

Research Article

Individual Travel Behavior Modeling of Public Transport Passenger Based on Graph Construction

Quan Liang,¹ Jiancheng Weng¹,¹ Wei Zhou,² Selene Baez Santamaria,³ Jianming Ma,⁴ and Jian Rong¹

¹Beijing Key Laboratory of Traffic Engineering, Beijing University of Technology, Beijing 100124, China

²Ministry of Transport of the People's Republic of China, Beijing 100736, China

³Computer Science Department, Vrije Universiteit Amsterdam, 1081 HV Amsterdam, Netherlands

⁴Texas Department of Transportation, Austin, TX 78717, USA

Correspondence should be addressed to Jiancheng Weng; youthweng@bjut.edu.cn

Received 14 January 2018; Revised 8 April 2018; Accepted 18 April 2018; Published 16 May 2018

Academic Editor: Luca D'Acerno

Copyright © 2018 Quan Liang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper presents a novel method for mining the individual travel behavior regularity of different public transport passengers through constructing travel behavior graph based model. The individual travel behavior graph is developed to represent spatial positions, time distributions, and travel routes and further forecasts the public transport passenger's behavior choice. The proposed travel behavior graph is composed of macronodes, arcs, and transfer probability. Each macronode corresponds to a travel association map and represents a travel behavior. A travel association map also contains its own nodes. The nodes of a travel association map are created when the processed travel chain data shows significant change. Thus, each node of three layers represents a significant change of spatial travel positions, travel time, and routes, respectively. Since a travel association map represents a travel behavior, the graph can be considered a sequence of travel behaviors. Through integrating travel association map and calculating the probabilities of the arcs, it is possible to construct a unique travel behavior graph for each passenger. The data used in this study are multimode data matched by certain rules based on the data of public transport smart card transactions and network features. The case study results show that graph based method to model the individual travel behavior of public transport passengers is effective and feasible. Travel behavior graphs support customized public transport travel characteristics analysis and demand prediction.

1. Introduction

Public transport has been one of the main travel modes in urban areas due to its comprehensive service for travelers and big influence on urban traffic systems. Taking Beijing as example, there are about 11.29 million public transport commute trips per day [1]. The average trip time and distance are 54 minutes and 19.4 kilometers, respectively [2]. There exist also many different kinds of public transport modes, such as metro, traditional bus, customized bus, and microcirculation shuttle bus. In order to better meet travel demands and further improve public transport service level, it is essential to obtain the demand of transport scheduling accurately and hierarchically. To achieve this, it is first necessary to extract the personal travel characteristics with respect to temporal and spatial distribution and their travel behavior choice.

In previous studies, many researchers have attempted to analyze public transport travel characteristics and to explore extraction methods for travel behavior regularity. Generally, these previous studies mainly focused on statistical analysis based on conventional trip survey of relatively small sample sizes or depended on limited public transport data [3–7]. On one hand, common feature parameters of public transport travel were extracted. Examples include the following: Walle and Steenberghen [4] analyzed the influence of temporal and spatial factors on public transport travel mode selection. Zhou et al. [5] presented comparative analysis of resident travel characteristics between typical cities in developed countries and in China, according to individual trip amount, purpose, and mode. On the other hand, travel behavior models of public transport were established to further analyze passengers' travel behavior choice. Structural equation model

was usually used to reflect public transport selection behavior based on trip chain extraction [6]. A travel mode selection model of trip chain was also developed and verified with high accuracy by taking travel time and cost into account [7].

As limited to data collection approaches, travel behavior characteristic parameters and travel behavior choice of public transport travelers were partially analyzed in previous studies. However, limited sample data does not sufficiently reflect the accurate travel behavior features in a public transport system. It is necessary to study the travel behavior features of public transport in larger scales. Fortunately, emerging technologies such as cloud computing and Internet+ have promoted the development of advanced public transport system (APTS). Combined with network communication, geographic information systems (GIS), global position systems (GPS), and electronic controls, APTS lays a good foundation for multimode data collection.

In fact, many researchers have attempted to apply public transport smart card data and GPS data for travel behavior characteristic analysis. Most of these relative studies focused on estimating public transport travelers' origin, transfer, and destination. These researches tried to explore travel features in local areas or for given travel modes. Zhang [8] analyzed the distribution and transfer of passenger flow of platform metro passenger at operational stages. Cao et al. [9] proposed models that demonstrated the characteristics of entering station, exiting station, and transferring and waiting of metro passengers. Meanwhile, an index algorithm was developed to optimize public transport vehicles operation [10]. In summary, compared to manual surveys, these studies have lower cost and higher accuracy. However, these statistical results were applied to obtain the average features and total attributes of public transport travelers; the characteristics of individual passenger were largely ignored.

As with different personal characteristics, such as travel habits, job type, and income, not all public transport travelers have the same travel regularity. Personal characteristics and habits would affect the regularity of passenger travel behavior apparently [11]. Recently, a few research studies have taken individuals travel characteristics into account, but the individual features were not studied adequately. For example, Yang et al. [12] constructed a binomial model to analyze the microcosmic factors that influence individual choice of bike sharing services. But influence factors were not comprehensive and accurate enough because of limited sample data. Besides, Xu [13] studied the individual passenger behavior in urban subway hub, but only walkways and downward and upward stairways were built to illustrate the relationships between individual passengers' walking speed and space. Many other types of infrastructure, such as the platform, were not fully considered.

Summarizing above, there is no doubt that grasping travel characteristics of individual passengers based on personal travel data could potentially help to predict personal trip behavior more accurately. However, as being subjected to the limitation of data sample, conventional statistical analysis method, and mathematic model, there is still lack of approaches to exhibit the space-time changing process of travel behavior intuitively, and thus it is difficult for visual

grasping or quantitative comparing of the whole and partial travel behaviors of a traveler. Therefore, it is challenging to obtain the travel demand of public transport passenger precisely.

In recent years, the knowledge graph, consisting of nodes, arcs, and/or transfer possibility, has been widely used in traffic, medicine, library, information, and many other areas recently [14–16]. For Chen et al.' study [14], driver behavior graph was constructed to identify the driving habit of different drivers (e.g., whether it is safe or not) and to express the process of his/her whole performance during driving intuitively. Integrating medical data from different languages, medical graph could seek the best knowledge presentation scheme [15]. Medical knowledge graph helped doctors make a rational diagnosis by using dynamic reasoning of graphs and patients understood the diagnosis results because of the intuitive information from graphs [15]. Similarly, in library and information studies, reference citation graph could be automatically generated to display the relationship between literatures and authors and further to analyze the structure, relationship, and evolutionary process of scientific knowledge [16]. Overall, three advantages of this graphical method were apparent. Firstly, knowledge graph can organize and express the features of mass and heterogeneous data effectively. Secondly, knowledge graph has excellent ability in deep reasoning depending on powerful knowledge base. Thirdly, knowledge graph could achieve cognitive ability when further combined with other methods (e.g., deep learning and dynamic fuzzy logic).

Based on these above analyses, the main innovation of our research is to model travel behavior from the aspect of individual passengers by adopting knowledge graph. On one hand, the whole travel characteristics of a given passenger during any time period could be completely and intuitively expressed in one graph. On the other hand, quantitative comparing, deep mining, forecasting, and reasoning of travel regularity would be achieved when establishing the database of travel behavior graphs from various passengers under the assist of computer programming. After this, the accurate travel regularity of one or more passengers could be easily got even if the initial travel data missed partially.

