

Domain-Independent Detection of Emergency Situations Based on Social Activity Related to Geolocations

Hernan Sarmiento
Department of Computer Science
University of Chile
Santiago, Chile
hsarmien@dcc.uchile.cl

Barbara Poblete
Department of Computer Science
University of Chile
Santiago, Chile
bpoblete@dcc.uchile.cl

Jaime Campos
Department of Geophysics
University of Chile
Santiago, Chile
jaime@dgf.uchile.cl

ABSTRACT

In general, existing methods for automatically detecting emergency situations using Twitter rely on features based on domain-specific keywords found in messages. This type of keyword-based methods usually require training on domain-specific labeled data, using multiple languages, and for different types of events (e.g., earthquakes, floods, wildfires, etc.). In addition to being costly, these approaches may fail to detect previously unexpected situations, such as uncommon catastrophes or terrorist attacks. However, collective mentions of certain keywords are not the only type of self-organizing phenomena that may arise in social media when a real-world extreme situation occurs. Just as nearby physical sensors become activated when stimulated, localized citizen sensors (i.e., users) will also react in a similar manner. To leverage this information, we propose to use self-organized activity related to geolocations to identify emergency situations. We propose to detect such events by tracking the frequencies, and probability distributions of the interarrival time of the messages related to specific locations. Using an off-the-shelf classifier that is independent of domain-specific features, we study and describe emergency situations based solely on location-based features in messages. Our findings indicate that anomalies in location-related social media user activity indeed provide information for automatically detecting emergency situations independent of their domain.

CCS CONCEPTS

• **Information systems** → **Data stream mining**; *Spatial-temporal systems*;

KEYWORDS

Emergency Situations; Citizen Sensors; Social Media

ACM Reference Format:

Hernan Sarmiento, Barbara Poblete, and Jaime Campos. 2018. Domain-Independent Detection of Emergency Situations Based on Social Activity

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 the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

WebSci'18, May 27–30, 2018, Amsterdam, Netherlands

© 2018 Copyright held by the owner/author(s). Publication rights licensed to the Association for Computing Machinery.

ACM ISBN 978-1-4503-5563-6/18/05...\$15.00
<https://doi.org/10.1145/3201064.3201077>

Related to Geolocations. In *WebSci'18: 10th ACM Conference on Web Science, May 27–30, 2018, Amsterdam, Netherlands*. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3201064.3201077>

1 INTRODUCTION

Social media has become a major communication channel during high-impact real-world events, such as elections, sports events, emergency situations, among others. In each case, users act as “citizen sensors” sharing and posting their mood, opinion, photos, videos and geographical points of interest.

According to United Nations Department of Humanitarian Affairs [8], an emergency situation can be defined as “a sudden and usually unforeseen event that calls for immediate measures to minimize its adverse consequences”. During these events, traditional media may suffer infrastructure issues and real-time communications could be disrupted. Instead, microblogging has played a critical role over the last fifteen years allowing users to share real-time information from people local to the incident, such as status updates, casualties, damages and alerts [15, 19, 29, 30]. For this reason, researchers have studied user behavior during these events to detect, summarize and classify messages with the goal of helping authorities, and the general public, with situational awareness to provide fast and conscientious responses during crisis situations.

Twitter is a microblogging platform that allows users to share short messages (called *tweets*) and is currently used worldwide by over 300 million people.¹ About 80% of Twitter users access from mobile devices, which contributes to the immediacy of diffusion of information, especially during crisis situations [7]. One of the most important tasks during these situations, is to timely detect incoming real-world events. This is because, as with most social media conversations, messages are often overridden with irrelevant and redundant noise. In current works [5, 14, 19, 23], these tasks are solved with methods that rely on keyword based filters on the *Twitter Public Streaming API*.² However, a drawback of keyword-based methods is that they require to train for each specific domain, and for different type of events. For instance, Olteanu et al. [26] generated a set of keywords for event detection based on different datasets. However, these keywords are not sufficient for cases in which specific, and previously unseen, terms arise for particular events (e.g., the hashtags: *#eqnz*, used for Earthquakes in New Zealand, and *#pabloph*, used for Typhoon Pablo in Philippines) [3, 17, 28]. Furthermore, domain-specific keywords are commonly extracted using a particular language and do not apply to others.

¹<http://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/>

²<https://developer.twitter.com/en/docs/tweets/filter-realtime/overview>

In addition, prior work related to crisis situations in social media, show a strong relationship between the type of event (the *what*) and spatio-temporal dimensions (the *when* and *where*). For example, the top-trends during earthquakes are usually related to location mentions [25], and that there is a strong relationship between the proximity to a hurricane path and hurricane-related social media activity [18]. Other works have addressed the extraction of locations and points of interest during floods [21], as well as mixing geographic information system (GIS) information with geo-tagged messages to improve disaster mapping and real-time event tracking [12].

In order to complement prior findings, we propose to address the following research question: *Can we detect automatically emergency situations, based solely on location-related information found in user messages?*

We propose a method based on recurring references to country-level locations in message metadata. For this task, we create a gazetteer tree based on the hierarchy of the specific country, divide the messages into fixed time-windows and compute the frequency and the probability distributions of the interarrival time of the messages for each geographical hierarchy. To detect an emergency situation, we train a SVM classifier in each hierarchy and apply a geographic spread to filter false positives detection considering that an emergency situation can be either *focalized* (i.e., that affects a reduced area) or *diffused* (i.e., that affects a large area).

Our main contribution is to create a methodology that detects instantaneous emergency situations just by using the location related information for a specific country. **Since an emergency situation is a high-impact real-world event, we use locations in different levels of geographic hierarchies for detecting anomalies in the lowest levels (as cities or states) and the highest levels (a country) at the same time.** Furthermore, our approach does not depend on language to detect a new event since it does not use textual features as input. In addition, we characterize focalized and diffused crisis situations.

