



# Early Tracking of People's Reaction in Twitter for Fast Reporting of Damages in the Mercalli Scale

Marcelo Mendoza<sup>1</sup>(✉), Bárbara Poblete<sup>2</sup>, and Ignacio Valderrama<sup>2</sup>

<sup>1</sup> Centro Científico y Tecnológico de Valparaíso,  
Universidad Técnica Federico Santa María, Valparaíso, Chile  
`marcelo.mendoza@usm.cl`

<sup>2</sup> Department of Computer Science, Universidad de Chile, Santiago, Chile  
`{bpoblete,ivalderr}@dcc.uchile.cl`

**Abstract.** The Modified Mercalli Intensity Scale is a measure of the severity of an earthquake for a nonscientist. Since the Mercalli scale is based on perceived effects, it has a strong dependence on observers. Typically, these reports take time to be prepared and, as a consequence, Mercalli intensities are published hours after the occurrence of an earthquake. The National Seismological Center of Chile needs to provide a preliminary overview of the observed effects of an earthquake. This has motivated us to create a system for early tracking of people's reaction in social networks to infer Mercalli intensities. By tracking people's comments about the effects of an earthquake, a collection of Mercalli point estimates is retrieved at county level of granularity. We introduce the concept of Reinforced Mercalli support that combines Mercalli point estimates with social support, allowing to discard social unsupported estimates. Experimental results show that our proposal is accurate providing early Mercalli reports 30 min after an earthquake, detecting the maximum Mercalli intensity of an event with high accuracy in terms of mean absolute error (MAE).

**Keywords:** Social networks · Disaster management  
Mercalli intensity · Social media during emergencies

## 1 Introduction

The Modified Mercalli intensity scale (from now on Mercalli) is a commonly used measure that summarizes the effects of an earthquake in public infrastructure and damages perceived by people. Unlike the *moment magnitude* scale, which quantifies the released energy during an earthquake. Energy and damages may differ due to a number of physical variables, as the depth of the seismic movement and the geological composition of the ground. In addition, energy and damages may differ due to the standard used to certify the quality of buildings.

Mercalli reports are prepared by observers providing ratings for earthquakes in a given location. Seismological centers keep groups of observers distributed

along territories providing these reports. Usually, Mercalli reports are released hours after an earthquake, as the strong dependence on local observers makes difficult to provide fast and fresh information. Many factors obstruct fast reporting, among them the quality of communications during a disaster or the observer availability. To provide a first fast report of damages, the National Seismological Center have pay attention to the information propagated through social networks.

In this paper we study how social networks can be used to infer damages in the Mercalli scale after an earthquake. The state of the art show some efforts in this direction with promising results in the problem of earthquake detection. We extend the state of the art providing fast Mercalli intensity reports, focusing our efforts in the estimation of the maximum intensity in the Mercalli scale of damages. The spatial dimension of the problem, namely how people distribute along a territory and how this piece of information is involved in the Mercalli inference process, is the key building block of our fast Mercalli intensity report method.

*Main contribution of the paper:* In this paper we address the problem of Mercalli intensity inference using the spatial dimension of the data. From our point of view, social tracking is naturally related to the description of the effects associated to a given earthquake, as the Mercalli scale is a scale of perceived damages. However, the spatial dimension of the problem, namely how people is distributed along a territory and how this information is included in the inference process, is a key aspect of the problem not addressed in the state of the art. We claim that this aspect can not be discarded. Experimental results will show that our method outperforms the state of the art in the specific task of maximum Mercalli intensity detection.

This paper is organized as follows. Related work is discussed in Sect. 2. Our method is introduced in Sect. 3. Experiments are discussed in Sect. 4. Finally, we conclude in Sect. 5 giving conclusions and outlining future work.

## 2 Related Work

Twitter has been a social network of much research along time, as it is a huge source of user-generated content, which currently reaches more than 300 million active users per month. This is the reason why several scientists have researched Twitter with the aim of exploiting the information available, such as finding correlations between Twitter and physical events [6]. These efforts have shown interesting results. For instance, during the Tohoku earthquake in 2011, a research highlighted high correlations between the amount of tweets and the intensity of the disaster in some locations [5].

