

Research Article

Destination Estimation for Bus Passengers Based on Data Fusion

Wusheng Liu,¹ Qian Tan², and Lisheng Liu³

¹*Engineering Research Center of Catastrophic Prophylaxis and Treatment of Road & Traffic Safety of Ministry of Education, Changsha University of Science & Technology, Changsha 410114, China*

²*School of Traffic & Transportation Engineering, Central South University, Changsha 410075, China*

³*School of Urban Construction, Wuchang Institute of Technology, Wuhan 430223, China*

Correspondence should be addressed to Qian Tan; tq.helen@csu.edu.cn

Received 6 July 2020; Revised 21 August 2020; Accepted 6 September 2020; Published 16 September 2020

Academic Editor: Hussein Abulkasim

Copyright © 2020 Wusheng Liu 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.

The planning and operation of urban buses depend heavily on the time-varying origin-destination (OD) matrix for bus passengers. In most cities, however, only boarding information is recorded, while the alighting information is not available. This paper proposes a novel method to predict the destination of a single bus passenger based on bus smartcard data, metro smartcard data, and global positioning system (GPS) bus data. First, the attractiveness of each bus stop in a bus line was evaluated, considering the attractiveness of nearby metro stations. Then, the exploration and preferential return (EPR) model was employed to estimate the probability of a bus stop to be the alighting stop, i.e., the destination, of a passenger. The estimation result was obtained through a simulation based on the Monte Carlo (MC) algorithm. The effectiveness of our method was proved through a case study on the bus network in Shenzhen, China.

1. Introduction

The origin-destination (OD) estimation of bus passengers is essential to the network planning and operation of buses. Traditionally, the ODs of bus passengers were estimated by questionnaire surveys, which are small in sample size and low in precision. In recent years, the OD estimation of bus passengers has progressed rapidly, owing to the proliferation and use of geoinformation systems and smartcard techniques. GPS and bus IC card data are widely used. A bus IC card provides a large amount of data for the travel characteristic analysis of bus passengers, which has large amount of data, low cost, and high accuracy. However, it can only provide the relevant information of passengers' boarding stops. In order to obtain the passenger travel OD, it is necessary to calculate the alighting stops.

Many scholars have investigated the OD estimation problem. The early studies mainly derived the OD matrix of public transit passengers from the passenger flow at each station. For instance, Ben-Akiva et al. [1, 2] proposed an OD derivation method based on the survey data on ODs. After analyzing the passenger flow at each bus stop, Navick et al.

[3, 4] estimated the travel patterns of bus passengers and constructed the OD matrix of these passengers. Tsygelnitsky [5] predicted the destinations of public transit passengers under the assumption that all the passengers boarding the same type of public transit vehicle at the same station have the same probability to alight at another station.

Moreover, Li and Cassidy [6] designed an algorithm that does not need a seed matrix to estimate the ODs of public transit passengers. Based on boarding and alighting counts at each stop along the route, the designed algorithm deduces an OD matrix for the entire trip and forecasts the probabilities for passengers to board and alight at each stop along the route. These probabilities tend to remain fixed throughout the trip. Compared with Tsygelnitsky's prediction technique, Li and Cassidy's algorithm is highly suitable for general use. Using boarding and alighting counts, Li [7] developed an efficient statistical inference method for a closed-form OD matrix of a travel route: the Markov chain model was adopted to capture the correlations between matrix elements and reduce the number of unknown parameters; then, the unknown parameters of the Markov chain model were inferred by Bayesian analysis.

In practice, the OD estimation is often realized through iteration proportional fitting, in which the OD matrix is adjusted continuously based on survey data, and the passenger flow of each bus stop is obtained by adding up the row and column vectors of the OD matrix [8, 9]. This approach is efficient and easy to implement but costly in terms of labor and money. In addition, a very few cities have adopted an automatic passenger counter (APC) system to collect the information of bus passenger flows in an automatic manner.

Recently, new methods have emerged to estimate the destinations of bus passengers based on global positioning system (GPS) data and smartcard data. Using the APC data, Barry et al. [10] estimated the destinations of bus passengers in New York City based on the trip chain theory. Zhao [11] derived the destinations of passengers that transfer between metro lines or between metro and bus. Seaborn et al. [12] explored the time span for transfer between public transit modes and relied on the time span to identify multimode transfers. Hofmann and O'Mahony [13] determined the transfer stations according to the time difference between two card swipes.

In addition, based on GPS data, Giannotti et al. [14] researched the detailed trajectory of vehicles and the frequent mode of urban residents' travel, the prediction method of traffic intensive areas, and the description method of traffic congestion; Jiang et al. [15] found that human travel distance obeyed power-law distribution using GPS data. Based on the travel survey data, Garske et al. [16] studied the residents' travel in two cities with different economic development levels in China. Kölbl and Helbing [17] found that under different travel modes, people's travel distance follows a general distribution law. In [18], the daily travel logs of 230 volunteers in Frauenfeld, Switzerland, were analyzed. The author found that the travel distance of the group obeyed a power-law distribution with exponential truncation, which was very close to the empirical research results based on mobile phone data, and most individual travel distance does not conform to power-law distribution. Cui [19] deduced the boarding stations of passengers from the data collected by the automatic fare collection system and the automatic vehicle location system [20]. Farzin [21] analyzed the boarding stations of passengers in São Paulo, Brazil, referring to information of integrated circuit (IC) bus cards, GPS data, and bus stop data. Xu et al. [22] estimated the destinations of public transit passengers, in the light of travel distance distribution and bus stop features. Xu et al. [23] clustered smart card data before estimating the alighting stops of passengers.