To validate our hypothesis, we create a ground truth based on random messages retrieving from Twitter Public Stream without specific keywords and geolocations. In contrast to previous works where they retrieve messages for a specific domain and type of crisis events, we get domain-independent messages to find and demonstrate our hypothesis that an anomaly in the activity related to locations can detect an emergency situation. Hence, we do not compare our detection method with others because they are centered in tracking specific crisis events. Besides, there are no jobs with independent domain to use as baseline related to emergency situations.

The paper is organized as follows: we first introduce an overview of relevant literature related to emergency situation and event detection in social media. Next we present a complete description of our proposal. Then, we provide a full description of our dataset and later we summarize our experimental validation over the ground truth and on-line evaluation. Finally, we deliver our discussions, conclusions and future work.

2 RELATED WORK

We discuss relevant work for our proposal, primary from two areas: (1) crisis-related social media monitoring and (2) event detection based on locations.

2.1 Crisis-Related Social Media Monitoring

Twitter has been used extensively during emergency situations to extract and identify relevant information. However, social media communications during emergency situations have become so abundant that it is necessary to sift through millions of data points to find information that is most useful during a given event [13].

One of the main tasks related to emergency situations is to detect a new real-crisis event in social media. Most existing event-detection methods described in the literature are based on keywords. For instance, *TweetTracker* [19] presented a case study using tweets that discuss a cholera outbreak in Haiti. The primary mechanism for monitoring tweets is through specific keywords and hashtags filters related to the Haiti. To detect a new event, emerging trends are identified based on the analysis of older tweets. In the same way, *EMERSE* [5] used a set of keywords related to the Haiti earthquake and applies a random forests algorithm for classifying messages. Likewise, the *Twicalli* system [23] introduced an unsupervised approach to detect earthquakes that only requires a general list of keywords.

Researchers at CSIRO Australia proposed *ESA* [4, 35], a system to detect disasters in Australia and New Zealand. This system is based on a probabilistic method to identify bursty keywords and historical data to build a language model of word occurrences. Alerts are identified if a term has a probability distribution that significantly deviates from the language model. Similarly, *Twitcident* [1] system was developed for detecting incidents which relies on emergency broadcasting services, such as the police, the fire department and other public emergency services. The *Twitcident* framework translates the broadcasted message into an initial incident profile that is applied as query to collect messages from Twitter, where an incident profile is a set of weighted attribute-value pairs that describe the characteristics related to the incident.

CrisisLex [26] is introduced to extract different crisis-related terms to filter messages in Twitter. In this work, authors collected six disasters which affected several millions of people. Data was collected from Twitter using two samples: a keyword-based sample and location-based sample. Based on datasets, they created a lexicon of the most frequent terms that appear in relevant messages posted during different types of crisis situations.

Our current work, on the other hand, is based on detecting events by monitoring *zscore* variations through fixed time-windows.

2.2 Event Detection Based on Locations

In addition to crisis-related social media monitoring, there are also some unsupervised event detection approaches based on location information. *Jasmine* [33] detected local real-world events using geolocation information from microblog documents. To detect such events, they identify a group of tweets that describe a particular theme, which are generated within a short time frame and a same geographic area. Similarly, Unankard, Li, and Sharaf [31] proposed an approach for early detection of emerging hotspot events in

uso
de
lo-
cali-
dades

sobre
el
base-
line

social networks with location sensitivity. In this work, the authors identified strong correlations between user locations and event locations when detecting emerging events using content similarity between clusters. In the work of Walther and Kaisser [32] real-world events were detected in a small scale with messages from the New York metropolitan area. There, clusters were created for each candidate event and evaluated using cluster score based on textual features (sentiment analysis, common theme, duplicate, etc) and other features (tweet count, unique coordinates, etc.).

Supervised approaches have also been proposed. *TEDAS* [20] detected, analyzed and identified events using refined rules (e.g., keywords, hashtags) and classify messages based on content as well Twitter specific features as URLs, hashtags and mentions. Besides, location information is extracted using both explicit geographical coordinates and implicit geographical references in the content. In the same way, Becker et al. [2] proposed an on-line clustering technique, which continuously clusters similar tweets and then classifies the clusters using Support Vector Machine algorithm. Finally, events (clusters) are classified into real-world events or nonevents.

Summary. Most of the existing work for detecting emergency situations, rely on keywords using probabilistic temporal models for specific domains (e.g., using keywords or locations). In general, these approaches require background knowledge about the event, and therefore, they will not identify new types of emerging crisis situations. In the case of existing approaches for event detection based on locations, these have been designed for small areas and are based on historical data. Regardless of the type of approach used (supervised or unsupervised), events are always characterized based on textual features, and on-line clustering is the most common technique used to create candidate events.

Based on the works presented by Guzman and Poblete [10] and Maldonado et al. [23], we extend the ideas of bursty keywords and *z-score* variation between fixed time-windows and we apply these proposals over locations to identify anomalies in social media activity. Furthermore, we do not use a set of keywords to detect specific types of events. In contrast, our proposal is focused on detecting bursty activity related to such events by tracking frequencies, and probability distributions of the interarrival time of the messages in specific locations.

3 PROPOSED APPROACH

Our focus in this proposal is to detect an emergency situation based on identifying anomalies in social media activity related to locations. In this way, the main task is to extract locations from messages by reducing the noise and irrelevant information. For this reason, data is pre-processed to allow a better analysis.

For location extraction task, we used a geographical dictionary (also known as gazetteer) to create a geographic hierarchy for a specific country. To understand the effect of an emergency situation inside to the country, we created signals in different levels of the geographic hierarchy and the tweet metadata. Afterwards, signals are divided into fixed time-windows and non textual features are computed for each of these. Finally and with the goals of filtering false positives detection, we created a geographic spread based on the proximity between locations by using an adjacency matrix.

