PAPER • OPEN ACCESS

Public Transport Passengers' Classification and Path Choice Characteristics Analysis by Using C&RT Model in Beijing

To cite this article: Jingjing Wang *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **688** 044018

Public Transport Passengers' Classification and Path Choice Characteristics Analysis by Using C&RT Model in Beijing

Jingjing Wang [1], Yanyan Chen [2*], Lizeng Mao[1], Rui Yang[1], Dongyue Li[1], and Biao Chen [2]

[1] Transportation Operations Coordination Center, Beijing, 100071, China

[2] Beijing Key Laboratory of Traffic Engineering, Beijing University of Technology, Beijing, 100020, China

[*]Corresponding author's e-mail: cdyan@bjut.edu.cn

Abstract. The mining of IC card data for commuters' classification and analysis of bus route choice plays an important role for public transport passengers' choice behavior analysis, and formulating a scientific and reasonable traffic planning strategy. The authors choose some indicators for passengers' classification in Beijing and use classification and regression tree model to build the classifier. Three parameters: departure time, travel distance and travel frequency, as the classification parameters, are input into the model and the results show 45 classification passengers; Among them, strong commuting passengers accounted for 7.25%, weak commuting passengers accounted for 30.27%, and accidental commuting passengers accounted for 62.48. This paper also analyzes the characteristics of the public transport route choice for all classification groups from four aspects: the overall characteristics, transfer characteristics, travel characteristics and subway travel characteristics. The conclusions of this paper can provide a theoretical basis for the analysis of passenger flow composition of multi-model public transport and the establishment of multi-mode bus route selection model.

1. Introduction

With the construction of transit information system for middle and large-sized cities, IC cards of transit are widely used. Extracting the behavior characteristics of travelers by mining IC card data provides important basic data support to trade management and operations for traffic management departments, bus route planning and adjustment for transportation planning departments. At present, IC cards mainly used for the calculation of loading on and off stops [1] - [3], bus passenger flow scale [4] and OD estimation bus trips [5] - [7]. For the travel characteristics of bus travelers, reference [8] used the occupations, income levels and trip purposes as the basis of population classification, and established multiple logit models for path choice. Reference [9] proposed a bus commuters identification method based on Naïve Bayesian Classifier (NBC). Reference [10] defined three types of travel according to the time and space characteristics of travelers participating in different activities: commuter trip, general trip and random trip. The standard deviation of travel frequency and departure time is used as a standard for bus travelers. The laws and demands of bus travelers can be explored through the classification study.

In addition, people have different psychological and decision-making behaviors while choosing public transportation, and different results are reflected in the transportation mode and path choice. For megacities, including subways and ground buses, the multi-modal combination travel paths are diverse. The choice psychology and behavioral decisions of people are different in choosing routes. This paper uses C&RT (Classification and Regression Trees) classification and regression tree model to choose indicators and classifiers for Beijing public transportation traveler classification, and analyzes the path choice characteristics of traveler classification results.

2. Decision tree and classification parameters

C&RT classification and regression tree is a type of data mining algorithm. The model uses a binary tree to divide the prediction space into several subsets. The leaf nodes in the tree correspond to the index parameters used to divide the region. The partition range is determined by the relevant branching rules (Spitting Rules) of each node. The objective and prediction range can be either range fields or categorical fields; all segments are binary segments (ie, split into two groups), using Gini Index. Each prediction sample moves from the root to the leaf node and finally vests a unique leaf node.

Many indicators, including the socio-economic attributes and the trip purpose of individuals, are set for classification, which not containing the personal and socio-economic attributes of passengers, such as age, gender, income, etc. In this paper, bus IC card data and subway AFC data are used to classify, and the following three indicators are selected as description parameters of psychological characteristics: (1) Departure time. The distribution of passengers' daily transit time in Beijing presents a saddle-shaped curve with obvious early and late peak characteristics, and public transportation has obvious congestion and punctuality rates in different periods. Therefore, this paper considers the departure time as one of the important psychological parameters for classification and decision-making of path choice. (2) Origin and destination. The origin and destination are the second important parameters affecting the path choice of travelers. The distance between the origin and the destination and the convenience of the bus and subway in the vicinity of the origin and destination greatly affect the result of the path choice for travelers, this paper identifies that this parameter is an important psychological parameter for the classification and path choice of travelers. (3) Trip purpose: The trips for work and leisure and entertainment produce different results of path choice through experience. The trip purpose affects the focus of the travelers while choosing paths. The difference of the focus eventually results in different path choices. Therefore, the trip purpose is also to be used as an important psychological parameter for the classification and path choice of travelers.

3. Decision tree and classification parameters

Since the public transportation charges are double-swipe in Beijing, recording the data of loading on and off. Reference [8] describes the various types of data of the bus and subway, the processing methods and results of each type of data in the travel chain. The calculation process of the three parameters is as follows:

3.1. Departure time

The departure time field is available through extracting the OD data of the public transportation travel chain using SQL. The characteristics of the departure time of the public transportation on working day in Beijing are shown in Fig.1, which has obvious early and late peak characteristics. The peak time (7:00-9:00) accounts for 24.69%, and the peak time (17:00-19:00) accounts for 21.46% of the whole trips in a day.

