Crowd4**SDG**

Study on timeliness of disaster-related data

Barbara Pernici

Politecnico di Milano

November 10, 2022

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 872944

Short introduction

Barbara Pernici

Background: Electrical engineering + Computer Science

Main research interests: information systems design, Information and Service quality, Emergency information management

Crowd4SDG European H2020 project https://crowd4sdg.eu

Citizen Science for Monitoring Climate Impacts and Achieving Climate Resilience

Call: H2020-SwafS-2018-2020 (Science with and for Society)
May 2020 - April 2023

Focus on SDG 13 (Climate Action) + SDGs 11, 5, 16

Crowd4SDG

Motivation and problem formulation

Motivation

- Assess how **social media** data can complement or corroborate spatio-temporal descriptions of large-scale events
- Focus on **natural disasters**: high **impact** in terms of human and economic loss
- Applications: ubiquitous monitoring, trigger demanding processes

Problem formulation

- Focus on **event detection**
- Investigate if an end-to-end, automated detection system for natural disasters can be designed from tweets
- Stringent quality, latency and computational **constraints**
- **Data quality:** accuracy, completeness, consistency, **timeliness**
- Multilingualism (e.g. language-transferable approaches)

Data collection from social media

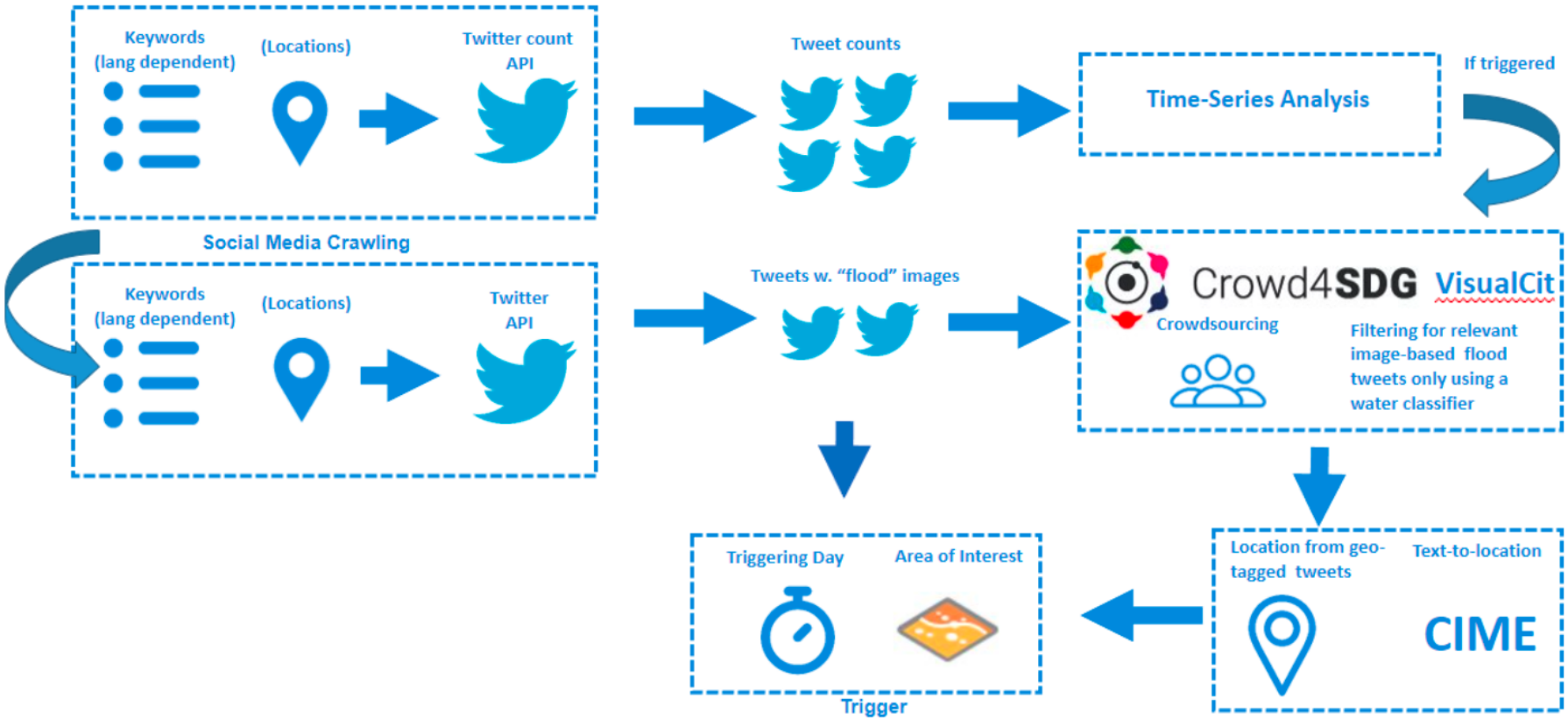
1) Studying the extraction of relevant images and info from social media to build indicators on a disaster in early stages

