Transportation Research Part C 87 (2018) 123–137

Contents lists available at ScienceDirect

Transportation Research Part C

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

Public transport trip purpose inference using smart card fare data

Azalden Alsger*, Ahmad Tavassoli, Mahmoud Mesbah, Luis Ferreira, Mark Hickman

School of Civil Engineering, The University of Queensland, Australia

A R T I C L E I N F O

Keywords:
Trip purpose
O-D trips
Public transport smart card fare data
Rule-based modelling

A B S T R A C T

Although smart card fare data has recently become more prevalent as a rich, comprehensive and continuous source of information, there is still some missing information which inhibits its capability in the research field. One key missing piece of information is the passengers' trip purpose. This paper investigates the potential of the smart card data to infer passengers' trip purpose, thereby reducing the use of the expensive and time-consuming Household Travel Surveys (HTS). On this basis, an improved model has been proposed, calibrated and validated for trip purpose inference by integrating different data sources, namely: HTS, a land use database, the South East Queensland Strategic Transport Model (SEQSTM), the General Transit Feed Specification (GTFS) data, O-D survey data, and most importantly the unique smart card fare data from Brisbane, Queensland. As smart card fare data does not record passengers' trip purpose, the calibration and validation procedures are performed on HTS data. Based on the validation results, the proposed methodology shows a strong capability to predict trip purpose at a high level of accuracy. The results show an overall 67% correct inference after applying spatial attributes, but the correct inference increases to 78% after applying temporal attributes. Different trip purposes show different sensitivities to the applied spatial and temporal attributes. Work and home trips have the highest correct inference results, at 92% and 96%, respectively. Furthermore, the results of correct inference for shopping and education trips improved after applying the temporal attributes.

1. Introduction

The influence of different trip purposes on public transport demand is varied in time and space. For example, most of the work and education trips start in the morning peak from home and return back to home in the evening peak. On the other hand, most shopping and recreational trips occur in the off peak time (Lee and Hickman, 2014, Kuhlman, 2015). Thus, it is important to segment origin-destination (O-D) flows by trip purpose, particularly when the time of day is considered (Alexander et al., 2015). In spite the fact that no information on trip purpose is recorded by Automated Fare Collection (AFC) systems (Bagchi and White, 2005), smart card data still provide a very rich source of information which can be used to analyse and understand passengers' travel behaviour (Alsger et al., 2016a). Acknowledging the importance of estimating an O-D matrix by trip purpose in transport planning, smart card fare data can be utilized with other data sources to reduce the use of the traditional and expensive data collection methods, such as household travel surveys (HTS), for inferring trip purpose.

Previous attempts have been introduced to utilize different data sources to infer individuals' trip purpose. The range of considered trip purposes varies from just work trips (Jun and Dongyuan, 2013, Zhou et al., 2014), to the basic trip purposes such as work and

* Corresponding author.
E-mail addresses: a.alsger@uq.edu.au (A. Alsger), a.tavassoli@uq.edu.au (A. Tavassoli), mahmoud.mesbah@uq.edu.au (M. Mesbah), l.ferreira@uq.edu.au (L. Ferreira), m.hickman1@uq.edu.au (M. Hickman).

https://doi.org/10.1016/j.trc.2017.12.016
Received 12 December 2016; Received in revised form 27 December 2017; Accepted 27 December 2017
0968-090X/

home trips (Chakirov and Erath, 2012) and education trips (Devillaine et al., 2012, Chu and Chapleau, 2010).

Lee and Hickman (2014) utilized different data sources, namely Automated Fare Collection (AFC), General Transit Feed Specification (GTFS) and parcel-level land use data, to infer individuals' trip purposes. A training set of behavioural and heuristic rules were built for deriving trip purpose using these data sources. The regularity and variability of time and space in day-to-day data were used to infer changes in trip purposes. The study focused on two main trip purposes (work- and school-related trips) and used 306 cardholders for their study. Kusakabe and Asakura (2014) used the concept of data fusion to estimate the probability distribution of trip purpose by analysing the behavioural features observed from smart card and trip survey data. That study integrated the available attributes from trip survey data (such as trip purpose, origin and destination) with attributes derived from smart card data (such as trip frequency) to estimate the trip purpose. The data was obtained at a single railway station.

