ESC2016-501, 2016
35th General Assembly of the European Seismological Commission

35th General Assembly of the European Seismological Commission
4–10 September 2016 Trieste

Rapid estimation of earthquake intensity from Twitter's social sensors

Carlo Meletti (1), Stefano Cresci (2), and Maurizio Tesconi (3)

(1) Istituto Nazionale di Geofisica e Vulcanologia, Pisa, Pisa, Italy (carlo.meletti@ingv.it), (2) Institute of Informatics and Telematics, National Research Council (IIT-CNR), Pisa, Italy (stefano.cresci@iit.cnr.it), (3) Institute of Informatics and Telematics, National Research Council (IIT-CNR), Pisa, Italy (maurizio.tesconi@iit.cnr.it)

The pervasive diffusion of online social networking platforms that are filled with user-generated content, has created a digital world that largely mirrors our physical world. In the paradigm of crowdsensing, the crowd of social network users becomes a distributed network of social sensors. Information spontaneously shared by these sensors can be exploited to monitor significant events in a quasi-real-time fashion. This approach, called nowcasting, can provide rapid estimates of the consequences of unfolding events and can be particularly useful in the aftermath of natural disasters such as earthquakes. Here, the authors explore feeding predictive models by tweets (short text messages shared on the Twitter microblogging platform) conveying on-the-ground social sensors observations to nowcast the perceived intensity of worldwide earthquakes. The models build on a dataset of almost 5 million tweets exploited to compute 45 distinct predictive variables, and data about more than 7,000 globally distributed earthquakes, acquired in a semi-automatic way from USGS and serving as our authoritative ground truth. The predictive variables used in our experiments fall into four different classes according to the nature of the information they aim at capturing. The first class of variables exploits the structure of tweets (e.g., tweet length, punctuation) and their metadata (e.g., tweet geotag metadata). The second class, built upon user metadata such as a user's home account location, can help understand whether a relation exists between earthquake intensity and the spatial distribution of users reporting the earthquake. The third class of variables leverages the publication timestamp of tweets to quantify their time distribution. Emergency communications are bursty in nature, and here we want to evaluate whether the quantification of this temporal characteristic contributes to the estimation of the intensity of earthquakes. The fourth class of variables is derived from linguistic features of tweets. Variables of this class are count of specific keywords among tweets.
Overall, trained models were able to estimate earthquake intensity with a percentage MAE (Mean Absolute Error) error of 5.3% and the best performing model shows a percentage MAE error as low as 4.1%. Among the predictive variables yielding the highest contributions to intensity estimations, 5 out of 10 variables are based on the account's location field or GPS geolocation associated to tweets. This further stresses the role of geographic information as a key contributing factor for this task and demonstrates the correlation existing between the geographic distribution of messages and the intensity. Besides the small error in intensity estimates, another promising result of the study is the responsiveness of our approach. The average delay of our estimations is in the order of 100 minutes, that could be further reduced by reconfiguring the system in those cases when the accuracy of the prediction can be traded off for responsiveness.