The current research focuses on two aspects: Firstly, the main method is to analyze the distribution characteristics of passenger travel distance to estimate alighting stop of a single bus route. Secondly, OD was obtained with using the difference of swipe card time in passenger travel to analyze the transfer between different modes of transportation. In terms of individual travel characteristics, it is found that the individual average travel distance is significantly different, the individual visit frequency to the location also follows the power-law distribution, and the individuals with different average travel distance have high similarity in spatial motion location distribution [24].

Therefore, it is reasonable to use the distribution characteristics of travel distance to study the single bus stop without transfer with subway, but this study uses the mobile phone signaling data; the traffic mode includes walking and nonmotor vehicle. The travel distance is a continuous variable, which is from the origination to the destination, not the next stop. When passengers travel by bus, the travel distance is a discrete variable, and the next station may not be the destination, so the transfer must be considered. In addition, according to the habits of human travel activities, travel behavior tends to choose to return to places that have been visited more times in history, such as home and office [25]. For the passengers with a large number of historical travel records, it is more accurate to analyze the historical travel law to predict the next stop.

In general, the destinations of bus passengers on a single bus line are estimated from the trip chain of boarding stations and the historical smartcard data, or from the land use attributes of all bus stops. In this paper, the historical travel features of a single passenger and the attractiveness of the nearby metro stations of a bus stop are analyzed in detail; the alighting stop, i.e., the destination, of that passenger was estimated by using the exploration and preferential return (EPR) model [26] and the Monte Carlo (MC) algorithm; and we also try to estimate the destinations of bus passengers in the future under connected and autonomous vehicle environment [27–31].

2. Data Sources

The research data can be divided into five datasets: bus line data, bus smartcard data, bus GPS data, metro smartcard data, and road network data.

2.1. Bus Line Data. Bus line data contain the information of each bus stop on a bus line, including but not limited to coordinates, name, and bus line number. In total, the data were collected from 1,516 bus lines in Shenzhen.

2.2. Bus Smartcard Data. The bus smartcard data refer to the transaction information captured by the smart card fare collector, once a passenger swipes his/her smartcard upon boarding a bus. The data cover passenger identity (ID), boarding time, bus ID, and bus line number. Each bus smartcard carries a unique passenger ID that can be identified easily. In this research, the bus smartcard data were provided by ShenZhenTong, the largest public transit service provider in Shenzhen, and collected in the 21 days from October 11th to 31st, 2014. These data are for the three weeks after the National Day holiday. The three weeks include 18 working days and 4 nonworking days. The average number of card swipes per week is the largest in the whole year in Figure 1, which is representative.

2.3. Bus GPS Data. Each bus has a GPS tracker that records the bus position in real time. The bus ID and bus line

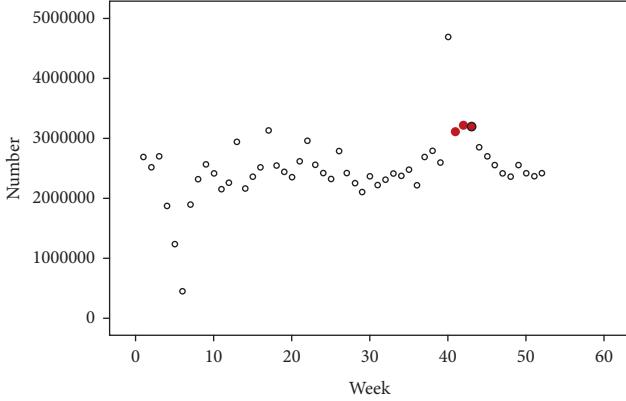


FIGURE 1: The average number of card swipes per week.

number recorded by the GPS tracker are unique, allowing us to match bus GPS data with bus smartcard number.

2.4. Metro Smartcard Data. Metro smartcard data include the name of boarding and alighting metro stations, passenger ID, and boarding and alighting time. In this research, the metro smartcard data were also provided by ShenZhenTong and collected in the same period.

2.5. Road Network Data. The road network data of Shenzhen (N: 22°46'–22°83'; E: 113°78'–114°59') was simplified into 21,115 sections.

3. Methodology

Based on the above data, this paper aims to estimate the destination for every single passenger boarding a bus line. The estimation process can be broken down into three tasks. The first task is to determine the arrival time and location of the bus at each stop of the bus line by matching bus line data with bus GPS data. The second task is to identify the boarding station of the passenger by matching GPS data with bus smartcard data. The third and final task is to estimate the destination of the passenger by using the EPR model, with metro smartcard data as the basis for exploration.

3.1. Data Matching. The bus smartcard only records the boarding time and vehicle of the passenger, but not the boarding location (station). To determine the boarding station, it is necessary to match bus smartcard data with bus GPS data, i.e., complete the first two tasks mentioned at the beginning of this section:

Task 1: matching bus line data with bus GPS data

In general, a bus line is composed of an upline and a downline; that is, a bus moves in one of the two opposite directions at a time. Considering the inevitable errors in bus GPS data and the close proximity between stops with similar names, it is impossible to match bus line data with bus GPS data based on the distance indicated by position feedbacks from the GPS trackers on the bus and the GPS device at each stop. Instead, the

moving direction of the bus should be identified in the light of its trajectory from the GPS, and then the locations be matched based on the direction. The matching process is illustrated in Figure 2, where L_u and L_d are the upline and downline, respectively.

Let S and S' be the departure stations of L_u and L_d , respectively. Ten consecutive tracking points were chosen from the GPS data of a bus M_1 that operates along line L . Then, the distance from each tracking point to S or S' was calculated. If the distance of these points increases from station S (S'), then the bus is on the up (down) line, that is, $L_{M_1} = L_u$ (L_d). Next, the bus GPS data were matched with stops along the right direction. The location matching was deemed successful, when the GPS location fell within 100 m of the boarding stop S_0 . The matching time was taken as the arrival time at the stop. In this way, the arrival time of bus M_1 at all stops along line L can be obtained as T_i , $i = 1, 2, 3, \dots, n$.

