Journal of Transport Geography 54 (2016) 81–90

Contents lists available at ScienceDirect

Journal of Transport Geography

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

Traffic incidents in motorways: An empirical proposal for incident detection using data from mobile phone operators

John Steenbruggen [a,*], Emmanouil Tranos [b], Piet Rietveld [a]

[a] VU University Amsterdam, Department of Spatial Economics, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands
[b] University of Birmingham, School of Geography, Earth and Environmental Sciences, Edgbaston B15 2TT, UK

ARTICLE INFO

Article history:
Received 20 October 2013
Received in revised form 17 May 2016
Accepted 18 May 2016
Available online 31 May 2016

Keywords:
Road traffic incident management
Mobile phone data
Data science
Collective sensing

ABSTRACT

This paper proves that mobile phone usage data is an easy to use, cheap and most importantly, reliable predictor of motorway incidents. Using econometric modelling, this paper provides a proof of concept of how mobile phone usage data can be utilised to detect motorway incidents. Greater Amsterdam is used here as a case study and the results suggest that mobile phone usage data can be utilised for the development of an early warning system to support road traffic incident management.

1. Introduction

Increased urbanisation does not come for free and car traffic related congestion and incidents are some of the most pronounced externalities. This paper aims to contribute to the management of these externalities by providing a proof of concept which can assist traffic Incident Management (IM). In brief, this paper proposes the use of mobile phone usage data within an IM system as a tool to detect motorway incidents. By adding a layer of information inferred by mobile phone usage data, which is easily accessible nowadays and free of charge, the efficiency of IM can be drastically increased.

IM involves the cooperation of many public and private actors. To support these tasks in an effective way, advanced information systems and the use of spatio-temporal data are becoming increasingly important (Steenbruggen et al., 2012, 2014a). Along with the growing ubiquity of mobile technologies, the extensive data logs produced in the course of their usage have helped researchers to create and define new methods of observing, recording, and analysing environments and their human dynamics (O'Neill et al., 2006). In effect, these personal devices create a vast, geographically-aware sensor web that accumulates tracks to reveal both individual and social behaviours in unprecedented detail (Goodchild, 2007). Steenbruggen et al. (2013a, 2013b) have identified this phenomenon as *collective sensing*, or, in other words, the reconstruction of "collective human behaviour from individual anonymous digital traces". These traces left by individuals are accumulating at an unprecedented scale (Zhang et al., 2010) resulting in very large data sets known as 'big data'. The usability of such data has been demonstrated in the relevant literature (Boyd and Crawford, 2012; Steenbruggen, 2014; Steenbruggen et al., 2014b). In this paper we use various (big) data sets to explore the relationship between motorway traffic incidents and mobile phone usage. Taking into account both the spatial and the temporal dimension, we model how mobile phone activity is related to road traffic incidents.

Traffic incidents may be sensitive to different weather conditions. Therefore, we also include in our models other variables such as meteorological measurements to control for the weather effect on motorway incidents in the Greater Amsterdam area. Within the environmental monitoring domain, the amount and the availability of digital information, based on near real-time sensor measurements, have been rapidly increasing (e.g. see Akyildiz et al., 2002; Hart and Martinez, 2006). Such sensor nodes include highly mobile and intelligent sensor pods (Resch et al., 2010), as well as fixed sensor stations (Alesheikh et al., 2005). Given the increasing accuracy of meteorological monitoring and forecasting, understanding the relationship between weather patterns and traffic incidents can potentially provide valuable insights into understanding and predicting mobility and traffic accidents (Sabir, 2011).

The main research question of this paper is to examine whether we can use mobile phone data to detect motorway traffic incidents (dashed line in Fig. 1). The underlying goal is to explain which factors affect mobile phone usage in area i at time t, and in particular the role of incidents in area i at time t (straight line in Fig. 1).

* Corresponding author.
E-mail addresses: john.steenbruggen@rws.nl (J. Steenbruggen), e.tranos@bham.ac.uk (E. Tranos), prietveld@vu.nl (P. Rietveld).

http://dx.doi.org/10.1016/j.jtrangeo.2016.05.008
0966-6923/

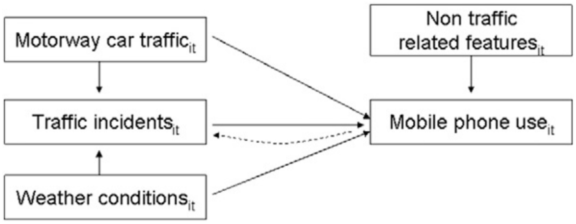

Fig. 1. Graphical representation of our research model.

More specifically, the mobile communication volume depends on the specific land use (e.g. business areas, shopping centres) and other non-traffic-related features, such as weather conditions. We focus on GSM (Global System for Mobile Communications) zones which strongly overlap the motorway infrastructure, which means that motorway traffic intensity and traffic incidents potentially have a substantial influence on mobile phone usage.

The dashed arrow in the figure, pointing in the reverse direction (from mobile phone usage to traffic incidents) is also addressed in the paper. This arrow is not meant to represent a causal relationship, but it is introduced to represent the notion that data on telecom use can be employed to detect the occurrence of incidents, and therefore contribute to a rapid detection of incidents on motorways.

The structure of the paper is the following. The next section describes the data used and then the empirical application is presented. A three step modelling strategy has been designed, which starts by modelling the relation between traffic and mobile phone usage, followed by a model explaining the relationship between car incidents and motorway traffic and finally presenting the marginal effects of this relationship. The paper ends with a conclusion section.

2. Data

The study area involves the city of Amsterdam and its surroundings, covering an area of about 1000 km^2 (see Fig. 2). We used four different types of data sets for our research: mobile phone usage data; traffic incident data; motorway traffic flow data; and meteorological sensor data. These data sets are described below in more detail.

2.1. Mobile phone usage data

The mobile phone usage data that we utilize for this paper was supplied by a major Dutch telecom operator, and provides aggregated information about mobile phone usage at the level of the (GSM) cell zones for the period 2007–2010. The most common format is the '*Call Data Record*' (CDR), according to which subscribers' mobile phone activities are recorded each time a user uses a service (Steenbruggen et al., 2014b). The project uses anonymized data of the mobile network. The raw data contains aggregated CDR information with a temporal dimension of a 1-hour time interval in a certain (GSM) cell zone. In the study area, over 1200 (GSM) cell zones were provided by a Dutch telecom operator (see Fig. 2). Based on our criteria, as described at the end of this section, only 109 (GSM) cell zones were used in our modelling exercise (see also Fig. 3). The telecommunication operator applied special scripts to extract the necessary data for the project. For the purpose of this case study, we select data from 1 January 2010 (00.00 h.) through to 20 November 2010 (07.00 h.). See also Fig. 3. The GSM cellular network is built on the basis of radio cells. They define the spatial dimensions of the two best serving cell maps generated by antennas with two different frequencies overlaid on each other: namely, 900 MHz coverage (the basic network with full area coverage), and 1800 MHz (capacity network only in densely populated areas). The

Fig. 2. Overview the Amsterdam test area. Note: Owing to the irregular boundaries of the sectors, the network coverage of the study area does not exactly correspond to the above box.