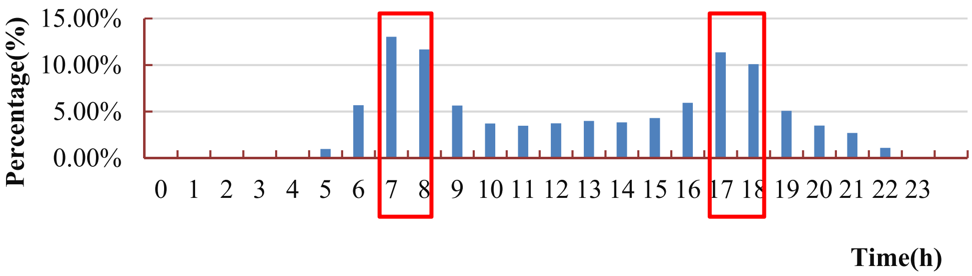

Figure.1 Departure time distribution

3.2. *Origin and destination*

The passenger's origin, reflecting in the IC card data, is the starting line and loading stop, the terminal line and off stop, whereas the information cannot be directly quantified as a number for analysis. Therefore, in this paper, the travel distance is used to replace the index of origin and destination, and a certain treatment method is adopted to make the travel distance represent the actual travel distance, thereby quantifying the index of origin and destination. The travel distance of public transportation on working day in Beijing is shown in Fig.2. The average travel distance is 13.4 km, the travel distance below 10 km accounts for 46.94%, and nearly half of the trips are concentrated below 10 km.

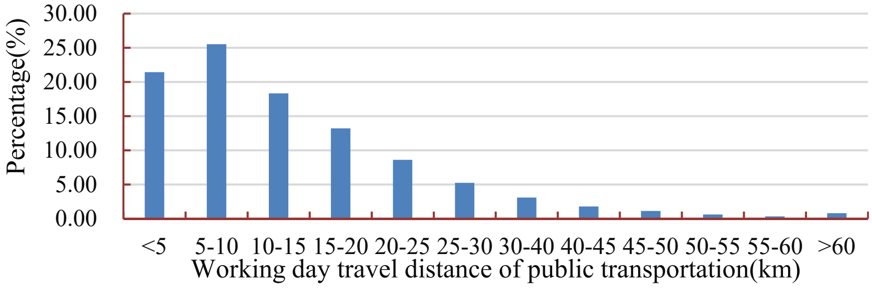

Figure.2 Characteristic of travel distance

3.3. *Trip purpose*

The trip purpose is also difficult to obtain from the IC card data. We translate this indicator into the frequency from the same origin stops to destination stops within the same time period within 5 working days. Different trip frequencies can reflect the travelers' trip purpose in a certain extent. The passenger flow with higher travel frequency during the early peak hours may be commuter flow. The passenger flow with lower travel frequency during the early peak hours is likely to be more freedom.

The trip frequency characteristics of public transportation in Beijing on the working day are shown in Fig. 3. The average trip frequency is 0.37, which means the frequency of repeated trips on the same path on the working day is nearly 2 times, and the frequency of 0.2 (the repetition trip frequency on the working day is 1) accounts for 62.1%, which means that the proportion of travelers who choose one path for working is the largest. In other words, there is not only one path chosen by the traveler for traveling.

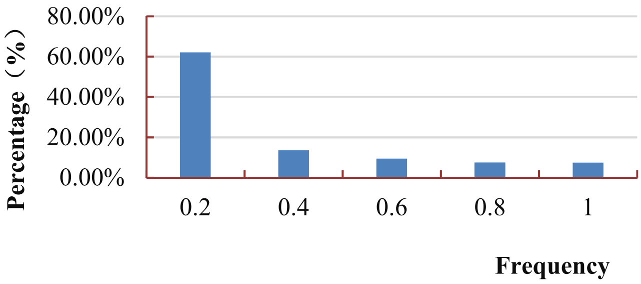

Figure.3 Characteristic of travel frequency

3.4. *Correlation test*

While inputting the classification parameters into the classification model training, this paper conducts a further correlation test on the three parameters of the classification, in order to guarantee the mutual independence of the parameters. The correlation coefficient r presents the correlation degree between the variables. The correlation coefficient ranges from -1 to +1. While describing the degree of correlation between variables, the degree of correlation can be divided into the following cases:

(1) If $|r| \geq 0.8$, it is regarded as high correlation;
(2) If $0.5 \leq |r| < 0.8$, it is regarded as moderate correlation;
(3) If $0.3 \leq |r| < 0.5$, it is regarded as low correlation;
(4) If $|r| < 0.3$, the correlation between variables is extremely low and can be regarded as irrelevant.

Many types of statistics are used to indicate the degree of correlation. We use Pearson correlation coefficient. The corrcoef function in Matlab is used to obtain the correlation matrix between variables. As shown in Table 1.

