WILEY | Hindawi

Research Article

Identification and Classification of Bus and Subway Passenger Travel Patterns in Beijing Using Transit Smart Card Data

Lewen Wang [iD],[1] **Yuan Chen,**[2] **Yu Wang,**[3] **Xiaofei Sun,**[3] **Yizheng Wu** [iD],[1] **Fei Peng** [iD],[1] **and Guohua Song** [iD][1]

[1]*Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport,*
Beijing Jiaotong University, Haidian District, Beijing 100044, China
[2]*Shanghai University of International Business and Economics, Shanghai, China*
[3]*CCCC Highway Consultants Co., Ltd., Dongcheng District, Beijing 100088, China*

Correspondence should be addressed to Yizheng Wu; wuyizheng@bjtu.edu.cn

Urban public transit has been rapidly developed in recent years. However, given increases in travel volume, the level of service still needs to be improved to meet the satisfaction of passengers. Transit service providers and researchers have focused on improving transit devices, but the service level of public transit has not yet been effectively improved, so more and more research is interested in analyzing the travel patterns of passengers. Compared with traditional survey methods, smart card collection systems—which can collect spatial-temporal information about passengers' trips—are convenient for the study of bus and subway passengers' travel patterns. However, the data provided by smart cards have not yet been fully explored. Therefore, this paper proposed a multistep methodology to gather information on the travel patterns of bus and subway passengers in Beijing, China. We conducted statistical analyses and used an unsupervised clustering method to study and classify passengers based on travel patterns. Four groups have been identified: standard commuters, flexible commuters, and two types of low-frequency passengers. Then, a comprehensive analysis was conducted. We also discussed the changes of passengers' travel time consumption before and after the implementation of customized bus for high-frequency passengers. The analyses indicated that passengers' travel patterns can provide useful information for transit service providers and can help improve the level of service of urban public transit by promoting the promulgation of local public transport policies and the implementation of customized services.

1. Introduction

Given increases in traffic congestion, the use of public transit is becoming more and more popular in many cities. The development of urban public transit infrastructure has also attracted more people to use public transit, which may help to reduce pollutant emissions from motor vehicles [1], accelerate population mobility, alleviate traffic congestion, and improve residents' living standards. With the advantages of convenience, affordability, and accessibility, more and more people are willing to choose public transit [2].

However, with the increased use of public transit and the demand for personalized and diversified travel, the public transit service is facing more challenges and several

problems have gradually emerged. According to the Beijing Transport Annual Report, the total passenger flow in the public transit network decreased by 13.69% between 2014 and 2018, with a precipitous decline of almost 38% in 2020 compared to 2018 because of COVID-19. A similar trend was observed in the United States, suggesting that public transit has become less attractive and is facing greater challenges [3]. To address declines in bus passenger flow, Nishiuchi et al. [4] analyzed temporal and spatial changes in passengers' travel behaviors in a month and revealed that there are differences in the travel patterns of different passengers. They pointed out that more attention should be paid to the travel patterns of passengers. Transit agencies have been studying ridership patterns to identify key travel

factors such as ridership preferences and spatial and temporal patterns [5]. By better understanding ridership patterns, transit agencies can identify gaps in their current services and better adjust their operating strategies and propose more relevant customized policies to improve public transit services and encourage more usage.

The advent of automated fare collection (AFC) systems has provided great convenience, allowing for accurate and extensive recordings of each passenger's travel transactions. Such data can be used for the extraction of passengers' boarding station [6], identification of commuters [7], and analysis of passengers' spatial-temporal dynamics [8, 9], among other possibilities. Therefore, to better explore the travel regularity of passengers, identify their travel preferences, and spatiotemporality patterns for proposing operational measures, we conducted a study using passenger transaction data, established a method for dividing high- and low-frequency passengers, identified commuters, analyzed the patterns of different group of passengers, proposed suggested measures, and briefly analyzed customized bus before and after implementation.

2. Literature Review

AFC systems are a valuable resource for studying passenger travel patterns in large cities. In recent years, using smart card data to analyze travel patterns has become a common technique employed by researchers.

Since the frequency of passenger usage can affect the development and profitability of urban public transit, many studies have focused on identifying high-frequency passengers for better understanding their travel patterns to improve the services of public transit since they are the main contributors. The identification of high- and low-frequency riders is usually based on researchers' definition to make an artificial distinction. Nishiuchi et al. [4] classified the top 40% as low-frequency passengers based on the cumulative curve of trip days, who traveled only 1 or 2 days during the period, while the remaining 60% were high-frequency passengers. Kieu et al. [10] identified frequent passengers by having 53 trips on weekdays for three months, i.e., at least one trip per day during at least 75% of the weekdays. Lathia et al. [11] and El Mahrsi et al. [12] used 2 out of 30 days and 10 out of 30 days as frequency thresholds to distinguish high-frequency passengers from low-frequency passengers, respectively. Liu and Cheng [13] defined high-frequency passengers as those who used their Oyster cards at least 37.5% of days during the four-week study period. While Zhao et al. [14] defined commuters as the 95th percentile ratio of routine weeks among all smart card holders.

In addition to trip frequency, the temporal and spatial indicators of passenger are often used for passenger classification and combined with clustering algorithms. The K-means algorithm is a frequently used algorithm with simple parameters, as the only required input is the number of clusters (K). Ortega-Tong [15] divided London public transport users into four categories based on travel frequency, travel hour, travel time consumption, and travel start and end stations. Zhao et al. [9] analyzed the spatial-temporal aspects

of passengers' travel patterns and clustered metro passengers into four groups using statistical and unsupervised cluster-based methods. Yang et al. [16] classified passengers into four types according to the number of routes and transfers and proposed a method to estimate passenger spatial-temporal trajectory. To better understand long-term patterns in passenger travel, Kaewkluengklom et al. [17] studied changes in the individual travel behavior by using three years of longitudinal smart card data from Shizuoka Prefecture, Japan, and classified passengers using the K-means method. Other clustering methods have also been used in passenger classification. Kieu et al. [18] adopted the DBSCAN algorithm, which can identify clusters of arbitrary shapes based on different parameters, to segment transit users into four groups. They stressed that passenger segmentation helps operators to provide customized information and services for different classes of transit users. Ouyang et al. [19] proposed a trip reconstruction algorithm and mined the travel patterns of smart card users in Beijing using the DBSCAN algorithm. Also, Cui et al. [20] proposed a method to classify users based on weekly boarding frequency using smart card data from Shenzhen collected over four consecutive weeks. Briand et al. [21] proposed a two-level model for passenger classification in Gatineau, Canada. Using Singapore as a model, Zhu [22] developed an activity type classification model by combining smart card data and traditional travel surveys to better understand urban travel demand and activity dynamics.

Despite the valuable insights provided by previous studies, the potential for smart card data to elucidate the travel patterns of passengers has yet to be fully explored. Firstly, identifying high-frequency passengers is usually based only on the travel days or the trips, which is often not comprehensive enough. In addition, the literature is less likely to analyze the travel patterns of low-frequency passengers after identifying high-frequency passengers yet improving the utilization rate of low-frequency passengers is one of the keys to improving service quality.

To build on previous studies, using smart card data to examine passenger patterns, this paper proposed a multistep methodology to mine the travel patterns of bus and subway passengers in Beijing, China. We studied passenger travel patterns by analyzing three main variables: trip frequency, travel time, and travel space. The approach used in this study is shown in Figure 1. First, we used smart card data and geographic information data to extract each passengers' travel indicators. Then, combined with the statistical analysis method, we analyzed passengers' general travel patterns. In addition to using the travel days and trips, we established a method to identify high- and low-frequency passengers by combining weekday and week-based travel frequency. We also used a clustering-based analysis method to group passengers based on their multidimensional travel indicators, then mined their travel patterns in detail. This paper makes the following contributions:

(i) We extracted the trip frequency index, temporal index, and spatial index for each passenger and analyzed the general travel patterns of passengers based on these three indicators.
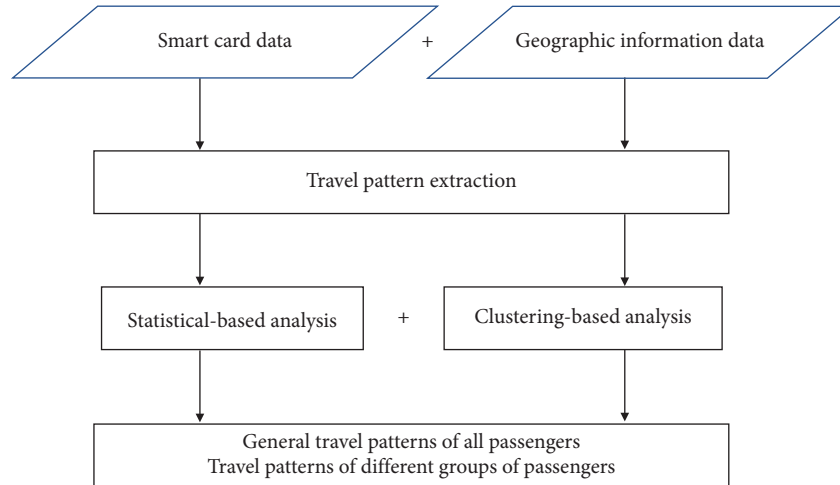
FIGURE 1: Overview of the approach used in this study to examine passenger travel patterns in Beijing.