Kuhlman (2015) applied a data fusion method to estimate trip purpose by analysing household travel survey data. The estimation success rate was dependent on the trip purpose. Although the method was not applied to actual smart card data, the validation results from HTS data supports the feasibility of trip purpose estimation with the attributes available from smart card data.

Different methods and techniques have been used to study passengers' travel behaviour and infer the trip purpose. User clustering has been used as one method for classifying passengers' travel behaviour (Trépanier et al., 2007, Agard et al., 2009, Ma et al., 2013). The combination of temporal and spatial variables has been used to define different clusters based on their travel behaviour. The temporal variables were classified as times during the day, frequency of travel during weekdays in a single week, and regularity of travel during a period of several weeks (Agard et al., 2006, Agard et al., 2009). As the spatial variable, stations and stops in the public transport network can be used to define spatial clusters (Ma et al., 2013). Although the clustering method has the capability to characterize trips based on the passengers' travel behaviour (Chu and Chapleau, 2010), clustering cannot capture the full complexity of travel patterns (Kim et al., 2014). In addition, clustering is based on temporal and spatial variables and does not reflect passengers' motivation for travelling, where a classification method based on trip purpose does (Kuhlman, 2015).

The rule-based processing approach has been also applied to trip purpose inference by integrating smart card data with other data sources (Kuhlman, 2015). On this basis, activity duration, departure time, frequency and card type have been the four attributes considered as the explanatory values of the trip purpose inference (Lee and Hickman, 2014, Kuhlman, 2015). Different rules and values have been used to identify these attributes. As examples, Chakirov and Erath (2012) identified the duration of work trips to be longer than six hours; Devillaine et al. (2012) identified the duration of work trips with adult cards in Gatineau, Canada to be longer than five hours; and, the duration for work trips with adult cards in Santiago, Chile was identified to be longer than two hours. However, trips that took place in the morning peak and had return trips in the evening peak were considered as work trips for employer-based smart cards (Lee and Hickman, 2014).

Although a number of studies in the literature used various approaches for trip purpose inference, these approaches cannot be generalised to other cases because:

1. There is a need for alighting estimation (as alighting information is not recorded in previous studies) before proceeding with trip purpose inference, which may incorporate cumulative errors in the inference. This allows the current study to focus on the main objective of the paper, which is trip purpose inference;
2. Only a few major trip purposes have been considered by different studies, namely work-related trips and education-related trips;
3. Few studies have considered passengers' travel patterns during weekdays as an important attribute (Langlois et al., 2016, Briand et al., 2017) for inferring trip purpose; and,
4. An aggregate inference approach has been chosen by most studies, while a disaggregate approach requires more investigation and provides better insight into the individual passenger's behaviour.

Hence, the aim of this paper is to build a trip purpose inference model based on the unique smart card dataset (Go Card data) available from South East Queensland (SEQ), Australia.

The remaining sections of this paper are organised as follows. The next section explains the research methodology, which comprises the logical inference framework, the data description and preparation and model structure. Then, the results of the model development and validation based on HTS are provided. Based on the validation results from HTS, the trip purpose inference results using Go Card data are presented and discussed. Finally, some conclusions and suggestions for future work are presented.

2. Methodology

The literature review has indicated attributes which might have explanatory value in the trip purpose inference. Several of these attributes are available in both HTS (temporal attributes) and the land use database (spatial attributes). In addition, the trip frequency is an additional attribute which can be extracted from public transport O-D surveys, which is used in the trip purpose inference model. As the purpose of the trip is not recorded by Go Card transactions, the trip purpose inference model is implemented, calibrated and validated using HTS data. Then, the same modelling procedure is applied to Go Card data. The following subsection discusses the logical inference framework that is used for the trip purpose inference as shown in Fig. 1.

2.1. Logical inference framework

Fig. 1 shows the logical framework of the proposed trip purpose inference model. The framework consists of four stages, namely: 'inputs', 'processes', 'modelling', and 'outcomes'. The modelling approach is developed based on the integration of different