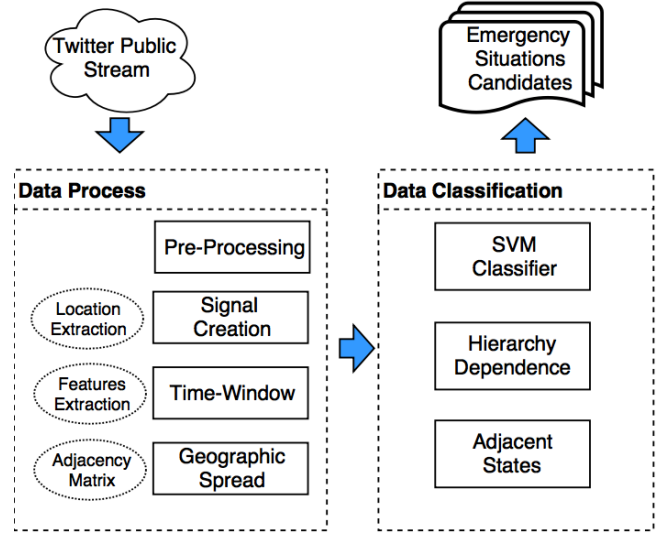


Figure 1: Key components of the proposed approach.

In order to provide a complete coverage of location-based detection of emergency situations, we divided our approach into four stages (depicted in the “data process” module in Figure 1). Next, we describe each stage:

3.1 Data Pre-Processing

Our focus is on localized bursty user activity. Therefore, initially, we filter messages according to the most common language used in each country (using the attribute *lang* in tweet metadata³). For example, when analyzing Italy, we only take into account messages written in Italian. This helps to filter noise of unrelated messages. We also remove user mentions, URLs, special characters, and apply text tokenization. We do not remove hashtags nor stopwords, because some locations can be included as hashtags, and some location names contain stopwords, which differentiate them from other locations or other terms.

3.2 Signal Creation

We create a set of discrete-time signals for each location, which indicates each time that a message related to a specific location was posted. In order to explain the effect of an emergency situation in a local and national scope, we use the lowest possible geographical hierarchy level available with the aim of comparing the impact in the highest level. Furthermore, we study the anomalies in different metadata levels to understand how locations are shared in Twitter, for instance, either based in the locations setted by users in their profile or sharing location in their messages.

3.2.1 Geographical Hierarchy. We use the idea of *gazetteer* as a *tree* presented in [34] in which each place is associated with a canonical taxonomy node. We create our gazetteer tree based

³<https://developer.twitter.com/en/docs/tweets/data-dictionary/overview/tweet-object>

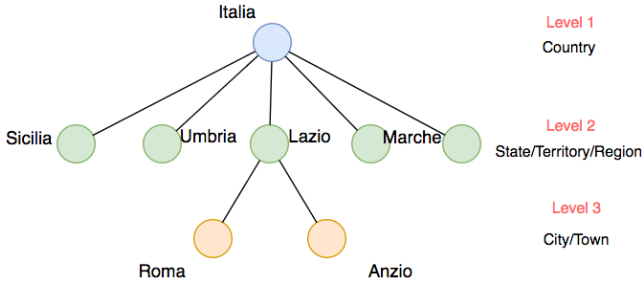


Figure 2: Example of gazetteer tree for Italy.

on Geonames⁴ and Wikipedia⁵. However, in [34] the gazetteer hierarchy presents four levels where the lowest level represents a specific point of interest. In our approach, we use a subset of the gazetteer hierarchy with only three levels: *city*, *state* and *country*. We do so because a large amount of users specify their location down to city level [11]. For example, if we have the *city:Manchester*, we associate this location with *region-state:North West* and also with *country:England*. As indicated in our data pre-processing stage, we consider only locations in the native language of the country. For instance, in the case of Italy locations, we consider *Roma* and not *Rome* (Figure 2).

3.2.2 Location Extraction. The structure of the tweet metadata contains information about the message and the user. Given a small portion of users sharing their current location using GPS coordinates [9], we do not consider this level of the tweet metadata in this work.

Considering the aforementioned geographical hierarchy, we extract locations from different parts of the metadata, creating 3 signals for each location:

- *tweet text*: the location is mentioned in the attribute text of the tweet, that is, on the body of the message.
- *user location*: the location is mentioned in the attribute location inside the user object, that is, the location set by the user in their profile.
- *tweet text - user location*: the location is mentioned in the attribute text of tweet object and also location is mentioned in the attribute location inside the user object. This means that the location is mentioned in the body of the message and the user who shares message has the same location in his profile. In this case, tweet text and user location can be different in the smallest hierarchy, but in the highest level can be equal.

In this way, mixing geographical hierarchy and locations in microblog metadata, we create $N \times M$ signals where N is the number of locations obtained by gazetteer tree and M is the number of metadata-levels extracted from the tweet object. For instance, we create a signal for *city:Manchester* and we find this hierarchy in *metadata:Tweet text* and also in *metadata:User location*. That means that we track the mention of *city:Manchester* at the level of the

body of message and at the level of the location of the user profile individually.

3.3 Time-Window

In this stage we address the problems of how to divide and determine the time-window size to detect a new emergency situation and what features by the time-window allow it.

3.3.1 Determining Optimal Window Size. According to Guzman and Poblete [10]: “If the window size is too small, the occurrence of empty windows for a term increases, making the noise rate increase and frequency rate tend towards zero. On the other hand, if the window size is too large, the stability of the signal becomes constant and bursty keyword detection is delayed”. Using this definition, we divide our signals into windows of six minutes because it divides a 24-hour day exactly, making the analysis easier to understand and to compare from different days.

3.3.2 Normalized Frequency. We compute the number of the messages of each time-window by signal. To normalize frequency, we compute z-score as following:

$$zscore = \frac{x_i - \mu_k}{\sigma_k} \quad (1)$$

where x_i is the frequency of the current i time-window, μ_k and σ_k are mean and standard deviation of the previous k time-windows respectively.