(ii) We established a method for identifying high- and low-frequency passengers considering multiple trip frequency indicators and classified passengers into different groups using statistical methods and unsupervised clustering methods.

(iii) We conducted a comprehensive analysis of both high-frequency and low-frequency passenger patterns based on trip frequency, travel time, travel space, transfers, and made targeted recommendations and analyzed the effects of the measures before and after implementation, which can provide a basis for improving urban public transit services.

3. Experimental Methods and Materials

3.1. Study Area. The smart card dataset used in this study was collected from Beijing, the capital of China, which is composed of 16 administrative regions. Beijing has a resident population of nearly 21.886 million, and the city contains about 1,200 bus lines and 27 subway lines. Due to the large population and the large daily demand on public transit in Beijing, the number of daily card swipes can reach the tens of millions. Thus, we restricted our data collection and analysis to two large residential areas: Tiantongyuan and Huilongguan, which are both located in the Changping District (Figure 2). The resulting dataset included more than 1.8 million travel records of 70,539 passengers from September 2019.

3.2. Dataset. The data used in this study were preliminarily processed, including station matching and transfer identification. A trip may include multiple subtrips, which is commonly referred to as a transfer; transfers may occur between the same modes of transportation (e.g., bus to bus; subway to subway cannot be recorded because passengers only swipe their cards in and out of the subway station) or between different modes (e.g., subway to bus or bus to subway). We set the transfer time threshold as 30 minutes (this threshold was set based on the 5th Comprehensive

Transport Survey Summary Report in Beijing [23]), meaning that if the time between the end time of the last trip and the start time of the current trip was less than 30 minutes, the two subtrips were considered as one trip. Also, the data fields and their definitions are shown in Table 1.

3.3. Travel Index Extraction and Analysis

3.3.1. Trip Frequency Index Extraction. We extracted five indicators for each passenger: monthly travel days (D_m), monthly trips (T_m), monthly travel days on working days (D_{mw}), monthly trips on working days (T_{mw}), and number of weeks with more than three travel days (C_m).

The distribution of travel days and trips showed a bimodal pattern (Figure 3). The majority of passengers with fewer travel days and trips, but there still existed some passengers who traveled very frequently. It is important to capture the travel patterns of these passengers, which will be explained in later sections.

3.3.2. Temporal Index Extraction. As passengers usually follow a fixed pattern on weekdays, we extracted the average departure hours for the first departure (H_{mw}) and the variance of the first departure time interval (V_{mw}) on weekdays.

From Figure 4, the total amount of trips spiked during peak hours and remained stable during nonpeak hours, which showed a doubled peak on each day, but more pronounced on weekdays. As passengers prefer to travel on weekdays, the trips on weekdays were higher than weekends.

3.3.3. Spatial Index Extraction. We mainly considered the working days with closed travel chains (TC_{mw}) and the average travel distance (Dis_m). Supposing that the origin (O) of the ith trip by public transit on day d is O_{di}, the destination (D) is D_{di}, and the total number of trips in the day is i_{max}, the set of O-D pairs on day d can be expressed as follows:
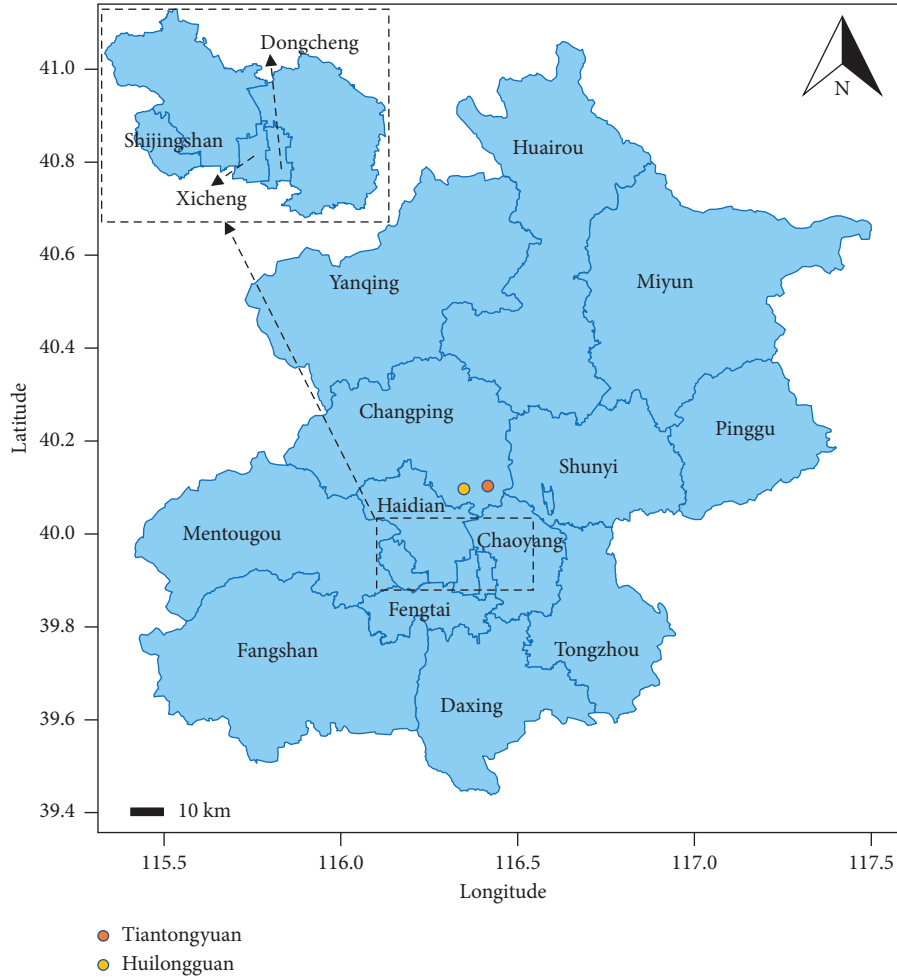
FIGURE 2: Study area.

TABLE 1: Smart card data format.

Field	Definition
CNID	Card number. Each smart card has a different card number.
Trip	Itinerary number. Each trip has a different itinerary number.
SubTrip	Subtrip number. Number of increments in each trip: if the trip does not include a transfer, the column will be 0; if it includes transfer, it will increase from 1.
TrTp	The mode of transportation used in the trip. 1 = bus only. 2 = subway only. 3 = bus and subway.
SubtrTP	The type of transportation mode used in the subtrip. b = bus, s = subway.
T1	The time that passengers swipe their cards at the boarding station.
STATION1	The name of the boarding station.
lon1	The longitude of the boarding station.
lat1	The latitude of the boarding station.
T2	The time that passengers swipe their cards at the alighting station.
STATION2	The name of the alighting station.
lon2	The longitude of the alighting station.
lat2	The latitude of the alighting station.

$$\left\{ O_{d1} - D_{d1}, O_{d2} - D_{d2}, O_{d3} - D_{d3}, \ldots, O_{di_{\max}} - D_{di_{\max}} \right\}. \quad (1)$$

As passengers often go back and forth between O and D, it may be through A_1 and B_1 of line 1 or A_2 and B_2 of line 2, where A_1 and A_2 are close to O and B_1 and B_2 are close to D; the number of lines will vary depending on the case.

