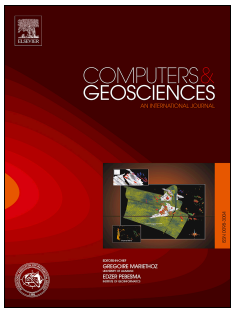
Accepted Manuscript



Hyper-resolution monitoring of urban flooding with social media and crowdsourcing data

Ruo-Qian Wang, Huina Mao, Yuan Wang, Chris Rae, Wesley Shaw

PII: S0098-3004(17)30609-X

DOI: 10.1016/j.cageo.2017.11.008

Reference: CAGEO 4049

To appear in: *Computers and Geosciences*

Received Date: 31 May 2017

Revised Date: 26 September 2017

Accepted Date: 5 November 2017

Please cite this article as: Wang, R.-Q., Mao, H., Wang, Y., Rae, C., Shaw, W., Hyper-resolution monitoring of urban flooding with social media and crowdsourcing data, *Computers and Geosciences* (2017), doi: 10.1016/j.cageo.2017.11.008.

Hyper-resolution Monitoring of Urban Flooding with Social Media and Crowdsourcing Data

Ruo-Qian Wang[1,2], Huina Mao[3], Yuan Wang[4], Chris Rae[5], Wesley Shaw[5]

1 Department of Civil and Environmental Engineering, University of California, Berkeley, CA 94720
2 School of Science and Engineering, University of Dundee, Dundee, UK DD2 1BW
3 Oak Ridge National Laboratory, Oak Ridge, TN 37831
4 Department of Civil and Environmental Engineering, Tufts University, Medford, MA 02155
5 Blue Urchin LLC, 419 11th Ave E, Seattle, WA 98102

Abstract

Hyper-resolution datasets for urban flooding are rare. This problem prevents detailed flooding risk analysis, urban flooding control, and the validation of hyper-resolution numerical models. We employed social media and crowdsourcing data to address this issue. Natural Language Processing and Computer Vision techniques are applied to the data collected from Twitter and MyCoast (a crowdsourcing app). We found these big data based flood monitoring approaches can complement the existing means of flood data collection. The extracted information is validated against precipitation data and road closure reports to examine the data quality. The two data collection approaches are compared and the two data mining methods are discussed. A series of suggestions is given to improve the data collection strategy.

1. Introduction

Urban flooding is a global problem that costs lives and money. In 2010 alone, 178 million people suffered from floods. The total economic losses in 1998 and 2010 both exceeded $40 billion (Jha et al., 2012). Urban floods can be caused by a variety of reasons, including natural hazards of river overflow, coastal storm surge, sea-level rise, flash floods, groundwater seepage, sewer overflow, lack of permeability, and lack of city management. As urbanization proceeds and climate change intensifies, urban planners and city managers are facing the challenge of preparing for and mitigating flood damage. They need tools to monitor and predict the event for emergency response and development planning.

Monitoring and predicting urban floods needs high-resolution data with good coverage. High-resolution data can capture the variation of flood flows among streets or parcels, so that the heterogeneity of flood flows caused by heterogenous urban landscape can be captured. In this study, we define data that can reflect the variation on the parcel and street scale as “hyper-resolution” data. In addition to resolution, it is important to have a good coverage of flood data to obtain complete information.

The traditional method of obtaining flood related data lack both resolution and coverage. Remote sensing is a commonly used data source. Aerial photography, for instance, is being conducted by many research teams, engineering companies, emergency response services as well as governmental departments, and has demonstrated its value (Marcus and Fonstad, 2008). However, the systematic application of aerial photography is limited by vegetation canopies and