Table.1 Correlation matrix

Variate	Departure time	Distance	Frequency
Departure time	1.000	-0.1026	-0.412
Distance	-0.1026	1.000	0.1812
Frequency	-0.412	0.1812	1.000

The correlation coefficient between the two variables is low through comparison in the matrix, and no both variables whose correlation coefficient is close to 1. Therefore, the correlation between the three variables is very low, and no merge processing is needed. Three variables can participate in the classification model as separate variables.

4. Decision tree classification process

In this paper, the C&RT module in SPSS Clementine 12.0 software is used for the decision tree classification. The maximum tree depth is designed to be 10 layers in the classifier design process. The first layer is mainly classified by ON_TIME field, and the subsequent layers are finally classified into 27 categories using distance DIS, travel frequency PINLV. The departure time ON_TIME is divided into one segment according to each hour; the travel frequency PINLV segmentation threshold is 0.3 and 0.9, representing once in five days and five trips in five days respectively; the distance DIS division threshold is more, including 3km, 5km, 9km, 11km, 11km, 21km.

In the process of tuning the classification results, the main problem is to reset the threshold of each parameter. The final adjustment result is: 1. set the departure time ON_TIME threshold value as 6, 10, 16, 20 time points, divide the time period into a flat peak, early peak, flat peak, late peak, flat peak; 2. set the trip distance DIS threshold as 5, 21. The distance is divided into short distance trip, medium distance trip, and long distance trip three types; 3. set trip frequency PINLV threshold is 0.3, 0.9. The trip frequency is divided into accidental trip, weak commute trip, strong commute trip.

According to the above three parameter threshold settings, the three types of parameter thresholds are recombined, and public transportation travelers are classified into 45 types. The classification identifiers and physical meanings of each type of traveler are shown in Table 2. Among them, strong commuter trip accounts for 7.25%, weak commuter trip accounts for 30.27%, and accidental trip accounts for 62.48%. It also shows that there is not only one path for travelers, and accidental travel exceeds half. In the flat peak period, the proportion of travelers in the medium distance and the travelers in the late peaks reached more than 10%, which are the previous two travelers.

Table.2 Physical significance and proportion of all types of people

Departure time threshold	Distance threshold	Frequency threshold	Identifier	Physical significance	Proportion
<6	<5	<0.3	<6|<5|<0.3	short distance accidental trip before 6:00	0.10%
<6	<5	0.3-0.9	<6|<5|0.3-0.9	short distance weak commute trip before 6:00	0.03%
<6	<5	>0.9	<6|<5|>0.9	short distance strong commute trip before 6:00	0.00%
<6	5-21	<0.3	<6|5-21|<0.3	medium distance accidental trip before 6:00	0.32%
<6	5-21	0.3-0.9	<6|5-21|0.3-0.9	medium distance weak commute trip before 6:00	0.20%
<6	5-21	>0.9	<6|5-21|>0.9	medium distance strong commute trip before 6:00	0.04%
<6	>21	<0.3	<6|>21|<0.3	long distance accidental trip before 6:00	0.24%
<6	>21	0.3-0.9	<6|>21|0.3-0.9	long distance weak commute trip before 6:00	0.12%
<6	>21	>0.9	<6|>21|>0.9	long distance strong commute trip before 6:00	0.02%
6-10	<5	<0.3	6-10|<5|<0.3	early peak short distance accidental trip	4.33%
6-10	<5	0.3-0.9	6-10|<5|0.3-0.9	early peak short distance weak commute trip	2.56%
6-10	<5	>0.9	6-10|<5|>0.9	early peak short distance strong commute trip	0.36%
6-10	5-21	<0.3	6-10|5-21|<0.3	early peak medium distance accidental trip	9.24%
6-10	5-21	0.3-0.9	6-10|5-21|0.3-0.9	early peak medium distance weak commute trip	8.72%
6-10	5-21	>0.9	6-10|5-21|>0.9	early peak medium distance strong commute trip	3.31%
6-10	>21	<0.3	6-10|>21|<0.3	early peak long distance accidental trip	3.43%
6-10	>21	0.3-0.9	6-10|>21|0.3-0.9	early peak long distance weak commute trip	2.74%
6-10	>21	>0.9	6-10|>21|>0.9	early peak long distance strong commute trip	1.22%
10-16	<5	<0.3	10-16|<5|<0.3	flat peak short distance accidental trip	5.53%
10-16	<5	0.3-0.9	10-16|<5|0.3-0.9	flat peak short distance weak commute trip	0.67%
10-16	<5	>0.9	10-16|<5|>0.9	flat peak short distance strong commute trip	0.02%
10-16	5-21	<0.3	10-16|5-21|<0.3	flat peak medium distance accidental trip	11.04%
10-16	5-21	0.3-0.9	10-16|5-21|0.3-0.9	flat peak medium distance weak commute trip	1.63%
10-16	5-21	>0.9	10-16|5-21|>0.9	flat peak medium distance strong commute trip	0.08%
10-16	>21	<0.3	10-16|>21|<0.3	flat peak long distance accidental trip	3.79%
10-16	>21	0.3-0.9	10-16|>21|0.3-0.9	flat peak long distance weak commute trip	0.39%
10-16	>21	>0.9	10-16|>21|>0.9	flat peak long distance strong commute trip	0.02%
16-20	<5	<0.3	16-20|<5|<0.3	late peak short distance accidental trip	5.40%
16-20	<5	0.3-0.9	16-20|<5|0.3-0.9	late peak short distance weak commute trip	1.86%
16-20	<5	>0.9	16-20|<5|>0.9	late peak short distance strong commute trip	0.13%
16-20	5-21	<0.3	16-20|5-21|<0.3	late peak medium distance accidental trip	10.64%
16-20	5-21	0.3-0.9	16-20|5-21|0.3-0.9	late peak medium distance weak commute trip	7.07%
16-20	5-21	>0.9	16-20|5-21|>0.9	late peak medium distance strong commute trip	1.30%
16-20	>21	<0.3	16-20|>21|<0.3	late peak long distance accidental trip	3.15%
16-20	>21	0.3-0.9	16-20|>21|0.3-0.9	late peak long distance weak commute trip	2.30%
16-20	>21	>0.9	16-20|>21|>0.9	late peak long distance strong commute trip	0.58%
>20	<5	<0.3	>20|<5|<0.3	short distance accidental trip after 20:00	1.47%
>20	<5	0.3-0.9	>20|<5|0.3-0.9	short distance weak commute trip after 20:00	0.37%
>20	<5	>0.9	>20|<5|>0.9	short distance strong commute trip after 20:00	0.02%
>20	5-21	<0.3	>20|5-21|<0.3	medium distance accidental trip after 20:00	3.07%
>20	5-21	0.3-0.9	>20|5-21|0.3-0.9	medium distance weak commute trip after 20:00	1.30%
>20	5-21	>0.9	>20|5-21|>0.9	medium distance strong commute trip after 20:00	0.11%
>20	>21	<0.3	>20|>21|<0.3	long distance accidental trip after 20:00	0.75%
>20	>21	0.3-0.9	>20|>21|0.3-0.9	long distance weak commute trip after 20:00	0.30%
>20	>21	>0.9	>20|>21|>0.9	long distance strong commute trip after 20:00	0.03%

