Zurich, Zurich, Switzerland; ^bFaculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Enschede, The Netherlands

ABSTRACT

Twitter is a well-known microblogging platform for rapid diffusion of views, ideas, and information. During disasters, it has widely been used to communicate evacuation plans, distribute calls for help, and assist in damage assessment. The reliability of such information is very important for decision-making in a crisis situation, but also difficult to assess. There is little research so far on the transferability of quality assessment methods from one geographic region to another. The main contribution of this research is to study Twitter usage characteristics of users based in different geographic locations during disasters. We examine tweeting activity during two earthquakes in Italy and Myanmar. We compare the granularity of geographic references used, user profile characteristics that are related to credibility, and the performance of Naïve Bayes models for classifying Tweets when used on data from a different region than the one used to train the model. Our results show similar geographic granularity for Myanmar and Italy earthquake events, but the Myanmar earthquake event has less information from locations nearby when compared to Italy. Additionally, there are significant and complex differences in user and usage characteristics, but a high performance for the Naïve Bayes classifier even when applied to data from a different geographic region. This research provides a basis for further research in credibility assessment of users reporting about disasters

ARTICLE HISTORY

Received 23 June 2017
Accepted 12 August 2017

KEYWORDS

Geographic feature granularity; Volunteered Geographic Information (VGI); Naïve Bayes; Twitter; credibility; Geonames

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, and Modafferi 2014). In particular, where events take the form of natural disasters additional information with respect to casualties, damage, situational updates, and evacuation plans has the potential to be extremely valuable (Verma et al. 2011).

However, not everything shared on social media can be considered as useful and actionable information with respect to natural disasters, since people also share spam, personal opinions, and material to explicitly harass other users (Senaratne et al. 2017). Even if we collect Tweets based on particular keywords related to a specific theme, the retrieved content may still not be relevant since many words and phrases are polysemous and may also be used as synonyms or metaphors (Sakaki, Okazaki, and Matsuo 2013). Thus, one may “tremble” in fear, “like an avalanche,” and we may be “flooded” with information, and “fire” is used in many metaphors about emotions. This makes the adoption of methods which can analyze

the semantics behind particular terms very important if we wish to categorize information harvested from social media as relevant or irrelevant pieces of information with respect to a particular class of events.

Twitter currently offers access to 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, we wrote a script in R to capture 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. This observation forms the basis of many event detection applications which claim to detect events in near real time (e.g. Sakaki, Okazaki, and Matsuo 2010).

The normal daily count of Tweets containing our keywords is around 50,000, but it rises tenfold to maxima of around 500,000 Tweets in case of disasters. 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 determine their geographical extent or magnitude (Sakaki, Okazaki, and Matsuo 2010), among other things. However, little attention appears to have been paid to issues relating either to the semantics of Tweets or the specific quality of information, as opposed to many more general studies on the quality of Twitter and Volunteered Geographic Information more generally. Especially, the potential geographic variability in the usage of Twitter remains a concern to be addressed, as it impacts on the potential transferability of methods to assess and evaluate Tweets.

In the case-study reported in this paper, which extends a workshop paper on the same topic (Zahra and Purves 2017), we selected two natural disasters which occurred on the same date in two different geographic regions of the world to explore the geographic variability of Tweets and its impact on information content, credibility-related characteristics, and trained models to classify Tweets. The first disaster was an earthquake which occurred in Italy on 24 August 2016 at 03:36 local time, and the second one was an earthquake in Myanmar on the same date at 17:04 local time. The two earthquakes were both of strong magnitudes (Italy 6.2 and Myanmar 6.8 on the Richter scale).

Since Tweets contain free text, Twitter users can report on disasters in many different ways. One critical feature in terms of information content that relates to fitness-for-purpose of Tweets is the granularity of the reported geographic location. We defined granularity with respect to a Tweet as referring to the specificity or 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 and higher information value.

There are four types of location information associated with a Tweet: GPS coordinates formatted as GeoJSON in the “coordinates” metadata field, a place indicated by the user in the “place” metadata field using Twitter’s database of places, a location mentioned in the user profile’s “location” free-form metadata field, and a location mentioned in the Tweet’s content. We focus on the latter, because only 1–2% of all Tweets have GPS coordinates, the “place” metadata is often too coarse at the country level, the user profile location is often incorrect and static (Hecht et al. 2011), and we are interested in the location being tweeted about. We consider any Tweet containing locational information about the earthquake to be a potential source of information.

Geonames is an open source gazetteer that offers a standardized administrative hierarchy for different countries of the world, thereby assisting in the analysis of

the granularity of place names (toponyms) used between different regions. Our first research question took advantage of this feature:

RQ1: How does the spatial granularity with which an event is reported in terms of toponym hierarchy according to Geonames vary in two different continents?

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, authoritative, objective, and current (Gupta and Kumaraguru 2012). Since Tweets are user-generated data, produced for many different reasons, they are also associated with varying quality with respect to particular contexts (Senaratne et al. 2017). We assume that the usage characteristics of contributors can help to assess the credibility of a particular Tweet. In our second research question, we therefore explore the different user-based features of Tweets.

