

Received April 15, 2020, accepted April 23, 2020, date of publication April 28, 2020, date of current version May 13, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2991069

Big Data Analysis of Beijing Urban Rail Transit Fares Based on Passenger Flow

HONGHU GAO^{1,2}, SHIFENG LIU^{1,2}, GUANGMEI CAO¹, PENGFEI ZHAO¹, JIANHAI ZHANG³, AND PENG ZHANG³

¹School of Economics and Management, Beijing Jiaotong University, Beijing 100044, China

²Beijing Logistics Informatics Research Base, Beijing 100044, China

³Beijing Jingtong Urban Utility Tunnel Investment Company, Ltd., Beijing 100027, China

Corresponding authors: Shifeng Liu (shfliu@bjtu.edu.cn) and Pengfei Zhao (16125370@bjtu.edu.cn)

This work was supported in part by the Beijing Social Science Foundation under Grant 19JDGLA002, in part by the Ministry of Education in China (MOE) Project of Humanities and Social Sciences under Grant 19YJC630043, in part by the National Natural Science Foundation of China under Grant J1824031, and in part by the Beijing Logistics Informatics Research Base.

ABSTRACT This paper proposed improved measures for the shortest path fare scheme of urban rail transit. Firstly, this paper simulated Beijing rail transit by using Anylogic simulation technology and shortest path algorithm. Then, in order to find the travel time between any originations and destinations, this research measured the inbound time, waiting time, interval time, section running time, transfer time and outbound time. In addition, this paper used big data analysis technology to obtain the actual travel time distribution between any originations and destinations by processing the basic data of passengers entering and leaving the station. Finally, by comparing the valid path travel time calculated by any originations and destinations with the actual travel time distribution of passengers, the path taken by majority of passengers was pushed back to determine the ticket price based on the mileage of the path taken by the majority of passengers. The results reduced the dependence on government subsidies by rail transit operation and made up for the operation and maintenance costs.

INDEX TERMS Shortest path, Anylogic simulation, travel time, time distribution, big data analysis, pricing scheme of urban rail transit.

I. INTRODUCTION

Rail transit operators are not actively controlling rail transit operating and maintenance costs due to huge government subsidies. In other words, the huge subsidy from the government does not meet its incentive measures. While social benefits are the primary task of urban rail transit, urban rail transit operators should also have a certain degree of self-sufficiency to maximize the use of public resources and ensure the sustainable and effective operation of public resources [1].

In this paper, some lines of Beijing rail transit are simulated to calculate the shortest path between any originations and destinations and the travel time of their valid paths. With the help of big data technology to analyze the time distribution of passengers entering and leaving O-D stations, we can obtain the actual paths for most passengers. Through the comparison, the irrationality of the existing fare strategy of

Beijing urban rail transit is verified, and the decision-making advice for the rules of rail transit fare is provided according to the specific travel conditions of passengers. In the exploration in this paper, the fare setting is based on the mileage of the path taken by all or the vast majority of the passengers, rather than the mileage of the shortest path between O-D stations. Compared with the established fares, the rail transit fares made by the method of this article may be increased to a certain extent between some O-D stations, but the fare increasing should not be too high. This method of fare setting can not only provide the maximum convenience for the public to travel, but also increase the income of rail transit operators, thus reducing the financial subsidies of the government. In addition, this method can also reduce the passenger flow during peak periods to some extent.

II. LITERATURE REVIEW

The subway network of Beijing runs through the entire urban area and even covers some suburbs, which meets the needs

The associate editor coordinating the review of this manuscript and approving it for publication was Sabah Mohammed.

of passengers of a certain level of comfort and satisfaction in transportation [2], [3]. The subway network can effectively relieve the congestion of the city, some scholars have introduced congestion factors into the fare strategy formulation of urban transportation system. Basso *et al.* [4] introduced public transport congestion and traffic design, used congestion fees, transit fees (the level of subsidies), and transit frequencies as optimization variables to model and analyze the optimal price (welfare maximization) and design of transport services in a bimodal context. Tirachini *et al.* [5] determined and analyzed the interaction between congestion and congestion externalities in the design of urban transportation systems, as well as their relationship with ticket price. Game theory-based pricing model is an important tool to solve the problem [6], so Ding *et al.* [7] established a congestion pricing model of urban roads based on game theory, which introduced riding comfort into the cost function of the traditional bottleneck model, and verified the correctness of the model with an example. Yang and Tang [8] introduced a fare reward scheme (FRS) to relieve queuing congestion at transit stations. They found that, comparing the original fare, FRS results in an optimal reward ratio up to 50% and the system total time costs and average equilibrium trip costs reduced by at least 25% and 20%, respectively. They put the congestion factor into the model of rail transit fare, which enriched the idea of urban rail transit fare strategy.

Meanwhile, some scholars began to pay attention on the formulation of urban transportation fare strategy. Borger *et al.* [9] used the numerical optimization model to evaluate the Nash equilibrium of transportation prices, and provided an empirical study on the optimal pricing of transportation externality benefits. Gkritza *et al.* [10] analyzed different fare structures and estimated public transportation fares from the perspective of fare structures. And through the use of game theory in a variety of modes of transportation to develop different pricing models. Zhu *et al.* [11] established a dynamic fare model based on the division of passenger groups and the probability of passengers purchasing tickets, and on the basis of optimization theory and decision tree analysis (DTA). Huang *et al.* [12] proposed a new bus fare structure based on non-linear distance. Based on a tripartite game (including transportation management departments, passengers, and transportation companies), they established an optimization model to determine the optimal fare function and frequency, and solved the model by artificial swarm algorithm. Zhao and Yang [13] established a bi-level programming model by considering the factors social and economic benefits of the urban rail transit company and the related benefits of the passengers. An improved particle swarm optimization algorithm is designed to solve the model and an example is given to verify the feasibility and effectiveness of the model and related measures. Gong and Jin [14] established a tripartite game model of price adjustment plans including government, operating companies and passengers. Through the analysis of the trilateral benefits, they drew a study on whether the adjustment plan is successful. The

research results provide a research method for studying the feasibility of the price adjustment plans for urban transportation. Zhao and Zhang [15] found that the effects of the metro fare increase would significantly increase the cost burden of vulnerable residents on metro use. The most significant innovation of this research is that the real characteristics of passengers, such as the passenger flow, transfer and travel path selection.

On the other hand, the travel characteristics of passengers have attracted the attention of scholars, such as transfer characteristics, etc. [16], [17]. Xu *et al.* [18] analyzed the method of handling large passenger flow caused by operation delay at interchange station and obtained the time section of the station where the large passenger flow occurs as the basis of early warning. Liu and Xu [19] proposed the concept of continuous cycle of large passenger flow and a model based on traffic interval coordination is established to optimize the disposal method of large passenger flow in the latency period. Zhang *et al.* [20] analyzed the passenger's transfer time and travel time, and established a new algorithm, the result showed that the accuracy of the algorithm was accuracy. Lu *et al.* [21] identified the tourists of common diligence and a model that considered their travel preferences was established to learn and predict their next trip. In addition, some scholars also analyze the big data in the field of transportation to release the travel behavior of passengers [22]–[24]. Although the existing literature considers the travel characteristics of passengers, previous research rarely studies the relationship between the travel characteristics of passengers and the strategy of urban rail transit fare.

The current scholars' research mainly focused on congestion factors, fare strategy and the travel characteristics of passengers. However, there is no scholar to study the travel path based on passenger flow by big data analysis and the impact of travel path on the pricing of urban rail transit. In this paper, the fare setting rules of urban rail transit explored through big data analysis technology and simulation technology.

III. ALGORITHMS AND FARE PRINCIPLE

A. ALGORITHMS

Definition 1: The passenger travel network is represented by directed graph $G = (N, A)$. N is the set of nodes, $R \subseteq N$ and $S \subseteq N$ are the sets of origination and destination, respectively. A is the set of paths, q is the set of travel demand, which is actually the set of OD flow and q_{rs} represents the OD flow from the origination r to the destination s ($r \subseteq R, s \subseteq S$).

