

Research Article

Estimation Method of Path-Selecting Proportion for Urban Rail Transit Based on AFC Data

Feng Zhou,^{1,2,3} Jun-gang Shi,^{1,2} and Rui-hua Xu^{1,2,3}

¹The Key Laboratory of Road and Traffic Engineering, Tongji University, 1239 Siping Road, Shanghai 200092, China

²School of Transportation Engineering, Tongji University, 1239 Siping Road, Shanghai 200092, China

³Intelligent Transportation Institute for Advanced Study, Tongji University, 1239 Siping Road, Shanghai 200092, China

Correspondence should be addressed to Feng Zhou; zhoufeng24@tongji.edu.cn

Received 4 January 2015; Accepted 4 March 2015

Academic Editor: Wei (David) Fan

Copyright © 2015 Feng Zhou 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.

With the successful application of automatic fare collection (AFC) system in urban rail transit (URT), the information of passengers' travel time is recorded, which provides the possibility to analyze passengers' path-selecting by AFC data. In this paper, the distribution characteristics of the components of travel time were analyzed, and an estimation method of path-selecting proportion was proposed. This method made use of single path ODs' travel time data from AFC system to estimate the distribution parameters of the components of travel time, mainly including entry walking time (ewt), exit walking time (exwt), and transfer walking time (tw). Then, for multipath ODs, the distribution of each path's travel time could be calculated under the condition of its components' distributions known. After that, each path's path-selecting proportion can be estimated. Finally, simulation experiments were designed to verify the estimation method, and the results show that the error rate is less than 2%. Compared with the traditional models of flow assignment, the estimation method can reduce the cost of artificial survey significantly and provide a new way to calculate the path-selecting proportion for URT.

1. Introduction

As the basis of network flow assignment calculation, the path-selecting proportion is directly related to the operation and management of urban rail transit (URT), including operation indicators calculating, train plan making, and fare clearing. Currently, there have been many research results on flow assignment for URT under the condition of network operation, most of which are multipath models based on path utility.

Nguyen et al. [1] developed a graph-theoretic framework for the passenger assignment problem, which encompassed simultaneously the departure time and the route choice dimensions. Also, a passenger equilibrium flow model was defined and a mathematical formulation was suggested. The research results can be used to solve the passenger flow distribution problem of URT network.

Poon et al. [2] also studied the dynamic traffic assignment model for congested networks and used time-increment simulation to calculate the passenger flow on the network.

The model assumed that the vehicle was running in full accordance with the schedule and the passengers had full predictive information about the present and future network conditions, so the model can be used to simulate the performance of an existing transit system operating with preannounced schedules or to evaluate the effects of changes in schedules, lines, or passenger demand on system performance.

Xu et al. [3] established a multipath assignment model of URT network, using travel time as paths' impedance. The basic characteristics of rail network and passenger travel behavior are fully considered in the model. With its high accuracy and strong practicality, the model has been successfully applied in Beijing Subway network. The survey results show that its error rate of section flow is less than 5%.

Si et al. [4] constructed the generalized cost function of URT network, considering the major factors (including the travel time and times of transfer) influencing the passenger flow assignment, and then proposed a mathematical optimization model of passenger flow assignment based

on the stochastic user equilibrium principle. The results of numerical example using Beijing URT network data showed that the model was feasible and effective.

The basic idea of such models above can be summarized as follows:

- (1) determine the impedance of each path, which can be time costs or mileage;
- (2) design the utility function based on the path impedance;
- (3) calculate the flow assignment proportion of each path;
- (4) distribute each OD's total passenger flow of one day to paths between the origin stations and the destination stations (OD).

Such models can basically guarantee the accuracy of flow assignment. However, the parameters of these models, including the entry and exit walking time at stations and transfer time, are calibrated by manual survey, which is really a time-consuming and costly work. In addition, when the network structure or operation organization changes, the parameters need to be recalibrated in order to keep accurate. Therefore, it is necessary to study a new method of network flow assignment.

In recent years, the AFC system is widely used in URT, which can accurately record the passengers' entry time at the origin stations and the exit time at the destination stations. With many years of application, the AFC system has accumulated vast amounts of passengers' travel information. However, such information has rarely been used to study the behavior of travelling and path-selecting.

Chapleau et al. [5] and Rahbee [6] used the AFC data to predict OD matrices and network flow distribution, but only the entry records were used, which could not support the analysis of passengers' path-selecting behavior.

Lee and Hickman [7] analyzed large amount of AFC data and found that activity and travel patterns differ significantly across the different farecard types, such as travel period, travel time, travel region (urban areas to the suburbs, urban areas to urban areas, etc.). The research results can be used to grasp the characteristics of different travel cardholders.

Kusakabe et al. [8] developed a methodology and an algorithm for estimating which train is boarded by each smart card holder based on long-term transaction data. The proposed method made a number of assumptions and distributed the uncertain smart card data to each possible train evenly, which needed further research.

Zhou and Xu [9] considered the process of the passenger travelling with the train running together and set up a model of calculating each passenger's path by AFC data. The proposed method could estimate each passenger's travel route and train choice. But the model needed a lot of manual investigation work to be also further verified by mathematical statistics.