RQ2: What is the difference in user attributes which can help assessing credibility of Tweets during natural disasters from Europe and Asia?

Another important requirement for using Twitter in the context of disaster is to be able to distinguish between signal and noise. We needed a simple, repeatable, and reproducible method to classify Tweets as disaster related and containing useful information. We therefore used a common approach in text classification, the supervised machine learning algorithm Naïve Bayes. The performance of any supervised machine learning algorithm is dependent on the training data-set used. During a real disaster, time is of the essence, and building a new training data-set for every event could result in a significant delay in classification (Spinsanti and Ostermann 2013).

One possible solution is crowdsourcing the labeling 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). While some researchers claim that classifiers trained for one disaster work well for another disaster of the same nature (Verma et al. 2011), others have shown that classification of specifically geographic information is a challenging task, often requiring local knowledge (Ostermann, Tomko, and 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 the 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?

The overall aim of this research is to analyze Twitter usage characteristics of users residing in two different continents of the world typically characterized as

developed (Italy) and developing (Myanmar) regions. The main contribution of this study is to analyze how users from a developed country and a developing country report similar kind of disasters and how different or similar are the user-based credibility assessment features of Twitter users who are reporting about the disaster. This paper also analyzes the granularity of toponyms in Tweets.

2. Related work

In the following, we briefly introduce related work with respect to each of our three research questions. The potential role of VGI in disaster management (Goodchild and Glennon 2010) has become more important as mobile technologies have become increasingly ubiquitous (Sarda and Chouhan 2017). Thus, emerging technologies and the increased use of social media have changed the speed and ways in which people use and share information during disasters (Hughes et al. 2008). Ostermann and Spinsanti (2011) highlight the main challenges in using such content including a lack of structure to generate information (particularly in case of Twitter), the huge volume of data and a lack of quality control. As governmental authorities and disaster response agencies as well as individuals continue to use such data for disaster management, these challenges need to be addressed (Haworth and Bruce 2015).

Our first research question therefore concerns the granularity of locations present in Tweet content. Despite a wealth of research attempting to georeference Tweets, including many approaches using language-based models where toponyms are treated as potential features (e.g. Kinsella, Murdock, and O'Hare 2011), there is a dearth of research exploring the specifics of locational information associated with Tweets. Many locational models are implicitly very coarse, for example measuring accuracy with respect to 0.1° grids (e.g. Wing and Baldrige 2011).

However, conversely, studies using georeferenced Tweets typically assume that the coordinates associated with a Tweet accurately reflect the location of the content (e.g. Li, Lei, and Khadiwala 2012) despite more recent work suggesting a weak relationship between the locations of points of interest (POI) and content associated with these POIs (Hahmann, Purves, and Burghardt 2014). In practice, disaster-related applications of Twitter often seem to assume that data delivered are of a granularity appropriate to the task at hand, without any clear analysis of the ways in which locations are described in Tweet content, the association between content and locations and any analysis of variation in granularity as a function of the region being studied. While such assumptions may be justifiable when Tweets are averaged to create, for example, density surfaces, the granularity of locational information with respect to individual Tweets and their information content is important if these are to be treated as actionable information.

Our second research question focuses on the extraction and analysis of attributes argued to be associated with credibility in Twitter. According to the Merriam Webster dictionary, credibility is defined as “the quality of being believed or accepted as true, real, or honest”¹. Despite the sheer volume of data shared on Twitter, not every Tweet provides information and facts related to an event (Gupta and Kumaraguru 2012). Rather, trending topics on Twitter, including disasters, can provide an opportunity for spammers to share spams using keywords associated with trending topics and generate revenue (Benevenuto et al. 2010). Such intrusions from spammers and other sources can make the credibility of information mined from social media platforms questionable (Morris et al. 2012). Senaratne et al. (2017) discussed possible quality indicators for VGI and argue that when International Standard Organization (ISO) standard measures are not applicable to assess quality, researchers tend to use more abstract indicators including credibility, trustworthiness, text content quality, etc. O'Donovan et al. (2012) suggest using features in Tweets such as “hashtags, reTweets and mentions” to predict the credibility of Tweet content. Conversely, Canini, Suh, and Pirolli (2011) use the approach of ranking individual social media users on the relevance and their expertise on the content they share to assess credibility of the content. Ostermann and Spinsanti (2012) successfully use geographic context information to assist in filtering relevant information on forest fires. Castillo, Mendoza, and Poblete (2011) combine four sets of features based on propagation, message, topic, and users to determine credibility of trending topics. They demonstrated that machine learning algorithms trained on these features can automatically classify credible and not credible trends with good precision and recall.

However, Gupta and Kumaraguru (2012) aim to assess credibility at the level of individual Tweets and argue that assessing credibility at a trending topic level is insufficient as trending topic about an earthquake may be true but Tweets about misleading magnitude can question the credibility. They used message and source-based features to determine credibility of individual Tweets. Castillo, Mendoza, and Poblete (2012) discuss different set of features such as Tweet length, friends count, followers count, etc. to determine credibility of Tweets. Becker, Naaman, and Gravano (2011) studied the techniques of assessing the quality of Tweets based on relevance to a particular topic instead of studying the truthfulness and factual credibility of Tweet content which is an important perspective in case of disasters. It is thus clear that credibility is a complex and important topic, where it is unclear which features and approaches are most appropriate in assessing credibility. Furthermore, it is also unclear how features thought to be associated with credibility vary in space as a result of, for example, different patterns in the use of Twitter in different locations.