$A_1 \longrightarrow B_1$ and $A_2 \longrightarrow B_2$, despite having different starting and ending stations, are both $O \longrightarrow D$ trips. Thus, in our study, if the distance between the first departure station (O_{d1}) and the last arrival station ($D_{di_{\max}}$) in a day is less than 1.2 km [24], it is considered to constitute a closed travel chain (more details can be found in the Supplementary Materials). In examining
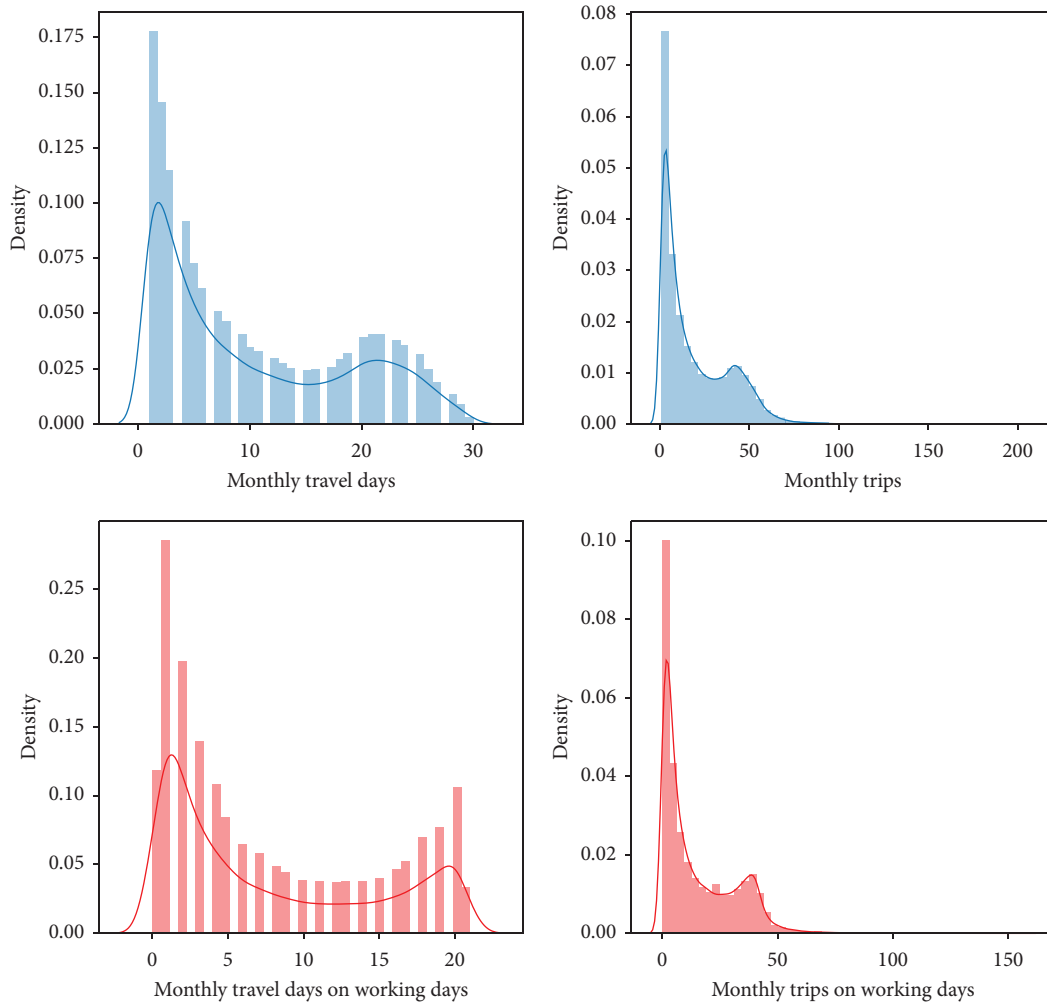
FIGURE 3: Trip frequency of passengers in a month, including monthly travel days, monthly trips, monthly travel days on working days, and monthly trips on working days.

average travel distance, we used the average travel time cost as a proxy for travel distance as has been done in previous studies [8].

Most passengers' travel time cost were about 50 minutes (Figure 5(a)). There was an approximately linear relationship between the number of transfers and the average travel time cost (Figure 5(b)), which suggested that transfers have an important impact on passengers' travel time cost.

There were differences in the distribution of passenger boarding and alighting stations during the morning and evening peak periods. From Figure 6(a), it can be found that passenger boarding stations during the morning peak were mainly concentrated in large residential areas—Tiantongyuan and Huilongguan areas. Figure 6(b) shows that alighting stations were mainly distributed in the central city, where many enterprises and schools are gathered.

Due to the daytime trips, passengers have spread all over the city. Therefore, from Figure 6(c), boarding stations in the evening peak were relatively dispersed. However, the alighting stations were mainly concentrated in large residential areas (Figure 6(d)). The movement of passengers in the morning and evening peaks showed different characteristics, with passengers moving from outside to inside the city in the morning peak, and vice versa in the evening peak. This was because many passengers were commuters who return to their residence after completing their trips during the day. This was also indicated the separation of workplace and residences that existed in Beijing, consistent with the study of Ma et al. [25].

4. Passenger Classification

4.1. Passenger Classification Based on Trip Frequency. From the previous analysis, we knew that passengers vary widely in their trip frequency. Thus, to better analyze and highlight the regularity of habitual passengers, we proposed a set of methods considering multidimensional travel frequency indicators. In this paper, we considered high-frequency passengers to those whose (a) monthly travel days $D_m \geq 14$ and monthly travel trips $T_m \geq 18$, or (b) travel days on weekdays $D_{mw} \geq 10$ and trips on weekdays $T_{mw} \geq 15$, or (c) under the condition of $D_m \geq 10$ and $T_m \geq 15$, the number of weeks with more than three travel days $C_m \geq 3$. After a preliminary analysis (the detail can be seen in the
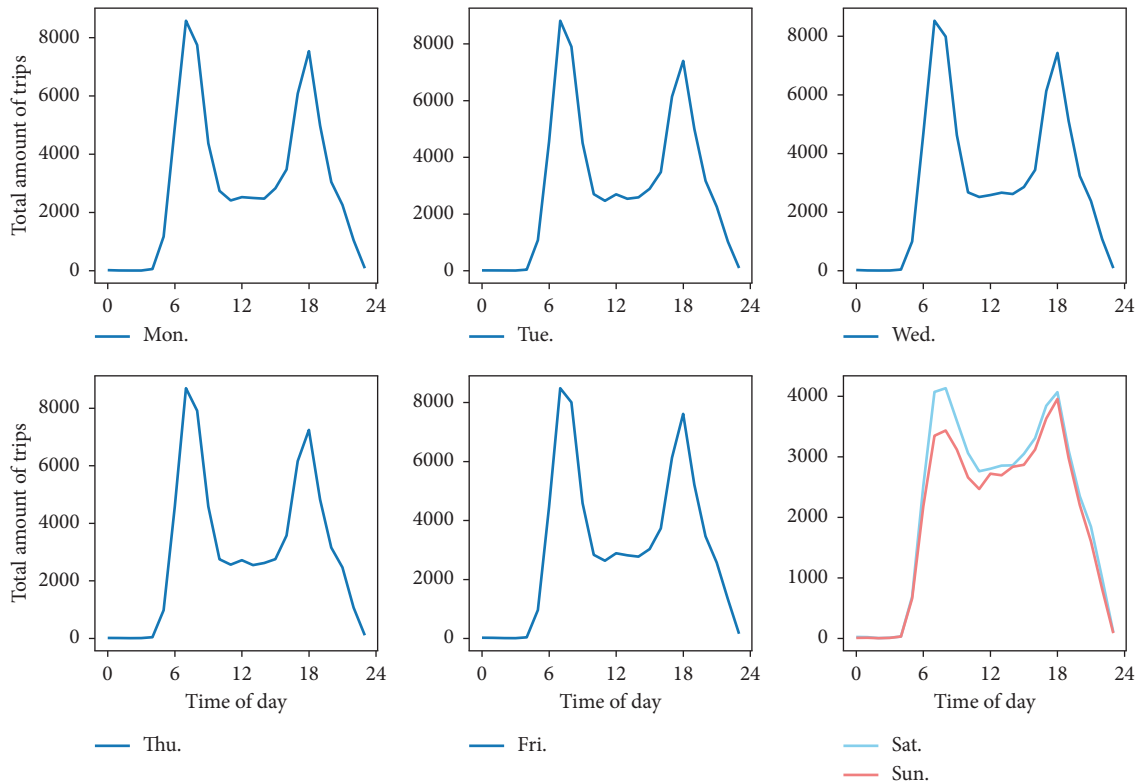
FIGURE 4: Total amount of trips per day of a week.

Supplementary Materials), 23,103 people met condition (a), 23,948 people met condition (b), and 24,530 people met condition (c). Ultimately, 26,100 high-frequency passengers and 44,439 low-frequency passengers were identified.

The classification results are shown in Table 2, and high-frequency riders traveled an average of 20.09 days and 40.51 trips per month, while low-frequency riders traveled 4.31 days and 7.31 trips per month. Although the percentage of high-frequency passengers (37%) was lower than low-frequency passengers (63%), they accounted for 76.5% of all passenger trips.

4.2. Passenger Clustering Based on Multiple Indicators. Since the K-means algorithm only needs to input the classifications K to get the classification labels of all passengers, it is very efficient and suitable for the huge data set of urban public transit. And the passenger classification method using K-means algorithm has been widely used in combination with some evaluation methods [9, 17]. Thus, we used the K-means algorithm combined with the average silhouette coefficients to select the best combination of indicators and the optimal number of clusters from the set of indicators (D_m, T_m, D_{mw}, T_{mw}, H_{mw}, V_{mw}, TC_{mw}, and Dis_m). The average silhouette coefficient is calculated as follows:

$$s_i = \frac{(a_i - b_i)}{\max(a_i, b_i)}, \tag{2}$$

where a_i is the average dissimilarity between the ith passenger and all other passengers within the same cluster and b_i is the lowest average dissimilarity between the ith

passenger and any other cluster. The value of the average silhouette coefficient is contained within (−1, 1); values closer to 1 indicate a better clustering effect.

By clustering high-frequency passengers and low-frequency passengers separately, we found that the clustering effect was optimal when the indicators (H_{mw}, V_{mw}, TC_{mw}, and Dis_m) and (D_m, T_m, D_{mw}, T_{mw}, H_{mw}, V_{mw}, TC_{mw}, and Dis_m) were selected for the high- and low-frequency passenger groups, respectively, and the numbers of clusters were 2.