3.3.3 Interarrival Time. To characterize the urgency of the messages during a time-window, we compute the *interarrival time* which is defined as $d_i = t_{i+1} - t_i$, where d_i denotes the difference between two consecutive social media messages i and $i + 1$ that arrived in moments t_i and t_{i+1} respectively. Using this definition, which follows the work of Kalyanam et al. [16], high-activity events have a high-frequency in the first bins represented by values $d_i \approx 0$.

To quantify a high-frequency in very small values of d_i , we compute the measures *skewness* and *kurtosis*, which represent the asymmetry and the tailedness of the shape of probability distribution respectively [24]. Finally, we apply the equation 1 over *skewness* and *kurtosis* to calculate variation based on previous values.

3.4 Geographic Spread

An emergency situation that affects and mobilizes response in a small area is defined as *focalized*, while a disaster with a large geographic impact is defined as *diffused* [27]. Using this definition, we extend this concept to represent neighborhoods between locations obtained from section 3.2.1. For that purpose, we create an *adjacency matrix* M , where $M_{i,j} = 1$ represents if two locations are geographically connected and $M_{i,j} = 0$ if they are not connected. For instance, if an event is diffused (e.g., earthquake), the detection should be in adjacent-locations independently of metadata-level. On the other hand, if an event is focalized (e.g., terrorist attack), just one location should be detected but in different metadata-levels simultaneously.

4 EXPERIMENTAL ANALYSIS

Our experiments aim to find empirical evidences that allow detecting an emergency situation. Thereby, we use the locations for

⁴<http://download.geonames.org/export/dump/>

⁵<http://www.wikipedia.org/>

Table 1: List of earthquakes studied as ground truth, sorted by date.

Country	Datetime (UTC)	Magnitude (Mw)	Language
Italy	2016-10-26 17:10:36	5.5	Italian
Italy	2016-10-30 06:40:17	6.6	Italian
Chile	2016-12-25 14:22:26	7.6	Spanish
Chile	2017-04-23 02:36:06	5.9	Spanish
Chile	2017-04-24 21:38:28	6.9	Spanish

a specific country in different levels of the geographic hierarchy and the tweet metadata. We construct our ground truth based on two publicly available earthquake catalogs and apply the proposal presented in the above section. Thereafter, we study the performance of our methodology mainly based on the concept of affected adjacent locations during an emergency situations.

4.1 Dataset Description

We collect data from Twitter Public Streaming API, which allows access to subsets equal to 1% of public status descriptions in real-time. With this tool, we can retrieve either messages using a set of keywords or messages from specific locations setting a bounding box. In our approach, we get entire subsets of messages without use of keywords or specific locations. Then, we retrieve random messages about any topic and any place in the world.

4.2 Ground Truth

We choose earthquakes as crisis situation for our ground truth since that according to Carr [6], these events are also known as *instantaneous-diffused* events. These events are characterized since that nobody could do anything about it and wrecked its effects on the entire community. This definition is relevant because an unexpected (or instantaneous) event generates an anomaly of frequency in the social media activity since that it disrupts the “normal life” of the users. Furthermore, diffused events affect a big portion of the users simultaneously generating a collective reaction in neighbors affected locations.

In this way, we analyze five earthquakes with magnitudes between $5.5Mw$ and $7.6Mw$ ⁶, occurring in Italian-speaking and Spanish-speaking countries between October 2016 and April 2017 (Table 1). For that purpose, we collect 20 million messages 12 hours before and after the emergency situation events.

According to our proposal we create both the gazetteer hierarchies⁷ for each country and construct the signals based on each hierarchy and metadata-level. As a result of the number of messages of each signal (Table 2), we discard all signals related to city hierarchy since a great amount of small cities have zero frequency in a normal situation unlike to capital or metropolitan cities.

4.2.1 Labeled Emergency Situations. The exact event datetime is obtained from National Seismology Agency in Chile⁸ and National

Table 2: Number of messages by signal.

Hierarchy	Metadata-level	Messages
All	All	87,291
Country	Tweet Text	11,584
	User Location	25,313
	Tweet Text - User Location	1,417
State	Tweet Text	4,110
	User Location	13,352
	Tweet Text - User Location	86
City	Tweet Text	1,415
	User Location	8,971
	Tweet Text - User Location	20

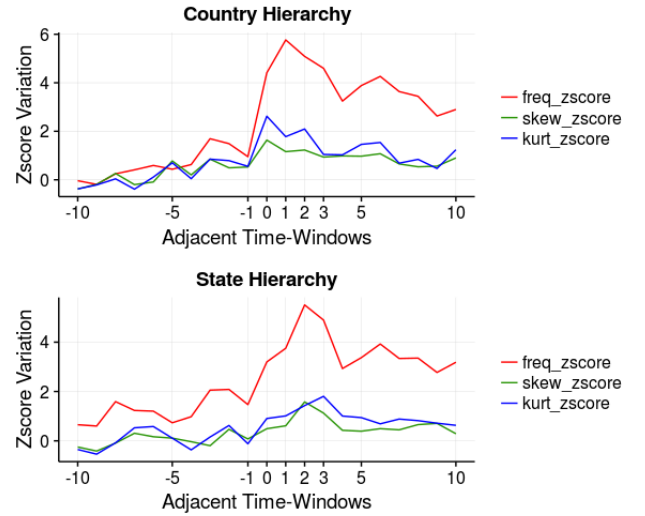


Figure 3: Average variation in emergency situations between time-windows. Positive and negatives values in x-axis represent the following and previous time-windows from the beginning of the event respectively.

Institute of Geophysics and Volcanology in Italy⁹. With the purpose of labeling a time-window as positive class (detection), we set as detection those time-windows with positive variation in frequency, skewness and kurtosis with respect to the normalization of the previous values. Moreover and according to (Figure 3), we include the three next time-windows after the event to compensate the imbalance between classes given that after these number of time-windows, the variation in the features decrease.