Sun and Xu [10] analyzed the travel time reliability and estimated passenger route choice behavior by AFC data. The proposed model could be used to calibrate the parameters of traditional passenger flow distribution models. But the model needed all the stations' walking time obtained by artificial

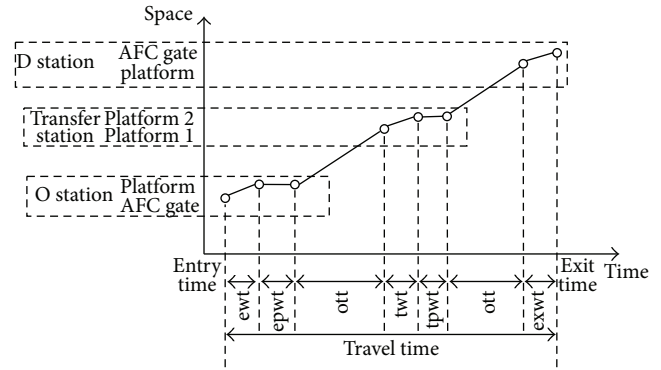


FIGURE 1: Components of travel time.

survey and lacked further excavation of AFC data, such as ticket type and travel period.

Based on the above background, this paper works on analyzing passengers' travel time data recorded by AFC system with the theory of mathematical statistics and proposes a new estimation method of path-selecting proportion. This method provides a new idea of network flow assignment for URT.

2. Analysis of Travel Time

Travel time of passengers in URT network mainly consists of the following six parts:

- (1) entry walking time (the time of passengers walking from the AFC gate to the platform in the origin station, denoted by ewt);
- (2) entry platform waiting time (the time of passengers waiting for the train on the platform in the origin station, denoted by $epwt$);
- (3) on-train time (the time of passengers travelling on the train, denoted by ott);
- (4) transfer walking time (the time of passengers walking from the arrival platform to the departure platform in the transfer station, denoted by twt);
- (5) transfer platform waiting time (the time of passengers waiting for the train on the platform in transfer station, denoted by $tpwt$);
- (6) exit walking time (the time of passengers walking from the platform to the AFC gate in the destination station, denoted by $exwt$).

twt and $tpwt$ only occur in the transfer station. The components of travel time are shown in Figure 1.

Lots of random factors can affect passengers' travel time, so this paper makes the following assumptions:

- (1) passengers arrive at the station randomly, dispersedly, and stably;
- (2) all passengers can board the first train after arriving at the platform;

(3) the trains run according to the plan strictly with a certain speed level and no abnormal condition occurs.

2.1. Walking Time (ewt, twt, and exwt). For one station, suppose the walk distances for different passengers are the same, and then the distribution of walking time only depends on the walking speed. For walking speed, most research results [11, 12] show that it can be considered to follow normal distribution. Thus, walking time follows a normal distribution as well; that is,

$$\begin{aligned} \text{ewt} &\sim N(\mu_{\text{ewt}}, \sigma_{\text{ewt}}^2), \\ \text{twt} &\sim N(\mu_{\text{twt}}, \sigma_{\text{twt}}^2), \\ \text{exwt} &\sim N(\mu_{\text{exwt}}, \sigma_{\text{exwt}}^2), \end{aligned} \quad (1)$$

where μ_{ewt} , μ_{twt} , and μ_{exwt} are the mean of ewt, twt, exwt and σ_{ewt}^2 , σ_{twt}^2 , σ_{exwt}^2 are the variance of ewt, twt, and exwt.

2.2. Entry Platform Waiting Time (epwt). epwt is the random variables between 0 and the interval of the trains. Based on the assumption of passengers arriving at the origin station randomly and dispersedly, they arrive at the platform randomly and dispersedly too. Thus, epwt can be considered to follow a uniform distribution; that is,

$$\text{epwt} \sim U[0, H], \quad (2)$$

where H is the interval of the trains.

The distribution can be verified by the following simulation experiment. The experimental environment is as follows: the interval of the trains at Station A is 180 s, between 8:00:00 and 10:00:00; $\text{ewt} \sim N(145 \text{ s}, 20^2)$.

The experimental procedure is as follows.

(1) Generate randomly and evenly 20,000 passengers between 8:00:00 and 9:50:00; for passenger i , let T_{Ei} be the arriving time of the station and let ewt_i be the entry walking time and then the time of arriving at the platform is $T_{Ei} + \text{ewt}_i$.

(2) Search for the first train arriving at Station A after $T_{Ei} + \text{ewt}_i$. Let T_{Di} be the departure time of the first train at Station A, and then $\text{epwt}_i = T_{Di} - (T_{Ei} + \text{ewt}_i)$.

MATLAB is used to do the simulation experiment. Assuming $\text{epwt} \sim U[0, 180]$, use SPSS to do Kolmogorov-Smirnov test (KS test for short) of the waiting time data from simulation, and the significance level α is generally taken to be 0.05. Frequency statistics is shown in Figure 2. The result shows that the significance is 0.626 ($> \alpha$). So, the assumption $\text{epwt} \sim U[0, 180]$ is tenable.

2.3. Transfer Platform Waiting Time (tpwt). Same as epwt, tpwt is the random variables between 0 and the interval of the trains. However, their distributions are different, that is because passengers arrive randomly and dispersedly at the origin station but intensively at the transfer station.

Generally speaking, when the interval of transfer line is big, the faster passengers walk and the longer passengers may wait; when the interval of transfer line is small, the faster

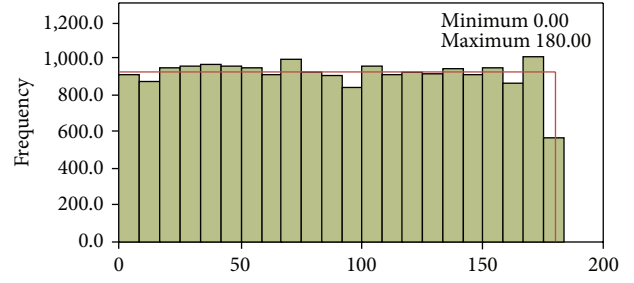


FIGURE 2: KS test.

