2nd International Conference
Citizen Observatories for natural hazards and Water Management
Venice, 27-30 November 2018

SOCIAL MEDIA OBSERVATIONS FOR FLOOD EVENT MONITORING IN ITALY OVER A ONE-YEAR PERIOD

Andreadis S.[1], Gialampoukidis I.[1], Fiorin R.[2], Lombardo F.[2], Norbiato D.[2], Karakostas A.[1], Ferri M.[2], Vrochidis S.[1] & Kompatsiaris I.[1]

[1] Information Technologies Institute, Centre for Research and Technology Hellas, Thessaloniki, Greece – e-mail: {andreadisst, heliasgj, akarakos, stefanos, ikom}@iti.gr
[2] Autorità di Bacino Distrettuale delle Alpi Orientali, Venice, Italy – e-mail: {roberto.fiorin, francesca.lombardo, daniele.norbiato, michele.ferri}@adbve.it

KEY POINTS

- *We collect Twitter posts to monitor flood events in Italy during a period of one year.*
- *We analyse the statistics of the collected content.*
- *We extract word clouds, popular keywords and we demonstrate events in a one-year timeline.*
- *A large part of the collection has been annotated by Italian experts on the relevance or not to flood events.*

1 INTRODUCTION

Nowadays, water authorities take measures aiming at reducing risks by minimizing the possible damages effects and losses that may result from a flood event. Monitoring a flood event requires not only weather, sensor, Earth Observation data, and messages from first responders, but also social data from social media platforms. The constantly growing popularity of microblogging, and particularly of the Twitter platform, has led to a collaborative network of news distribution between interested users (Bruns et al., 2012b). At the same time, organizations have developed a new communication channel with their public using Twitter (Saffer et al., 2013). The wide adoption of Twitter by both individuals and authorities can also be reflected in the case of natural disasters (Bruns & Burgess, 2014) and the large amount of posts generated during such events has motivated the research community to investigate on how this data can be proven useful for crisis management. Focusing on real flood incidents, Bruns et al. (2012a) and Takahashi et al. (2015) conclude that Twitter has a leading role in crisis communication due to the timely dissemination of critical information. Regarding the analysis of tweets that are produced during floods, Saravanou et al. (2015) use geotagging and visual analytics tools to discover flood-stricken areas, Vieweg et al. (2010) employ information extraction strategies to detect the intention of a tweet, i.e. an advice, an evacuation order, etc., and Cheong & Cheong (2011) perform social network analysis techniques to identify active players and how they affect the sharing of crisis information. Other works, e.g. Kongthon et al. (2012) and Moumtzidou et al. (2018), try to estimate whether text or images from tweets are relevant to floods, while Reuter & Schröter (2015) examine the retweet ratio to mine related tweets. In this work we present novel analytics and flood event detection methods from social media streams on an Italian case study. Our target is to exploit actual tweets in order to detect if and when a flooding event is occurring, but also to reveal more insights on the event. This will enhance the flood situational awareness and support the authorities' preparedness.

2 METHODOLOGY

In order to accumulate a large number of social media data that refer to a specific topic, i.e. floods in Italy in our study case, we have utilized Twitter's Streaming API[1]. This service grants real-time access to public data flowing through Twitter that contain any keyword of a predefined set. Our list of keywords can be seen in Table 1, along with their English translation for a better understanding. The selected terms focus mainly on flood events in Italy. This crawling procedure lasted from April 01, 2017 until March 31, 2018, resulting to a wide collection of related tweets over a complete year.

[1] https://dev.twitter.com/streaming/overview

Table 1. Terms used to track relevant tweets

Keywords	English translation
alluvione	flood
alluvionevicenza	flood Vicenza
allagamento	flooding
bacchiglione	Bacchiglione
fiumepiena	full river
allertameteo	weather alert
sottopassoallagato	underpass flooded
alluvione2017	flood 2017
allertameteovicenza	weather alert Vicenza
esondazione	flooding

3 RESULTS

After one year of crawling tweets that concern flooding incidents in Italy, the collection counts 43,352 tweets. It is anticipated that part of this data will also include irrelevant posts, thus we proceeded with human annotation, e.g. users that tag tweets as relevant or not. This feedback also serves the development of automatic mechanisms to distinguish related posts, where ground-truth annotation is required for building robust machine learning algorithms that can automatically filter out irrelevant social media posts. Figure 1 displays three time series regarding the number of the crawled tweets per each date of the year; the first refers to the total set, the second to a total of 16,749 annotated tweets by Italian experts and the third to the 4,701 tweets that were marked as relevant. The number of annotated posts is always larger than the relevant ones, which means that there are indeed many irrelevant items in the collection and, therefore, the necessity of a classification method is highlighted. However, solely by examining the uncharacterized data, it is evident that two important events were detected throughout the last year: one on September 10, 2017 and one on November 04, 2017. In fact, it can be confirmed that they connect to the Livorno floods[2] and the anniversary of the 1966 flood of the Arno in Florence[3], respectively.

