Hindawi
Journal of Advanced Transportation
Volume 2017, Article ID 5824051, 12 pages
https://doi.org/10.1155/2017/5824051

Research Article

Estimating Train Choices of Rail Transit Passengers with Real Timetable and Automatic Fare Collection Data

Wei Zhu,[1] Wei Wang,[2] and Zhaodong Huang[3]

[1] *College of Transportation Engineering, Tongji University, Shanghai 201804, China*
[2] *School of Mathematical Sciences, Tongji University, Shanghai 200092, China*
[3] *College of Maritime and Transportation, Ningbo University, Ningbo 315211, China*

Correspondence should be addressed to Wei Zhu; zhuweimail@163.com

Received 20 March 2017; Revised 14 June 2017; Accepted 5 July 2017; Published 15 August 2017

Academic Editor: Wai Yuen Szeto

Copyright © 2017 Wei Zhu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

An urban rail transit (URT) system is operated according to relatively punctual schedule, which is one of the most important constraints for a URT passenger's travel. Thus, it is the key to estimate passengers' train choices based on which passenger route choices as well as flow distribution on the URT network can be deduced. In this paper we propose a methodology that can estimate individual passenger's train choices with real timetable and automatic fare collection (AFC) data. First, we formulate the addressed problem using Manski's paradigm on modelling choice. Then, an integrated framework for estimating individual passenger's train choices is developed through a data-driven approach. The approach links each passenger trip to the most feasible train itinerary. Initial case study on Shanghai metro shows that the proposed approach works well and can be further used for deducing other important operational indicators like route choices, passenger flows on section, load factor of train, and so forth.

1. Introduction

Passenger flow is the foundation of making and coordinating operation plans for an urban rail transit (URT) system, while assigning passenger flows on the URT network plays a paramount role in analyzing (calculating, predicting, and simulating) passenger flows. A number of transit assignment models have been developed using both theory and practical experience, and thorough reviews were presented in some of the literature [1–3]. However, different from urban road traffic systems, a URT system is operated according to relatively punctual schedule, which is an important constraint for a URT passenger's travel. Thus, the passenger flow distribution on the network is subjected to not only passengers' physical route choices but also their individual train choices especially in peak hours (Figure 1), which may be a more important issue [4]. For analyzing passenger flows on a schedule-based URT network, it is the key to estimate passengers' train choices for threefold reasons:

(1) On a schedule-based URT network, passenger route choices as well as flow distribution on the network can be deduced if the train choices of passengers are obtained, but that is not so either.

(2) It can give more precise estimation results for both spatial and temporal dimensions, since URT passengers may fail to board on a train in certain conditions especially in peak hours because of the overcrowding.

(3) These pieces of information would be further useful for improving the customer relationship management of a URT company and for improving train timetables, if each passenger's train choice can be identified over a long period of time. For example, URT companies can check how passengers select trains after timetable improvements.

As mentioned, there are a number of transit assignment models developed for analyzing passengers flows on the network. In those models, in order to obtain passenger route choice preference data, a conventional approach is to conduct field surveys in rail stations, asking passengers about the exact route they took to reach their destinations. However, the shortcomings of these methods have been identified

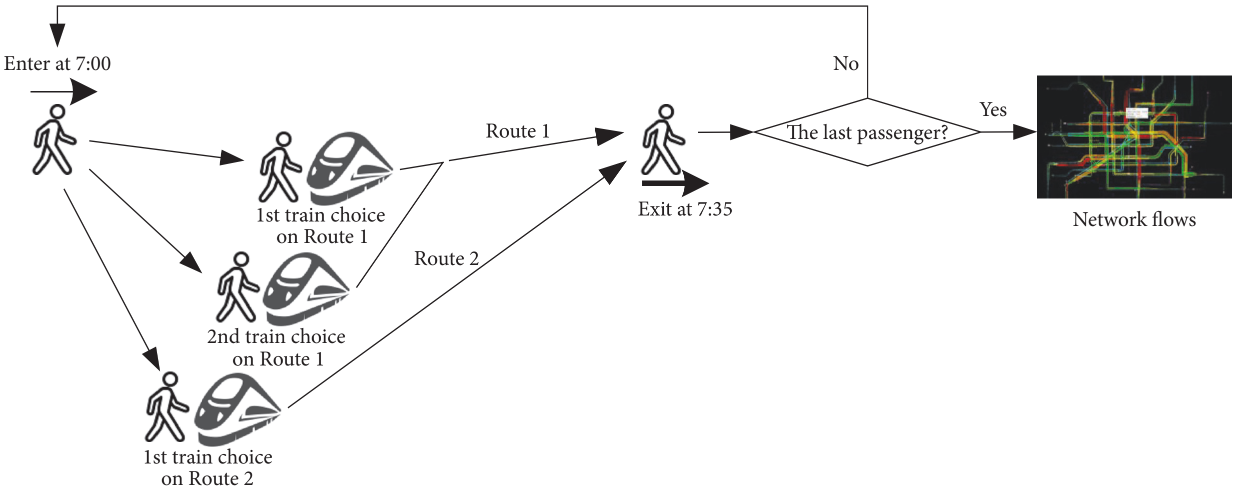

Enter at 7:00
No
Yes
Route 1
The last passenger?
1st train choice on Route 1
Route 2
Exit at 7:35
Network flows
2nd train choice on Route 1
1st train choice on Route 2

Figure 1: Relationship among train choice, route choice, and flow assignment.

by more and more researchers. For example, the resulting data from these manual methods may be subject to bias and error and is expensive and time consuming both to collect and to process [5]. In addition, the manual methods usually focus only on particular location and time [6, 7]. As a result, alternative concepts and methods need to be developed.

In recent years, automatic fare collection (AFC) data such as smart card data have been used by transit service providers to analyze passenger demand and system performance. These data have been used for O-D matrices estimation [8, 9], demand analysis [10, 11], travel behavior analysis [12], operational management, and public transit planning [13–15], and so forth. In particular, there are emerging studies dealing with AFC data of URT systems. Some impressive publications include works by Chan in 2007 [16], Kusakabe et al. in 2010 [17], Xu et al. in 2011 [7], Sun et al. in 2012 and 2016 [18, 19], Zhou and Xu in 2012 [20], Fu et al. in 2014 [21], Zhu et al. in 2014 [22], and Sun et al. in 2015 [23]. However, in spite of the widespread attention on the use of AFC data, there are fewer studies dealing with the passenger train choice behavior in a URT system. Kusakabe et al. [17] developed a methodology for estimating which train would be boarded by each smart card holder using long-term transaction data. Their approach was based on the assumption that smart card data that could not be identified to the possible train choices would be assigned with equal probability. Zhou and Xu [20] developed a passenger flow assignment model based on entry and exit time constraints from AFC data. The model includes an algorithm for generating path's boarding plan which is similar to passenger train choice. However, the matching degree employed in this algorithm is more intuitive than rigorously defined. Sun and Schonfeld [19] proposed a schedule-based passenger's path-choice estimation model using AFC data. The model uses the train schedule connection network (TSCN) which considers passengers' behaviors of boarding on and alighting from the train. However, a weighted assignment used by the model may be not appropriate for a factual travel choice process which uses only one route at the same time rather than multiroutes. And the problem will further become more obvious for those O-D pairs with fewer passenger trips.

For better understanding of passenger flows on network, the objective of this paper is to propose a methodology that can estimate passenger train choices with real timetable and AFC data. The contributions of this paper are presented as follows:

(1) We formulate the addressed problem using Manski's [24] paradigm on modelling choice, which consists of generating consideration choice set and calculating corresponding choice probability.