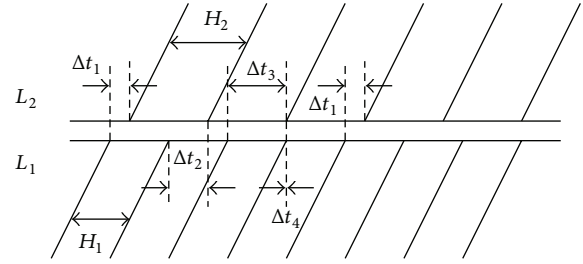


FIGURE 3: Factors affecting tpwt.

passengers walk and the shorter passengers may wait. It can be seen that there is strong correlation between tpwt and twt. Thus, this paper analyzes the distribution of $(\text{twt} + \text{tpwt})$, which is denoted by $\text{twt} + \text{tpwt}$ hereinafter.

The factors affecting distribution of $\text{twt} + \text{tpwt}$ are as follows: the interval of the before-transfer line (H_1); the interval of the after-transfer line (H_2); the coordination time between the before-transfer line L_1 and the after-transfer line L_2 (Δt), as shown in Figure 3; the distribution of twt.

Obviously, Δt is related to H_1 and H_2 , and its value has certain periodicity. To facilitate the presentation, let Δt_1 be the first coordination time, and it can be calculated by the following formula:

$$\Delta t_1 = \begin{cases} \Delta t_{\min}, & H_1 \leq H_2, \\ \Delta t_{\max}, & H_1 > H_2, \end{cases} \quad (3)$$

where Δt_{\min} is the minimum coordination time and Δt_{\max} is the maximum one.

Let $[H_1, H_2]$ be the least common multiple of H_1 , H_2 . Then, the calculation formula of Δt_i is designed as follows:

$$\Delta t_i = \Delta t_1 + (i - 1)(H_2 - H_1) \quad i = 1, 2, \dots, \frac{[H_1, H_2]}{H_1}. \quad (4)$$

Suppose that the passengers arrive at the transfer station with the coordination time Δt_i . Then, if the transfer walking time of one passenger is smaller than Δt_i , his/her $\text{twt} + \text{tpwt}$ is Δt_i ; otherwise, the passenger will wait for one or some more intervals of Line 2 and his/her $\text{twt} + \text{tpwt}$ will be $\Delta t_i + n \cdot H_2$.

TABLE 1: Structure of AFC data.

Records	O station code	Entry time	D station code	Exit time	Ticket type
Record 1	000102	8:05:12	000425	8:45:43	Single card
...

TABLE 2: Distribution characteristics and parameters of the components of travel time.

Components of travel time	Probability density function	Parameters to be estimated	Mean	Variance
ewt	$\text{ewt} \sim N(\mu_{\text{ewt}}, \sigma_{\text{ewt}}^2)$	$\mu_{\text{ewt}}, \sigma_{\text{ewt}}^2$	μ_{ewt}	σ_{ewt}^2
epwt	$\text{epwt} \sim U[0, H]$	—	$H/2$	$H^2/12$
ott	—	—	TT	0
twt&tpwt	As formula (5)	$\mu_{\text{twt}}, \sigma_{\text{twt}}^2$	$\mu_{\text{twt\&tpwt}}$	$\sigma_{\text{twt\&tpwt}}^2$
exwt	$\text{exwt} \sim N(\mu_{\text{exwt}}, \sigma_{\text{exwt}}^2)$	$\mu_{\text{exwt}}, \sigma_{\text{exwt}}^2$	μ_{exwt}	σ_{exwt}^2

As a result of $\text{twt} \sim N(\mu_{\text{twt}}, \sigma_{\text{twt}}^2)$, the probability distribution of twt\&tpwt can be obtained by formula (5) as follows:

$$P\{\text{twt\&tpwt} = k\} = \begin{cases} \int_0^{\Delta t_i} \frac{1}{\sqrt{2\pi}\sigma_{\text{twt}}} e^{-(x-\mu_{\text{twt}})^2/2(\sigma_{\text{twt}})^2} dx & k = \Delta t_i \\ \int_{\Delta t_i+(n-1)H_2}^{\Delta t_i+nH_2} \frac{1}{\sqrt{2\pi}\sigma_{\text{twt}}} e^{-(x-\mu_{\text{twt}})^2/2(\sigma_{\text{twt}})^2} dx & k = \Delta t_i + n \cdot H_2, n = 1, 2, \dots \end{cases} \quad (5)$$

In the above formula, n can be infinite theoretically, but, in fact, the probability is almost 0 when $n \geq 5$.

2.4. On-Train Time (ott). Based on the assumption, all the trains run according to the timetable strictly with a certain speed level. So, between two certain stations (Station i , Station j) on the same line, the total running time of different trains is a constant, which is denoted by TT. Obviously, the on-train time of passengers traveling between Station i and Station j is equal to TT and is a constant as well.

2.5. Analysis of Independence. Whether the components of travel time are independent is very important to analyze distribution characteristics of the path's travel time. According to the analysis in the previous section, ott is a constant, so ott is independent of other components. ott divides travel time into three independent parts: ewt and epwt, twt&tpwt, and exwt. Thus, only the independence of ewt and epwt needs to be analyzed. In fact, one passenger's waiting time is not related to his/her walking speed. As the passengers arrive at the station randomly and dispersedly, they arrive at the platform randomly and dispersedly too, which makes their waiting time random variable. That is the reason why the passengers walking fast may wait longer at the origin station. Therefore, ewt and epwt can be considered independent.