Definition 2: The actual distribution of the actual travel of passengers is represented by $T = (t, k)$. τ is the set of estimated time period, and $\tau \subseteq t$ is the estimated time period for a specific path. The time range is k and $k \in [(k-1)\tau, k\tau]$.

The Variables are shown as follows:

τ : the unit length of the estimated time period; $I_i(k)$: the traffic flow of node i flowing into the road network during time period k ; $O_j(k)$: the traffic outflow of node j to the road network during time period k ; $F_r(k)$: the traffic volume

on section r (the amount of traffic that passes through this section) during time period k ; $q_{ir}(k)$: the traffic volume (O-D flow) from point i to point j during time period k ; $\delta_{ir}(k)$: an distribution matrix used to establish the spatial and temporal influence relationship between O-D flow and section flow, indicating the proportional coefficient of O-D flow $q_{ij}(k)$ flowing through section r during the time period k ;

From the above information, we can establish the following basic relationships:

(1) The relationship between the nodal inflow traffic and O-D flow is shown in Equation 1:

$$I_i(k) = \sum_{j=1, j \neq i}^N q_{ij}(k), \quad \forall i \in N, K \in T \quad (1)$$

(2) The relationship between the nodal outflow traffic and O-D flow is shown in Equation 2:

$$O_j(t) = \sum_{k=1}^T \sum_{i=1, i \neq j}^N q_{ij}(k) \delta_{ij}^k(t), \quad k \leq t, \forall j \in N, t \in T \quad (2)$$

(3) The relationship between the amount of traffic that passes through this section and O-D flow is shown in Equation 3:

$$F_r(t) = \sum_{k=1}^T \sum_{i=1, i \neq j}^N q_{ij}(k) \delta_{ir}^k(t), \quad k \leq t, \forall r \in A, t \in T \quad (3)$$

There are slight differences for different modes of transportation, the appropriate flow Equation constraint can be selected according to the actual situation. In general, the unknown variable in the flow relationship is $N(N - 1)$ (the logarithm of O-D), where $N(N - 1) \gg R$ (the volume of flow on the observed section). The number of unknown variables in the Equations are much larger than the number of Equations, the unique O-D matrix cannot be determined only by Equations (1), (2) and (3). Therefore, a model must be used to approximate the O-D matrix. The process of O-D matrix estimation is to find a set of solutions closest to the actual O-D matrix from the feasible solution space of the O-D matrix that satisfy the flow constraint relationship. The general expression of the model can be expressed as Equation 4:

$$\begin{aligned} & \min f(D, F(1), F(2)) \\ & s.t. F = \text{assignment}(D) \end{aligned} \quad (4)$$

where D is the estimated O-D matrix, $F(1)$ is the actual observed traffic volume of the section; and $F(2)$ is the traffic volume obtained according to the estimated O-D matrix.

After more than 30 years of development, the dynamic O-D matrix estimation method has been constantly improved, forming different types of estimation models and algorithms. According to whether the model contains DTA (Dynamic Traffic Assignment) module, it can be divided into DTA-based model and non-DTA-based model. According to the types of collected data source, it can be divided into fixed data source and mobile data source estimation model. According to the forms of the model, it can be divided into mathematical optimization model and statistical analysis model. According

to the network structure, it can be divided into closed network model and open network model, etc. At present, there is still no unified classification standard for the estimation model. This paper introduces several representative, referential and widely used models to provide reference for dynamic O-D matrix estimation of rail transit network. Maximum Likelihood Model is adopted in this paper. Maximum likelihood estimation is one of the commonly used methods in O-D matrix estimation, which is based on the maximum likelihood theory in statistics. This method maximizes the conditional likelihood of the target O-D matrix and the observed traffic flow in the real O-D matrix. The basic model of the maximum likelihood method is shown as follows (Equation 5):

$$\begin{aligned} \text{Max } S &= - \sum_{i=1}^N \sum_{j=1}^N q_{ij} * \ln \left(r_{ij} * \left(\frac{q_{ij}}{Q} \right) \right) \\ s.t. F_r &= \sum_{i=1}^N q_{ij} * \delta_{ij}^r, \quad r_{ij} = \frac{\hat{q}_{ij}}{\sum_{i=1}^N \sum_{j=1}^N \hat{q}_{ij}}, \quad Q = \sum_{i=1}^N \sum_{j=1}^N q_{ij} \end{aligned} \quad (5)$$

where r_{ij} is the prior probability of the occurrence of O-D on (i, j) which is obtained according to the statistics of the historical O-D quantity, and Q is the total O-D quantity.

The principle of maximum likelihood method is simple, it does not need to determine the weight matrix, and it is a method with application value. For example, Spiess (1983, 1987) proposed the maximum likelihood model of O-D estimation based on the maximum likelihood principle. Cascetta and Nguyen (1988) proposed a method to estimate the O-D matrix by using the traffic volume of the section, using the idea of reviewing traffic distribution graph and maximum likelihood theory. Nihan and Davis (1989) used the maximum likelihood model to estimate the O-D matrix of intersections.

B. FARE PRINCIPLE

Beijing rail transit fares are mainly divided into two stages. Before the implementation of the new Beijing rail transit fare regulations, Beijing rail transit has always implemented welfare-friendly low prices from October 2007 to December 2014, Beijing rail transit fare was reduced to 2 RMB except for the airport express rail line. At this time, the fare setting method lacked dynamics, the price was too low and the government subsidies for rail transit also increased every year. In the second phase, Beijing rail transit started to implement new fares since December 28, 2014. And the urban rail transit fares have changed relatively, from the previous unified fare of 2 RMB to the starting price of 3 RMB. The new fares of Beijing urban rail transit are shown in Table 1: 3 RMB within 6 km (inclusive); 6 km to 12 km (inclusive) is 4 RMB; 12 km to 22 km (inclusive) is 5 RMB; 22km to 32km (inclusive) is 6 RMB; for the part above 32km, you can ride 20km for every additional 1 RMB.

The mileage between any O-D stations is calculated according to the distance of the shortest path (excluding the subway transfer distance), that is, the algorithm of the shortest

TABLE 1. Fare principle of Beijing rail transit.

The mileage(km)	Ticket price(RMB)
0-6	3
6-12	4
12-22	5
22-32	6
32-52	7
52-72	8

path is used to calculate the shortest path between O-D stations as the origination and destination of the passenger are known, and then use the mileage of the shortest path (excluding the subway transfer distance) as the basis for calculating the fare. However, the calculation method of Beijing rail transit using the shortest path for mileage brings many problems, such as: charging by the shortest path makes the rail transit fare lower, and the low fare is not conducive to the shunting of passenger flow during peak hours, causing too large local traffic flow and causing potential safety hazards. At the same time, it also reduces the comfort of passengers during the ride, which is particularly obvious especially during peak hours. Meanwhile, because the full load rate is too high, the passageways, elevators, platforms, and compartments are crowded, and everyone is shoulder to shoulder, passengers' comfort is not guaranteed and safety accidents are prone to occur.

According to the query data of Beijing Transport Institute, during the morning and evening peak hours since 2009, the full load rate of Beijing rail transit has reached more than 100%. For example, the maximum load rates from 2009 to 2011 were 133%, 135% and 138%, respectively. Although Beijing urban rail transit has been in the profit stage in the past one or two years, if the operating costs and maintenance costs are included [25], Beijing urban rail transit is still in a huge loss stage. So the government still needs to invest a lot of subsidies every year to make up for the losses. The low fare formed according to the shortest path is also not conducive to the influx of private capital and the motivation of operating enterprises to reduce the cost of rail transit. Therefore, the establishment of a fare formulation system that matches the financial subsidies and the scale of the rail transit network is favorable conditions to ensure the sustainable development of Beijing rail transit.

IV. DATA ANALYSIS OF URBAN RAIL TRANSIT

A. THE PROBABILITY OF ENTERING EACH STATION

According to the basic passenger flow data provided by the rail transit command center, the average daily passenger volume can reach 10 million. However, the number of passengers entering and leaving from various stations every day is quite different. For example, the number of passengers entering and leaving from Xizhimen station of line 4 and Wangfujing station of line 1 is large, while the passenger traffic from Haojiafu station of line 6 and the Huagong station of line 7 is very small.