4.3 Methodology

According to the “data classification” module (Figure 1), we first train a classifier to identify emergency events. Also, we introduce the hierarchy dependence to understand the local and national

⁶Mw: the moment magnitude scale

⁷<http://users.dcc.uchile.cl/~hsarmien/gazetteer.html>

⁸<http://www.sismologia.cl/>

⁹<http://www.ingv.it/>

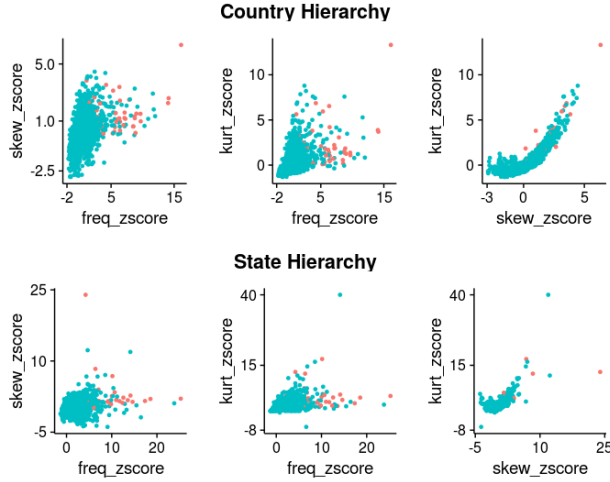


Figure 4: Relationship between features in country and state hierarchy. Red circles represent positive class (*detection*) and blue circles represent negative class (*nothing*).

impact when an high-impact real-world event occurs. Besides, diffused and focalized events are identified with the goal of filtering false positives detections.

Our filtering task can be seen as binary classification task. The positive class (*detection label*) corresponds to messages related to instantaneous emergency situations, while the negative class (*nothing label*) corresponds to the remaining or non-related to crisis situations.

To classify messages we employed traditional binary classifier Support Vector Machine (SVM). As a result of the analyzed data scattering (Figure 4), we separated country and state in different datasets and set both kernels and classification parameters independently. On the one hand, *country* classifier uses a polynomial kernel and strict-parameters for gamma, cost and weights since a great amount of messages are included in country hierarchy as an effect of the minor hierarchies. On the other hand, *state/region* classifier uses a linear kernel with soft-weights and cost.

Given that an emergency situation is not a usual event, we have an highly unbalanced data respect to the classes after labeled ($1 \approx 2\%$ of positive class corresponding to *detection*). Therefore, we used *under-sampling* [22] over country and state datasets increasing our positive class to $15 \approx 18\%$. Additionally, to validate our model, we used *5-fold cross-validation* where one earthquake dataset is used as testing and the remaining earthquakes dataset as training.

Table 3 shows the average results of our model applying 5-fold cross validation. In order provide a extended analysis about incorrect labels and time-windows, we include the metric *False Positive Rate (FPR)*.

4.3.1 Independent Analysis of Hierarchies. Our first analysis is just considering the hierarchies as isolated detections. The top of the Table 3 shows the results considers only the prediction over each label in our datasets. As noted above, the assignation from the lowest level (*city*) to the highest (*country*) in the gazetteer hierarchy

Table 3: Average performance of 5-fold cross-validation by hierarchy and geographic spread (G.S.)

	Hierarchy	P	R	F1	FPR
label	Country	0.3	0.83	0.45	0.14
	State	0.35	0.83	0.5	0.08
time-window	Country	0.15	0.77	0.25	0.15
	State	0.17	0.88	0.29	0.12
	Country-State	0.35	0.7	0.47	0.03
	Country(2)-State with G.S.	1	0.64	0.78	0
	Country(3)-State with G.S.	1	0.47	0.64	0

generated high frequency of messages which cause multiple *bursts* in our country signal for non emergency situations. This concept can explain the values of Precision (*P*) and *FPR*.

In addition to the analysis of number of detections by labels, we also study the number of detections by time-windows. For this analysis, we search the time-windows for each hierarchy where the all metadata-levels are well classified with correct class. According to the results shown on the middle of the Table 3, when we analyze country and state independently the values of Precision, F1 and FPR have worst values than the analysis by label .

4.3.2 Dependent Analysis of Hierarchies. Our second analysis considered the hierarchies as non-isolated detections. In the results explained above, we considered country and state hierarchy independently, which is not a correct analysis because an emergency situation affects states and country at the same time. For this reason, we inspected the time-windows where all metadata-level for country and state hierarchy have a correct detection simultaneously. The results are shown in the row *Country-State* in the Table 3. In contrast to the independent analysis of country and state, we improved the Precision, F1 and FPR values as a consequence of a smaller amount of the time-windows related to non emergency situations are assigned as detection. However, when we see the value obtained for FPR ($FPR = 0.03$), this rate represents an incorrect number of time-windows assigned as detection equal to 23. This means that we have 23 new emergency situations detected by our classifier.

4.3.3 Geographic Spread Analysis. Our third analysis considered the hierarchies as non-isolated detections and applies the Geographic Spread (G.S.). Using the *Adjacency Matrix* to represent neighborhoods between regions/states, we considered as a correct detection those time-windows where the state/s classified as detection are defined as *Focalized* or *Diffused* and exist dependency between hierarchies.

In addition to the results of the dependency analysis explained above, we see that a large amount of time-windows for country hierarchy ($\approx 82\%$) have more than one metadata-level when exist a correct detection. This can be explained since an emergency situation produce a collective reaction on the level of body of the message (*tweet text*), users sharing any messages with profile location in a specific country (*user location*) or mixing both concepts (*tweet text - user location*).

Table 4: On-line evaluation of events occurred in England by time-windows (T-W) using Country(2)-State with G.S. method. The table shows the total number of detected time-windows, the number of detected time-windows before to the beginning and after to the end of the event. Last two columns show the detection delay time respect to the beginning of the event and the top 3 bigrams when the detection occurs.