Based on the above analysis, all the components of travel time are independent.

3. Model and Algorithm

3.1. AFC Data. In China, smart cards and AFC system are applied in most cities, which can record part of the passengers' traveling information on the URT network. The basic structure of AFC data is shown in Table 1.

Thus, based on the AFC data structure, any passenger's travel time can be calculated.

3.2. Distribution Parameter Estimation for Components of Travel Time. According to the analysis in Section 2, the distribution characteristics and parameters of the components of travel time are shown in Table 2.

From Table 2, it can be seen that the parameters to be estimated are the mean and variance of ewt of each station ($\mu_{\text{ewt}}, \sigma_{\text{ewt}}^2$); the mean and variance of twt in each transfer direction of each transfer station ($\mu_{\text{twt}}, \sigma_{\text{twt}}^2$); the mean and variance of exwt of each station ($\mu_{\text{exwt}}, \sigma_{\text{exwt}}^2$).

3.2.1. Estimation Method of $\mu_{\text{ewt}}, \sigma_{\text{ewt}}^2$ and $\mu_{\text{exwt}}, \sigma_{\text{exwt}}^2$. Take the OD with single path and no transfer (the origin station and the destination station are on the same line) as the research object, and the travel time is only comprised of ewt, epwt, ott, and exwt. Let m be the origin station and let n be the destination station. Then, large samples of passengers' travel time data ($T^{mn} = \{T_1^{mn}, T_2^{mn}, \dots, T_i^{mn}, \dots, T_S^{mn}\}$) can be obtained by AFC system, where T_i^{mn} is the actual travel time of passengers i and S is the sample size.

Based on T^{mn} , the mean and variance of the path's travel time ($E(t^{mn}), D(t^{mn})$) can be estimated by moment estimation. Moment estimation is a commonly used method of parameter estimation, proposed by K. Pearson in 1894 [13]. According to Wiener-Khinchin law of large numbers, the sample moment converge to the population moment when the sample size is large. The principle of moment estimation is as follows: (1) estimate the corresponding population moment by the sample moment and (2) estimate the parameters by making use of the relationship between the unknown parameters and the population moment.

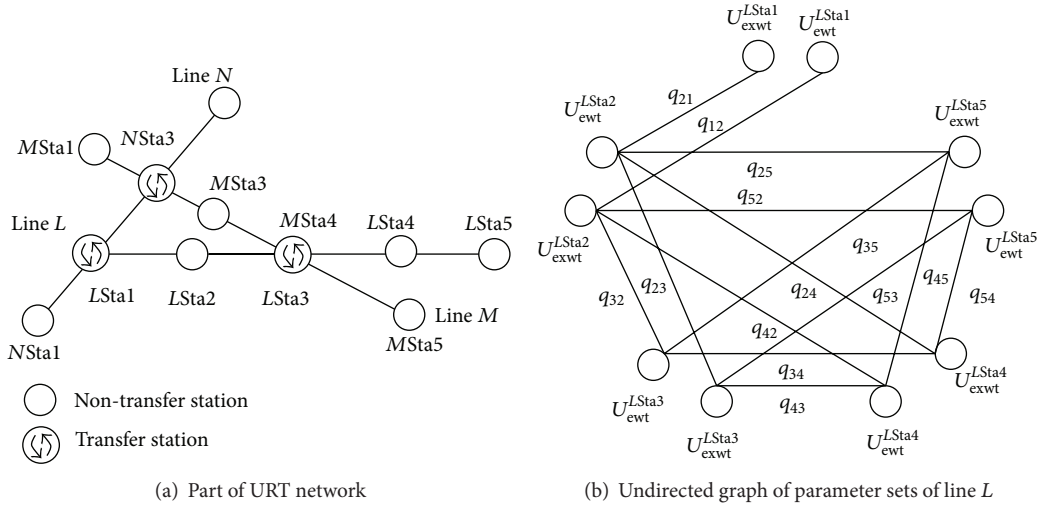


FIGURE 4: Relationship graph of parameter sets.

Because the components of travel time are independent, the following equations are established:

$$\begin{aligned}
 E(t^{mn}) &= E(ewt^m) + E(epwt^m) + E(ott^m) + E(exwt^n) \\
 &= \mu_{ewt}^m + \frac{H^m}{2} + TT^{mn} + \mu_{exwt}^n, \\
 D(t^{mn}) &= D(ewt^m) + D(epwt^m) + D(ott^m) + D(exwt^n) \\
 &= (\sigma_{ewt}^m)^2 + \frac{(H^m)^2}{12} + (\sigma_{exwt}^n)^2,
 \end{aligned} \tag{6}$$

where μ_{ewt}^m is the mean of ewt of station m ; $(\sigma_{ewt}^m)^2$ is the variance of ewt of station m ; H^m is the interval of station m ; TT^{mn} is the train running time between station m and station n ; μ_{exwt}^n is the mean of exwt of station n ; H^m and TT^{mn} can be obtained from the timetable directly.

Equation (6) are suitable for any two different stations on the same line and can be converted to the following:

$$\begin{aligned}
 \mu_{ewt}^m + \mu_{exwt}^n &= E(t^{mn}) - \frac{H^m}{2} - TT^{mn} \\
 \forall m, n \in U_S^L, \quad m \neq n, \\
 (\sigma_{ewt}^m)^2 + (\sigma_{exwt}^n)^2 &= D(t^{mn}) - \frac{(H^m)^2}{12} \\
 \forall m, n \in U_S^L, \quad m \neq n,
 \end{aligned} \tag{7}$$

where U_S^L is the station set of line L .