TABLE 2. The probability of passengers inbound and outbound at each station (partial data).

A	B	C	D	E	F
1	Jianguomen	31399	0.00423	427	0.00886
2	Pingguoyuan	43577	0.00672	438	0.00909
3	Babaoshan	29456	0.00436	391	0.00811
4	Yuquanlu	29470	0.00408	301	0.00624
5	Wukesong	37840	0.01062	362	0.00751
6	Fuxingmen	33784	0.00470	565	0.01172
7	Xidan	32402	0.00936	159	0.00330
8	Wangfujin	33932	0.01325	331	0.00687

where A=Number, B=Station, C= Inbound passengers, D= Inbound probability, E= Outbound passengers, F= Outbound probability.

In order to simulate the actual situation of Beijing urban rail transit more realistically, the passenger traffic inbound and outbound data for the week from May 8, 2017 to May 14, 2017 were taken before the modeling. First of all, data cleansing was performed to the passenger flow data, individual abnormal data was filtered out, and 6,228,567 data were randomly selected as the research object from the processed data. From these more than 6 million data, the number of passengers entering each station was calculated, and according to this, the probability of passengers entering at each station is calculated as shown in the table 2 (only some data are displayed here). In addition, according to the number of passengers entering from one station and leaving from another station, the corresponding outbound probability is shown in the Table 2 (only part of the data is shown here). Since the data is randomly selected from the data of passenger flow in a week, and the data processing is completed by computer without human interference, and the amount of selected data is huge, the researched data has a certain representativeness and higher accuracy.

B. TRAVEL TIME OF THE PASSENGERS

Travel time is the total time for passengers from the origination to the destination, including the walking time from the ticket gate of the entrance to the subway door, the section running time, the parking interval of the intermediate stations, the subway transfer time (transfer walking time and transfer waiting time), and the walking time from the subway door to the ticket gate of the exit. And the shortest travel time between O-D stations is the most important factor affecting the passengers' choice of the path. In general, the path corresponding to the shortest travel time is the shortest path. At the same time, the path corresponding to the shortest travel time is also the most likely path that passengers may choose when taking rail transit. In the actual road network, the following two situations may also exist: the long mileage (excluding the subway transfer distance) corresponds to a short travel time of all valid paths for the start stations, and the short mileage corresponds to a long travel time. In order to measure the travel time accurately, this paper obtains the calculation method of travel time from Beijing Metro Operation Administration Corporation Limited. In addition, due to the large number of stations on Beijing rail transit lines, complicated

TABLE 3. The rail transit section distance and section running time (partial data).

Line	Section distance(m)	Running time(s)	Line	Section distance(m)	Running time(s)	Line	Section distance(m)	Running time(s)
1	400	61	10	700	69	13	1100	98
1	800	87	10	800	73	13	1200	110
1	900	90	10	900	75	13	1400	131
1	1200	109	10	1000	83	13	1800	145
1	1300	114	10	1100	89	13	2100	159
1	1400	118	10	1200	99	13	2200	161
1	1500	122	10	1300	103	13	2300	164
1	1600	128	10	1500	115	13	2500	184
1	1700	133	10	1600	121	13	2800	201
1	1800	138	10	1700	124	13	3600	244
1	1900	145	10	1800	135	13	4800	306
1	2000	147	10	2400	163	13	4900	309
1	2600	196				13	6700	421

TABLE 4. The parking interval of different lines (partial data).

Line	Parking interval(s)	Line	Parking interval(s)
1	33	9	45
2	32.5	10	37.5
4	37	13	33.3
5	35	15	41.3
8	42	Yizhuang	30

road conditions and heavy workload, only some lines such as line 1, line 2, line 4, and line 5 are surveyed on-site. The survey data of the rail transit is as follows.

1) RAIL TRANSIT SECTION DISTANCE AND SECTION RUNNING TIME

According to the section distance provided by the Beijing Urban Rail Transit Clearing and Settlement Center, we can get the distance between adjacent stations on any lines. At the same time, we obtained the time required between any adjacent stations of the rail transit through on-site measurement. Then, the interval running time of the same distance interval on the same line is obtained by processing the abnormal data, see Table3.

2) PARKING INTERVAL OF DIFFERENT LINES

Due to the different busy degree of the various rail transit lines during peak hours, the parking interval of each line is different. For example, the morning peak passenger flow of line 9 is large, so the average parking interval of each station on line 9 is 55 seconds, while the average parking interval of line 1 is 33 seconds. And by comparing the parking interval of different stations on the same line, it is found that the parking interval of different stations on the same line are the same or have little difference. Therefore, this paper cleans the measured parking interval data to obtain the average parking interval of each line as shown in Table 4.

3) DEPARTURE INTERVAL OF DIFFERENT LINES

According to the departure interval of different lines provided by the Beijing Urban Rail Transit Clearing and Settlement

TABLE 5. The parking interval of different lines (partial data).

Line	Departure interval(s)	Line	Departure interval(s)
1	200	9	180
2	140	10	240
4	180	13	120
5	130	15	220
8	200	Yizhuang	530

Center, we can get the departure interval of different lines (see Table 5).

4) SUBWAY TRANSFER TIME OF DIFFERENT LINES

The interchange station is a key node in the urban rail transit network. Interchange stations play an extremely important role in the networked operation of Beijing urban rail transit. A convenient and fast interchange system can not only fully meet the needs of passengers, but also help to further the social and economic benefits of urban public transportation [26]. There are many interchange stations in Beijing rail transit, and the transfer time in different directions at the same interchange station is different. For example, the transfer subway time from line 4 to line 2 at Xizhimen station is different from the time from line 2 to line 4. Through the on-site measurement of the subway transfer time, although the time in different directions of the same interchange station is different, the difference is very small, which has little impact on the final verification and the study of fare formulation. Basic transfer time between different lines of Beijing urban rail transit interchange stations are obtained, as shown in Table 6.

C. ACTUAL TIME BETWEEN O-D STATIONS

1) THE CLEANING OF ACTUAL TIME BETWEEN O-D STATIONS

Due to large amount data of the passengers' time of entering and leaving the station, some incomplete data records may appear in the data, for example, through data collation, it is found that the data of individual passengers entering or leaving the station is lost, and the travel time of some passengers

