

EARS (Earthquake Alert and Report System): a Real Time Decision Support System for Earthquake Crisis Management

Marco Avvenuti
Department of Information
Engineering
University of Pisa, Pisa, Italy
m.avvenuti@iet.unipi.it

Stefano Cresci
Institute for Informatics and
Telematics (IIT)
National Research Council
(CNR), Pisa, Italy
stefano.cresci@iit.cnr.it

Andrea Marchetti
Institute for Informatics and
Telematics (IIT)
National Research Council
(CNR), Pisa, Italy
andrea.marchetti@iit.cnr.it

Carlo Meletti
National Institute of
Geophysics and Volcanology
Pisa, Italy
carlo.meletti@pi.ingv.it

Maurizio Tesconi
Institute for Informatics and
Telematics (IIT)
National Research Council
(CNR), Pisa, Italy
maurizio.tesconi@iit.cnr.it

ABSTRACT

Social sensing is based on the idea that communities or groups of people can provide a set of information similar to those obtainable from a sensor network. Emergency management is a candidate field of application for social sensing. In this work we describe the design, implementation and deployment of a decision support system for the detection and the damage assessment of earthquakes in Italy. Our system exploits the messages shared in real-time on Twitter, one of the most popular social networks in the world. Data mining and natural language processing techniques are employed to select meaningful and comprehensive sets of tweets. We then apply a burst detection algorithm in order to promptly identify outbreaking seismic events. Detected events are automatically broadcasted by our system via a dedicated Twitter account and by email notifications. In addition, we mine the content of the messages associated to an event to discover knowledge on its consequences. Finally we compare our results with official data provided by the National Institute of Geophysics and Volcanology (INGV), the authority responsible for monitoring seismic events in Italy. The INGV network detects shaking levels produced by the earthquake, but can only model the damage scenario by using empirical relationships. This scenario can be greatly improved with direct information site by site. Results show that the system has a great ability to detect events of a magnitude in the region of 3.5, with relatively low occurrences of false positives. Earthquake detection mostly occurs within seconds of the event and far earlier than the notifications shared by

INGV or by other official channels. Thus, we are able to alert interested parties promptly. Information discovered by our system can be extremely useful to all the government agencies interested in mitigating the impact of earthquakes, as well as the news agencies looking for fresh information to publish.

Categories and Subject Descriptors

H.2.8 [Database Management]: Database Applications—*Data mining*; K.4.1 [Computers and Society]: Public Policy Issues—*Human safety*

General Terms

Design, Evaluation, Reliability

Keywords

Social sensing, social mining, decision support system, disaster management, event detection

1. INTRODUCTION

Social Media (SM) is the most effective, sophisticated and powerful way to gather preferences, tastes and activities of groups of users in the context of Web 2.0 [23]. This amount of information generates an in-depth knowledge of one or more specific issues¹. Therefore, SM users could be regarded as social sensors, namely as a source of information about situations and facts related to the users (e.g., their preferences or experiences) and their social environment [31]. Emergency Management is a promising field of application for Social Sensing since it is possible to exploit the content shared on SM to gather up to date information on emerging situations of potential danger. These techniques allow for the acquisition of greater situational awareness which can be used to promptly alert interested parties. In the case of an emergency or a disaster, critical information needs to

¹<http://beautifuldata.net/2013/01/social-sensors/>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or to publish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
KDD'14, August 24–27, 2014, New York, NY, USA.
Copyright 2014 ACM 978-1-4503-2956-9/14/08 ...\$15.00.
<http://dx.doi.org/10.1145/2623330.2623358>.

be urgently collected so as to speed up decision making by emergency services, advising affected people, and notifying government authorities. We refer to Early Warning Systems (EWS) as those information systems able to timely detect occurring events of social concern and thus able to deliver appropriate warnings [27]. By “event” we mean a potentially dangerous situation that takes place in a finite period of time.

We aim to use spontaneous reports from users on SM as our source of information. Our focus is on reports about events causing social and safety concerns, such as natural disasters. This study is based on Social Networks (SN), a specific type of SM particularly suitable for hosting this kind of analysis. The reason for this choice is due to the large number of users involved in SN and their high level of interaction. The advantage of exploiting SN compared to traditional methods of investigation lies in the spontaneous participation of the users, in that their contribution is made without pressure or influence from others. Twitter is particularly suitable as a source of data for social sensing platforms. This is due to the fact that its users generally talk about what they are doing, and therefore what is happening around them. Analysing messages shared by Twitter users is a good way to quickly obtain details of the impact of the event.

In this paper, we will focus on reports related to earthquakes in Italy. In the Euro-Mediterranean region Italy is one of the countries with the highest seismic hazard together with Greece and Turkey. In Italy, the relationship between the damage caused by earthquakes and the energy released is much higher than in other countries with high seismicity, such as California or Japan. This is mainly due to the high population density and the considerable fragility of Italian artistic and monumental heritage². As suggested in [24], the highly seismic nature of Italy makes it one of the best countries where to carry out a study on the detection of earthquakes based on Twitter data. In addition, the growing use of Twitter during the last few years seems to favour a successful study (in 2013 active users increased by 40% in the world and 50% in Italy [25]).

The National Institute of Geophysics and Volcanology (INGV) is the reference scientific institution of the Italian Civil Protection for seismic and volcanic risks, as it has overall responsibility for seismic and volcanic monitoring in Italy. On the basis of specific agreements, it is in charge of detecting and communicating the occurrence of earthquakes. In addition, a protocol exists which indicates which actions must be carried out by INGV at 2, 5, 10 and 30 minute intervals after the earthquake according to the magnitude of the event. In the case of an earthquake with magnitude equal to or greater than 4, one of the products to be released is the so-called ShakeMap. A ShakeMap is an estimation of the shaking level distribution in the epicentral area in terms of peak ground-motion parameters and of instrumentally derived intensities. The application of a ShakeMap is the identification of the area where damage is probable in order to define where to concentrate the rescue teams and organize a prompt emergency response. The model is constrained with real observations coming from seismic stations connected in real-time. The limited number of stations in Italy does not allow to obtain an accurate indication of the real shaking.