5. Characteristics of selection path characteristics of various groups of people

For the classification results of the above 45 categories of people, from the overall path characteristics (travel expenses, travel time), transfer characteristics (transfer times, transfer time), bus travel characteristics (bus load rate, bus waiting time), subway travel characteristics (the ratio of subway travel, subway station congestion) four pieces of content analysis of various types of people.

5.1. Overall path characteristics

5.1.1. Travel expenses

The travel cost indicator characterizes the cost of the travel after using the public transportation paths. The process of calculation needs to separate the subway from the regular bus. The specific charging rules are as follows: (1) Subway fare rules: 3 yuan within 6 kilometers (including); 4 yuan for 6 kilometers to 12 kilometers (including); 5 yuan for 12 kilometers to 22 kilometers (including); 6 yuan for 22 kilometers to 32 kilometers (including); each additional 1 yuan can take 20 kilometers for 32 kilometers or more. (2) Bus fare rules: 2 yuan within 10 kilometers (including), each additional 1 yuan can take 5 kilometers for 10 kilometers or more. Using the municipal traffic card to take the bus can enjoy a discount of 50%. The impact of travelers' cumulative consumption on travel expenses in a month is not considered. In order to facilitate the representation in the graph, we use the abbreviated expression to represent the physical meaning of the passenger flow (such as the early peak long distance strong commuter flow, expressed as "early/long/strong"), the specific travel cost characteristics are shown in Fig. 4. The average travel cost of 45 categories of people is 3.4 yuan. The average travel expenses of travelers in each period are relatively close, all of which are 3-4 yuan. The average travel expenses are relatively close for each type of trip frequency. The average strong commuter flow is 3.8 yuan, average weak commuter flow is 3.3 yuan, and the average accidental flow is 3.1 yuan; the average long-distance travels have the highest average cost, and the short-distance travels have the lowest average cost, which is nearly three times less. It shows that the travel expenses increase with the increase of the travel distance, no significant correlation with the travel time, and no obvious correlation with the travel frequency.