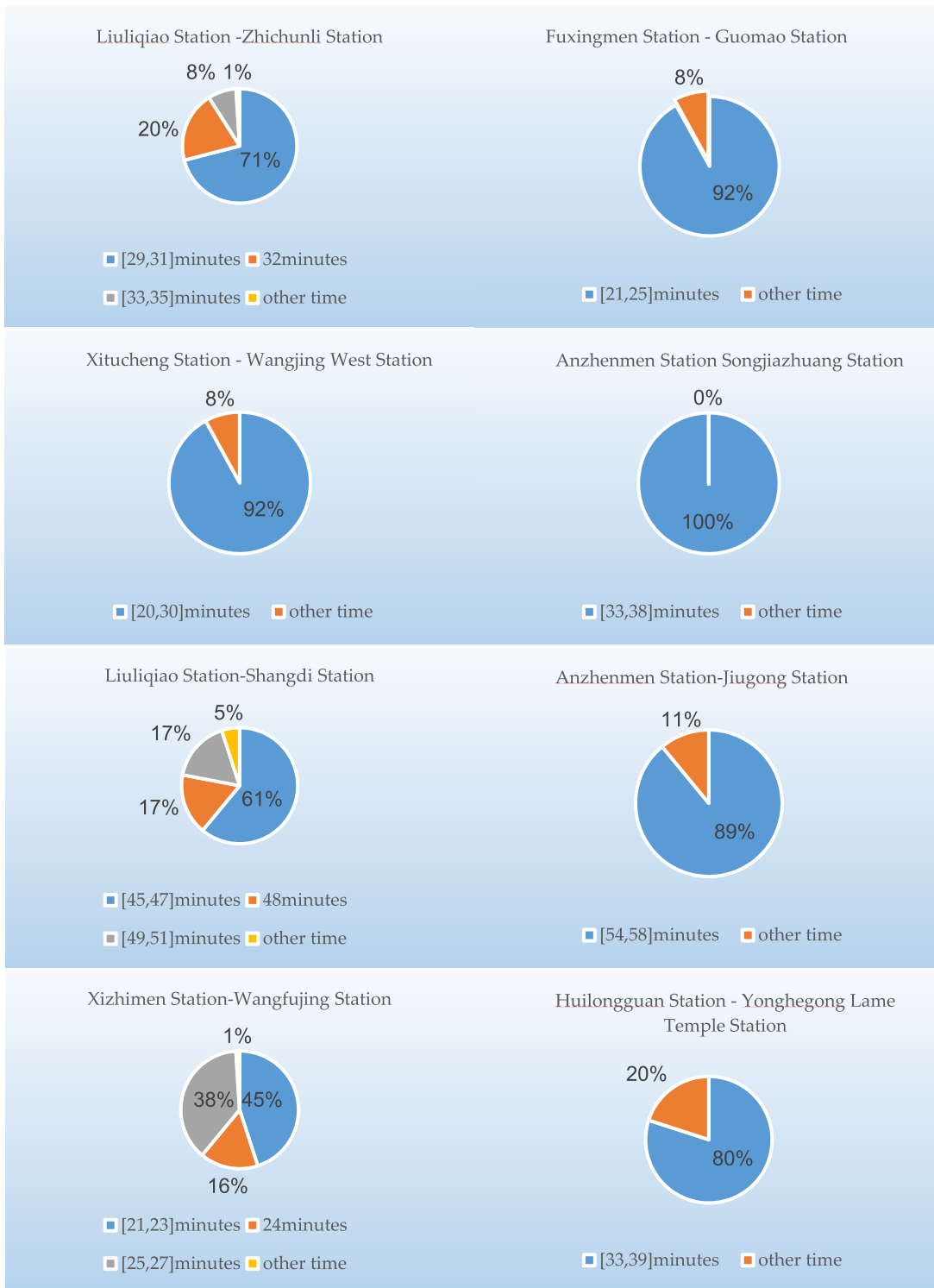


FIGURE 1. Distribution of actual time between O-D stations.

is inevitably far beyond normal among a large number of passengers. In these cases, big data analysis technology must be used to perform preliminary cleaning of the primary data to avoid the interference of abnormal data on the real time distribution between O-D stations, thereby making the data

more accurate and reliable. This lays the foundation for later accurate verification of the irrationality of fare setting based on the mileage of the shortest path and the exploration of a more reasonable fare setting scheme between any origins and destinations.

TABLE 6. The transfer time of different lines (partial data).

Number	Interchange station	Line	Line	Transfer time(s)
1	Gongzhufen	1	10	120
2	Military museum	1	9	180
3	Fuxingmen	1	2	60
4	Xidan	1	4	180
5	Dongdan	1	5	180
6	Jianguomen	1	2	120
7	Guomao	1	10	240
8	Xizhimen	2	4	120

2) DISTRIBUTION OF ACTUAL TIME BETWEEN O-D STATIONS
 It is found from the cleaned data that the travel time of most passengers or almost all passengers between some O-D stations are concentrated in a smaller interval. That is to say, the vast majority of passengers follow the same path between any O-D stations, which provides a good support for fare verification and inquiry. Due to the large amount of data, this paper just shows parts of the travel time data distribution related to subsequent research (see Figure 1).

V. THE SIMULATION MODEL AND RESULT ANALYSIS

A. THE SIMULATION MODEL

The main function in the simulation model is Main which contains the following information: (1) the data of the probability of passengers entering any of the 328 stations of the rail transit, the correlation set and the corresponding function of reading the contents of the file; (2) the data of the outbound probability of all reachable terminals after entering station, the correlation set, and the corresponding function to read the file; (3) the data of the location and information of the rail transit gate, the correlation set and the corresponding function of reading the location and information of the gate [27]; (4) the set data of distance between any two adjacent stations of rail transit, and the output file that represents the simulation result. In addition, an object class representing the information of each station of rail transit is added to the main function, and an object class named Passenger is stored to simulate the state diagram of the system [28].

Passanger contains a state diagram that shows the entire process from the origination to the destination of the passengers according to the shortest path and related functions. Firstly, we insert a graph object in the Passanger object, and then establish variable edges, model cyclenumbers (simID), simulation numbers (CycleCount), and other variables. Among them, Vertexs represents the vertex set of the stations, all the stations on the shortest path are represented by a vertex. Finally, the function that related the shortest path is generated. For example, rand SelectStation indicates the originations and the destinations that are randomly generated in the system, generate_trip indicates the line generated in the system from the originations and the destinations, generate_vertexs represents all the stations represented by vertices on the line, init_edges represents initializing the previously generated edges, set_vert_std represents the stations on the shortest path obtained by the

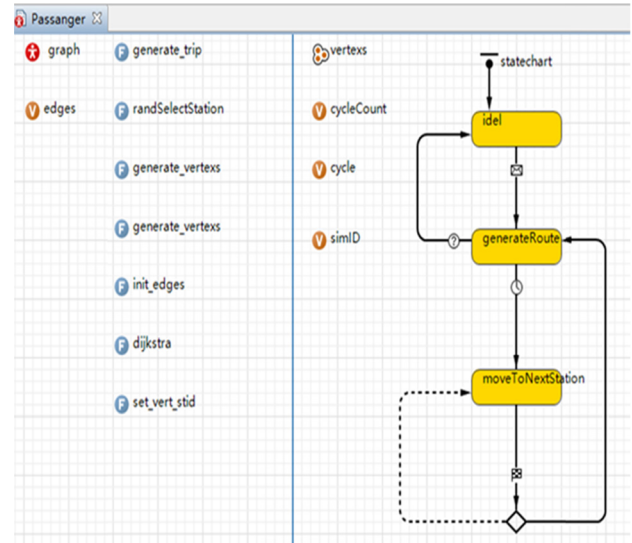


FIGURE 2. Passanger object of the Rail transit simulation model.

corresponding id number, and dijkstra indicates that the shortest path that required by the shortest path algorithm (see Figure 2).

Through the collation of a large amount data, the coordinates of each station of the rail transit, the distance between the stations, and the probability of entering and leaving each O-D stations can be obtained. And then, we import the collated rail transit data into the Anylogic simulation system. In order to represent the current status of the shortest path between the originations and the destinations completely, the simulation result includes the number of simulation cycles representing the number of simulation times, the names of all the stations on the shortest path obtained by the shortest path algorithm between the originations of each simulation, the distances between all adjacent stations on the shortest path. The simulation results are shown in table 7.

B. ANALYSIS OF THE RESULTS

In Beijing rail transit, the fare is calculated according to the mileage of the shortest path between O-D stations, but the subway transfer distance is not considered, so the travel time of the shortest path is not necessarily the shortest, nor is it necessarily the best path for passengers.

The verification of rail transit fares in this section includes three cases. The first case is that the origination and the destination of the O-D station are on the same line. In this case, among all valid paths between O-D stations, one of the valid paths between O-D stations does not require transfer. The second case is that the O-D stations are on different lines, that is, the origination and the destination are on different lines. In other words, the valid path from the origination to the destination requires at least one subway transfer. When O-D stations are different, it can be also divided into two situations: one is that all valid paths from the origination and the destination require the same number of subway transfers.

TABLE 7. Part of the simulation results.