(2) An integrated framework for estimating passenger train choices is developed. The approach links each AFC transaction (a passenger trip) to the most feasible train itinerary (a boarding plan).

(3) Real timetable and AFC data are investigated as the inputs to the proposed methodology, instead of relying on manual methods.

The remainder of this paper is organized as follows. In Section 2, the estimation problem of passenger train choices is described and formulated. Section 3 presents the integrated estimation framework. In particular, methods of deducing passenger boarding plan, choice probability, and travel behavior parameters are developed with real timetable and AFC data. Section 4 demonstrates a case study of the proposed approach. Finally, Section 5 concludes the paper.

2. Formulating the Problem

The topic discussed in this paper falls in the scope of choice modelling. From a variety of studies [24, 25] it is well known that the size and composition of choice sets do matter in cases

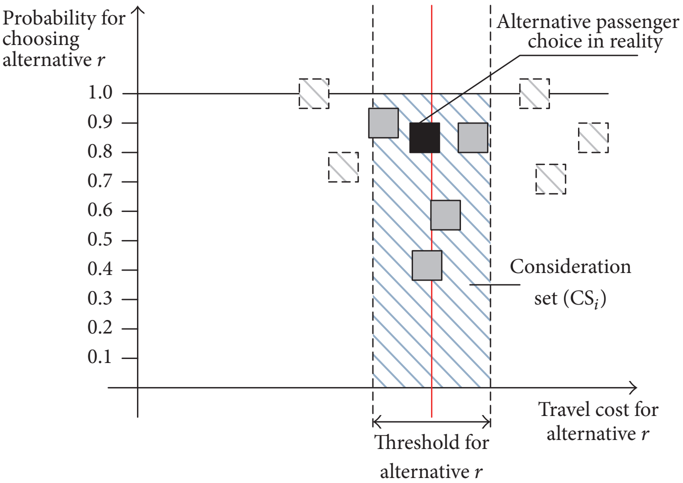

Probability for choosing alternative r
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
Alternative passenger choice in reality
Consideration set (CS_i)
Travel cost for alternative r
Threshold for alternative r

Figure 2: Formulating the estimation problem of URT passenger train choices.

of choice model estimation. Incorrect choice sets can lead to misspecification of choice models [26, 27]. And, furthermore, for a variety of reasons the specification of train choice sets for train choice modelling is different from and more complex than mode choice and route choice, which is why this topic deserves our special attention.

To clearly formulate the estimation problem, Manski's [28] paradigm on predicting choice is used. The essential conceptual contribution of this paradigm lies in its explicit treatment of the processes making perfect predictions of choice behavior unattainable. Up to date, most of the existing literature on random utility models still generally imposes distributional assumptions directly and consequently this practice has often caused researchers to remain unaware of the restrictiveness of their models because it leaves so much implicit information.

Manski's paradigm states that the probability of passenger i to choose alternative r from the choice set CS_i, which is also called his/her consideration set, is given by the following expression:

$$P_i(r \mid \mathrm{US}_i) = \sum_{\mathrm{CS}_i \in \mathrm{US}_i} p_i(r \mid \mathrm{CS}_i)\, p(\mathrm{CS}_i \mid \mathrm{US}_i), \tag{1}$$

where $P_i(r \mid \mathrm{US}_i)$ is the probability that passenger i will choose alternative r from the universal set US_i of all alternatives available to i, $p_i(r \mid \mathrm{CS}_i)$ is the conditional probability that passenger i will choose alternative r given that CS_i is his/her consideration set where CS_i is a subset of US_i, and $p(\mathrm{CS}_i \mid \mathrm{US}_i)$ is the probability that CS_i is the consideration set of passenger i given his/her universal set US_i.

Thus, the corresponding solution for estimating an individual passenger's train choices with schedule and AFC data can consist of two works: one is generating the consideration set (CS_i) of his/her train choices. And the other is calculating the probability $p_i(r \mid \mathrm{CS}_i)$ that he/she will choose alternative r from CS_i.

The above addressed solution also can be depicted as in Figure 2. The horizontal axis indicates the travel cost for alternative r, and the vertical axis indicates the probability that passenger i chooses alternative r. As we use travel time as cost measure in this study, "cost" and "time" are treated the same (interchangeable) throughout the paper. The *red* vertical line indicates the observed travel time of passenger i extracted from his/her AFC transaction record. Each alternative in his/her consideration set (CS_i) can be plotted as a dot in the figure. Then, *how to estimate which train itinerary the passenger chose in reality? It seems natural that the alternative, which is close to the red vertical line with higher probability, is most likely to be used by the passenger.*

3. Methodology

3.1. Overview of Estimation Procedure. For an individual passenger, his/her *train choice solution* during the travel can be depicted as *a boarding plan* which is the order of trains that he/she can take to complete his/her travel. The overall framework of our estimation procedure for this kind of boarding plan is shown in Figure 3. At the beginning of the algorithm, denoted by "a," the AFC data are extracted from the original transaction data and sorted with fields of origin station, entry time, destination station, and exit time, which will be used later. After these data are sorted, several travel behavior parameters of passengers are extracted from abundant timetable and AFC data, which is denoted by "b." Then, one of the records of the AFC transaction data (which is also a passenger trip) is extracted for estimation. To generate the consideration set, boarding plan generation algorithm is applied at "c." And at "d," calculating choice probability of boarding plan is executed. At "e," the train choice solution (which equals a boarding plan), the passenger choice is determined based on the probability of each alternative in the consideration set. These processes are repeated until all of the records are estimated.