²<http://www.protezionecivile.gov.it/jcms/en/homepage.wp>

As a result, the opportunity to check whether damage has in fact occurred is limited. The integration of seismic station data with information discovered from social sensors will increase the number of “stations” and the reliability of a ShakeMap. This could help the Civil Protection to better evaluate the type and the degree of assistance required in the earthquake-stricken area.

The system described in this paper is aimed at providing a valuable decision support tool for INGV researchers and analysts. Our objective is to leverage experiences described in previous works in order to design and implement a system for the detection, alerting and assessment of the consequences of earthquakes. We plan to integrate the EARS (Earthquake Alert and Report System) platform with the other tools that INGV already exploits such as the ShakeMaps. Information discovered with this system can be used to acquire a greater and more fine-grained context awareness during earthquake emergencies. It can help model damage scenarios with fresh data automatically collected site by site and exploit on-the-ground community knowledge which would be otherwise inaccessible.

The remainder of this paper is organized as follows: Section 2 describes the related work. Section 3 details our design and implementation choices, Section 4 describes the frontend application. Section 5 describes the results of our study and Section 6 draws the conclusions and describes future work.

2. RELATED WORK

Several initiatives, both in scientific and in application environments, have been developed with the aim of exploiting information available on SM. Anyway despite the interest in the fields of crisis and emergency management, to date the task of the automatic assessment of the consequences of disasters has not been studied yet. To the best of our knowledge previous works have only focused on the event detection and information dissemination tasks. In literature descriptive and general approaches are opposed to practical, sector-based experiences.

2.1 Descriptive approaches

In [3] we highlighted the lack of a comprehensive approach to the problem of emergency management on SM. We described an architectural approach, we designed the components necessary for an event detection system and finally we evaluated the feasibility of the proposed solution. In [3] we did not address implementation and deployment challenges but the promising results we have achieved paved the way for the design and deployment of the EARS system which is discussed in this paper.

Researchers in [24] and [25] had the goal of creating an early warning system (EWS) for the real-time detection of earthquakes and tornadoes in Japan based on Bayesian statistics. The proposed system was able to timely detect 67.9% (53 of 78) of the earthquakes with JMA (Japan Meteorological Agency) scale 2 or more which occurred in two months. In these works data acquisition is performed via the Twitter Search API³ which accesses only a portion of all the tweets produced. This limitation can be negligible for large scale events but can impair event detection for events felt by a small number of social sensors. As described in Sec-

³<https://dev.twitter.com/docs/api/1.1/get/search/tweets>

tion 3 our approach to the data acquisition exploits the Twitter Streaming API⁴ which delivers all the tweets produced worldwide that match the search criteria. Moreover the system described in [24] and [25] only focuses on the event detection task.

In [9] is described another system for the detection of earthquakes based solely on Twitter data. Again data acquisition is performed via the Twitter Search API and has the same limitations previously described. The system is evaluated with different settings according to the sensitivity of the event detection module. However, even with its best configuration the system could only detect 48 globally distributed earthquakes out of the 5,175 earthquakes reported during the same time window by the USGS.

The SMART-C project [1,2] describes a high level, multi-modal framework for emergency detection and alert dissemination. The proposed solution is capable of collecting and integrating data from different sources such as SN, blogs, telephone land line communications, SMS, MMS. The project aims to improve two-way communications between the emergency response personnel and the population. Unfortunately the work presented in [1,2] mainly focuses on architectural and privacy issues without dealing with the implementation and deployment of the proposed solution.

Other works related to the emergency management have studied communication patterns and information diffusion in SN in the aftermath of disasters. The study described in [7] shows how SN can be used as a reliable source of spatio-temporal information. Researchers investigate Twitter activity during a major forest fire in the south of France in July 2009. Other similar studies have been carried out in [8,14,18] showing the importance of SN in the communications after a disaster. These studies encourage the exploitation of this information and motivate the development of systems such as the one that we are proposing.

As a generic approach, we can mention the European initiative Alert4All⁵. This project aims to create a framework to improve the effectiveness of warning messages and communications with the population in case of disasters at pan-European level. Again the focus is on the role of SM in the emergency communications [15,19].

2.2 Practical experiences

Together with the scientific studies previously described, in the last few years there has been an increasing number of applications encouraging participatory sensing in the fields of urban management and personal safety. These applications are mainly developed for mobile devices and allow users to share concise reports of civilian issues. Such tools generally perform simple tasks, lack a solid scientific background and don't employ techniques of information analysis. While a small number of these applications have become widely used in some cities or regions, the vast majority never managed to attract a significant user base. Moreover information shared on these tools is fragmented among the various applications and cannot be exploited to acquire full knowledge about the reported issues.

One of the most interesting local initiatives is represented by Emergenza24⁶, the experimental version of the italian

“Social Network for Emergency Management”. This platform exploits a dedicated Twitter account to gather spontaneous reports of emergencies in Italy. Although reports directed to Emergenza24 are fairly common, it is clearly stated in the official website that only messages with a specific syntax are automatically captured. In Section 3 we highlight how the vast majority of emergency reports do not follow any specific format or syntax and often present grammatical mistakes or slang words. We believe that this poses a serious limitation to the effectiveness of the initiative. Similarly, the italian SMEM platform (Social Media Emergency Manager)⁷ tried to promote the use of the `#smem` hashtag to report emergencies or other social issues. This initiative did not receive much interest by Twitter users and was aborted in 2012.

Authors of [5,29] describe a system to improve situational awareness during emergencies in Australia. The proposed system is interesting, but [5,29] lacks an in-depth explanation of the methodologies adopted and any kind of system evaluation. The SHIELD system [26] exploits wi-fi and bluetooth technologies to track the frequency and duration of encounters between users of the application. SHIELD exploits this information to infer the trust level between users. The system automatically selects a set of users to call with the aim of reducing the response and rescue times for victims of micro-criminality in the U.S. university campuses. Although proximity-based applications can be very effective for small scale events, this approach is hardly applicable to the field of earthquake emergency management.

