Huerta, Schade, Granell (Eds): Connecting a Digital Europe through Location and Place. Proceedings of the AGILE'2014 International Conference on Geographic Information Science, Castellón, June, 3-6, 2014. ISBN: 978-90-816960-4-3

Exploring twitter georeferenced data related to flood events: an initial approach

Maria Antonia Brovelli
University/Institute
Via Natta 12/14.
Como, Italy
maria.brovelli@polimi.it

Giorgio Zamboni
Politecnico di Milano
Via Valleggio 11
Como, Italy
giorgio.zamboni@polimi.it

Carolina Arias Muñoz
Politecnico di Milano
Via Valleggio 11
Como, Italy
carolina.arias@polimi.it

Alexander Bonetti
Politecnico di Milano
Via Valleggio 11 Como,
Italy
alexander.bonetti@mail.
polimi.it

Abstract

The purpose of this research-in-progress paper is to present an initial data exploration of twitter georeferenced data related to flood events, in order to determine the potential of these type of data in flood damage assessment. Data exploration aimed at determine: (1) if the features of information associated to a flood event can be generated using Twitter, (2) What are the most frequently used hash tags related to flood events? and (3) If it is possible to detect post flood information from Twitter messages. We developed a script to gather data using the Twitter Search API. The data collected gave an idea of the main feature information, spatial distribution and better search strategies.

Keywords: twitter data , flood event, damage assessment

1 Introduction

Twitter data have been playing also great role on disaster management in the last years, where research has focused mostly on disaster respond than disaster relief or post event[1]. Twitter research is quickly evolving to include more in-depth studies of social interactions and message content, that involves mostly Twitter data exploration during emergencies [2] as well as tools and methods for capturing Twitter data during and after disasters events [3]. The studies have shown that Twitter data can clearly be useful in coordinating resources and efforts and also in preparing and planning for disaster relief.

This paper presents a first step at determining the potential of Twitter georeferenced data for flood damage assessment, focusing on the following initial analysis questions: (a) What are the features of information that can be generated using Twitter associated to a flood event? (b) What are the most frequently used hash tags related to flood events? and (c) Is it possible to detect post flood information from Twitter messages?. To solve those questions, we decided to gather twitter messages with the scope of finding paths for following data analysis.

2 Twitter data collection and exploration

For twitter data collection we developed a script using PHP as a language and the Twitter Search API[1] (GET search/tweets). The script is instructed to search for a specific case-insensitive keyword, to execute queries to Twitter and to save the results in a PostgreSQL database.

The term alluvion was used for pulling Italian tweets in a half/full width forms (i.e. alluvione, alluvionato); the terms flood, inundacion, inundation and hochwasser representing the main official languages of the European Union, were also use in the query to obtain Italian tweets also in this languages. We collected Twitter data from all over the world for 69 days, from a total of 8242 unique tweet users. The results can be seen on table 1: the percentage of georeferenced tweets is very low – no more than 3% -. Among the georeferenced tweets, the ones that include an hyperlink can give very useful information, but they represent an even lower amount of the total tweets collected. The features of information contained in the hyperlinks were: online news articles, photos and videos of flooded areas, or other type of information such as maps or link to maps applications.

Table 1. Overall results

				Georeferenced Messages					
Keyword	# Messages	# GeoMessages	% GeoMessages	http Messages	Video	Photo	News	Others	No relevant info
flood	336462	7903	2,35	2192	10	643	1058	0	481
alluvion	9964	121	1,21	43	2	8	22	0	11
inundacion	6632	104	1,57	49	0	8	7	28	6
inondation	3089	91	2,95	39	0	21	14	0	4
hochwasser	565	8	1,42	4	0	1	1	0	1