Task 2: matching bus smartcard data with bus GPS data Suppose a passenger P_1 boards M_1 at station S_i and swipes his/her smartcard at time T_p . Then, T_p was taken as the boarding time of passenger P_1 . Comparing T_p with T_i , the boarding stop S_0 (as shown in Figure 3, i.e., origin, of passenger P_1 can be identified as the station S_i with the lowest time difference from T_p . Any of the stops following the origin could be the destination of the passenger.

3.2. Destination Estimation of a Single Bus Passenger. The destination of a bus passenger depends on various factors, namely, travel distance and attractiveness of each stop. The latter, referring to the possibility that a passenger alights at the stop, encompasses the attractiveness of the stop itself and extra attractiveness. Here, extra attractiveness is measured by the attractiveness of metro stations near the stop because buses often serve as the feeders for metro in the large and dense transit network in Shenzhen. The two mechanisms of the EPR model and the MC algorithm are introduced to estimate the destination of a single bus passenger.

3.2.1. Explore Mechanism. The explore mechanism applies to the scenario where there is no card swiping record at the previous boarding stops of the passenger. For such a passenger, the possibility that he/she alights at a stop is the probability P_{ij} for the stop to be the destination. According to the definition and composition of stop attractiveness, the probability P_{ij} covers two items, namely, the probability F_{ij} arising from the attractiveness of the stop itself and the probability D_{ij} stemming from the attractiveness of nearby metro stations.

The travel distance can be approximated by the number of stops passed by the bus. Without considering probability D_{ij} , the number of stops passed by a bus follows the Poisson distribution. In this case, the probability F_{ij} can be expressed as

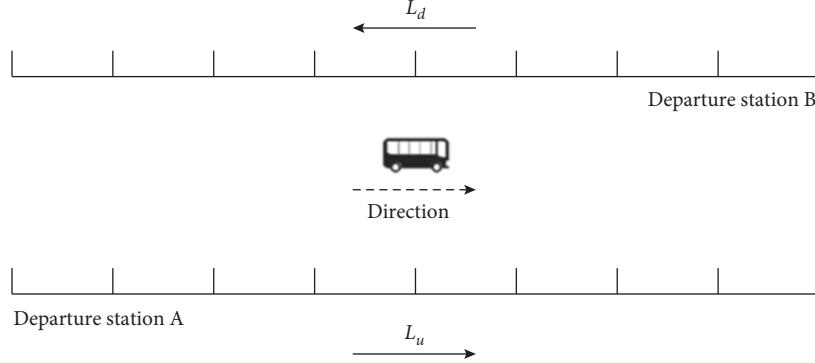


FIGURE 2: The matching between bus line data and bus GPS data.

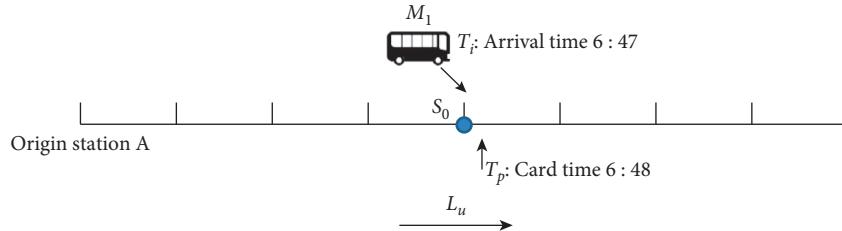


FIGURE 3: The matching between bus smartcard data and bus GPS data.

$$F_{ij} = \begin{cases} \frac{e^{-\lambda} \lambda^{(j-i)}}{(j-i)!}, & i < j, \\ 0, & i \geq j, \end{cases} \quad i, j = 1, 2, \dots \quad (1)$$

where i and j are the serial numbers of the boarding and alighting stops, respectively (the serial numbers were assigned from the departure station of the bus line); n is the number of stops on the bus line; and λ is the mean number of stops in a bus line (λ was set to 10 for Shenzhen [32]). If the number of remaining stops after the boarding stop is fewer than λ , then $\lambda = n - i$.

The Poisson distribution can be normalized as

$$F_{ij}^* = \frac{(e^{-\lambda} \lambda^{(j-i)}) / (j-i)!}{\sum_{i=i+1}^m (e^{-\lambda} \lambda^{(j-i)}) / (j-i)!}. \quad (2)$$

However, another determinant factor of bus passengers' destination is the nature of land use. If there are shopping malls and entertainment sites nearby, the attraction is greater; especially in the stops near the transportation hub, the number of people boarding and alighting the bus is the largest. Due to the round-trip characteristics of residents' public transportation, the volume of generation and attraction of stops are basically in a balanced state; that is, the more the people get on the bus, the more the people get off the bus. The attraction intensity of each stop is calculated by counting the total number of passengers at each stop from the judgment of the previous stop.

The metro records near the boarding stop were statistically analyzed. The boarding stop was considered near a metro station, if their distance is smaller than 1,000 m. This distance may lead to a time difference in the metro records of the transfer passenger. Here, the time difference is set to 30 min; that is, the metro records generated 30 min after the passenger swiped his/her smartcard on the target bus are counted. The records at the nearby metro stations were used to measure the attractiveness of the bus stop. If there is no metro station nearby, the attractiveness of the bus stop was set to zero.

Then, the probability D_{ij} of the bus stop j can be computed as

$$D_{ij} = \frac{d_j^*}{\sum_{j=i+1}^m d_j^*}, \quad (3)$$

where d_j^* is the record at the nearby metro stations generated 30 min after the passenger swiped his/her smartcard on the target bus.

