

Harvesting Twitter

GRS-3386 GeoScripting

20-01-2016 Arend Ligtenberg



WAGENINGEN UNIVERSITY
WAGENINGEN **UR**

Why

- People Tweet about where and where
- Rich source of Geo-information
- Spatial events, processes
- Use to analyse and monitor
- Humans as sensor

Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors

Takeshi Sakaki
The University of Tokyo
Yayoi 2-11-16, Bunkyo-ku
Tokyo, Japan
sakaki@biz-model.t.u-
tokyo.ac.jp

Makoto Okazaki
The University of Tokyo
Yayoi 2-11-16, Bunkyo-ku
Tokyo, Japan
m.okazaki@biz-
model.t.u-tokyo.ac.jp

Yutaka Matsuo
The University of Tokyo
Yayoi 2-11-16, Bunkyo-ku
Tokyo, Japan
matsuo@biz-model.t.u-
tokyo.ac.jp

ABSTRACT

Twitter, a popular microblogging service, has received much attention recently. An important characteristic of Twitter is its real-time nature. For example, when an earthquake occurs, people make many Twitter posts (*tweets*) related to the earthquake, which enables detection of earthquake occurrence promptly, simply by observing the tweets. As described in this paper, we investigate the real-time interaction of events such as earthquakes, in Twitter, and propose an algorithm to monitor tweets and to detect a target event. To detect a target event, we devise a classifier of tweets based on features such as the keywords in a tweet, the number of words, and their context. Subsequently, we produce a probabilistic spatiotemporal model for the target event that can find the center and the trajectory of the event location. We consider each Twitter user as a *sensor* and apply Kalman filtering and particle filtering, which are widely used for location estimation in ubiquitous/pervasive computing. The particle filter works better than other compared methods in estimating the centers of earthquakes and

currently estimated as 44.5 million worldwide¹. Monthly growth of users has been 1382% year-on-year, which makes Twitter one of the fastest-growing sites in the world².

Some studies have investigated Twitter: Java et al. analyzed Twitter as early as 2007. They described the social network of Twitter users and investigated the motivation of Twitter users [13]. B. Huberman et al. analyzed more than 300 thousand users. They discovered that the relation between friends (defined as a person to whom a user has directed posts using an "@" symbol) is the key to understanding interaction in Twitter [11]. Recently, Boyd et al. investigated *retweet* activity, which is the Twitter-equivalent of e-mail forwarding, where users post messages originally posted by others [5].

Twitter is categorized as a micro-blogging service. Microblogging is a form of blogging that allows users to send brief text updates or multimedia such as photographs or audio clips. Microblogging services other than Twitter include Tumblr, Plurk, Enote.in, Squeed, Jaiku, identi.ca, and so on³. They have their own characteristics. Some examples

Geo-spatial Event Detection in the Twitter Stream

Maximilian Walther and Michael Kaiser

AGT International,
Jägerstraße 41, 10117 Berlin, Germany
{mwalther,mkaiser}@agtinternational.com

Abstract. The rise of Social Media services in the last years has created huge streams of information that can be very valuable in a variety of scenarios. What precisely these scenarios are and how the data streams can efficiently be analyzed for each scenario is still largely unclear at this point in time and has therefore created significant interest in industry and academia. In this paper, we describe a novel algorithm for geo-spatial event detection on Social Media streams. We monitor all posts on Twitter issued in a given geographic region and identify places that show a high amount of activity. In a second processing step, we analyze the resulting spatio-temporal clusters of posts with a Machine Learning component in order to detect whether they constitute real-world events or not. We show that this can be done with high precision and recall. The detected events are finally displayed to a user on a map, at the location where they happen and while they happen.

Keywords: Social Media Analytics, Event Detection, Twitter.

1 Introduction

The rise of Social Media platforms in recent years brought up huge information streams which require new approaches to analyze the respective data. At the time of writing, on Twitter⁴ alone, more than 500 million posts are issued every day. A large part of these originate from private users who describe how they