In this study, we mainly discussed how to establish individual travel behavior graph referencing the elements, rules, and other requirements of knowledge graph, especially the reference citation graph and measurement visualization analysis in Chinese National Knowledge Infrastructure (CNKI, <http://www.cnki.net>). Firstly, the travel chain extraction method of individual public transport traveler was proposed based on multimode public transport data. Then, the travel association map, including individual space position, travel time, and route, was constructed and integrated to form the structure of travel behavior graph. Probabilities of the graph arcs were considered to estimate the transfer probability of passenger's next trip. The proposed travel behavior graph of public transport intends to provide new insight into individual travel behavior modeling, which supports the intuitive individual public transport travel presentation and implicit regularity mining. The study results lay a foundation to more accurate travel demand prediction considering

TABLE 1: Effective fields of public transport smart card transaction data.

	Bus	Metro	Bike sharing services
Smart card transaction	(i) Line code	(i) Card code	(i) Card code
	(ii) Card code	(ii) Entry line number	(ii) Rent station number
	(iii) Boarding station number	(iii) Entry station number	(iii) Rent time
	(iv) Boarding time	(iv) Entry station time	(iv) Return station number
	(v) Alighting station number	(v) Exit line number	(v) Return time
	(vi) Alighting time	(vi) Exit station number	
		(vii) Exit station time	

TABLE 2: Effective fields of public transport network features data.

	Bus	Metro	Bike sharing services
Network features	(i) Arc length	(i) Line code	(i) District name
	(ii) Arc code	(ii) Station code	(ii) Station code
	(iii) Station longitude and latitude	(iii) Station name	(iii) Station name
	(iv) Station spaces	(iv) Station longitude and latitude	(iv) Station latitude and longitude
		(v) Station spaces	

the personal characteristics of public transport individuals.

2. Multimode Data Collection and Matching

In order to construct individual travel behavior graph and further extract travel characteristics of public transport travelers in depth, two types of public transport base data were collected and matched in this study. Namely, multimode data of public transport smart card transaction and network features were used.

2.1. Smart Card Transaction Data of Public Transport. The smart card transaction data contains bus Integrated Circuit (IC) card data, metro automated fare collection (AFC) data, and bike sharing services card data. The data used in the current study were obtained from the government agency of the Beijing Transportation Operations Coordination Center (TOCC). Each day, there are about eight million bus IC card data, five million metro AFC data, and sixty thousand bike sharing services card data, respectively. Public transport data from April 2017 was used in this study.

Six effective fields of bus IC card data, seven effective fields of AFC data, and five effective fields of bike sharing services card data were extracted from smart card terminal. The detailed effective fields of smart card transaction data were shown in Table 1.

2.2. Data of Multimode Public Transport Network Features. The data of network features used in this study contains bus line and station data, metro line and station data, and the station data of bike sharing services. The detailed fields of network features data are shown in Table 2.

Bus line and station data were collected by geographic information system (GIS) map analysis. The arc length and code were obtained directly by GIS map. The longitude

and latitude of stations were acquired through station aggregation. The space between each station and the other was calculated by searching bus route.

For the metro line and station data, AFC station code was specified by uniform rules in Beijing and the station names were confirmed by combining the line code with the station code. Station latitude and longitude were obtained by coordinate aggregation of each exit and entry of the station. The travel distance of any metro station pairs is acquired through the shortest route searching method.

For the station data of bike sharing services, district name, station code, and station name were obtained directly from base data. Station latitude and longitude were calculated by coordinate aggregation.

2.3. Data Matching. As multimode public transport data is obtained from different approaches, it is necessary to reconstruct the travel process of a given passenger by using corresponding rules. Thus, the main objective of data matching is to get integration data of public transport smart card from different travel modes and further to obtain the travel chain of individual passenger.

Firstly, public transport smart card data are integrated from different travel modes together by using the same card code of an individual public transport passenger.

Secondly, public transport smart card integration data are ordered by departure time. Departure time represents bus boarding time, metro entry station time, and bike rent time.

Thirdly, travel chain data of an individual passenger are obtained to record the whole travel process per travel through travelers' transfer judgment. The travel chain consists of one or more travel stages. The method used for travel chain extraction from smart card integration data was mainly based on the approach proposed by Wang [17]. Table 3 is an example of the travel chain data for a passenger in April 2017 in Beijing.

TABLE 3: Example of travel chain data of one passenger.

Card code	24050273	24050273	...	24050273
Travel mode	Metro	Metro	...	Bus
Departure time	2017/4/1 8:28	2017/4/1 17:53	...	2017/4/30 17:04
Arrival time	2017/4/1 8:55	2017/4/1 16:29	...	2017/4/30 17:39
Departure line	4	1	...	114
Arrival line	1	4	...	114
Path distance (km)	8.115	8.115	...	8.620
Departure station name	BeiJingNan Zhan	Mu Xi Di	...	Bai Yun Qiao Xi
Arrival station name	Mu Xi Di	BeiJingNan Zhan	...	Kai Yang Qiao Nan
Longitude of departure station (°)	116.3779	116.3369	...	116.3395
Latitude of departure station (°)	39.8641	39.9075	...	39.8973
Longitude of arrival station (°)	116.3369	116.3779	...	116.3471
Latitude of arrival station (°)	39.9075	39.8641	...	39.8666

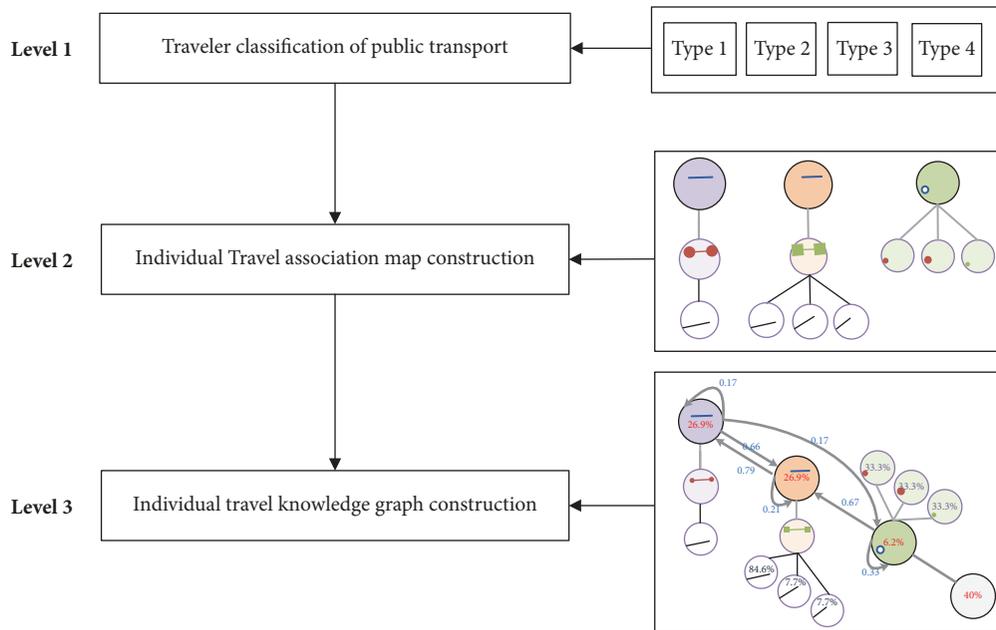


FIGURE 1: The flowchart of individual travel behavior graph construction.

