Research Article

Identifying the Service Areas and Travel Demand of the Commuter Customized Bus Based on Mobile Phone Signaling Data

Jingyuan Wang[1] **and Meng Zhang** [1,2]

[1]*College of Civil and Transportation Engineering, Shenzhen University, Shenzhen 518060, China*
[2]*Guangdong Planning and Designing Institute of Telecommunications, Guangzhou, Guangdong 510630, China*

Correspondence should be addressed to Meng Zhang; zhangmeng2018@email.szu.edu.cn

In recent years, customized bus (CB), as a complementary form of urban public transport, can reduce residents' travel costs, alleviate urban traffic congestion, reduce vehicle exhaust emissions, and contribute to the sustainable development of society. At present, customized bus travel demand information collection method is passive. There exist disadvantages such as the amount of information obtained is less, the access method is relatively single, and more potential travel demands cannot be met. This study aims to combine mobile phone signaling data, point of interest (POI) data, and secondary property price data to propose a method for identifying the service areas of commuter CB and travel demand. Firstly, mobile phone signaling data is preprocessed to identify the commuter's location of employment and residence. Based on this, the time-space potential model for commuter CB is proposed. Secondly, objective factors affecting commuters' choice to take commuter CB are used as model input variables. Logistic regression models are applied to estimate the probability of the grids being used as commuter CB service areas and the probability of the existence of potential travel demand in the grids and, further, to dig into the time-space distribution characteristics of people with potential demand for CB travel and analyze the distribution of high hotspot service areas. Finally, the analysis is carried out with practical cases and three lines are used as examples. The results show that the operating companies are profitable without government subsidies, which confirms the effectiveness of the method proposed in this paper in practical applications.

1. Introduction

As a new innovative public transport mode, the CB advocates energy saving and emission reduction, green travel, alleviating urban traffic congestion, and providing people with high-quality travel services in a "point-to-point" way [1, 2]. CB originated from the idea of "car-sharing." It was introduced in 1948 by the organization "Sefage" in Sweden to save transportation costs for families who did not own a car [3]. Travel demand is an important part of customized bus route planning. Before most scholars study the route planning framework, they need to analyze the travel demand initially. K Tsubouchi et al. [4] applied the Internet and big data to develop a demand-responsive bus system that could be adapted to different city types. Qiu et al. [5] investigated a method to improve the performance of flexible route buses in an operational environment with uncertain travel demand. Scott et al. [6] researched both 'point-to-point' and 'round-trip' modes in London and predicted future demand for customized buses in London. ANand Lo [7] proposed a two-stage solution algorithm, compared to the traditional robustness formulation to determine the service with reliability using a two-stage formulation. Liu et al. [8] proposed a new commuter minibus transit system with on-demand interaction. The authors evaluated and compared the performance of CB, PC, and conventional public transportation systems through travel cost, travel time, and fuel consumption. Lyu et al. [9] proposed a CB-Planner method for a bus line planning framework with multiple travel data sources and designed a heuristic solution framework.

China's CB development started late and is still in the development stage. Zhong et al. [10] collected passenger travel demand information through online questionnaires and a mobile phone app and identified a suitable passenger

flow catchment area division method. By considering the station traffic volume and regional capacity allocation, a suitable regional clustering method for passenger flow distribution is established. Cheng et al. [11] used the data from the public bus smart card to mine potential CB demand points. Yu et al. [12] planned CB stops and routes based on large amounts of demands data. Liu et al. [13] proposed a visual analysis method. They evaluated the actual, dynamically changing travel demand and planned the routes for the nighttime CB system. The reliability of the method was verified with cases.

At present, many scholars mainly research line optimization, station location, and price strategy and have achieved certain research achievements [14–16]. And the research on commuter CB travel demand is rather inadequate. Existing ways of collecting information on CB travel demand are mainly through online collection (e.g., Ma et al. proposed a framework of CB methods based on online questionnaires to obtain travel demand [17]) or through offline questionnaires in some large residential areas, commercial areas, transportation hubs, and other areas (e.g., Li et al. used RP and SP questionnaires to research the factors of influencing the potential travel demand for CB in Shanghai, China [18]). However, this passive way of collecting travel demand information is time-consuming and costly. In addition, due to the incomplete coverage and low audience level of the current CB travel demand information collection, the mining of the potential commuter CB travel demand population is neglected. Only by collecting data online or offline for a certain region, it is inevitable that the data collected for the study of travel demand is not large enough and the coverage is not extensive. There are more potential travel demands that cannot be met.