Our final research question concerned the extraction of Tweets containing information using machine learning approaches. Extracting useful Tweets using crowdsourcing is a key task when dealing with large volumes of data which have been argued to be rapid and effective ways of collecting data for time-sensitive events such as natural disasters (e.g. Wald et al. 2011; Imran et al. 2014; Haubrock et al. 2017).

3. Methods

In this section following we first explain how our datasets were collected, before describing our methods for exploring geographic granularity, study of credibility related features, and classification of information.

3.1. Data collection

We collected the Twitter data based on disaster-related keywords from the Twitter Streaming API. This API allows retrieval of Tweets in real time. The streaming API provides access to some 1–40% of Tweets.² 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 and the data-set detailed in Table 2.

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

- English is one of the most frequently learned and spoken second languages worldwide.
- Many researchers have used English Tweets in their research.
- We were not familiar with all regional languages spoken in earthquake hit areas, making analysis in local languages difficult.

3.2. 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 analyzed the content and

identified every geographic location (place name) reported in the Tweet text and the number of times it appeared in the sample data-set. These geographic locations were then identified in Geonames gazetteer, and we added feature classes to every location as per gazetteer on the list. While searching Geonames for geographic locations we came across ambiguous cases typically during geocoding:

- (1) Presence of the same geographic location in different countries.
- (2) Same geographic location categorized in different administrative hierarchies, e.g. Deoghar in India, is categorized as second-order administrative division as well as a populated place (City or Town).

To resolve the first ambiguity, we went through the full content of Tweet text to try to resolve the appropriate country. For the second case, we assumed that users are talking about finer granularity locations in the Geonames hierarchy (thus are more likely to be naming towns or villages than a containing administrative region of the same name).

We retrieved all the Tweets reporting on Myanmar and Italy earthquake from our database using these toponyms with “earthquake” as keywords which resulted in 47,557 Tweets for Myanmar and 234,620 Tweets for Italy. We counted the number of times a toponym occurred in whole data and compared them with number of occurrences of toponyms in sample data. We made this comparison to know the difference between the facts drawn from sample data vs. actual data. In sample data, there were some toponyms such as Tyrrhenian Sea, 66.6 miles from Vatican City, Himalayas, and South Indian Ocean which were not considered for toponyms count.

We performed a second comparison between hierarchies of toponyms according to Geonames gazetteer to analyze how users report about earthquake location in different times of during and post-disaster phases. We divided our data into 2-h intervals to make the difference more visible and counted the occurrence of every location in the group of geographic hierarchy occurring in the data. We have post disaster data for Italy, since the earthquake occurred at 01:36 UTC and first Tweet in our data is at 08:57 UTC, but for Myanmar, the data are during and post-disaster as earthquake occurred at 10:34 UTC and first Tweet in our data is at 10:36 UTC.

3.3. User-based attribute assessment

We adopted user-based features (Castillo, Mendoza, and Poblete 2011) for this case study to assess user-based attributes which are important for credibility assessment of Tweets (Table 3).

We selected the user provided “location” field to filter Tweets from our data-set for Italy and Myanmar. This field is entered by users at the time of creating their account, or may be added later, and is a free-text format

Table 1. Set of Keywords used to query Twitter Streaming API.

List of keywords used for the querying		
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	–

Table 2. Data-set details.

Size	Tweets	Start time	End time
2.54 GB	488,175	Wednesday 24 August 2016 08:57	Thursday 25 August 2016 08:57

Table 3. User-based attributes.

User-based features	Description
Registration age	The time passed since the author registered their account
Statuses count	The number of Tweets 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

field. We wrote a new query for our research question two, because we wanted to collect only the Tweets for which the users claimed to be in earthquake hit location. For the Italian earthquake, we filtered our data-set 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 16,797 records for Myanmar earthquake. We selected 500 records by random sampling for each event to analyze credibility related user-based attributes of Tweets originating from these two regions. We assume that credibility is a function of user-based features, as follows:

$$C = f(FrC, SC, FoC, AG, U, D, V) \quad (1)$$

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.

3.4. 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 used a supervised machine learning algorithm, Naïve Bayes, to classify Tweets according to frequency of earthquake-related terms in the sample corpus. We

used a ratio of 7:3 for training and test data and tested Naïve Bayes on three different cases. We annotated 350 Tweets as “not information” class to train the classifier on not information class and also prepared 300 Tweets (150 information, 150 not information) as Italy earthquake test data and 300 Tweets (150 information, 150 not information) as Myanmar earthquake test data. This data remained the same in all three cases to train Naïve Bayes on “not information” class and test the classifier on Italy and Myanmar earthquake event.

For the first case, we annotated 350 Tweets from the Italian earthquake as information class, coupled with 350 Tweets as not information class to train the classifier, and we tested it on 300 Tweets prepared as Italy test data. Then we replaced independent entity geographic feature “Italy” in Italian training data-set for information class with “Myanmar” keeping rest of the content same and used this new data to train the classifier to run on Myanmar test data to explore the ability of the classifier to identify Tweets containing information when trained on annotated Tweets from a different region only with same geographic location in text.