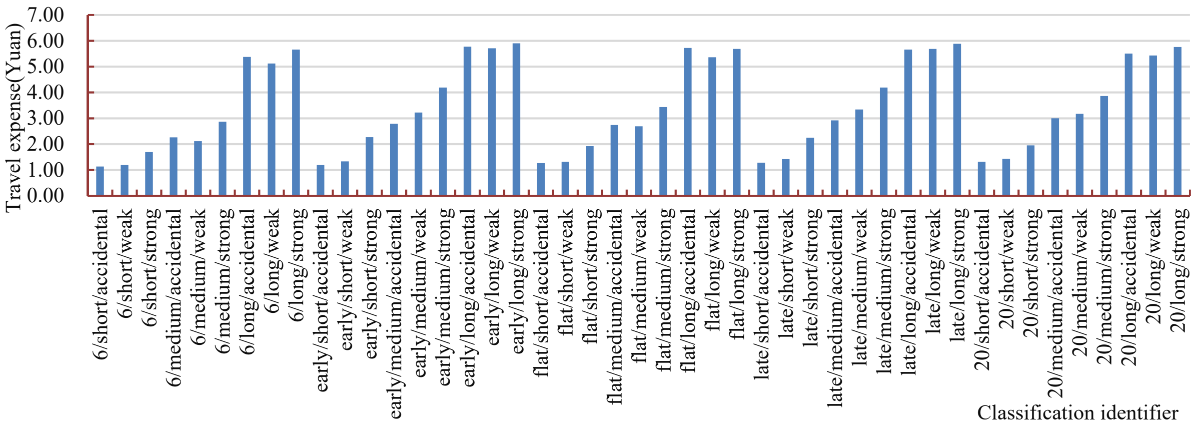

Figure.4 Characteristic of travel cost

5.1.2. Travel time

The travel time represents the total time with the public transportation paths, which determined by the arrival time of the travel chain OD data subtracting from the departure time in internal public transportation system. The travel time and its characteristics of various groups of people are shown in Fig.5. The average travel time of 45 groups people is 40 minutes, and the travel time increases with the increase of travel distance. The average travel time of long-distance travel is 3.6 times higher than that of short-distance travel. The average accidental travel time is 45 minutes, the average weak commuter travel time is 36 minutes, the average strong commuter travel time is 40 minutes, and the frequency affects the travel distance; the average duration of each travel time is also significantly different. Results express that the travel time have a certain correlation with the travel frequency. In the case of the same travel distance, the travel time with high travel frequency is smaller than the travel with lower frequency obviously.

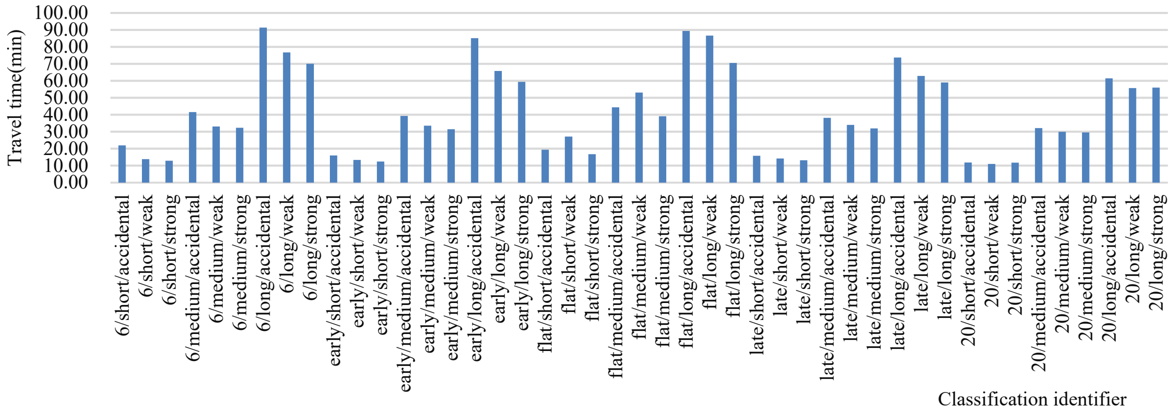

Figure.5 Characteristic of travel time

5.2. Transfer characteristics

5.2.1. Transfer times

The transfer times indicator characterizes the number of times the passengers need to transfer during the process of selecting the public transportation route. The transfer data of the passenger flow of the travel chain can be obtained according to the sum of the travel codes. The characteristics of the transfer times of various types of people are shown in Fig. 6. It can be seen that the number of transfer times is not related to the departure time, has a certain correlation with the travel distance, and has the greatest correlation with the travel frequency. It can be clearly seen that in the case of the same departure time travel distance, the passenger flow with a larger frequency of travel tends to select a path with a smaller number of transfer times.

