Detecting Emergency Situations in Twitter Using Self-Organized Citizen Sensors*

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

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

ABSTRACT

Most methods to detect emergency situations using Twitter rely on keyword. The problem of keyword-based methods is the need to train models in specific domains, in multiple languages, and for different types of events, for example: earthquakes, terrorist attacks, floods, etc. In contrast, our proposal is to rather use the recurring references of a location in the metadata of microblogging messages, where users act self-organized citizen sensors in real-locations. In order to identify an emergency situation, we characterize such events through frequencies, and probability distributions of the interarrival time of the messages. Our method uses a SVM classifier which is independent of the language of the text, and textual features. Finally, we describe an emergency situation according to geographic impact based on the propagation of the messages in social media between neighboring locations.

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 Sergio Campos. 1997. Detecting Emergency Situations in Twitter Using Self-Organized Citizen Sensors. In *Proceedings of ACM Woodstock conference (WOODSTOCK'97).* ACM, New York, NY, USA, 8 pages. https://doi.org/10.475/123_4

1 INTRODUCTION

Social media has become a major channel for communication during high-impact real-world events, for example: elections, sports events, emergency situations, etc. In any case, users act as "citizen sensors" where they share and post their mood, opinions, photos, videos and geographical point of interest (POI).

During emergency situations, traditional media suffer structural damage and real-time communications could be disrupted. Instead, microblogging has played a critical role in the last fifteen years since it has allowed users share real-time information from people local to the incident as current status, casualties, damages and alerts [12, 16, 26, 28]. For this reason, researchers have studied the behaviour during these events for to detect, summarize and classifyng messages with the goal to help authorities and the general public with situational awareness to act fast and conscientiously during a crisis situations.

Twitter, a microblogging which allows to share short messages (called *tweets*) and is currently used world-wide by over 300 million users[1]. About 80% of Twitter users access from mobile services which contribute to the inmediacy of diffusion of information specially during a crisis situations [6]. One of the main tasks during these situations is to detect a new real-event because as with most social media conversations, messages are often overcomed by irrelevant and redundant noise. In current works [5, 11, 16, 20], these tasks are solved with methods rely on keywords over *Twitter Public Streaming API*[2]. The problem with these keyword-based methods is the need to train in specific domains for different type of events. [23] generate set of keywords with different datasets but sometimes arise specific terms for one event, for example *#eqnz* for Earthquake in New Zealand or *#pabloph* for Typhoon Pablo in Philippines [3, 14, 25]. Furthermore, these set of keywords are trained for a unique language and can not be used in others.

Additionally, several works related to crisis situations show a strong relationship between type of event (what) and the spatiotemporal dimensions (when and where). For example, we can see top trends during an earthquake related to locations [22], strong relationship between proximity to hurricane path and hurricane-related social media activity [15], extraction of locations and POI during floods [18] and mixing geographic information system (GIS) with geo-tagged messages to improve disaster maping and real-time event tracking [9].

Based on previous works, the question we address here is: *Is there evidence to detect an emergency situation based on anomaly frequency of messages that contain locations?.*

Here, we propose a method based on recurring references of country locations in message's 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 inter-arrival time of the messages for each geographical hierarchy. To detect an emergency situation, we train a SVM classifier by each hierarchy and apply a geographic spread to filter false positives detection considering that an emergency situation can be *Focalized* or *Diffused.*

*Produces the permission block, and copyright information

WOODSTOCK'97, July 1997, El Paso, Texas USA

ACM ISBN 123-4567-24-567/08/06.
https://doi.org/10.475/123_4

[1] http://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/
[2] https://developer.twitter.com/en/docs/tweets/filter-realtime/overview

Our main contribution is to create a methodology that detecting instantaneous emergency situations just using the locations for a specific country. Furthermore, our classifier do not depend of the language to detect a new event since that do not use textual features as input. Moreover, we characterize crisis situations that affect a small or large geographic area.

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 divide in four parts. Thereafter, we provide a full description of our dataset and later we summarize our experimental validation over the ground truth and online evaluation. Finally, we deliver our discusions, conclusions and future work.

2 RELATED WORK

Work related to this paper primarily arises 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 used extensively during emergency situations to extract and identify relevant information in recent years. However, social media communications during emergency situations are now so abundant that it is necessary to sift through millions of data point to find information that is most useful during a given event [10].

