Transportation Research Part C 52 (2015) 116–131

Contents lists available at ScienceDirect

Transportation Research Part C

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

An integrated Bayesian approach for passenger flow assignment in metro networks

Lijun Sun [a,b], Yang Lu [c,*], Jian Gang Jin [d], Der-Horng Lee [b], Kay W. Axhausen [a,e]

[a] Future Cities Laboratory, Singapore-ETH Centre, Singapore 138602, Singapore
[b] Department of Civil & Environmental Engineering, National University of Singapore, Singapore 117576, Singapore
[c] Singapore-MIT Alliance for Research and Technology (SMART), Singapore 138602, Singapore
[d] School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
[e] Institute for Transport Planning and Systems (IVT), ETH Zürich, Zürich CH-8093, Switzerland

ARTICLE INFO

Article history:
Received 20 September 2014
Received in revised form 2 January 2015
Accepted 2 January 2015
Available online 13 February 2015

Keywords:
Bayesian inference
Metro network
Travel time
Smart card
Route choice

ABSTRACT

This paper proposes an integrated Bayesian statistical inference framework to characterize passenger flow assignment model in a complex metro network. In doing so, we combine network cost attribute estimation and passenger route choice modeling using Bayesian inference. We build the posterior density by taking the likelihood of observing passenger travel times provided by smart card data and our prior knowledge about the studied metro network. Given the high-dimensional nature of parameters in this framework, we apply the variable-at-a-time Metropolis sampling algorithm to estimate the mean and Bayesian confidence interval for each parameter in turn. As a numerical example, this integrated approach is applied on the metro network in Singapore. Our result shows that link travel time exhibits a considerable coefficient of variation about 0.17, suggesting that travel time reliability is of high importance to metro operation. The estimation of route choice parameters conforms with previous survey-based studies, showing that the disutility of transfer time is about twice of that of in-vehicle travel time in Singapore metro system.

1. Introduction

With the increasing demand and range of urban mobility, metro systems are playing more and more important roles in urban transportation, particularly in high-density mega-cities. Taking Singapore's Mass Rapid Transit (MRT) system as an example, around two million metro trips were made daily in the year 2012. Compared with other transport modes, metro systems have dedicated and exclusive rail-based infrastructures, making it possible to provide superior service with higher speeds and larger capacity. Due to their superiority, metro systems not only attract but also suffer from high passenger demand – especially during rush hours when passenger demand exceeds its designed capacity for not only trains, but also platforms – experiencing over-crowdedness, disturbances and disruptions time and again. These factors bring about negative effects on passenger's traveling experience and therefore should be minimized. From operators' point of view, understanding passenger demand and flow assignment patterns in a complex metro network becomes crucial to maintaining service reliability and developing efficient failure response strategies.

* Corresponding author at: Singapore-MIT Alliance for Research and Technology (SMART), 1 Create Way, #09-01, CREATE Tower, Singapore 138602, Singapore

E-mail addresses: lijun.sun@ivt.baug.ethz.ch (L. Sun), luyang@smart.mit.edu (Y. Lu), jiangang.jin@sjtu.edu.cn (J.G. Jin), dhl@nus.edu.sg (D.-H. Lee), axhausen@ivt.baug.ethz.ch (K.W. Axhausen).