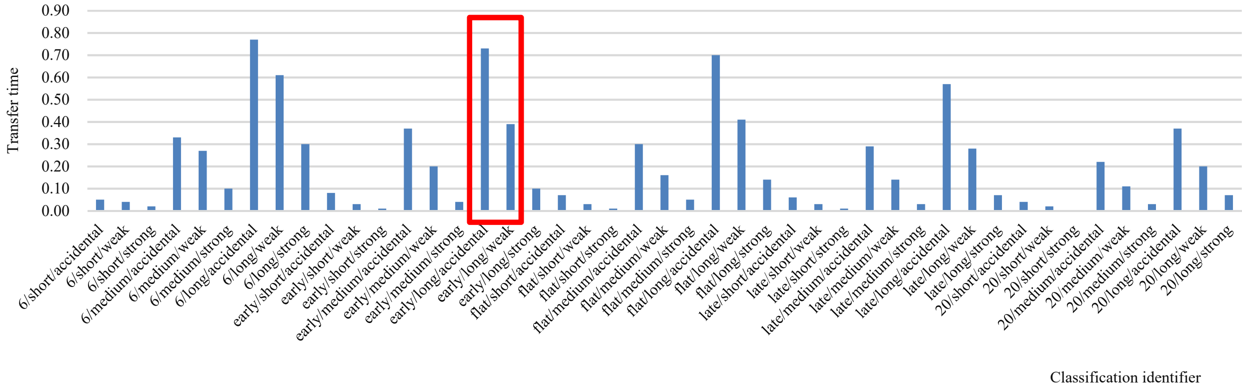

Figure.6 Characteristic of transfer time

5.2.2. Transfer time

The transfer time indicates the total transfer time required by the traveler choosing path. The transfer data of the passenger flow in the travel chain can be determined according to the sum of travel code and statistics. The transfer time characteristics of various groups of people are shown in Fig.7. The average transfer time is 3 minutes, the average transfer time for accidental travel is 4 minutes, the average transfer time for weak commute travel is 2 minutes, and the average travel time for strong commute travel is 0.6 minutes. The average transfer time for long-distance travel is 4 minutes. The average transfer time of different periods is relatively close. The passenger flow of transfer time during 10:00 to 16:00 is higher than that of other periods. Passenger flows with a frequency above 0.9 tend to choose a path with a smaller transfer time. Therefore, there is a strong correlation between the travel frequency and the transfer time, the distance.

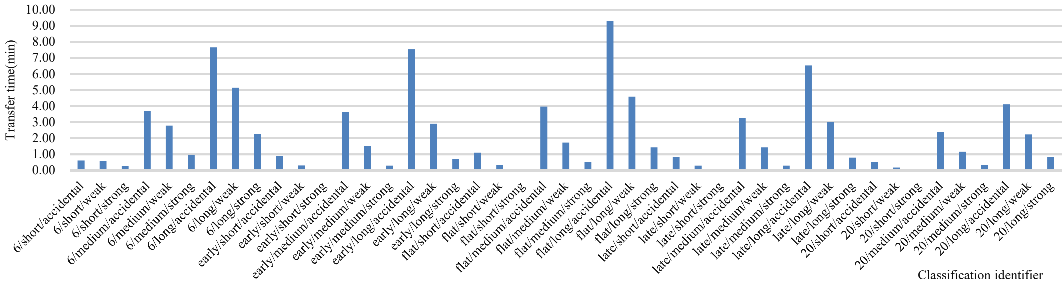

Figure.7 Characteristic of transfer time

5.3. Bus travel characteristics

5.3.1. Bus loading rate

The bus loading rate represents the average loading rate of all buses in the public transportation paths chosen by the travelers, and its value indicates the crowdedness of the bus when the bus arrives at the bus stops. The loading rate can be determined by section passenger flow dividing rated capacity (single carriage vehicles and articulated vehicles, 80 loads for single carriage vehicles, 120 loads for articulated vehicles). The loading rate characteristics of various types of bus travelers are shown in Fig.8. The average loading rate is 0.76, the average loading rate of the early peak is above 1 which is the highest, and the average loading rate of the late peak is 0.95, which is significantly higher than other time periods except early peak; the average loading rate difference of the three types of travel distance is small; the average loading rate of strong commuter travel is slightly higher than that of weak commuter travel and accidental travel. It shows that the passenger flow with high frequency has a high acceptability for the loading rate of the bus, and the average loading rate of the bus is highly correlated with the departure time.

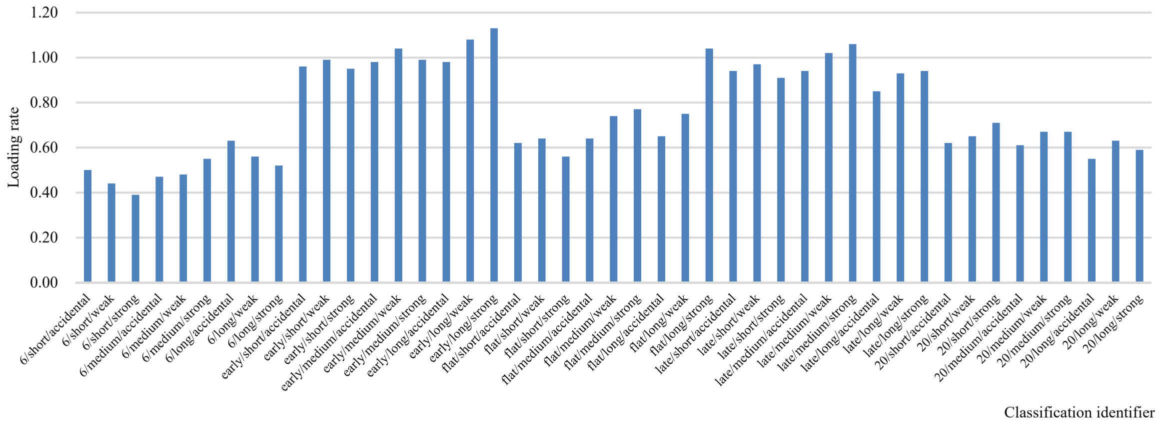

Figure.8 Characteristic of bus load rate

5.3.2. Waiting time

The waiting time characterizes the waiting time of all the travelers taking buses. For public transportation with a fixed frequency, the maximum waiting time of the traveler is headway of each line. The average waiting time of the travelers is set as half of headway. The waiting time characteristics are shown in Fig.9. The average waiting time is 3.15 minutes, the minimum waiting time for the travelers at the early peak is the lowest, followed by the late peak, and the maximum waiting time is emerged after 20:00; the long-distance travel time is the smallest, but no obvious difference compared with the other two types of travel with different distances. It can be explained that the waiting time and the departure time have greater correlation, which is closer to the frequency regularity of the whole day. Besides, the long-distance, strong commute passenger flow has the lowest acceptable level for waiting time, and is likely to choose a bus with a shorter waiting time.