2) Focus on automated event detection with Machine Learning

Two aspects:

- Citizen Science
  "passive CS", possibly also with active citizens' contributions in data preparation
- Needs preprocessing to guarantee quality of the data ("fit for use")

Timely collection of relevant posts analyzing text and images



(a) Removed items

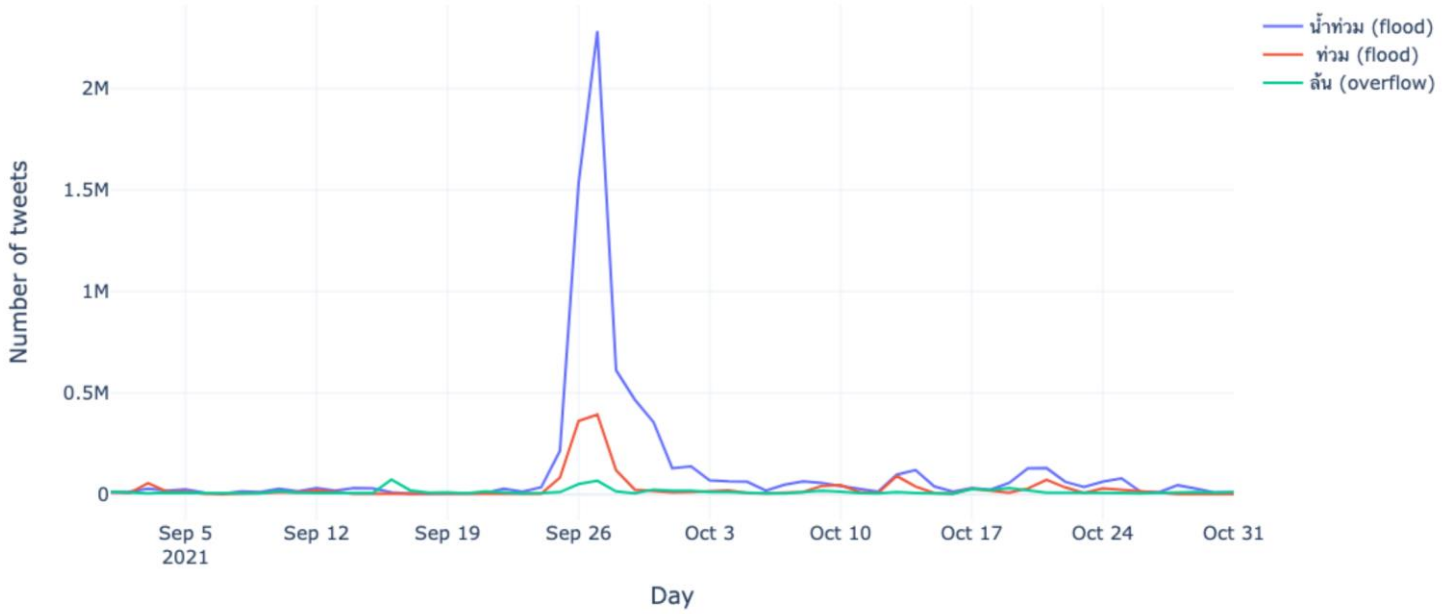
(b) Kept items

Bono et al, TriggerCit, ISCRAM, 2022