Twitter API

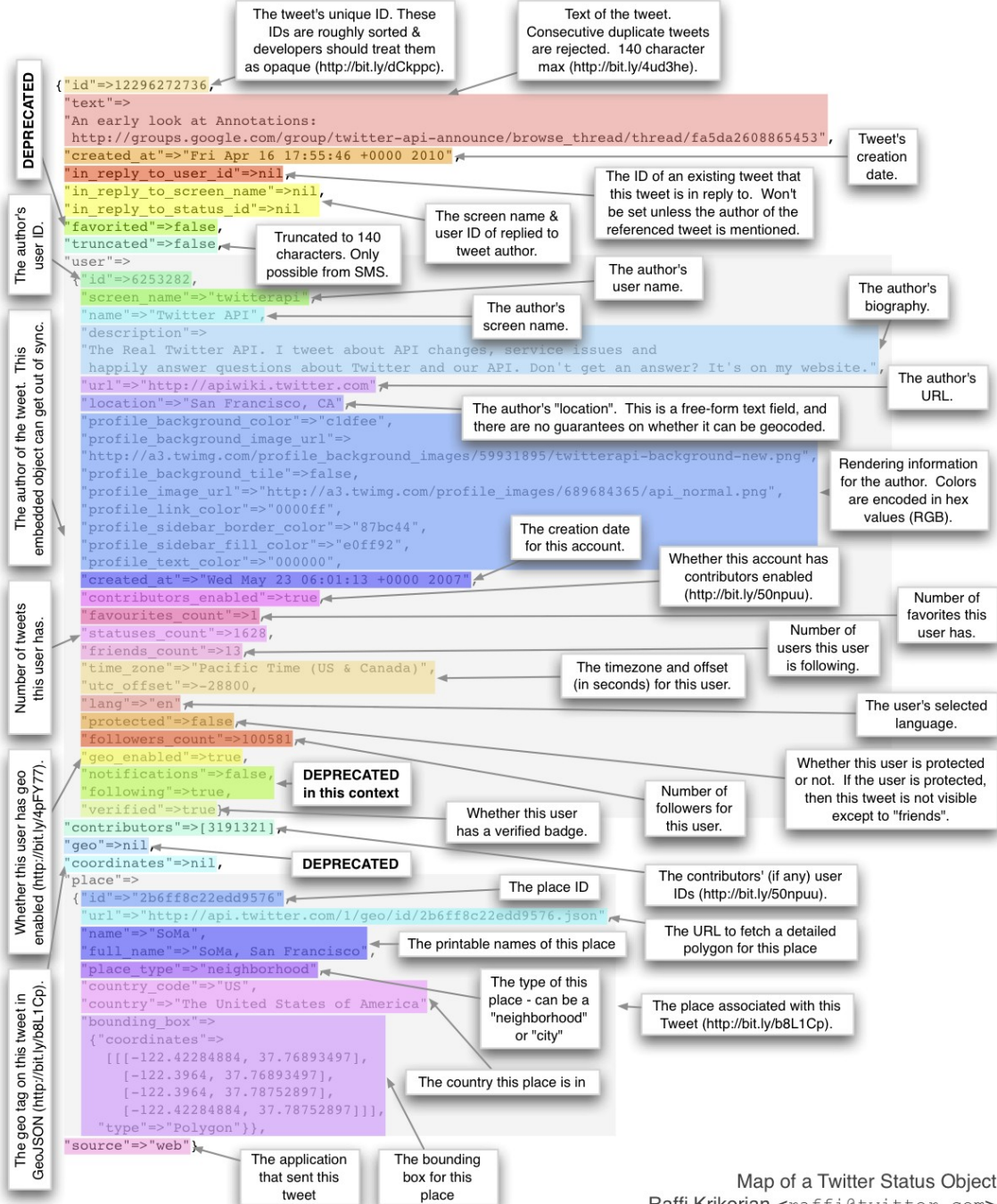
- Access to twitter through webservices
- REST & STREAMING
- Various streams
 - Garden Hose (public stream)
 - Deca Hose (10%)
 - Fire Hose (full)
- Various Python libraries (Twython, Tweepy..)
- Probably also for R

