Multilingual Analysis of Twitter News in Support of Mass Emergency Events

Andrea Zielinski
Fraunhofer IOSB
andrea.zielinski@iosb.fraunhofer.de

Ulrich Bügel
Fraunhofer IOSB
ulrich.buegel@iosb.fraunhofer.de

ABSTRACT

Social media are increasingly becoming a source for event-based early warning systems in the sense that they can help to detect natural disasters and support crisis management during or after disasters.

In this work-in-progress paper we study the problems of analyzing multilingual twitter feeds for emergency events. The present work focuses on English as "lingua franca" and on under-resourced Mediterranean languages in endangered zones, particularly Turkey, Greece, and Romania Generally, as local civil protection authorities and the population are likely to respond in their native language. We investigated ten earthquake events and defined four language-specific classifiers that can be used to detect earthquakes by filtering out irrelevant messages that do not relate to the event. The final goal is to extend this work to more Mediterranean languages and to classify and extract relevant information from tweets, translating the main keywords into English.

Preliminary results indicate that such a filter has the potential to confirm forecast parameters of tsunami affecting coastal areas where no tide gauges exist and could be integrated into seismographic sensor networks.

Keywords

Multilingual Event Detection, Cross-lingual Information Extraction, Social Sensor, Twitter.

INTRODUCTION

Emergency information processing of social media can contribute effectively to identify regions affected by natural hazards such as earthquakes or tsunami, given that the feeds are real-time and often contain location information (ca. 12% with exact coordinates; ca. 50% city or state derived from the user profile). Due to the massive growth of Twitter data and its increasing number of users, it is however, a challenge to access and interpret the stream of data efficiently. Within the last years, there have been major achievements to make use of such "weak" human sensors as a complement to seismic sensors in some early warning systems (see Sakaki, 2010, Guy, 2010), focusing on English and Japanese. At present, there is no similar alerting system for the Mediterranean region. We try to fill this gap within the European TRIDEC project (www.tridec-online.eu) by adapting state of the art algorithms to the common Twitter languages in the endangered zones.

Social media often play a crucial role in disaster management during and after the crisis: citizens generally use Twitter postings or SMS messages to report emergencies. In this case, the information contained in them might be relevant for crisis management (relief and medical care for those affected, repair of broken infrastructure), so that there is a strong need to classify, cluster and extract such information effectively from large-scale noisy and unstructured data. As the messages are very short (max. 140 characters), NLP analysis is particularly difficult.

A number of text mining tools have been applied to recognize tactical, actionable information in tweets (Verma, 2011), to find messages that contain real-world or real-event information (Becker, 2011; Naaman ,2011), or to extract Named Entities (Neubig, 2011) or other news content (Sankaranarayanan, 2009) for one single language (mostly Japanese or English).

In some cases, though, it is crucial to cross language boundaries. For instance, when the epicenter is near the border of a country (e.g., Western Turkey and Greece), or when a twitter user reports an event in his/her native language (e.g., Romanian) that needs to be translated into a different language (e.g. English, German, Spanish).

Therefore, our long-term goal within TRIDEC is to support the access to relevant information across languages, focusing on the translation of under-resourced Mediterranean languages like Turkish/Greek/Romanian into English.

The multilingual nature of the blogosphere has been a major hindrance during the Haitian earthquake, where reports ranged from Japanese, to English and Spanish. Caragea (2011)'s work is one of the few that deals with multilinguality, classifying either English or Spanish messages into one of 10 emergency classes.

For the same event, a Statistical Machine Translation (SMT) System has been built that supports crisis management for Haitian Creole (Lewis, 2010) and profits from crowdsourced human translations (Munro, 2010). From the author's experience, a "codebook" is established that would enable and support the development of an SMT system for other less-resourced languages in short time and which could be integrated into an instant messaging system like twitter – provided, however, that a minimum of domain-specific bilingual parallel or comparable corpora and/or terminologies are available (Lewis et al., 2011). The need for rapid translation to aid the relief efforts has also come up during the tsunami event in Japan (Neubig et al., 2011).

EMERGENCY EVENTS

The most recent earthquakes with a magnitude > 4 in the Mediterranean shown in Fig. 1 (See map offered by EMSC - European Mediterranean Seismological Centre) took place in regions for which few language resources exist. These are, in descending order, Greek, Hungarian, Bulgarian, Croatian, Bosnian, Albanian, Slovenian, Slovak, Turkish, Romanian, and Serbian (cf. http://matrix.statmt.org/resources/matrix?set=euro_all).