3. Individual Travel Behavior Graph Construction

The main function of a travel behavior graph is to transform the low-level numerical data into a high-level abstracted form and thus to display the individual travel characteristics intuitively. The flowchart for constructing a travel behavior graph is shown in Figure 1, which includes three levels from top to bottom. After the individual travel chain data is preprocessed, the typical type of passengers can be classified (for level 1). Next, for level 2, the individual travel association maps are constructed mainly based on the travel space position, travel time, and routes. Each travel association map can be regarded as a macronode of the travel behavior graph for level 3. The arcs connecting each macronode and its corresponding possibility representing travel behavior transfer are created in level 3. Thus, the travel behavior graph can be used to analyze

the temporal and spatial characteristics and travel behavior choice of an individual passenger.

3.1. Traveler Classification of Public Transport. The dependency degree of an individual passenger on public transport is an important indicator to reflect his or her travel characteristics [18], which are typically represented by travel days and travel times. Thus, passengers were classified into different types based on their dependency on public transport in this study. On one hand, applying passenger type classification provides a uniform rule for establishing travel graphs (i.e., the layout of travel behavior graph). On the other hand, it becomes easy to grasp and compare the travel features of various public transport passengers.

Passengers' travel chain data in April 2017 in Beijing was analyzed to set the indicators' thresholds for passenger classification. For travel days, the threshold used for

TABLE 4: Symbols used for travel time classification.

Symbol	Morning			Noon	Afternoon				
	●	●	●	▲	■	■	■	■	■
Size (pound)	9	12	15	9	7.5	10.5	14	16	22
Travel time	5–7:00	7–9:00	9–11:00	11–13:00	13–15:00	15–17:00	17–19:00	19–21:00	21–23:00

identifying higher or lower public transport travel frequency for a passenger was defined as average 4 days in a week and thus 16 days in one month, while for travel times, the threshold for higher frequency of public transport travel times was supposed to be at least twice in one day and thus 32 times in a month.

As passengers are divided into two groups by each indicator, a total of four types of passengers could be classified, such as passengers with high frequency of both travel days and times or passengers with high frequency of travel days and low frequency of travel times.

3.2. Travel Association Map Construction. The travel association map, which consists of passengers’ travel spatial positions, time distributions, and routes, represents macronodes of the individual travel behavior graphs. This part will discuss the method for building travel association maps from travel chain data.

3.2.1. Individual Travel Space Position Clustering. Based on the travel chain data ordered by time of occurrence, the longitude and latitude of the travel space position could be obtained through matching the departure and arrival stations with the geolocation data. Then, the method of hierarchical system clustering is applied to merge spatial position clusters according to the longitude and latitude data. The cluster method is based on connections between groups and the measurement standard was Euclidean distance. Thus, the positions are merged into the same set if the ODs of travel chains fall within a certain range.

In the first and second layers, the lines in the node were represented by the origin and destination points that were confirmed through travel data of longitudes and latitudes. According to data of public transport network features in Beijing, the longitude within the Five Rings is from 116.22 to 116.54, and the latitude within the Five Rings is from 39.76 to 40.02. So the horizontal and vertical coordinates were set as the range of longitude and latitude, respectively. After setting the node size for these two layers and removing the horizontal and vertical axes, the point positions were uniquely confirmed.

Figure 2 shows an example of cluster distribution results for a given passenger (P1). A total of three OD groups are created. The horizontal and vertical axes represent that the longitudes and latitudes changed from origin to destination. From group 1 (13 travel chains) and group 2 (13 travel chains), it is intuitive that two major travels existed for this passenger and the public transport travel was a round-trip. For group 3 (3 travel chains), as the distance of this paired OD is

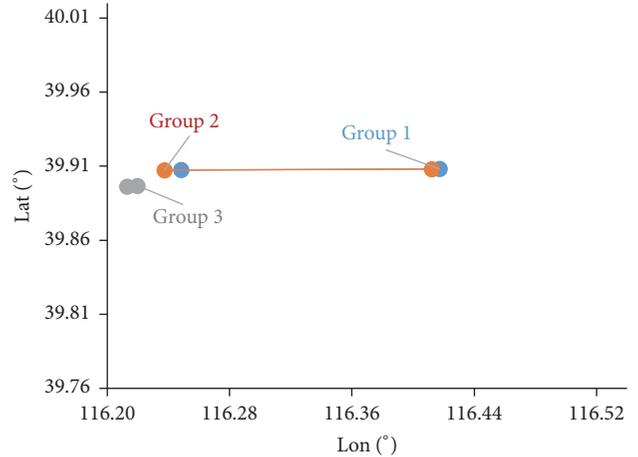


FIGURE 2: Example of cluster results of departure and arrival stations position.

relatively closer, the average geolocation is nearly overlapping.

3.2.2. Individual Travel Time Classification. Individual travel time is further classified based on the results of travel space position clustering. In each OD pair, travel time is classified by the departure time because of its typical representation of travel behavior. The departure time of each trip falls into one of nine different intervals, which are divided into two-hour periods from 5:00 a.m. to 23:00 p.m. The time periods cover the main operating time of public transport in Beijing. To illustrate the complete travel time, the classification result refers to passenger’s both departure and arrival times in the OD groups.

Table 4 shows the symbols used to mark the individual travel time classification results. Different shapes (circle, triangle, and square), sizes (7.5 to 22 pounds), and colors (red, orange, and green) are defined to represent the different travel time for each OD pair.

Figure 3 shows the results of individual travel time classification of three ODs for P1 in Figure 2. The travel time classifications, respectively, correspond to groups (1), (2), and (3) shown in Figure 2. Figure 3(a) illustrates that the departure and arrival times are both from 7:00 to 9:00 when the passenger’s travel belongs to the first OD pairs (group 1). Figure 3(b) displays that the departure and arrival times are both from 17:00 to 19:00 when the passenger’s travel is in group 2. For the third group, the travel time is irregular, which is distributed among three intervals as 5:00 to 7:00, 9:00 to 11:00, and 13:00 to 15:00 (see Figure 3(c1–c3)).