One of the main task related to emergency situations is to detect a new real-crisis event in social media. Most existing event-detection method described in the literature are based on keywords. For instance, *TweetTracker* [16] presents a case study using tweets that discuss the Cholera outreak 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] uses a set of keywords related to Haiti earthquake and applies random forests algorithm for classifyng messages. Likewise, *Twicalli* system [20] introduces an unsupervised approach to detect earthquakes that only require a general list of keywords. In this work, detection is made monitoring *zscore* variation through fixed time-windows.

Researchers at CSIRO Australia propose *ESA* [4, 33], a system to detect disasters in Australia and New Zealand. This system are based on probabilistic method to identify bursty keywords and historical data to build a language model of word occurences. Alerts are identified if a term has a probability distribution that significantly deviates from the language model. Similarly to specific domains for a country, *Twitcident* [1] system is developed for detecting incident which relies on emergency broadcasting service as police, fire departament or other public emergency services. *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.

In order to describe different crisis-related terms to filter messages in Twitter, *CrisisLex* is presented in [23]. In this work, authors collected six disasters which affected up to several million people. Data is 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.

2.2 Event Detection Based on Locations

Additionally to crisis-related social media monitoring, event detection based on locations for other types of real-world events have been developed for which several unsupervised approaches have been proposed in the literature. *Jasmine* [31] detects local events in the real-world using geolocation information from microblog documents. To detect such events, they identify a group of Twitter documents describing the same theme that were generated within a short time and a small geographic area. Similarly, [29] proposes an approach for the early detection of emerging hotspot events in social network with location sensivity. In this work, the authors identify strong correlations between user locations and event locations in detecting the emerging events using content similarity between clusters. In [30] real-world events are detected in a small scale with messages from the New York metropolitan area. There, clusters are 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* [17] detects, analyzes and identifies 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, locations information is extracted using both explicit geographical coordinates and implicit geographical references in the content. In the same way, [2] proposes an online clustering technique, which continuosly clusters similar tweets and then classifies the clusters using Support Vector Machine algorithm. Finally, events (clusters) are classified into real-world events or nonevents.

Summarizing, most previous works to detect an emergency situations rely on keywords using probabilistic temporal models for specific domains (eg. keywords or location). In general terms, this approach needs knowledge about the event and a new emerging topic referring to a crisis can not identify to detect an emergency situation. In the case of the works about event detection based on locations, they are designed to be applying in a small area and based on historical data. Regardless of which approach is implemented (supervised or unsupervised), textual features are the most used attribute to characterize an event and online clustering is the most common technique to create candidate events.

3 PROPOSED APPROACH

In order to provide a complete coverage of location-based detection of emergency situation, we propose an approach which has four stages as we can in Data Process module in Figure 1. The following information provides details of each stage.

3.1 Data Pre-Processing

Since that we consider users as citizen sensors, we filter messages depending of native language of each country using the attribute

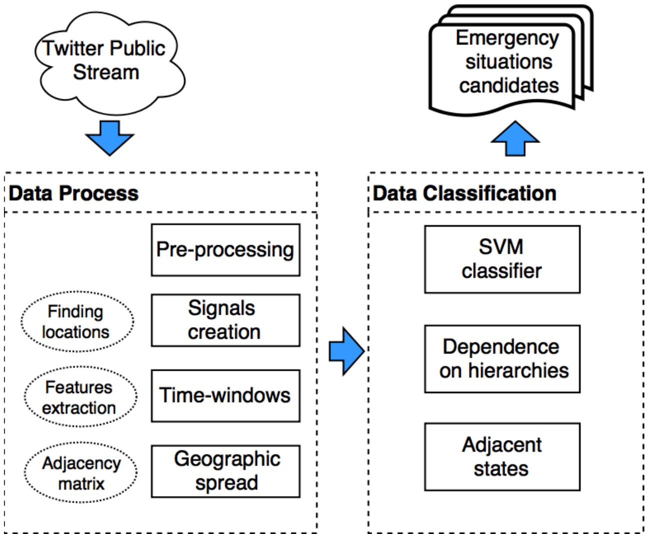

Figure 1: Key components of the proposed approach.

lang retrieved from Tweet Metadata[3]. For example, if we analyze Italy, we just consider messages in italian. Finally, we remove user mentions, urls, special characters and apply text tokenization . Also we do not remove hashtags or stopwords because some locations can be included in hashtags terms and some places can be contained stopword that can differentiate them from other similar non-related to locations terms.

3.2 Signals Creation

The problems that we address in this stage is how to infer locations from micro-blog messages to track signal over time in the Stream and what is the lowest hierarchy level to find and assign to this signal.