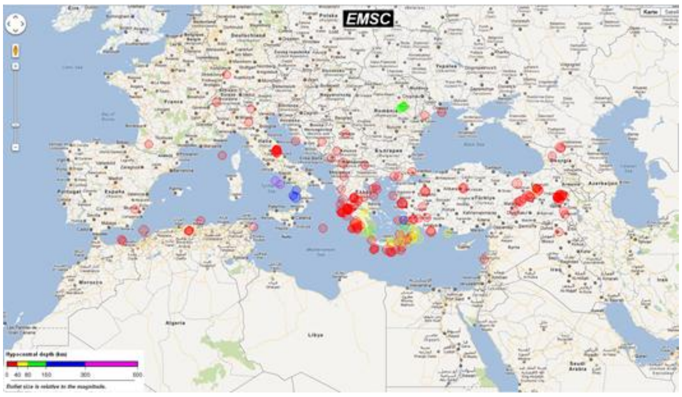

Figure 1. Recent earthquake events (2011) around the Mediterranean

Most earthquakes were felt across national borders. While some events caused severe damage, e.g. the earthquake of Van, Turkey, for others no damage or injuries were reported (e.g. Romania).

Although the seismic system in the Mediterranean is quite dense and, in general, earthquakes can be accurately located and quantified, in the case of a tsunami, the network of tide gauges is quite sparse, and a human-sensor-based system can help to confirm forecast parameters of a tsunami (estimated arrival time and wave height) affecting coastal areas where no tide gauges exist.

Moreover, in the case of earthquakes, it generally takes less time to confirm such an event (less than 1 min.) as opposed to the official verification by Earthquake or Tsunami Early Warning Systems (5-15 min.).

The EMSC has implemented a response system for people who experienced the earthquake. Results of the survey along with statistics show that in each country responses are given in a variety of languages.

TASK: EARTHQUAKE DETECTION IN THE MEDITERRANEAN

Goal:

Our goal is to test if it is possible to detect an earthquake by observing a rapid increase in twitter messages around the Mediterranean. We therefore investigate the number of tweets before and after such an event, trying to filter out spam messages automatically. Our hypothesis is that only features that pertain to the native language of the area where the earthquake is felt are good indicators, ignoring, e.g., relevant English messages.

Related Work:

In related studies (Sakaki et al., 2010) it has been shown that earthquakes can be detected reliably provided there are a large number of twitter users. The authors found out that the accuracy of prediction is increased by semantic analysis which tries to classify tweets that refer to an actual earthquake occurrence. This is achieved by devising a SVM-based classifier based on statistical and linguistic features with an F-measure of ca. 73%.

Experiment:

We implemented and tested a number of language-specific earthquake detection classifiers for this task.

As a first common filter prior to classification, we defined a keyword-based query using the twitter streaming API. The keywords are translation equivalents or synonyms of "earthquake", "earth" and "shake" in a variety of languages. This way, only a small portion of twitter messages, i.e. about 0.1% of the whole amount of daily tweets (e.g. 200 thousand of a total of 200 Million, in 2011), is delivered (cf. Fig. 2).

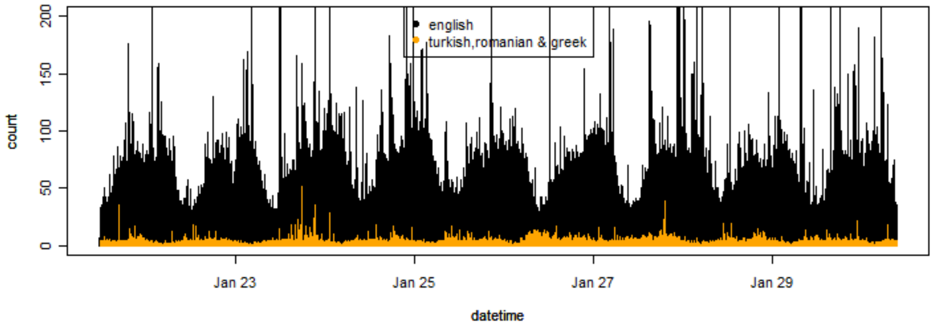

Figure 2. Throughput in tweets/minute

The average throughput of tweets per minute is different for each language. For the languages considered in our study we can observe that (in a normal case) the English language is predominant. English tweets have a share of 98% as opposed to Greek, Turkish and Romanian with less than 2%.