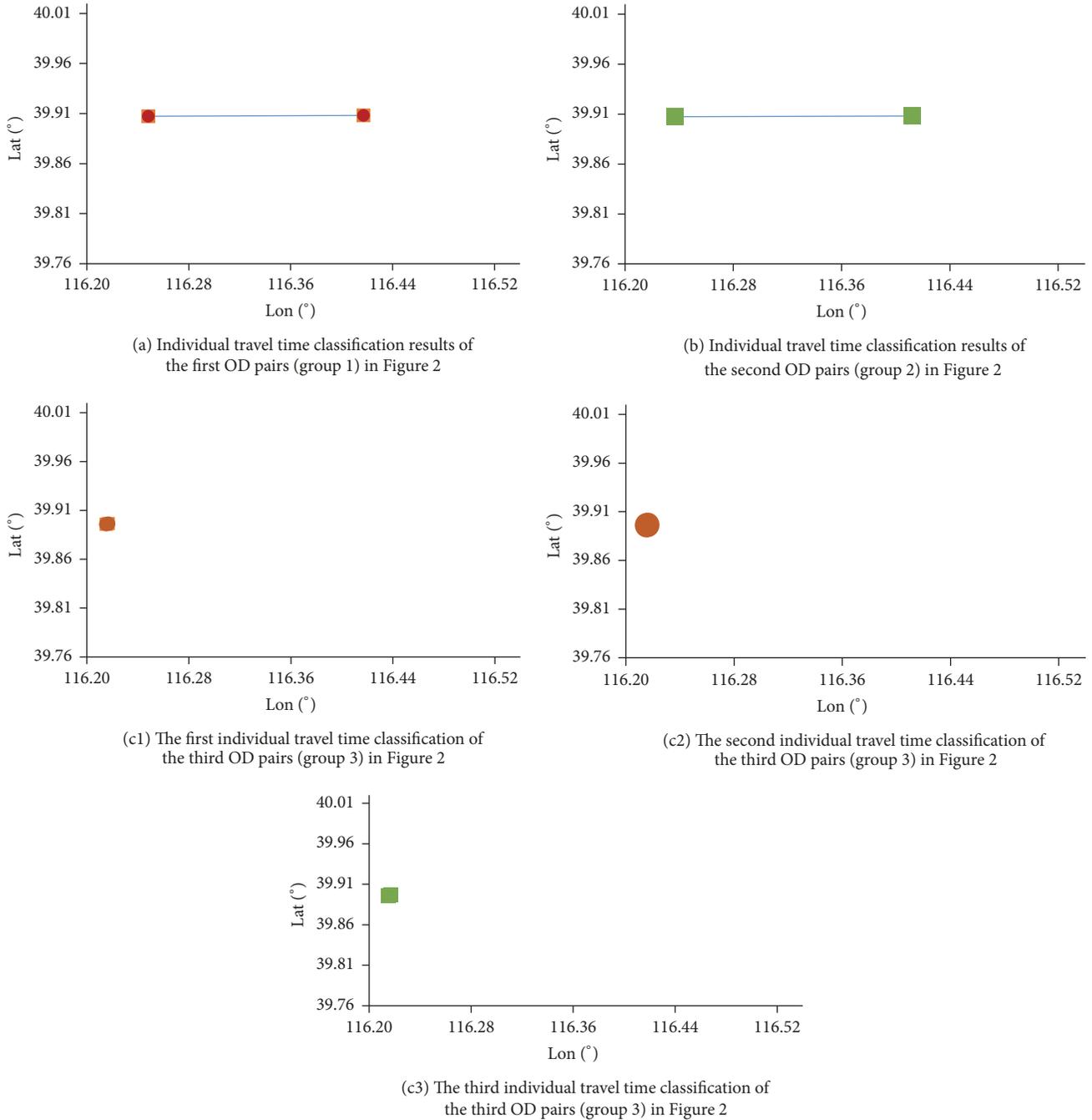


FIGURE 3: Examples of passenger travel time classification results of P1 in Figure 2.

Also, as the distance of this paired OD is relatively closer, the time distribution in Figure 3(c1–c3) is almost overlapping.

3.2.3. Travel Route Clustering. The result of travel route clustering represents travel modes of individual passengers. It is a further step based on space position clustering and travel time classification. Travel distance and route direction are selected to reflect travel modes. If both the route distance and directions are close in a certain range, the travel chains

will be merged into the same cluster. This step is repeated until the difference between the two closest travel chains is not significant. Different travel route clusters would be created for each travel time classification subset.

The travel distance is the actual path distance of each travel stage, such as the distance from departure to transfer stations and transfer to arrival stations. Travel distance is obtained by matching station information to network features data. Similarly, the travel direction is also calculated for each travel stage. Given a travel chain n , let lo_{no} and lo_{nd} be the

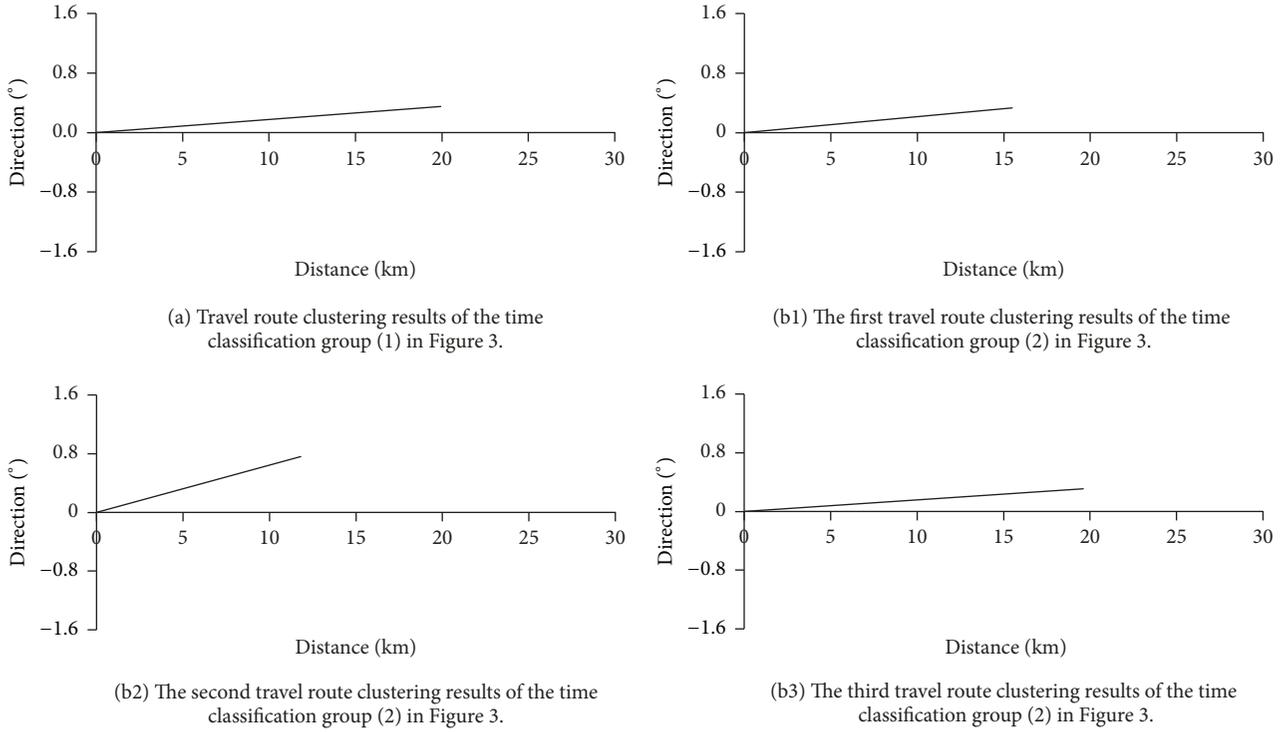


FIGURE 4: Travel route clustering results of P1 in Figure 3.

longitude of origin and destination and let la_{no} and la_{nd} be the latitude of origin and destination; the route direction can be obtained by the following equation:

$$Ang = \frac{la_{nd} - la_{no}}{7.5 * 10^{-6}} \times \frac{10^{-5}}{lo_{nd} - lo_{no}} \quad (1)$$

$$D_{chain}(n) = \text{degree}(\text{ATAN}(Ang)),$$

where Ang is the tangent value of the distance between the origin and destination of travel stage. D_{chain} is the direction of each travel stage.