For the second case, we annotated 350 Tweets from Myanmar earthquake event as information class to train the classifier on information class, and we tested it on Myanmar test data. Then we replaced independent entity geographic feature in Myanmar training data-set with “Italy” keeping rest of the content same and used this new data to train the classifier to run on Myanmar test data.

For the third case, the information class contained 175 Tweets from Italy and 175 Tweets from Myanmar earthquake event and tested it on Italy and Myanmar test data.

4. Results and interpretation

4.1. Geographic granularity according to Geonames

Figures 1 and 2 show the places named and their frequencies in Myanmar and Italy sample data-set. We attempted to use the hierarchy of administrative regions as used by Geonames to explore the granularity of the spatial information available.

However, though Myanmar appears to have information of finer granularities as Italy, it is clear that the toponyms used in Italy cover a much more tightly defined region, while for Myanmar, many Tweets appear to be from the surrounding countries.

Since our initial results are based on stratified sampling of Tweets, we also retrieved all Tweets containing these toponyms and explored their relative distribution in our corpus as a whole. The approach taken means that we only retrieve Tweets identified by our manual annotation, but gives insight into the usage of these place names in a much larger sample.

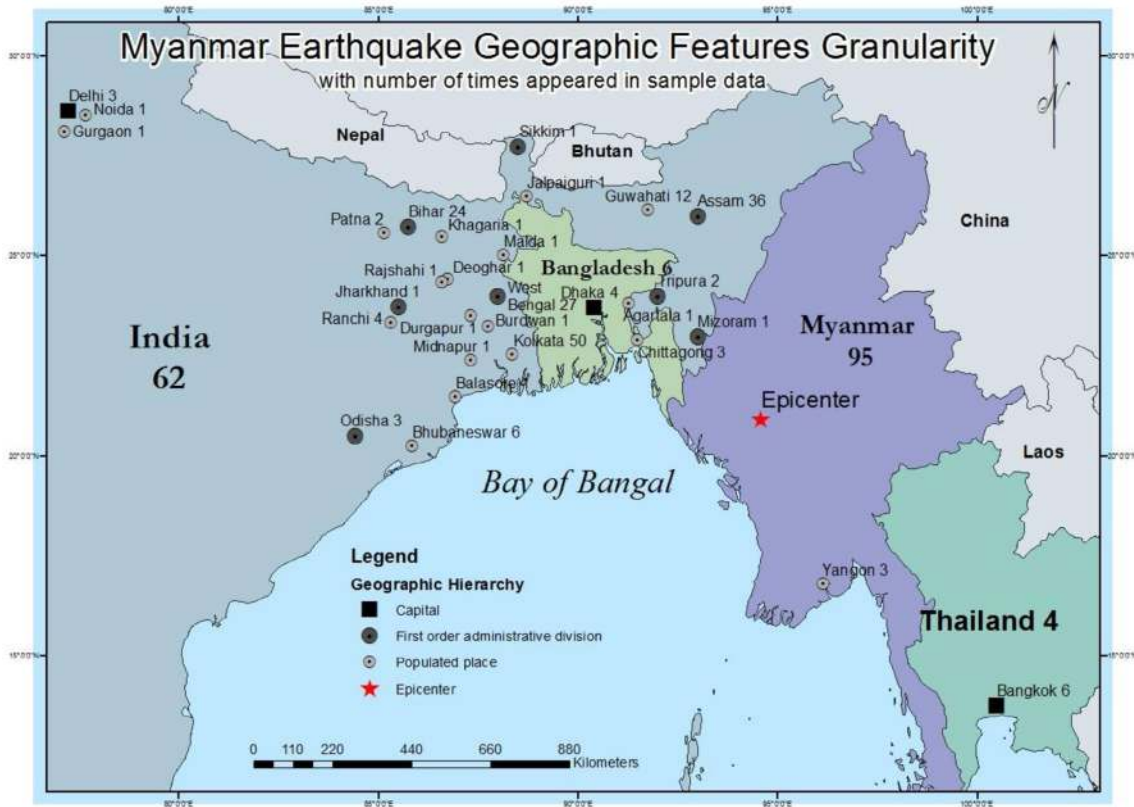


Figure 1. Myanmar earthquake geographic feature granularity according to Geonames.