After the first filtering step, we applied a second filter (which is more dynamic and subject to modification in this preparatory phase) and then classified and analyzed the remaining data as follows:

- Eliminate re-tweets, tweets of certain @Users and #topics
- Classify tweets by language (e.g., "tr", "el", "en", "sl")
    - Apply language-specific tokenizers
    - Eliminate language-specific high-frequency words (stopword list)
    - Normalization of embedded links
    - Compute frequency of all remaining words

We reused existing language-specific tools whenever possible. A tokenizer has been implemented to parse non-formalized input, so that special symbols (e.g. emoticons) can be identified correctly. We experimented with two classification methods, a knowledge-engineering (KB) and a Machine Learning (ML) approach using

- Classification by regular expressions
- Classification by Naïve Bayes

Dataset:

For our training set, we harvested real-time tweets by establishing a continuous streaming connection to twitter between 09/01/2012 and 15/01/2012 with a list of multilingual keywords related to "earthquake". We also specified that those messages tagged with a geo-location should be in a bounding box that corresponds to the investigated areas. There were 10 earthquake events at a magnitude greater 4 during that time period (cf. ESCM) for which we gathered approx. 1000 tweets for training. By manually inspecting these tweets and classifying them as positive or negative, we found out that, on average ca. 55% of the messages were relevant (see Fig. 4). While in an empirical investigation on recent Japanese earthquakes, a strong correlation between the frequency of English and Japanese tweets could even be observed (Collier, 2011), this is different in our scenario, because we do not require that tweets are geolabeled to the restricted area only.

For our training set, we over sampled the "observed earthquake" event, in order to be able to handle the skewed class distribution (as the proportion of non-earthquakes events is extremely large).

Events	Positive Samples (Relevant)	Negative Samples
Earthquakes in Turkey	287 (tu)	153 (tu)
Earthquakes in Greece	152 (el)	230 (el)
Earthquakes in Romania	121 (ro)	117 (ro)

Figure 3. Training set of positive and negative tweets

A further set of more than 250 positively categorized messages from reported earthquake events from EMSC (http://www.emsc-csem.org/Earthquake/Testimonies/) has been added to our training set, as it is similar in style and register. Most testimonies have been assembled for English (80%), following Greek (10%), Romanian (10%) and Turkish (5%).

Methodology:

We assume that in the stream of posted tweets all items are independent of each other. We use a combination of two different binary classification methods which have different strengths and weaknesses. Both are trained and tested language-specifically so that the characteristics of each language are accounted for.

Classification by Regular Expressions

This classification method is based on expert knowledge which was acquired manually by inspecting the training data. It performs well, whenever a clear decision is possible. The sequence of words in the tweet is thereby preserved and can be exploited for analysis. For instance, certain lexemes like *σεισμός* (el) or *titremek* (tr) (engl. "shake") tend to be used in a variety of contexts that are not necessarily related to the topic "earthquake". In order to avoid too many false hits, it is required that the tweet matches a (filter of) regular expressions; e.g. *σεισμός* AND γη (engl. "earth"). In contrast to the first filter defined on the streaming API, we have a greater flexibility at this stage, as regular expressions of arbitrary complexity can be defined.

Classification by Naïve Bayes

We chose a robust NB classifier based on a simple bag of word model which can be computed in advance from our training data. In this model, word ordering is ignored and only the word occurrences are counted. Because of scarcity of data, we use a unigram model and refrain from building more complex n-grams at this stage. In contrast to decision trees (i.e., based on regular expressions), the strength of the classification method is that many "subtle" factors can be taken into account to make the classification. For instance, the fact that in most cases sentiment adjectives (e.g., *funny*), smiley icons or re-tweets tend to be classified as non-relevant.

Results:

For testing, we used a second set of twitter messages related to the following observed earthquakes. The result is best for the official language and worst for the English language (see Fig.4):

Event	**Country**	**Time and Location of Event**	**Accuracy (of classifiers)**			
			en	ro	el	tu
Observed Earthquake	..in Turkey	Date Time 2012-01-14 05:52:32.0 UTC Location40.08 N ; 38.35 E	31%	-	-	68%
	..in Greece	Date time 2012-01-09 09:39:29.0 UTC Location 36.40 N ; 25.51 E	54%	-	63%	-
	..in Romania	Date time 2012-01-07 05:43:25.0 UTC Location 45.70 N ; 21.19 E	27%	83%	-	-

Figure 4. Test result for different classifiers

While it is quite obvious that language-specific classifiers are needed for the task, it is an open question, if English twitter messages help to resolve this issue, or if they even blur the distinction (earthquake-related or not). When the aim is to detect an earthquake for a certain region, classification of English tweets is a more difficult task (referring to earthquakes around the whole world), so that such a classification is prone to error.