3.2.1 Geographical Hierarchy. Using the idea of *Gazetteer as a Tree* presented in [32] in which each place is associated with a canonical taxonomy node, we construct our gazetteer tree based on Geonames[4] and Wikipedia[5] for each country to analyze. However, in [32] the gazetteer hierarchy presents four levels where the lowest level represents a specific POIs. In our approach, we use a subset of the gazetteer hierarchy with three levels, in which the lowest level is represented by City since a large amount of users specify their location down to the city scale [8]. For example, if we have the *city:Manchester*, we associate this location with *region-state:North West* and also with *country:England*. Furthermore, the name of locations are considered just in the native language of the country. For instance, in the case of Italy locations, we consider *Roma* and not *Rome*.

[3]https://developer.twitter.com/en/docs/tweets/data-dictionary/overview/tweet-object
[4]http://download.geonames.org/export/dump/
[5]http://www.wikipedia.org/

3.2.2 Finding Locations in Micro-Blog Metadata. The structure of the Tweet Metadata allows to obtain information about the message and the user who shares a message. Considering the geographical hierarchy explained in the previous section, we search these locations on different levels of metadata and create one signal for each as following:

- Tweet Text: location is mentioned in the attribute `text` of tweet object, that is, on the body of the message.
- User Location: location is mentioned in the attribute `location` inside the `user object`, that is, the location set by the user in his profile.
- Tweet Text - User Location: 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 on the body of the message and the user who share 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 micro-blog metadata, we create NxM signals where N is the number of locations obtained by gazetteer tree and M is the number of metadata-levels extracted from Tweet Object. In 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-Windows

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

3.3.1 Determining Optimal Window Size. According to [7]: "If the window size is to small, the ocurrence of empty windows for a term increases, making the noise rate increate and frequency rate tend to 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 Inter-arrival Time. To characterize the urgency of the messages during a time-window, we use our previous results [27] to compute the *Inter-arrival Time* which is defined as $d_i = t_{i+1} - t_i$, where d_i denotes the difference between two consecutive social

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

media messages i and $i+1$ that arrived in moments t_i and t_{i+1} respectively. Using this definition and according to [13], high-activity events have a high-frequency in the first bins represented by values $d_i \approx 0$.

In order 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 [21]. 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* [24]. Using this definition, we extend this concept to represent neighborhoods between locations obtained from section 3.2.1. For that purpose, we create an *Adjancency 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 (eg. earthquake), the detection should be in adjacent-locations independently of metadata-level. In the other hand, if an event is focalized (eg. terrorist attack), just one location should be detected but in different metadata-levels simultaneously.

4 DATASET

Data is collected 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 keywords or specific locations. Then, we retrieve random messages about any topic and any place in the world.

4.1 Ground Truth

We analyze five earthquakes with magnitudes between 5.5Mw and 7.6Mw[6], ocurring in Italian-speaking and Spanish-speaking countries between October 2016 and April 2017 (Table 1). For that purpose, we collect 20 millions of messages 12-hour before and after the emergency situation events.

According our proposal, we create both the gazetteer hierarchies[7] for each country and construct the signals based on each

[6]Mw: the moment magnitude scale

[7]http://users.dcc.uchile.cl/~hsarmien/gazetteer.html

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

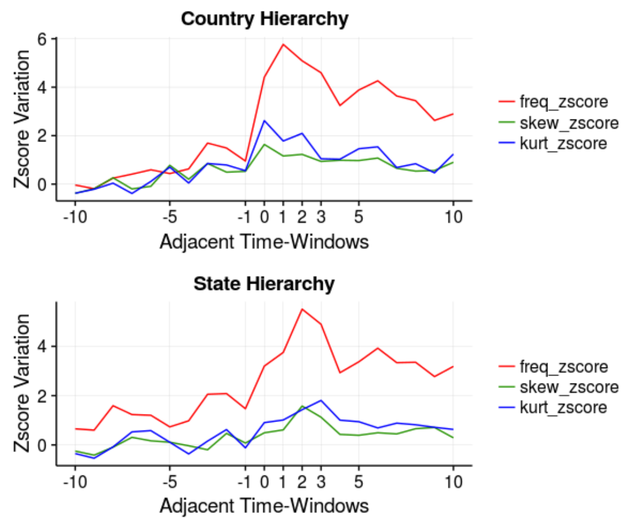

Figure 2: Average variation in emergency situations between time-windows.

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.