Therefore, for any station on the line L , if its distribution parameters (mean and variance) of ewt or exwt are known (by survey), the distribution parameters of all the stations on line L can be calculated by (7). According to the theory of moment estimation, the larger the sample size is, the more accurate the parameters are estimated. Thus, in order to improve the

accuracy of parameter estimation, the sequence of stations on line L for parameters estimation (called PES problem hereinafter) can be made according to the passenger flow.

To describe and solve PES problem, define U_P^{LS} as the set of distribution parameters of all stations on line L , and then

$$U_P^{LS} = \{U_{ewt}^{LSta1}, U_{exwt}^{LSta1}, \dots, U_{ewt}^{LStai}, U_{exwt}^{LStai}, \dots, U_{ewt}^{LStak}, U_{exwt}^{LStak}\}, \tag{8}$$

where U_{ewt}^{LStai} is the set of ewt's distribution parameters of station i on line L and $U_{exwt}^{LStai} = \{\mu_{exwt}^{LStai}, (\sigma_{exwt}^{LStai})^2\}$; U_{ewt}^{LStai} is the set of ewt's distribution parameters of station i on line L and $U_{exwt}^{LStai} = \{\mu_{exwt}^{LStai}, (\sigma_{exwt}^{LStai})^2\}$.

Take the five stations of line L of the network in Figure 4(a); for example, the relationship between the sets of each station's distribution parameters can be established as the undirected graph in Figure 4(b). In the figure, two sets are connected by an edge if there is only one path between the two stations; otherwise, two sets are unconnected. q_{ij} is the passenger flow between station i and station j .

In order to estimate the parameters accurately, the sample size should be as large as possible. Therefore, the PES problem can be summarized to seek the maximum spanning tree in Figure 4(b), and the model is described as follows:

$$\max q(t) = \max \left(\sum_{(u,v) \in t} q(u,v) \right), \tag{9}$$

where t is the spanning tree of the undirected graph and $q(u,v)$ is the weight value of the edge between vertex u and v .

The most common algorithm used to calculate the optimal spanning tree is Kruskal algorithm [14]. Let V be the set of all vertices, let E be the set of all edges, let s be the number of all edges, let $T = (V, TE)$ be the spanning tree, let TE be the set of edges of the spanning tree, and let x be the number of edges of the spanning tree. Then, the steps of Kruskal algorithm for solving the maximum spanning tree are as follows.

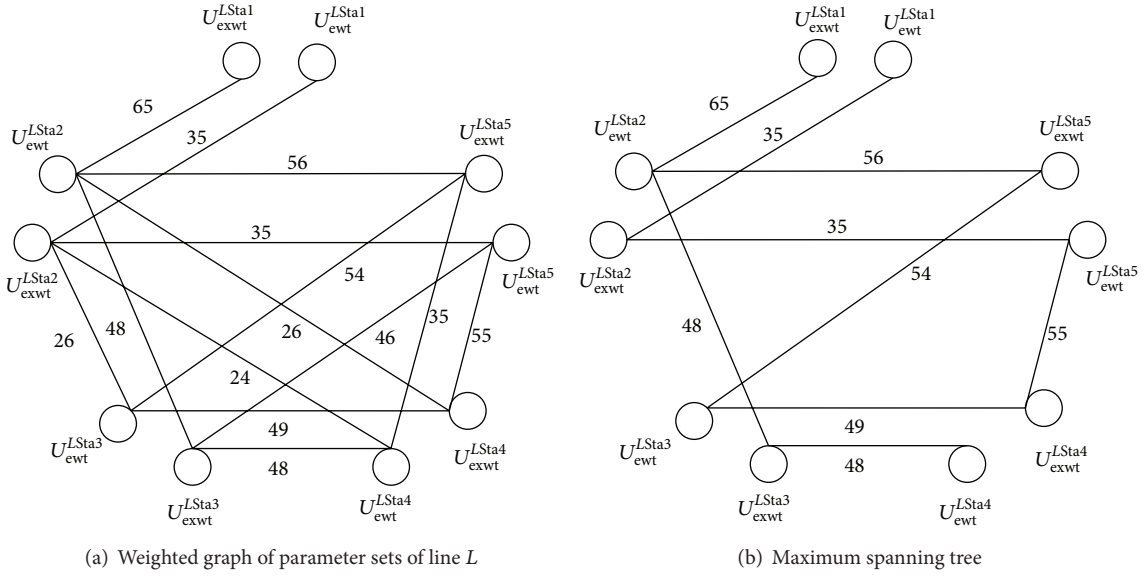


FIGURE 5: Results of maximum spanning tree.

Step 1. Set $TE = \Phi$, $x = 0$.

Step 2. Choose the edge e^* with the maximum value in the graph, $e^* = \{\max q(u, v) \mid u, v \in V\}$.

Step 3. If there is any loop in the graph $T = (V, TE \cup e^*)$, then $E = E \setminus e^*$ and then back to Step 2; otherwise, $TE = TE \cup e^*$, $E = E \setminus e^*$, $x = x + 1$.

Step 4. If $x = s - 1$, the algorithm ends; otherwise, repeat Step 2.

For example, suppose the passenger flows of ODs in Figure 4(b) are shown as in Figure 5(a) and the maximum spanning tree calculated by Kruskal algorithm is shown in Figure 5(b).