3. DESIGN AND IMPLEMENTATION

In this section, we discuss the EARS system as shown in Figure 1. The system's architecture is derived from the studies we carried out in [3]. Here we turn our attention to the implementation and deployment issues and we discuss decisions and trade-offs made when making design choices for the EARS system. The main components of the EARS backend system are described in the remainder of this section. The web application frontend is discussed in Section 4.

3.1 Data Acquisition

The role of this module is very important because the system only operates on data collected in this phase. Errors at this stage, especially regarding the loss of data, have to be minimized since they will propagate throughout the system thus impairing the ability to detect events. The data acquisition module must meet both data completeness and data specificity requirements. This clearly represents a trade-off between the number of messages gathered and their relevance to the earthquake detection and damage assessment tasks. Although this module does not perform any analysis on collected data, it is possible to overcome this trade-off by fine-tuning the keywords used to collect data.

In order to select the best set of keywords we started monitoring terms reported in the literature [24], [9], [3] together with other words related to earthquakes in the italian language. We progressively restricted the initial set of 9 keywords by eliminating the ones that did not show a correlation between their frequency of usage and the seismic events reported by INGV. We discarded those keywords, such as “crollo” (wreckage) and “crepa” (crack), specifically related

⁷<http://www.socialmediaemergencymanager.com>

⁴<https://dev.twitter.com/docs/streaming-apis>

⁵<http://www.alert4all.eu/>

⁶<http://www.emergenza24.org/>

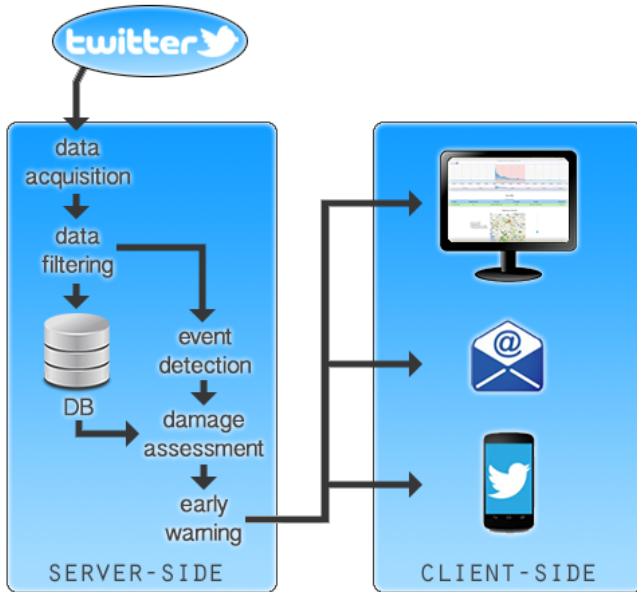


Figure 1: System architecture

to the damage assessment task; those keywords, such as “sisma” (seism) and “magnitudo” (magnitude), that are often used in official communications rather than in spontaneous user reports; and those keywords, such as “trema” (shakes) and “tremando” (shaking), that are too generic and therefore not specifically related to earthquakes. The keyword fine-tuning process lasted roughly two months. At the end of the process we concluded that the most selective italian words to collect tweets about ongoing earthquakes are “terremoto” (earthquake) and “scossa” (tremor). It is worth noticing that our set of keywords for earthquake detection differs from the sets used in [24] and [9]. Even though the keyword selection process in previous works is not thoroughly detailed, we can conclude that the best keywords for the event detection task are language specific. We believe that this poses a limitation to the possibility to easily redeploy this kind of systems to other regions of the world.

Among the methods provided by Twitter for information extraction, the EARS system exploits the Streaming API to open a persistent connection with a stream of tweets. By using this connection, new tweets containing the selected keywords can be collected. In contrast with the Search API used in the studies described in [24] and [9], which gives access only to a subset of all the tweets produced, the Streaming API potentially makes it possible to capture all the tweets matching the search criteria.

A limitation can arise considering that, as the connection can potentially access the entire flow of tweets produced in the world, Twitter delivers at most 1% of the total traffic and automatically cuts off the excess. However, our system never suffered from such a limitation over the seven month period it was used. To guarantee the robustness and the reliability of the system we also implemented additional mechanisms that manage rate-limit and generic connection problems in the use of the APIs.

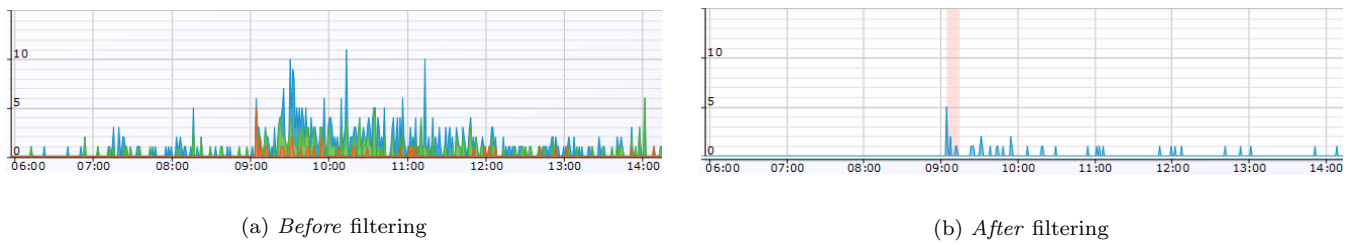
3.2 Data filtering

Using keywords to query Twitter makes it possible to gather messages potentially related to an event. However, not all the messages gathered in this process relate to an ongoing earthquake. Some messages can be misleading for the event detection module and must be filtered out as noise [9]. By noise we refer to the messages containing the query keywords but which are not related to the type of event to detect. In [3] we have identified two different sources of noise: (i) messages in which the keyword is used with a different meaning from the one related to the searched event and (ii) messages in which the keyword refers to a past event. Excessive levels of noise in collected messages lead to false detections by the system. However filtering too much may result in the loss of useful messages and thus in the impossibility to detect important events. Therefore this task presents another crucial trade-off related to the accuracy of the filtering process. In EARS we overcame this trade-off by employing data mining techniques.

