Transportation Research Part B 91 (2016) 270–291

Contents lists available at ScienceDirect

Transportation Research Part B

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

Inferring origin-destination pairs and utility-based travel preferences of shared mobility system users in a multi-modal environment

Anshuman Anjani Kumar[a], Jee Eun Kang[a,*], Changhyun Kwon[b], Alexander Nikolaev[a]

[a] Department of Industrial and Systems Engineering, University at Buffalo, SUNY, United States
[b] Department of Industrial and Management Systems Engineering, University of South Florida, United States

ARTICLE INFO

Article history:
Received 22 July 2015
Revised 21 May 2016
Accepted 24 May 2016
Available online 9 June 2016

Keywords:
Origin-destination estimation
Traveler preferences
Expectation maximization
Probabilistic inference
Multimodal route choice
Bike sharing systems
Shared mobility systems

ABSTRACT

This paper presents a methodological framework to identify population-wide traveler type distribution and simultaneously infer individual travelers' Origin-Destination (OD) pairs, based on the individual records of a shared mobility (bike) system use in a multimodal travel environment. Given the information about the travelers' outbound and inbound bike stations under varied price settings, the developed Selective Set Expectation Maximization (SSEM) algorithm infers an underlying distribution of travelers over the given traveler "types," or "classes," treating each traveler's OD pair as a latent variable; the inferred most likely traveler type for each traveler then informs their most likely OD pair. The experimental results based on simulated data demonstrate high SSEM learning accuracy both on the aggregate and dissagregate levels.

1. Introduction

Bike sharing systems are gaining prominence in the United States and world-wide as a viable shared mobility option. In driving green transportation initiatives, such systems are expected to alleviate the congestion in urban areas and provide commuters with additional travel utility as well as health and socio-demographic benefits. Researchers position bike sharing systems as a solution to the "first and last mile problem," stimulating users to switch to public transit modes and avoid relying on personal vehicles for reaching transit stations (Liu et al., 2012). The main challenge of shared mobility system operation is that, as many travelers tend to follow similar routes, the decreasing vehicle counts in trip origin areas (and parking spot counts in destination areas) cause vehicle imbalance across multiple stations in these areas. These operational issues are currently handled by system managers in a reactive manner; however, recent research (Haider et al., 2014) suggests a more pro-active solution—a strategic offering of incentives to travelers so as to reduce the imbalance build-up. The challenge is that in such efforts, the knowledge of travel demand and traveler preferences is critical for calculated incentive (pricing) planning.

Extensive analyses of network data and customer surveys have been conducted to understand traveler needs on an aggregate level (Vogel and Mattfeld, 2010). However, analytical and mathematical optimization models found in the literature

* Corresponding author.
E-mail address: jeeeunka@buffalo.edu (J.E. Kang).

http://dx.doi.org/10.1016/j.trb.2016.05.015
0191-2615/

have had limited success inferring *individual* traveler behavior (Vogel and Mattfeld, 2010). Descriptive analyses of shared vehicle usage patterns, reported in the recent past (Borgnat et al., 2009; Froehlich et al., 2009; Vogel and Mattfeld, 2010), may help this cause.

In order to parametrize a Mixed Multinomial Logit model (Hensher and Greene, 2003), often used to describe traveler routing decisions (Ben-Akiva and Lerman, 1985; Bovy and Hoogendoorn-Lanser, 2005; Wardman, 2004), true OD pair information is required. The main deficiency of OD pair estimation approaches that neglect the multimodal nature of transit is that they lose the information of demand elasticity and flexibility. Prior research has examined the problem of disaggregate multimodal (bus and metro) OD matrix estimation at the stop level (Munizaga and Palma, 2012), using automatic fare collection system records of boarding counts for two different transit mode systems. Stop-level OD pair estimation based on system data (e.g., traffic count, passenger count) has also been carried out (Abrahamsson, 1998; Lam et al., 2003; Li and Cassidy, 2007; Wong and Tong, 1998), including a study that exploited fare card transaction data (Lee and Hickman, 2014). However, no method exists that does disaggregate (i.e., individual-based) inference of traveler preferences from stop-level OD pair information or automated fare collection, and then uses these inferred preferences to distill “true” (i.e., not stop-level) OD pairs.