According to the literature [33], based on more than 230000 pieces of data analysis, research on the transfer time between conventional bus and subway, and analyzing the transfer data with the interval of 5 minutes, it can be found that most of the transfer behavior takes less than 20 minutes and only 2–4% of the total transfer amount exceeds 20 minutes (Figure 4). In order to ensure the integrity of transfer sample identification, 30 is selected as the transfer time threshold.

According to the agglomeration effect of public transit stations, the attractiveness of a metro station decreases with

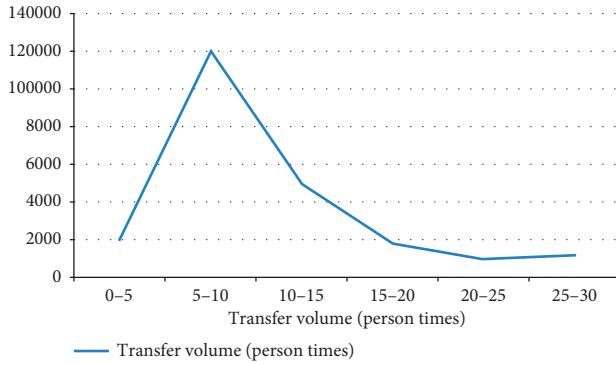


FIGURE 4: Transfer time distribution.

its distance to the bus stop. Hence, the D_{ij} value obeys the exponential distribution:

$$E_j = E_0 e^{-s}, \quad (4)$$

where E_j is the intensity of the agglomeration effect; $E_0 E_j = 1 E_0$ is the peak agglomeration effect; and s is the distance from a metro station to the bus stop.

Then, the d_j^* value can be obtained by

$$d_j^* = E_j \times d_j. \quad (5)$$

On this basis, the probability P_{ij} a stop to be the destination for a single bus passenger can be calculated by

$$P_{ij} = \frac{F_{ij}^* + \alpha D_{ij}}{\sum \alpha D_{ij} + \sum F_{ij}^*}, \quad (6)$$

where α is the coefficient of the attractiveness of nearby metro stations ($0 < \alpha \leq 1$), i.e., the $0 < \alpha \leq 1$ weight of D_{ij} , and the weight of F_{ij}^* is 1. The value of α is positively correlated with the proportion of passengers taking metro instead of bus.

3.2.2. Preferential Return Mechanism. Based on the research results of the literature [25], it is found that in terms of residents' travel destination, when the historical data of residents' travel increases with increase of time, the number of residents visiting new places follows $s(t) = t^{-\mu}$, $\mu \sim 0.6$; the frequency of residents' visit to the place (k) is subject to $f_k = k^{-\xi}$, $\xi \sim 1.2$; and the authors pointed out that the accuracy of the CTRW model is not good. The main idea of the exploratory regression model is that individuals return to the previously visited places with the probability $(\rho s^{-\gamma})$ of exploring the location, and the probability $(1 - \rho s^{-\gamma})$ of visiting a place is directly proportional to the probability of individuals being found in the location, as shown in Figure 5.

Next, the preferential return mechanism was employed to predict the destination under the scenario that there is no card swiping record on the bus line boarded by the passenger. In general, passengers prefer to alight at frequently visited places, such as home and workplace. Thus, the basic idea of the preferential return mechanism is that passengers tend to alight at stops with more historical card swiping records. In other words, the probability for a stop to be the

destination is proportional to the historical record count of the passenger at that stop.

To eliminate the interference of stops with similar names, the smartcard records within 100 m were counted as the records of one stop, where 100 m is the return range. The stop with many historical records has a high probability of being returned, which is directly proportional to the number of historical records.

Based on the historical records of a passenger, the probability f_i of a stop to be the destination can be described as

$$f_i = \frac{m_i}{\sum_{i=1}^n m_i}, \quad (7)$$

where i is the serial number of stops following the boarding stop; m_i is the number of historical records of stop i ; and n is the number of stops with a probability of being returned.

3.2.3. MC Algorithm. Finally, the destination of a single bus passenger was predicted by using the MC algorithm. Based on probability theory, the MC algorithm relies on a random probability model to approximate the probability through simulation and statistical testing on random variables. As shown in Figure 6, the MC algorithm is implemented in the following steps [20]:

Step 1: construct and describe the probability process as formulae (2) and (6)

Step 2: determine the sample size and samples from the probability distribution, and set the number of simulations to 1,000 for each passenger

Step 3: confirm the estimation, i.e., the alighting stop

The estimated destination is the stop with the largest number of occurrences in the 1,000 simulations. If a passenger boards at stop S_0 of bus line L , then the stops after S_0 are numbered as S_1, S_2, \dots, S_n in turn. The number of occurrences of each stop in the 1,000 simulations is denoted as x_1, x_2, \dots, x_n , respectively, and the estimated destination as S_i , where $x_i = \max(x_1, x_2, \dots, x_n)$.

4. Case Study

4.1. Estimation of the Boarding Station. The proposed method was applied to predict the destination of every single bus passenger in the 1,516 bus lines across Shenzhen. The smartcard bus data, bus GPS data, and smartcard metro data were collected by ShenZhenTong in 21 days from October 11th to 31st, 2014, including 14,109 trip records for 23 passengers, as well as the trajectories and stop locations of the 1,516 bus lines.

During the 21 days from October 11 to 31, 2014, the total number of card swipes was 268623 (excluding subway passengers). There were 113625 passengers. The method of random sampling was used to collect the sample, and the sample size is calculated as shown in Table 1.