The elaboration of early reports of seismic event based on Twitter has been of growing interest during the last five years. Quake alert systems have been developed in different places of the world such as Australia [7] or Italy [1]. These systems use a burst detection algorithm to report an earthquake, where a burst is defined as a large number of occurrences of tweets within a short time window [9].

Despite the fact that these systems just report that an earthquake happened in a given location, they have shown that it is possible to infer more information. Maybe the most salient result on seismic event report relies on the estimation of the epicenter of an earthquake event using only information recovered from Twitter [8] as tweets counts and tweets rates.

Burks *et al.* [2] proposed the first approach to estimate the Mercalli intensity of an earthquake using Twitter. Conditioned on a set of reports retrieved from seismological recording stations provided by the Japanese seismological center, the area around each recording station is segmented into 9 radial areas, mapping tweets that mention the word 'earthquake' to these areas. Lexical features in each areal disc are calculated to study the correlation of these features with the Mercalli intensity. To do this, the authors explored a number of linear regression models, showing good results in terms of accuracy. We take some inspiration from the ideas explored by Burks *et al.* to design our method, but discarding the use of data recovered from seismological stations. The point here is that Japan has a huge network of seismological recording stations distributed along its territory, providing valuable information to the method. The aim of our study is to explore the predictive power of the social network itself, in specific Twitter, in absence of seismological recording stations reports, to provide a first fast Mercalli intensity report that does not depend on the quality and coverage of the seismological sensor network.

An on-line system named TwiFelt [4] has exploited the Twitter stream to provide an estimation of the extension area where an earthquake was felt in Italy. The system only use geo-located tweets to infer the area showing promising results for high intensity earthquakes. One limitation of this system relies on the availability of geo-located tweets, as a great proportion of the tweets recovered in our country from the stream does not include the tweet location.

Maybe the closest work to our proposal is the one authored by Cresci *et al.* [3]. In that paper the authors studied how to use Twitter to estimate the maximum intensity of an earthquake using only Twitter features. Using linear regression models over a huge collection of aggregated features (45 features were tested in that proposal), the authors showed that Twitter has enough predictive power to infer the maximum intensity of an earthquake in the Mercalli scale. The set of features tested by the authors comprises features extracted from user profiles, tweets contents and time-based features of the tweet stream (e.g. tweet interval rates). Our proposal can be considered as an extension of this work but focusing only on tweets contents. We will use a linear regression model for a first estimation of the Mercalli intensity. The main difference in the maximum intensity task between our method and the method proposed by Cresci *et al.* is that our proposal works over a reduced set of features (only 12 lexical features) in comparison with the 45 features used in [3]. We will show in our experiments that our method performs well in this specific task, taking advantage of the spatial dimension of the data boosting the results achieved by Cresci *et al.* [3].

### 3 Early Inference of Mercalli Intensities

We use information gathered from social networks, in specific from Twitter, to infer damages in the Mercalli scale. As this information can be collected and summarized in short periods of time, it is possible to infer at the early stages of an emergency the Mercalli intensities of a given earthquake. We divide the inference process in 3 stages: (1) Social tracking of earthquake effects, (2) Estimation of a region of interest, and (3) Inference of the maximum Mercalli intensity.

The first stage corresponds to the social tracking of an earthquake's effects. Each event of interest is characterized at county level, the finer level of geolocation considered in our method. The second stage of the process starts with a regression process applied to infer the region of interest of a given earthquake. The last stage of our method takes the collection of point estimates to infer the maximum intensity in the Mercalli scale for a given seism. We introduce a Reinforced Mercalli variable that is used to adjust the Mercalli estimate according to the level of support of the point estimate. Finally, we look for the maximum intensity at the area of interest, fixing this value as the maximum intensity of the earthquake.

#### 3.1 Social Tracking of Effects

Posts are collected to extract features of the event that characterize the social perception of the earthquake. In our study we use Twitter to collect the data. Each perceived event is characterized at a level of aggregation that describes the perception of the earthquake in a county. For each county batch, a set of features is calculated to describe the earthquake.

