Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 154 (2016) 176 – 183

Procedia Engineering

www.elsevier.com/locate/procedia

12th International Conference on Hydroinformatics, HIC 2016

Harvesting social media for generation of near real-time flood maps

Dirk Eilander[a*], Patricia Trambauer[a], Jurjen Wagemaker[b], Arnejan van Loenen[a]

[a]*Deltares, Boussinesqweg 1, 2600 MH, Delft, The Netherlands*
[b]*Floodtags, Binckhorstlaan 36, 2516 BE, The Hague, The Netherlands*

Abstract

Social media are a new, big and exciting source of data. Rather than from traditional sensors and models, this data is from local people experiencing real-world phenomena, such as flood events. During floods, disaster managers often have trouble getting an accurate overview of the current situation. At the same time, people affected by the floods Tweet how they are affected, if they need help and how deep the flood water is, providing an important source of information. Tweets about actual floods and containing a reference to a location, can be considered as flood observations. However, the observations are not made by validated instruments or reliable observers. Therefore, a single observation has to be considered as being unreliable. Multiple unique observations reporting the same flood severity however increase the probability of the observations being correct.

In this paper we show how these observations can be used for decision support during floods. The approach is based on the "wisdom of the crowd" principle: a group of independent consistent observations is relatively more reliable than an inconsistent group of observations. The approach uses filtering and geo-statistical methods to take into account that observations in tweets are inherently unreliable. We developed a concept that exploits observed information of the physical characteristics of a flood, such as flood depth and the location. If plotted on a Digital Elevation Map, the flooded area around one flood depth observation can be calculated. With multiple observations over the total affected area, for each of which the reliability is assessed, a flood probability map can be constructed. The method was tested in a pilot project in Jakarta, a city suffering from frequent recurring floods, but also called Twitter capital of the world.

Peer-review under responsibility of the organizing committee of HIC 2016

Keywords: Social Media; Disaster Management; Disaster Response; Near Real-Time; Floods; Big Data

* Corresponding author. Tel.: +31 (0)88 335 7672; fax: +31 (0)88 335 8582.
E-mail address: Dirk.Eilander@deltares.nl

Peer-review under responsibility of the organizing committee of HIC 2016
doi:10.1016/j.proeng.2016.07.441

1. Introduction

In the first hours of a flood event it is difficult to obtain accurate information about the extent and severity of the hazard. This information is very important for disaster management (people in need, property in danger, availability of evacuation routes). Currently, information for disaster management is derived from a few sources such as field reports, traffic cameras, satellite images and aerial images. However, getting a near real-time and accurate picture of the situation on the ground remains a problem.

At the same time, people affected by floods increasingly share their observations and needs through digital social media. It has been recognized that hazards leave a footprint on social media. Guan and Chen [1] show that the ratio between hazard related tweets and average number of tweets (The Disaster-Related-Ratio) can be used to identify hazards. Hazard related tweets have been used to gain information about the social impact [2], the relative impacts [1] and temporal behaviour of floods [3]. However, data from Twitter also contain a large number of real-time observations with physical hazard characteristics such as water depth and location, which reveal the ground truth during a flood event. Although the use of these data is still in its infancy, multiple authors have researched its possible application for disaster monitoring. One of the first to discuss the use of social media for disaster monitoring was Muralidharan at al. [4]. They used Twitter and Facebook during the 2010 earthquake in Haiti. Several applications of using social media for flood monitoring have been published since. Sun et al. [5] assessed the suitability of using geotagged Flickr images to support remote sensing based flood maps. They found that 95 % of selected Flickr observations coincided with their remote sensing derived flood map and used, along with other data sources, the intensity of flood related Twitter messages to map flood likelihood in the city of Calgary, Canada. Jongman et al. [6] researched the suitability of Twitter data for early detection of floods in the Philippines and Pakistan. Floods were mentioned one to several days earlier on Twitter then reported to humanitarian aid organizations. However, they found that the pre-processing of social media data needs to be improved for operational use. Fohringer et al. [7] was the first to utilize flood depth information in Twitter messages for rapid flood inundation mapping. By manually estimating water depths from photographs in the Tweets in combination with a Digital Elevation Model (DEM), water level estimates were generated. These water level observations were interpolated throughout the area. Results showed that social media contain additional and potentially even exclusive information that is useful for flood mapping. They found the uncertainty of interpolated inundation depth maps and the uncontrollable availability of the information to be major threats to the utility of these data for flood mapping. The PetaJakarta project [8] tried to limit the uncertainty involved by building a local community to confirm flood observations from Twitter. Although much more reliable, a limitation to this approach is the necessity of a local community which limits the coverage and scalability of the approach.

In this paper we present results of a feasibility study to use Twitter data for disaster response in the city of Jakarta, Indonesia. We developed a concept that exploits observed information about the physical characteristics of floods, such as flood depth and the location. These observations are used in combination with a Digital Elevation Model (DEM) to derive flood extent in near real-time. The data mining and processing is automated and uncertainties in the data are taken into account. The results could be used in the first hours of an event to trigger action and allocate people and money for disaster response.

2. Pilot Case

Floods pose a continuous thread to Jakarta. Every year, an increasing number of floods are observed during the rainy season, from December to March. The areas mostly struck by the repetitious floods are local dwellings along the river beds and lower parts of polder areas. Less frequently, like in 1996, 2002, 2007 and 2013, higher areas are also flooded, resulting in large scale disruption of life in extensive areas of the city. Around 240 km^2 of Jakarta is estimated to be below sea level. With an ever growing population of now more than 10 million inhabitants there is a lot at stake. The floods of February 2013 alone left 20 fatalities and more than 45,000 people displaced. The total loss was estimated to be more than USD 3.3 billion, of which a quarter was spent on economic recovery [9].

The most regularly occurring floods in Jakarta are caused by intense rain showers. Precipitation rates can reach up to 30-100 $mm \bullet h^{-1}$. Daily precipitation volumes vary from 90-265 mm for an every year to 100 year return period [10].