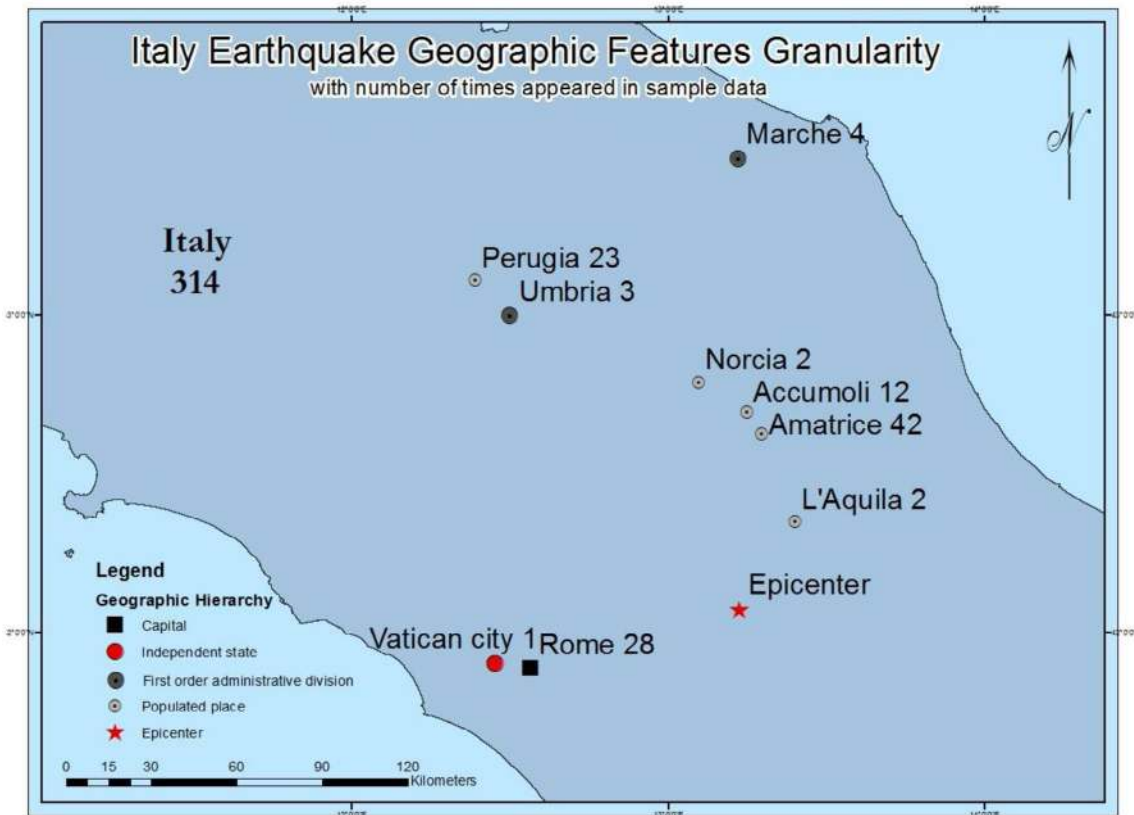


Figure 2. Italy earthquake geographic feature granularity according to Geonames.

The relatively small number of toponyms in the sample for Italy clearly shows that our sample data represent the overall distribution of toponyms well,

and indeed this is confirmed by a Kendall's Tau Rank Correlations of 0.7 for Italy ($p < 0.05$). Notable are the prominence of very coarse grained toponyms (e.g.

Italy) and Rome (as the capital city) which provide very limited spatial information with respect to the earthquake.

In the case of Myanmar the picture is more complex, since very little data actually come from the country itself, other than in the form of the country name, and a large number of Tweets are from India. These Tweets appear to be primarily from regions where the earthquake was physically felt, but demonstrate a clear bias away from the areas most seriously affected by this event toward those where, we speculate, engagement with social media in general, and Twitter in particular is higher. Nonetheless, our sample appears to reflect overall behavior well with Kendall's Tau Rank Correlations of 0.67 for Myanmar ($p < 0.01$).

Figures 3 and 4 show the usage of toponyms over time for the two incidents. Notable is the relatively constant ratio of usage, with in both cases the country name being by far the most common, followed by populated places. Both data-sets also show a slow decline in Tweets using these place names immediately after the event.

4.2. User-based attributes 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 4. We tested significance of differences using a Mann–Whitney U test, and found that the count of friends, followers,

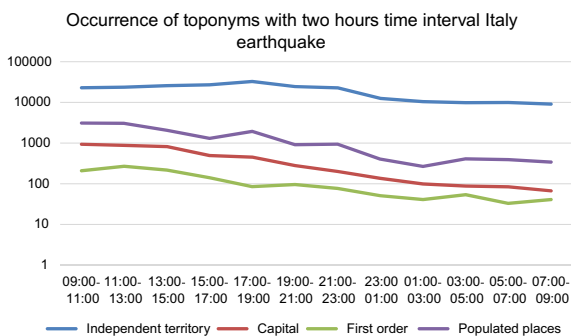


Figure 3. Occurrence of toponyms over time in Italy.

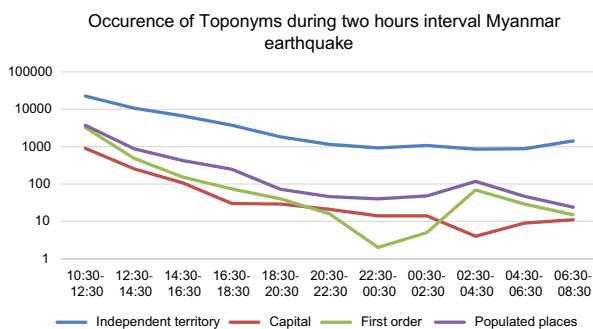


Figure 4. Occurrence of toponyms over time in Myanmar.

Table 4. Differences between credibility-related attributes.