Crowd4SDG

Experimental results

The event onset of large-scale events can be clearly identified



Tweet counts for seed dictionary entries
Thailand case study September 2021

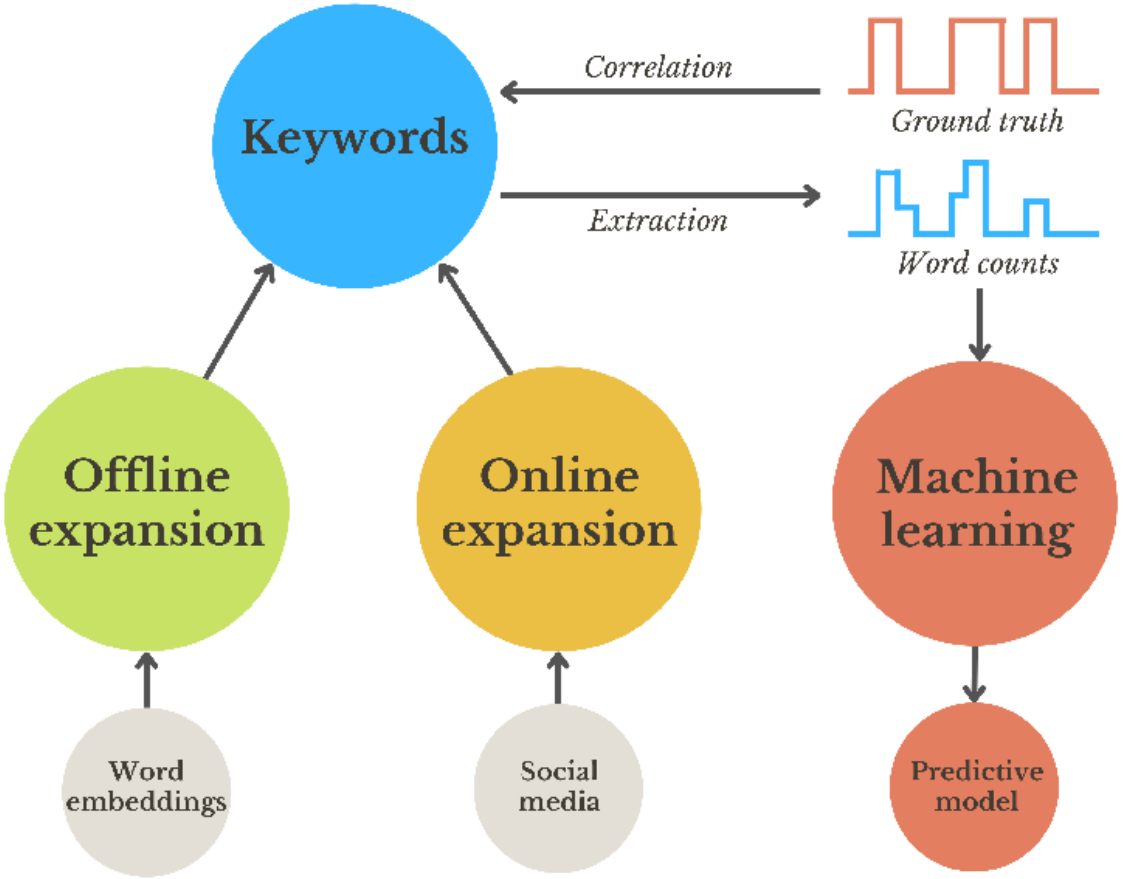
Data	Count
All tweets, September 26-27	4’145’447
No retweets	66’868
Containing images	6’292
Native Twitter locations	227
Overall images	8’774
Passed VisualCit filters	3’056
Places geolocated by CIME	1’671

