

The challenge of collecting and analyzing information from citizens and social media in emergencies: the Crowd4SDG experience and tools^{*}

Barbara Pernici¹[0000-0002-2034-9774], Carlo Bono¹[0000-0002-5734-1274], Jose Luis Fernandez-Marquez²[0000-0002-0254-0794], and Mehmet Oğuz Mülâyim³[0000-0002-3993-5597]

¹ Politecnico di Milano, DEIB, Via Giuseppe Ponzio, 34, 20133 Milano, Italy
`name.lastname@polimi.it`

² University of Geneva, Geneva, Switzerland
`joseluis.fernandez@unige.ch`

³ Artificial Intelligence Research Institute (IIIA-CSIC), Cerdanyola del Vallès, Spain
`oguz@iiia.csic.es`

Tutorial Abstract

Every year more than 150 million people worldwide are affected by natural disasters. As declared by the United Nations Office for the Coordination of Humanitarian Affairs, “The first 72 hours after a disaster are crucial; response must begin during that time to save lives”. Social media has been demonstrated to be a potential data source to provide actionable data just as a disaster happens and develops, thus allowing emergency responders to better coordinate their activities. However, social media data also presents many challenges regarding data quality and geolocation (i.e., the geographical location of a post). Over the years, several technologies enabled the retrieval and processing of high volumes of data, with artificial intelligence often employed as a replacement for human intelligence for data classification tasks. Nevertheless, the need to deliver high-quality results within a critical response time is still a major challenge.

In this tutorial, we will see how crowdsourcing assisted by artificial intelligence can make a significant contribution, especially where critical thinking and decision making are needed, in extracting valuable information from unconventional data sources. The tutorial will introduce the basics for extracting and analyzing information from social media, with a specific focus on retrieving images in an emergency after a natural disaster. We will provide the basics about social media crawling and analysis. A specific focus will be given to fine-grained geolocalization of tweets and the combination of AI and crowdsourcing to filter relevant images and confirm or improve geolocations, which are needed to deliver high-quality information. Our experiences with social media

^{*} This work was funded by the EU H2020 project Crowd4SDG “Citizen Science for Monitoring Climate Impacts and Achieving Climate Resilience”, #872944.

analysis (e.g., [5]), geolocalization (e.g., [1, 6]), and crowdsourcing (e.g., [2, 3]) obtained in a recently concluded H2020 project E2mC (Evolution of Emergency Copernicus services) [4] and in the on-going H2020 project Crowd4SDG (Citizen Science for Monitoring Climate Impacts and Achieving Climate Resilience, <https://crowd4sdg.eu>) will be illustrated.

The objective of the tutorial is to provide an introduction and hands-on experience in some of the tools available in the field of emergency information systems. In particular, we focus on the tools that enable the search and analysis of social media posts, mainly on Twitter but also on other social media. The analysis of posts includes approaches for selecting relevant images based on image contents and text analysis techniques for information extraction. We also show how we could leverage citizen scientists by setting up a crowdsourcing environment, based on the PyBossa open-source platform (<https://pybossa.com>), and we demonstrate how we evaluate the quality of crowdsourcing results. We will also discuss the methods and processes for using such tools in a sudden emergency to gather different types of information to support first responders and decision makers.

The tutorial is intended for participants who represent organizations looking for emergency data, who can benefit from collective intelligence, especially where there is a data gap in their research using traditional data sources, but are skeptical in reliability of this kind information; communities and agencies looking for tools to analyze the data; individuals interested in learning about available tools that can enrich and ensure reliability and usability of data obtained from social media.

References

1. Bono, C., Pernici, B., Fernandez-Marquez, J.L., Shankar, A.R., Mülâyim, M.O., Nemni, E.: TriggerCit: Early Flood Alerting using Twitter and Geolocation – a comparison with alternative sources (2022). <https://doi.org/10.48550/arxiv.2202.12014>
2. Cerquides, J., Mülâyim, M.O.: crowdanalysis: A software library to help analyze crowdsourcing results (2022). <https://doi.org/10.5281/zenodo.5898579>
3. Cerquides, J., Mülâyim, M.O., Hernández-González, J., Ravi Shankar, A., Fernandez-Marquez, J.L.: A Conceptual Probabilistic Framework for Annotation Aggregation of Citizen Science Data. *Mathematics* **9**(8) (2021). <https://doi.org/10.3390/math9080875>
4. Havas, C., Resch, B., Francalanci, C., Pernici, B., Scalia, G., Fernandez-Marquez, J.L., Van Achte, T., Zeug, G., Mondardini, M.R.R., Grandoni, D., et al.: E2mC: Improving emergency management service practice through social media and crowdsourcing analysis in near real time. *Sensors* **17**(12), 2766 (2017). <https://doi.org/10.3390/s17122766>
5. Negri, V., Scuratti, D., Agresti, S., Rooein, D., Scalia, G., Shankar, A.R., Marquez, J.L.F., Carman, M.J., Pernici, B.: Image-based Social Sensing: Combining AI and the Crowd to Mine Policy-Adherence Indicators from Twitter (2020). <https://doi.org/10.48550/ARXIV.2010.03021>
6. Scalia, G., Francalanci, C., Pernici, B.: CIME: Context-aware geolocation of emergency-related posts. *GeoInformatica* **26**(1), 125–157 (2021). <https://doi.org/10.1007/s10707-021-00446-x>