County batches are built as follows. After each earthquake, a set of tweets that matches the keywords “quake”, “earthquake” or “seismic” are retrieved from Twitter. The time considered to collect the data is a parameter of our system, with a window length of 30 min by default. Shorter windows can be considered but at the cost of less accurate Mercalli predictions. Tweets that are mapped to counties are aggregated into county batches. This piece of data is the basic unit of earthquake characterization used for feature extraction.

We map tweets to counties using the user location field. We were forced to use this field as only a very small fraction of the tweets in our country is geo-located. The user location field is retrieved from the user profile and then, using a fuzzy string matching procedure, it is mapped to a specific county. We understand that many tweets will be effectively posted from a location matching the user location, giving us a trace of the tweet spatial distribution. More accurate methods for tweet geolocation can improve this aspect of our method but to the best of our knowledge, this task is challenging and it is still open.

Twelve features are considered at this level of aggregation, as is shown in Table 1. These features are calculated in each county data batch, characterizing the set of tweets mapped to each specific county for a given seism.

**Table 1.** Features used for our tracking system

| Feature             | Description  |
|---------------------|--|
| NUMBER OF TWEETS    | Number of tweets in the data batch                   |
| TWEETS NORM         | Fraction of tweets over county population            |
| AVERAGE WORDS       | Average length of tweets in number of words          |
| AVERAGE LENGTH      | Average length of tweets in number of chars          |
| QUESTION MARKS      | Fraction of tweets with question marks               |
| EXCLAMATION MARKS   | Fraction of tweets with exclamation marks            |
| UPPER WORDS         | Fraction of tweets with uppercase words              |
| HASHTAG SYMBOLS     | Fraction of tweets containing the # (hashtag) symbol |
| MENTION SYMBOLS     | Fraction of tweets containing the @ (mention) symbol |
| RT SYMBOLS          | Fraction of tweets containing the “RT” symbol        |
| CONTAINS EARTHQUAKE | Fraction of tweets containing the word “earthquake”  |
| POPULATION          | Number of inhabitants in the county                  |

### 3.2 Estimation of a Region of Interest

Our method starts detecting the region of interest from where county data batches will be used to infer Mercalli intensities. This step of the method separate counties into two classes. We do this using a 0/1 classifier trained over county-seismic data batches pairs. These data batches were labeled according to the actual Mercalli intensity reported into two disjoint classes. The 0 class represents an earthquake that was not perceived (not reported in the Mercalli scale) and the 1 class represents an earthquake that was effectively perceived by people with an intensity value in the Mercalli scale. Each data batch is represented by a vector of features, using the features defined in Table 1. Once the 0/1 classifier was trained, our method is ready to detect the region of interest on new earthquakes at county level.

The set of counties labeled in the 1 class is used as an input to characterize the event in the third and last stage of our method: the inference of the maximum intensity in the Mercalli scale. Data batches labeled in the 1 class are provided to our method for a regression procedure, where the Mercalli intensity at each county will be estimated.

### 3.3 Maximum Mercalli Intensity Estimation

**Reinforced Mercalli Estimation.** The maximum Mercalli intensity estimation starts by inferring a reinforced Mercalli variable at county level. Let  $i$  be the index of a county in the set of counties distributed in the region of interest of the seism. The social support  $s(i) \in [0, 1]$  at the  $i$ -th county is defined as the ratio between earthquake observers (people who have posted at least one message related to the seismic movement during the period of observation) and Twitter users at the county. We combine  $m(i)$  with its support  $s(i)$  to provide a reinforced Mercalli support estimation. The Reinforced Mercalli support estimation

consist only of supported Mercalli point estimates. Unsupported Mercalli point estimates will be discarded combining both factors in a soft minimum bivariate function defined as follows:

$$\text{REINFORCED MERCALLI SUPPORT}(i) = \frac{2 \cdot \overline{m}(i) \cdot s(i)}{\overline{m}(i) + s(i)}, \quad (1)$$

where  $\overline{m}(i)$  is the Mercalli point estimate at the  $i$ -th county, constrained to the  $[0, 1]$  interval. To achieve a Mercalli point estimate in the  $[0, 1]$  interval, we normalize the estimate from the Mercalli intensity scale in  $\{1, 12\}$  to  $[0, 1]$  as  $\overline{m}(i) = \frac{m(i)-1}{11}$ . Then, as  $\overline{m}(i)$  and  $s(i)$  range in  $[0, 1]$ , the Reinforced Mercalli support function also ranges in  $[0, 1]$ .