4.1.1 *Labeled Emergency Situations.*

The exactly datetime event is obtained from National Seismology Agency in Chile[8] and National Institute of Geophysics and Volcanology in Italy[9]. With the purpose of to labeling a time-window as positive class (detection), we set as detection those time-windows with positive variation in frequency, skewness and kurtosis respect to the normalization of the previous values. Moreover and according to (Figure 2), we include the three next time-windows after the event to compensate the imbalance between classes given that later to these number of time-windows, the variation in the features decrease.

[8]http://www.sismologia.cl/

[9]http://www.ingv.it/it/

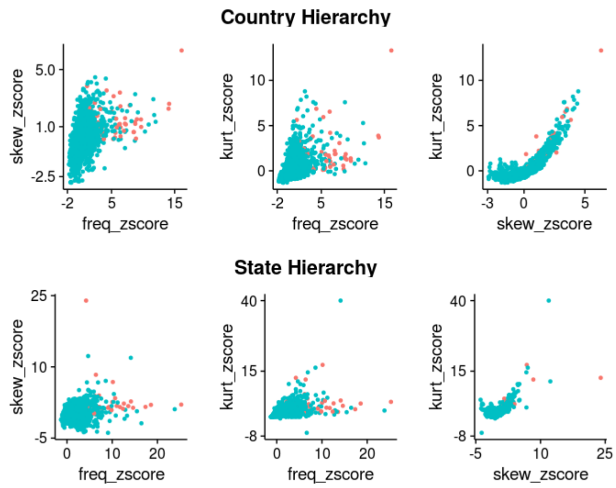

Figure 3: Relationship between features in country and state hierarchy. Red circles represent positive class (*detection*) and blue circles represent negative class (*nothing*).

5 EXPERIMENTS

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.

In order to classifier messages, we employ traditional binary classifier Support Vector Machine *(SVM)*. As a result of the analyzed data scattering (Figure 3), we separate 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 that 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 an 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 use *under-sampling* [19] over country and state datasets increasing our possitive class to $15 \approx 18\%$. Additionally to validate our model, we use *5-fold cross-validation* where one earthquake dataset is used as testing and the reamining 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).*

5.0.1 Independent Analysis of Hierarchies. Our first analysis is just considering the hierarchies as isolated detections. On the top of the Table 3, we show the results considering 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 generates high frequency of messages which cause multiples *bursts*

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

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 independentely the values of Precision, F1 and FPR have worst values than the analysis by label .

5.0.2 Dependent Analysis of Hierarchies. Our second analysis is considering the hierarchies as non-isolated detections. In the results explained above, we consider country and state hierarchy independently, which is not a correct analysis because an emergency situation affect to states and country at the same time. For this reason, we inspect 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 improve 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.

5.0.3 Geopraphic Spread Analysis. Our third analysis is considering the hierarchies as non-isolated detections and applying the Geographic Spread (G.S.). Using the *Adjacency Matrix* to represent neighborhoods between regions/states, we consider 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*).

Considering the geographic spread by states and the number of metadata-levels by country hierarchy, we analyze the results

Table 4: Online evaluation by time-windows (T-W) using Country(2)-State with G.S. method

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 at Wembley	9	7	2	-	(elland, road), (new, times), (phil, jackson)
England vs Eslovenia Soccer Match	4	4	0	-	(simon, brodkin), (join, us), (theresa, may)
Metallica Live at London	4	4	0	-	(always, said), (chance, win), (carabao, cup)

Table 5: Online evaluation by time-windows (T-W) using Country(3)-State with G.S. method

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 at Wembley	0	0	0	-	
England vs Eslovenia Soccer Match	1	1	0	-	(per, day), (menswear, sample), (closed, roads)
Metallica Live at London	2	2	0	-	(happy, birthday), (chance, win), (always, said)

shown on the bottom of the Table 3. On the one hand, the row with value equal to *Country(2)-State with G.S.* represent the detection when we consider at least two metadata-levels for the country hierarchy and the geographic spread for states. In contrast to the the previous analyzes, we improve the values of the Precision, F1 and FPR. The last metrics 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 a 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 consider three metadata-levels for country hierarchy and the geographic spread for states. Similar to *Country(2)-State with G.S.*, we improve 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 a average delay equal to 11.5 minutes ($min = 8$, $max = 14$) from the impact of the event to the first detection.

5.1 Online Evaluation

For our evaluation in the Twitter Public Stream, we training a classifier with five earthquakes identified in our ground truth. Futhermore, our evaluation dataset is formed by eight different events that ocurred in England between December 2016 and October 2017. For each event we consider the full-day in which they ocurred. The main goals of this evaluation is to know the capacity of our method to detect emergency situations and discard non-related to emergency situations events that involve references of locations.