Our system performs data filtering by cleaning data in 2 steps. A pre-filtering phase applies raw rules to discard tweets that clearly do not refer to an ongoing seismic event. Tweets not discarded in the pre-filtering phase are then analyzed with data mining techniques to perform a more fine-grained selection. Studying the characteristics of the messages shared on Twitter in the aftermath of seismic events lead us to the observation that genuine reports of earthquakes do not follow any information diffusion model and are not influenced by other reports. This scenario rapidly evolves over time as the news of the earthquake spreads over the different medias and subsequent reports are in growing percentage influenced by other news. We concluded that the best results for the event detection task could be achieved by considering only spontaneous and independent messages. Following these guidelines we created a set of rules to apply during the pre-filtering phase. This means that EARS discards retweet messages, reply messages and messages shared by accounts belonging to a blacklist of 345 Twitter profiles that periodically publish information about past seismic events.

During the implementation and deployment of the system we also faced a challenge because the keyword “terremoto” means earthquake not only in italian but also in spanish and portuguese. Twitter natively provides a language detection mechanism that EARS exploits. Unfortunately Twitter language detection follows a best effort approach and proved to be highly inaccurate [12]. This initially lead our system to produce false detections based on messages in spanish and portuguese which contained the word “terremoto”. To mitigate this issue in EARS we employ natural language processing techniques for the language detection and we discard those messages that do not appear to be in italian language.

Another possible flaw for all social mining systems lies in the vulnerability to intentional attacks performed by malicious users. Security concerns can arise if groups of people collude to generate fictitious tweets referring to an earthquake. In [3] we discussed the problem and proposed a solution based on fake account detection in order to mitigate the impact of such attacks. In the EARS system we exploit a classifier trained to distinguish between fake and real accounts [6]. The classifier has been trained on a set of 3900 equally distributed fake and real accounts and was able to correctly classify more than 95% of the accounts of the



(a) *Before filtering* (b) *After filtering*
 Figure 2: Impact of the filtering phases on the number of analyzed tweets.

training set. We bought around 3000 fake accounts from 3 different sellers and randomly chose 1950 of them to be part of the training set. Similarly we created a set of 1950 verified real accounts by asking volunteers to fill a captcha or by manually checking them. The trained model is passed to the Weka tool [13] every time a new message is collected so as to infer the class (fake, real) of the user who posted it. Messages posted by fake users are automatically discarded by the system. In addition users repeatedly triggering false detections are added to the account blacklist. To further protect the system from harmful attacks we only consider for the event detection task a maximum of one message per user. In addition, messages which contain exactly the same text are only taken into consideration once. While we understand that these solutions do not fully solve the problem of malicious attacks, we are confident that our efforts represent a first response to security concerns in social mining systems. As a consequence of the adopted solutions potential attackers are required to put more effort into the creation of plausible accounts.

After the pre-filtering phase, a more sophisticated filtering process is performed by a classifier that infers the class of a tweet starting from a trained model. Again we exploited Weka to train and generate our classifier. During the offline training phase the classifier has been trained using two distinct sets of messages: tweets related and tweets not related to a seismic event in progress. Tweets of the training set were manually classified using an ad-hoc interface. Our analysis on the messages reporting earthquakes has highlighted a few interesting characteristics that help distinguish between tweets related and tweets not related to an ongoing seismic event. Tweets referring to an earthquake are generally very short, they present fewer punctuation than normal tweets and often contain slang or offensive words. This is because people reporting an earthquake are usually scared about the event and the content of the messages they write tend to represent this situation. Instead, tweets referring to official news of an earthquake or talking about a past earthquake present a longer, more structured message. Tweets not related to a recent earthquake also include a higher number of mentions and URLs than spontaneous earthquake reports. Thus we defined the following set of features that takes into account the results of the previous analysis:

- character count;
- word count;
- punctuation count;
- URL count;
- mention count;
- slang/offensive word count;

Training the classifier with this set of features produced correct classifications in more than 90% of the tweets of the

training set. The classifier was obtained using the decision tree J48, corresponding to the Java implementation of the C4.5 algorithm [22] with a 10-fold cross validation [28]. The prediction is performed at run-time by invoking the classifier every time a message passes the pre-filtering phase. As Weka generally needs less than a second to predict the class of a new tweet, it is feasible to use the fine-grained classifier filter in the EARS real-time system.

Figure 2 shows the comparison between the numbers (y axis) of all collected tweets (Figure 2a) and the tweets that passed the filtering phase (Figure 2b) during a seismic event of low magnitude. The earthquake occurred in Modena, Italy, September the 4th, 2013 at 09.03 a.m. The blue plot is related to all the collected tweets. The green plot shows the tweets that passed the pre-filtering phase and, finally, the red plot represents the tweets after the filtering phase with the classifier. As shown in Figure 2 without an accurate filtering process it would have been difficult to notice the peak in the messages related to users reporting the earthquake. In fact reports for small scale events are usually overwhelmed by the number of noisy messages. In our experience the filtering process eliminated 88% of the collected messages.

3.3 Event Detection

The detection of an event is triggered by an exceptional growth in the frequency of the messages that have passed the filtering phase. The better the filtering phase the easier is the task of event detection. Event detection in social media is a topic that has been widely studied in literature. To accomplish this task researchers in [24] exploit a temporal model based on bayesian statistics, the use of peak detection algorithms is investigated in [17]. The work discussed in [21] introduces the concept of Corrected Conditional Entropy (CCE) and shows how it can be exploited to detect irregularity in time series.