Attribute	Italy	Myanmar
Friends count	1320 ± 3839	1073 ± 2369
Statuses count	31,498 ± 79,831	52,067 ± 101,762
Followers count	2082 ± 5479	3966 ± 22,622
Account age	5.32 ± 2.25	3.45 ± 2.53

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. 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.

These results point to the difficulty of assessing credibility using simple measures which are not normalized for local differences, since it appears that for events of the same class in different locations we find users with very different average behaviors, implying that a globally applied credibility metric is likely to capture differences in the local properties of Twitter users rather than differences in the credibility of content at these locations.

4.3. Classification results

For the first case, we used our test data to evaluate our classifier's performance on data from Italy (Table 5). 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 a recall of 93% means that a small number of Tweets were falsely discarded. When running the classifier on a different geographical region by replacing only geographical features in text, the performance decreased somewhat but remained relatively high (Table 6). This result has important implications, as it suggests that training data from other regions may help us extract information from Tweets.

For the second case, we applied the same approach as for case one but swap Italy with Myanmar. The results

Table 5. Confusion matrix for Italy (Case 1).

Predicted class	Class	Actual class		Precision
		Information	Not information	
Information	Information	147	3	98%
	Not information	11	139	92.667%
Recall		93.038%	97.887%	–

Table 6. Confusion matrix for Myanmar (Case 1).

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

Table 7. Confusion matrix for Myanmar (Case 2).

Predicted class	Class	Actual class		Precision
		Information	Not information	
	Information	149	1	99.333%
	Not information	11	139	92.667%
	Recall	93.125%	99.286%	–

Table 8. Confusion matrix for Italy (Case 2).

Predicted class	Class	Actual class		Precision
		Information	Not information	
	Information	141	9	94%
	Not information	37	113	75.333%
	Recall	79.213%	92.623%	–

Table 9. Confusion matrix for Italy (Case 3).

Predicted class	Class	Actual class		Precision
		Information	Not information	
	Information	146	4	97.333%
	Not information	36	114	76%
	Recall	80.22%	96.61%	–

Table 10. Confusion matrix for Myanmar (Case 3).

Predicted class	Class	Actual class		Precision
		Information	Not information	
	Information	146	4	97.333%
	Not information	37	113	75.333%
	Recall	79.781%	96.581%	–

(Table 7) for Myanmar show very high precision 99% with classifier trained on Myanmar data. For Italy (Table 8), the classifier was trained on Myanmar data but performed very well with 94% precision for Tweets reporting on Italian earthquake.

For the third case, results show (Tables 9 and 10) again very high precision 97% for both Italy and Myanmar as the classifier was trained on 50% data from Italy and 50% data from Myanmar.

5. 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 of finer granularities were more common than in Italy. However, mapping the data clearly show 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. Murthy 2011; Graham et al. 2014) when analyzing such data, and also reflect the difficulties of using VGI itself (here in the form of Geonames and Twitter) to do so. Bagan, the city in Myanmar which was

the most affected city in terms of damages and casualties, was not reported in our sample data even once. It could be argued that this is simply a function of the sample of data which we annotated.

However, by extending our analysis using toponyms found as search terms in Tweets referring to earthquakes, we found that Kendall Tau rank correlations were high and statistically significant. This result has important implications, as it suggests, firstly, that a well-stratified data-set is adequate for the exploration of toponym usage. However, in terms of actionable information the slow decline of toponym usage immediately after such significant events suggests the challenges present in identifying truly actionable and local information from such data streams.

Our results exploring attributes commonly used in the assessment of credibility suggest an equally 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). Since our results point to differences in the use of Twitter in different locations, and these differences are reflected in attributes previously associated with credibility, we suggest that efforts on understanding credibility would be better focused on content rather than proxy information. One promising approach to such problems is that proposed by Truelove, Vasardani, and Winter (2015) who aim to identify first-person witness accounts in Twitter.

Underpinning the importance of looking at content when trying to understand the nature of informational content was the performance of simple, off-the-shelf machine learning methods in classifying Tweets. Here we found that, independent of the location of Tweets being classified or the training data used we could identify Tweets containing information with a precision of the order of between 88 and 99%. Recall, while the lower value (with a minimum of 79%) was also satisfying, and again we argue that this result points to the importance of analyzing content when assessing the quality of information provided by Twitter with respect to natural hazards.

This paper contributes to our understanding of how social media sources are being used in different geographic regions, and the important question which analytical approaches may be suitable to transfer and reproduce methods from one geographic region to another. While our results once again highlight the often underestimated digital divide, the successful use of relatively straightforward analytical methods to both data-sets promises that global body of knowledge and methodological toolkit is possible. However, it is important to also be clear in stating that

our approaches currently are far from being capable of identifying actionable and additional information suitable for use in applications of such data. We suggest that in the rush to exploit social media to produce academic output, there is also an urgent need for more thoughtful and critical work such as the analysis of Mission 4636 after the Haiti earthquake by Munro (2013) and we repeat verbatim his important conclusion “It is recommended that future humanitarian deployments of crowdsourcing focus on information processing within the populations they serve, engaging those with crucial local knowledge wherever they happen to be in the world.”

