Transportation Research Part A 113 (2018) 55–74

Contents lists available at ScienceDirect

Transportation Research Part A

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

Identifying road user classes based on repeated trip behaviour using Bluetooth data

F. Crawford[a,*], D.P. Watling[b], R.D. Connors[b]

[a] *Centre for Transport and Society, University of the West of England, UK*
[b] *Institute for Transport Studies, University of Leeds, UK*

ARTICLE INFO

Keywords:
Intrapersonal variability
Bluetooth data
Sequence alignment
Model based clustering

ABSTRACT

Analysing the repeated trip behaviour of travellers, including trip frequency and intrapersonal variability, can provide insights into traveller needs, flexibility and knowledge of the network, as well as inputs for models including learning and/or behaviour change. Data from emerging data sources provide new opportunities to examine repeated trip making on the road network. Point-to-point sensor data, for example from Bluetooth detectors, is collected using fixed detectors installed next to roads which can record unique identifiers of passing vehicles or travellers which can then be matched across space and time. Such data is used in this research to segment road users based on their repeated trip making behaviour, as has been done in public transportation research using smart card data to understand different categories of users. Rather than deciding on traveller segmentation based on a priori assumptions, the method provides a data driven approach to cluster together travellers who have similar trip regularity and variability between days. Measures which account for the strengths and weaknesses of point-to-point sensor data are presented for (a) spatial variability, using Sequence Alignment, and (b) time of day variability, using Model Based Clustering. The proposed method is also applied to one year of data from 23 fixed Bluetooth detectors in a town in northwest England. The data consists of almost 7.5 million trips made by over 300,000 travellers. Applying the proposed methods allows three traveller user classes to be identified: infrequent, frequent, and very frequent. Interestingly, the spatial and time of day variability characteristics of each user class are distinct and are *not* linearly correlated with trip frequency. The frequent travellers are observed 1–5 times per week on average and make up 57% of the trips recorded during the year. Focusing on these frequent travellers, it is shown that these can be further separated into those with high spatial and time of day variability and those with low spatial and time of day variability. Understanding the distribution of travellers and trips across these user classes, as well as the repeated trip characteristics of each user class, can inform further data collection and the development of policies targeting the needs of specific travellers.

1. Introduction

While considering daily snapshots of transport networks is sufficient for many purposes, the benefits of considering the patterns and variability in each individual's behaviour over days, months and even years is receiving increasing research attention. It can inform us about traveller habits (Minnen et al., 2015), predictable differences in travel patterns (Cherchi et al., 2017) and traveller

* Corresponding author.
E-mail address: fiona.crawford@uwe.ac.uk (F. Crawford).

https://doi.org/10.1016/j.tra.2018.03.027
Received 5 April 2017; Received in revised form 21 January 2018; Accepted 25 March 2018
0965-8564/

flexibility (Kitamura et al., 2006), all of which are important for developing new policies and modelling traveller *response* to those policies, for example using day-to-day dynamical models which include micro-level learning mechanisms (Chen and Mahmassani, 2004; Liu et al., 2006). Understanding the current behaviour of travellers, not just on a single day but over days, weeks and months, also provides information about traveller needs and knowledge of the network.

A common assumption is that urban traffic, particularly the morning peak, consists of commuters who drive between home and work at the same time each weekday. This assumption is often made implicitly and largely for convenience but is rarely challenged despite increases in part time, flexible and home working in recent years. In Great Britain, a 2013 survey (Department for Business, Innovation and Skills, 2014) found that 80% of workplaces with at least 5 employees had part time staff, and other forms of flexible working such as reduced hours, flexitime and compressed hours had all increased since the first comparable survey in 2000. Assumptions about the regularity of travellers is likely to influence the types of policies formulated to reduce morning peak congestion, some of which may perform differently based on the repeated trip making behaviour of travellers. For example, if the proportion of frequent travellers is overestimated, then the benefits to travellers of switching to public transportation due to savings from weekly or monthly tickets would also be overestimated. Similarly, making an assumption that all travellers have very little departure time flexibility would underestimate the impact of interventions such as time of day specific congestion charging.