In our system we adopt a novel approach based on a burst detection algorithm. A burst is defined as a large number of occurrences of a phenomenon within a short time window [30]. Burst detection techniques are commonly applied to various fields such as the detection of topics in data streams. Our system triggers the detection of a seismic event when it identifies a burst of messages referring to an earthquake. Figure 4 shows the arrival times of messages during an earthquake occurred at 00:09 a.m, August 27, 2013 in Umbria and Marche regional districts. After T1, the occurrence time of the earthquake, a big burst of tweets was recorded by our system.

Works in [10], [32] and [16] discuss various burst detection algorithms. In EARS we chose to employ a simplified version of the hierarchical algorithm proposed in [10] since it is computationally light and can adapt well to both big and small bursts. An efficient algorithm is necessary because of

2014-01-20 08:21:16	2.6	CAMPOBASSO,CASERTA,BENEVENTO	10.8Km	2014-01-20 08:38:02	16m, 46s
2014-01-20 08:12:40	4.2	CASERTA,BENEVENTO	11.1Km	2014-01-20 08:36:32	23m, 52s
2014-01-20 07:04:43	2.1	VIBO VALENTIA	18.9Km	2014-01-20 07:50:33	45m, 50s
2014-01-20 02:11:53	2.5	MASSA,LUCCA	5.3Km	2014-01-20 02:35:02	23m, 9s

2014-01-20 08:13:35	22m, 57s	97%	Terremoti	2393
2014-01-20 02:25:16	-	26%	Terremoti	3
2014-01-20 02:13:12	21m, 50s	26%	Terremoti	9

Figure 3: Chronological summary of the events

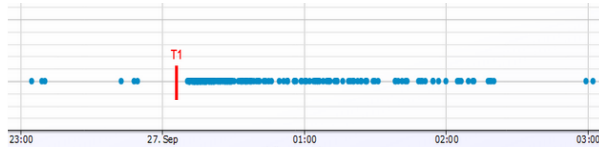


Figure 4: A burst of tweets registered after an earthquake (occurred at T1)

the real-time nature of our system and the ability to detect both big and small bursts fits well with the need to identify large scale and small scale events.

The detection of a burst is based on the calculation of the frequency (C_{freq}) of messages in a short-term time window. Assuming $L_{SW} = \text{Length of the short-term time window}$ and $N_{SW} = \text{Messages in the last short-term time window}$, then:

$$C_{freq} = \frac{N_{SW}}{L_{SW}}$$

A burst is detected when such frequency exceeds a given threshold Th . The threshold to trigger a burst depends on a reference frequency (R_{freq}) calculated over a long-term time window. Assuming $L_{LW} = \text{Length of the long-term time window}$ and $N_{LW} = \text{Messages in the last long-term time window}$, then:

$$R_{freq} = \frac{N_{LW}}{L_{LW}}$$

The size of both the short and long time windows along with the threshold represent implementation trade-offs and are closely linked to the characteristics of the events to detect. Shortening the short-term window makes the system respond faster since less time is required to detect a burst. However it also limits the possibility to detect events that receive a small number of reports. Widening the short-term window increases the number of events detected by the system since it is easier to reach the threshold over a longer period of time. Anyway it also exposes the system to more false detections and increases response times. Similarly, lowering the threshold allows for the detection of smaller scale events but also increases the risk of false detections.

During the fine tuning phase of the system we tried different combinations of settings. We achieved the best results with the following settings:

$$L_{SW} = 1 \text{ minute}$$

$$L_{LW} = 1 \text{ week}$$

$$Th = \max(3, 10 * R_{freq})$$

Setting a minimum threshold of only 3 messages per minute greatly increases the sensitivity of the system. As discussed in Section 5, this setting allowed EARS to detect 100% of the earthquakes with a magnitude of 3.5 or more reported by at least one user.

3.4 Damage Assessment

Damage assessment (DA) is the process that allows emergency management personnel to determine the impact and the consequences of an emergency on communities and infrastructures. Typically, the DA process takes place in the aftermath of an emergency and involves specialized personnel who assess the consequences by visiting the location of the event. EARS exploits information shared by people directly involved in the event to automatically perform this task without the need to transport personnel over the various potentially damaged locations.

Every new message collected after the detection of an earthquake is associated to the event and contributes to the creation of a corpus of reports. As the event unfolds, the system repeatedly analyzes the growing corpus of reports to extract valuable information from it. Among the information that EARS is capable of extracting are coordinates and toponyms of locations and emerging n-grams. Locations mentioned in the reports are most likely places struck by the seismic event. To accomplish this task we make use of TagMe, a service of text annotation and disambiguation developed at the University of Pisa [11]. This tool can be specifically set to work with tweets and outputs a list of tags with the description of the terms and other related information. We then extract the coordinates associated to the toponyms disambiguated by TagMe via SPARQL [20] queries to DBpedia [4]. To further expand the set of messages associated to an earthquake the system automatically opens a new streaming connection to Twitter servers. We exploit this adaptive connection to collect messages produced in or mentioning the possible locations struck by the earthquake. This augmented message corpus is used to perform the damage assessment task by mining both the textual descriptions and the multimedial content retrieved (eg: videos, photos).

4. WEB APPLICATION

EARS is a decision support tool designed for INGV researchers and analysts. Its web frontend enables temporal, geographical and content analyses both at event and message level.