In view of the existing problems and combined with big data processing technology, this paper proposes commuter CB service areas and travel demand identification method based on mobile phone signaling data. With the following main contributions: (1) Combining mobile phone signaling data and using big data processing technology, the distribution characteristics of commuters' workplace and residence are identified. Based on the above, a time-space potential model of commuter CB travel is established and an algorithm is designed to solve it. (2) Using the unit grid as the fundamental unit, we choose the factors affecting passengers' choice of the commuter CB as the input parameters of the model. The logistic regression model is constructed and solved by SPSS software, to study the time-space distribution characteristics of people with potential commuter CB travel demand and to further identify the service areas of commuter CB and travel demand.

The rest of the paper is organized as follows. In Section 2, a brief description of the data types used in the paper is given. In Section 3, the commuter CB service areas and travel demand identification method are proposed. The central city of Chongqing, China, is used as a case study for demonstration in Section 4. The main findings of the paper are briefly summarized, and further perspectives on the following research on CB travel demand are discussed in Section 5.

2. Data Description

The data used in this paper involve three parts: mobile phone signaling data, POI data of rail stations and bus stops, and data of secondary housing prices around where commuters reside.

(i) Mobile phone signaling data: it is provided by the operator of China Unicom in Chongqing, China. It has covered 38 districts and counties in the city for mobile phone signaling monitoring, with signaling collection interval of 30–60 min. The average number of daily subscribers is 4.7 million. The average number of valid signaling data records for a single user is 26. In this paper, about 43 million data pieces of China Unicom in August 2019 are selected as the research data to identify the space-time distribution characteristics of commuters' occupation and residence. And 175,794 users from 7:00 a.m. to 9:00 a.m. on a working day in August are chosen as the research data for potential travel demand mining.

(ii) POI data of rail stations and bus stations: the POI data of the study area including 10,780 bus stations and 158 rail stations are crawled in Python programming language by retrieving the Gaode API interface. The POI attributes information included station ID, longitude, and latitude.

(iii) Secondary house prices data: by crawling the second-hand house prices on the websites of 58 TongCheng and LianJia in China, we obtain the name of each community, convert it to latitude and longitude coordinates, and obtain its spatial geographic information. The mean value of the secondary house price near the commuter's residence is used as the input parameter of the model, and this feature is used to represent the income of the commuter.

3. Identification Method of Service Areas and Travel Demand

In the process of generating mobile phone signaling data, the natural environment, interference from human factors, and other conditions can lead to error in the location of cellular cells, and there may be missing data and duplication. At first, the abnormal data are cleaned, and on this basis, the origin (O) and destination (D) of commuters in the study area are identified using the training method proposed in [19]. The characteristics of commuters' occupational and residential distribution are obtained. Based on the time-space distribution characteristics of commuter travelers' occupations and residences, a time-space potential model of commuter CB is established. We considered the influence factors as the input parameters of the model and established logistic regression model. We use the model to predict the study area and select the areas that meet the conditions as the commuter CB service area. Based on this, we further identify the potential commuter CB travel demand population.

3.1. Time-Space Potential Model. In this paper, based on mobile phone signaling data, the travel regularity of commuters, the similarity of travel time and spatial distribution of work and residence, and the possibility of taking commuter CB in time-space distribution are comprehensively considered. Based on the shared travel model framework, the distribution characteristics in two dimensions of time and space are considered, and based on the literature [20], the time-space potential model of commuter CB is proposed.