According to the maximum spanning tree, the node $U_{\text{ewt}}^{L\text{Sta2}}$ converges the maximum number of edges, of which the parameters are designed, to be estimated by artificial survey. Then, the parameters of other nodes will be obtained in succession by formula (7). Therefore, the estimation sequence of parameters can be made as shown in Figure 6.

Based on the above analysis, the estimation method of μ_{ewt} , σ_{ewt}^2 , and μ_{exwt} , σ_{exwt}^2 of each station on one line is summarized as follows.

Step 1. Build the relationship graph of parameter sets, of which the value of each edge is the passenger flow between its two vertices collected by AFC system.

Step 2. Use Kruskal algorithm to calculate the maximum spanning tree of the relationship graph.

Step 3. Find the node with the maximum number of edges and use the artificial survey to collect its samples of walking time (ewt or exwt); then, its parameters of $(\mu_{\text{ewt}}, \sigma_{\text{ewt}}^2)$ or $(\mu_{\text{exwt}}, \sigma_{\text{exwt}}^2)$ can be estimated by moment estimation.

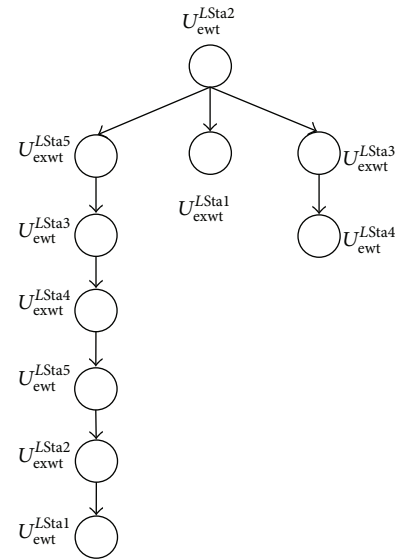


FIGURE 6: Sequence of parameters estimation.

Step 4. According to the maximum spanning tree, estimate the parameters $(\mu_{\text{ewt}}, \sigma_{\text{ewt}}^2, \mu_{\text{exwt}}$, and $\sigma_{\text{exwt}}^2)$ of other stations in succession by formula (7).

3.2.2. Estimation Method of $\mu_{\text{twt}}, \sigma_{\text{twt}}^2$. In URT system, transfer path means the path from the platform of one line to the platform of another line at the transfer station. Therefore, for a two-line transfer station, there are four transfer paths in total. Based on the analysis of most URT networks in China, for any transfer path of one transfer station, certain OD with single path and one transfer can be always found to contain the transfer path, of which the travel time only includes ewt, epwt, twt&tpwt, ott, and exwt.

Suppose that a certain OD (mn) with single path and one transfer contains transfer path i , as the components of travel time are independent. Formulas (10) and (11) are established as follows:

$$E(t^{mn}) = \mu_{\text{ewt}}^m + \frac{H^m}{2} + \text{TT}^{mi} + \mu_{\text{twt\&tpwt}}^i + \text{TT}^{in} + \mu_{\text{exwt}}^n, \quad (10)$$

$$D(t^{mn}) = (\sigma_{\text{ewt}}^m)^2 + \frac{(H^m)^2}{12} + (\sigma_{\text{twt\&tpwt}}^i)^2 + (\sigma_{\text{exwt}}^n)^2, \quad (11)$$

where m, n are the origin station and the destination station; $E(t^{mn}), D(t^{mn})$ are the mean and the variance of mn 's travel time; $\mu_{\text{twt\&tpwt}}^i, (\sigma_{\text{twt\&tpwt}}^i)^2$ are the mean and the variance of twt\&tpwt of transfer path i ; TT^{mi} is the train running time from station m to the transfer station; TT^{in} is the train running time from the transfer station to station n .

$E(t^{mn}), D(t^{mn})$ can be estimated by moment estimation of AFC data; H^m, TT^{mi} , and TT^{in} can be obtained from the timetable; $\mu_{\text{ewt}}^m, \sigma_{\text{ewt}}^m, \mu_{\text{exwt}}^n$, and σ_{exwt}^n can be estimated from Section 3.2.1. Therefore, only $\mu_{\text{twt\&tpwt}}^i, (\sigma_{\text{twt\&tpwt}}^i)^2$ are unknown in formulas (9) and (10), which can be calculated easily.

Under the conditions of $\mu_{\text{twt\&tpwt}}^i, (\sigma_{\text{twt\&tpwt}}^i)^2$ known, $\mu_{\text{twt}}, \sigma_{\text{twt}}^2$ can be calculated by formulas (5).

3.3. Estimation Method of Path-Selecting Proportion. Let mn be a multipath OD and let t^{mn} be the variable of travel time. Then, the probability density function of mn 's travel time $f(t^{mn})$ can be described as follows:

$$f(t^{mn}) = \sum_{i=1}^k w_i f_i(t^{mn}), \quad (12)$$

$$\sum_{i=1}^k w_i = 1, \quad \forall m, n \in U_s^N,$$

where w_i is the path-selecting proportion of the i th path between station m and station n ; $f_i(t^{mn})$ is the probability density function of the i th path's travel time; k is the number of paths between station m and station n ; U_s^N is the set of stations on the URT network.

Based on (12) and (13), the following can be deduced as well:

$$E(t^{mn}) = \mu_{mn} = \sum_{i=1}^k w_i \mu_{mni},$$

$$D(t^{mn}) = (\sigma_{mn})^2 = \sum_{i=1}^k (w_i (\sigma_{mni})^2 + (\mu_{mni})^2) - \left(\sum_{i=1}^k w_i \mu_{mni} \right)^2, \quad (13)$$

where μ_{mn} is the mean of mn 's travel time; $(\sigma_{mn})^2$ is the variance of mn 's travel time; μ_{mni} is the mean of the i th path's travel time; $(\sigma_{mni})^2$ is the variance of the i th path's travel time.