Figure 3 contains a screenshot of the web application showing a chronologically ordered summary of events of interest. Events detected by EARS are listed on the right hand side of the screen while events reported by INGV are listed on the left hand side. This screen allows for a quick evaluation of the system behavior by highlighting confirmed and unconfirmed events. Specifically, a red background is associated to events reported by INGV but not detected by EARS, an orange background is associated to events detected by our system but not yet confirmed by INGV. Finally a green background is associated to events detected by EARS and confirmed by INGV. Every listed event, both confirmed or unconfirmed, can be expanded in another window showing

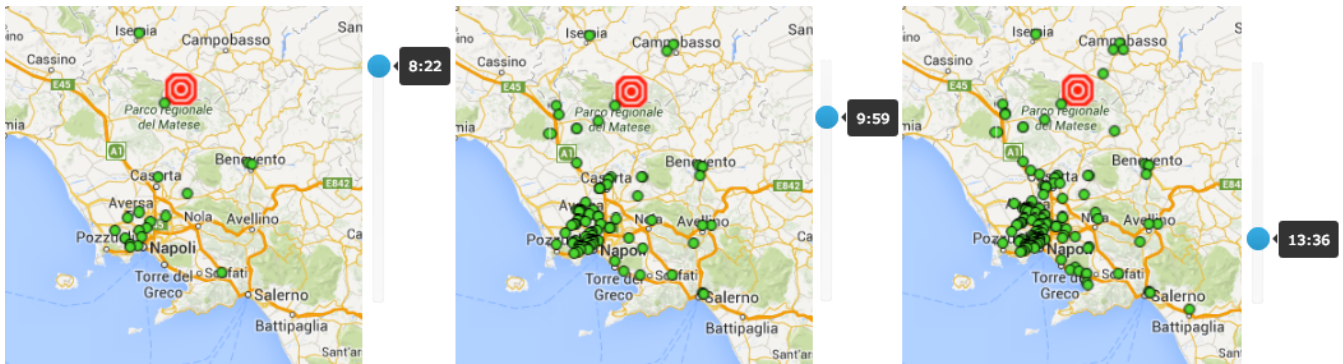


Figure 5: Geographical view: epicenter and spatial distribution of relevant messages

2014-01-20 08:30:28	Alex_Giordano: Ci mancava solo il terremoto di prima mattina	
2014-01-20 08:30:28	_manuhappy: No perché dovete sapere che io sto da giorni che tremo per il terremoto #vaffanculo	
2014-01-20 08:30:31	rinogargano: Se c'ero dormivo. Davvero. #terremoto	
2014-01-20 08:30:35	DavideRoccio: Di nuovo terremoto al sud. Confermate? #terremoto	
2014-01-20 08:30:39	ilfuzz: e con questa scossa, #gennarodauria si conferma il Sommo Vegente del nostro cuore	
2014-01-20 08:30:43	asollazzo98: PorcoDio il terremoto per la prima volta a scuola. Mi son messo paura ye	
2014-01-20 08:30:49	animaesiliata: Come fate a dormire ancora tutti??? Non vi ha svegliati il #terremoto??	
2014-01-20 08:30:53	laccio: mappa del #terremoto di magnitudo 4.6 http://t.co/Igw1Mo3VXS	

Figure 6: Messages view: chronological list of relevant messages

all the details associated to it.

The event details window is organized in three views:

- Temporal view;
- Geographical view;
- Messages view.

The temporal view, shown in Figure 7, contains a graph of the message frequency over time. The red colored area in the graph identifies the event duration as estimated by EARS. This is not the duration of the physical event but instead the time interval in which the system collects relevant messages for the event itself. This temporal view immediately indicates the number of relevant messages and their distribution over time.

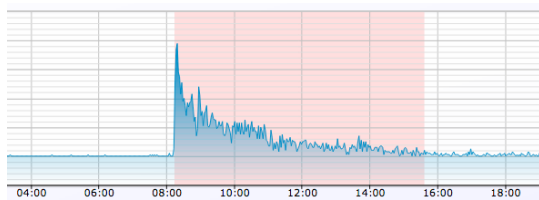


Figure 7: Temporal view: message frequency over time

The geographical view displays a map of the region affected by the event. The map contains a red marker for the epicenter of the event and a small green marker for each geocoded message. Geocoded messages include both messages carrying Twitter GPS data and messages geocoded by

EARS as explained in Section 3.4. Beside the map a time slider allows to reproduce the evolution of the event from its detection time to its end. Moving the slider's handle affects information displayed in the geographical and messages views. This component shows the geographical distribution of collected messages as the event evolved in time. It also lists the the messages collected during the "lifespan" of the event. Figure 5 shows an analysis of the spatio-temporal evolution of the earthquake that struck Caserta and Benevento the 20th January 2014.

The message view provides a chronological ordered list of the relevant messages collected by EARS. Every message is described by its time of creation, the author's username and its textual content. Multimedia content is automatically displayed inside the message box. In addition an icon highlights geocoded messages. Clicking on the icon highlights the selected message in the map provided by the geographical view. Viceversa clicking on a marker in the geographical view shows the content of the related message. Figure 6 shows an excerpt of the messages view. In addition to the chronological list of messages, this view contains a list of the most used n-grams and a list of the most used annotated tags. These two lists are extracted from the whole corpus of messages associated to an event.

5. TESTING AND RESULTS

Together with the EARS system we have developed a simulator designed to read and analyze data already collected by our real-time application. The advantage of using this simulator lies in the possibility to progressively tune the system and rapidly test it against already collected data.

<i>Magnitude</i>	<i>Occurred earthquakes</i>	<i>TP</i>	<i>FP</i>	<i>FN</i>	<i>Precision</i>	<i>Recall</i>	<i>F-Measure</i>
> 2.0	404	17	30	387	36.17%	4.21%	7.54%
> 2.5	102	16	30	86	34.78%	15.69%	21.62%
> 3.0	26	13	17	13	43.33%	50%	46.43%
> 3.5	11	9	3	2	75%	81.82%	78.26%
> 4.0	7	5	0	2	100%	71.43%	83.33%
> 4.5	2	2	0	0	100%	100%	100%

Table 1: System evaluation against all the earthquakes registered by INGV

This approach also makes it possible to evaluate our system with stronger metrics than the ones used in other similar works [24, 25], [9] and [5, 29].