The Reinforced Mercalli support only rates high supported events with high intensity in the Mercalli scale. The rationale of the Reinforced Mercalli support is to limit the effect of false positives. Then, only relevant events will be included in the tracking system at the cost of diminishing the effect of supported low intensity earthquakes. This cost is marginal for the tracking system as Mercalli intensity reports are only critical for high intensity earthquakes. Low intensity earthquakes can be characterized using only the Richter magnitude scale. As for low intensity events the impact in terms of damages is very limited, it is not necessary to characterize this kind of earthquakes on the damage scale.

**Adjusted Mercalli at County Level and Maximum Intensity Detection.** Our method considers each county as a sensor. Data aggregation at county level unleashes people reactions in front of an earthquake aggregating different reaction signals. We model the activation of a sensor at county level using an activation function. We do this using a sigmoid function  $\frac{1}{1+e^{-x}}$  to model the level of activation of a county in front of a given earthquake. The activation of the function is fixed to Mercalli intensity 3, as is defined as the first level of the Mercalli scale where the event is felt. Then, a county is considered as active starting from level 3, as from this level to up, the earthquake will be reported. We do this by applying the sigmoid function to the Reinforced Mercalli support, scaling the function to  $\{1, 12\}$  and shifting it to 3 ( $11 \cdot \text{RE.M.S.} + 1$ ) $_{\{1,12\}} - 2$ . As the sigmoid function ranges in  $[0, 1]$ , by combining it with the Mercalli point estimate of the county  $m(i)$  we obtain an Adjusted Mercalli intensity at the county, denoted by  $M_{adj}(i)$  and obtained from the following expression:

$$M_{adj}(i) = m(i) \cdot \text{SIGMOID} \left( 11 \cdot \left[ \frac{2 \cdot \overline{m}(i) \cdot s(i)}{\overline{m}(i) + s(i)} \right] - 1 \right). \quad (2)$$

The value of the Adjusted Mercalli intensity is retrieved from a surface that comprises a collection of sigmoid functions in the Mercalli scale, stretching the sigmoid according to the Mercalli point estimate.

Adjusted Mercalli intensities are at some extent noisy signals of the event, as the quality and quantity of information per county varies. However, by looking for the maximum intensity, our method is able to detect the county with the highest support and consequently, the high confidence data to provide a fast

estimation of the maximum intensity in the Mercalli scale. Then, our method for maximum intensity detection ends as follows, looking for the maximum adjusted Mercalli in the region of interest (ROI) for a given earthquake:

$$\text{MAX. INTENSITY} = \text{MAX}\{M_{adj}(i) \mid i \in \text{ROI}\}.$$

## 4 Experiments

### 4.1 Dataset

A collection of 825310 tweets was retrieved from Twitter. These tweets were collected using keywords as “quake”, “earthquake” and “seismic movement” (in Spanish). The collection comprises a year and a half of Twitter data, matching the keywords during 2016 and the first semester of 2017. From these tweets, only 2200 include the geolocation field, representing only the 0.26% of the data. The collection was posted by 309749 users where 207015 records a location field in their profiles, representing the 66.8% of the users recorded in the data.

As only a very small fraction of the tweets is geo-located, we inferred the tweet location using the user location. From the set of 207015 users with user location in our dataset, 57546 matched Chile in the country field. Then we used approximate matching to associate this field with a Chilean county. To do this we used fuzzy string matching, implemented in *Fuzzy wuzzy*<sup>1</sup>. Using an 80% of fuzzy confidence level, a total of 41885 Chilean users were mapped to Chilean counties. These users record in the dataset a total of 190249 tweets mapped to the 345 different counties in Chile.

A second database was used to conduct a cross match between Twitter and earthquake records. We used data collected by the National Seismological Center of Chile, comprising 331 records of earthquakes in Chile during the observation period, ranging magnitudes in Richter from 2.2 Mw to 7.6 Mw. In addition, the National Seismological Center of Chile provided Mercalli reports for these events along the Chilean territory.