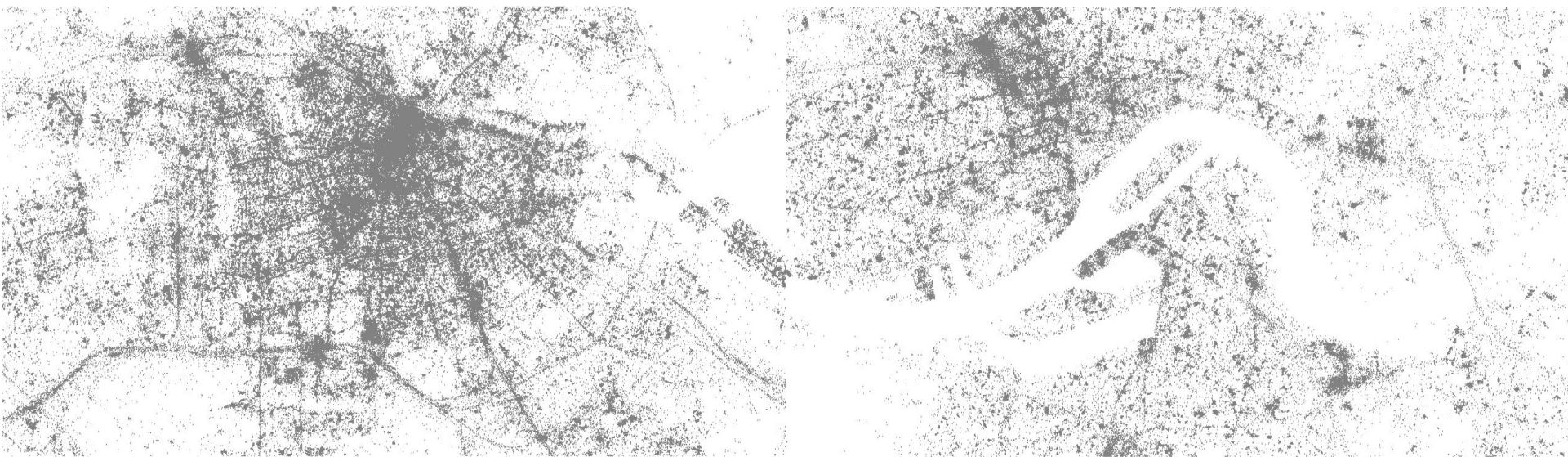


Twitter Data

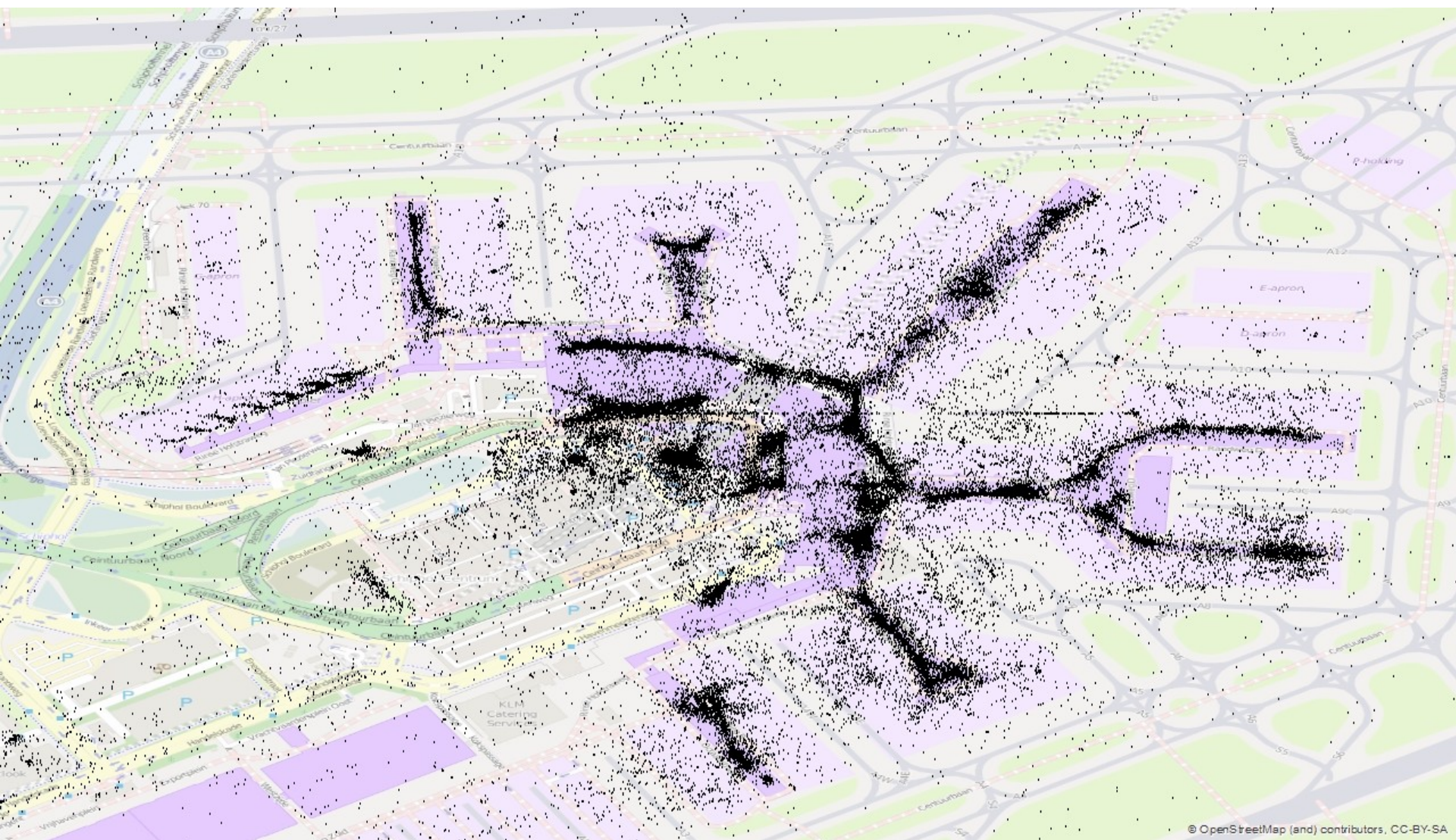
- JSON
- Nested structure consisting of objects
 - Tweet
 - Users
 - Places
 - Bounding box
 - Coordinates
 - Entities







WAGENINGEN UNIVERSITY
WAGENINGEN **UR**



© OpenStreetMap (and) contributors, CC-BY-SA



WAGENINGEN UNIVERSITY
WAGENINGEN UR

Assignment

- Create an tweet harvesting script
- Use REST or STREAMING
- Extract useful information
- Create a spatial dataset / map



Questions?