The model takes commuter travelers as the research object. We take the time difference between commuters leaving their places of residence and the distance difference between commuters' places of work and residence as independent variables. Due to the difference between time and distance units, maximum-minimum normalization is used to convert them into dimensionless expressions and introduce weighting factors. The objective function is to calculate the value of time-space potential between commuters. The model takes into account the shorter time difference between commuters in terms of travel time and the smaller distance between commuters' residence and workplace in the spatial dimension. To a certain extent, it indicates the greater potential of commuters who can travel by the same transportation mode. Therefore, when certain conditions are met, it is considered that there is a potential similar travel demand between commuters in both temporal and spatial dimensions. The formula of the model is defined as

$$\text{TPV}(i,j) = \alpha \left[\frac{T(i,j) - \min(T)}{\max(T) - \min(T)} \right] + \gamma \left[\frac{S(i,j) - \min(S)}{\max(S) - \min(S)} \right]$$
$$+ \lambda \left[\frac{L(i,j) - \min(L)}{\max(L) - \min(L)} \right].$$

(1)

Equation (1) constraint is

$$T(i,j) = \frac{t(i,j)}{\Delta T},$$
$$t(i,j) = t_i - t_j,$$
$$T < \delta,$$
$$S(i,j) < \varepsilon,$$
$$L(i,j) < \varepsilon,$$

(2)

where $\text{TPV}(i,j)$ denotes the time-space potential between the commuter and the commuter, and the magnitude of the value indicates the likelihood that the commuter will travel in time and space by commuter CB. i,j are commuters. ΔT is time period of study. $S(i,j)$ denotes the difference in distance between commuter i and the place of residence of j. $L(i,j)$ denotes the difference in distance between commuter i and the place of job of j. $t(i,j)$ denotes the time difference between commuters i and j when leaving their place of residence. S is the sets composed by $S(i,j)$. L are the sets composed by $L(i,j)$. T are the sets composed by $t(i,j)/\Delta T$, and ε is the distance threshold, which takes the value of

300–500 m in general. δ is the time threshold. α, β, γ are weighting factors.

According to equation (1), the time-space potential value $\text{TPV}(i,j)$ of commuter CB between commuters i and j is inversely proportional to $S(i,j)$, $L(i,j)$, and $t(i,j)$. Therefore, the smaller the value of $\text{TPV}(i,j)$, the greater the potential for commuters between i and j to take commuter CB travel together. Passengers are similar in space and time of travel, showing a more similar time space of commuting travel. The likelihood that they will share commuter CB travel is higher.

3.2. Solution of Time-Space Potential Model. Firstly, the study area is gridded and the boundaries of the study area are adjusted to generate 5729 1 km × 1 km grids. Secondly, a time window constraint is established to calculate the time-space potential values between commuters in the grid with each cell grid. Finally, all grids in the study area are iterated to obtain the potential value between any commuters. The steps are as follows.

> Step 1: the study area is divided into a unit grid of 1 km × 1 km, denoted by U_c, and the unit grid within the entire study area is defined as a set U, and the commuters located in the unit grid form a set P_{C_k}, where $P_{C_k} \subseteq P$, and P is the set of commuters.
> Step 2: establish time window constraint TW_t.
> Step 3: iterate over all the grids in the study area in terms of the unit cell grid U_c and calculating the values of $S(i,j)$, $L(i,j)$ and $t(i,j)$ among the commuters in each grid.
> Step 4: if $T > \delta$ or $S(i,j), L(i,j) > \varepsilon$, then it indicates that i and j do not have the potential for commuter CB.
> Step 5: if $T \leq \delta$ and $S(i,j), L(i,j) \leq \varepsilon$, then calculate the time-space potential values between i and j. The entire algorithm process is iterated through all grids until all the time-space potential values of CB between commuters in the study area are calculated.

Algorithm 1 for calculating the time-space potential values of commuter CB is designed according to the calculation process.