Simulation number	Origination	Destination	Line	Line	Distance(m)
1	Fuxingmen	Xindan	1	1	1600
1	Xindan	Tiananmen West	1	1	1200
1	Tiananmen West	Tiananmen East	1	1	900
1	Tiananmen East	Wangfujin	1	1	900
1	Wangfujin	Dongdan	1	1	800
1	Dongdan	Jianguomen	1	1	1200
1	Jianguomen	Yonganli	1	1	1400
1	Yonganli	Guomao	1	1	800
2	Liuliqao	Liuliqao East	9	9	1300
2	Liuliqao East	Beijingxi Railway	9	9	1200
2	Beijingxi Railway	Military Museum	9	9	1400
2	Military Museum	Baiduizi	9	9	1900
2	Baiduizi	BaishiQiao Nan	9	9	900
2	BaishiQiao Nan	National Library	9	4	1100
2	National Library	Weigongcun	4	4	1700
2	Weigongcun	Renmin University	4	4	1100
2	Renmin University	Haidianhuangzhuang	4	4	1100
2	Haidianhuangzhuang	Zhichunli	4	10	1000
3	Anzhenmeng	HuixinxijieNankou	10	5	1000
3	HuixinxijieNankou	Hepingxiqiao	5	5	1000
3	Hepingxiqiao	HepingliBeijie	5	5	1100
3	HepingliBeijie	Yonghegong Lama Temple	5	5	1200
3	Yonghegong Lama Temple	Beixingqiao	5	5	900
3	Beixingqiao	Zhangzizhonglu	5	5	800
3	Zhangzizhonglu	Dongsi	5	5	1000
3	Dongsi	Dongshikou	5	5	800
3	Dongshikou	Dongdan	5	5	900
3	Dongdan	Chongwenmen	5	5	800
3	Chongwenmen	Ciqikou	5	5	900
3	Ciqikou	Tiantandongmen	5	5	1200
3	Tiantandongmen	Puhuangyu	5	5	1900
3	Puhuangyu	Liujiaoyao	5	5	900
3	Liujiaoyao	Songjiazhuang	5	5	1700
4	Xitucheng	Mudanyuan	10	10	1300
4	Mudanyuan	Jiandemen	10	10	1000
4	Jiandemen	Beitucheng	10	10	1100
4	Beitucheng	Anzhenmen	10	10	1000
4	Anzhenmen	HuixinxijieNankou	10	10	1000
4	HuixinxijieNankou	Shaoyaoju	10	13	1700
4	Shaoyaoju	Wangjing West	13	13	2200

The other case is that among all the valid paths from the origination to the destination, some have few subway transfers, while others have more.

1) THE VALID PATH DOES NOT REQUIRE SUBWAY TRANSFER

In this situation, the origination and the destination are on the same line, and at least one valid path does not require a subway transfer, while other valid paths require. Among all valid paths, the shortest path require subway transfer. However, in the fare formulation of Beijing urban rail transit, the subway transfer distance is not calculated when calculating the shortest path between the origination and the destination. Therefore, when the travel time of the shortest path is almost the same as that of the non-shortest path without subway transfer, or when the travel time of the former is much greater than that of the latter, almost most passengers choose the path without subway transfer. We can get that the mileage

of the shortest path as the basis of O-D station pricing is not reasonable.

For example, there are two valid paths between Liuliqiao station and Zhichunli station. The first valid path (see Table 8) which is the shortest (excluding the subway transfer distance) consists of taking line 9 from Liuliqiao station to National Library station, and then taking line 4 from National Library station to Haidianhuangzhuang station, finally, taking line 10 from Haidianhuangzhuang station to the destination at Zhichunli station. Among them, parking 6 times on line 9, 3 times on line 4 and 1 time on line 10, each parking interval is 45 seconds, 37 seconds and 37.5 seconds, respectively, that is to say, the total parking interval on this path is approximately 7 minutes. We know that the inbound time, waiting time, interval time, transfer time, transfer waiting time and outbound time of this path add up to about 26 minutes, therefore, with the time between parking, the total travel time of passengers on this path is about 33 minutes. In another valid path (see Table 9), passengers take line 10 at Liuliqiao station

TABLE 8. The valid path between Liuliqiao station and Zhichunli station (1).

Line	Station	Station	Distance(m)	Time(s)	Time(s)
			Inbound time	90	
			Waiting time	90	
9	Liuliqiao	Liuliqiao East	1300	104	45
9	Liuliqiao East	Beijingxi Railway	1200	101	45
9	Beijingxi Railway	Military Museum	1400	109	45
9	Military Museum	Baiduizi	1900	135	45
9	Baiduizi	BaishiQiao Nan	900	89	45
9	BaishiQiao Nan	National Library	1100	97	45
			Transfer time	0	
			Waiting time	90	
4	National Library	Weigongcun	1700	121	37
4	Weigongcun	Renmin University	1100	90	37
4	Renmin University	Haidianhuan	1100	90	37
			Transfer time	60	
			Waiting time	90	
10	Haidianhuan	Zhichunli	1100	83	37.5
			Outbound time	90	

TABLE 9. The valid path between Liuliqiao station and Zhichunli station (2).

Line	Station	Station	Distance(m)	Time(s)	Time(s)
			Inbound time	90	
			Waiting time	90	
9	Liuliqiao	Liuliqiao East	1300	104	45
9	Liuliqiao East	Beijingxi Railway	1200	101	45
9	Beijingxi Railway	Military Museum	1400	109	45
9	Military Museum	Baiduizi	1900	135	45
9	Baiduizi	BaishiQiao Nan	900	89	45
9	BaishiQiao Nan	National Library	1100	97	45
			Transfer time	0	
			Waiting time	90	
4	National Library	Weigongcun	1700	121	37
4	Weigongcun	Renmin University	1100	90	37
4	Renmin University	Haidianhuan	1100	90	37
			Transfer time	60	
			Waiting time	90	
10	Haidianhuan	Zhichunli	1100	83	37.5
			Outbound time	90	

and directly arrive at Zhichunli station, stop for 11 times, with a total parking interval about 4 minutes. Similar to that of the travel time calculation method in the first valid path, the total travel time of passengers on this path is about 31 minutes.

According to the statistics of the travel time distribution between Liuliqiao station and Zhichunli station, the travel time distribution ratio in the range [29], [31] is 61%, the travel time distribution ratio for 32 minutes is 20%, the travel time distribution ratio in the range [33], [35] is 8%, and the travel time distribution ratio for the remaining is 1%. Therefore, it can be determined that most passengers choose line 10 as the best path to travel without subway transfer. At this point, it is unreasonable to set the price according to the mileage of the shortest path (excluding the subway transfer distance) in the regulations of the rail transit fare, so the mileage of the travel path of most passengers should be taken as the basis of the fare.

2) THE VALID PATH REQUIRES THE SAME NUMBER OF SUBWAY TRANSFERS

In this situation, the origination and the destination are not on the same line. In all valid paths, passengers need to make the same number of subway transfers to reach the destination.

For example, there are two valid paths between Xizhimen station and Wangfujing station. The first valid path (see Table 10) consists of taking line 2 from Xizhimen station to Fuxingmen station, and then taking line 1 from Fuxingmen

station to the destination at Wangfujing station. Among them, parking 3 times on line 2 and 4 times on line 1, each parking interval is 32.5 seconds and 33 seconds, respectively, that is to say, the total parking time on this path is approximately 4 minutes. We know that the inbound time, waiting time, interval time, transfer time, transfer waiting time and outbound time of this path add up to about 19 minutes, therefore, with the time between parking, the total travel time of passengers on this path is about 23 minutes. Another valid path (see Table 11) includes taking line 2 from Xizhimen station to Xidan station and line 1 from Xidan station to Wangfujing station. Similar to that of the travel time calculation method in the first valid path, the total travel time of passengers on this path is about 25 minutes.

According to the statistics of the travel time distribution between Xizhimen station and Wangfujing station, the travel time distribution ratio in the range [21], [23] is 45%, the travel time distribution ratio for 24minutes is 16%, the travel time distribution ratio in the range [25], [27] is 38%, and the travel time distribution ratio for the remaining is 1%. Therefore, it can be determined that relatively more passengers choose the path of line 2 and then take line 1 as the best choice. However, the fare strategy of Beijing rail transit from Xizhimen station to Wangfujing station is based on the shortest path (excluding the subway transfer distance) for passengers taking line 4 at Xizhimen station and then taking line 1 at

TABLE 10. The valid path between Xizhimen station and Wangfujing station (1).