Data volumes through processing steps

Bono et al, TriggerCit, ISCRAM, 2022

Detecting events with self-configuring dictionaries

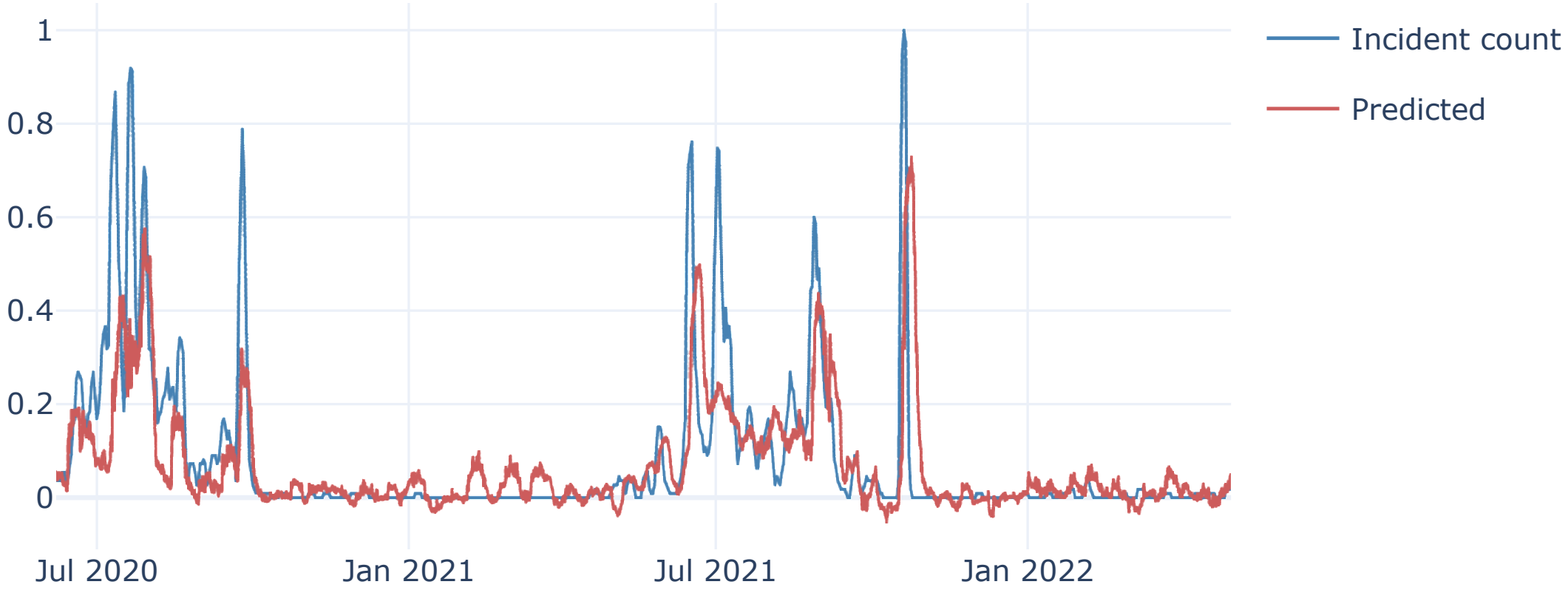
- An agnostic method for building a tailored language- and event-centered **lexicon**
- A supervised learning setup for the actual event **triggering**



Bono et al., IFIP WG5.15 Working Conference ITDRR, Oct. 2022

Crowd4SDG

Event detection based on word counts



Predicted incident count, normalized
Nepal case study (July 2020-Jan. 2022)

Bono et al., IFIP WG5.15 Working Conference ITDRR, Oct. 2022

Considerations on timeliness

- Evaluation of results and
- Supervised machine learning

both require a **ground truth**

When was the **onset** of the event?