The evaluation dataset consists of all the messages collected by EARS over a 70 days period. By using the simulator we were able to analyze the whole evaluation dataset in less than 3 hours. The simulator performs the same analyses as the EARS system and in addition outputs a report of the performance of the system. For the system evaluation we exploited official data published by the National Institute of Geophysics and Volcanology (INGV), the authority responsible for monitoring seismic events in Italy. INGV uses different channels, including Twitter, to distribute detailed information about seismic events having magnitude 2 or more, which have been detected by their equipment. We cross-checked the events detected by the simulator against the official reports provided by INGV. We classified earthquake detection results as in the following:

- True Positives (TP), events detected by the system and confirmed by INGV;
- False Positives (FP), events detected by the system, but not confirmed by INGV;
- False Negatives (FN), events reported by INGV but not detected by the system.

True Negatives are meaningless, as it would mean counting the number of earthquakes that did not happen and that our system did not detect.

We also computed the following evaluation metrics:

- Precision, ratio of correctly detected events among the total number of detected events:

$$Precision = \frac{TP}{TP + FP}$$

- Recall, ratio of correctly detected events among the total number of occurred events:

$$Recall = \frac{TP}{TP + FN}$$

- F-Measure, harmonic mean of Precision and Recall:

$$F-Measure = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Table 1 summarizes event detection evaluation against all the earthquakes registered by INGV from 2013-07-19 to 2013-09-23.

Results show that the detection of earthquakes with magnitude lower than 3 is very difficult. This is because the majority of these earthquakes are only detected by equipment and not by people. For events with a magnitude equal to or greater than 3.5, results show a good performance of the system in terms of F-Measure. This is especially significant given that seismic events of a magnitude around 3 are generally detected only by a very small number of people. In our study, for such earthquakes we always registered less than 10 tweets per event.

As the ultimate goal of the EARS system is to assess the consequences of earthquakes, light seismic events only detected by seismographs clearly do not pose any threat to communities and infrastructures. Earthquakes of interest are those actually felt by the population at large. Therefore we re-evaluated the system against those earthquakes that generated at least one report on Twitter. Results are displayed in Table 2 and show an overall improvement in the system performances. It is worth noticing that EARS achieves flawless results for earthquakes of magnitude 4.0 or more and also performs very well on earthquakes which have a magnitude in the region of 3.5.

6. CONCLUSIONS AND FUTURE WORK

EARS (Earthquake Alert and Report System) is a decision support environment for earthquake crisis management deployed for INGV researchers and analysts. The proposed system can clearly provide useful information on the consequences of seismic events. Such information is inferred from on-the-ground social sensors seconds after the detection of an earthquake. EARS can be integrated with other already established systems in order to help identify the areas where damage is probable. Information discovered by this system helps define where to concentrate the rescue teams and organize a prompt emergency response.

In Section 3 we discussed the design, implementation and deployment challenges faced during the development of EARS. These challenges are common to the majority of social mining systems even though they are not fully addressed in similar works. We proposed technical solutions for the most relevant issues which emerged from our work. Results suggest the effectiveness of such solutions although we are aware that further improvements may lead to even better performances. Among the issues we discussed is the problem of security. To the best of our knowledge EARS is the first emergency management system to employ security mechanisms in order to deter and minimize the effect of malicious attacks. Although we understand that these mechanisms are

<i>Magnitude</i>	<i>Tweeted earthquakes</i>	<i>TP</i>	<i>FP</i>	<i>FN</i>	<i>Precision</i>	<i>Recall</i>	<i>F-Measure</i>
> 2.0	128	17	30	111	36.17%	13.28%	19.43%
> 2.5	55	16	30	39	34.78%	29.09%	31.68%
> 3.0	21	13	17	8	43.33%	61.90%	50.98%
> 3.5	9	9	3	0	75%	100%	85.71%
> 4.0	5	5	0	0	100%	100%	100%
> 4.5	2	2	0	0	100%	100%	100%

Table 2: System evaluation against earthquakes that generated at least one report on Twitter

not sufficient on their own, we believe that they represent a good starting point towards addressing security issues of such applications. False detections are probably the most critical issue of EARS and require further investigation and effort. We manually verified that language detection problems have been the cause of several false detections by the system. These problems are primarily caused by Twitter’s best effort language detection approach and by the limited length of tweets. We adopted a solution based on natural language processing but we plan to exploit user account information as well to further improve the detection. Other problems contribute to the number of false detections registered, the vast majority of which is related to mistakes in the data filtering process.

Over the past seven months we have been developing and testing the EARS system against earthquakes reported by INGV. During this period of time none of the earthquakes that struck Italy have caused serious consequences on communities or infrastructures. The highest magnitude registered since July 2013 was 4.9 degrees when two earthquakes struck Ancona and Macerata on the 21st of July 2013 and Caserta and Benevento on the 29th of December 2013. Despite the lack of significant data reporting injuries to the population or damages to the infrastructures we have developed a system capable of inferring the consequences of earthquakes. However access to historical Twitter data, especially related to the earthquakes in Emilia Romagna in 2012 and L’Aquila in 2009, could prove crucial for future improvements in the damage assessment area. For this reason we are looking forward to the upcoming Twitter Data Grants program for research⁸.

In the future we will apply our system to other contexts such as wildfires, traffic jams, landslips and floods. We also plan to experiment with other sources of data like Facebook public posts and Google search activity data.