We also looked at the number of tweets classified as relevant in the time frame before and after the earthquake. In all three events, we could observe a peak in the stream after the specified time (until 6 hours later) but as the earthquakes were not felt that intensively (with the given magnitude of 3-4), the increase was not dramatic.

CONCLUSION

So far we defined three classifiers for Romanian, Greek and Turkish to label twitter messages as relevant or not, which is a prerequisite for the detection of earthquakes or a tsunami in the Mediterranean. Despite of the limited set of training data, the classifiers performed above random choice. In the course of the project, we will elaborate on our baseline classifiers and experiment on the impact of certain features (e.g. hashtags, re-tweets). We plan to integrate a component for identifying Named Entities which ground the event in time and space.

We also plan to use twitter analysis for Crisis Management. For this task, multiple-class classifiers are needed which can identify the topic of the tweet and extract specific information. To this end, a multilingual corpus of Twitter messages related to crises is being assembled, and domain-specific language resources like multilingual terminology lists or language-specific Natural Language Processing (NLP) tools are built up to cross the language barrier, particularly to support the automatic translation of tweets into English.

ACKNOWLEDGMENTS

This research work was funded by the European Commission's Seventh Framework Programme for Intelligent Information Management under contract TRIDEC IP FP7-258723. The authors acknowledge the continuing research collaboration with partners from the TRIDEC consortia, particularly IT Innovation.

REFERENCES

Becker, H. Naaman, M., Gravano, L. (2010) Learning similarity metrics for event identification in social media. In WSDM, 2010.

Caragea, C., McNeese N., Jaiswal, A., Traylor, G., Kim, H.-W., Mitra, P., Wu, D., Tapia, A.H., Giles, L., Jansen, B.J., Yen, J. (2011) Classifying Text Messages for the Haiti Earthquake. In *Proceedings of the 8th International ISCRAM Conference*, Lisbon, Portugal.

Collier,N., Doan, S., Vo, B.-K. H. (2011) An Analysis of Twitter message in the 2011 Tohuku Earthquake, *Proceedings of eHealth*, Malaga, Spain.

Guy, M, Earle, P., Ostrum, Ch., Gruchalla, K., Horvath, S. (2010) Integration and Dissemination of Citizen Reported and Seismically Derived Earthquake Information via Social Network Technologies (2010). In *Proceedings in Intelligent Data Analysis IX*, Tuscon, AZ, USA.

Lewis, W.D. (2011) Crisis MT: Developing a Cookbook for MT in Crisis Situations, *Proceedings of the Workshop on Statistical Machine Translation*, Edinburgh, Scotland.

Munro, R. (2010) Crowdsourced Translation for emergency response in Haiti: the global collaboration of local knowledge, *AMTA Workshop on Collaborative Crowdsourcing for Translation*, Denver, Colorado.

Naanman, M., Becker, H. Gravano, L. (2011) Beyond trending topics: Real-world event identification on twitter. In *ICWSM*, 2011.

Neubig, G., Matsubayayashi, Y., Hagiwara, M., Marukami, K. (2011) Safety Information Mining – What can NLP do in a disaster. In *Proceedings of the 5th IJCNLP*. Chiang Mai, Thailand.

Sakaki, T., Toriumi, F., and Matsuo, Y. (2011) Tweet Trend Analysis in an Emergency Situation, *Proceedings of the Special Workshop on Internet and Disasters*, ACM, New York, NY, USA.

Sakaki, T., Okazaki, M., and Matsuo, Y. (2010) Earthquake shakes Twitter users: real-time event detection by social sensors. In *Proc. WWW 2010,* ACM Press (2010), 851-860.

Sankaranarayanan, J., Samet, H., Teitler, B.E., Lieberman, M.D. and Sperling, J. (2009) TwitterStand: news in tweets. In Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in GeographicInformation Systems, pages 42-51, Seattle, Washington, USA, November

Verman, S., Vieweg,S., Corvey, W.J., Palen, L, Martin, J.H., Palmer, M., Schram, A., Anderson, K.M. (2011), Natural Language Processing to the Rescue?: Extracting "Situational Arwareness" Tweets during Mass Emergency. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, *7th-21 July 2011,* Barcelona, Spain