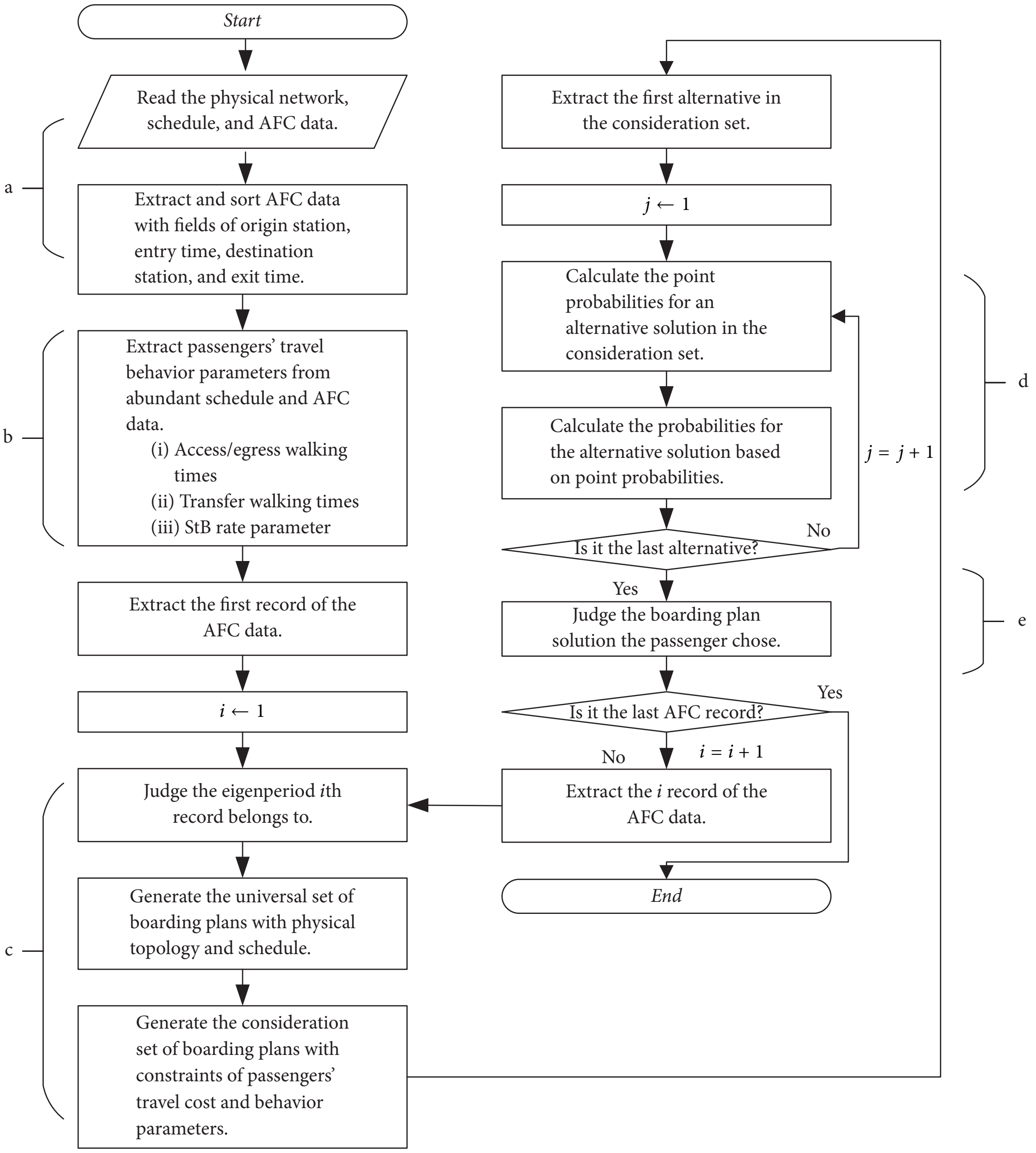

Start
Read the physical network, schedule, and AFC data.
a
Extract and sort AFC data with fields of origin station, entry time, destination station, and exit time.
b
Extract passengers' travel behavior parameters from abundant schedule and AFC data.
(i) Access/egress walking times
(ii) Transfer walking times
(iii) StB rate parameter
Extract the first record of the AFC data.
$i \leftarrow 1$
c
Judge the eigenperiod ith record belongs to.
Generate the universal set of boarding plans with physical topology and schedule.
Generate the consideration set of boarding plans with constraints of passengers' travel cost and behavior parameters.
Extract the first alternative in the consideration set.
$j \leftarrow 1$
Calculate the point probabilities for an alternative solution in the consideration set.
d
Calculate the probabilities for the alternative solution based on point probabilities.
$j = j + 1$
Is it the last alternative?
No
Yes
Judge the boarding plan solution the passenger chose.
e
Is it the last AFC record?
Yes
No
$i = i + 1$
Extract the i record of the AFC data.
End

Figure 3: Overall framework of estimation procedure.

3.2. Generating Boarding Plan Set

3.2.1. Universal Set Generation. This is a two-step part as shown in Figure 4. Due to the URT system's networked operation, there may be several alternative routes for a given O-D pair, and passengers in practice will choose not only the shortest route but also the second, third, . . ., kth shortest route for their imperfect knowledge of the network, individual differences, factor of congestion, and so forth. First, an improved Deletion Algorithm (DA) [29] based on Depth-First Traversal (DFT) is introduced to find the kth shortest route, and the initial route choice set of the O-D pair is obtained. Second, for each route in the initial route choice set, all the boarding choices of a given passenger at each boarding station (origin, destination, or transfer station) on the route are deduced with the corresponding schedule data. And then the universal set of the passenger's boarding plans can be obtained.

The improved DA based on DFT is provided as follows. Different from other k-shortest path algorithm, it will not miss any possible route including ring routes.

Step 1. Determine the shortest tree of directed graph (N, A) rooted at origin s based on the Dijkstra Algorithm. Let P_k be the shortest path from origin s to destination t in (N, A). Note that $k = 1$.