TABLE 3: Detailed information of paths: "NSta1 \rightarrow MSta5"

Paths	Stations	Transfer stations
Path 1	NSta1, LSta1, NSta3, MSta3, LSta3, MSta5	NSta3
Path 2	NSta1, LSta1, LSta2, LSta3, MSta5	LSta1, LSta3

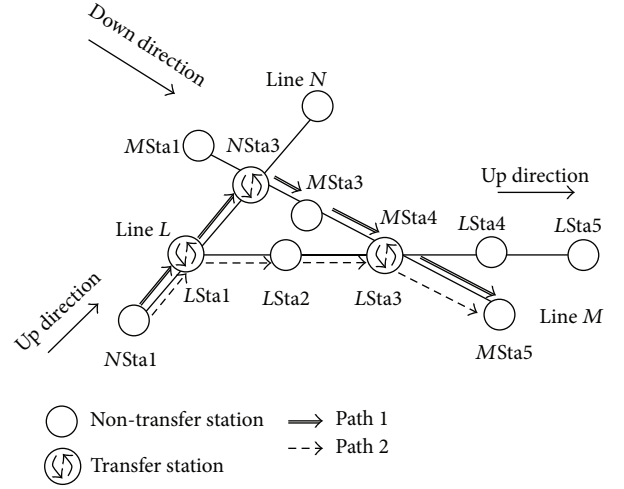


FIGURE 7: Part of URT network.

μ_{mn} and $(\sigma_{mn})^2$ can be estimated from the AFC data; while μ_{mni} and $(\sigma_{mni})^2$ can be estimated by the following equations, in which each parameter can be obtained by the methods described in Section 3.2:

$$\mu_{mni} = \mu_{\text{ewt}}^m + \frac{H^m}{2} + \sum_{\text{tr} \in U} \mu_{\text{tr}}^{\text{twt\&tpwt}} + \mu_{\text{exwt}}^n + \text{TT}^{mn},$$

$$(\sigma_{mni})^2 = (\sigma_{\text{ewt}}^m)^2 + \frac{(H^m)^2}{12} + \sum_{\text{tr} \in U} (\sigma_{\text{tr}}^{\text{twt\&tpwt}})^2 + (\sigma_{\text{exwt}}^n)^2, \quad (14)$$

where $\mu_{\text{tr}}^{\text{twt\&tpwt}}$ is the mean of transfer station tr 's twt\&tpwt ; $(\sigma_{\text{tr}}^{\text{twt\&tpwt}})^2$ is the variance of transfer station tr 's twt\&tpwt ; U is the set of transfer stations in the i th path.

Combining (13)~(14), each path's path-selecting proportion can be calculated. It is worth noting that only k -order ($k \leq 2$) central moment statistics (mean and variance) are used in the derivation above. So, the estimation method can only apply to the ODs with no more than three paths. When there are more than three paths, the idea in this paper is applicable as well, but k -order ($k > 2$) central moment statistics should be introduced.

4. Numerical Example

The OD "NSta1 \rightarrow MSta5" in Figure 4(a) is used as an example to verify the model and algorithm. Obviously, there are two paths between the OD, as shown in Figure 7 and Table 3.

Based on the method proposed in this paper, the process for estimating the proportion of two paths is as follows.

TABLE 4: Results of parameter estimation.

Parameters	LSta2			LSta4		
	Actual value	Estimation value	Error rate	Actual value	Estimation value	Error rate
Mean	145	144.91	-0.06%	165	165.21	0.13%
Standard deviation	20	19.93	-0.35%	20	19.81	-0.95%

TABLE 5: Values of parameters.

Parameters	Mean (s)	Standard deviation
ewt of NSta1	125	20 ²
exwt of MSta5	145	20 ²
twt of NSta3 (transfer path: from the platform of line N's up direction to the one of line M's down direction)	205	20 ²
twt of LSta1 (transfer path: from the platform of line N's up direction to the one of line L's up direction)	170	20 ²
twt of LSta3 (transfer path: from the platform of line L's up direction to the one of line M's down direction)	192	20 ²

TABLE 6: Estimation results of path-selecting proportion.

	Path-selecting proportion				Error rate of path 1
	Actual value		Estimation value		
	Path 1	Path 2	Path 1	Path 2	
Simulation 1	0.7	0.3	0.6884	0.3116	-1.66%
Simulation 2	0.5	0.5	0.5050	0.4950	0.10%
Simulation 3	0.3	0.7	0.2954	0.7046	-1.53%

Step 1 (estimate the distribution parameters of each station's walking time μ_{ewt} , σ_{ewt}^2 and μ_{exwt} , σ_{exwt}^2). Build the relationship graph of parameter sets and use Kruskal algorithm to calculate the maximum spanning tree of the graph. Based on the maximum spanning tree, the parameter estimation sequence of stations is obtained.

Take line L and its five stations; for example, as shown in Figure 6, $U_{\text{ewt}}^{L\text{Sta}2}$ of LSta2 should be estimated by artificial survey, and then other parameters can be estimated by formula (7).

To verify the parameter estimation method, a simulation of station LSta2 and LSta4 is designed as follows: generate randomly 10,000 passengers from LSta2 to LSta4 with the entry time between 8:00:00 and 9:50:00; $\text{ewt} \sim N(145 \text{ s}, 20^2)$; the interval of line L is 150 s; train running time is 320 s; $\text{exwt} \sim N(165 \text{ s}, 20^2)$.

