Transportation Research Part C 73 (2016) 91–104

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

Mining and correlating traffic events from human sensor observations with official transport data using self-organizing-maps ☆

Enrico Steiger [a,*], Bernd Resch [b,c], João Porto de Albuquerque [d], Alexander Zipf [a]

[a] GIScience Research Group, Institute of Geography, Heidelberg University, Germany
[b] Z_GIS, Department of Geoinformatics, University of Salzburg, Austria
[c] Center for Geographic Analysis, Harvard University, USA
[d] Centre for Interdisciplinary Methodologies, University of Warwick, United Kingdom

ARTICLE INFO

Article history:
Received 28 August 2015
Received in revised form 16 October 2016
Accepted 17 October 2016

Keywords:
Traffic data
Twitter
Self-organizing map
Point pattern analysis
Human mobility

ABSTRACT

Cities are complex systems, where related Human activities are increasingly difficult to explore within. In order to understand urban processes and to gain deeper knowledge about cities, the potential of location-based social networks like Twitter could be used a promising example to explore latent relationships of underlying mobility patterns. In this paper, we therefore present an approach using a geographic self-organizing map (Geo-SOM) to uncover and compare previously unseen patterns from social media and authoritative data. The results, which we validated with Live Traffic Disruption (TIMS) feeds from Transport for London, show that the observed geospatial and temporal patterns between special events ($r = 0.73$), traffic incidents ($r = 0.59$) and hazard disruptions ($r = 0.41$) from TIMS, are strongly correlated with traffic-related, georeferenced tweets. Hence, we conclude that tweets can be used as a proxy indicator to detect collective mobility events and may help to provide stakeholders and decision makers with complementary information on complex mobility processes.

1. Introduction

The complexity of cities with related human activities is becoming an increasingly tough challenge for policy makers and modelers to explore urban dynamics and the study of city-scale mobility patterns. One promising example of available high-granularity information sources is the Transport for London's (TfL) Traffic Information Management System (TIMS). It provides high resolution, up-to-date disruption information regarding congestions, traffic incidents, events, construction works and other issues affecting traffic. However, these existing traffic measuring systems (e.g., road-side detectors, video surveillance, floating car data, etc.) are resource intensive in terms of ongoing operating and maintenance costs. Furthermore, a complete detection of all traffic and road conditions is simply not feasible.

At the same time, in recent years an increasing amount of information has been generated through mobile devices, becoming a potentially powerful data resource for (geographic) knowledge discovery and human behavior analysis from

☆ This article belongs to the Virtual Special Issue on "Smart City".

* Corresponding author at: Berliner Strasse 48, 69221 Heidelberg, Germany.
E-mail address: enrico.steiger@geog.uni-heidelberg.de (E. Steiger).

http://dx.doi.org/10.1016/j.trc.2016.10.010
0968-090X/

crowdsourced data (Goodchild, 2007). For a number of disciplines this development opens up enormous potential for various applications, including urban- and traffic planning, disease- and disaster management.

In particular, harnessing human mobility information from social media platforms such as Twitter can potentially lead to new insights into the human mobility process. Due to the high spatiotemporal resolution this may provide complementary information when compared with existing traffic data sources. However, one main challenge when analyzing mobility with officially acquired data is the spatiotemporal complexity of latent processes within traffic events (e.g., effects of incidents such as roadworks on the traffic flow and the correlations with traffic disruptions), hampering the detection of patterns in large road networks (Asif et al., 2014).

Simultaneously, when using crowdsourced information it is uncertain how representative and trustworthy these new types of geodata are for the inference of human mobility patterns (Steiger et al., 2015c). Thus, research in this area requires new methodological approaches, which consider the high dimensionality and uncertainty of crowdsourced geographic information in the context of a data-driven geography (Miller and Goodchild, 2014). In a previous study, we therefore applied and have demonstrated the efficiency of self-organizing maps (SOMs) to abstract and cluster information from multidimensional Twitter data in a trans-disciplinary approach (Steiger et al., 2015b). However, it has not been analyzed whether spatiotemporal information from social media data is a suitable proxy for inferring certain traffic-related events. Further the question is whether the results in comparison with official traffic disruption reports lead to new insights regarding the study of human mobility patterns.

In this paper, we use a non-geographic and a geographic self-organizing map (SOM/Geo-SOM) to discover collective human mobility clusters by analyzing similar variances within geospatial, temporal and disruption characteristics from live traffic feeds. The results are correlated with traffic-related georeferenced tweets for a case study in London. We intend to answer the following research questions (RQ):

(RQ1): What is the correlation between inferred spatiotemporal clusters from tweets, as a proxy of collective human mobility patterns and the real time traffic information provided by TIMS?

(RQ2): Which official traffic events along with their individual traffic disruption characteristics (category, severity, duration) are reflected in traffic-related tweets and have a dissimilar/similar spatiotemporal distribution?

2. Background

This section summarizes the characteristics of both datasets used in this analysis (Sections 2.1 and 2.2). Then, related work in the area of spatiotemporal mobility analysis is depicted in Section 2.3, followed by a description of the current state of the art regarding the application of SOMs for mobility analysis in Section 2.4.

2.1. Comparative reference dataset: TIMS disruption messages

The Transport for London (TfL) authorities provide real time open traffic disruption data for the area of Greater London as part of their Traffic Information Management System (TIMS). TIMS contains a broad range of a priori known information concerning road disruptions, such as the location of occurrence, details regarding road closures and more in-depth categorization of the cause of a disruption. Our research solely focuses on analyzing active persisting traffic event messages within the five categories of traffic incidents, traffic volume, hazards, and special and planned events (Transport for London, 2007). As persisting traffic disruptions are repeated until cleared, we can compute the length of every incident and have, additional categorical information regarding the severity of a traffic event by combining it with the provided level of interest and priority status. All observed traffic disruption messages and the identified general spatiotemporal patterns of occurrences in London provide a reliable ground truth of mobility events. Thus, they serve as a reference dataset for our comparison with Twitter data.

2.2. Social media dataset: Twitter

Within the online social network and microblogging system Twitter, currently more than 288 million users post 500 million short status messages (tweets) with up to 140 characters per day.[1] As a further option, users can geotag their tweets with a geo-location acquired by their mobile devices. Therefore tweets are provided in high spatiotemporal resolution (geolocation and timestamp of the tweet) and include a semantic information layer (message content of the tweet). Since georeferenced tweets are to a certain extent a proxy of real-world observations (Hawelka et al., 2014), they represent a valuable opportunity for studying human mobility dynamics. Frequently repeating patterns of contextually similar tweets over geographic space and time might serve as an indicator to characterize human activity and to detect traffic-related events (exemplary georeferenced tweet in Fig. 1).

[1] https://about.twitter.com/de/company.