PSCE White Paper

Using Crowdsourced Information to Improve Disaster Management: The E2MC Project

This white paper is based on a series of publications on the analysis of social media and involvement of crowdsourcing communities in supporting crisis management operations emanating from the E2mC project.

The E2mC project (Evolution of Emergency Copernicus services) intends to demonstrate the technical and operational feasibility of integrating social media analysis and crowd-sourced information within the Mapping and Early Warning Components of Copernicus Emergency Management Service (EMS) with the aim to improve the timeliness and accuracy of geo-spatial information provided to Emergency responders (National and International Organisations, government agencies or NGOS) in crisis management contexts on the first hours after an event.

To achieve this, different social media data streams (Twitter, Flickr & YouTube) and various forms of data (text, image, video) are analysed and validated by combining automatic techniques and contributions of communities (crowd) active in disaster management.

This combination will foster the engagement of a large number of people in supporting and improving crisis management.

The aims and objectives of the E2mC project are therefore:

- To design and develop a prototype of an innovative and scalable social & crowd Platform, which will act as a technological enabler of the new Copernicus Witness Service Component. This platform will allow the analysis of social media data streams and federate crowdsourcing communities to provide necessary data to emergency responders
- To demonstrate the operational usefulness of the Copernicus Witness, the new Copernicus EMS Component which intends to improve the timeliness and accuracy of geospatial information provided to emergency response authorities
- To demonstrate the tangible benefits of the S&C platform to the Copernicus EMS within realistic and operational scenarios
- To assess the quality and credibility of the information generated through the analysis of social media data or gathered through crowdsourcing mechanisms
- To define a roadmap for the full uptake of Copernicus Witness within the Copernicus EMS

Date of publication: 26/04/2018

Examples of results deriving from the E2mC project

Geolocating social media posts for emergency mapping

A geolocation framework focusing on image extraction, called CIME, was developed with the goal of improving the geolocation of posts and within the purpose of the E2mC project. This uses both the local context of the tweet (text and associated metadata) and its global context (relations to other tweets, common hashtags, retweets, mentions, etc), with images and videos extracted from the tweets and following linked social media.

Based on an example of the 2014 UK floods, the platform showed the information associated to the tweet (text and media) with a link to the original post, and information associated to its analysis, such as : geolocation type, whether it has been manually validated by the crowd and tags obtained by analysing the image.

Identifying the Footprint of Natural Disasters

Georeferenced social media posts do not only contain information about a specific event such as a natural disaster, but also information about other topics that are not necessarily related to the event. A common approach to extract relevant social media posts is to apply algorithms that filter posts according to predefined keywords. However, this approach is limited to the identification of posts that contain the predefined keywords, which may often change due to the dynamic nature of social media networks (for instance, through the generation of ad-hoc hashtags), and it requires time for preparation and maintenance of keyword sets for multiple languages.

The E2mC project develops a workflow based on a topic modelling algorithm to detect latent semantic topics in the social media corpus that needs little to no a-priori information about the underlying structure of the corpus. The results of multiple use cases, including the Napa Valley Earthquake 2014 or Hurricane Harvey 2017, show that distinct disaster-relevant topics can be identified. These topics are then used for semantically classifying the social media posts. The accuracy of the topic model is assessed using pre-labelled social media posts and is used to determine the parameter values for optimum performance of the algorithm.

In a next step, the disaster-relevant tweets are clustered in a spatial hot spot analysis to investigate if a local neighbourhood is affected by the disaster with a defined confidence interval. This allows authorities to identify areas that have been affected by a disaster in a short period of time after the natural disaster occurred. Apart from this output, the results can be used for validating/complementing results of other data sources like remote sensing-based information products.

 Date of publication: 26/04/2018

Architecture for Leveraging User-Generated Data in Event Detection and Disaster Management

The general architecture of the EMS Witness system comprises components for data acquisition, storage, management and analysis and graphical user interfaces for the EMS operators and the crowd.

The workflow is triggered either by the Early Warning System (EWS) component in case of a flood or manually by an EMS operator. Various parameters have to be selected for the different analysis components and the crawlers before start. In case of a flood the EWS component provides information including probability, timing, location and magnitude of an event. The EMS operator chooses the parameters for Witness workflow and can use the parameters obtained by EWS. When the EMS operator starts the Witness workflow, crawlers begin to collect social media posts. The collected data are stored via a centric RESTful API with six resources e.g., user, event, keyword, post, image and tag in a relational database model.

Overall, the EMS portfolio is enriched by a new product that increases the production capability of the system. The citizens are involved in the process of helping authorities during a disaster that increases the awareness of Copernicus and enlarges the service usage. In the mid-term future, the architecture will be integrated into the EMS workflow.

Improving Rapid Mapping with Social Network Information

Associating a location to images and video can bring significant benefits to rapid mapping but this association is often challenging as most social media items are not geotagged and the location associated to a content does not necessarily coincide with the location of the post. This means that a possible solution will be to infer the location of an image based on the location associated with the related post, as originally geotagged content is too limited for rapid mapping purposes and many messages in social media contain implicit references to names of places.

Using the August 2016 earthquake in Italy as a case study, less than 4% of tweets had locations that could be disambiguated by building a local context – the algorithm therefore needs to be extended to connect tweets based on implicit interactions among messages. For example, tweets that share a similar content or belong to the same conversation.

A global context for each post is built based on its “neighbours” and their respective locations, which allows to overcome the limits of the local context and to improve the disambiguation of the location – in the case study of the 2016 Italy earthquake, this was observed in more than 20% of the tweets.

Incorporating automated analysis speeds up the process and assists in improving the accuracy of the data gathered. Information gathered from Social Media using such automated process can then be validated using crowdsourcing and by eliminating false positives. This can be done through volunteers who answer whether a specific information is relevant or not. Volunteers can also contribute to reduce uncertainty regarding the geo-localization of an image and find a more accurate location.

For more information, please look into the following related publications:

Havas, C., Resch, B., Francalanci, C., Pernici, B., Scalia, G., Fernandez-Marquez, J. L., … Rüping, S. (2017). E2mC: Improving Emergency Management Service Practice through Social Media and Crowdsourcing Analysis in Near Real Time. *Sensors*, *17*(12).

Resch, B., Usländer, F., & Havas, C. (2017). Combining machine-learning topic models and spatiotemporal analysis of social media data for disaster footprint and damage assessment. *Cartography and Geographic Information Science*, 1–15.

The aforementioned publications, as well as other elements of information, can be found on the E2mC website at: https://www.e2mc-project.eu/results