MATLAB is used to do the simulation experiment. The entry walking time data are collected to be the ewt data of artificial survey, and the entry and exit time data of simulation are collected to be AFC data. The parameters of LSta2 are estimated by moment estimation method, while the ones of LSta4 are calculated by formula (7). The results are shown in Table 4.

From Table 4, we can see that the error rate of parameter estimation is less than 0.2%, and the estimation method is verified.

Step 2 (estimate μ_{twt} , σ_{twt}^2 of each transfer path of all transfer stations concerned). With the walking time parameters of all the stations on the URT network known, calculate $\mu_{\text{twt\&tpwt}}$,

$\sigma_{\text{twt\&tpwt}}^2$ of each transfer path of transfer stations using the travel time data of AFC system by (10), (11). Then, μ_{twt} , σ_{twt}^2 can be estimated by (5). The parameter estimation method in this step can also be verified by similar simulation experiments in Step1.

Step 3 (estimate each path's path-selecting proportion). Suppose the values of parameters estimated in Step 1 and Step 2 are partly shown in Table 5.

Design the simulation experiment as follows: generate randomly 10,000 passengers from NSta1 to MSta5 with the entry time between 8:00:00 and 9:50:00; the interval of each line is 180 s; the train running time of Path 1 is 21 min, while 18 min of Path 2.

Do the simulation three times with different path-selecting proportions between Path 1 and Path 2 (0.7:0.3, 0.5:0.5, 0.3:0.7). Also, the entry and exit time data of simulation are collected to be AFC data.

Then, each path's path-selecting proportion can be estimated by formula (13)~(14). The results are shown in Table 6.

The results in Table 6 shows that the error rate of path's path-selecting proportion estimation is less than 2%, which verifies the model and algorithm in this paper.

5. Conclusion

In the network operation phase of URT, path-selecting proportion is the key to network flow assignment and fare clearing. This paper analyzed the distribution characteristics of the components of travel time and then proposed an

estimation method of path-selecting proportion, making use of the travel time data from the AFC system. Also, simulation experiments were created to verify the estimation method, and the results show that the error rate is less than 2% and the method is reliable.

Compared with the traditional models based on path utility, the estimation method of path-selecting proportion has the following advantages: (1) by making full use of the AFC data, the sample size for parameter estimation is large and the results have the good feature of accuracy and (2) the estimation method relies on data analysis and processing, reducing the cost of artificial survey significantly.

In fact, the estimation method in this paper is being used to analyze and validate the URT network flow assignment results in Shanghai.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This research is supported by China Postdoctoral Science Foundation (Project no. 2014M551454). The authors also wish to acknowledge Shanghai Shentong Metro Group Co., Ltd., for providing basic data during the research.

References

- [1] S. Nguyen, S. Pallottino, and F. Malucelli, "A modeling framework for passenger assignment on a transport network with timetables," *Transportation Science*, vol. 35, no. 3, pp. 238–249, 2001.
- [2] M. H. Poon, S. C. Wong, and C. O. Tong, "A dynamic schedule-based model for congested transit networks," *Transportation Research Part B: Methodological*, vol. 38, no. 4, pp. 343–368, 2004.
- [3] R.-H. Xu, Q. Luo, and P. Gao, "Passenger flow distribution model and algorithm for urban rail transit network based on multi-route choice," *Journal of the China Railway Society*, vol. 31, no. 2, pp. 110–114, 2009.
- [4] B.-F. Si, B.-H. Mao, and Z.-L. Liu, "Passenger flow assignment model and algorithm for urban railway traffic network under the condition of seamless transfer," *Journal of the China Railway Society*, vol. 29, no. 6, pp. 12–18, 2007.
- [5] R. Chapleau, M. Trepanier, and K. K. Chu, "The ultimate survey for transit planning: complete information with smart card data and GIS," in *Proceedings of the 8th International Conference on Survey Methods in Transport: Harmonisation and Data Comparability*, Annecy, France, May 2008.
- [6] A. B. Rahbee, "Farecard passenger flow model at Chicago transit authority, Illinois," *Transportation Research Record*, vol. 2072, pp. 3–9, 2008.
- [7] S. G. Lee and M. D. Hickman, "Travel pattern analysis using smart card data of regular users," in *Proceedings of the 90th Annual Meeting of the Transportation Research Board*, Washington, DC, USA, 2011.
- [8] T. Kusakabe, T. Iryo, and Y. Asakura, "Estimation method for railway passengers' train choice behavior with smart card transaction data," *Transportation*, vol. 37, no. 5, pp. 731–749, 2010.
- [9] F. Zhou and R.-H. Xu, "Model of passenger flow assignment for Urban rail transit based on entry and exit time constraints," *Transportation Research Record*, no. 2284, pp. 57–61, 2012.
- [10] 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.
- [11] L. F. Henderson, "The statistics of crowd fluids," *Nature*, vol. 229, no. 5284, pp. 381–383, 1971.
- [12] D.-W. Li, *Modeling and Simulation of Microscopic Pedestrian Flow in MTR Hubs*, Beijing Jiaotong University, Beijing, China, 2007.
- [13] Z. Huang, W. Feng, and Z. Hu, *Probability and Statistics*, People's Education Press, Beijing, China, 1982.
- [14] M.-Z. Li, *Graph Theory and Its Algorithms*, China Machine Press, Beijing, China, 2010.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