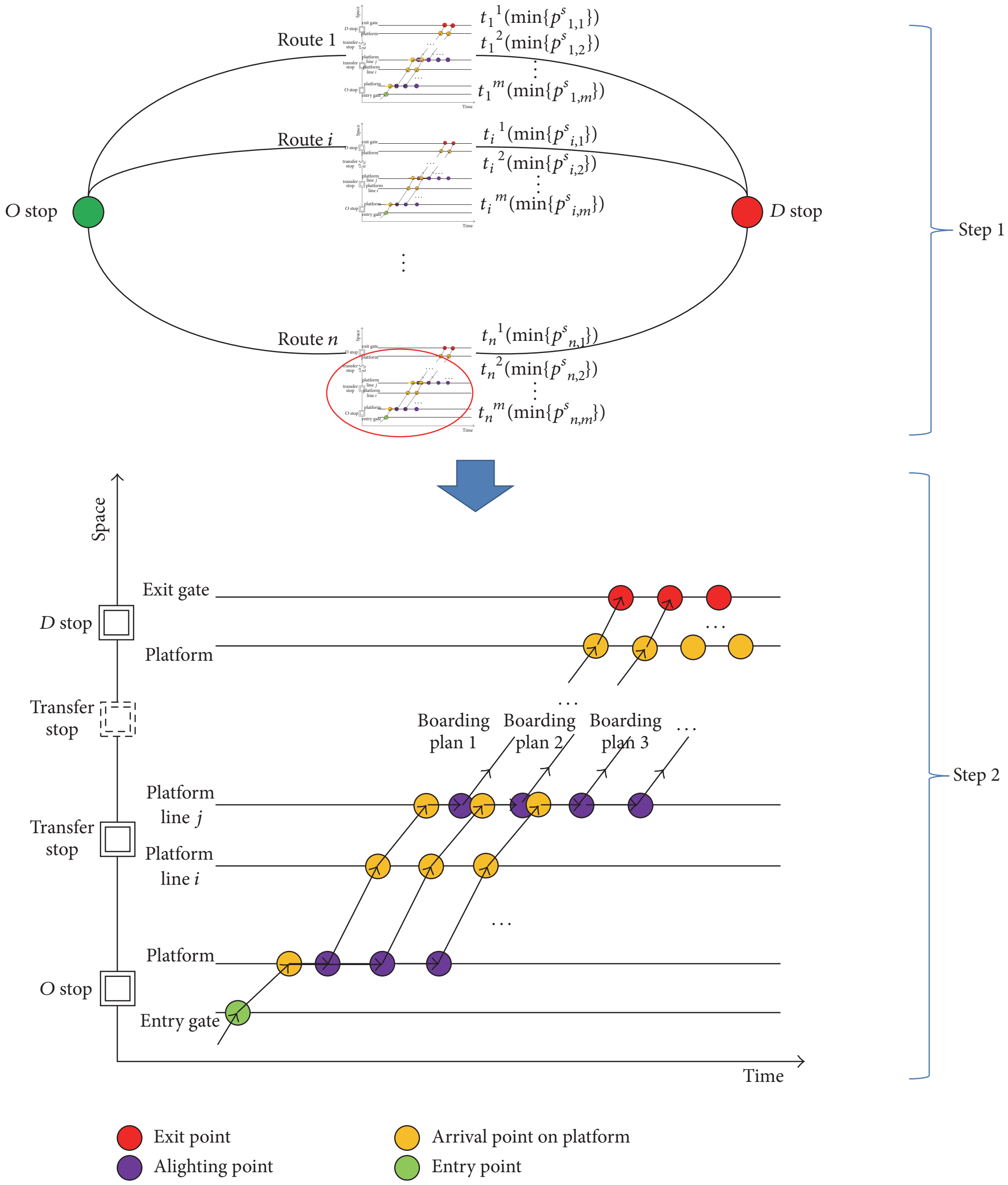

Route 1
$t_1{}^1(\min\{p^s{}_{1,1}\})$
$t_1{}^2(\min\{p^s{}_{1,2}\})$
$t_1{}^m(\min\{p^s{}_{1,m}\})$
Route i
$t_i{}^1(\min\{p^s{}_{i,1}\})$
$t_i{}^2(\min\{p^s{}_{i,2}\})$
$t_i{}^m(\min\{p^s{}_{i,m}\})$
O stop
D stop
Route n
$t_n{}^1(\min\{p^s{}_{n,1}\})$
$t_n{}^2(\min\{p^s{}_{n,2}\})$
$t_n{}^m(\min\{p^s{}_{n,m}\})$
Step 1
Space
Exit gate
D stop
Platform
Transfer stop
Boarding plan 1
Boarding plan 2
Boarding plan 3
Platform line j
Transfer stop
Platform line i
Platform
O stop
Entry gate
Time
Step 2
Exit point
Alighting point
Arrival point on platform
Entry point

FIGURE 4: Illustration of two-step universal set generation.

Step 2. If k does not exceed K, which is the maximum number of the kth shortest paths, and there is still an alternative path in (N, A), let $P = P_k$ and proceed to Step 3; otherwise, the algorithm stops.

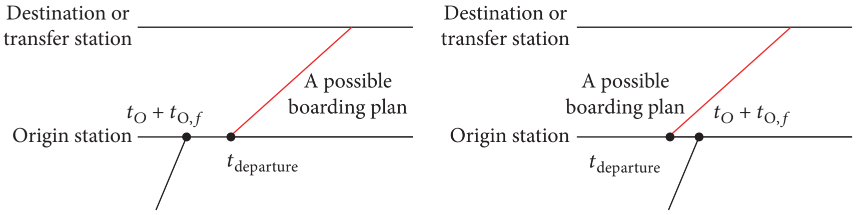
Step 3. Let $I(x)$ denote the set of incoming arcs to node x. Let n_h denote the first node of current path p for which $I(n_h) > 1$. If the primed node n'_h of node n_h is not in N, proceed to Step 4; otherwise, let n_j denote the first node of current path P without n_h if the node's primed node is in $\{N\}$ and proceed to Step 5.

Step 4. Add n'_h to N and $\{(x, n'_h) \mid (x, n_h) \in A \text{ and } x \neq n_h - 1\}$ to A. Let d_x denote the value of the shortest distance from s to x. Compute dn'_h and find the shortest path from s to n'_h. Let $n_i = n_{h+1}$.

Step 5. Let n_j denote any note following $n_j \in P$. Then execute as follows.

Step 5.1. Add the primed node n'_j of node n_j to N.

Step 5.2. Add $\{(x, n'_j) \mid (x, n_j) \in A \text{ and } x \neq n_{j-1}\} \cup \{(n'_{j-1}, n'_j)\}$.



(a) The boarding plan is reasonable ($(t_O+t_{O,f}) < t_{\text{departure}}$)

(b) The boarding plan is unreasonable ($(t_O + t_{O,f}) > t_{\text{departure}}$)

Figure 5: Comparisons of the calculated departure time of the possible boarding plan versus the passenger's arrival time on the platform. *Note.* t_O is the entry time of the given passenger at the origin station; $t_{O,f}$ is the walking time of passenger with the fastest speed from the entry gate to the platform at the origin station.

Step 5.3. Compute dn'_j and find the shortest path from s to n'_j.

Step 6. Let P be the shortest path from s to the primed node $t'_{(k)}$ of node t in (N, A), so that P is the best alternative path of P_{k-1}. Set $k = k + 1$ and proceed to Step 2.

Moreover, considering the influence from congestion, a passenger may fail to board and has to wait for the next train. The maximum "fail to board" (FtB) number is set to 3 based on investigations in China, which means a passenger can board on a train within four runs even if the congestion during peak hours makes the passenger be unable to board on the first train.

3.2.2. Consideration Set Generation. A boarding plan for a given passenger is the order of trains that the passenger can take to complete his/her travel. Obviously, it is difficult to determine which train the passenger board in reality. However, usually the passenger is rarely delayed in the process of walking out of the destination station, and consequently the train he or she alighted from can be determined accurately. Thus, we can calculate from the destination station to the origin station backward. For a given trip data (AFC transaction data) obtained from the URT system, a boarding plan is considered unreasonable and should be removed from the universal set if its boarding time at origin station is impossible for the passenger given the constraint of his/her entry time (Figure 5).

Therefore, a filtering algorithm can be developed to further narrow the universal set and get the consideration set. The algorithm is described as follows.

Step 1. Obtain possible boarding plans (universal set). For an actual passenger trip, with the corresponding train diagram, the passenger's exit time, and walking time at the destination station, possible boarding plans for each route can be easily deduced.

Step 2. Calculate the departure time of a possible boarding plan. Based on the passenger's travel chain combined with train diagrams, the departure time $t_{\text{departure}}$ of the possible boarding plan of each route can be calculated from the destination station to the origin station backward.

Step 3. Compare and remove. As shown in Figure 5, the calculated departure time $t_{\text{departure}}$ of the possible boarding plan at the origin station is compared with the passenger's arrival time $(t_O+t_{O,f})$ on the platform. If $(t_O+t_{O,f}) < t_{\text{departure}}$, the boarding plan is reasonable for the passenger to choose; otherwise, the boarding plan is unreasonable and removed from the universal set.

3.3. Calculating Choice Probability of Boarding Plan

3.3.1. Point Probability Calculation. For a given boarding plan in the obtained consideration set, we name a boarding station (origin, destination, or transfer station) in the boarding plan as a boarding point. So, the point probabilities of a boarding plan need to be calculated firstly.

It should be noted that passengers may fail to board the train in certain conditions especially in peak hours because of the overcrowding, though they are usually inclined to board on the first train as we know. Therefore, without loss of generality, we use "point probability" to present the probability for a passenger to board on the train within a given boarding plan. For a boarding point w in plan i, the probability of leaving with the train for a passenger is p_{wj} that can be obtained directly from the StB (success to board) rates as shown in Figure 6.

3.3.2. Plan Probability Calculation. The plan probability is the function of the point probabilities. Considering that the boarding point with minimum probability is the bottleneck for the boarding plan to be chosen, instead of the product of those probabilities at all boarding points, we adopt the following function:

$$p_w = \min\left\{p_{wj}\right\}, \tag{2}$$

where p_w is the probability of plan w.

