Transportation Research Part C 68 (2016) 490–506

Contents lists available at ScienceDirect

Transportation Research Part C

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

Validating and improving public transport origin–destination estimation algorithm using smart card fare data ☆

Azalden Alsger *, Behrang Assemi, Mahmoud Mesbah, Luis Ferreira

School of Civil Engineering, The University of Queensland, Australia

ARTICLE INFO

Article history:
Received 22 August 2015
Received in revised form 18 March 2016
Accepted 6 May 2016

Keywords:
O–D estimation
Validation
Public transport smart card fare data
Trip-chaining method

ABSTRACT

Smart card data are increasingly used for transit network planning, passengers' behaviour analysis and network demand forecasting. Public transport origin–destination (O–D) estimation is a significant product of processing smart card data. In recent years, various O–D estimation methods using the trip-chaining approach have attracted much attention from both researchers and practitioners. However, the validity of these estimation methods has not been extensively investigated. This is mainly because these datasets usually lack data about passengers' alighting, as passengers are often required to tap their smart cards only when boarding a public transport service. Thus, this paper has two main objectives. First, the paper reports on the implementation and validation of the existing O–D estimation method using the unique smart card dataset of the South-East Queensland public transport network which includes data on both boarding stops and alighting stops. Second, the paper improves the O–D estimation algorithm and empirically examines these improvements, relying on this unique dataset. The evaluation of the last destination assumption of the trip-chaining method shows a significant negative impact on the matching results of the differences between actual boarding/alighting times and the public transport schedules. The proposed changes to the algorithm improve the average distance between the actual and estimated alighting stops, as this distance is reduced from 806 m using the original algorithm to 530 m after applying the suggested improvements.

1. Introduction

The use of smart card fare data has been rapidly increased in the field of transit network planning, behaviour analysis and demand forecasting (Pelletier et al., 2011). These data have become a valuable source of information for public transport origin–destination (O–D) estimation, allowing a better understanding of individuals' travel patterns and analysing the variability of transit users' behaviour (Morency et al., 2007; Kusakabe and Asakura, 2014; Kieu et al., 2015; Langlois et al., 2016).

Recently, a number of studies have used different methodologies to infer the O–D matrices for public transport trips using smart card fare data (Barry et al., 2009; Alfred Chu and Chapleau, 2008; Munizaga et al., 2010; Wang, 2010; Nassir et al., 2011; Gordon et al., 2013). Most automated fare collection systems record passengers' boarding information but not their alighting information. The lack of details on alighting stops is therefore the result of the currently used automated fare collection systems in which passengers are not required to tap off their cards upon alighting. Given this limitation of the data

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

* Corresponding author.
E-mail addresses: a.alsger@uq.edu.au (A. Alsger), b.assemi@uq.edu.au (B. Assemi), mahmoud.mesbah@uq.edu.au (M. Mesbah), l.ferreira@uq.edu.au (L. Ferreira).

http://dx.doi.org/10.1016/j.trc.2016.05.004
0968-090X/

produced by the majority of the existing systems, the O–D estimation results based on the smart card fare data needs to be re-evaluated, before its use in the analysis of individuals' travel behaviour (Bagchi and White, 2004; Munizaga et al., 2014).

The trip-chaining method, described later in this paper, is normally used to construct a passenger's travel sequence by connecting trip-legs recorded by his/her smart card usage. A few studies have attempted to evaluate this method and its assumptions. Farzin (2008) validated the 2006 estimated O–D results obtained from 5% of all transit trips in São Paulo, Brazil with the 1997 O–D household survey results. The results were not convincing due to the data limitations and the 10-year time-lapse between the datasets used for the comparison. Barry et al. (2009) validated the results obtained from the analysis of smart card data, with the data collected by passenger counting at the exit and entrance of subway stations as well as boarding and alighting at bus stops. Their study is based on two major assumptions: a very high percentage of passengers return to their previous alighting station to start their next trip, and a high percentage of passengers finally return to the first station they started their first trip of the day. Gordon et al. (2013) used the assumption of last destination as the closest to the first origin in their methodology. However, these assumptions need to be validated with a more reliable data source which includes accurate boarding and alighting details of public transport passengers.

The accessibility and quality of the additional data required for further evaluation of the trip-chaining method have been usually a big challenge. Devillaine et al. (2012) proposed a method for evaluating smart card analysis results with travel surveys, where the users' smart card IDs are recorded as a part of the survey. Chow (2014) evaluated an online approach to conduct customer surveys at a public transit agency by linking prior trip history into the survey. Munizaga et al. (2014) applied exogenous validation (information from travel surveys and personal interviews of a small sample of volunteers), in addition to endogenous validation (information from the same dataset), to validate the assumptions of the trip-chaining method, given the lack of alighting information in the main dataset.

Alsger et al. (2015) evaluated the common trip-chaining method assumptions using a unique smart card fare dataset obtained from TransLink, the public transport authority of South-East Queensland (SEQ), Australia. The important advantage of this dataset for the evaluation of the trip-chaining method assumptions is that it includes both boarding and alighting times and locations for each passenger of the public transport services that comprise buses, trains and ferries. The study focused on the individual assumptions (allowable transfer time, allowable walking distance and last destination of a given day being the same as the first origin of that day) of the trip-chaining method, in a situation where actual boarding and alighting information were known. Table 1 summarises the findings and gaps of the existing literature:

However, none of the above-mentioned studies has implemented and validated the whole estimation algorithm with a reliable dataset. Hence, the objective of this paper is to validate and improve the accuracy of the existing trip-chaining method through an in-depth evaluation of the public transport O–D matrices based on passengers' actual boarding and alighting data. The results highlight the impact of the method's assumptions on the accuracy of O–D estimation. Furthermore, a revised algorithm is proposed and empirically evaluated to improve the accuracy of the trip-chaining method.

The remaining sections of this paper are organised as follows. The next section explains the data description and preparation procedure. The research methodology is then described, which comprises the implementation of the existing O–D estimation method, the validating procedure and the improvement of the method by suggesting a revised algorithm. The results of the implementation and evaluation of the existing trip-chaining method are provided next. These results are then

Table 1
Summary of the findings and gaps of the existing literature.

Component		Studies	Findings	Gaps
Estimation assumptions	Walking distances (buffer zones)	Cui (2006), Wang (2010), Nassir et al. (2011), and Munizaga and Palma (2012)	Different walking distances were chosen to infer alighting stops (e.g., 400, 800, 1000 and 1100 m)	Different values were used for the assumptions of the O–D estimation. None of these studies have implemented and validated the whole estimation algorithm with a reliable dataset
	Transfer times	Bagchi and White (2004), Nassir et al. (2011), Kieu et al. (2013), Ma et al. (2013), and Hofmann and O'Mahony (2005)	Different transfer times were chosen to connect trip-legs to infer O–D trips (e.g., 30, 60 and 90 min)	
	Last destination assumptions	Barry et al. (2002), Nassir et al. (2011), Munizaga and Palma (2012), and Gordon et al. (2013)	Some studies assumed the last destination as the first origin, where others assumed it as the closest stop to the first origin	
Validation attempts	Additional data requirement for validation	Farzin (2008), Barry et al. (2009), Devillaine et al. (2012), Munizaga et al. (2014) and Chow (2014)	Additional data (e.g., travel survey, personal interviews, and passenger counting) were used for validation	The accessibility and quality of the additional data required for further evaluation of the trip-chaining method are usually a concern