3.3. Service Areas and Potential Travel Demand. This section is the core of the paper. Based on the results of the time-space potential value calculation of CB and referring to the literature [21], the threshold of time-space potential value is 0.5. When the time-space potential value is less than 0.5, the distance difference between commuters' residence, workplace, and time difference from home is the smallest. At that time, the commuters have more potential to travel together and the possibility of using the same transportation mode is higher. The unit grids with time-space potential values less than 0.5 are sorted in descending order by the number of commuters. The top 30% of the sorted grids and the last 30% of the sorted grids are taken as the sample set. It is assumed that the 30% unit grids with the higher number of commuters are the high demand area, so that it is equal to "1."

```
Input left_lng, left_lat, right_lng, right_lat, sample set D = {x₁, x₂, x₃, ..., xₙ}
(1) for i, j in grid do (i = 1, 2, 3, ..., len (grid))
(2)    if t [i], t [j] in TW [i] do
(3)        S (i, j) = 2 × asin (sqrt (a)) × 6371 × 1000
(4)        L (i, j) = 2 × asin (sqrt (b)) × 6371 × 1000
(5)        t (i, j) = timestamp1 [i] − timestamp2 [j]
(6)        if S (i, j) < ε and L (i, j) < ε and t (i, j) < δ do
(7)            TPV [i, j] = a × S (i, j) + b × L (i, j) + c × t (i, j)
Output TPV
```

ALGORITHM 1: CBTPV algorithm.

The 30% unit grids with lower number of commuters are the low demand area, so that it is equal to "0". Considering the factors that influence commuters' choice of commuter CB travel as the input parameters of the model, construct a logistic regression grid model. Based on the model results, the commuter CB initial service areas and potential travel demand are obtained.

3.3.1. Logistic Regression Model. Logistic regression model is a classification algorithm of machine learning. The algorithm predicts in a classification way and can calculate the probability of each category, which fits the filtering of the grid in the study area of this paper. Firstly, based on the time-space potential model of commuter CB, we initially selected commuters with time-space potential value less than 0.5 and identified their geographical location in the unit grids. Secondly, we choose the average commuting distance, average commuting time, average income, number of bus stations, number of subway stations, average distance from neighboring bus stations, and average distance from neighboring rail stations of commuters in the grids as the input parameters of the logistic regression model. Finally, a binary logistic regression grid model is constructed to predict the unit grid, and the model is solved by SPSS software. The unit grids of high hotspots are filtered and probability values are obtained to mine the potential population of commuter CB.

(i) Logistic regression model theory: logistic regression is the search for the vector of independent variables $X = (X_1, X_2, ..., X_n)$ and the binary response Y [21]. The probability of Y belonging to a particular class is modeled.

$$P(X) = \Pr(Y = 1 \mid X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 +, ..., + \beta_p X_p. \tag{3}$$

In fact, logistic regression classification is the process of finding a function, mapping the function values for the 0 to 1 interval, and then classifying the data into two categories. Based on continuous exploration, an ideal "unit-step function" is eventually found, and the function value $P(X)$ is mapped to a 0 or 1 class label according to its positivity or negativity.

However, the direct design of the step function value in this way is discontinuous, and it is not possible to perform some relevant derivations, which is not conducive to the optimization calculation later. Thus, the Sigmoid function is chosen as the classification function in the Logistic Regression algorithm, and the function expression is as follows:

$$g(z) = \frac{1}{1 + e^{-z}}. \tag{4}$$

The Sigmoid function is an s-shaped curve, with $g(z)$ taking values in the interval $[0, 1]$; when $z = 0$, $g(z) = 0.5$, when $z \longrightarrow +\infty$, $g(z)$ tends to 1, and when $z \longrightarrow -\infty$, $g(z)$ tends to 0.

Then we have

$$P(X) = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 +, ..., + \beta_p X_p}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 +, ..., + \beta_p X_p}}. \tag{5}$$

The coefficients of the logistic regression model are usually estimated by the maximum likelihood estimation method.

$$L(\beta) = \prod_{i:y_i} p(X_i) \prod_{x':\acute{y}_i} p(X_{x'}), \tag{6}$$

where

$$\begin{aligned} \beta &= (\beta_0, \beta_1, \beta_2, ..., \beta_p), \\ X_i &= (X_{i1}, X_{i2}, ..., X_{ip}). \end{aligned} \tag{7}$$

(ii) Characteristic values: based on the existing basic data, the study is carried out to fully explore the travel demand and service areas of CB. We choose seven important factors as input parameters for the Logistic Regression model, which are strongly influencing commuters to take commuter CB travel.