Sample size n is shown below:

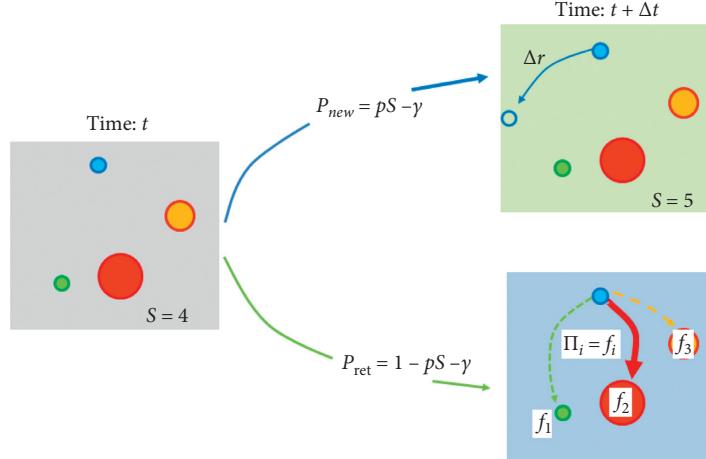


FIGURE 5: Preferential return mechanism [25].

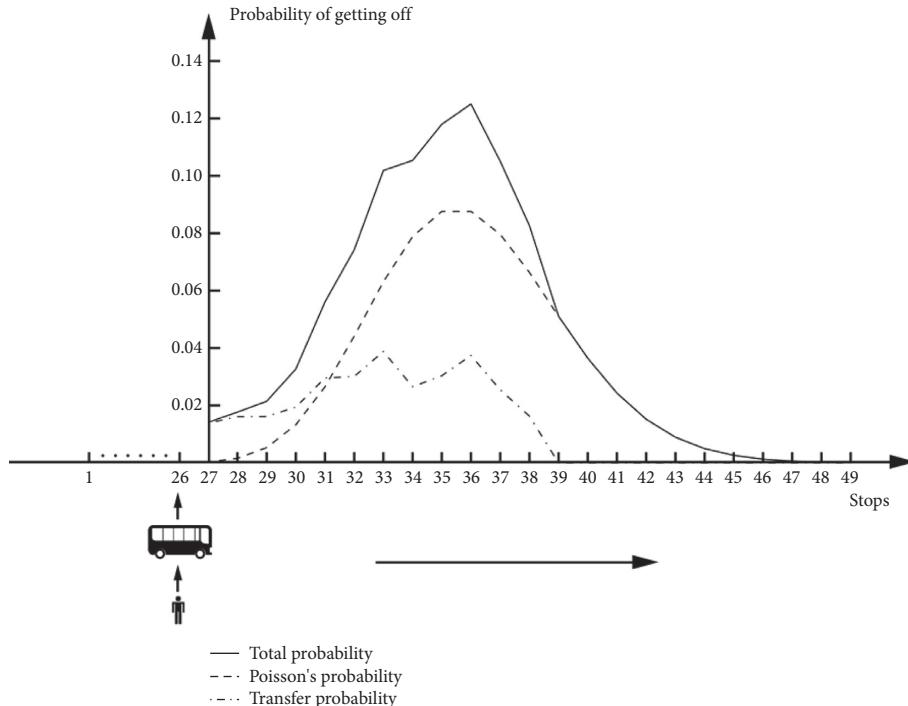


FIGURE 6: The MC-based estimation of destination.

TABLE 1: Sample-related parameters.

Sample characteristics	Standard deviation of total trips σ	Expected error E	Confidence interval (%)
Average number of trips per person	31.5	5.0	95

$$n = \frac{(z_{(\alpha/2)})^2 \sigma^2}{E^2} = \frac{1.96^2 \times 31.5^2}{5^2} = 152. \quad (8)$$

As mentioned before, the road network data of Shenzhen (N: $22^{\circ}46' - 22^{\circ}83'$; E: $113^{\circ}78' - 114^{\circ}59'$) was simplified into 21,115 sections.

For simplicity, the downtown of Shenzhen was divided into 74×162 zones, each of which is 555 m by latitude and 515 m by longitude. Then, the smartcard records, GPS bus data, and stop locations were fused to obtain spatial distribution of the boarding stops for bus passengers in Shenzhen during the five days from October 27th to 31st,

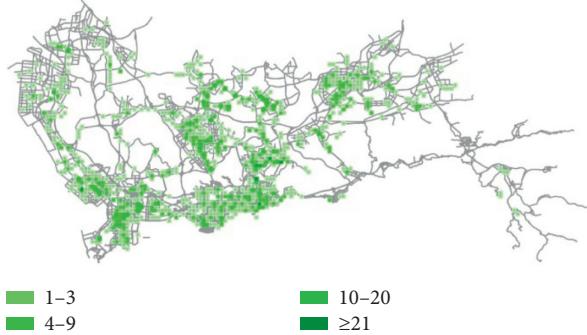


FIGURE 7: Spatial distribution of boarding stops.

2014. As shown in Figure 7, most passengers board buses at the center of the city; that is, the bus stops at the central area of Shenzhen have relatively high attractiveness.

To differentiate the estimated results of exploration mechanism from those of the preferential return mechanism, the boarding stops of the 293 bus passengers with and without historical records are displayed in Figures 8 and 9, respectively.

4.2. Destination Estimation

4.2.1. Determination of α Value. The value of α , that is, the weight of D_{ij} or the coefficient of the attractiveness of nearby metro stations, was set to 0.5 and 0.7 during the destination estimation. The results show that the α value has a limit effect on the estimation. Since the ratio of metro trips to bus trips in Shenzhen is 3:7, the value of α was set to 0.7.

4.2.2. Destination Estimation. Our method was adopted to estimate the alighting stop, i.e., destination, of every single bus passenger from October 27th to 31st, 2015. The estimated destinations are displayed as the heat map in Figure 10. It can be seen that the destinations concentrated in the central area of the city, revealing a correlation between boarding and alighting locations.

Next, the exploration mechanism and the preferential return mechanism were separately adopted to estimate the destinations of each of the 293 passengers without historical records on the bus line, with $\alpha = 0.7$. The estimated results of the two mechanisms are presented in Figures 11 and 12, respectively.