Line	Station	Station	Distance(m)	Time(s)	Time (s)
			Inbound time	90	
			Waiting time	70	
2	Xizhimen	Chegongzhuang	900	89	32.5
2	Chegongzhuang	Fuchengmen	1000	93	32.5
2	Fuchengmen	Fuxingmen	1800	135	32.5
			Transfer time	60	
			Waiting time	80	
1	Fuxingmen	Xidan	1600	128	33
1	Xidan	Tiananmen West	1200	109	33
1	Tiananmen West	Tiananmen East	900	90	33
1	Tiananmen East	Wangfujing	900	90	33
			Outbound time	90	

TABLE 11. The valid path between Xizhimen station and Wangfujing station (2).

Line	Station	Station	Distance(m)	Time (s)	Time (s)
			Inbound time	90	
			Waiting time	90	
4	Xizhimen	Xinjiekou	1000	85	37
4	Xinjiekou	Pinganli	1100	90	37
4	Pinganli	Xisi	1100	90	37
4	Xisi	LijingHutong	900	76	37
4	LijingHutong	Xidan	1000	85	37
			Transfer time	180	
			Waiting time	80	
1	Xidan	Tiananmen West	1200	109	33
1	Tiananmen West	Tiananmen East	900	90	33
1	Tiananmen West	Wangfujing	900	90	33

Xidan station to Wangfujing station. According to the time distribution between O-D stations, most passengers will not choose the shortest path, so it is unreasonable to set the ticket price according to the mileage of the shortest path (excluding the subway transfer distance).

3) THE VALID PATH REQUIRES THE SAME NUMBER OF SUBWAY TRANSFERS

In this situation, the origination and the destination are not on the same line, and the valid path requires different number of subway transfers.

For example, there are two valid paths between Liuliqiao station and Shangdi station. The first valid path (see Table 12) consists of taking line 10 from Liuliqiao station to Zhichunlu

station, and then taking line 13 from Zhichunlu station to the destination at shangdi station. Among them, parking 12 times on line 10 and 2 times on line 13, each parking interval is 37.5 seconds and 33.3 seconds, respectively, that is to say, the total parking time on this path is approximately 8 minutes. We know that the arrival time, inbound time, interval time, transfer time, transfer waiting time and outbound time of this path add up to about 39 minutes, therefore, with the total parking time, the total travel time of passengers on this path is about 47 minutes. Another valid path (see Table 13) includes taking line 9 from Liuliqiao station to National library station, changing to line 1 at National library station to Zhichunlu station, and taking line 13 from Zhichunlu station to shangdi station. Similar to that of the travel time calculation method in the first valid line, the total travel time of passengers on this path is about 49 minutes.

According to the statistics of the travel time distribution between Liuliqiao station and Shangdi station, the travel time distribution ratio in the range [45], [47] is 61%, the travel time distribution ratio for 48 minutes is 17%, the travel time distribution ratio in the range [49], [51] is 17%, and the travel time distribution ratio for the remaining is 5%. Therefore, it can be determined that relatively more passengers choose the path of line 10 and then taking line 13 as the best choice. However, the fare strategy of Beijing rail transit from Liuliqiao station to Shangdi station is based on the shortest path (excluding the subway transfer distance) for passengers to take line 9 at Liuliqiao station and change to line 4, line 10 and line 13 to Shangdi station. According to the time distribution between O-D stations, we can get that most passengers will not choose the shortest path. Therefore, it is unreasonable to set the ticket price based on the mileage of the shortest path (excluding the subway transfer distance)

VI. FARE STRATEGY OF URBAN RAIL TRANSIT

According to a questionnaire survey by the clearing center of Beijing rail transit, nearly 90% of passengers in rush hour are familiar with Beijing’s road network. They can accurately estimate the travel time between O-D stations based on their previous subway experience and Baidu map. At the same time, when investigating the travel paths between any O-D stations, nearly 80% of passengers would choose the shortest travel time instead of the shortest path under the condition of few subway transfers. In addition, the subway transfer distance between different paths is not taken into account when calculating the path mileage, so choosing the shortest path is not necessarily the shortest travel time.

Considering that there are many Beijing rail transit lines, more stations passing by each line, and interchange stations between each line, therefore, this paper only conducts field data research on line 1, line 2, line 4, line 5, line 8, line 9, line 10, line 13, line 15 and Yizhuang line. We compare the travel time of each valid path between any O-D stations with the actual travel time of passengers between O-D stations, and then we can derive the travel path for most or all passengers between any O-D stations. We think that it is scientific and

TABLE 12. The valid path between Liuliqiao station and Shangdi station (1).

Line	Station	Station	Distance(m)	Time (s)	Time (s)
			Inbound time	90	
			Waiting time	120	
10	Liuliqiao	Lianhuaqiao	2400	163	37.5
10	Lianhuaqiao	Gongzhufen	1000	83	37.5
10	Gongzhufen	Xidiaoyutai	2400	163	37.5
10	Xidiaoyutai	Cishousi	1200	99	37.5
10	Cishousi	Chedaogou	1600	121	37.5
10	Chedaogou	Changchunqiao	1200	99	37.5
10	Changchunqiao	Huoqiying	1000	83	37.5
10	Huoqiying	Bagou	1500	115	37.5
10	Bagou	Suzhoujie	1100	89	37.5
10	Suzhoujie	Haidianhuangzhuang	1000	83	37.5
10	Haidianhuangzhuang	Zhichunli	1000	83	37.5
10	Zhichunli	Zhichunli	1100	89	37.5
			Transfer time	240	
			Waiting time	60	
13	Zhichunli	Wudaokou	1800	145	33.3
13	Wudaokou	Shangdi	4900	309	33.3
			Outbound time	90	

reasonable to set the ticket price according to the travel path for most or all passengers between any O-D stations. In this paper, the fare strategies of urban rail transit are shown as follows:

1. The Originations and the Destinations are on the Same Line

(1) The shortest valid path does not require subway transfer
 This situation is also the simplest. The origination and the destination are on the same line, and the valid path which is also the shortest path. For example, there is only one valid path between Fuxingmen station and Guomao station. In this valid path, passengers take line 1 at Fuxingmen station and directly arrive at Guomao station, the total travel time of passengers on this path is about 22 minutes. According to the statistics of the travel time distribution between O-D stations, the travel time distribution ratio in the range [21], [25] is 92%. Therefore, in this situation, the fare strategy should be formulated according to the mileage of the shortest path without subway transfer.

(2) The shortest valid path requires subway transfer and the travel time is not much different from the non-transfer path

At this point, there are at least two valid paths from the origination and the destination, and the mileage of the transfer path (the shortest) is similar to the travel time of the non-

TABLE 13. The valid path between Liuliqiao station and Shangdi station (2).

Line	Station	Station	Distance(m)	Time (s)	Time (s)
			Inbound time	90	
			Waiting time	90	
9	Liuliqiao	Liuliqiao East	1300	104	45
9	Liuliqiao East	Beijingxi Railway	1200	101	45
9	Beijingxi Railway	Military Museum	1400	109	45
9	Military Museum	Baiduizi	1900	135	45
9	Baiduizi	BaishiQiao Nan	900	89	45
9	BaishiQiao Nan	National Library	1100	97	45
			Transfer time	0	
			Waiting time	90	
4		Weigongcun	1700	121	
4	National Library	Renmin University	1100	90	
4	Weigongcun	Haidianhuangzhuang	1100	90	
			Transfer time	60	6
			Waiting time	120	3
10		Zhichunli	1000	83	37.5
10	Haidianhuangzhuang	Zhichunli	1100	89	37.5
			Transfer time	240	
			Waiting time	60	
13		Wudaokou	1800	145	33.3
13	Zhichunli	Shangdi	4900	309	33.3
			Outbound time	90	
			Inbound time	90	

transfer path (excluding the subway transfer distance). For example, there are two valid paths between Liuliqiao station and Zhichunli station (see3.2.1). Therefore, when the travel time of the transfer path (shortest path) is almost the same as that of the path without subway transfer, the majority of passengers will choose the path without subway transfer. In this situation, we know that the rail transit fare strategy should be based on the mileage of the path without subway transfer rather than the mileage of the shortest path.