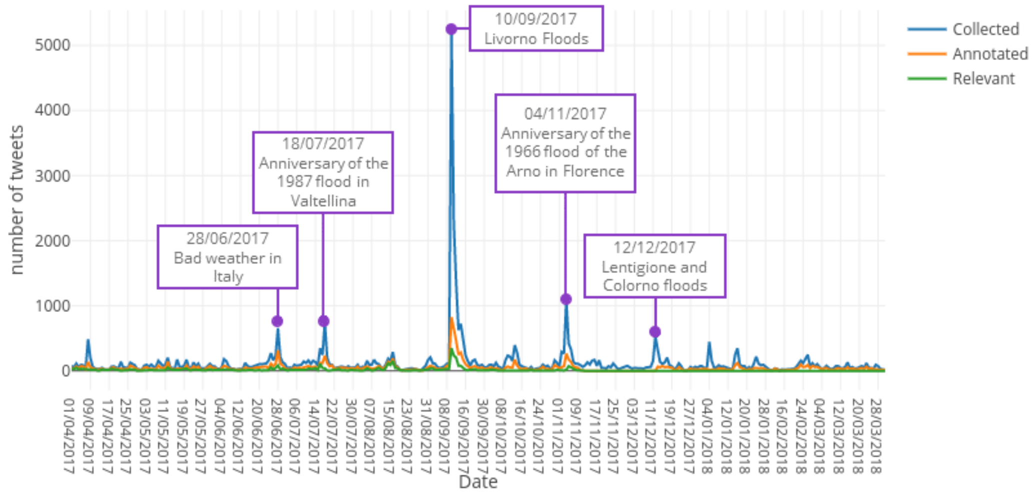

Figure 1. Fluctuation of the number of tweets during last year, grouped as collected, annotated and relevant

The content of the tweets was further analysed (e.g. removal of punctuation, URLs, and stop words) in order to discover the words that are most frequently used. The top ten non-location terms and the top ten mentioned locations are gathered in Table 2, together with their number of appearances and their English translation, if needed. Amongst the most repeated non-location words, there is only one term unrelated to

[2] https://en.wikipedia.org/wiki/2017_Livorno_floods
[3] https://en.wikipedia.org/wiki/1966_flood_of_the_Arno

floods (the music band Benji & Fede), while the most frequent locations are all places in Italy, including the country itself. These lists are also illustrated as word clouds in Figure 2.

Table 2. Most frequently mentioned terms inside the collected tweets, separated in locations and non-locations

	Non-locations			Locations	
#	**Appearances**	**Word**	**English translation**	**Appearances**	**Word**
1	30820	alluvione	flood	9542	Livorno
2	5418	colpire	to hit	1869	Roma
3	5388	maltempo	bad weather	1532	Firenze
4	4915	vittima	victim	999	Italia
5	4774	allagare	to flood	968	Genova
6	4100	allertameteo	weather alert	932	Valtellina
7	3751	famiglia	family	878	Toscana
8	3484	pensiero	thought	563	Milano
9	3407	tenere	to hold	414	Sardegna
10	3360	benjiefede	Benji & Fede (band)	386	Parma

Figure 2. Top ten most recurrent words and locations, in the form of word clouds

Using the top five non-location concepts and the top five locations, we have examined their frequency during the complete period of crawling. The time series of the number of appearances of each word are shown in Figure 3 and in Figure 4. The higher usage of the words “bad weather” and “victim” on September 10 compared to November 4 can be interpreted as the difference between an occurring flood and an anniversary. Furthermore, the increase on the appearances of “Livorno” and “Firenze” (Florence) on the same dates agrees with the afore-mentioned events.

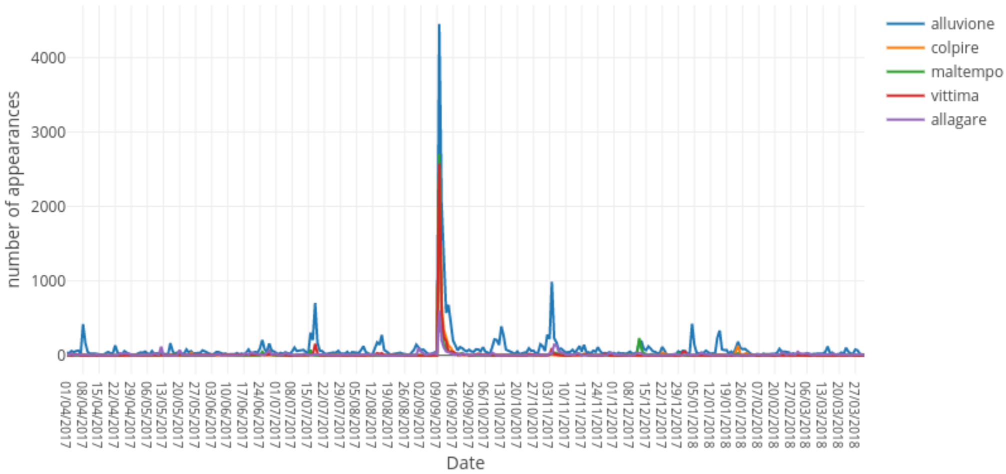

Figure 3. Appearances of the top five most used non-location words during last year