For the third layer, the diagram was demonstrated by travel distance (x -axis) and travel direction (y -axis). According to the statistical data in Beijing, the average travel distance is about 16 kilometers, and the 85th percentiles value is about 26 kilometers. Thus, the range of x -axis was defined as 0 to 30 kilometers in this study. The direction of each travel stage was calculated by (1) in the paper, the range of which would be 0 to 360 degrees. Correspondingly, the antitangent value of the travel direction would be about $(-1.6, 1.6)$. In order to display travel distance and direction in the same figure with more clear visibility, the antitangent value of real angle was set as y -axis. Similarly, after setting the node size for the third layer and removing the horizontal and vertical axes, the point positions were uniquely confirmed.

In addition, the travel modes are represented by different line attributes. The single and black solid lines represent bus travel, the dotted and red lines represent metro travel, and the double and blue solid lines indicate bike sharing service. Figure 4 shows an example of travel route clustering result. These route clustering results in Figure 4 correspond

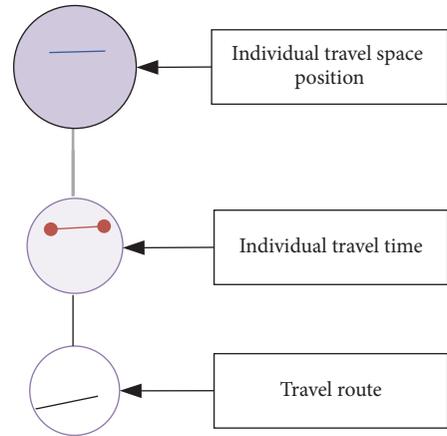


FIGURE 5: Example of one travel association map of P1.

to groups (1) and (2) shown in Figure 3. In Figure 4, group (1) has only one path and group (2) has three different paths (see Figure 4(b1–b3)). In addition, the distance of group (3) is too short to display and thus its route clustering result was not considered.

3.2.4. Individual Travel Association Map Construction. The travel association map is regarded as a macronode of the whole travel behavior graph. Travel spatial positions, time distributions, and the routes are zoomed out and, respectively, merged into three layers of a travel association map (see Figure 5). Referring to the multilayer programming theory,

TABLE 5: Four layouts of travel behavior graph.

Travel times		Travel days	
High	Low	High	Low
High	Low	High	Low
Low	High	Low	High

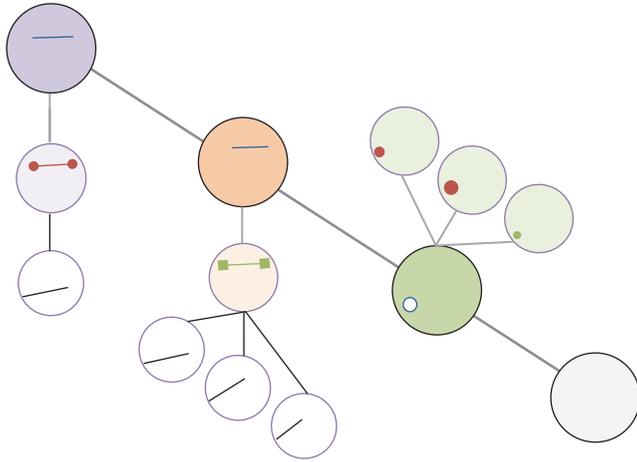


FIGURE 6: Travel association map integration of P1.

travel space position (big circle) is set as the first layer of the travel behavior graph, and travel time (median circle) and route (small circle) are treated as the second and third layers. Every layer is considered as a node of travel association map. In each layer, the horizontal and vertical coordinates are uniformed as well as the symbols and lines. After further abstraction processing, one of the travel association maps of passenger P1 can be drawn as Figure 5. According to the information displayed in Figure 2, there should be a total of three travel association maps of P1. Each travel association map represents a specific travel behavior.

3.3. Travel Behavior Graph Construction. After each travel association map constructed, the travel behavior graph is developed by using arcs to connect and arrange every association map. Meanwhile, the statistical probability within and among each travel association map was also calculated to represent travel behavior choice.

3.3.1. Travel Association Map Integration. Firstly, according to these four types of passengers defined by travel days and times in this study, the travel behavior graph is arranged as four different layouts as below, shown in Table 5.

For P1 mentioned earlier, the travel days and times are 18 and 29 in one month, which means that the traveler has high frequency of travel days and times. The arrangement of integrated travel association map is shown in Figure 6 (i.e., diagonally downwards from left to right). As the space positions of P1 consist of four categories, three public transport (solid circle) and one nonpublic transport (hollow circle) travel, the first layer is presented by four uniformed big

circles connected and abstracted space position coordinate axis within each circle. The information of the first layer exhibited is consistent with Figure 2. Then, the second layer is set as middle circles and abstracted time classification coordinate axes are within it. The information of circles in the second layer is corresponding to the findings in Figure 3. For the third layer, small circles are selected and abstracted route clustering coordinate axes are within these circles. It is worth noting that as travel distance for the third space position (shown as green circle) is too short to display, the travel route clustering is not considered in its third layer. The information in Figure 4 is displayed in the third layer.

3.3.2. Probabilities Calculation within and among Travel Association Map. In the travel behavior graph, probabilities of nodes (within circles) illustrate the occurrence proportion, and the weight of arcs among association maps indicates the probability of transformation from one association map to another.

For the transfer probability, it is calculated through three steps. Firstly, the travel space position of individual passengers is clustered to get his/her origin-destination categories (ODs). Secondly, the individual travel chain data of one month are ordered by departure time. Thirdly, the transfer probabilities are calculated by statistically analyzing the occupancy percentages from one OD to any others and itself, for example, if there are 11 times when the second OD occurred after the first OD and 3 times when the first OD followed itself; but there was inexistence when the third OD occurred after the first OD. Thus, the transfer probability is 0.79 (11/14) from the first to the second OD, is 0 (0/14) from the first to the third OD, and is 0.21 (3/14) from the first OD to itself.

The total occurrence proportion of each node per layer is assumed as 100 percent. In the first layer, the proportion of nonpublic transport and total public transport proportion are calculated by travel days during a month, respectively. Then proportions of each public transport OD cluster group are obtained through travel times of different ODs. In the second layer, the proportion is the occurrence frequency of different travel time. Similarly, the proportion in the third layer reflects travel routes selection.

As shown in Figure 7, probabilities within and between two travel association maps were displayed. For the first travel association map, the space position accounted for 26.9% in the first layer for all ODs. When it occurred, the departure time was between 7:00 and 9:00 a.m. in the second layer. The travel route took bus all the way in the third layer. It means that the travel time and route are stable for the first OD, while for the second travel association map, the occurrence probabilities were also 26.9% in its first layer. As with the high similarity to the first OD, it implied that these two travels were round-trips. For the second OD, the travel time was concentrated on 17:00–19:00 p.m. The route was focused on the first node in the third layer. In addition, the probability between these two association maps is 0.79, which indicates that one macronode would occur after the other in a large degree.

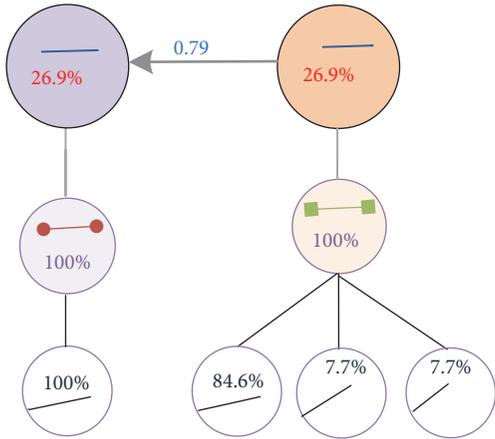


FIGURE 7: A simple example of the probabilities within and among travel association map.