Event	Detected T-W	T-W Before Event	T-W After Event	Delay (min)	Top 3 Bigrams
Premier League Soccer Matches	2	-	-	-	(man, utd), (new, year), (happy, new)
Westminster Terrorist Attack	13	0	13	32	(stay, safe), (terror, attack), (safe, everyone)
Manchester Terrorist Attack	12	1	11	23	(ariana, grande), (incident, arena), (grande, concert)
London Terrorist Attack	14	7	7	36	(stay, safe), (incident, bridge), (borough, market)
U.K. Elections	5	-	-	-	(theresa, may), (vote, labour), (van, dijk)
Adele Live in Wembley	9	7	2	-	(elland, road), (new, times), (phil, jackson)
England vs Slovenia Soccer Match	4	4	0	-	(simon, brodikin), (join, us), (theresa, may)
Metallica Live in London	4	4	0	-	(always, said), (chance, win), (carabao, cup)

Table 5: On-line evaluation of events occurred in England by time-windows (T-W) using Country(3)-State with G.S. method. The table shows the total number of detected time-windows, the number of detected time-windows before to the beginning and after to the end of the event. Last two columns show the detection delay time respect to the beginning of the event and the top 3 bigrams when the detection occurs.

Event	Detected T-W	T-W Before Event	T-W After Event	Delay (min)	Top 3 Bigrams
Premier League Soccer Matches	0	-	-	-	
Westminster Terrorist Attack	4	0	4	32	(terror, attack), (stay, safe), (terrorist, attack)
Manchester Terrorist Attack	2	0	2	23	(ariana, grande), (praying, everyone), (everyone, affected)
London Terrorist Attack	1	1	0	-	(ariana, grande), (around, world), (lady, gaga)
U.K. Elections	0	-	-	-	
Adele Live in Wembley	0	0	0	-	
England vs Slovenia Soccer Match	1	1	0	-	(per, day), (menswear, sample), (closed, roads)
Metallica Live in London	2	2	0	-	(happy, birthday), (chance, win), (always, said)

Considering the geographic spread by states and the number of metadata-levels by country hierarchy, we analyzed the results shown on the bottom of the Table 3. On the one hand, the row with value equal to *Country(2)-State with G.S.* represents the detection when we considered at least two metadata-levels for the country hierarchy and the geographic spread for states. In contrast to the the previous analyses, we improved the values of the Precision, F1 and FPR. The last metric is very important because there are no time-windows incorrectly assigned as emergency situations.

Consequently, the Recall values decrease which means that our method remove some time-windows classified as detection. Beside the percent of emergency situations detected is equal to 100% with an average delay equal to 10.4 minutes ($min = 6$, $max = 14$) from the impact of the event to the first detection.

On the other hand, *Country(3)-State with G.S.* represent the detection when we considered three metadata-levels for country hierarchy and the geographic spread for states. Similar to *Country(2)-State with G.S.*, we improved the values of Precision, F1 and FPR but our

recall decrease from $R = 0.58$ to $R = 0.47$, detecting 80% of the emergency situations with an average delay equal to 11.5 minutes ($min = 8$, $max = 14$) from the impact of the event to the first detection.

5 RESULTS

In this section, we present an on-line evaluation of our methodology for emergency situation detection. We identify several events occurred in England that includes emergency situations and non-emergency situations. In this way, our main goal is to detect emergency events in non labeled data and using a different language with respect to the languages used in our classifier. To reduce (as much as possible) the false positive detections, we use the analysis with the best performance presented in the Section 4.3.

5.1 On-line Evaluation

For our evaluation in the Twitter Public Stream, we trained a classifier with five earthquakes identified in our ground truth. Furthermore, our on-line evaluation dataset is formed by eight different events that occurred in England between December 2016 and October 2017. For each event we considered the full-day in which they occurred. The main goals of this evaluation is to know the capacity of our method to detect emergency situations and discard those non-related to emergency events that involve location references. Geographic spread analysis is used to evaluate our method because decrease the number of false positives detection. In the same way of the experiments in Section 4.3.3, we compare the results using the two presented methods respect to the number of metadata-levels by country hierarchy.

As can be noted on Table 4 and Table 5, we study three terrorist attacks and five high-impact real-world events related to soccer matches, music concerts and political elections. **In the case of the terrorist attacks, we study this type of events since that according to Carr [6], these crisis are identified as *instantaneous-focalized* events, where an unexpected event affects to the community but in a reduced area. Unlike to earthquakes (identified as *instantaneous-diffused* events), we study the capacity of our classifier to detect another type of event where the number of affected people are smaller than earthquakes, tsunami, and others *instantaneous-focalized* events.**

For *Premier League Soccer Matches* and *UK Elections*, we can not identify the beginning of the event, since in the first one there are many soccer matches during the analyzed day and in the second one there is no a specific start time. In order to know the topics when our method detects an event, we computed the Top 3 Bigrams in the detected time-windows. Also, we calculate the delay time for emergency events since the beginning of the event until the first detection.

On the one hand, the first evaluation *Country(2)-State with G.S.* has full detection of the terrorist attacks with average delay time equal to 30.3 minutes. These detections are related to the event given that the bigrams represent terms associated with crisis situations. However, the *London Terrorist Attack* has 50% of the detected time-windows after the event, which means that there are seven time-windows non-related to emergency situations. Besides the crisis situations analysis, we also study the number of detected



Figure 5: Delay time and number of locations in the first detection for diffused and focalized emergency situations.

time-windows in non-related to emergency situation events. In the same way, we have a large amount of misclassified time-windows that do not represent crisis situations as we can see in the Top 3 Bigrams for each non-related to event.

On the other hand, the second evaluation *Country(3)-State with G.S.* decreases the number non-related to emergency situations events detected as crisis situations. We can see three time-windows in two events detected as emergency situations (*England vs Slovenia*, and *Metallica Live in London*). In these cases, the time-windows are detected before the event and corresponding a non emergency situations according to the bigrams. Furthermore, when we analyzed the number of the detected emergency situations, two-thirds (66%) of the events are detected correctly with average delay time equal to 30.3 minutes. In the case of *London Terrorist Attack*, our method detects one time-window before the event but the bigrams describe that the detections do not correspond to crisis situations.