4.3. OD Distributions in Different Periods. Based on the above estimation, the distribution of origins (boarding stops) and destinations (alighting stops) was illustrated for different periods of the day (Figures 13–18).

As shown in Figures 13–18, the origin distribution in the morning peak is consistent with the destination distribution in the evening peak, and both origins and destinations are clustered in residential areas. It is learned that the bus trips of passengers in the morning and evening peaks leave from and return to their homes, respectively.

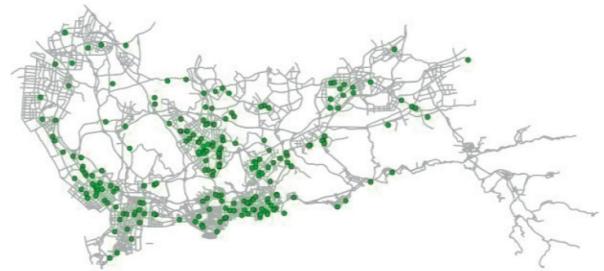


FIGURE 8: Spatial distribution of boarding stops for bus passengers with historical records.

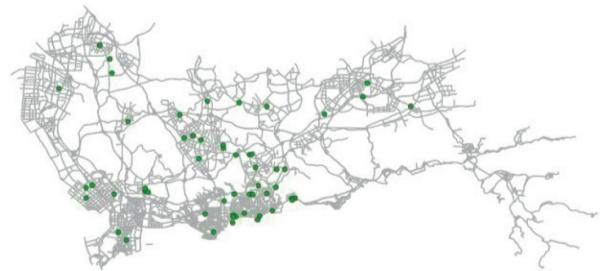


FIGURE 9: Spatial distribution of boarding stops for bus passengers without historical records.

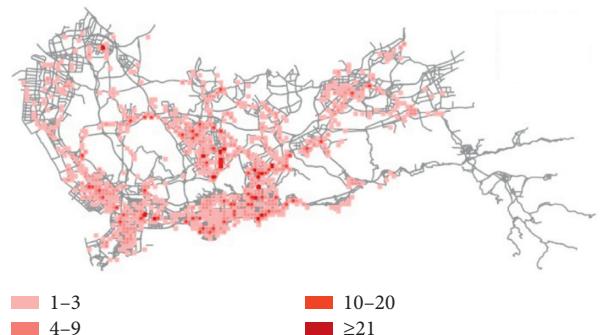


FIGURE 10: Heat map of destinations.

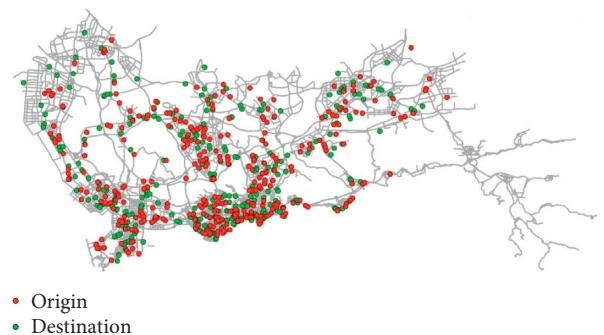
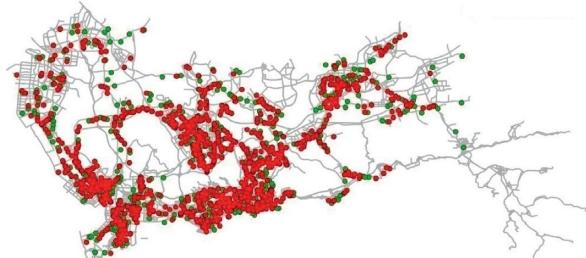
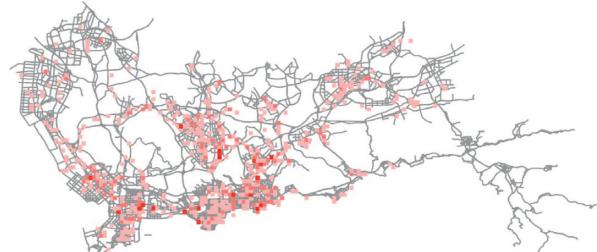


FIGURE 11: Spatial distribution of destinations with exploration mechanism.



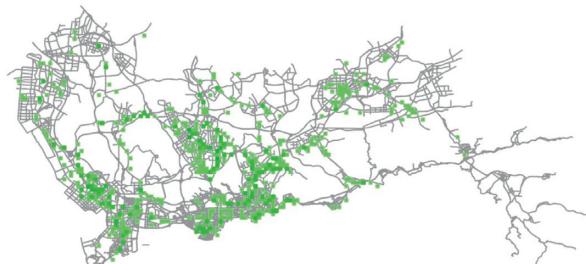
- Origin
- Destination

FIGURE 12: Spatial distribution of destinations with preferential return mechanism.



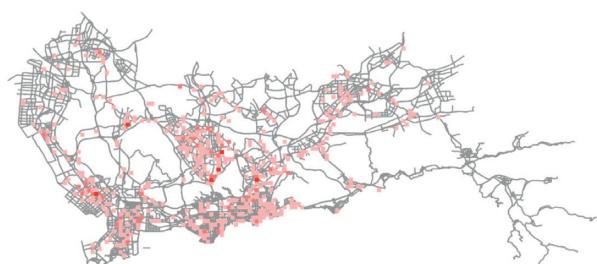
- | | |
|-----|-----------|
| 1–3 | 10–20 |
| 4–9 | ≥ 21 |

FIGURE 16: Destination distribution in the morning peak.



