Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 11 (2015) 381 – 398

Transportation Research Procedia

www.elsevier.com/locate/procedia

10th International Conference on Transport Survey Methods

Passive mobile phone dataset to construct origin-destination matrix: potentials and limitations

Patrick Bonnel [a]*, Etienne Hombourger [b], Ana-Maria Olteanu-Raimond [c], Zbigniew Smoreda [d]

[a] *Laboratoire d'Economie des Transports, ENTPE, Lyon, France, patrick.bonnel@entpe.fr*
[b] *DTecITM, CEREMA, Paris, France, Etienne.hombourger@cerema.fr*
[c] *COGIT, IGN, Paris, France, Ana-Maria.Raimond@ign.fr*
[d] *SENSE, Orange Labs, Paris, France, zbigniew.smoreda@orange.com*

Abstract

Mobile phone operators produce enormous amounts of data. In this paper we present applications performed with a dataset (communication events + handover and Location Area Up-date) collected by the operator Orange from 31 March to 11 April 2009 for the whole Paris Region. Trips are deduced from the spatio-temporal trajectory of devices through a hypothesis of stationarity within a Location Area in order to define activities. Trips are then aggregated in an origin-destination matrix which is compared with traditional data (census data and household travel survey).

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Peer-review under responsibility of International Steering Committee for Transport Survey Conferences ISCTSC

Keywords: origin-destination matrix; mobile phone data; travel survey; passive data

1. Introduction

Data on spatial mobility are essential in order to build and use travel demand forecasting models, for transport planning purposes and for the appraisal of transport policies… (Arentze et al., 2000; Ortuzar, Bates, 2000). They must also be of good quality and, in particular, accuracy, to ensure that investment or transport policy decisions are

* Corresponding author. Tel.: +0-000-000-0000 ; fax: +0-000-000-0000 .
E-mail address: author@institute.xxx

2352-1465 © 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Peer-review under responsibility of International Steering Committee for Transport Survey Conferences ISCTSC
doi:10.1016/j.trpro.2015.12.032

based on reliable analyses. While travel surveys provide extremely useful data in order to formalize and estimate behavioural choice models (for example the choice of a destination or mode of transportation), they are much less useful for constructing origin-destination (O-D) matrices due to an inadequate number of trips in many of the matrix elements. In addition, surveys are increasingly confronted by issues during the sample construction phase (Stopher, Greaves, 2007), by falling response rates (Atrostic, Burt, 1999; Ampt, 1997; Bonnel, 2003; Zmud, 2003) and by unreported trips (Wolf et al., 2003), which reduce even further the quality of the resulting matrices. Consequently, trip matrices are also generated from other sources, in some cases in combination with survey data, examples being roadside traffic counts, cordon or screen-line surveys and public transport surveys. The resulting data are useful for improving the quality of matrices, but do not always contain the necessary information. This applies, for example, to road traffic counts which provide information on the traffic volumes at a given point on a road, but not on trip origins and destinations. A variety of techniques have been developed for processing and combining the data from different sources. However, the reliability of the resulting matrices is uncertain and cannot always be measured statistically.

The advent of large volumes of data that are produced automatically and passively such as ticketing data (Arana et al., 2014, Morency et al., 2007; Munizagua et al., 2010; Pelletier et al., 2011), bank cards… and mobile phone data makes it possible to identify the presence of individuals in both space and time in a way which, while admittedly irregular is becoming less so. A number of techniques have been developed for converting these data into trips, but little research has attempted to "validate" them by comparing them with data from other sources in order to identify possible biases and gain a clearer idea of their potential. The aim of this paper is therefore to test the potential of these data for producing origin-destination matrices compared with other sources of available data. The analysis has been conducted within the Greater Paris Region (Ile-de-France) for which we were able to analyse the mobile phone data from the operator Orange and compare them on the one hand with the census data on commuting trips from home to place of work or study, and on the other hand with the data obtained from the "*Enquête Globale Transport*" (EGT), which is the name given to the household travel surveys conducted in the Ile-de-France Region.

We shall begin this paper with a literature survey (Section 1) before presenting the data we have used (Section 2) and (Section 3) the data processing methodology employed to produce the origin-destination matrices, which allow us to make the comparison with external data (Section 4). Finally, we shall present the principal lessons from this research, and some suggested directions for the future (Section 5).

2. Literature survey

Cell phone networks have existed for two decades, and mobile phones have achieved a high rate of penetration: there were 76.8 million active SIM (Subscriber Identity Module) cards in France at the end of 2013, for a total population of 65 million (ARCEP, 2014). Mobile devices (mobile phones, smartphones and tablets) have become indispensable tools, bearing witness to our activities and trips. Mobile phone operators, who are obliged for legal or billing purposes to record information about the use of these devices, therefore find themselves with increasingly informative databases. The reason for this is that each time a mobile terminal is used to make a call, send an SMS (Short Message Service), the operator generates a call detail record (CDR) that contains the timestamp, the terminal's identifier of the base station to which the user is connected and quantitative data about the call (call duration, volume of data exchanged).

As a result of the size of the samples, which in the case of some operators can involve as many as 40-50% of a country's population, and the non-intrusive way the data is collected, the exploitation of mobile phone data logs has enormous potential. Recent cases include using the data to analyse behavioural differences between men and women (Frias-Martinez et al., 2010), studying the propagation of an epidemic (Tizzoni et al., 2013), mapping activities within a city (Noulas et al., 2013), or improving the paging efficiency of the cellular network (Zhang, Bolot, 2007).

But the usefulness of mobile phone data has above all been proven for the study of human mobility, in spite of the fact that the localisation data associated with each log is limited to the position of the base station used, which results in a positioning uncertainty ranging from approximately a hundred metres in a dense urban zone (Calabrese et al., 2013) to several kilometres in rural zones. Gonzalez et al. (2008) were amongst the first scholars to carry out a large-scale study of the mobility of users, with a sample of over 100,000 individuals. This study demonstrated that human mobility may be modelled using a random technique and that trips follow a truncated power-law distribution.

Download English Version:

https://daneshyari.com/en/article/1106793

Download Persian Version:

https://daneshyari.com/article/1106793

Daneshyari.com