The cross match between our tweet collection and the Mercalli earthquake records was conducted over the county field. Only county batches that record tweets until 30 min after an earthquake were studied, accounting for a total of 6790 county-Mercalli pairs with Twitter activity. A total amount of 6548 county batches unmatched a Mercalli report, indicating the presence of tweets that mention earthquake keywords in counties where it was unperceived. In summary, our Twitter-Mercalli dataset comprises 331 earthquakes with 187317 tweets distributed over 345 Chilean counties during 18 months of Twitter activity, with county-earthquake pairs separated into 6790/6548 perceived/not-perceived earthquake data batches.

From the total amount of 331 earthquakes, 264 were selected for training and exploratory issues, reserving the remaining 68 earthquakes for testing and

<sup>1</sup> Fuzzy wuzzy is a Python string matching library that uses the Levenshtein Distance to calculate differences between string sequences. It is available in: <https://github.com/seatgeek/fuzzywuzzy>.

validation tasks, representing a training/testing split of 80/20 percent. The training/testing splitting process was conducted using random sampling over earthquakes according to each Mercalli level, keeping the same proportions between intensities in training and testings folds, avoiding over/under representations of low/high intensity earthquakes in training and/or testing folds. Training/testing proportions of instances according to the maximum Mercalli intensity report of each earthquake are shown in Table 2.

**Table 2.** Training/testing instance partitions according to the maximum Mercalli intensity of each seism

| Partition | II | III | IV  | V  | VI | VII |
|-----------|----|-----|-----|----|----|-----|
| Training  | 11 | 105 | 103 | 39 | 4  | 2   |
| Testing   | 3  | 26  | 26  | 10 | 2  | 1   |
| Overall   | 14 | 131 | 129 | 49 | 6  | 3   |

#### 4.2 Estimating the Region of Interest

Training/testing county data batches accounts for 10491/2847 instances at county level. To study the problem of perceived/not-perceived earthquakes at county level, we train a 0/1 classifier. In the training fold 5021 instances accounts of the 0 class (unreported Mercalli) and 5470 for the 1 class (reported Mercalli). Training was conducted using 5 folds cross validation, using an SVM of C-SVC type for classification with a radial basis function as a kernel implemented in Weka 3.7. As the focus of the problem is the detection of the 1 class, we used cost sensitive learning, penalizing false negatives in the 1 class to maximize the recall, at the cost of a high FP rate. More learning algorithms were tested among them naive Bayes or a Multilayer Perceptron but SVM was the one with the best results, with 7325 correctly classified instances, representing in overall a 69.82% of accuracy. The detailed accuracy by class is shown in Table 3

Once model selection is evaluated, applying the model built on training instances on the testing instances, the performance of the classifier reaches 1867 correctly classified instances over a total amount of 2847 instances, achieving a 65.57% of accuracy. These results show that the classifier generalizes well,

**Table 3.** Training accuracy by class using 5-folds cross validation

| Class          | FP rate | Precision | Recall | F-measure | ROC area |
|----------------|---------|-----------|--------|-----------|----------|
| 0 (unreported) | 0.189   | 0.736     | 0.575  | 0.646     | 0.693    |
| 1 (reported)   | 0.425   | 0.675     | 0.811  | 0.737     | 0.693    |
| Weighted avg.  | 0.312   | 0.705     | 0.698  | 0.693     | 0.693    |



as overall accuracies between training and testing partitions are similar. However, what is really important is that the recall in the testing partition remains high, showing good properties in terms of predictability for the 1 class. The results disaggregated by class are shown in Table 4.

**Table 4.** Testing accuracy by class

| Class          | FP rate | Precision | Recall | F-measure | ROC area |
|----------------|---------|-----------|--------|-----------|----------|
| 0 (unreported) | 0.184   | 0.765     | 0.517  | 0.617     | 0.667    |
| 1 (reported)   | 0.483   | 0.594     | 0.816  | 0.687     | 0.667    |
| Weighted avg.  | 0.323   | 0.685     | 0.656  | 0.650     | 0.667    |

Table 4 shows that the 0/1 classifier is able to recover the region of interest for each earthquake. The results show that each region of interest is over-estimated as the low precision for class 1 shows but achieving a good coverage of the actual region as its high recall shows. To better understand how the 0/1 classifier behaves, we disaggregate matching/mismatching testing instances according to the actual level of Mercalli intensity.