7. REFERENCES

- [1] N. Adam, J. Eledath, S. Mehrotra, and N. Venkatasubramanian. Social media alert and response to threats to citizens (smart-c). In *Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom), 2012 8th International Conference on*, pages 181–189. IEEE, 2012.
- [2] N. R. Adam, B. Shafiq, and R. Staffin. Spatial computing and social media in the context of disaster management. *Intelligent Systems, IEEE*, 27(6):90–96, 2012.
- [3] M. Avvenuti, S. Cresci, M. La Polla, A. Marchetti, and M. Tesconi. Earthquake emergency management by social sensing. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2014 IEEE International Conference on*. IEEE, 2014.
- [4] C. Bizer, J. Lehmann, G. Kobilarov, S. Auer, C. Becker, R. Cyganiak, and S. Hellmann. Dbpedia-a crystallization point for the web of data. *Web Semantics: Science, Services and Agents on the World Wide Web*, 7(3):154–165, 2009.
- [5] M. A. Cameron, R. Power, B. Robinson, and J. Yin. Emergency situation awareness from twitter for crisis management. In *Proceedings of the 21st international conference companion on World Wide Web*, pages 695–698. ACM, 2012.
- [6] S. Cresci, R. Di Pietro, M. Petrocchi, A. Spognardi, and M. Tesconi. A Criticism to Society (as seen by Twitter analytics). In *1st International Workshop on Big Data Analytics for Security*. IEEE, 2014.
- [7] B. De Longueville, R. S. Smith, and G. Luraschi. Omg, from here, i can see the flames!: a use case of mining location based social networks to acquire spatio-temporal data on forest fires. In *Proceedings of the 2009 International Workshop on Location Based Social Networks*, pages 73–80. ACM, 2009.
- [8] P. Earle, M. Guy, R. Buckmaster, C. Ostrum, S. Horvath, and A. Vaughan. Omg earthquake! can twitter improve earthquake response? *Seismological Research Letters*, 81(2):246–251, 2010.
- [9] P. S. Earle, D. C. Bowden, and M. Guy. Twitter earthquake detection: earthquake monitoring in a social world. *Annals of Geophysics*, 54(6), 2012.
- [10] R. Ebina, K. Nakamura, and S. Oyanagi. A real-time burst detection method. In *Tools with Artificial Intelligence (ICTAI), 2011 23rd IEEE International Conference on*, pages 1040–1046. IEEE, 2011.
- [11] P. Ferragina and U. Scaiella. Fast and accurate annotation of short texts with wikipedia pages. *Software, IEEE*, 29(1):70–75, 2012.
- [12] S. Hale, D. Gaffney, and M. Graham. Where in the world are you? geolocation and language identification in twitter. *Proceedings of ICWSM 2012*, pages 518–521, 2012.
- [13] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The weka data

⁸<https://engineering.twitter.com/research/data-grants>

- mining software: an update. *ACM SIGKDD Explorations Newsletter*, 11(1):10–18, 2009.
- [14] A. L. Hughes and L. Palen. Twitter adoption and use in mass convergence and emergency events. *International Journal of Emergency Management*, 6(3):248–260, 2009.
- [15] F. Johansson, J. Brynielsson, and M. N. Quijano. Estimating citizen alertness in crises using social media monitoring and analysis. In *Intelligence and Security Informatics Conference (EISIC), 2012 European*, pages 189–196. IEEE, 2012.
- [16] J. Kleinberg. Bursty and hierarchical structure in streams. *Data Mining and Knowledge Discovery*, 7(4):373–397, 2003.
- [17] A. Marcus, M. S. Bernstein, O. Badar, D. R. Karger, S. Madden, and R. C. Miller. Twitinfo: aggregating and visualizing microblogs for event exploration. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 227–236. ACM, 2011.
- [18] A. Murakami and T. Nasukawa. Tweeting about the tsunami?: mining twitter for information on the tohoku earthquake and tsunami. In *Proceedings of the 21st international conference companion on World Wide Web*, pages 709–710. ACM, 2012.
- [19] S. Nilsson, J. Brynielsson, M. Granåsen, C. Hellgren, S. Lindquist, M. Lundin, M. Narganes Quijano, and J. Trnka. Making use of new media for pan-european crisis communication. In *Proceedings of the Ninth International Conference on Information Systems for Crisis Response and Management (ISCRAM 2012)*. ISCRAM, 2012.
- [20] J. Pérez, M. Arenas, and C. Gutierrez. Semantics and complexity of sparql. *ACM Transactions on Database Systems (TODS)*, 34(3):16, 2009.
- [21] A. Porta, G. Baselli, D. Liberati, N. Montano, C. Cogliati, T. Gneccchi-Ruscione, A. Malliani, and S. Cerutti. Measuring regularity by means of a corrected conditional entropy in sympathetic outflow. *Biological cybernetics*, 78(1):71–78, 1998.
- [22] J. R. Quinlan. *C4.5: programs for machine learning*, volume 1. Morgan kaufmann, 1993.
- [23] A. Rosi, M. Mamei, F. Zambonelli, S. Dobson, G. Stevenson, and J. Ye. Social sensors and pervasive services: Approaches and perspectives. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2011 IEEE International Conference on*, pages 525–530. IEEE, 2011.
- [24] T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake shakes twitter users: real-time event detection by social sensors. In *Proceedings of the 19th international conference on World wide web*, pages 851–860. ACM, 2010.
- [25] T. Sakaki, M. Okazaki, and Y. Matsuo. Tweet analysis for real-time event detection and earthquake reporting system development. *Knowledge and Data Engineering, IEEE Transactions on*, 25(4):919–931, 2013.
- [26] G. S. Thakur, M. Sharma, and A. Helmy. Shield: Social sensing and help in emergency using mobile devices. In *Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE*, pages 1–5. IEEE, 2010.
- [27] N. Waidyanatha. Towards a typology of integrated functional early warning systems. *International Journal of Critical Infrastructures*, 6(1):31–51, 2010.
- [28] I. H. Witten and E. Frank. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2005.
- [29] J. Yin, A. Lampert, M. Cameron, B. Robinson, and R. Power. Using social media to enhance emergency situation awareness. *IEEE Intelligent Systems*, 27(6):52–59, 2012.
- [30] X. Zhang and D. Shasha. Better burst detection. In *Data Engineering, 2006. ICDE'06. Proceedings of the 22nd International Conference on*, pages 146–146. IEEE, 2006.
- [31] A. Zhou, W. Qian, and H. Ma. Social media data analysis for revealing collective behaviors. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '12*, pages 1402–1402, New York, NY, USA, 2012. ACM.
- [32] Y. Zhu and D. Shasha. Efficient elastic burst detection in data streams. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 336–345. ACM, 2003.