5. Analysis of Different Groups of Passengers

5.1. Analysis of Travel Frequency. Figure 7 shows that low-frequency passengers traveled less frequently than high-frequency passengers. The first group of high-frequency passengers traveled most frequently; they usually traveled on working days and their trips were twice as frequent as travel days, which meant they often traveled twice a day, likely between residence and workplace, and they were usually commuters. As for transfers, nontransfer trips accounted for the majority (Figure 8). This suggested that most passengers prefer to travel without transferring, indicating urban public operators should design their routes with the goal of minimizing transfers.

5.2. Temporal Analysis. The first departure hour for two group high-frequency passengers was concentrated in the morning peak. However, the wide distribution for two low-frequency groups suggested that they were more flexible in

(a)

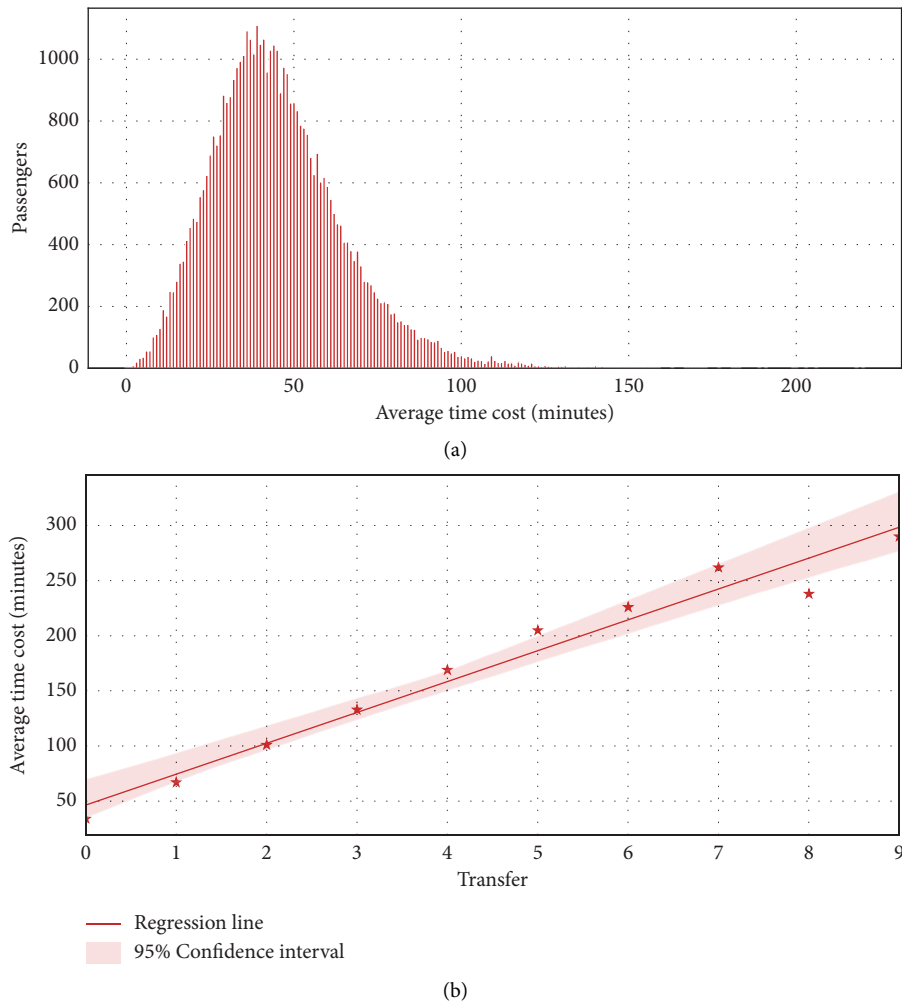Regression line

95% Confidence interval

(b)

FIGURE 5: Travel time cost pattern of all passengers. (a) The average travel time cost for all passengers. (b) Relationship between average travel time cost and transfer.

their travel times (Figure 9). Figure 10 shows high-frequency passengers exhibit a clear bimodal travel pattern on both weekdays and nonweekdays. In contrast, low-frequency passengers did not show a clear pattern and their travel times were relatively scattered. In addition, trips made by high-frequency passengers on weekends decreased to about half of on weekdays, while low-frequency riders remained constant.

According to Figure 11, the median of the first group of high-frequency passengers was the lowest and showed a single-peaked distribution, mostly distributed around 0, implying that this group started their travel plans at a fixed time each weekday. The second group of high-frequency passengers had a more even distribution, but also had a higher percentage of passengers with lower travel time variances. In addition, the median of the first low-frequency passengers was lower, indicating that they maintained stability in travel time for lower frequency trips. The second group showed a uniform distribution. However, there were passengers with large travel time interval variance among low-frequency passengers.

5.3. *Spatial Analysis.* Average travel time cost was similar among the different types of passengers (median around 50 minutes and similar shape of distribution), as they tended to pursue the shortest travel time (by reducing transfers, Figure 12). The first type of high-frequency passengers had the most closed travel chains on weekdays, which meant they were more stable in travel space (Figure 13), and more analysis can be seen in the Supplementary Materials. As analyzed above, these group passengers also stable in the temporality dimension. Therefore, the first group of high-frequency passengers were mostly standard commuters, who traveled to and from the fixed area (residence and workplace) every day during peak hours. The second type of high-frequency passengers were flexible commuters and their last trip was freer. We used *OD* diagrams to visualize the travel records of typical passengers in each group (Figure 14). The *OD* of high-frequency passengers were more clustered into a circle, especially for the first type of high-frequency passengers, while the *OD* of low-frequency passengers were more scattered, which forms multiple circles.
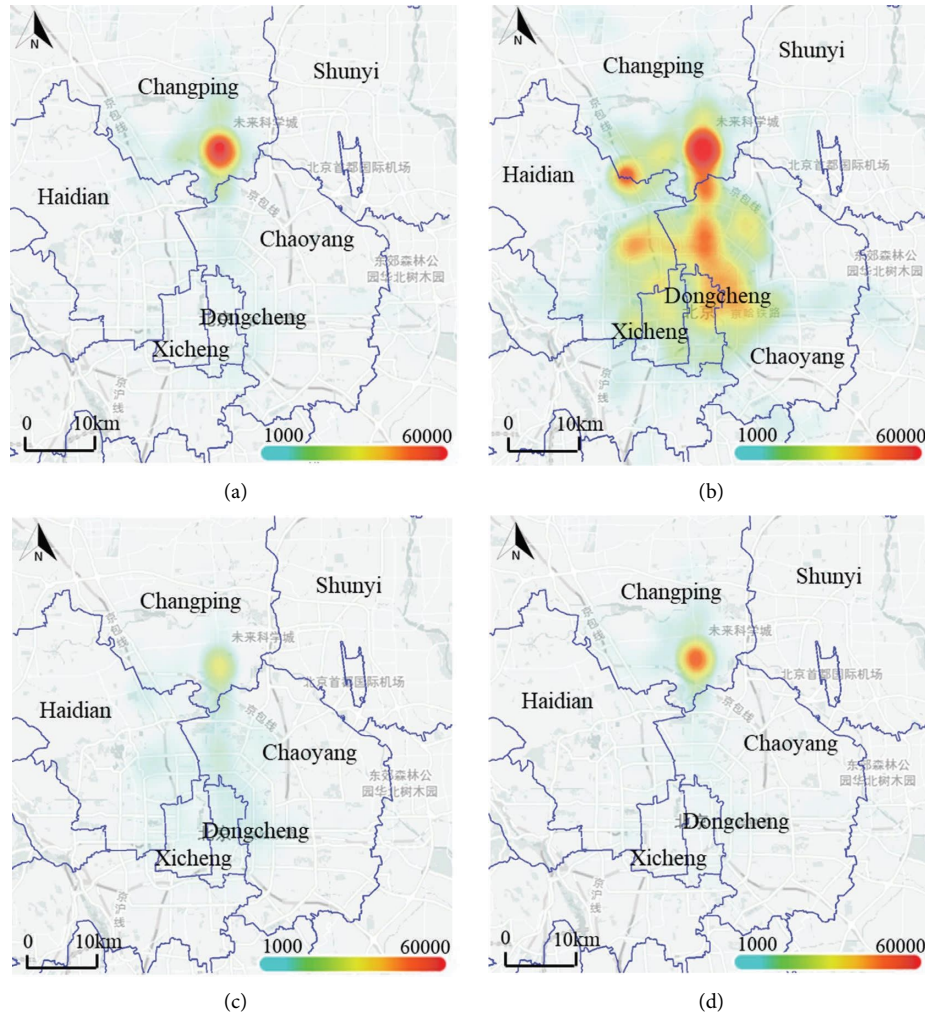
(a)

(b)

(c)

(d)

FIGURE 6: Heat maps for boarding and alighting stations in the morning and evening peaks. (a) Boarding stations in the morning peak. (b) Alighting stations in the morning peak. (c) Boarding stations in the evening peak. (d) Alighting stations in the evening peak.

TABLE 2: Classification results for high-frequency and low-frequency passengers.