(3) The shortest valid path requires subway transfer and the travel time is much different from the non-transfer path

At this point, there are at least two valid paths from the origination and the destination, in addition, the shortest path

requires subway transfer and the travel time is much different from the non-transfer path. That is to say, the travel time on the non-transfer path is significantly longer than the time on the subway transfer path. For example, there are two valid paths between Anzhenmen station and Songjiazhuang station. The first valid path which is the shortest path (excluding the subway transfer distance) consists of taking line 10 from Anzhenmen station to HuixinxijieNankou station, and then taking line 5 from HuixinxijieNankou station to the destination at Songjiazhuang station. In another valid path, passengers take line 10 at Anzhenmen station and directly arrive at Songjiazhuang station. The travel time on the two valid paths is about 35 minutes and 43minutes, respectively. According to the statistics of the travel time distribution between O-D stations, the travel time distribution ratio in the range [33], [38] is 100%. In other words, the first valid path is the passenger's first choice. In this situation, we know that the rail transit fare strategy should be based on the mileage of the shortest path with subway transfer.

2. The Originations and the Destinations are not on the Same Line: different number of subway transfers

(1) The path with few number of subway transfers is the shortest

In this situation, the path (excluding the subway transfer distance) is the shortest and the number of transfers is small, so the travel time may be less. For example, there are three valid paths between Xitucheng station and Wangjing West station. The first valid path which is the shortest (excluding the subway transfer distance) consists of taking line 10 from Xitucheng station to Shaoyaoju station, and then taking line 13 from Shaoyaoju station to the destination at Wangjing West station. In another valid path consists of taking line 10 from Xitucheng station to Beitucheng station, and then taking line 8 from Beitucheng station to Olympic Green station, finally, taking line 15 from Olympic Green station to the destination at Wangjing West station. The third valid path consists of taking line 10 from Xitucheng station to HuixinxijieNankou station, and then taking line 5 from HuixinxijieNankou station to Datunlu East station, finally, taking line 15 from Datunlu East station to the destination at Wangjing West station. The travel time on the three valid paths is about 25 minutes, 35 minutes and 35 minutes, respectively. According to the statistics of the travel time distribution between O-D stations, the travel time distribution ratio in the range [20], [30] is 92%, the travel time distribution ratio in the range [30], [40] is 8%. In other words, the first valid path is the passenger's first choice. In this situation, we know that the rail transit fare strategy should be based on the mileage of the shortest path with few number of subway transfers.

(2) The path with more number of subway transfers is the shortest and the travel time is not much different from the path with few number of subway transfers

In this situation, the shortest valid path requires more number of subway transfers. Moreover, the travel time of the path with more number of subway transfers is similar to the path with few number of subway transfers. For example, there are

two valid paths between Liuliqiao station and Shangdi station (see3.2.3). Therefore, when the path with more number of subway transfers is the shortest and the travel time is not much different the path with few number of subway transfers, the rail transit fare strategy should be based on the mileage of the path with few number of subway transfers rather than the mileage of the shortest path with more number of subway transfers.

(3) The path with more number of subway transfers is the shortest and the travel time is much different from the path with few number of subway transfers

In this situation, the shortest valid path requires more number of subway transfers. And even though the path has more number of subway transfers, its total travel time is significantly less than that of the path with few number of subway transfers. For example, there are two valid paths between Anzhenmen station and Jiugong station. The first valid path which is the shortest (excluding the subway transfer distance) consists of taking line 10 from Anzhenmen station to HuixinxijieNankou station, and then taking line 5 from HuixinxijieNankou station to Songjiazhuang station, finally, taking Yizhuang line from Songjiazhuang station to the destination at Jiugong station. Another valid path consists of taking line 10 from Anzhenmen station to Songjiazhuang station, and then taking Yizhuang line at Songjiazhuang station to the destination at Jiugong station. The travel time on the two valid paths is about 55 minute and 61 minutes, respectively. According to the statistics of the travel time distribution between O-D stations, the travel time distribution ratio in the range [53], [57] is about 90%. In other words, the first valid path is the passenger's first choice. In this situation, we know that the rail transit fare strategy should be based on the mileage of the shortest path with more number of subway transfers.

3. The Originations and the Destinations are not on the Same Line: same number of subway transfers

(1) The travel time difference of the valid path between O-D stations is very small

In this situation, the travel time difference is not significant. For example, there are two valid paths between Xizhimen station and Wangfujing station (see 3.2.2). We can concluded that the fare strategy should be determined according to the mileage of the first path that passengers choose more or half of the mileage of the two paths.

(2) The travel time difference of valid path between O-D stations is relatively large.

In this situation, the travel time difference is relatively large. For example, there are two valid paths between Huilongguan station and YonghegongLama Temple station. The first valid path consists of taking line 13 from Huilongguan station to Dongzhimen station, and then taking line 2 from Dongzhimen station to the destination at YonghegongLama Temple station. Another valid path consists of taking line 13 from Huilongguan station to Lishuiqiao station and then taking line 5 from Lishuiqiao station to the destination at Yonghegong Lama Temple station. According to the statistics

of the travel time distribution between O-D stations, the travel time distribution ratio in the range [36], [39] is about 80%. Therefore, it can be determined that relatively more passengers choose the path of line 13 and then taking line 5 as the best choice. In other words, the fare strategy should be determined according to the mileage of the paths with significantly less travel time.

Based on the above analysis, we can draw the following conclusions:

When the origination and the destination are on the same line, and there is a shortest path without subway transfer to achieve the minimum travel time, almost all passengers will choose the path without subway transfer. In this situation, the fare strategy should be determined according to the mileage of the paths without subway transfer. When the shortest path requires subway transfer, there are two situations as follows: the first is that the travel time of the non-transfer path is not different from that of the subway transfer path, at this point, most passengers will choose the path without subway transfer. In this situation, the fare strategy should be determined according to the mileage of the paths without subway transfer. The second is that the travel time of the non-transfer path differs greatly from that of the subway transfer path, and most passengers choose the path that requires subway transfer. In this situation, the fare strategy should be determined according to the mileage of the paths with subway transfer.

When the origination and the destination are not on the same line, and the valid path requires different number of subway transfers. If the shortest path is the one of the least number of subway transfers, almost all passengers will choose this one. In this situation, the fare strategy should be determined according to the mileage of the path with fewer subway transfers. When the shortest path requires more number of subway transfers, there are two situations as follows: the first is that the travel time of the path with few number of subway transfers is not different from that of the path with more number of subway transfers, at this point, most passengers will choose the path with few number of subway transfers. In this situation, the fare strategy should be determined according to the mileage of the paths with few number of subway transfers. The second is that the travel time of the path with few number of subway transfers differs greatly from that of the path with more number of subway transfers, and most passengers choose the path with more number of subway transfers. In this situation, the fare strategy should be determined according to the mileage of the paths with more number of subway transfers.

When the origination and the destination are not on the same line, and the valid path requires same number of subway transfers. If the travel time difference of all valid paths is not large, and the difference of passengers' choice of each valid path is small, the fare strategy should be determined according to the mileage for which the passenger chooses more or equal mileage for each valid path. Meanwhile, if the travel time of all valid paths is significantly different, the fare

strategy should be determined according to the mileage of the paths for which the passenger chooses more.

VII. CONCLUSION

In the paper, the shortest path between an origination and a destination is firstly obtained through simulation technology, then, the travel time between any O and D is obtained. Thirdly, we used big data analysis technology to obtain the actual travel time between the O and D stations. By comparing the actual travel time with the travel time of the shortest path, we can conclude that the best path chosen by most passengers is not necessarily the shortest path, thus, the irrationality of the fare principle based on the mileage of the shortest path can be verified.

There are still some shortcomings in this paper, such as the ability to expand the amount of data sample analysis. In the future research, more data and simulation technologies will be used to solve other problems in rail transit, and these technologies will also be used to solve problems in other industries in the society.