Considerations on timeliness

Activation-based

Manual

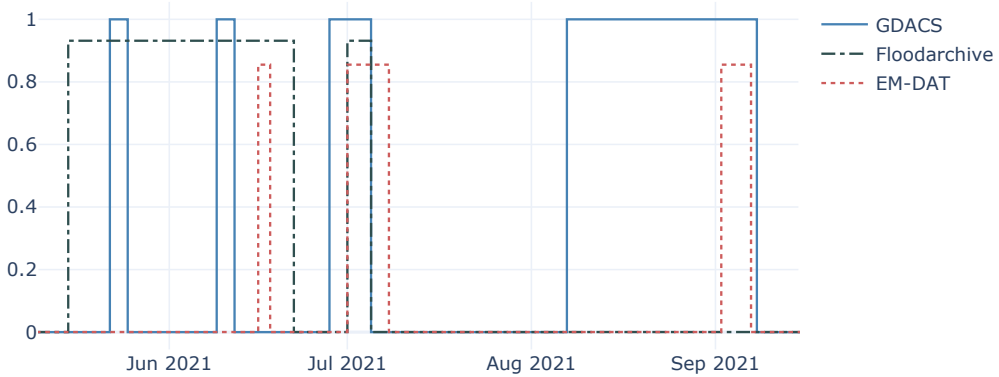
Forecast

Manual

Dates / Countries	Thailand	Nepal
UNOSAT activation	28/09/2021	30/06/2021
Copernicus EMS activation	None	None
GDACS Disaster Alerts	27/09/2021 (Green Alert)	28/06/2021 (Green Alert)
GloFAS	24/09/2021	28/06/2021
FloodList reported news	27/09/2021	04/07/2021
TriggerCit	26/09/2021	02/07/2021

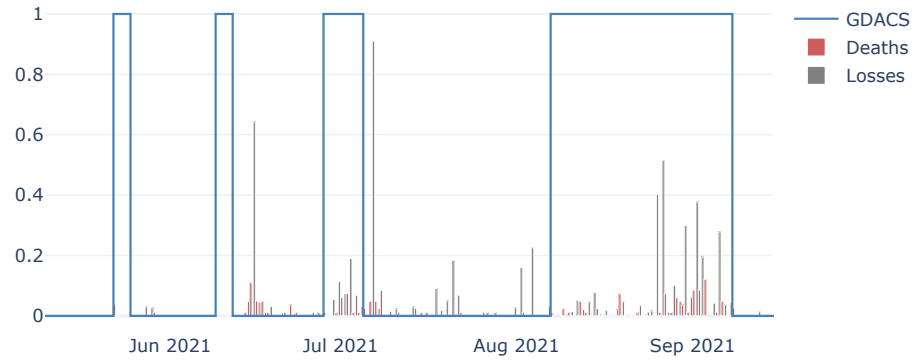
Time of the events as reported from different sources

Bono et al, TriggerCit, ISCRAM, 2022



Different event dates reported on the same time frame
https://www.emdat.be
https://floodobservatory.colorado.edu

Accuracy of temporal data



Incident data plotted against GDACS events
http://drrportal.gov.np

Crowd4SDG

Lessons learned

- Automated event **triggering from social media posts** can provide timely data
- Supervised approaches require good quality training data on the **timeliness** dimension.
  Authoritative sources such as incident data from previous incidents can be a reliable source
- Similar considerations apply to the **spatial dimension**

References

- Carlo Bono, Barbara Pernici, Jose Luis Fernandez-Marquez, Amudha Ravi Shankar, Mehmet Oğuz Mülâyim, Edoardo Nemni, TriggerCit: Early Flood Alerting using Twitter and Geolocation - a comparison with alternative sources, 2022, ISCRAM 2022, preprint: https://doi.org/10.48550/arXiv.2202.12014
- Bono, Carlo A., Mehmet Oğuz Mülâyim, and Barbara Pernici. Learning early detection of emergencies from word usage patterns on social media. IFIP WG5.15 Working Conference on Information Technology for Disaster Risk Reduction ITDRR, Norway, Oct. 2022 link
- V. Negri, D. Scuratti, S. Agresti, D. Rooein, G. Scalia, J. L. Fernandez-Marquez, A. Ravi Shankar, M. Carman and B. Pernici, Image-based Social Sensing: Combining AI and the Crowd to Mine Policy-Adherence Indicators from Twitter, ICSE, Track Software Engineering in Society, May 2021
- G. Scalia, C. Francalanci, B. Pernici, CIME: Context-aware geolocation of emergency-related posts, Geoinformatica, Issue 1/2022

Thank you!

www.crowd4sdg.eu