For example (as shown in Figure 6), suppose there are two boarding plans in the consideration set. For plan 1, the point probability is 0.66 for the train within the given boarding plan at origin station and 0.27 at transfer station. For plan 2, the point probability is 0.34 for origin station and 0.73 for

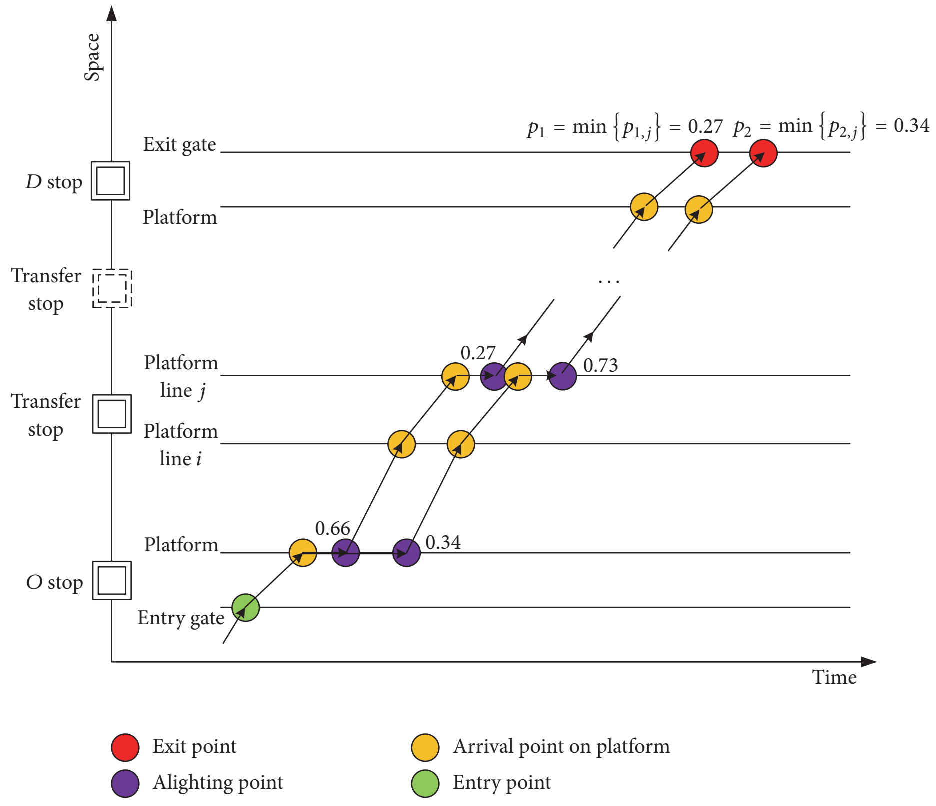


Figure 6: Illustration to choice probabilities of boarding plans.

transfer station. Then, the probabilities for the two plans can be calculated easily as follows:

$$\begin{aligned}p_1 &= \frac{\min\left\{p_{1j}\right\}}{\sum_i \min\left\{p_{ij}\right\}} \\ &= \frac{\min\{0.66, 0.27\}}{\min\{0.66, 0.27\} + \min\{0.34, 0.73\}} \\ &= \frac{0.27}{0.27 + 0.34} = 0.443, \\ p_2 &= \frac{\min\left\{p_{2j}\right\}}{\sum_i \min\left\{p_{ij}\right\}} \\ &= \frac{\min\{0.34, 0.73\}}{\min\{0.66, 0.27\} + \min\{0.34, 0.73\}} \\ &= \frac{0.34}{0.27 + 0.34} = 0.557.\end{aligned} \tag{3}$$

3.4. Extracting Travel Behavior Parameters. As mentioned, we also need to extract in advance several travel behavior parameters of passengers using abundant AFC data resource, for both of boarding plan set generation and choice probability calculation. These parameters include, for each station on the network, minimum access walking time ($t_{\min}^{\text{access}}$), maximum access walking time ($t_{\max}^{\text{access}}$), minimum egress walking time ($t_{\min}^{\text{egress}}$), maximum egress walking time ($t_{\max}^{\text{egress}}$), minimum transfer walking time ($t_{\min}^{\text{transfer}}$), maximum transfer walking time ($t_{\max}^{\text{transfer}}$), and "success to board" (StB) rate (r_{StB}). Walking time parameters are used for generating the consideration sets, while StB rate parameter is used for calculating the choice probabilities of boarding plans.

3.4.1. Access/Egress Walking Time Extraction. First, we deduce parameters of $t_{\min}^{\text{egress}}$ and $t_{\max}^{\text{egress}}$ at every station on the network based on AFC data. It should be noticed that passengers may be delayed at the origin station and transfer stations by passenger flow, the capacity utilization rate of the train, and other factors but are rarely delayed in the process of walking out of the destination station. Thus, it is easier to deduce the parameters of $t_{\min}^{\text{egress}}$ and $t_{\max}^{\text{egress}}$. By matching the train's arrival time derived from schedule data and passengers' exit time derived from AFC data, passengers' egress walking times can be obtained and its distribution can be extracted too:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \times \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]. \tag{4}$$

It is a kind of normal distribution and can be calibrated with the AFC data. Then, we set the minimum egress walking time ($t_{\min}^{\text{egress}}$) using the 5th percentile of the calibrated distribution and the maximum egress walking time ($t_{\max}^{\text{egress}}$) using the 95th percentile of the calibrated distribution.

Second, we try to get parameters of $t_{\min}^{\text{access}}$ and $t_{\max}^{\text{access}}$ at every station on the network. It is noticed that passengers may be delayed during their walking process of access to platform,

Table 1: Distribution of passengers boarding on different trains during 8:00 AM~9:00 AM at Yanchang Rd.

O-D	First train	Second train	Third train	Fourth train	Sum
Yanchang Rd. → Zhongshan Bei Rd.	100	80	50	20	250
Yanchang Rd. → Shanghai Railway Station	200	100	50	30	380
Yanchang Rd. → Hanzhong Rd.	250	200	40	25	515
Yanchang Rd. → Xinzha Rd.	150	100	30	10	290
Yanchang Rd. → People's Square	200	160	35	25	420
Sum	900	640	205	110	1855

Table 2: The parameter of StB of down direction during 8:00 AM~9:00 AM at the station of Yanchang Rd.

StB									
First train		Second train		Third train		Fourth train		Sum	
Trips	p_1	Trips	p_2	Trips	p_3	Trips	p_4	Trips	P
900	0.485	640	0.345	205	0.111	110	0.059	1855	1.000

which makes distribution of access walking times different from egress walking times, and passengers' exact arrival times on platform also cannot be obtained directly. However, we can still suppose $t_{\min}^{\text{access}} = t_{\min}^{\text{egress}}$ and $t_{\max}^{\text{access}} = t_{\max}^{\text{egress}}$, since there is some symmetrical characteristic between the processes of a passenger's access and egress, and we just want to obtain the threshold rather than the exact distribution.

3.4.2. Transfer Walking Time Extraction. In order to extract parameters of $t_{\min}^{\text{transfer}}$ and $t_{\max}^{\text{transfer}}$, two assumptions are adopted in advance as follows:

(1) The walking speed of the same passenger should be on the same level in his or her trip train. In other words, for a given passenger, the walking speeds at stations (origin station, destination station, or transfer station) should not be different from each other to a great extent.

(2) The delay caused by crowding, high-capacity utilization of the train, and similar factors for an individual passenger happens in the origin station as well as transfer stations with equal probability.

(3) Last but not least, we just try to extract the threshold rather than the exact distribution.

Then, the minimum transfer walking time ($t_{\min}^{\text{transfer}}$) and maximum transfer walking time ($t_{\max}^{\text{transfer}}$) at a transfer station can be calculated as follows.

Step 1. Aggregate the AFC data whose O-D flows use the given transfer station as their unique transfer point.