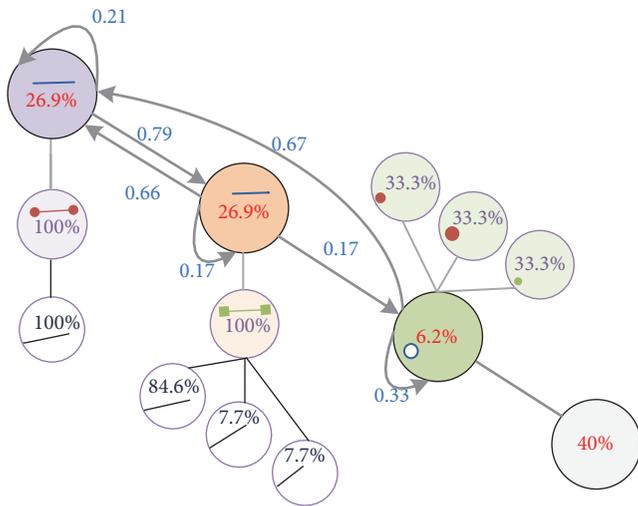


FIGURE 8: Individual travel behavior graph of passenger P1.

4. Case Studies

To illustrate and verify the effectiveness of travel behavior graph based method in modeling individual travel behavior of public transport travelers, four passengers of each type were randomly selected to construct their individual travel behavior graphs (shown in Figures 8–11).

Figure 8 is the travel behavior graph of P1. It shows that the public transport travel accounts for 60% of this individual traveler’s entire trips. For this passenger’s public transport travel, there are mainly three travel behaviors (i.e., three travel association maps). The first two travel behaviors displayed as purple and pink circles are the main travels and they accounted for equal percentages. For the first travel behavior, both the travel time (morning peak) and path are unique (bus). For the second travel behavior, the travel time (evening peak) is also fixed, while there are three travel different routes and the first route is most likely to be adopted by this traveler (i.e., 84.6%). For the third travel behavior, the travel time is averagely distributed to three periods, and the

travel distance is very short. In addition, according to the transfer probabilities among each travel behavior, it could be drawn that the first two behaviors occurred subsequently between each other in high degree. It is implied that these two travel behaviors might be round-trips. Besides, the second behavior might be generated with possibility of about 70% if the third behavior existed. Summarizing the regularities analysis above, it is implied that this passenger would be a commuter.

Figure 9 demonstrates the travel behavior graph of P2, which is more complicated and irregular than P1. This travel (16 travel days and 25 travel times in one month) belongs to high frequency of travel days and low frequency of travel times. The public transport travel accounts for 53.3% of the entire trips. This passenger’s public transport travel could be classified as four travel behaviors. The first two travel behaviors displayed as purple and pink circles are the main travels (i.e., 25.6% and 17.1%). For the first travel behavior, the travel time is concentrated on morning peak and the route is unique (metro). For the second travel behavior, the travel time is mainly in the afternoon especially during evening peak, and metro is the main travel mode. The travel time of his third travel behavior is focused on evening and travel path is also fixed by metro. For the fourth travel behavior, the travel time is equally distributed to morning and evening, and the route includes single metro or the combination of metro and bus. In addition, based on the transfer probabilities among each travel behavior, it could be drawn that the first behavior could follow the second behavior with half probability. Additionally, the first behavior would be generated if the third behavior existed, and the third behavior would occur after the fourth behavior.

The travel behavior graph of P3 is displayed in Figure 10. The travel days are low frequency (13 travel days in one month) and travel times are high frequency (39 travel times in one month). Public transport is not the dominant mode (only 43.3%). There are mainly three public transport travel behaviors for this passenger. The first two travel behaviors displayed as purple and pink circles are the main public transport travels. For the first travel behavior, the travel time is averagely divided into morning and afternoon periods, and bus is the only travel mode. The travel time is also mainly focused on morning and noon (47.6% for each) for the second travel behavior, and travel path with fixed travel mode (i.e., bus) is unique for certain time periods. For the third travel behavior, the travel time is averagely distributed to two periods, and the travel route is not the same between these two periods. In addition, depending on the transfer probabilities among each travel behavior, it is indicated that the second behavior would occur subsequently after the first behavior, while the first behavior might also follow the second one with more than half possibility. Besides, if the third behavior existed, the first and second behaviors might happen with equal probabilities of 50% for each.

Figure 11 shows the travel behavior graph of P4. This travel belongs to low frequency of both travel days and times (15 travel days and 25 travel times in one month). Public transport and nonpublic transport, respectively, account for half of this passenger’s travels. For this passenger’s public

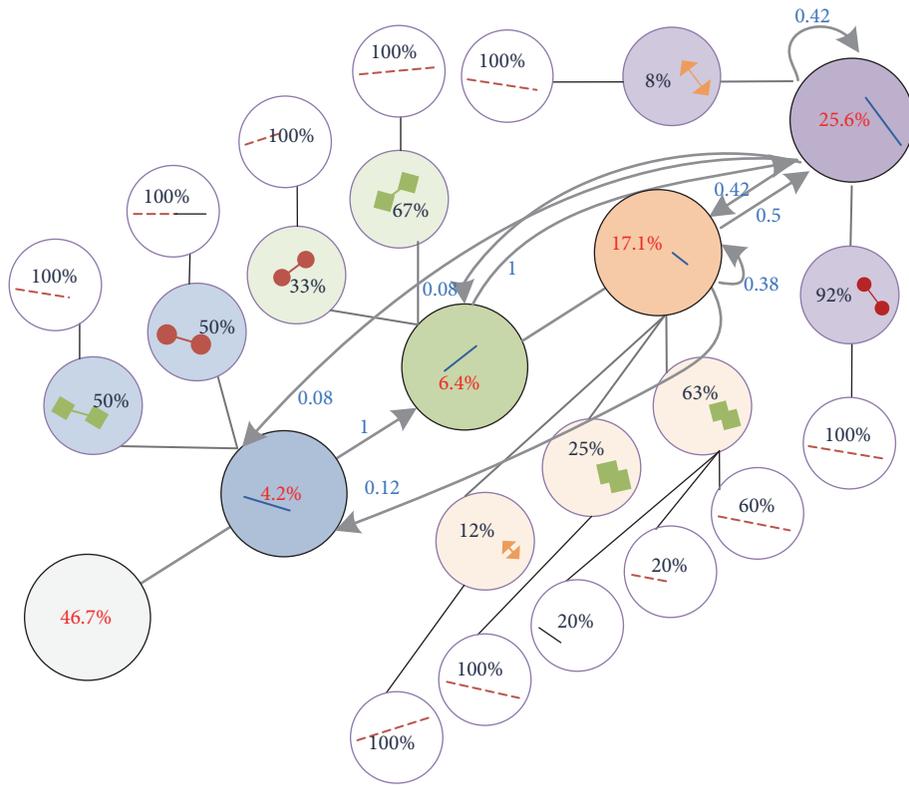


FIGURE 9: Individual travel behavior graph of passenger P2.