The OD estimation problem is paid much attention in transportation modeling and planning research. This problem is often referred to as the trip demand estimation problem, where an estimate of OD trip demand matrix is to be computed using traffic flow data and other available information (Cascetta and Nguyen, 1988). Several model formulations and heuristic methods have been proposed so far: they employ diverse theoretical approaches including the minimization of the sum of squares of the predicted and observed OD matrix value differences (Bell, 1991; Cascetta and Nguyen, 1988), column generation (Garcia-Rodenas and Verastegui-Rayo, 2008; Sherali and Park, 2001), bi-level formulations (Lundgren and Peterson, 2008; Yang, 1995), entropy/information based inference (Xie et al., 2010; 2011; van Zuylen, 1978), and path flow estimation (Chen et al., 2010; Nie et al., 2005; Ryu et al., 2014). Bayesian methods, exploiting the properties of certain families of parametrized distributions (Castillo et al., 2008; Hazelton, 2008; Maher, 1983; Mahmassani and Sinha, 1981; Tebaldi and West, 1998), have found use for updating the trip generation parameters and generating the trip matrix. Notably, recent work using Bayesian inference combines OD estimation with route choice behaviour (Sun et al., 2015) to assign passenger flows in networks. There, a route choice model is first developed, and then, using observable passenger data, the model parameters for the route choice model are calibrated.

The problem attacked in the present paper is even more complex, as it involves two unknowns: the trip ODs and the traveler preferences expressed via a set of feasible traveler types. In addition, the existing approaches are based on *aggregate* system (zonal/station/stop) level OD estimation. Meanwhile, for more effective pricing/incentive program implementation, the information of *individual*, i.e., *disaggregate*, traveler ODs and preferences is crucial, particularly for bike sharing systems and more broadly shared-mobility systems. The presented framework is developed to estimate both the individual trip OD and traveler types, using Bayesian logic, which is enabled by collecting multiple responses of same individuals (through user id or cards). It allows one to infer the OD-demand not at the station level but at the more granular traveler (“true” OD) level, exploiting a data driven approach, coupled with Bayesian learning, that enables the mining of trip details for each individual traveler.

This paper presents a methodological framework for traveler preference (represented as a traveler type) and OD pair inference in complex multimodal transportation systems. The developed Selective Set Expectation Maximization (SSEM) algorithm allows for estimating the unknown traveler type distribution by treating the OD pairs as latent variables, using the information about the changes in traveler route choices under varied circumstances. Due to the flexibility of system operation and the dynamic nature of the bike sharing systems, such systems offer a convenient test bed for implementing the presented estimation method. The estimation framework relies on the observed traveler responses to pricing incentives, road closures or extreme weather events. Such partial route information pieces can be collected from automatic fare collection system-type data (or from GPS or user pass-card data) as travelers respond to system perturbations. Using the partial information about a traveler’s route under multiple price settings, it becomes possible to identify eligible OD points from a set of points in a geographical zone with certain belief/probability. We thereby manage to gauge the sensitivity of travelers to incentives by learning the travel utility—traveler type—distribution over the population of travelers. This distribution indicates how the travelers valuate travel options by time, price and convenience: a combination of the disutility weights for these respective measures defines a traveler type; the feasible range of the traveler types is assumed to be discrete, finite and given. Then, one can infer the bikers’ origins and destinations, by first inferring the distribution of the traveler types in the entire population. A Bayesian model is used in the SSEM algorithm to quantify the likelihood of the observed data and the model parameters are adjusted in an iterative manner to maximize this likelihood.

The SSEM algorithm works to assess the sensitivity of bike-using travelers to pricing incentives, offered at outbound and inbound bike stations and varied over multiple scenarios. The algorithm can be viewed as an extension of the conventional Expectation Maximization (EM) algorithm. The EM algorithm has an issue of favoring the extreme traveler types: from each pricing scenario, it tends to conclude that travelers are either too sensitive or not-at-all sensitive to pricing incentives. The SSEM algorithm, on the other hand, infers *a range* of suitable traveler types from each scenario; an intersection of these ranges for a given traveler over all the scenarios is taken as the most likely traveler type for this traveler. The SSEM algorithm is also enhanced by the data pre-processing stage that effectively screens out infeasible route choice solutions for each traveler. In the reported computational studies, the SSEM algorithm is able to infer the true OD pairs for the travelers with the accuracy of about 75% by correctly learning the population’s sensitivity to pricing incentives.