Group	Passenger proportion (%)	Trip proportion (%)	Monthly trips		Monthly travel days	
			Average	Standard deviation	Average	Standard deviation
Low-frequency	63	23.50	7.31	14.75	4.31	4.75
High-frequency	37	76.50	40.51	5.89	20.09	3.05

Point *D* in Figure 15 was Tiantongyuan and Hui-longguan area in the Changping District, which is a large residential area in Beijing. We inferred that most of the passengers in the dataset lived in this area. From Figure 15(a), the first group of low-frequency passengers mainly traveled between Tiantongyuan and Huilongguan (point *D*) and Beijing West Station (point *A*), Beijing South Station (point *B*), or Beijing Station (point *C*), and bar chart showed the top 10 *OD* pair trips among 3,543 *OD* pairs. The second group showed a similar pattern, but it was not as pronounced as the first group, with the bar chart showing the top 10 *OD* pair trips among 4532 *OD* pairs. High-frequency passengers have different patterns than low-frequency passengers. They showed more travel activity in the center of the city. For example, they often traveled between Tiantongyuan and Huilongguan areas and Xierqi (Figure 15(b)) where there was a Zhongguancun Software Park with a lot of office buildings and enterprises. And the bar chart showed the top 10 *OD* pair trips among 10,473 and 21,158 *OD* pairs for two group of high-frequency passengers, respectively. This also showed that high-frequency riders travel between residence and workplace in high volumes and travel over a wider range of travel spaces than low-frequency riders.
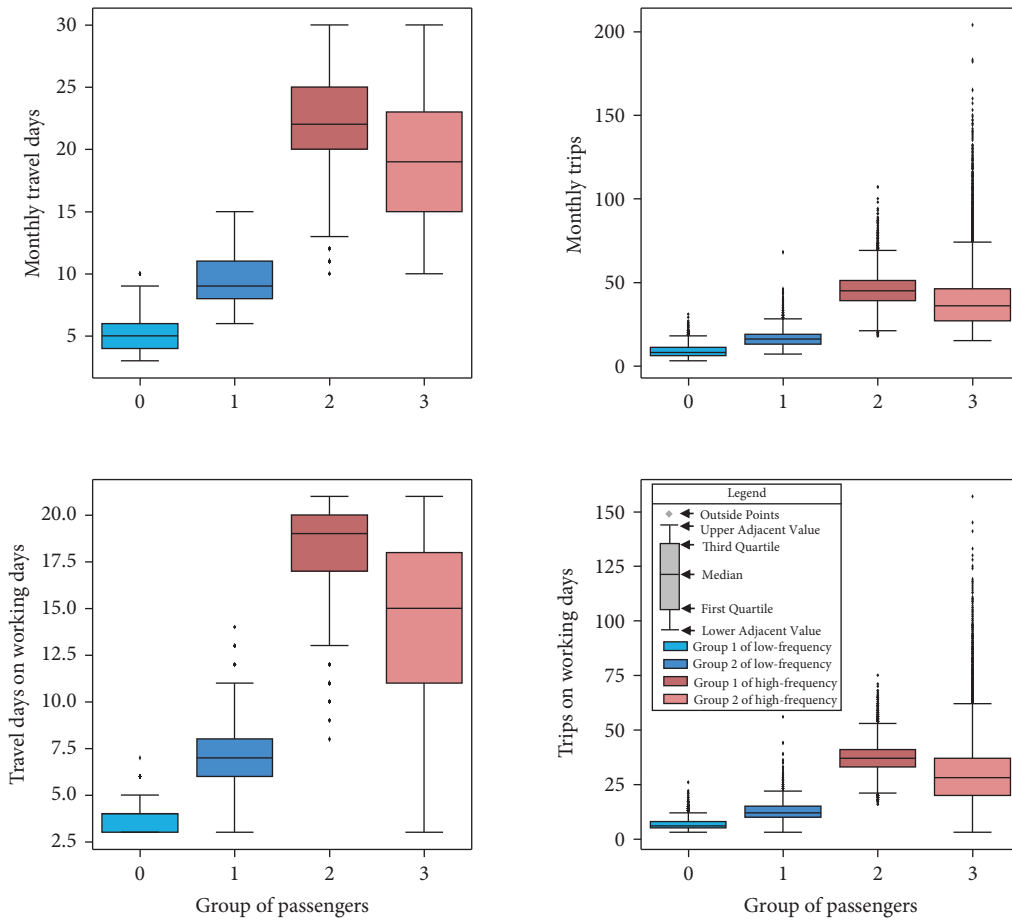
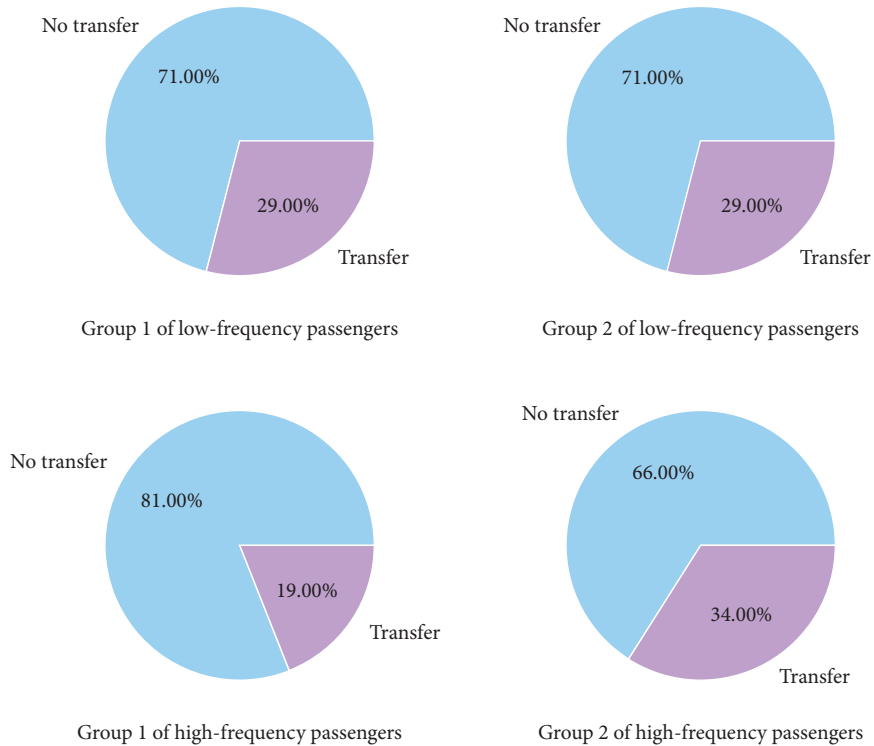Figure 7: Travel frequency for four types of passengers.

Figure 8: The proportion of different groups of passengers with or without transfer trips.
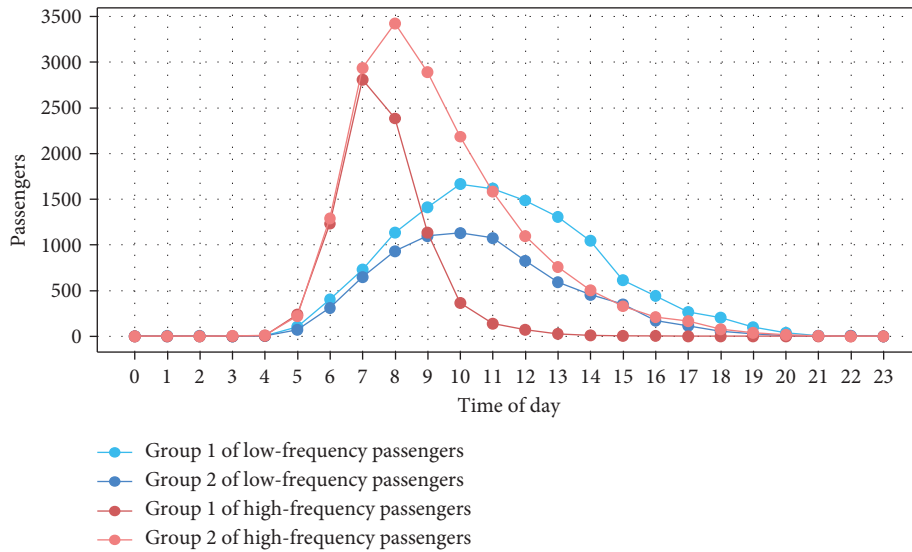
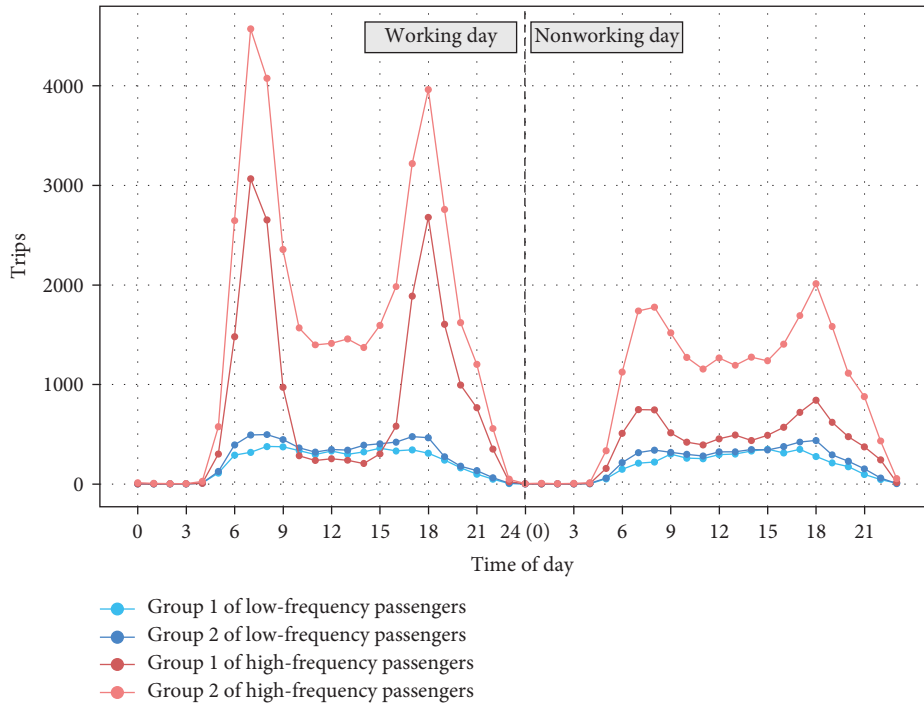FIGURE 9: Average departure hours for four types of passengers.