4 CONCLUSION

This paper focuses on information that can be extracted by collecting social media data about a particular topic. After gathering thousands of Twitter posts during the period of one year for floods in Italy, we have observed that the increase of the number of tweets can be perceived as event detection. Moreover, studying the number of appearances of words can indicate more details on events, such as the location where they take place or their severity level. Our work contributes to the flood management procedures before the crisis and can be integrated in relevant flood management and decision support systems. In the future, we intend to investigate whether our techniques regarding the storage and usage of tweets comply with the General Data Protection Regulation (https://www.eugdpr.org/) on data protection. We also plan to develop an automatic classification method that will estimate if a posted text is related to floods, by exploiting the annotation that has already been performed by Italian experts.

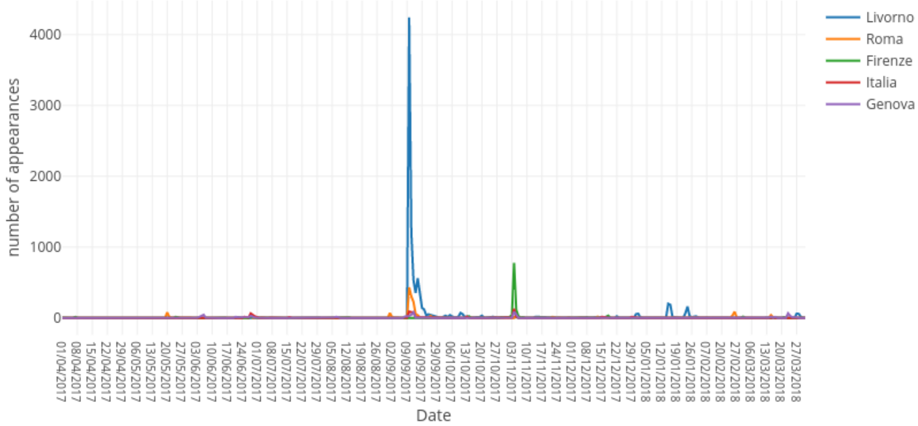

Figure 4. Appearances of the top five most used locations during last year

ACKNOWLEDGEMENTS

This work has been supported by the EC-funded projects beAWARE (H2020-700475) and EOPEN (H2020-776019).

REFERENCES

Bruns, A. & J. Burgess (2014): Crisis communication in natural disasters: The Queensland floods and Christchurch earthquakes. In: *Twitter and society*. Peter Lang, pp. 373–384.

Bruns, A., J.E. Burgess, *et al.* (2012):(a): # qldfloods and@ QPSMedia: Crisis communication on Twitter in the 2011 south east Queensland floods.

Bruns, A., T. Highfield, *et al.* (2012):(b): Blogs, Twitter, and breaking news: The produsage of citizen journalism. *Produsing theory in a digital world: The intersection of audiences and production in contemporary theory* (**80**)2012: 15–32.

Cheong, F. & C. Cheong (2011): Social Media Data Mining: A Social Network Analysis Of Tweets During The 2010-2011 Australian Floods. *PACIS* (**11**): 46–46.

Kongthon, A., C. Haruechaiyasak, *et al.* (2012): The role of Twitter during a natural disaster: Case study of 2011 Thai Flood. In: *Technology Management for Emerging Technologies (PICMET), 2012 Proceedings of PICMET'12:* IEEE, pp. 2227–2232.

Moumtzidou, A., S. Andreadis, *et al.* (2018): Flood Relevance Estimation from Visual and Textual Content in Social Media Streams. In: *Companion of the The Web Conference 2018 on The Web Conference 2018*. International World Wide Web Conferences Steering Committee, pp. 1621–1627.

Reuter, C. & J. Schröter (2015): Microblogging during the European floods 2013: What Twitter may contribute in German emergencies. *International Journal of Information Systems for Crisis Response and Management* (**7**)1: 22.

Saffer, A.J., E.J. Sommerfeldt, *et al.* (2013): The effects of organizational Twitter interactivity on organization–public relationships. *Public Relations Review* (**39**)3: 213–215.

Saravanou, A., G. Valkanas, *et al.* (2015): Twitter Floods when it Rains: A Case Study of the UK Floods in early 2014. In: *Proceedings of the 24th International Conference on World Wide Web*. ACM, pp. 1233–1238.

Takahashi, B., E.C. Tandoc Jr, *et al.* (2015): Communicating on Twitter during a disaster: An analysis of tweets during Typhoon Haiyan in the Philippines. *Computers in Human Behavior* (**50**): 392–398.

Vieweg, S., A.L. Hughes, *et al.* (2010): Microblogging during two natural hazards events: what twitter may contribute to situational awareness. In: *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, pp. 1079–1088.