Step 2. Calculate the egress walking speeds with egress walking times and distances (d^{egress}) and set the transfer walking speeds using the egress walking speeds; that is,

$$\begin{aligned} v_{\max}^{\text{transfer}} &= v_{\max}^{\text{egress}} = \frac{d^{\text{egress}}}{t_{\min}^{\text{egress}}}, \\ v_{\min}^{\text{transfer}} &= v_{\min}^{\text{egess}} = \frac{d^{\text{egress}}}{t_{\max}^{\text{egress}}}. \end{aligned} \tag{5}$$

Step 3. Calculate the transfer walking times at the transfer station with the calculated transfer walking speeds and distances (d^{transfer}); that is,

$$\begin{aligned} t_{\max}^{\text{transfer}} &= \frac{d^{\text{transfer}}}{v_{\min}^{\text{transfer}}}, \\ t_{\min}^{\text{transfer}} &= \frac{d^{\text{transfer}}}{v_{\max}^{\text{transfer}}}. \end{aligned} \tag{6}$$

3.4.3. StB Rate Parameter Extraction. At last, we deduce the parameter of StB (success to board) rate. Assuming StB is a direct outcome of overcrowding which is mostly true in peak periods, we can conclude that as long as passengers depart from the same station in the same direction and period, the StB parameter is the same. In that case we can use those O-D flows without any transfers (and hence no alternative route) to estimate the StB parameter. And then, we can consequently apply those parameters to O-D flows with transfers.

The parameter of StB can be defined as a vector as follows:

$$R = (p_0, p_1, p_2, p_3), \tag{7}$$

where p_0, p_1, p_2, and p_3 are the probabilities that passengers succeed to board on the first, second, third, and fourth train, and obviously all items in the vector sum up to 1.

Taking a case from the Shanghai metro network, for example, if we want to calculate the StB of down direction during 8:00 AM~9:00 AM at Yanchang Rd. Station of Line number 5, we can use the data of those O-D flows without any transfers, including Yanchang Rd. → Zhongshan Bei Rd., Yanchang Rd. → Shanghai Railway Station, Yanchang Rd. → Hanzhong Rd., Yanchang Rd. → Xinzha Rd., and Yanchang Rd. → People's Square. Table 1 shows the distribution of passengers boarding on different trains during 8:00 AM~9:00 AM at Yanchang Rd. And based on Table 1, the StB of down direction during 8:00 AM~ 9:00 AM at the station of Yanchang Rd. can be deduced (Table 2).

Table 3: Samples of passenger trip records.

Number	Origin station	Destination station	Entry time	Exit time
1	Xingzhi Rd.	Xinzhuang	07:00:06	07:50:47
2	Xingzhi Rd.	Xinzhuang	07:00:05	07:51:00
3	Xingzhi Rd.	Xinzhuang	07:21:07	08:12:45
4	Xingzhi Rd.	Xinzhuang	07:25:51	08:19:05
5	Xingzhi Rd.	Xinzhuang	07:26:05	08:18:45
⋮	⋮	⋮	⋮	⋮

Table 4: The real timetable of trains on Route 1.

Xingzhi Rd. of Line 7		Changshu Rd. of Line 7		Changshu Rd. of Line 1		Xinzhuang of Line 1	
Arrival	Departure	Arrival	Departure	Arrival	Departure	Arrival	Departure
07:01:30	07:01:45	07:19:42	07:20:06	07:24:43	07:25:01	07:46:47	—
07:07:30	07:07:45	07:25:42	07:26:06	07:29:00	07:29:18	07:51:04	—
07:13:30	07:13:45	07:31:42	07:32:06	07:31:26	07:31:44	07:53:30	—
07:19:30	07:19:45	07:37:42	07:38:06	07:33:52	07:34:10	07:55:56	—
07:22:45	07:23:00	07:40:57	07:41:21	07:36:18	07:36:36	07:58:22	—
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table 5: The real timetable of trains on Route 2.

Xingzhi Rd. of Line 7		Dongan Rd. of Line 7		Dongan Rd. of Line 4		Shanghai Indoor Stadium of Line 4		Shanghai Indoor Stadium of Line 1		Xinzhuang of Line 1	
Arrival	Departure	Arrival	Departure	Arrival	Departure	Arrival	Departure	Arrival	Departure	Arrival	Departure
07:01:30	07:01:45	07:24:27	07:24:42	07:28:41	07:28:56	07:32:32	07:33:00	07:36:00	07:36:27	07:51:04	—
07:07:30	07:07:45	07:30:27	07:30:42	07:33:41	07:33:56	07:37:32	07:38:00	07:38:26	07:38:53	07:53:30	—
07:13:30	07:13:45	07:36:27	07:36:42	07:38:41	07:38:56	07:42:32	07:43:00	07:40:52	07:41:19	07:55:56	—
07:19:30	07:19:45	07:42:27	07:42:42	07:43:41	07:43:56	07:47:32	07:48:00	07:43:18	07:43:45	07:58:22	—
07:22:45	07:23:00	07:45:42	07:45:57	07:48:41	07:48:56	07:52:32	07:53:00	07:45:44	07:46:11	08:00:48	—
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

4. Case Study

4.1. Test O-D Pair. For the purpose of approach test, a case study is conducted on a specific O-D pair (from Xingzhi Rd. to Xinzhuang) on the Shanghai metro network. As shown in Figure 7, there are three routes connecting the original station (Xingzhi Rd.) and the destination station (Xinzhuang), which can be obtained by improved DA based on DFT. The first route moves through Line 7 and Line 1, with the transfer station: Changshu Rd. The second route moves through Line 7, Line 4, and Line 1, with the transfer stations: Dongan Rd. and Shanghai Indoor Stadium. The third route travels through Line 7, Line 9, and Line 1, with the transfer stations: Zhaojiabang Rd. and Xujiahui. The theoretic travel times of the three routes are 2970 seconds, 3485 seconds, and 3488 seconds, respectively.

4.2. Data Used in the Test. In the test, 57 passenger trips records between 07:00 AM and 08:00 AM and obtained from the AFC system are used to verify the proposed approach. Table 3 gives a sample record from these 57 passenger trips.

Moreover, as another important input of the proposed approach in this paper, the corresponding real timetables of the relevant URT lines (e.g., Line 1, Line 4, Line 7, and Line 9) were obtained from automatic train supervision (ATS) system and used too. Tables 4–6 show the samples of this data.

4.3. Results and Discussions. Using the above input data, the boarding plan estimation for these 57 passenger trips is performed with the proposed approach. Table 7 gives a sample of the estimation results. As can be seen in the table, each passenger trip (which equals an AFC transaction record) derived from the AFC system can be assigned to the unique boarding plan by the proposed approach.

As mentioned, for a schedule-based URT system, the result in Table 7 is the key for passenger flow analysis, based on which other important indicators (e.g., route choices, passenger flows on section, and load factor of train, as shown in Table 8 and Figure 8) can be deduced furthermore.

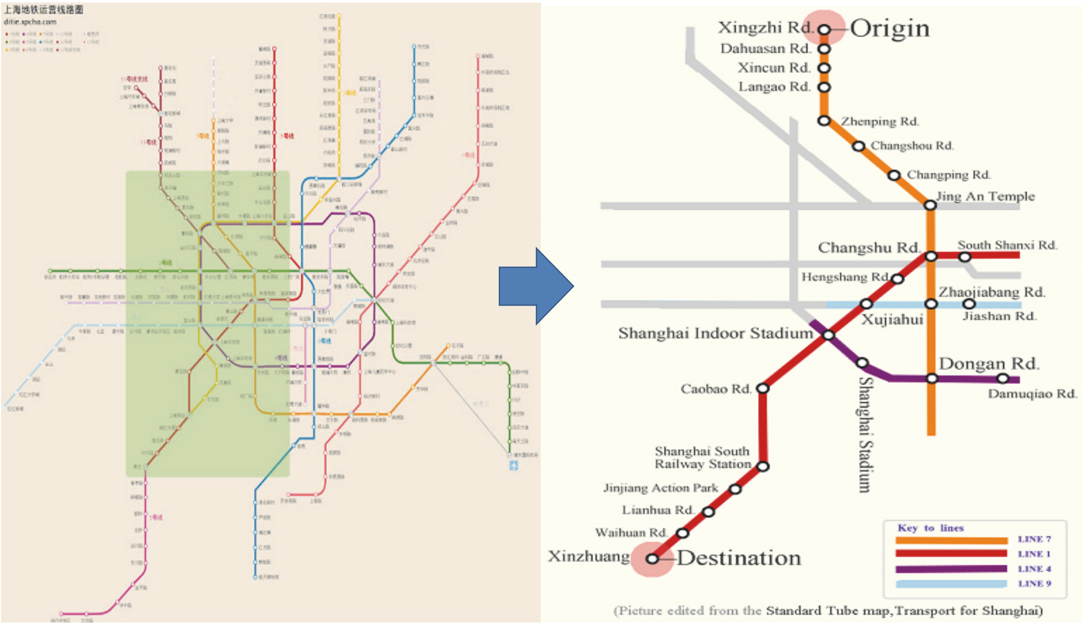
Companying with the above case study, some extended discussions can be further made. Previous studies use discrete choice analysis extensively to predict passenger choice