① Average commuting distance: based on the longitude and latitude information of mobile phone signaling data, we calculate the difference between the Euclidean distance of commuters leaving their place of residence and arriving at their place of job.

② Average commuting time: based on the time difference between the user's departure from the

place of residence and arrival at the place of work recorded by the mobile phone signaling data, we consider personal business trips or out of work, etc., and take the average commuting time of three working days in a week as the average commuting time. Then, counting the number of commuters in each unit grid, we calculate the average commuting time of each unit grid.

③ Secondary house prices: considering that the prices of secondary houses can characterize people's income to some extent, based on this, secondary house prices are used as a substitute variable for people's income. The mean value of the price of second-hand houses nearby where commuters reside is calculated as a characteristic to represent the income of commuters.

④ Number of bus stops: invoke Gaode map API interface, use the Python programming language to crawl the latitude and longitude of bus stops in the study areas, and count the number of bus stops in the unit grids.

⑤ Number of rail stops: similar to ④, the Gaode map API interface is retrieved and the Python programming language is used to crawl the latitude and longitude of rail stations in the study area and count the number of rail stations in the unit grids.

⑥ Distance of commuters' neighboring bus stops: the distance of commuters from bus stops and rail stops will influence whether they choose to take CB for commuting. The average value of the shortest distance between bus stops and rail stops in the grid of commuters' neighboring cells is considered as the input parameter of the logistic regression model.

⑦ Distance of commuters' neighboring rail stations: the distance of commuters from the rail station platform will influence whether they choose to take CB for commuting. The average value of the shortest distance of rail stations in the grids of commuters' neighboring units is considered as the input parameter of the logistic regression model.

3.3.2. Service Areas and Potential Travel Demands.

Based on the Logistic Regression model, the parameters of the model are input to predict the grids in the study area. Through the theory of the Logistic Regression model, it is known that when $P \geq 0.5$, the prediction result has good predictive value, and the grids are considered as high hotspots grids; on the contrary, when $P < 0.5$, the unit grids are low hotspots grids. Thus, the high hotspots grid area can be used as the commuter CB service areas. And, the commuters that exist in the high hotspot grids are considered as the potential commuter CB travel demand people.

4. Case Study

4.1. Background of the Case.

In this study, the commuter CB travel demand and service areas identification method is proposed in the paper. The method is applied to a real case in the central city of Chongqing, China. The distribution of commuters' occupational and residential locations is identified and visualized based on the commuter OD identification algorithm. In Figure 1, it can be seen that commuters' residence is mainly concentrated in the central area of the central city, and the areas are also the commuters' work gathering area.

4.2. Case Results

4.2.1. Analysis of the Results of Calculating the Time-Space Potential Value of Commuter CB.

Algorithm 1 is designed in Python to calculate the potential values between commuters in the unit grids between 7:00 a.m. and 9:00 a.m. The results are shown in Figure 2. The average value of potential values between commuters in the unit grids is statistically analyzed. And the grids with potential values less than 0.5 in the unit grids are chosen to prepare for the logistic regression model to be established below.

4.2.2. Analysis of Logistic Regression Model Prediction Results.

Based on the calculation results of the commuter CB travel potential model, the unit grids with an average travel potential value less than 0.5 (471 units) are chosen and sorted in descending order by the number of commuters in the unit grids. The upper 30% and the lower 30% of the sorted units are taken as the sample set. Since the number of commuters in the upper 30% of the unit grids is higher, they are identified as $Y = 1$, and similarly, the lower 30% of the unit grids are identified as $Y = 0$. The total number of unit grids is 282.

The binary logistic regression model is solved by SPSS software. The fitted results show that the average commuting time, the average distance of neighboring bus stations, the number of bus stations, and the income level had positive effects on the identification of the areas served by commuter CB. The summary table of parameters of the model is shown in Table 1, and the table of prediction accuracy is shown in Table 2.

