SMDRM - Social Media for Disaster Risk Management

Valerio Lorini[1], Peter Salamon[1], and Carlos Castillo[2]
[1]European Commission, Joint Research Centre, Ispra, Italy
[2]Universitat Pompeu Fabra, Barcelona, Spain

Social media has been described as a form of distributed cognition, a mechanism for understanding a situation using information spread across many minds. The interactions among people in social media are a form of collective intelligence, as they allow people to make sense of a developing event collectively. Social media users can contribute to creating a "sensor" for citizen-generated data that modelling or monitoring systems can assimilate during a crisis. Gaining situational awareness in a disaster is critical and time-sensitive. Social media presents the possibilities of a growing data source to help improve response in the early hours and days of a crisis. However, social media platforms may not provide the functionality of summarising the information that is useful for crisis responders.

SMDRM is a software platform that streamlines the processing of text and images extracted from Twitter in near real-time during a specific event. The data is collected using a combination of keywords and locations based on daily forecasts from the early warnings systems of the Copernicus Emergency Management Service such as EFAS, GloFAS and EFFIS (emergency.copernicus.eu) or triggered manually in case of earthquakes or not-forecasted events. Text is automatically "annotated" using a binary multilingual classifier trained on 12 languages and extended with multilingual embeddings. Simultaneously, a multi-class convolutional neural network labels relevant images for floods, storms, earthquakes and fires. The information that doesn't embed coordinates is geolocated in a two-step algorithm where location candidates are first selected using a multilingual named-entity recognition tool and then searched on available gazetteers. The last step of the SMDRM data processing is the aggregation of relevant information in spatial (administrative areas) and temporal (daily) units. Social media activity about an event can finally be distributed as a data map and visualised on a map server and made available to users.

SMDRM could offer timely information useful for reducing the hazard models' uncertainty and providing added-value information such as reports or descriptions of the situation on the ground or in the vicinity. Other stakeholders, such as research groups could access new data to complement the ones extracted from traditional sensors or earth observation.

The platform can adapt to cope with the varying workload as it uses scalable software containers. If the number of tweets is higher during an impactful event, the platform can use more containers to annotate them. SMDR code, together with the tens of thousands of annotated social media messages used for training its models, will be released as an open-source platform whose modules can be adapted to serve other research projects. We describe the platform's architecture and implementation details, and two use cases where images and text were used as a use-case to

test the system's modules.