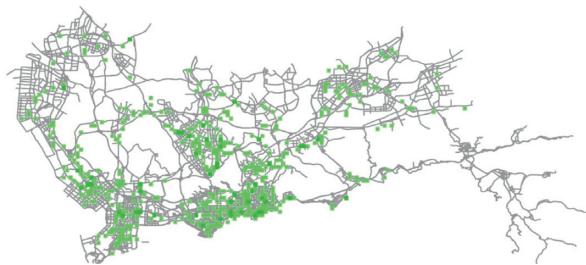
- | | |
|-----|-----------|
| 1–3 | 10–20 |
| 4–9 | ≥ 21 |

FIGURE 13: Origin distribution in the morning peak.



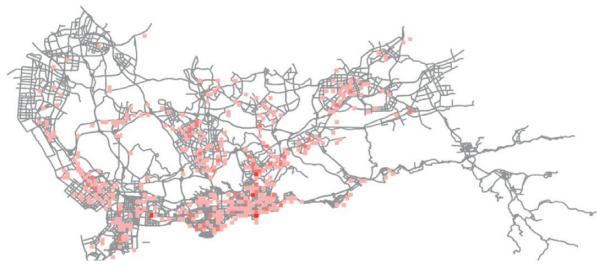
- | | |
|-----|-----------|
| 1–3 | 10–20 |
| 4–9 | ≥ 21 |

FIGURE 17: Destination distribution in the evening peak.



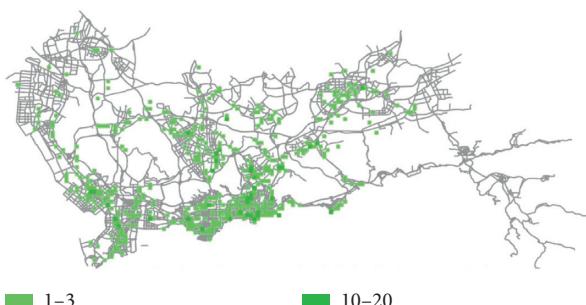
- | | |
|-----|-----------|
| 1–3 | 10–20 |
| 4–9 | ≥ 21 |

FIGURE 14: Origin distribution in the evening peak.



- | | |
|-----|-----------|
| 1–3 | 10–20 |
| 4–9 | ≥ 21 |

FIGURE 18: Destination distribution in off-peak hours.



- | | |
|-----|-----------|
| 1–3 | 10–20 |
| 4–9 | ≥ 21 |

FIGURE 15: Origin distribution in off-peak hours.

Besides, the destinations in the morning peak mostly fell in commercial and school areas, indicating that most passengers go to work or school in the morning. The origins in the evening peak was slightly scattered, yet mainly from commercial and school areas, which are the main destinations for business and school travels.

Finally, the origins and destinations were relatively decentralized in the off-peak hours, suggesting that a certain portion of the travels are nondomestic.

Through our data analysis, 3,023 destinations (82.8%) were derived by the preferential return mechanism from the data of 3,651 passengers, while 62 (17.2%) were derived by the exploration mechanism. It can be speculated that urban residents tend to return to places they have visited before.

With the growing number of trips, the residents are more likely to prefer the historical locations over new locations. This agrees with the rule of preferential return. Our estimation shows that the travels of bus passengers concentrated in the morning and evening peaks: going downtown in the morning peak and returning home in the evening peak. This is clearly in line with the situation of urban commuters.

5. Conclusions

This paper fully integrates bus line data, bus smartcard data, bus GPS data, and metro smartcard data, with road network data, and introduces the EPR model and MC algorithm to estimate the alighting stops of bus passengers in Shenzhen. Firstly, the boarding location and time of every single passenger were estimated based on the integrated data. Then, the alighting station of the passenger was predicted under the exploration mechanism and the preferential return mechanism, which is based on the features of travel activities. During the prediction, the metro smartcard data were innovatively employed to evaluate the extra attractiveness of each bus stop. Considering the features of historical trips and bus-metro transfer, the proposed method was found to effectively solve the destination estimation problem through a case study. The future research will further explore the multimode traffic and OD estimation of multitransfer trips; based on the results of this study, we can obtain the origination and the destination of passengers and estimate the travel OD combined with multimode traffic and multiple transfer, so as to obtain the relevant characteristics of passengers' travel, including travel time, travel OD, travel distance, and station passenger flow; the travel distance and station passenger flow can provide more accurate data support for urban bus dispatching and schedule more reasonable departure time and interval of peak and flat peak vehicles. In addition, the passenger flow and travel OD can provide the basis for public transportation network adjustment.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors thank Shenzhen Urban Transport Planning Center for providing part of our research data and also to other scholars and research teams for their scientific contributions. This project was supported by Open Fund of Engineering Research Center of Catastrophic Prophylaxis and Treatment of Road & Traffic Safety of Ministry of Education (Grant no. kfj180401, Changsha University of Science & Technology), Natural Science Foundation of Hunan Province, China (Grant no. 2019JJ40306), and the National Natural Science Foundation of China (Grant no. 61773077).