Table 6: The real timetable of trains on Route 3.

Xingzhi Rd. of Line 7		Zhaojiabang Rd. of Line 7		Zhaojiabang Rd. of Line 9		Xujiahui of Line 9		Xujiahui of Line 1		Xinzhuang of Line 1	
Arrival	Departure	Arrival	Departure	Arrival	Departure	Arrival	Departure	Arrival	Departure	Arrival	Departure
07:01:30	07:01:45	07:22:25	07:22:43	07:27:43	07:28:07	07:30:48	07:31:06	07:33:03	07:33:30	07:51:04	—
07:07:30	07:07:45	07:28:25	07:28:43	07:31:43	07:32:07	07:34:48	07:35:06	07:35:29	07:35:56	07:53:30	—
07:13:30	07:13:45	07:34:25	07:34:43	07:35:43	07:36:07	07:38:48	07:39:06	07:37:55	07:38:22	07:55:56	—
07:19:30	07:19:45	07:40:25	07:40:43	07:39:43	07:40:07	07:42:48	07:43:06	07:40:21	07:40:48	07:58:22	—
07:22:45	07:23:00	07:43:40	07:43:58	07:43:43	07:44:07	07:46:48	07:47:06	07:42:47	07:43:14	08:00:48	—
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table 7: Samples of estimated boarding plans for passenger trips.

Number	Passenger entry time	Passenger exit time	Boarding plan		Route	
			Train ID	Boarding time	Route number	Route description
1	07:00:06	07:50:47	031-411	07:01:30–07:24:43	1	Line 7 → Line 1
2	07:00:05	07:51:00	031-411	07:01:30–07:24:43	1	Line 7 → Line 1
3	07:21:07	08:12:45	321-211-197	07:22:45–07:48:41–07:55:28	2	Line 7 → Line 4 → Line 1
4	07:25:51	08:19:05	067-503	07:29:15–07:53:20	1	Line 7 → Line 1
5	07:26:05	08:18:45	067-503	07:29:15–07:53:20	1	Line 7 → Line 1
6	07:02:06	07:52:48	042-432	07:07:30–07:29:00	1	Line 7 → Line 1
7	07:00:58	07:53:18	042-432	07:07:30–07:29:00	1	Line 7 → Line 1
8	07:07:02	07:55:34	042-476	07:07:30–07:31:26	1	Line 7 → Line 1
9	07:20:11	08:09:58	321-533	07:22:45–07:43:36	1	Line 7 → Line 1
10	07:33:15	08:24:40	164-236-097	07:35:45–07:59:43–08:04:41	3	Line 7 → Line 9 → Line 1
11	07:39:33	08:31:33	344-047-113	07:42:15–08:08:41–08:14:56	2	Line 7 → Line 4 → Line 1
12	07:41:07	08:30:58	403-136	07:45:30–08:05:30	1	Line 7 → Line 1
13	07:38:41	08:36:32	403-047-075	07:45:30–08:08:41–08:17:22	2	Line 7 → Line 4 → Line 1
14	07:40:34	08:36:52	344-369	07:42:15–08:07:43–08:14:25	3	Line 7 → Line 9 → Line 1
15	07:58:09	08:53:26	401-059-119	08:01:45–08:27:43–08:33:53	3	Line 7 → Line 9 → Line 1
⋮	⋮	⋮	⋮	⋮	⋮	⋮



(Picture edited from the Standard Tube map, Transport for Shanghai)

Figure 7: Rail network connecting the test O-D pair (Xingzhi Rd. → Xinzhuang).

Table 8: Route choices deduced from estimated boarding plans.

Route number	Route description	Passengers	Proportion (%)
1	Line 7 → Line 1	34	59.6
2	Line 7 → Line 4 → Line 1	12	21.1
3	Line 7 → Line 9 → Line 1	11	19.3

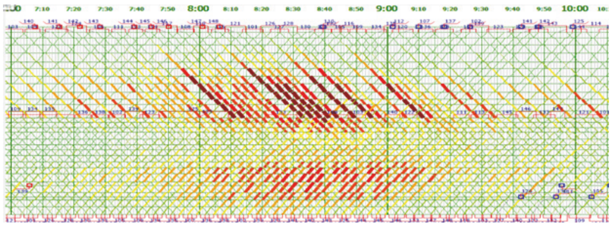


Figure 8: Example of factual train diagrams with passenger flows on sections and load factors of trains.

behavior. Such a model requires preference data and still displays great variability in real-world estimation. Recently, in this context, some researchers try to reveal route choice from observed passenger travel time derived from smart card system (Kusakabe et al., 2010; Sun and Xu, 2012; Zhou and Xu, 2012; Zhu et al., 2014; Fu et al., 2014; Sun et al., 2015). As demonstrated in the case study, we figure out the key issue of estimating passenger boarding plans, based on which all the route choice, section flow, load factor, and so forth can be deduced, furthermore, and no longer depend on the assumption that smart card data that could not be identified to the possible train choices would be assigned with equal probability (Kusakabe et al., 2010). Furthermore, the proposed approach improves the methodologies of Sun and Schonfeld [19] and Zhou and Xu [20] on calculating passenger boarding plans. On the other hand, compared to the study efforts presented in [21, 23], our approach models the problem of interest considering the temporal dynamics induced by demand profiles, service timetables, and crowdedness.

5. Conclusions

A URT system is operated based on its schedules. Different from those urban road traffic systems, it is more important to estimate passengers' train choices based on which passenger route choices as well as flow distribution on network can be deduced. Developments in the application of AFC systems have made the collection of detailed passenger trip data in a URT network possible and can be used to obtain more in-depth understanding to passenger travel behaviors. In this paper, we aim to formulate the problem of estimating passenger train choices and subsequently propose an integrated approach for the addressed estimation combining real timetable and AFC data.

Advantages of the proposed approach include the following:

(1) A posteriori estimation framework, which uses revealed information combining real timetable and AFC data of URT systems rather than the a priori knowledge, was proposed.

(2) The approach links each AFC transaction (a passenger trip) to the most feasible train itinerary (a boarding plan). It is more appropriate for a factual travel choice process which uses only one route at the same time rather than multiroutes.

(3) The travel behavior parameters used in the approach are exacted from abundant timetable and AFC data rather than the manual surveys. Meanwhile, those exact pieces of information, which are difficult to be measured such as distributions of passengers' walking speeds and times, are also avoided to be obtained.

