Transportation Research Part C 58 (2015) 178–192

Contents lists available at ScienceDirect

Transportation Research Part C

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

Transit passenger origin–destination flow estimation: Efficiently combining onboard survey and large automatic passenger count datasets

Yuxiong Ji [a], Rabi G. Mishalani [b,*], Mark R. McCord [b]

[a] Tongji University, The Key Laboratory of Road and Traffic Engineering, Ministry of Education, Caoan Road 4800, Shanghai 201804, China
[b] The Ohio State University, Department of Civil, Environmental and Geodetic Engineering, 2070 Neil Ave., Rm 470, Columbus, OH 43210, USA

A R T I C L E I N F O

Article history:
Received 1 June 2014
Received in revised form 21 April 2015
Accepted 21 April 2015
Available online 4 August 2015

Keywords:
Transit route origin–destination (OD) flow estimation
Expectation Maximization (EM) method
Automatic Passenger Count (APC) data
Onboard OD flow surveys
Empirical evaluation
Estimation accuracy
Computational performance

A B S T R A C T

As transit agencies increasingly adopt the use of Automatic Passenger Count (APC) technologies, a large amount of boarding and alighting data are being amassed on an ongoing basis. These datasets offer opportunities to infer good estimates of passenger origin–destination (OD) flows. In this study, a method is proposed to estimate transit route passenger OD flow matrices for time-of-day periods based on OD flow information derived from labor-intensive onboard surveys and the large quantities of APC data that are becoming available. The computational feasibility of the proposed method is established and its accuracy is empirically evaluated using differences between the estimated OD flows and ground-truth observations on an operational bus route. To interpret the empirical differences from the ground-truth estimates, differences are also computed when using the state-of-the-practice Iterative Proportional Fitting (IPF) method to estimate the OD flows. The empirical results show that when using sufficient quantities of boarding and alighting data that can be readily obtained from APC-equipped buses, the estimates determined by the proposed method are better than those determined by the IPF method when no or a small sample sized onboard OD flow survey dataset is available and of similar quality to those determined by the IPF method when a large sample sized onboard OD flow survey dataset is available. Therefore, the proposed method offers the opportunity to forgo conducting costly onboard surveys for the purpose of OD flow estimation.

1. Introduction

Route-level origin–destination (OD) flow matrices indicate the number of passengers traveling from specific origin stops to specific destination stops on a single bus route. Such matrices are often considered in ridership forecasting, service planning (e.g., route extensions, route splitting or combination, and designing new routes), and operations control (e.g., short turning, expressing, and holding) (Cortes et al., 2011; Tirachini et al., 2011). Route-level OD matrices could also serve as inputs to network-level OD estimation methods (Cui, 2006) used in transit network planning and design applications.

Transit passenger OD flows are traditionally collected through onboard surveys. Such surveys are labor-intensive, time-consuming, and costly. As a result, the sample size is usually small, potentially resulting in large sampling errors in the survey OD flow data. Boarding and alighting counts at bus stops contain information useful in inferring OD flows.

* Corresponding author. Tel.: +1 (614) 292 5949.
E-mail addresses: yxji@tongji.edu.cn (Y. Ji), mishalani@osu.edu (R.G. Mishalani), mccrod.2@osu.edu (M.R. McCord).

http://dx.doi.org/10.1016/j.trc.2015.04.021
0968-090X/

Specifically, on a bus trip (the movement of a transit bus from its scheduled departure terminal at one end of a route to its scheduled destination terminal at the other end of a route), boarding and alighting counts at a bus stop are equal to the sum of OD flows originating from and destined to that bus stop, respectively. Since the "ride-check" surveys used to collect boarding and alighting count data are easier to implement than onboard surveys, several methods have been proposed to estimate transit route-level OD flow matrices from boarding and alighting counts (Ben-Akiva, 1987; Ben-Akiva et al., 1985; Hazelton, 2010; Ji et al., 2014; Kikuchi and Perincherry, 1992; Li, 2009; Li and Cassidy, 2007). The Iterative Proportional Fitting (IPF) procedure (Ben-Akiva et al., 1985) is one such method that is straightforward to implement, is computationally efficient, has been shown to produce similar or identical estimates as several other OD flow estimation methods (Ben-Akiva et al., 1985; Furth and Navick, 1992; Ji, 2011), has been found to perform surprisingly well in empirical studies (Lu, 2008; McCord et al., 2010; Mishalani et al., 2011), and could be considered to represent the state-of-the-practice in OD flow estimation.