**Table 5.** Matching/mismatching instances according to the actual Mercalli intensity

| Actual     | Predicted | -    | I    | II   | III  | IV   | V    | VI | VII |
|------------|-----------|------|------|------|------|------|------|----|-----|
| 0          | 0         | 790  | -    | -    | -    | -    | -    | -  | -   |
| 0          | 1         | 737  | -    | -    | -    | -    | -    | -  | -   |
| 1          | 0         | -    | 66   | 85   | 62   | 25   | 5    | -  | -   |
| 1          | 1         | -    | 130  | 234  | 351  | 198  | 65   | 95 | 4   |
| Instances  |           | 1527 | 196  | 319  | 413  | 223  | 70   | 95 | 4   |
| Error rate |           | 0.48 | 0.33 | 0.26 | 0.15 | 0.11 | 0.07 | -  | -   |

As Table 5 shows, the false negative rate is very low, and as the intensity of the earthquake increases, the error rate decreases. High intensity earthquakes (V to up) show an almost perfect performance. The thick part of this error occurs in low intensity earthquakes (III to down), which is natural for this kind of phenomena as in this part of the Mercalli scale many people do not recognize the event as an earthquake, being felt only under very favorable conditions (for instance, on upper floors of buildings). On the other hand, for the 0/1 classifier it is hard to distinguish counties where the earthquake is reported but it is unperceived. The over estimation of the region of interest will not affect the performance in the task of maximum intensity detection in the Mercalli scale of damages, as by looking for the maximum intensity in this area, the method will discard noisy data recorded in many counties.

## Detecting the maximum intensity of an earthquake in the Mercalli scale

Now we compare our method with the state of the art in the specific task of maximum intensity prediction in the Mercalli scale of damages. Results based on mean absolute error measures for each earthquake in the testing set are shown in Table 6

**Table 6.** Averaged MAE of the maximum Mercalli intensity for each earthquake.

|     | Proposal                           | Cresci <i>et al.</i> [3]            | Baseline                           |
|-----|------------------------------------|-------------------------------------|------------------------------------|
| II  | $\langle 1.00, +1.0, -1.0 \rangle$ | $\langle 2.00, +0.0, -0.0 \rangle$  | $\langle 4.12, +0.6, -0.4 \rangle$ |
| III | $\langle 0.69, +1.3, -0.6 \rangle$ | $\langle 1.05, +0.9, -1.0 \rangle$  | $\langle 3.04, +1.0, -0.8 \rangle$ |
| IV  | $\langle 1.25, +2.7, -1.2 \rangle$ | $\langle 0.20, +0.7, -0.2 \rangle$  | $\langle 1.27, +1.4, -0.8 \rangle$ |
| V   | $\langle 0.55, +0.4, -0.5 \rangle$ | $\langle 0.73, +1.2, -0.7 \rangle$  | $\langle 0.94, +0.8, -0.6 \rangle$ |
| VI  | $\langle 1.00, +1.0, -1.0 \rangle$ | $\langle 0.5, +0.5, -0.0 \rangle$   | $\langle 1.00, +1.0, -1.0 \rangle$ |
| VII | $\langle 1.00, +0.0, -0.5 \rangle$ | $\langle 11.00, +0.0, -0.0 \rangle$ | $\langle 1.00, +0.0, -0.0 \rangle$ |

As Table 6 shows, our method performs well in the specific task of maximum intensity prediction, being very competitive with the state of the art. The method of Cresci *et al.* [3] performs better than our method in IV level intensity earthquakes but at the cost of poor results on low and high energy earthquakes. The improvement of our method over the baseline is important. Note that the baseline correspond to a regression over the lexical features at county level, picking the maximum value detected in each earthquake. Our proposal applies the adjusted Mercalli at county level to improve the baseline, picking the maximum over the set of adjusted Mercalli estimates defined in Eq. 2. The results show that adjusted Mercalli variable is useful for maximum intensity detection.