FIGURE 10: Comparison of travel hours between working days and nonworking days for different groups of passengers.

6. Discussion

The purpose of studying passenger travel characteristics is to identify passengers' travel patterns of different groups of passengers. This is helpful to propose suggestions for improvement, formulate corresponding fare and service policies, and make urban public transit more attractive. Therefore, based on the results of the above analysis, we illustrated how the ridership pattern analysis provided the basis for the relevant customized services.

The spatial pattern of high-frequency commuters showed that they travel more between large residential areas and typical workplaces, so customized buses can be considered, and the departure time and frequency can be set based on the identified travel patterns of commuters. Figure 15(b) shows that some commuters travel frequently between Tiantongyuan and Xierqi, so we used the example of setting up the customized bus from Tiantongyuan to Xierqi for the discussion of individualized measures.
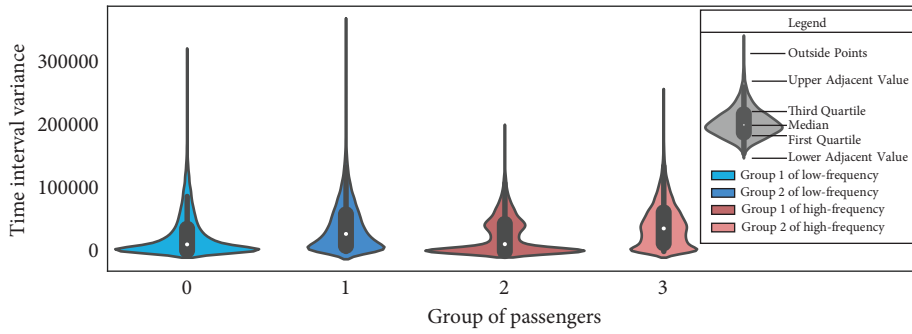
FIGURE 11: Variance in the first departure time interval for various passengers.
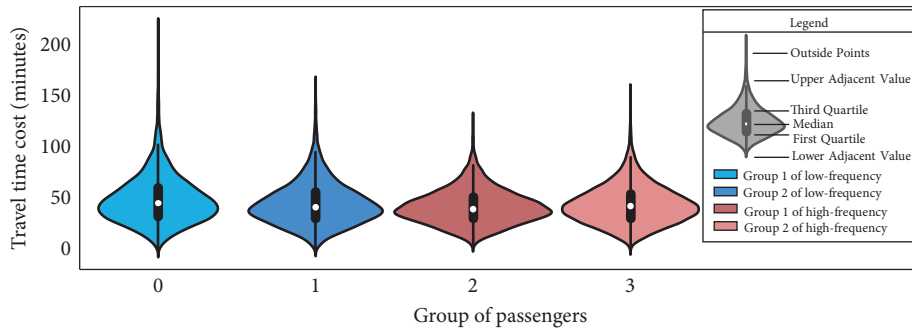

FIGURE 12: Distribution of average travel time for four groups of passengers.
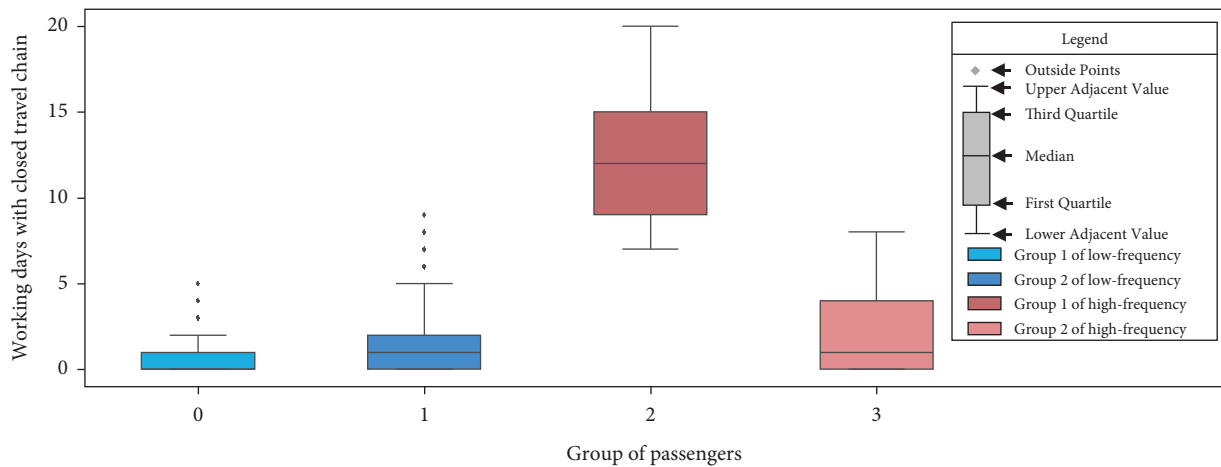

FIGURE 13: Closed trip chain distribution for four groups of passengers.