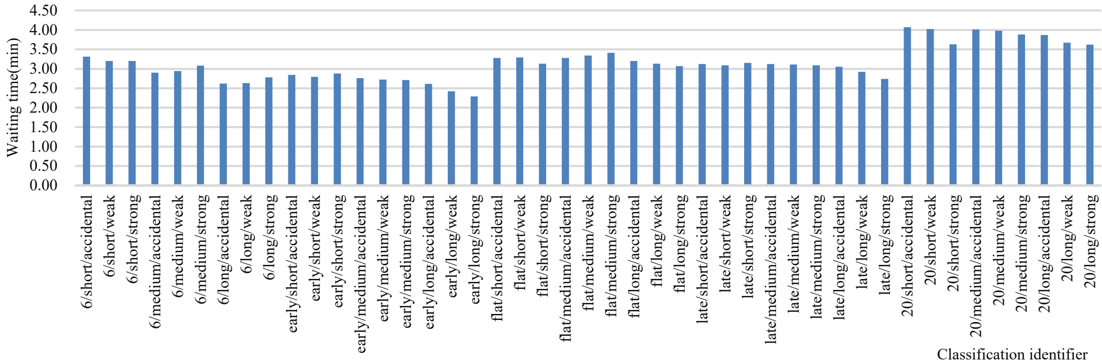

Figure.9 Characteristic of waiting time for bus

5.4. Subway travel characteristics

5.4.1. The proportion of subway travel

The subway travel ratio indicates the proportion of subway in the paths chosen by travelers, which can be determined by counting the proportion of subway trips in the travel chain OD data. The subway trip characteristics of various classification identifiers are shown in Fig.10. The average subway travel ratio of all types of people accounts for 46.33%, the proportion of subway trips in strong commute travel reached 63%, the weak commute travel is 40.7%, and the accidental travel is 35%. The long-distance travel for the subway trips accounts for 69.7%. The medium-distance travel accounts for 47.7%, and the short-distance travel is 21.5%; the difference in the proportion of subway trips for each type of departure time is small. The travel frequency and travel distance are more relevant to whether to choose subway travel, and the tendency choosing subway travel for high-frequency or long-distance travelers is greater.

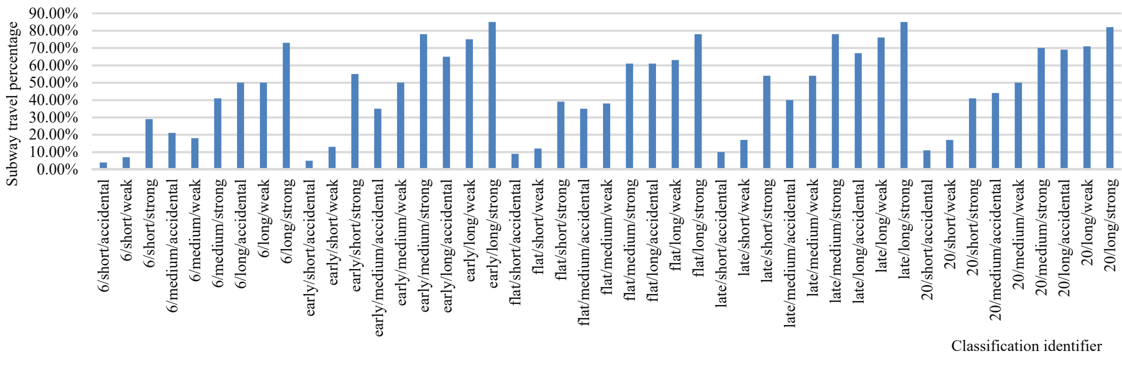

Figure.10 Characteristic of subway travel proportion

5.4.2. Congestion degree of the subway station

The congestion degree of the subway station indicates the congestion degree of the boarding station for the subway travels, which can be obtained by processing the OD data of the subway. The formula is as follows:

$$C_{i,k} = \frac{Q_{k,p}}{Q_{k,avg}}$$

where *C*_ (*i*,*k*) = congestion degree of subway station *i* at period *k* (%); *Q*_(*i*,*k*) = the number of travelers entering the subway station i at period *k* (s); *Q*_(*i*, *avg*) = The average number of entering subway station *i* of all the periods in the whole day.