One of the reasons why behaviour over multiple days is often overlooked may be the difficulty in collecting data. Detailed information about repeated trip making behaviour has typically been collected using multi-day travel diaries (Muthyalagari et al., 2001; Schlich and Axhausen, 2003; Elango et al., 2007). Such surveys provide data of great depth, but at a cost – both financially and in terms of respondent burden. For example, the National Travel Survey in England involves face to face interviews and 7 day travel diaries for individuals in 7000 households and costs approximately £2.1 million per year to collect and process (Data.gov.uk, 2012). Respondent burden can be decreased by using GPS devices to track participants (Muthyalagari et al., 2001), but costs remain high, resulting in surveys which often are for short periods of time and/or have small sample sizes. For example, the 7 day travel diaries undertaken annually in England have a relatively large sample size, but sample sizes are usually much smaller for longer surveys, for example the six week Mobi*drive* survey collected in 1999 in Karlsruhe and Halle in Germany had 317 participants in 139 households (Axhausen et al., 2002).

More recently, emerging data sources have been explored to determine their usefulness with respect to measuring repeated trip making behaviour. Mobile phone data has been used to examine activity spaces, as in Järv et al. (2014), but the spatial precision is relatively low. In public transportation research, the availability of smart card data has resulted in researchers identifying different types of user based on their travel behaviour over time (Chu and Chapleau, 2010; Kieu et al., 2015; Goulet Langlois et al., 2016). Goulet Langlois et al. (2016) analysed four weeks of smart card data from London and identified four types of regular commuter. The daily and weekly activity sequences constructed using the smart card data had distinct patterns for each of these four groups: 'typical' commuters, commuters who sometimes did not take public transportation home at night, commuters who used public transport as their main mode at the weekend and commuters who travelled less during school holiday periods.

The current paper examines data which could be considered the road network counterpart to smart card data, namely point-to-point sensor data, which includes Bluetooth and Automatic Number Plate Recognition (ANPR) data. Point-to-point 'sensors' or 'detectors' collect unique identifiers, either for vehicles or travellers, at fixed locations. It is this "re-identification and tracking" ability which defines this type of data (Antoniou et al., 2011, p140) and as the unique identifiers can be matched over space and time, the data is ideal for examining repeated trip making. Where point-to-point sensors are permanently installed, the amount of data collected can quickly become very large. For example, in Section 3 an application to one year of data from 23 detectors is presented, and that data contains almost 7.5 million trips. These trips are obtained from processing 29.7 million observations, each of which corresponds to a Bluetooth device passing a detector.

Analysing such data with respect to repeated trip behaviour as a whole is not straightforward, however. Previous research on repeated trip making has usually focused on a single aspect, for example trip frequency (Elango et al., 2007; Tarigan and Kitamura, 2009), spatial variability (Buliung et al., 2008; Järv et al., 2014), time of day variability (Kitamura et al., 2006; Chikaraishi et al., 2009) or mode choice (Cherchi and Cirillo, 2014; Heinen and Chatterjee, 2015). Other research has combined different aspects to create a single measure of intrapersonal variability (see Schlich and Axhausen (2003) for an overview). Calculating a single Similarity Index for travellers can be limiting, however, as it cannot account for travellers which differ in terms of different aspects of variability, for example travellers whose trips are spatially predictable but unpredictable in terms of the time of day at which they occur. The current paper uses cluster analysis to segment travellers based on measures relating to multiple aspects of intrapersonal variability, as has been done for public transport users (Goulet Langlois et al., 2016). The methods proposed to measure the different aspects are distinctive from previous work, however, due to the nature of the data available from point-to-point sensors. Firstly, point-to-point sensor data does not generally provide origin or destination information due to limited network coverage and the possibility that many trips start and/or end outside the monitored area. It does not provide information about trip purpose either. This means that existing approaches for measuring spatial variability are not suitable. Existing approaches include measuring the distance travelled from home (Bayarma et al., 2007) and comparing daily activity sequences (Goulet Langlois et al., 2016). Secondly, point-to-point sensor data can provide some route choice information, depending on sensor locations, and it would be preferable to have a methodology which takes this additional information into account. Thirdly, for time of day variability, adjustments need to be made since the observations are not departure times.

There is, therefore, a research gap as user classes based on repeated trip behaviour have not, to the authors' knowledge, been considered for *road* users. Addressing this lack of empirical evidence is not trivial since the methods used to measure intrapersonal variability on other modes are not directly transferable. There is therefore a methodological gap in addition to the empirical one; in the present paper a methodology is proposed which takes into account the strengths and weaknesses of such point-to-point sensor