Figure 16(a) shows the average travel time consumption of high-frequency passengers who traveled between Tiantongyuan and Xierqi. Most passengers traveling in the 30–40 minute range, with an average travel time cost of 39.21 minutes. Customized bus is designed primarily for buses to provide a higher level of service, faster and more comfortable for passengers [26]. According to the real-time prediction function of Google Maps, the minimum time cost from Tiantongyuan to Xi'erqi in the morning peak is 30 minutes. Therefore, we assumed that the travel time consumption of the customized bus is 30 minutes. Figure 16(b) shows the change of time cost per passenger after choosing customized bus, with the majority of passengers experiencing a significant reduction, for a total reduction of 2,550 minutes for all passengers.
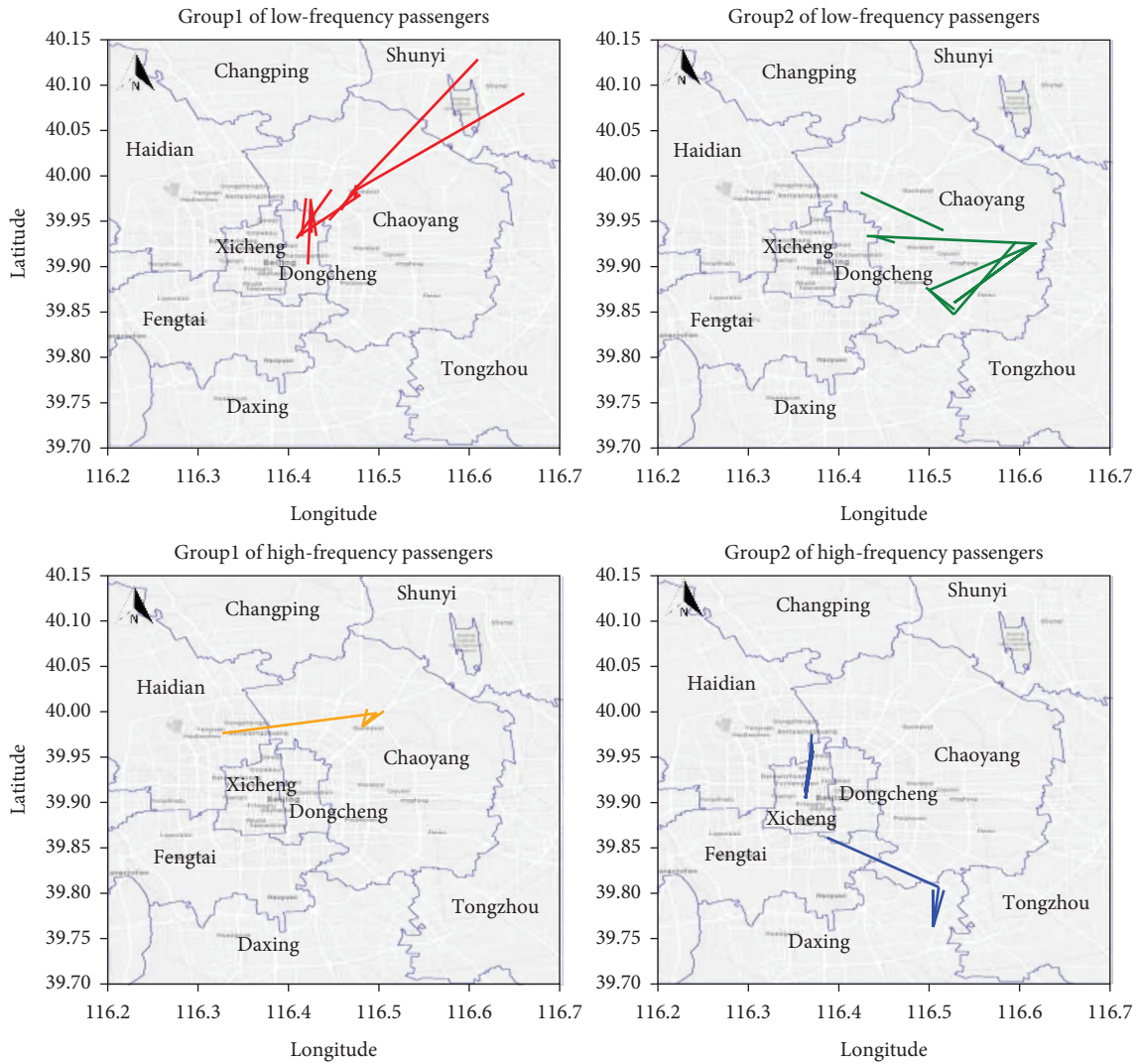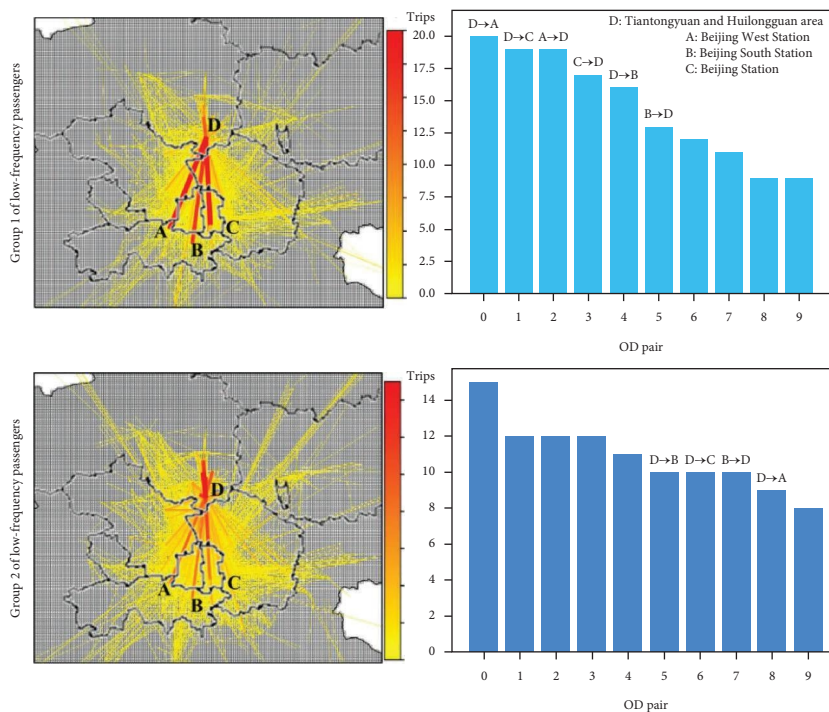
FIGURE 14: *OD* diagram for a typical passenger selected from each group in a month.
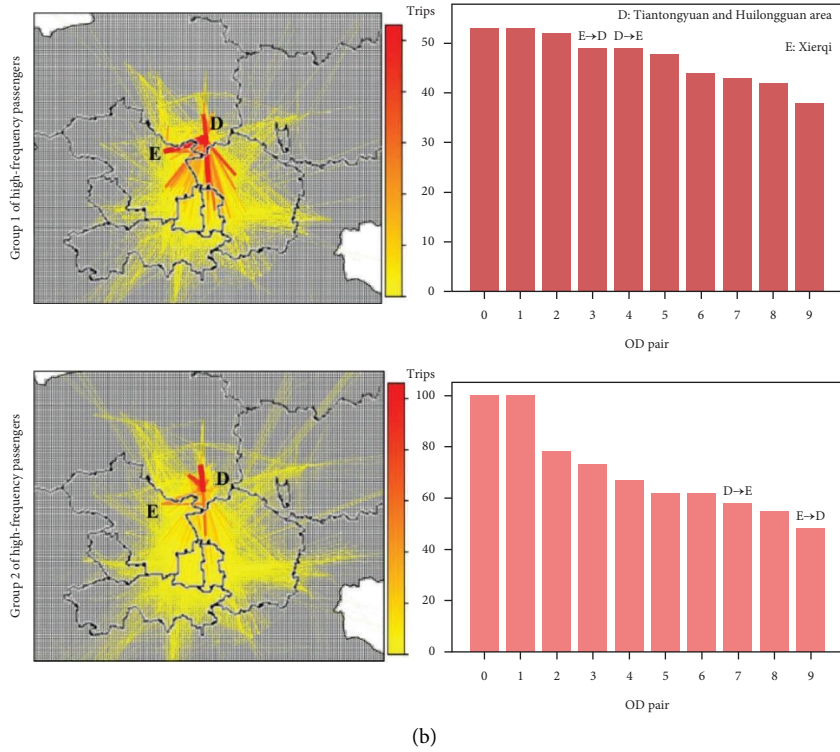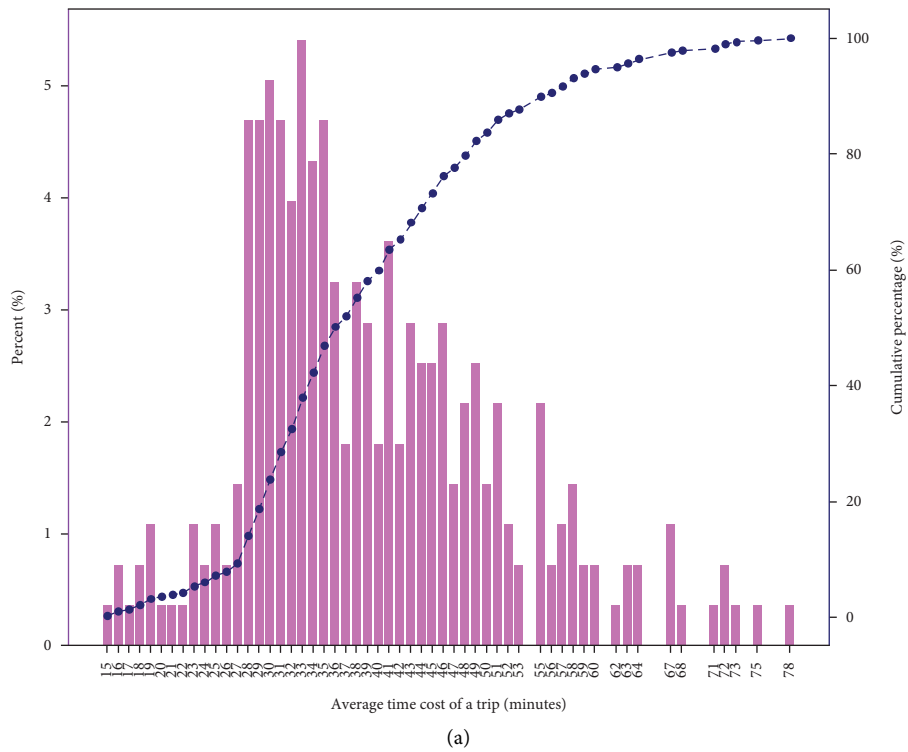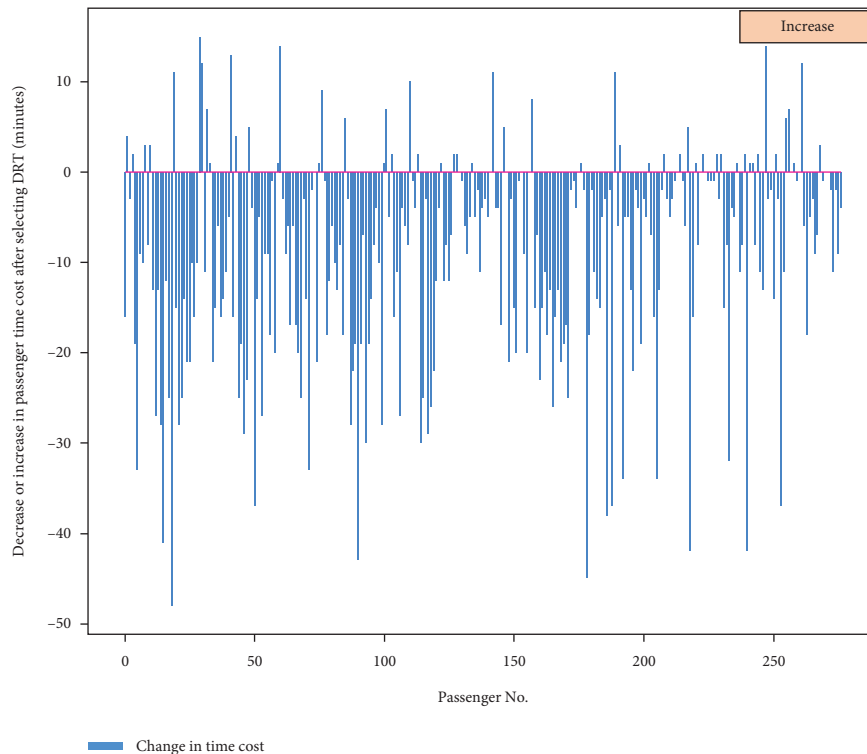

(a)

FIGURE 15: Continued.

