Transportation Research Part C xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Transportation Research Part C

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

Inferring left behind passengers in congested metro systems from automated data

Yiwen Zhu[a,*], Haris N. Koutsopoulos[a], Nigel H.M. Wilson[b]

[a] *Northeastern University, Boston, MA 02115, USA*
[b] *Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

ARTICLE INFO

Keywords:
Left behind
Automated data
Passenger assignment
Maximum likelihood estimation
Bayesian estimation
MCMC sampler

ABSTRACT

With subway systems around the world experiencing increasing demand, measures such as passengers left behind are becoming increasingly important. This paper proposes a methodology for inferring the probability distribution of the number of times a passenger is left behind at stations in congested metro systems using automated data. Maximum likelihood estimation (MLE) and Bayesian inference methods are used to estimate the left behind probability mass function (LBPMF) for a given station and time period. The model is applied using actual and synthetic data. The results show that the model is able to estimate the probability of being left behind fairly accurately.

1. Introduction

In recent years, transit ridership has experienced large increases, especially in major metropolitan areas. The Mass Transit Railway (MTR) in Hong Kong for example, experienced an 83.8% ridership increase in the past 10 years (MTR Coporation, 2016), and Transport for London (TfL) has seen an increase of 55.7% in public transportation ridership since 2000 (Mayor of London, 2015). With rapid population growth, London anticipates a further 21% increase in ridership by 2041 even if the trip rate per person remains the same. As a result, many systems operate at (or near) capacity during peak hours. Even with very short headways, systems cannot cope with the large demand and are often congested with passengers left behind, unable to board the first train due to the capacity constraints. Level of service can deteriorate significantly with overcrowding on station platforms and trains, long queues at escalators, unreliable journey times, and potential safety concerns.

To better deal with the increasing demand, operators are interested in: (i) measuring the impact on passengers due to near capacity operations, (ii) better understanding the performance of the system especially when it is stressed by large demand, and (iii) developing short-term strategies to improve capacity utilization.

The availability of Automated Fare Collection (AFC) and Automated Vehicle Location (AVL) data affords the opportunity to monitor system operations and facilitate the development of relevant metrics to measure passenger experience, such as crowding levels, excess waiting time due to limited capacity, etc. (Bagchi and White, 2005; Agard et al., 2006; Zhao et al., 2007; Chan, 2007; Pelletier et al., 2011; Ortega-Tong, 2013; Kieu et al., 2015; Langlois et al., 2016). Left behind passengers due to overcrowding has become a major concern for many transit operators (Delgado et al., 2012; Sun and Xu, 2012).

To address some of the questions raised above, detailed models using automated data have been proposed, aiming at "assigning" passengers to individual trains. The models reproduce the passenger flow through the network and estimate the capacity utilization of the system (see for example, (Buneman, 1984; Kusakabe et al., 2010; Paul, 2010; Zhu, 2014; Zhu et al., 2017b)).

* Corresponding author.
E-mail address: zhu.yiwen@microsoft.com (Y. Zhu).

https://doi.org/10.1016/j.trc.2017.10.002
Received 29 August 2017; Received in revised form 27 September 2017; Accepted 1 October 2017
0968-090X/

However, the problem of estimating aggregate metrics for peak hour operations, such as passengers left behind, has not received a lot of attention in the literature. While aggregate metrics can be calculated from the output of the passenger-to-train assignment models mentioned above, they are still too complicated and time consuming to apply. Moreover, for peak hour applications, when passengers may be left behind at origin/transfer stations by several trains, passenger assignment models become very difficult to apply because of the large number of possible itineraries. Agencies sometimes hire surveyors to count the number of passengers left behind after a train departs, however this process is expensive, inefficient and inaccurate.

The objective of this paper is to propose an efficient method to directly infer passenger movements and left behind probabilities from AFC and AVL data, for systems in which entry and exit station transaction data are recorded. The output of the model supports various applications:

1. Developing performance metrics from the passenger's point of view, such as the probability distribution of the number of times a passenger is left behind.
2. Providing important input to some passenger assignment models (Sun and Schonfeld, 2015), such as the left behind probabilities, which can potentially improve the accuracy of those models and simplify their structure. These assignment models can also be used to analyze train-loading and crowding levels from the operator's point of view (Zhu et al., 2017b).

The remainder of this paper is organized as follows. Section 2 develops the model and describes the estimation approaches. Section 3 validates the model using synthetic data generated from actual AFC transactions and train movement data. Section 4 applies the model using actual data during the peak period, and compares the results with manual observations. It also evaluates the system performance under sudden increases in demand (stress test). Section 5 concludes the paper.

2. Methodology

This paper examines the problem of estimating the probability distribution of the number of times passengers are left behind at a station in a given time period. Two data sources are used: (i) fare transaction records from a closed AFC system (i.e. a system where passengers both tap-in and tap-out), and (ii) train tracking data from the AVL system which provides station arrival and departure times. Assuming that the number of times a passenger is left behind is a random variable with an unknown probability mass function, two approaches are used to estimate the left behind probability mass function (LBPMF). The first method is based on a maximum likelihood formulation of the problem and the second uses Bayesian inference to estimate the posterior distribution of the LBPMF.

The problem can be viewed as identifying the underlying groups of passengers based on how many times they were unable to board a train. In the transportation literature, a number of problems have been studied using similar approaches. Barnhart et al. (2014) and Yan et al. (2016) for example, developed discrete choice models to assign passengers to flight itineraries in air transportation. Zhou and Xu (2012) and Zhou et al. (2015) developed passenger assignments models based on recorded entry and exit times from AFC data in Metro systems. Zhang and Yao (2015) utilized AFC data to estimate various travel times including walking, waiting, transfer, and in-vehicle travel times. Kazagli and Koutsopoulos (2013), used a mixture model to identify latent driver populations with different travel behavior using observations of travel times from an Automated Number Plate Recognition (ANPR) system. Fu et al. (2014) estimated the posterior probabilities of choosing different routes by formulating the travel times as mixture distributions using smart card data (each corresponding to a different route). Lee and Sohn (2015) introduced the number of routes used as an unknown variable in a Bayesian framework and used reversible-jump Markov Chain Monte Carlo simulation for the estimation of the posterior distribution of different routes. A potential drawback of using only travel times to infer route shares is the inability to capture crowding effects. In practical applications, during peak hours, longer journey times may be caused by severe congestion instead of passenger choosing an alternative route with a longer journey time. In this paper, train tracking data is used along with AFC data to estimate the left behind probabilities given capacity constraints.

2.1. Problem definition

We assume a closed AFC system, where the tap-in/out times of passengers are known. Train arrival/departure times at stations are also known from the train control and signaling systems (AVL). Fig. 1 shows the movement of a passenger who enters the system at t^{in} and exits at t^{out}. The estimation uses data from trips without transfers and route choice. For further applications, it can be assumed that passengers with and without transfers experience the same left behind probabilities at the same station (Zhu et al., 2017a). The problem involving route choice is discussed in Zhu et al. (2017a). Access time is defined as the time for a passenger to walk from the (tap-in) fare gate to the platform; waiting time is the time waiting on the platform; and egress time is the time to walk to the (tap-out) fare gate after alighting.

To be conservative, the minimum access and egress times are set to zero. The passenger in Fig. 1 can board one of three trains (1,2,3). In this paper, we define, for passenger i, a train j to be feasible if it satisfies the following conditions:

1. Departs the origin station after the passenger's arrival at the platform:

$$t_i^{in} + \tau_i^a \leqslant DT_j. \quad (1)$$