Most applications of estimating transit OD flows from boarding and alighting counts were developed when obtaining boarding and alighting count data was still relatively labor-intensive and costly. As such, limited data were available for estimation, and the methods tended to rely on aggregating the counts by bus stop across bus trips to determine time-of-day period OD flow matrices. As a result, these methods (such as the IPF method) mainly rely on the sample means (i.e., the first moments) of boarding and alighting counts for OD estimation. Such methods may fail to take into account the complete information in the boarding and alighting count data when estimating passenger OD flows. Some studies considered additional information about the distribution of link flows (i.e., boarding and alighting counts when considering transit route-level networks), such as the variance-covariance matrix of such flows across time intervals (or boarding and alighting counts across bus trips) (Cascetta, 1984; Maher, 1983). However, these studies introduced limiting assumptions, such as the assumption that the variance-covariance matrix is known *a priori*.

Recently, transit agencies have been adopting various automated technologies for a variety of purposes. The operational use of these technologies leads to the amassing of large and varied data. For example, Automatic Fare Collection (AFC) systems have been implemented on many transit networks. Transit passenger OD flows could be derived from AFC data. However, many AFC systems, notably those for bus transit, are access-based (i.e., swipe-on or tap-on) only and, thus, only record the stops where passengers board and do not record the stops where passenger alight. As a result, assumptions, some of which are difficult to verify, need to be made for inferring individual passenger OD flows (Chan, 2007; Wang et al., 2011) from such AFC data.

Automatic Passenger Count (APC) technologies have also been deployed by many transit agencies to collect boarding and alighting counts at stops, predominantly for the purpose of measuring ridership and determining bus loads. The availability of the large amounts of boarding and alighting data collected via on-going use of APC systems makes it possible to utilize these data more effectively in attempts to arrive at better OD flow estimates. For example, a recently proposed IPF-based method (Ji et al., 2014) takes advantage of the availability of APC data on a large number of bus trips to iteratively improve the initially used uninformative base matrix. The large quantity of boarding and alighting data amassed by APC technologies was shown to be essential for this method to lead to improved OD flow estimates.

Regarding the consideration of more information about the distribution of link flows discussed above, the availability of large quantities of link flow data (such as boarding and alighting counts on transit routes) has motivated other recent studies to propose formulations and solution methodologies to estimate OD flows based on the full distribution of link flows, where depending on the application, the distribution is considered across time intervals or disaggregate trip-level boarding and alighting counts across bus trips (Hazelton, 2000, 2010; Li, 2005; Tebaldi and West, 1998; Vardi, 1996). A critical obstacle to implementing OD flow estimation methods based on the distribution of passenger counts is the need to enumerate the extremely large number of feasible OD flow matrices that satisfy the observed boarding and alighting counts for each bus trip (Hazelton, 2010; Li, 2005; Vardi, 1996). Such enumeration is computationally prohibitive for practical applications, which include realistically long transit routes, where the OD flow matrices are large (Hazelton, 2010; Li, 2005; Vardi, 1996). While a limited number of solutions to this difficulty have been proposed, computational challenges remain, and as far as the authors are aware, none has been demonstrated to be feasible in the presence of large APC datasets.

In this paper, a computationally feasible method that takes advantage of both large APC datasets and information derived from onboard survey data is proposed to estimate OD flow matrices. An empirical study is conducted that demonstrates the computational feasibility of the proposed method in combining large APC datasets with onboard survey data, evaluates the accuracy of the OD flow estimates against reliable ground-truth observations, explores the effect of the sample size of the APC data and the quality of the onboard survey data on the accuracy of the OD flow estimates, and investigates the computational performance of the proposed method. To put the empirical results in perspective, similar analyses are conducted using the easy-to-implement, computationally efficient, and extensively tested IPF method representing the state-of-the-practice. The IPF method for determining time-of-day period probability OD flow matrices is described first. The proposed method that incorporates APC and onboard OD flow survey data and the algorithm adapted to arrive at OD flow estimates are then presented. The data used in the empirical study, the experimental set-up, and the results follow. Finally, a discussion of the results and directions for future research are offered.

2. IPF method

The IPF method (Ben-Akiva et al., 1985) uses boarding and alighting counts for every stop along a bus route and a base OD flow matrix as inputs. The boarding and alighting counts are assumed to be measurement error-free and no assumptions are made on the distributions of the bus trip-level volume OD flows. As discussed previously, given boarding and alighting counts