(b)

FIGURE 15: Travel spatial patterns for four types of passengers on weekdays. (a) Low-frequency passengers. (b) High-frequency passengers.

(a)

FIGURE 16: Continued.

(b)

FIGURE 16: The travel time cost of high-frequency passengers with trips from Tiantongyuan to Xierqi. (a) Average travel time cost per trip for passengers using public transit. (b) Change in passenger travel time cost after selecting customized bus.

In addition, high-frequency passengers travel much more frequently than low-frequency passengers, so discounted fares and souvenirs can be offered to passengers who meet the high-frequency ridership classification, which can also motivate more low-frequency passengers who prefer to choose other travel modes (e.g., owned vehicles) to use public transit. Also, low-frequency passengers can be encouraged by reducing fares at nonpeak hours as they chose their travel time more freely.

7. Conclusions

In this paper, we studied the travel patterns of public transit passengers in Beijing, including both general patterns and the different patterns for four groups of passengers. First, we extracted travel indicators for each passenger and conducted a general analysis. Then, we used an unsupervised clustering algorithm to classify passengers and analyzed the travel frequency, travel time, travel space, transfers, and travel time cost for the four resulting groups of passengers. We found that there were some passengers with high travel frequency whose travel exhibited stable temporal and spatial patterns. Most of these passengers were commuters (standard commuters and flexible commuters), which comprise the main users of urban public transit. And it is feasible to improve public transit service levels based on passenger travel patterns. In addition, there are many limitations in our study. For example, we only studied the travel frequency and travel spatial and temporal characteristics of passengers, but not other aspects of travel characteristics. In the future, we will

combine other data for passenger profiling to analyze passenger characteristics from a multidimensional perspective [27, 28].

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Supplementary Materials

The supplementary material consisted of three sections. The first was an explanation of the choice of 1.2 km in the closed travel chain. The second was a methodological supplement to the classification of high- and low-frequency passengers.

The third was a supplement to the spatiotemporal patterns of different groups of passengers. (*Supplementary Materials*)

References

[1] M. Chester, S. Pincetl, Z. Elizabeth, W. Eisenstein, and J. Matute, "Infrastructure and automobile shifts: positioning transit to reduce life-cycle environmental impacts for urban sustainability goals," *Environmental Research Letters*, vol. 8, no. 1, Article ID 15041, 2013.

[2] Y. Sun and Y. Cui, "Evaluating the coordinated development of economic, social and environmental benefits of urban public transportation infrastructure: case study of four Chinese autonomous municipalities," *Transport Policy*, vol. 66, pp. 116–126, 2018.

[3] M. E. G. Parker, M. Li, M. A. Bouzaghrane et al., "Public transit use in the United States in the era of COVID-19: transit riders' travel behavior in the COVID-19 impact and recovery period," *Transport Policy*, vol. 111, pp. 53–62, 2021.

[4] H. Nishiuchi, J. King, and T. Todoroki, "Spatial-temporal daily frequent trip pattern of public transport passengers using smart card data," *International Journal of Intelligent Transportation Systems Research*, vol. 11, no. 1, pp. 1–10, 2013.

[5] X. Ma, Y. J. Wu, Y. Wang, F. Chen, and J. Liu, "Mining smart card data for transit riders' travel patterns," *Transportation Research Part C: Emerging Technologies*, vol. 36, pp. 1–12, 2013.

[6] X. Ma, Y. Wang, F. Chen, and J. Liu, "Transit smart card data mining for passenger origin information extraction," *Journal of Zhejiang University - Science C*, vol. 13, no. 10, pp. 750–760, 2012.

[7] C. M. Schneider, C. Rudloff, D. Bauer, and M. C. González, "Daily travel behavior: lessons from a week-long survey for the extraction of human mobility motifs related information," in *Proceedings of the 2nd ACM SIGKDD international workshop on urban computing*, pp. 1–7, Chicago, Illinois, August 2013.

[8] 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," *Journal of Transport Geography*, vol. 41, pp. 21–36, 2014.

[9] J. Zhao, Q. Qu, F. Zhang, C. Xu, and S. Liu, "Spatio-temporal analysis of passenger travel patterns in massive smart card data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 11, pp. 3135–3146, 2017.

[10] L. M. Kieu, A. Bhaskar, and E. Chung, "Mining temporal and spatial travel regularity for transit planning," in *Proceedings of the Australasian Transport Research Forum 2013 Proceedings*, pp. 1–12, Australasian Transport Research Forum, Brisbane, QLD Australia, October 2013.

[11] N. Lathia, D. Quercia, and J. Crowcroft, "The hidden image of the city: sensing community well-being from urban mobility," in *Proceedings of the International Conference on Pervasive Computing*, pp. 91–98, Springer, Berlin, Heidelberg, June 2012.

[12] M. K. El Mahrsi, E. Côme, J Baro, and L Oukhellou, "Understanding passenger patterns in public transit through smart card and socioeconomic data: A case study in rennes, france," *ACM SIGKDD workshop on urban computing*, New York, NY, USA, 2014, https://www.comeetie.fr/pdfrepos/urbcomp2014.pdf.

[13] Y. Liu and T. Cheng, "Understanding public transit patterns with open geodemographics to facilitate public transport planning," *Transportmetrica: Transportation Science*, Taylor & Francis, vol. 16, no. 1, pp. 76–103, 2020.

[14] X. Zhao, M. Cui, and D. Levinson, "Exploring temporal variability in travel patterns on public transit using big smart card data," *Environ Plan B Urban Anal City Sci*, SAGE Publications Sage UK, London, England, Article ID 23998083221089664, 2022.

[15] M. A. Ortega-Tong, *Classification of London's Public Transport Users Using Smart Card Data*, Massachusetts Institute of Technology, Cambridge, MA 02139, USA, 2013.

[16] T. Yang, P. Zhao, and X. Yao, "A method to estimate URT passenger spatial-temporal trajectory with smart card data and train schedules," *Sustainability*, vol. 12, no. 6, p. 2574, 2020.

[17] R. Kaewkluengklom, F. Kurauchi, and T. Iwamoto, "Investigation of changes in passenger behavior using longitudinal smart card data," *International Journal of Intelligent Transportation Systems Research*, vol. 19, no. 1, pp. 155–166, 2021.

[18] L. M. Kieu, A. Bhaskar, and E. Chung, "Transit passenger segmentation using travel regularity mined from Smart Card transactions data," *Transportation Research Board (TRB) 93rd Annual Meeting Compendium of Papers*, pp. 1–17, Washington, D.C, USA, 2014.

[19] Q. Ouyang, Y. Lv, Y. Ren, J. Ma, and J. Li, "Passenger travel regularity analysis based on a large scale smart card data," *Journal of Advanced Transportation*, vol. 2018, Article ID 9457486, 2018.

[20] C. Cui, Y. Zhao, and Z. Duan, "Research on the stability of public transit passenger travel behavior based on smart card data," in *Proceedings of the CICTP 2014: Safe, Smart, and Sustainable Multimodal Transportation Systems*, pp. 1318–1326, Changsha , China, July 2014.

[21] A. S. Briand, E. Côme, M. Trépanier, and L. Oukhellou, "Analyzing year-to-year changes in public transport passenger behaviour using smart card data," *Transportation Research Part C: Emerging Technologies*, vol. 79, pp. 274–289, 2017.

[22] Y. Zhu, "Estimating the activity types of transit travelers using smart card transaction data: a case study of Singapore," *Transportation*, vol. 47, no. 6, pp. 2703–2730, 2020.

[23] Y. Zhang, Z. Zhou, W. Wang, and R. Zhang, *The 5th Comprehensive Transport Survey Summary Report*, Beijing China, 2016.

[24] Y. Zhao, G. Yin, and S. Wu, *2021 China Principal Cities Sharing Bikes and Sharing Electric Bikes Riding Report*, Beijing China, 2021.

[25] X. Ma, C. Liu, H. Wen, Y. Wang, and Y. J. Wu, "Understanding commuting patterns using transit smart card data," *Journal of Transport Geography*, Pergamon, vol. 58, pp. 135–145, 2017.

[26] A. Huang, Z. Dou, L. Qi, and L. Wang, "Flexible route optimization for demand-responsive public transit service," *J Transp Eng A Syst*, vol. 146, no. 12, Article ID 04020132, 2020.