The 24-hour period is segmented into a 15-minute period, totaling 96 segments per day, and the total number of travelers entering the station is counted in each 15-minute period. The congestion degree of the subway station is characterized by the total number of travelers in the corresponding period dividing the average number of travelers for 15 minutes. The congestion characteristics of the subway stations of various types of travelers are shown in Fig.11. The average congestion degree of all types of travelers is 1.47, and the congestion degree of strong commuter travel is up to 1.6, but the difference of congestion degree between the other two types of travels is small; the congestion degree of long-distance travel is up to 1.58. For the other two types of travels, the difference on congestion degree is small; the early peak is the highest in congestion degree, followed by the late peak, which is much higher than the average congestion degree. It can be seen that the congestion degree has the highest correlation with the departure time, and is associated with the travel frequency and travel distance

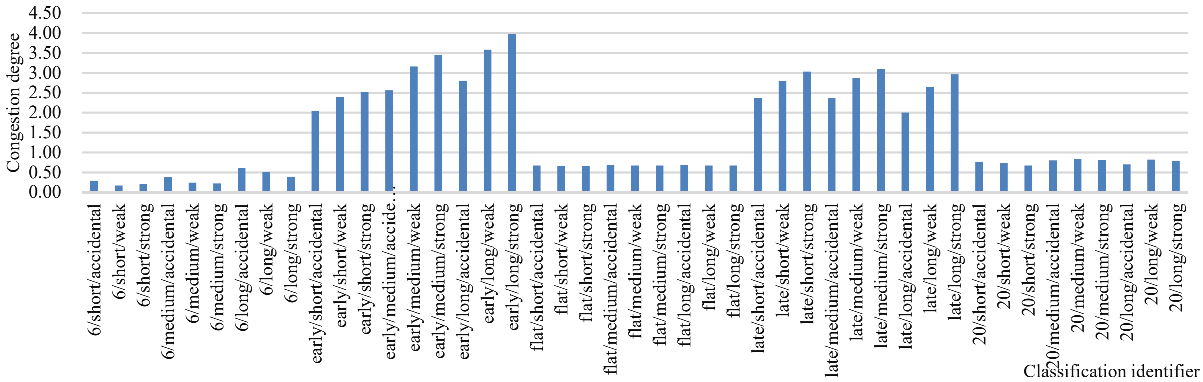

Figure.11 Characteristic of subway station congestion

6. Conclusions

This study uses the bus IC card data to classify the bus travelers, making full use of the characteristics of the IC card data sample, and the ability to carry out multi-mode combined path analysis. Furthermore, the global features, transfer features, bus travel features, and subway travel features are analyzed in the paper. Some conclusions are summarized as follows:

(1) Travel expenses have a strong correlation with travel distance; travelers with high travel frequency is more affordable for bus full loading rate, and they are inclined to paths with less waiting time, less transfer, and less travel time. Travel paths with subway participation are obvious in long-distance travel especially in the early peak.

(2) Travelers with a long distance travel tend to choose the path of the subway participation and the lower waiting time. No obvious unacceptability is presented for the full loading rate, the congestion degree of the subway station and the transfer times.

(3) The travelers in the early and late peaks tend to choose a path with subway participation, low travel time, and less waiting time. The acceptability of full loading rate, the congestion degree of the subway station is higher.

The research results and related conclusions of this paper can provide a theoretical basis for the choices of influencing factors in the process of multi-model bus traveler analysis and the establishment of bus path choice model.

References

[1] Xu W, Deng C, and Liu B. (2013) Approach on Public Traffic Passenger Flow Statistics Based On IC Data[J]. China Journal of Highway and transport, 26(5): 158-163.

[2] Ma X, Liu C, Liu J, CHEN Feng, and YU Hai-yang. (2015) Boarding Stop Inference Based on Transit IC Card Data[J]. Journal of Transportation Systems Engineering and Information Technology, 15(4): 78-84.

[3] Chen J, Lu Y, and Cui M. (2018) Estimating alighting stops of smart card public transportation passengers based on travel patterns[J]. J.Xi'an Univ. of Arch.&Tech.(Natural Science Edition), 50(1): 23-29.
[4] Li L, Jiang Y, Wang J, and Wu B. (2016) Passenger Volume Estimation Based on Data Fusion[J]. Urban Transport of China, 14(1): 43-50.
[5] Zhao J. (2007) Estimating a rail passenger trip origin destination matrix using automatic data collection systems[J]. Computer-Aided Civil and Infrastructure Engineering. 22(1):376-387.
[6] Chen J, Yang D. (2013) Estimating smart card commuter's origin-destination distribution based on APTS data[J]. Journal of Transportation Systems Engineering and Information Technology, 13(4): 47-53.
[7] Munizaga M.A.，Palma C. (2012) Estimation of a disaggregate multimodal public transport origin-destination matrix from passive smartcard data from Santiago，Chile[J]． Transportation Research Part C: Emerging Technologies，24: 9-18.
[8] Di D, Yang D. (2016) A Passenger-Classification Transportation Assignment Model for Urban Public Traffic Corridor[J]. Journal of TongJi University (Natural Science), 44(2): 235-242.
[9] Sun S, Yang D. (2015) Identification of Transit Commuters Based on Naïve Bayesian Classifier[J]. Journal of Transportation Systems Engineering and Information Technology, 15(6): 46-53.
[10] Li J, Deng H. (2016) Classification of Passenger's Travel Behavior Based on IC Card Data[J]. Journal of Chong Qing Jiao Tong University (Natural Scicenc), 35(6): 109-114.