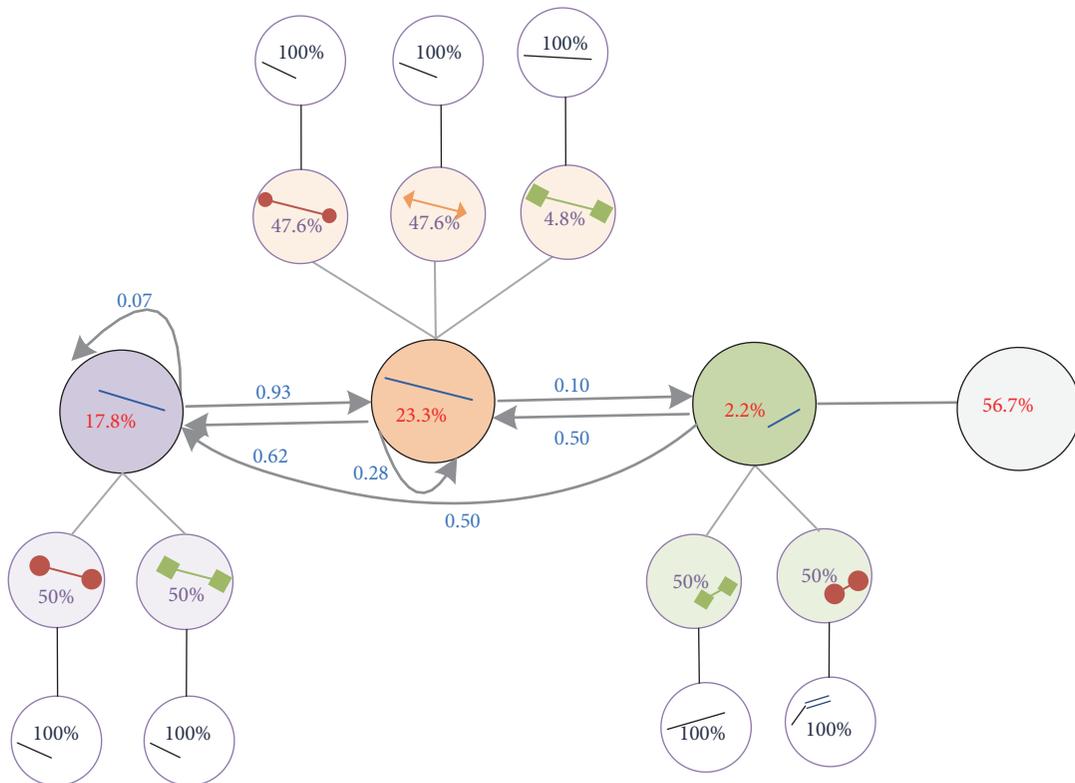


FIGURE 10: Individual travel behavior graph of passenger P3.

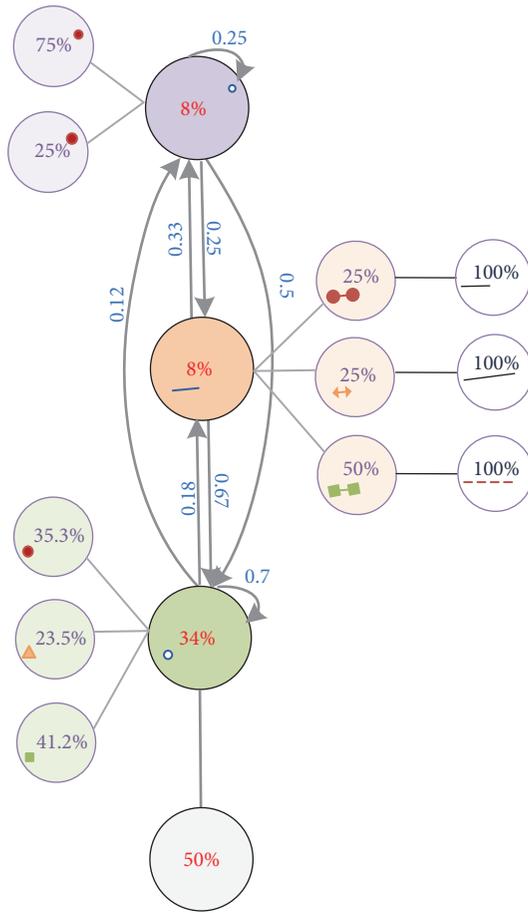


FIGURE 11: Individual travel behavior graph of passenger P4.

transport travel, there are three types of travel behaviors. For the first travel behavior, the travel time is centralized on the early morning, and the travel distance is very short. For the second travel behavior, the travel times are scattered, and half travels are during early afternoon with metro. For the third travel behavior with highest proportion of public transport travel, the travel times are scattered in the morning, at noon, and in the afternoon. The route distance is short as well. In addition, the transfer probabilities among each travel behaviors illustrate that the third travel behavior might be highly possible to happen if the first behavior, second behavior, or itself existed.

5. Discussion

In this study, an individual travel behavior graph constructing method was proposed and then followed by case studies of four passengers. The travel behaviors of these four passengers are not the same and thus the individual travel behavior graphs have different shapes. Travel behavior graph could intuitively display the difference of individual travel pattern. Examples of different travel information drawn from travel behavior graphs include the following: (1) P1 would be a commuter with fixed round-trip characteristics; (2) the travel behavior is relatively complicated for P2 because of

more dispersive travel time and routes; (3) bus is the main travel mode for P3; and (4) the travel destination by public transport is single and the travel distance is short for P4. In addition, although the overall travel features are dissimilar for different passengers, some specific travel behaviors (i.e., travel association map) are similar. For example, the fourth travel association map of P2 is similar to the third travel association map of P3. Both travel times are equally divided into two different periods, and the travel modes are mixed. The third travel association maps of P1 and P4 are also similar, both of which belong to short distance travel.

Besides its advantage in intuitive exhibition of travel behavior information, six feature indexes for travel characteristics description could be also extracted from individual travel behavior graphs. These quantitative indexes could be concluded as follows:

- (1) Dependency on public transport: passengers' dependency on public transport is represented by travel days and travel times, which is displayed as different layouts of travel behavior graphs. For example, the dependency on public transport of P1 is higher than that of P4
- (2) OD classifications: the first layer of graph includes OD clusters. Taking P1 and P4 as examples, both of them have three main public transport destinations, and two destinations for each occupy the same proportions
- (3) Round-trip or not: if public transport round-trip exists, two OD clusters would be significantly similar. Round-trips only existed for P1 for these four travelers
- (4) Peak periods: the travel time concentration degree is reflected by peak periods. The travel time of P1 is focused on morning peak and evening peak, but the travel time of P4 is flexible
- (5) Route selection: for individual travel with similar travel destination and departure time, the travel route might be unique or multiple. For instance, the travel path is fixed for the first travel behavior of both P1 and P2, while there exist different routes for the second travel association map of P2
- (6) Transfer probability of the next trip: this indicator represents the occurrence probability of the subsequent trip. For P2, if the fourth travel association map exists, the next trip would be the third and then follows by the first travel behavior.

Indeed, the advantage of travel behavior graph was not limited to extracting feature indexes of different passengers. Moreover, the graph can exhibit a complete individual travel behavior intuitively, and it contributes to judging the travel behavior's similarity and difference among different passengers. In addition, after generating travel behavior automatically depending on computer assistance, these graphs are the foundation for further intelligent applications (e.g., identifying, forecasting, and reasoning) when combined with other methods (e.g., deep learning).