http://dx.doi.org/10.1016/j.trc.2015.01.001

To characterize a passenger flow assignment model for metro network, two factors are of the most importance: origin–destination (O–D) demand matrix and route choice behavior. Because of the widely adopted tap-in-tap-out fare collection system, the station-to-station O–D matrix in a metro network is known; however the route choice decisions are usually not directly observable, therefore a widely used approach is to first develop a route choice model – characterized by some critical cost attributes influencing passenger perception, such as in-vehicle time, number of transfers and fare paid – and then employ observed preference data to calibrate model parameters. Despite the mathematical modeling, in principle there are two crucial issues to be solved in this approach before applying it on metro networks. The first is to accurately measure each attribute in the model, such as different stages of travel time and transit fares mentioned above. These values are used as input and assumed to be known in advance. In practice, experimenters need to determine such network properties by using train operation data and field surveys. However, accurate evaluation of route attributes, such as in-vehicle time, waiting time and transfer time, could be challenging considering possible congestion or interruption scenarios. The second issue is to obtain enough field observations, which register individual route choice preferences to support parameter estimation. However, in practice one may encounter many difficulties. On one hand, in the absence of detailed train operation logs recording train departure/arrival time and trajectories, it is difficult to measure exact network attributes, such as in-vehicle time, waiting time and transfer time. On the other hand, as most metro networks are designed as closed systems where passengers only leave traces at boarding/alighting stations for the purpose of fare collection, operators have limited knowledge on passenger route choice and trajectory within the system. In other words, we know little about which train or which transfer station an individual passenger has taken during his/her trip in the case where multiple alternative routes exist. In order to obtain passenger route choice preference data, a conventional approach is to conduct field surveys in train stations, asking people the exact route they will take to reach their destinations. However, some shortcomings of these methods have been identified, such as being subject to bias and errors and being both time-consuming and labor-intensive in conducting surveys and processing the data. In addition, since most surveys are conducted with focus on particular location and time, the results are often limited in scale and diversity. As a result, developing alternative methods to reveal individual route choice preference in large-scale networks remains challenging.

The emergence and wide deployment of automated fare collection (AFC) systems open a new data-driven approach for metro network analysis. Taking advantage of smart card-based fare collection systems, in which individual passenger's tapping-in/out transactions are recorded, researchers are now able to better understand metro operation with large quantities of real-world observations (Pelletier et al., 2011). Such data set also provides us with a good opportunity to study passenger behavior in a data-driven approach. In doing so, researchers have tried to combine passenger travel time information with train operation logs (Kusakabe et al., 2010; Sun and Xu, 2012; Zhou and Xu, 2012). However, without further investigating travel time variability, these approaches essentially assume that train services are always punctual to timetables and hence network cost attributes are assumed to be deterministic, even though there is clear evidence showing that train services can be delayed/disrupted by excessive passenger demand. On the other hand, owing to the uncertainty in travel time, the difficulties in revealing individual trajectory from tap-in/tap-out information still remain, preventing us from collecting accurate preference data. In view of these unsolved issues, this paper presents the development and empirical verification of a new integrated metro assignment framework using Bayesian inference approach. Taking advantage of large quantities of real-world observations provided by smart card data, the suggested model simultaneously estimate network attributes and passenger route choice preference. Consequently, the proposed framework utilizes only travel time observations along with static network data to construct the passenger flow assignment model in a closed metro network. With low social-economic cost and implementation convenience, such approach is appealing for metro operations and maintenance.

Bayesian inference method is a well established statistical model which has been applied to various transportation applications, including O–D estimation, route choice modeling and flow assignment inference (Hazelton, 2008, 2010; Wei and Asakura, 2013). It enables us to find a posterior distribution which integrates all our prior knowledge with the available observations. Although in this sense it is a powerful tool for our inference problem, in practice it is difficult to implement such models owing to the difficulty in computing the Bayesian posterior analytically. However, thanks to the rapid increase of computational power, nowadays we can characterize properties of the Bayesian posterior using computational approaches, of which the most notable one is Markov Chain Monte Carlo (MCMC) methods (Robert and Casella, 2004; Robert, 2014). The proposed framework in this paper is also based on solution algorithms provided by MCMC methods. In general, the MCMC approach is used for complex models where maximum likelihood is difficult to calculate using conventional optimization methods (e.g., Newton–Raphson algorithm and Expectation–Maximization (EM) algorithm) or prior knowledge is important and should be integrated.

The contribution of this paper is threefold. First, we construct an integrated network characterization and flow assignment framework through a data-driven approach, allowing us to better understand passenger route choice behavior from large quantities of smart card observations. Second, by taking travel time variability caused by possible interruption during metro operation into consideration, our model can better characterize network travel time and its uncertainty given any O–D pairs, providing better travel information to metro users. Finally, as will be shown in the following, the Bayesian formulation has the capacity to estimate network cost attributes and characterize passenger route choice model in a simultaneous manner, showing great potential in practice, in particular in cities with large/complex metro networks such as Beijing, London, New York, Seoul and Tokyo.

This remainder of this paper is organized as follows: in Section 2, we review previous studies on several related topics, including travel time reliability, passenger route choice behavior, the use of smart card data in understanding metro