## 5 Conclusion

In this paper we have proposed the method that predicts the maximum Mercalli intensity of an earthquake using social media features. The state of the art shows efforts in earthquake detection, namely where an earthquake was felt and which was its maximum intensity. Our proposal performs well in this specific task, being very competitive with the state of the art [3] in earthquake detection and maximum intensity prediction tasks using less features. However, the specific contribution of our proposal is to provide a new Mercalli estimate, named adjusted Mercalli, combining supported estimates at county level for the regression method. Our proposal discards the use of geological models of the ground or the inclusion of signals captured from spatially distributed seismographs as was done in previous work [2]. The simplicity of our method favors its application in many countries, avoiding the need to build huge networks of seismographs to track the effects of earthquakes. Our method shows that social media provides valuable information helpful for the task of Mercalli damages reports, providing accurate and fast reports of maximum Mercalli intensity.

Currently, we are extending our method to work with more features. The inclusion of time-based features helps to characterize the tweet stream (e.g. tweet interval rate), a valuable source of information for earthquake detection task. We think that these features will also be helpful in the elaboration of spatial intensity reports. In addition, we are working with network-based features (e.g. RT depth). Preliminary experiments using these features show promissory results.

At last but not least, the design of a system for early tracking of earthquake damages is the next step of this project. How to efficiently use our method to provide spatial real-time damage reports is one of our most challenging tasks in the near future. The pursuit of this goal involves efforts in data integration and visualization, among other challenging tasks for our group.

**Acknowledgements.** M. Mendoza was funded by Conicyt PIA/Basal FB0821. This work was also supported by the Millennium Nucleus Center for Semantic Web Research under Grant NC120004.

## References

1. Avvenuti, M., Cresci, S., Marchetti, A., Meletti, C., Tesconi, M.: Ears (earthquake alert and report system): a real time decision support system for earthquake crisis management. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1749–1758. ACM (2014)
2. Burks, L., Miller, M., Zadeh, R.: Rapid estimate of ground shaking intensity by combining simple earthquake characteristics with tweets. In: 10th US National Conference on Earthquake Engineering, Frontiers of Earthquake Engineering, Anchorage, AK, USA, 21–25 July 2014
3. Cresci, S., La Polla, M., Marchetti, A., Meletti, C., Tesconi, M.: Towards a timely prediction of earthquake intensity with social media. IIT TR-12/2014. Technical report. IIT: Istituto di Informatica e Telematica, CNR (2014)
4. D'Auria, L., Convertito, V.: Real-time mapping of earthquake perception areas in the Italian region from Twitter streams analysis. In: D'Amico, S. (ed.) Earthquakes and Their Impact on Society, pp. 619–630. Springer, Switzerland (2016). <https://doi.org/10.1007/978-3-319-21753-6>
5. Doan, S., Vo, B.-K.H., Collier, N.: An analysis of Twitter messages in the 2011 Tohoku earthquake. In: Kostkova, P., Szomszor, M., Fowler, D. (eds.) eHealth 2011. LNCS, vol. 91, pp. 58–66. Springer, Heidelberg (2012). [https://doi.org/10.1007/978-3-642-29262-0\\_8](https://doi.org/10.1007/978-3-642-29262-0_8)
6. Earle, P., Guy, M., Buckmaster, R., Ostrum, C., Horvath, S., Vaughan, A.: Omg earthquake! Can Twitter improve earthquake response? *Seismol. Res. Lett.* **81**(2), 246–251 (2010)
7. Robinson, B., Power, R., Cameron, M.: A sensitive Twitter earthquake detector. In: Proceedings of the 22nd International Conference on World Wide Web, pp. 999–1002. ACM (2013)
8. Sakaki, T., Okazaki, M., Matsuo, Y.: Earthquake shakes Twitter users: real-time event detection by social sensors. In: Proceedings of the 19th International Conference on World Wide Web, WWW 2010, pp. 851–860. ACM (2010)
9. Zhang, X., Shasha, D.: Better burst detection. In: Proceedings of the 22nd International Conference on Data Engineering, ICDE 2006, pp. 146–146. IEEE (2006)