As stated above, in recent years, knowledge graph has been widely applied in driving behavior identification in traffic areas [14], state of illness analysis in modern medicine [15], and relationship mining between literature and authors in information science [16]. In different research areas, the implication and contents of knowledge information might vary significantly, but the elements and structure were almost consistent. In this research, the shape of travel behavior graph was mainly referenced to the reference citation graph and measurement visualization analysis in CNKI. Thus, travel behavior graph also constituted of nodes and arcs. Besides, transfer possibility was added to travel behavior graph to forecast passengers' travel behavior choice.

In fact, the reference citation graph and measurement visualization analysis could be generated automatically in CNKI searching website within a few seconds. As we defined the uniformed standard of the elements of travel behavior graph (i.e., the method for obtaining node, arc, and transfer possibility of graph), it is believable to generate travel behavior graph automatically in our future work. In addition, we also discussed this issue of automatic generation of travel behavior graph with professional computer programmers. In the future and next study, we think that achieving the automatic generating of travel behavior graph supported by computer programming is practicable. The computing time of an individual travel behavior graph could be controlled into several seconds.

6. Conclusions

This paper develops a novel method for modeling an individual travel behavior based on knowledge graph to mine the travel regularity of different public transport passengers. The study results indicated that travel behavior graph is effective for intuitive presenting and further forecasting individual passengers' travel behavior characteristics, which support customized public transport travel characteristics analysis and demand prediction.

Based on multimode data of public transport smart card transaction and network features of one month in Beijing, travel chain data was obtained to reflect the entire travel process of different trips for an individual passenger by data matching. Depending on travel chain data, individual travel behavior graphs composed of macronodes, arcs, and transfer probability were constructed to represent travel spatial positions, time distributions, routes, and their possible behavior choice of next trip. In addition to intuitive illustration of travel information, the whole and specific travel behavior similarities or differences among different passengers were compared. Besides, six feature indexes were extracted from graphs to analyze the hidden characteristics of individual travel behavior.

In this study, we focused on finding a way to establish individual travel behavior graph referencing the elements, rules, and other requirements of knowledge graph, especially the reference citation graph and measurement visualization analysis in Chinese National Knowledge Infrastructure (CNKI, <http://www.cnki.net>). In addition, the reasonability and practicability were tested and verified by case studies of

several passengers. In the next studies, we will make efforts to generate travel behavior graph automatically depending on computer programming. The database of individual travel graph from various and large amount of passengers would be established in future researches. Finally, we hope to provide a convenient and precise way to grasp travel regularity through combining travel behavior graph and artificial intelligence. For example, combining with semantic analysis technology, individual travel behavior graphs can be used to further classify the types of public transport passengers in the future, such as high, moderate, and low stability of taking public transport of commuters and noncommuters. Thus, different types of passengers can be estimated to provide better public transport service with more suitable modes, such as rapid bus, customized shuttle bus, and mini bus. These findings could provide a foundation for transportation agencies to predict the public transport scheduling demand more accurately and optimize the transport operating network more reasonably.

Data Availability

The data used in the current study was obtained from the Beijing Transportation Operations Coordination Center (TOCC). The data generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest

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

Acknowledgments

The authors would like to show great appreciation for support from the National Natural Science Foundation of China (project: *Multimode Travel Demand Identification Methods Based on Individual Travel Feature Atlas of Public Transport Commuters* (no. 51578028)) and "Beijing Nova" Program by the Beijing Municipal Science and Technology Commission: *Study on the Feature Extraction Method and Demand Mechanism of Public Transport Travel with Multimodes* (no. Z171100001117100).

References

- [1] J. F. Guo, X. Li, G. C. Wang, H. M. Wen, and Y. Liu, "Beijing transport annual report," Research Report, 2017.
- [2] J. Rong and J. C. Weng, *Features Extraction of Public Transport Commute Travel Based on Multi-Mode Data*, Beijing University of Technology, 2014.
- [3] M. V. Sezhian, C. Muralidharan, T. Nambirajan, and S. G. Deshmukh, "Attribute-based perceptual mapping using discriminant analysis in a public sector passenger bus transport company: a case study," *Journal of Advanced Transportation*, vol. 48, no. 1, pp. 32–47, 2014.
- [4] S. V. Walle and T. Steenberghen, "Space and time related determinants of public transport use in trip chains," *Transportation Research Part A: Policy and Practice*, vol. 40, no. 2, pp. 151–162, 2006.

- [5] Z. Zhou, Y. Li, C. Meng, and H. P. Lu, "Analysis of travel demand based on a structural equation model," *Journal of Tsinghua University (Science and Technology)*, vol. 48, no. 5, pp. 879–882, 2008.
- [6] H. R. Zhu, "Research on sensitivity of public transport priority in central districts of Beijing," in *Proceedings of Urban and Rural Planning in the Perspective of Conservation Culture, Chinese Urban Planning Conference*, Dalian, China, 2008.
- [7] R. Zhang, *Mode Split Model and Its Application on Urban Passenger Transportation System [Master, thesis]*, Beijing Jiaotong University, Beijing, China, 2011.
- [8] C. Zhang, *Analysis on the Characteristics of Urban Rail Transit Passenger Flow [Ph.D. thesis]*, Southwest Jiaotong University, Chengdu, China, 2006.
- [9] S.-H. Cao, Z.-Z. Yuan, C.-Q. Zhang, and L. Zhao, "LOS classification for urban rail transit passages based on passenger perceptions," *Jiaotong Yunshu Xitong Gongcheng Yu Xinxil/ Journal of Transportation Systems Engineering and Information Technology*, vol. 9, no. 2, pp. 99–104, 2009.
- [10] S. H. Chen, Y. Y. Chen, and C. Y. Yin, "Calculation method for bus vehicle performance indexes based on IC data," *Traffic standardization*, no. 21, pp. 52–55, 2011.
- [11] J. Wa, *The Travel Characteristics Analysis of Passengers Based on the Bus Data [Master, thesis]*, South China University of Technology, Guangzhou, China, 2016.
- [12] C. Yang, J. Lu, W. Wang, and Q. Wang, "A study on the influencing factors of bicycle transportation based on individual mode choice," in *Transportation Systems Engineering and Information Technology*, vol. 7, pp. 131–136, 4 edition, 2007.
- [13] Q. W. Xu, *Analysis and Modeling of Individual Passenger Behavior in Urban Subway Hub [Master, thesis]*, Beijing Jiaotong University, Beijing, China, 2008.
- [14] S. Chen, C. Fang, and C. Tien, "Driving behaviour modelling system based on graph construction," *Transportation Research Part C: Emerging Technologies*, vol. 26, pp. 314–330, 2013.
- [15] K. Q. Yuan, Y. Deng, D. Y. Chen, B. Zhang, K. Lei, and Y. Shen, "Construction techniques and research development of medical knowledge graph," *Application Research of Computers*, vol. 35, no. 7, pp. 1–12, 2017.
- [16] X. Jiao, L. Wang, and F. Han, "Application model construction of mapping knowledge domains in the science and technology intelligence research," in *Knowledge, Learning & Management*, vol. 3, pp. 118–129, 2017.
- [17] C. Wang, "Extraction method of public transportation trip chain based on the individual travel data," *Transportation Systems Engineering and Information Technology*, vol. 17, no. 3, pp. 68–73, 2017.
- [18] R. L. Wang, *Public Transit Passenger Division and Analysis of Trip Characteristics Based on Smart Card Data [Master, thesis]*, Tongji University, Shanghai, China, 2014.