References

- [1] M. E. Ben-Akiva, "Methods to combine different data sources and estimate origin-destination matrices," in *Proceedings of the 10th international symposium on transportation and traffic theory*, Elsevier, Cambridge, UK, pp. 459–481, July 1987.
- [2] M. Ben-Akiva, P. Macke, and P. Hsu, "Alternative methods to estimate route-level trip tables and expand on-board surveys," *Transportation Research Record*, vol. 1037, p. 1, 1985.
- [3] D. S. Navick and P. G. Furth, "Distance based model for estimation a bus route OD matrix," *Transportation Research Record*, vol. 1433, p. 16, 1994.
- [4] H. L. Dou, H. D. Liu, and X. G. Yang, "OD matrix estimation method of public transportation flow based on passenger boarding and alighting," *Computer and Communications*, vol. 25, no. 135, p. 79, 2007.
- [5] S. Tsyglinsky, "Simplified methods for transportation planning," Massachusetts Institute of technology, Cambridge, UK, 1977.
- [6] Y. Li and M. J. Cassidy, "A generalized and efficient algorithm for estimating transit route ODs from passenger counts," *Transportation Research Part B: Methodological*, vol. 41, no. 1, pp. 114–125, 2007.
- [7] B. Li, "Markov models for Bayesian analysis about transit route origin-destination matrices," *Transportation Research Part B: Methodological*, vol. 43, no. 3, pp. 301–310, 2009.
- [8] M. Bacharach, "Biproportional matrices and input-output change," Cambridge University Press, London, UK, 1970.
- [9] B. Beirami and M. Mokhtarzade, "Spatial-spectral random patches network for classification of hyperspectral images," *Traitemet du Signal*, vol. 36, no. 5, p. 399, 2019.
- [10] J. J. Barry, R. Newhouser, A. Rahbee, and S. Sayeda, "Origin and destination estimation in New York City with automated fare system data," *Transportation Research Record*, vol. 1817, no. 1, pp. 183–187, 2002.
- [11] J. Zhao, "The planning and analysis implications of automated data collection systems: rail transit OD matrix inference and path choice modeling examples," Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA, USA, 2004.
- [12] C. Seaborn, J. Attanucci, and N. H. Wilson, "Using smart card fare payment data to analyze multi-modal public transport journeys in London," 2009.
- [13] M. Hofmann and M. O'Mahony, "Transfer journey identification and analyses from electronic fare collection data," in *Proceedings of the 2005 IEEE Intelligent Transportation Systems*, pp. 34–39, Vienna, Austria, September 2005.
- [14] F. Giannotti, M. Nanni, D. Pedreschi et al., "Unveiling the complexity of human mobility by querying and mining massive trajectory data," *The VLDB Journal*, vol. 20, no. 5, p. 695, 2011.
- [15] B. Jiang, J. Yin, and S. Zhao, "Charactering the human mobility pattern in a large street network," *Physical Review E*, vol. 80, Article ID 021136, 2009.
- [16] T. Garske, H. Yu, and Z. Peng, "Travel patterns in China," *PLoS One*, vol. 6, Article ID e16364, 2011.
- [17] R. Kölbl and D. Helbing, "Energy laws in human travel behavior," *New Journal of Physics*, vol. 5, p. 1, 2003.
- [18] X. Yan, X. Han, and B. Wang, "Diversity of individual mobility patterns and emergence of aggregated scaling laws," *Scientific Reports*, vol. 3, Article ID 2678, 2013.
- [19] A. Cui, "Bus passenger origin-destination matrix estimation using automated data collection systems," Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA, USA, 2006.

- [20] X. Song, S. Gao, and C. Chen, "A novel vehicle feature extraction algorithm based on wavelet moment," *Traitemen du Signal*, vol. 35, no. 3-4, p. 223, 2018.
- [21] J. M. Farzin, "Constructing an automated bus origin-destination matrix using farecard and global positioning system data in São Paulo, Brazil," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2072, no. 1, pp. 30–37, 2008.
- [22] W. Y. Xu, C. Y. Deng, and B. Y. Liu, "Approach on public traffic passenger flow statistics based on IC data," *China Journal of Highway and Transport*, vol. 26, no. 5, pp. 158–163, 2013.
- [23] J. M. Xu, W. H. Xiong, and F. You, "The method of signals-line public transportation OD base on GPS and intelligent card," *Control & Automation*, vol. 24, no. 8, pp. 221–223, 2008.
- [24] D. Brockmann, I. Hufnagel, and T. Geisel, "The scaling laws of human travel," *Nature*, vol. 439, no. 7075, p. 462, 2006.
- [25] P. Wang, Z. R. Huang, and G. Hang, "Transportation engineering in the big data era," *Journal of University of Electronic Science and Technology of China*, vol. 42, p. 806, 2013.
- [26] C. Song, T. Koren, P. Wang, and A.-L. Barabási, "Modelling the scaling properties of human mobility," *Nature Physics*, vol. 6, no. 10, pp. 818–823, 2010.
- [27] W. Wu, L. Huang, and R. Du, "Simultaneous optimization of vehicle arrival time and signal timings within a connected vehicle environment," *Sensors*, vol. 20, no. 1, p. 191, 2020.
- [28] W. Wu, F. Zhang, W. Liu, and G. Lodewijks, "Modelling the traffic in a mixed network with autonomous-driving expressways and non-autonomous local streets," *Transportation Research Part E: Logistics and Transportation Review*, vol. 134, Article ID 101855, 2020.
- [29] S. Veeramalla and V. Talarji, "Estimation of neural sources from EEG measurements using sequential Monte Carlo method," *Ingénierie des Systèmes d'Information*, vol. 24, no. 4, p. 411, 2019.
- [30] S. Ourrad, Y. Houmadi, A. Ziadi, S. Mamoune, and A. Lousdad, "Probabilistic analysis for estimating the hydrogen desorption time from steel wire rods using Monte Carlo simulation," *Annales de Chimie-Science des Matériaux*, vol. 43, no. 4, p. 241, 2019.
- [31] H. M. Reda and B. Abdelylah, "Numerical investigation and solar flux distribution analysis of parabolic trough solar collector by adding secondary reflector," *Instrumentation Mesure Métrologie*, vol. 18, no. 3, p. 275, 2019.
- [32] Shenzhen Traffic and Transport Committee Integrated traffic and transportation system planning of Shenzhen (2013-2030). 2013.
- [33] N. N. Guan and B. Wang, "Characteristics of transfer passenger flow between metro and conventional bus based on IC data in Chengdu city," *China Metros*, vol. 2018148 pages, 2018.