Source: the authors

122 tweet messages published on Italian territory were collected, 18 of them corresponding to the word *flood* and 3 corresponding to the word *inundación*. Distribution of tweets varied over time by both distance and geographical locations thought Italy, showing a non-strong correlation between tweets and places were a flood event have occurred. The most recent big floods events near the experiment time window (26/11/2013 till 03/02/2014): were on Province of Modena (15th of Jannuary 2014) and in Sardegna island (18th of November 2013) Sardegna, but the tweets clustered on cities such as Cagliari (5 tweets), Roma (8 tweets), Milano (4 tweets). The most frequently used hash tags used by italian

[1] https://dev.twitter.com/docs/api/1.1/get/search/tweets

users were: #siamobloccati (we are blocked), #aiuto (help), #disastro (disaster), #alluvione (flood), #rischio (risk), #Sardegna, #SOS, #river.

Post flood information from Twitter messages was found mostly in the forms of news and photos that showed damage to buildings or infrastructure. No water heights information was found.

3 Conclusions and paths for future work

The data collected until now gave us an idea of the main feature information, spatial distribution and better search strategies, but further analysis are needed to determine if it is possible to detect post flood information to give a better flood event scenario.The next logical step seems to execute other queries using the most common hash tags found, but in this case using other Twitter API resources like *Streaming* or *Trends* that retrieve non only relevant tweets, but all free available tweets. Regarding data analysis can be useful to understand if there are interactions between users to find citizens networks; or to look at the message content to identify emergent themes and categories using the available data analysis software.

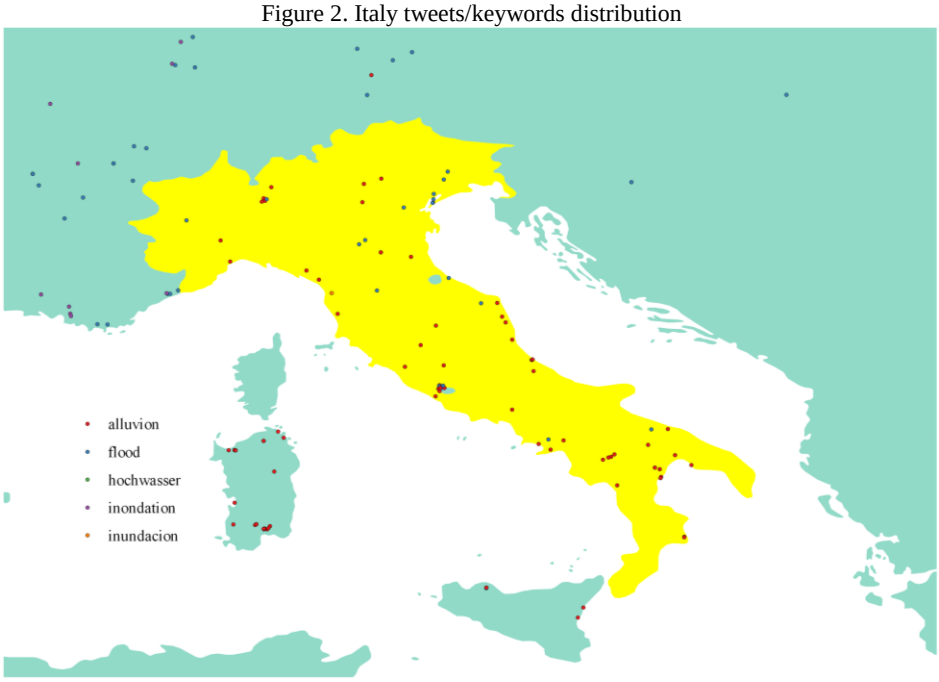
4 References

[1] M. Goodchild. Crowdsourcing geographic information for disaster response: a research frontier. *International Journal of Digital Earth*3.3: 231-241, 2010.

[2] L. Palen. Twitter based information distribution during the 2009 Red River Valley flood threat. Bulletin of the American Society for Information Science and Technology, 2010, 36.5: 13-17.

[3] A. Bruns, # qldfloods and@ QPSMedia: Crisis communication on Twitter in the 2011 south east Queensland floods. 2012.

Figure 2. Italy tweets/keywords distribution



Source: The authors