Notes

1. <https://www.merriam-webster.com/dictionary/credibility>
2. <https://brightplanet.com/2013/06/twitter-firehose-vs-twitter-api-whats-the-difference-and-why-should-you-care/>

Acknowledgments

The first author of this paper would like to thank her colleagues at the University of Zurich Ms. Flurina Wartmann and Mr. Oliver Burkhard for helping at various stages of analyzing the data and results.

Notes on contributors

Kiran Zahra is a PhD student in department of Geography at the University of Zurich. Her research interests are Twitter, natural disaster management, geographic information retrieval, and data quality.

Frank O. Ostermann joined the ITC in 2014 as an assistant professor for Cloud and Crowd geo-information processing. His main research interests are the opportunities and challenges of distributed collection, storage, processing and sharing of crowd-sourced and volunteered geographic information. Since 2009, he holds a PhD (Dr. Sc. Nat.) in Geographic Information Science from the University of Zurich. Previous work includes three years as a post-doctoral researcher at the Joint Research Center of the European Commission, as well as several years as a research assistant at the University of Zurich and Hamburg on several EU-funded projects on user-generated geographic content, and spatio-temporal data analysis and visualization in urban contexts.

Ross S. Purves is concerned with methods to analyze and understand geography through a range of different approaches, including the analysis of user generated content and text more generally. He is based at the Department of Geography of the University of Zurich, where he leads the Geocomputation Group in the GIScience Centre.

References

Becker, H., M. Naaman, and L. Gravano. 2011. “Selecting Quality Twitter Content for Events.” Paper presented at

- The International Conference on Weblogs and Social Media, Barcelona, Spain, July 17–21.
- Benevenuto, F., G. Magno, T. Rodrigues, and V. Almeida. 2010. “Detecting Spammers on Twitter.” Paper presented at The Seventh annual Collaboration, Electronic messaging, AntiAbuse and Spam Conference, Redmond, WA, USA, July 13–14.
- Canini, K. R., B. Suh, and P. L. Pirolli. 2011. “Finding Credible Information Sources in Social Networks Based on Content and Social Structure.” Paper presented at The IEEE Third International Conference on Privacy, Security, Risk and Trust, Boston, MA, USA, October 9–11.
- Castillo, C., M. Mendoza, and B. Poblete. 2011. “Information Credibility on Twitter.” Paper presented at The 20th International Conference on World Wide Web, Hyderabad, India, March 28–April 1, 675–684.
- Castillo, C., M. Mendoza, and B. Poblete. 2012. “Predicting Information Credibility in Time-sensitive Social Media.” *Internet Research* 23 (5): 560–588. doi:10.1108/IntR-05-2012-0095.
- Goodchild, M. F., and J. A. Glennon. 2010. “Crowdsourcing Geographic Information for Disaster Response: A Research Frontier.” *International Journal of Digital Earth* 3 (3): 231–241.
- Graham, M., B. Hogan, R. K. Straumann, and A. Medhat. 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., and P. Kumaraguru. 2012. “Credibility Ranking of Tweets during High Impact Events.” Paper presented at The 1st Workshop on Privacy and Security in Online Social Media, Lyon, France, April 17. doi:10.1145/2185354.2185356.
- Hahmann, S., R. S. Purves, and D. Burghardt. 2014. “Twitter Location (Sometimes) Matters: Exploring the Relationship between Georeferenced Tweet Content and Nearby Feature Classes.” *Journal of Spatial Information Science* 2014 (9): 801–802.
- Haubrock, S., C. Little, S. McBride, and N. Balfour. 2017. “Keeping up with the Citizens-collecting Earthquake Observations in New Zealand.” Paper presented at AGILE Conference 2017, Wageningen, The Netherlands, May 9–12. <https://www.cs.nuim.ie/~pmooney/VGI>.
- Haworth, B., and E. Bruce. 2015. “A Review of Volunteered Geographic Information for Disaster Management.” *Geography Compass* 9 (5): 237–250.
- Hecht, B., L. Hong, B. Suh, and E. H. Chi. 2011. “Tweets from Justin Bieber’s Heart: The Dynamics of the Location Field in User Profiles.” Paper presented at The ACM Conference on Human Factors in Computing Systems (CHI 2011), Vancouver, BC, Canada, May 7–12, 237–246.
- Hossmann, T., P. Carta, D. Schatzmann, F. Legendre, P. Gunningberg, and C. Rohner. 2011. “Twitter in Disaster Mode: Security Architecture.” Paper presented at SWID ’11 – The Special Workshop on Internet and Disasters, Tokyo, Japan, December 6–9. doi:10.1145/2079360.2079367.
- Hughes, A. L., L. Palen, J. Sutton, S. B. Liu, and S. Vieweg. 2008. “Site-seeing in Disaster: An Examination of Online Social Convergence.” Paper presented at The 5th International ISCRAM Conference, Washington, DC, May 5–7.
- Imran, M., C. Castillo, J. Lucas, P. Meier, and S. Vieweg. 2014. “AIDR: Artificial Intelligence for Disaster Response.” Paper presented at The 23rd International Conference on World Wide Web, Seoul, April 7–11, 159–162. doi:10.1145/2567948.2577034.