6 DISCUSSION

Our findings suggest that there is evidence to detect an emergency situation based on anomaly frequency of messages that contain locations for a specific country. Indeed, our method based on the number of metadata-levels by country hierarchy and geographic spread by state, detects a 80% of the events related to emergency situations as we could demonstrate in our ground truth. Also, our method is independent of the textual features because we apply the model over different languages as Spanish, Italian and English. Furthermore, we test our model in different types of crisis events such as earthquakes (EQ) and terrorist attacks (TA), **where these are identified in the literature as *instantaneous-diffused* and *instantaneous-focalized* events respectively.** Also, we apply our methodology on different magnitudes (in the case of earthquakes) and number of affected people (e.g., *Manchester Terrorist Attack* vs *Westminster Terrorist Attack*).

However, when we apply our method in the on-line evaluation, we detect 66% of the emergency situations that affected England. This explains that the signals, and for various reasons: the number of active users in United Kingdom¹⁰ which can affect the anomaly frequency of the messages since there exists a high daily average activity of the messages; similar locations in other countries (York \approx New York); and soccer teams with names of cities (Manchester United, Liverpool). These issues also can affect the number of false positive detection in which in the case of England was 30% of the non-related to emergency events.

Regarding the geographic spread where we define an emergency situation as diffused or focalized, we find some evidence that differentiates them. In the case of diffused events, the delay time of the our first detection was less than 12 minutes and in focalized events was greater than 30 minutes (Figure 5). This explains that, in diffused events such as earthquakes, a high number of people are affected (thousands or millions) at the same time by an event which generates a collective reaction in social media in the locations where the event impacted. In Figure 5, we can see that earthquakes have at least two detected locations in the first detection (except Italy EQ2). In contrast, focalized events have less amount of eyewitness (hundreds or thousands) then when the users share messages in social media, the frequency does not affect the average daily message of the country in the first minutes. This can be explained in Figure 5 where the terrorist attacks have just 1 detected location in the first detection.

Additionally, the delay time can be different for many reasons: datetime of the event (for example, during the early hours), few differences with the end of the current time-window, type of the affected locations (rural, urban cities) and the number of active users by locations.

7 CONCLUSION

In this paper we have presented a methodology for detecting an emergency situation based on location for a specific country. This approach is independent of the textual features and can be used in different types of events and languages. We show that the users act as self-organized in the affected locations like citizen sensors when an emergency situation occurs. We furthermore have presented an analysis of geographic spread for different types of events that can be categorized. However, our experiment considers just a small portion of emergency situations, which is not representative for all types of crisis situations according to either the hazard type (natural or human-induced), temporal development (instantaneous or progressive) or geographic spread (diffused or focalized).

There are many things that can improve our results. We will add Point of Interest to our gazetteer tree to increase the frequency by time-windows in each hierarchy. Furthermore, we will add more non-textual features as number of retweets and tweets, unique locations detected and special locations. We also plan to study the relevance of the different metadata-levels and assign weights for each. Finally, we will create a web application to visualize events in real-time.

¹⁰<http://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/> visited on January 2018

ACKNOWLEDGMENTS

This work has been partially funded by the Millennium Institute for Foundational Research on Data and FONDECYT 1170218.