Furthermore, the proposed approach in this paper can be used for other challenges in the field of URT operation and management such as validation of rail transit assignment models, time-dependent train load estimation, and integrated simulation of passenger flows on network.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The study was financially supported by the National Natural Science Foundation of China (71271153), Scientific Research Foundation for Returned Scholars (Ministry of Education of China), Program for Young Excellent Talents in Tongji University (2014KJ015), and Fundamental Research Funds for the Central Universities of China (1600219249). The authors are also appreciative of Shanghai Shentong Metro Co., Ltd., for providing useful information.

References

[1] Y. Sheffi, *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*, Prentice-Hall, Inc, Englewood Cliffs, NJ, USA, 1985.

[2] M. G. H. Bell and Y. Iida, *Transportation Network Analysis*, John Wiley & Sons, Inc, West Sussex, UK, 1997.

[3] H. Kato, Y. Kaneko, and M. Inoue, "Comparative analysis of transit assignment: evidence from urban railway system in the Tokyo Metropolitan Area," *Transportation*, vol. 37, no. 7, pp. 775–799, 2010.

[4] W. Zhu and R. Xu, "Generating route choice sets with operation information on metro networks," *Journal of Traffic and Transportation Engineering (English Edition)*, vol. 3, no. 3, pp. 243–252, 2016.

[5] J. Attanucci and N. H. M. Wilson, *Bus Transit Monitoring Manual: Volume 1: Data Collection Program Design*, US Department of Transportation, 1981.

[6] J. Zhao, A. Rahbee, and N. H. M. Wilson, "Estimating a rail passenger trip origin-destination matrix using automatic data collection systems," *Computer-Aided Civil and Infrastructure Engineering*, vol. 22, no. 5, pp. 376–387, 2007.

[7] R. H. Xu, W. F. Zhu, Zhou., and J. G. Shi, *Research on the Clearing Model and Simulation System based on Passengers' Travel Times*

and Train Plans, School of Transportation Engineering, Tongji University, Shanghai, China, 2011.

[8] M. A. Munizaga and C. Palma, "Estimation of a disaggregate multimodal public transport Origin-Destination matrix from passive smartcard data from Santiago, Chile," *Transportation Research Part C: Emerging Technologies*, vol. 24, no. 12, pp. 9–18, 2012.

[9] M. Munizaga, F. Devillaine, C. Navarrete, and D. Silva, "Validating travel behavior estimated from smartcard data," *Transportation Research Part C: Emerging Technologies*, vol. 44, no. 6, pp. 70–79, 2014.

[10] C. Morency, M. Trépanier, and B. Agard, "Measuring transit use variability with smart-card data," *Transport Policy*, vol. 14, no. 3, pp. 193–203, 2007.

[11] C. Seaborn, J. Attanucci, and N. H. M. Wilson, "Analyzing multimodal public transport journeys in London with smart card fare payment data," *Transportation Research Record*, no. 2121, pp. 55–62, 2009.

[12] M. Bagchi and P. R. White, "The potential of public transport smart card data," *Transport Policy*, vol. 12, no. 5, pp. 464–474, 2005.

[13] M. Lehtonen, M. J. Rosenberg, Rasanen., and A. Sirkia, "Utilization of the smart card payment system (SCPS) data in public transport planning and statistics," in *Proceedings of the 9th World Congress on Intelligent Transport Systems*, Chicago, Ill, USA, 2002.

[14] M. Utsunomiya, J. Attanucci, and N. H. M. Wilson, "Potential uses of transit smart card registration and transaction data to improve transit planning," *Transportation Research Record: Journal of the Transportation Research Board*, pp. 119–126, 2006.

[15] Z. Guo and N. H. M. Wilson, "Transfer behavior and transfer planning in public transport systems: a case of the london underground," in *Proceedings of the 11th International Conference on Advanced Systems for Public Transport*, Hong Kong, China, 2009.

[16] J. Chan, *Rail Transit OD Matrix Estimation and Journey Time Reliability Metrics Using Automated Fare Data*, Massachusetts Institute of Technology, Cambridge, Mass, USA, 2007.

[17] T. Kusakabe, T. Iryo, and Y. Asakura, "Estimation method for railway passengers' train choice behavior with smart card transaction data," *Transportation*, vol. 37, no. 2, pp. 731–749, 2010.

[18] Y. Sun and R. Xu, "Rail transit travel time reliability and estimation of passenger route choice behavior," *Transportation Research Record*, no. 2275, pp. 58–67, 2012.

[19] Y. Sun and P. M. Schonfeld, "Schedule-based rail transit path-choice estimation using automatic fare collection data," *Journal of Transportation Engineering*, vol. 142, no. 10, Article ID 04015037, pp. 36–51, 2016.

[20] F. Zhou and R.-H. Xu, "Model of passenger flow assignmentfor Urban rail transit based on entryand exit time constraints," *Transportation Research Record*, no. 2284, pp. 57–61, 2012.

[21] Q. Fu, R. Liu, and S. Hess, "A Bayesian modelling framework for individual passengers probabilistic route choices: a case study on the london underground," in *Proceedings of the 93rd TRB Annual Meeting*, Washington, DC, USA, 2014.

[22] W. Zhu, H. Hu, and Z. Huang, "Calibrating rail transit assignment models with genetic algorithm and automated fare collection data," *Computer-Aided Civil and Infrastructure Engineering*, vol. 29, no. 7, pp. 518–530, 2014.

[23] L. Sun, Y. Lu, J. G. Jin, D.-H. Lee, and K. W. Axhausen, "An integrated Bayesian approach for passenger flow assignment in metro networks," *Transportation Research Part C: Emerging Technologies*, vol. 52, pp. 116–131, 2015.

[24] J. Swait and M. E. Ben-Akiva, *Analysis of the Effects of Captivity on Travel Time and Cost Elasticities. Behavioural Research for Transport Policy*, VNU Science Press, Utrecht, The Netherlands, 1985.

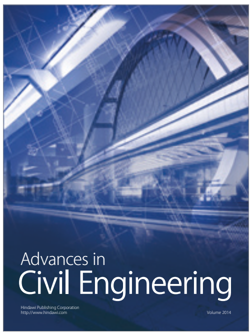
[25] J. Swait and M. Ben-Akiva, "Incorporating random constraints in discrete models of choice set generation," *Transportation Research Part B*, vol. 21, no. 2, pp. 91–102, 1987.

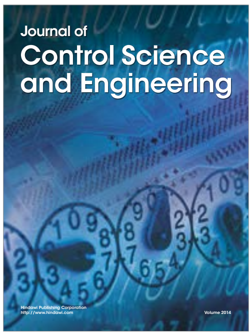
[26] H. C. W. L. Williams and J. D. Ortuzar, "Behavioural theories of dispersion and the mis-specification of travel demand models," *Transportation Research Part B*, vol. 16, no. 3, pp. 167–219, 1982.

[27] J. d. Ortúzar and L. G. Willumsen, *Modelling Transport*, John Wiley & Sons, Ltd, Chichester, UK, 2011.

[28] C. F. Manski, "The structure of random utility models," *Theory and Decision. An International Journal for Philosophy and Methodology of the Social Sciences*, vol. 8, no. 3, pp. 229–254, 1977.

[29] J. A. Azevedo, J. J. E. R. S. Madeira, E. Q. V. Martins, and F. M. A. Pires, "A Shortest Paths Ranking Algorithm," in *Proceedings of the Annual Conference AIRO'90, Models and Methods for Decision Support*, Operational Research Society of Italy, 1990.





Hindawi

Submit your manuscripts at
https://www.hindawi.com