REFERENCES

- [1] D. Zhang, "High-speed train control system big data analysis based on the fuzzy RDF model and uncertain reasoning," *Int. J. Comput. Commun. Control*, vol. 12, no. 4, pp. 577–591, 2017.
- [2] M. Bagchi and P. R. White, "The potential of public transport smart card data," *Transp. Policy*, vol. 12, no. 5, pp. 464–474, Sep. 2005.
- [3] L. Kang, J. Wu, H. Sun, X. Zhu, and Z. Gao, "A case study on the coordination of last trains for the Beijing subway network," *Transp. Res. B, Methodol.*, vol. 72, pp. 112–127, Feb. 2015.
- [4] L. J. Basso and S. R. Jara-Díaz, "Integrating congestion pricing, transit subsidies and mode choice," *Transp. Res. A, Policy Pract.*, vol. 46, no. 6, pp. 890–900, Jul. 2012.
- [5] A. Tirachini, D. A. Hensher, and J. M. Rose, "Multimodal pricing and optimal design of urban public transport: The interplay between traffic congestion and bus crowding," *Transp. Res. B, Methodol.*, vol. 61, pp. 33–54, Mar. 2014.
- [6] D. Gong, S. Liu, J. Liu, and L. Ren, "Who benefits from online financing? A sharing economy E-tailing platform perspective," *Int. J. Prod. Econ.*, vol. 222, Apr. 2020, Art. no. 107490, doi: 10.1016/j.ijpe.2019.09.011.
- [7] L. L. Ding, Z. W. Wang, and L. Y. Li, "Model of urban road congestion pricing based on game theory," *Appl. Mech. Mater.*, vols. 253–255, pp. 1895–1899, Dec. 2012.
- [8] H. Yang and Y. Tang, "Managing rail transit peak-hour congestion with a fare-reward scheme," *Transp. Res. B, Methodol.*, vol. 110, pp. 122–136, Apr. 2018.
- [9] B. De Borger, C. Courcelle, and D. Swysen, "Optimal pricing of transport externalities in an international environment: Some empirical results based on a numerical optimization model," *Regional Sci. Urban Econ.*, vol. 34, no. 2, pp. 163–201, Mar. 2004.
- [10] K. Gkritza, M. G. Karlaftis, and F. L. Mannering, "Estimating multimodal transit ridership with a varying fare structure," *Transp. Res. A, Policy Pract.*, vol. 45, no. 2, pp. 148–160, Feb. 2011.
- [11] Y.-T. Zhu, F.-Z. Wang, X.-Y. Lv, and Y. Pan, "Dynamic pricing for railway tickets with demand-shifted passenger groups," in *Proc. Int. Conf. Manage. Sci. Eng., 21th Annu. Conf. Proc.*, Aug. 2014, pp. 256–262.
- [12] D. Huang, Z. Liu, P. Liu, and J. Chen, "Optimal transit fare and service frequency of a nonlinear origin-destination based fare structure," *Transp. Res. E, Logistics Transp. Rev.*, vol. 96, pp. 1–19, Dec. 2016.
- [13] Z. Xueyu and Y. Jiaqi, "Research on the bi-level programming model for ticket fare pricing of urban rail transit based on particle swarm optimization algorithm," *Procedia Social Behav. Sci.*, vol. 96, pp. 633–642, Nov. 2013.
- [14] H. Gong and W. Jin, "Analysis of urban public transit pricing adjustment program evaluation based on trilateral game," *Procedia Social Behav. Sci.*, vol. 138, pp. 332–339, Jul. 2014.

- [15] P. Zhao and Y. Zhang, "The effects of metro fare increase on transport equity: New evidence from Beijing," *Transp. Policy*, vol. 74, pp. 73–83, Feb. 2019.
- [16] M.-P. Pelletier, M. Trépanier, and C. Morency, "Smart card data use in public transit: A literature review," *Transp. Res. C, Emerg. Technol.*, vol. 19, no. 4, pp. 557–568, Aug. 2011.
- [17] X. Ma, Y.-J. Wu, Y. Wang, F. Chen, and J. Liu, "Mining smart card data for transit riders' travel patterns," *Transp. Res. C, Emerg. Technol.*, vol. 36, pp. 1–12, Nov. 2013.
- [18] R. H. Xu, J. M. Ye, and H. C. Pan, "Method for early warning of heavy passenger flow at transfer station of urban rail transit network under train delay," *China Railway Sci.*, vol. 35, no. 5, pp. 127–133, 2014.
- [19] T. Liu and R. H. Xu, "Disposal of large passenger flow at transfer station based on coordinated headways adjustment," *Urban Mass Transit*, vol. 17, pp. 50–53, 2014.
- [20] F. Zhang, J. Zhao, C. Tian, C. Xu, X. Liu, and L. Rao, "Spatiotemporal segmentation of metro trips using smart card data," *IEEE Trans. Veh. Technol.*, vol. 65, no. 3, pp. 1137–1149, Mar. 2016.
- [21] Y. Lu, H. Wu, X. Liu, and P. Chen, "TourSense: A framework for tourist identification and analytics using transport data," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 12, pp. 2407–2422, Dec. 2019.
- [22] S. Dabiri, C.-T. Lu, K. Heaslip, and C. K. Reddy, "Semi-supervised deep learning approach for transportation mode identification using GPS trajectory data," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 5, pp. 1010–1023, May 2020.
- [23] S. Tao, D. Rohde, and J. Corcoran, "Examining the spatial-temporal dynamics of bus passenger travel behaviour using smart card data and the flow-comap," *J. Transp. Geography*, vol. 41, pp. 21–36, Dec. 2014.
- [24] J. Wang, X. Chen, Z. Chen, and L. Mao, "Cluster algorithm based on LDA model for public transport passengers' trip purpose identification in specific area," in *Proc. IEEE Int. Conf. Intell. Transp. Eng. (ICITE)*, Aug. 2016, pp. 186–192.
- [25] D. Gong, S. Liu, M. Tang, L. Ren, J. Liu, and X. Liu, "Revenue sharing or profit sharing? An Internet production perspective," *Adv. Prod. Eng. Manage.*, vol. 13, no. 1, pp. 81–92, Mar. 2018.
- [26] D. Gong, M. Tang, S. Liu, G. Xue, and L. Wang, "Achieving sustainable transport through resource scheduling: A case study for electric vehicle charging stations," *Adv. Prod. Eng. Manage.*, vol. 14, no. 1, pp. 65–79, Mar. 2019.
- [27] X. Chu, J. Liu, D. Gong, and R. Wang, "Preserving location privacy in spatial crowdsourcing under quality control," *IEEE Access*, vol. 7, pp. 155851–155859, 2019.
- [28] D. Gong, M. Tang, B. Buchmeister, and H. Zhang, "Solving location problem for electric vehicle charging Stations—A sharing charging model," *IEEE Access*, vol. 7, pp. 138391–138402, 2019.



SHIFENG LIU received the Ph.D. degree in industry economics from Beijing Jiaotong University, Beijing, China. He is currently a Professor of computer science with Beijing Jiaotong University. His current researches involve logistics management and information management.



GUANGMEI CAO received the B.S. degree in engineering management from Zhengzhou University, Zhengzhou, China, and the master's degree in the field of industrial engineering from Zhengzhou University, Zhengzhou, where he is currently pursuing the Ph.D. degree in management science.

His research interests include big data and project management.



PENGFEE ZHAO received the M.S. degree in industrial engineering from Beijing Jiaotong University, Beijing, China. His current research is logistics management.



JIANHAI ZHANG received the B.S. degree in professional traffic engineering from the Liaoning University of Science and Technology. He currently committed to research the planning, design, investment, construction, and operation management of utility tunnel.



HONGHU GAO received the B.Eng. degree from the Zhengzhou University of Light Industry, Zhengzhou, China, and the master's degree in the field of industrial engineering from Zhengzhou University, Zhengzhou, where he is currently pursuing the Ph.D. degree in management science.

His research interests include intelligent transportation and big data.



PENG ZHANG received the B.S. degree in information and computing science from Beijing Information Science and Technology University. He is currently pursuing the M.S. degree in economics with the Renmin University of China. His current research interests include computer graphics, artificial intelligence, and block chain.

• • •