REFERENCES

- [1] Fabian Abel, Claudia Hauff, Geert-Jan Houben, Richard Stronkman, and Ke Tao. 2012. Twitcident: fighting fire with information from social web streams. In *Proceedings of the 21st International Conference on World Wide Web*. ACM, 305–308.
- [2] Hila Becker, Mor Naaman, and Luis Gravano. 2011. Beyond Trending Topics: Real-World Event Identification on Twitter. *ICWSM 11, 2011 (2011)*, 438–441.
- [3] Axel Bruns and Jean E Burgess. 2012. Local and global responses to disaster: #eqnz and the Christchurch earthquake. In *Disaster and emergency management conference, conference proceedings*, Vol. 2012. AST Management Pty Ltd, 86–103.
- [4] Mark A Cameron, Robert Power, Bella Robinson, and Jie Yin. 2012. Emergency situation awareness from twitter for crisis management. In *Proceedings of the 21st International Conference on World Wide Web*. ACM, 695–698.
- [5] Cornelia Caragea, Nathan McNeese, Anuj Jaiswal, Greg Traylor, Hyun-Woo Kim, Prasenjit Mitra, Dinghao Wu, Andrea H Tapia, Lee Giles, Bernard J Jansen, et al. 2011. Classifying text messages for the Haiti earthquake. In *Proceedings of the 8th international conference on information systems for crisis response and management (ISCRAM2011)*. Citeseer.
- [6] Lowell Juilliard Carr. 1932. Disaster and the sequence-pattern concept of social change. *Amer. J. Sociology* 38, 2 (1932), 207–218.
- [7] Carlos Castillo. 2016. *Big crisis data: social media in disasters and time-critical situations*. Cambridge University Press.
- [8] UN DHA. 1992. Internationally agreed glossary of basic terms related to disaster management. *UN DHA (United Nations Department of Humanitarian Affairs)*, Geneva (1992).
- [9] Mark Graham, Scott A Hale, and Devin Gaffney. 2014. Where in the world are you? Geolocation and language identification in Twitter. *The Professional Geographer* 66, 4 (2014), 568–578.
- [10] Jheser Guzman and Barbara Poblete. 2013. On-line relevant anomaly detection in the Twitter stream: an efficient bursty keyword detection model. In *Proceedings of the acm sigkdd workshop on outlier detection and description*. ACM, 31–39.
- [11] Brent Hecht, Lichan Hong, Bongwon Suh, and Ed H Chi. 2011. Tweets from Justin Bieber’s heart: the dynamics of the location field in user profiles. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 237–246.
- [12] Quynh Huang, Guido Cervone, Duangyang Jing, and Chaoyi Chang. 2015. DisasterMapper: A CyberGIS framework for disaster management using social media data. In *Proceedings of the 4th International ACM SIGSPATIAL Workshop on Analytics for Big Geospatial Data*. ACM, 1–6.
- [13] Muhammad Imran, Carlos Castillo, Fernando Diaz, and Sarah Vieweg. 2015. Processing social media messages in mass emergency: A survey. *ACM Computing Surveys (CSUR)* 47, 4 (2015), 67.
- [14] Muhammad Imran, Carlos Castillo, Ji Lucas, Patrick Meier, and Sarah Vieweg. 2014. Aidr: Artificial intelligence for disaster response. In *Proceedings of the companion publication of the 23rd international conference on World wide web companion*. International World Wide Web Conferences Steering Committee, 159–162.
- [15] Muhammad Imran, Shady Elbassuoni, Carlos Castillo, Fernando Diaz, and Patrick Meier. 2013. Extracting information nuggets from disaster-Related messages in social media. In *ISCRAM*.
- [16] Janani Kalyanam, Mauricio Quezada, Barbara Poblete, and Gert Lanckriet. 2016. Prediction and Characterization of High-Activity Events in Social Media Triggered by Real-World News. *PLoS one* 11, 12 (2016), e0166694.
- [17] Sarvnaz Karimi, Jie Yin, and Cecile Paris. 2013. Classifying microblogs for disasters. In *Proceedings of the 18th Australasian Document Computing Symposium*. ACM, 26–33.
- [18] Yury Kryvasheyeu, Haohui Chen, Nick Obradovich, Esteban Moro, Pascal Van Hentenryck, James Fowler, and Manuel Cebrian. 2016. Rapid assessment of disaster damage using social media activity. *Science advances* 2, 3 (2016), e1500779.
- [19] Shamanth Kumar, Geoffrey Barbier, Mohammad Ali Abbasi, and Huan Liu. 2011. TweetTracker: An Analysis Tool for Humanitarian and Disaster Relief. In *ICWSM*.
- [20] Rui Li, Kin Hou Lei, Ravi Khadiwala, and Kevin Chen-Chuan Chang. 2012. Tedas: A twitter-based event detection and analysis system. In *Data engineering (icde), 2012 IEEE 28th international conference on*. IEEE, 1273–1276.
- [21] John Lingad, Sarvnaz Karimi, and Jie Yin. 2013. Location extraction from disaster-related microblogs. In *Proceedings of the 22nd international conference on world wide web*. ACM, 1017–1020.
- [22] Nicola Lunardon, Giovanna Menardi, and Nicola Torelli. 2014. ROSE: A Package for Binary Imbalanced Learning. *R Journal* 6, 1 (2014).

- [23] Jazmine Maldonado, Jheser Guzman, and Barbara Poblete. 2017. A Lightweight and Real-Time Worldwide Earthquake Detection and Monitoring System Based on Citizen Sensors. In *Proceedings of the Fifth Conference of Human Computation and Crowdsourcing*. AAAI, 137–146.
- [24] Kanti V Mardia. 1970. Measures of multivariate skewness and kurtosis with applications. *Biometrika* 57, 3 (1970), 519–530.
- [25] Marcelo Mendoza, Barbara Poblete, and Carlos Castillo. 2010. Twitter Under Crisis: Can we trust what we RT?. In *Proceedings of the first workshop on social media analytics*. ACM, 71–79.
- [26] Alexandra Olteanu, Carlos Castillo, Fernando Diaz, and Sarah Vieweg. 2014. CrisisLex: A Lexicon for Collecting and Filtering Microblogged Communications in Crises.. In *ICWSM*.
- [27] Alexandra Olteanu, Sarah Vieweg, and Carlos Castillo. 2015. What to expect when the unexpected happens: Social media communications across crises. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 994–1009.
- [28] Liza Potts, Joyce Seitzinger, Dave Jones, and Angela Harrison. 2011. Tweeting disaster: hashtag constructions and collisions. In *Proceedings of the 29th ACM international conference on Design of communication*. ACM, 235–240.
- [29] Christian Reuter and Marc-André Kaufhold. [n. d.]. Fifteen years of social media in emergencies: a retrospective review and future directions for crisis informatics. *Journal of Contingencies and Crisis Management* ([n. d.]).
- [30] Kevin Stowe, Michael J Paul, Martha Palmer, Leysia Palen, and Kenneth Anderson. 2016. Identifying and categorizing disaster-related tweets. In *Proceedings of The Fourth International Workshop on Natural Language Processing for Social Media*. 1–6.
- [31] Sayan Unankard, Xue Li, and Mohamed A Sharaf. 2015. Emerging event detection in social networks with location sensitivity. *World Wide Web* 18, 5 (2015), 1393–1417.
- [32] Maximilian Walthert and Michael Kaisser. 2013. Geo-spatial Event Detection in the Twitter Stream.. In *ECIR*. Springer, 356–367.
- [33] Kazufumi Watanabe, Masanao Ochi, Makoto Okabe, and Rikio Onai. 2011. Jasmine: a real-time local-event detection system based on geolocation information propagated to microblogs. In *Proceedings of the 20th ACM international conference on Information and knowledge management*. ACM, 2541–2544.
- [34] Jie Yin, Sarvnaz Karimi, and John Lingad. 2014. Pinpointing locational focus in microblogs. In *Proceedings of the 2014 Australasian Document Computing Symposium*. ACM, 66.
- [35] Jie Yin, Sarvnaz Karimi, Bella Robinson, and Mark Cameron. 2012. ESA: emergency situation awareness via microbloggers. In *Proceedings of the 21st ACM international conference on Information and knowledge management*. ACM, 2701–2703.