- Kinsella, S., V. Murdock, and N. O'Hare. 2011. "I'm Eating a Sandwich in Glasgow': Modeling Locations with Tweets." Paper presented at The 3rd International Workshop on Search and Mining User-generated Contents, Glasgow, UK, October 24–28, 61–68.
- Li, R., K. H. Lei, R. Khadiwala, and C. C. Chang. 2012. "TEDAS: A Twitter-based Event Detection and Analysis System." Paper presented at The 28th IEEE International Conference on Data Engineering, Washington, DC, USA, April 1–5, 1273–1276. doi:10.1109/ICDE.2012.125.
- Middleton, S. E., L. Middleton, and S. Modafferi. 2014. "Real-time Crisis Mapping of Natural Disasters Using Social Media." *IEEE Intelligent Systems* 29 (2): 9–17. doi:10.1109/MIS.2013.126.
- Morris, M. R., S. Counts, A. Roseway, A. Hoff, and J. Schwarz. 2012. "Tweeting is Believing? Understanding Microblog Credibility Perceptions." Paper presented at The ACM 2012 Conference on Computer Supported Cooperative Work, Seattle, WA, USA, February 11–15. 441–450.
- Munro, R. 2013. "Crowdsourcing and the Crisis-affected Community." *Information Retrieval* 16 (2): 210–266.
- Murthy, D. 2011. "Twitter: Microphone for the Masses?" *Media Culture and Society* 33 (5): 779–789.
- O'Donovan, J., B. Kang, G. Meyer, T. Höllerer, and S. Adalii. 2012. "Credibility in Context: An Analysis of Feature Distributions in Twitter." Paper presented at 2012 ASE/IEEE International Conference on Privacy, Security, Risk and Trust, Washington, DC, USA, September 3–5, 293–301.
- Ostermann, F., and L. Spinsanti. 2012. "Context Analysis of Volunteered Geographic Information from Social Media Networks to Support Disaster Management: A Case Study on Forest Fires." *International Journal of Information Systems for Crisis Response and Management* 4 (4): 16–37.
- Ostermann, F. O., and L. Spinsanti. 2011. "A Conceptual Workflow for Automatically Assessing the Quality of Volunteered Geographic Information for Crisis Management." Paper presented at AGILE Conference 2011. Utrecht, The Netherlands, April 18–21.
- Ostermann, F. O., M. Tomko, and R. Purves. 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): 480–499.
- Sakaki, T., M. Okazaki, and Y. Matsuo. 2010. "Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors." Paper presented at The 19th international conference on World Wide Web, Raleigh, NC, USA, April 26–30, 851–860. doi:10.1145/1772690.1772777.
- Sakaki, T., M. Okazaki, and Y. Matsuo. 2013. "Tweet Analysis for Real-time Event Detection and Earthquake Reporting System Development." *IEEE Transactions on Knowledge and Data Engineering* 25 (4): 919–931. doi:10.1109/TKDE.2012.29.
- Sarda, P., and R. L. Chouhan. 2017. "Extracting Non-situational Information from Twitter During Disaster Events." *Journal of Cases on Information Technology* 19 (1): 15–23.
- Senaratne, H., A. Mobasheri, A. L. Ali, C. Capineri, and M. Haklay. 2017. "A Review of Volunteered Geographic Information Quality Assessment Methods." *International Journal of Geographical Information Science* 31 (1): 139–167.
- Spinsanti, L., and F. Ostermann. 2013. "Automated Geographic Context Analysis for Volunteered Information." *Applied Geography* 43: 36–44.
- Terpstra, T., A. D. Vries, R. Stronkman, and G. L. Paradies. 2012. "Towards a Realtime Twitter Analysis during Crises for Operational Crisis Management." Paper presented at The 9th International ISCRAM Conference, Vancouver, Canada, April 22–25.
- Truelove, M., M. Vasardani, and S. Winter. 2015. "Towards Credibility of Micro-blogs: Characterising Witness Accounts." *GeoJournal* 80 (3): 339–359.
- Verma, S., S. Vieweg, W. Corvey, L. Palen, J. H. Martin, M. Palmer, A. Schram, and K. M. Anderson. 2011. "NLP to the Rescue?: Extracting 'Situational Awareness' Tweets during Mass Emergency." Paper presented at The Fifth International AAAI Conference on Weblogs and Social Media, Barcelona, July 17–21.
- Wald, D. J., V. Quitarano, C. B. Worden, and M. Hopper. 2011. "USGS 'Did You Feel It?' Internet-based Macroseismic Intensity Maps." *Annals of Geophysics* 54 (6): 688–707.
- Wing, B. P., and J. Baldrige. 2011. "Simple Supervised Document Geolocation with Geodesic Grids." Paper presented at The Meeting of the Association for Computational Linguistics: Human Language Technologies, Portland, OR, USA, June, 19–24.
- Zahra, K., and R. Purves. 2017. "Analysing Tweets Describing during Natural Disasters in Europe and Asia." Paper presented at AGILE Conference 2017. Wageningen, The Netherlands, May 9–12. <https://www.cs.nuim.ie/~pmooney/VGI>.