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 5.0.3, we compare the results using the two presented methods respect to the number of metadata-levels by country hierarchy.

As can be appreciated 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 Premier League Soccer Matches and U.K Elections, we can not identify the beginning of the event, since that 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 detect an event, we compute the Top 3 Bigrams in the detected time-windows. Also, we calculate the delay time just for emergency situations events.

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 to the crisis situations analysis, we also study the number of detected 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 Eslovenia, and Metallica Live at 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-third (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 detection do not correspond to crisis situation.

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, detect 80% of the events related to emergency situations as we could demostrate 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 events such as earthquakes and terrorist attacks, and also 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 online evaluation , we detect 66% of the emergency situations that affected to England. This can explain because the signal are noise but many reason: the number of active users in United Kingdom[10] which can affect the anomaly frequency of message since that exist 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 situations events.

Figure 4: Delay time and number of locations in the first detection for diffused and focalized emergency situacions.

Reggarding the geographic spread where we define an emergency situation as diffused or focalized, we find some evidences that differentiate 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 4). This can be explain given that, in diffused events such as earthquakes, a large amount of people are affected (thousand or millions) at the same time by an event which generate a collective reaction in social media in the locations where the event impacted. In Figure 4, 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 (hundred or thousands) then when the users share messages in social media, the frequency not affect the average daily message of the country in the first minutes. This can be explain en Figure 4 where the terrorist attacks have just 1 detected locations in the first detection.

Additionally, the delay time can be different according to many reason: datetime of the event (for example, during the early hours), little 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.

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

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 self-organized in the affected locations like citizen sensors when an emergency situation occur. We furthermore have presented an analysis of geographic spread for different types of events, allowing characterize them. However, our experiment consider just a small portion of emergency situations which is not representive for all type of crisis situations according 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 be improved 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 study the relevance of the different metadata-levels and assign weights for each. Finally, we will create a web application to visualize in real-time events.

ACKNOWLEDGMENTS

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] Carlos Castillo. 2016. *Big crisis data: social media in disasters and time-critical situations*. Cambridge University Press.

[7] 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.

[8] 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.

[9] Qunying 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.

[10] 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.

[11] 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.

[12] Muhammad Imran, Shady Elbassuoni, Carlos Castillo, Fernando Diaz, and Patrick Meier. 2013. Extracting information nuggets from disaster-Related messages in social media.. In *ISCRAM*.

[13] 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.

[14] Sarvnaz Karimi, Jie Yin, and Cecile Paris. 2013. Classifying microblogs for disasters. In *Proceedings of the 18th Australasian Document Computing Symposium*. ACM, 26–33.

[15] 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.

[16] Shamanth Kumar, Geoffrey Barbier, Mohammad Ali Abbasi, and Huan Liu. 2011. TweetTracker: An Analysis Tool for Humanitarian and Disaster Relief.. In *ICWSM*.

[17] 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.

[18] 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.

[19] Nicola Lunardon, Giovanna Menardi, and Nicola Torelli. 2014. ROSE: A Package for Binary Imbalanced Learning. *R Journal* 6, 1 (2014).

[20] 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.

[21] Kanti V Mardia. 1970. Measures of multivariate skewness and kurtosis with applications. *Biometrika* 57, 3 (1970), 519–530.

[22] 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.

[23] Alexandra Olteanu, Carlos Castillo, Fernando Diaz, and Sarah Vieweg. 2014. CrisisLex: A Lexicon for Collecting and Filtering Microblogged Communications in Crises.. In *ICWSM*.

[24] 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.

[25] 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.

[26] 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.]).

[27] Hernan Sarmiento. 2017. Detecting Emergency Situations by Inferring Locations in Twitter. In *Seventh BCS-IRSG Symposium on Future Directions in Information Access, FDIA 2017, 5 September 2017, Barcelona, Spain*. https://doi.org/10.14236/ewic/FDIA2017.7

[28] 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.

[29] 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.

[30] Maximilian Walther and Michael Kaisser. 2013. Geo-spatial Event Detection in the Twitter Stream.. In *ECIR*. Springer, 356–367.

[31] 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.

[32] Jie Yin, Sarvnaz Karimi, and John Lingad. 2014. Pinpointing locational focus in microblogs. In *Proceedings of the 2014 Australasian Document Computing Symposium*. ACM, 66.

[33] 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.