From Table 1, Wald is 84.817, $P \leq 0.01$. According to the logistic regression theory, it is known that it passed the significance level test and the model is statistically significant. While Cox–Snell R Square is 0.260 and Nagelkerke R Square is 0.346, the fit of the model is high and the model explains the original data at a desirable level.

As can be seen from Table 2, the Sigmoid function takes values in the range of 0-1 interval, with 0.5 as the dividing line. The prediction cannot be used as a commuter CB unit grid in the prediction accuracy rate of 71.6%, the prediction as the service areas has 100 unit grids, and the prediction

(a)   (b)
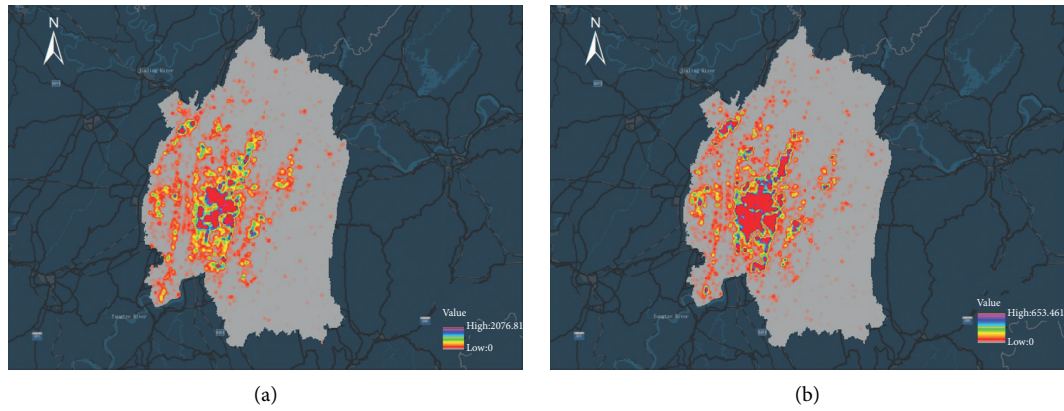
FIGURE 1: Heat map of where commuters reside and where they work. (a) Heat map of population distribution in the place of residence. (b) Heat map of the population distribution of the workplace.
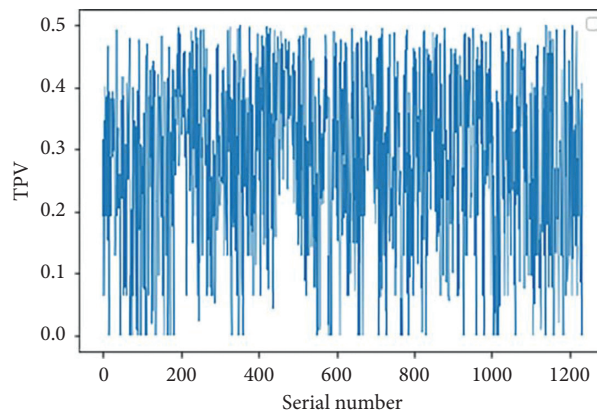

FIGURE 2: Distribution of time-space potential values of CB.

TABLE 1: Summary table of model parameters.

Parameters	Wald	P value	Cox–Snell R square	Nagelkerke R square
Values	84.817	≤0.01	0.260	0.346

TABLE 2: Prediction accuracy.

Actual prediction		Prediction		Accuracy rate (%)
		Y		
		0	1	
Y	0	101	40	71.6
	1	41	100	70.9
Total accuracy rate (%)				71.3

correct rate is 70.9%. The total prediction accuracy rate is 71.3%, the accuracy rate is 71.43%, the recall rate is 70.92%, and AUC value is 0.811 (as shown in Figure 3). These indicators show that the prediction model is more ideal and the prediction effect is perfect.

Based on the learned model, logistic regression is applied to predict 5729 grids in the central city of Chongqing, China.

The machine learning model is solved by SPSS software, and the prediction results are shown in Figure 4.

4.2.3. High Hotspot Grids and Potential Travel Demand. Based on the above analysis of the model results, it can be learned that the prediction results for the area of high
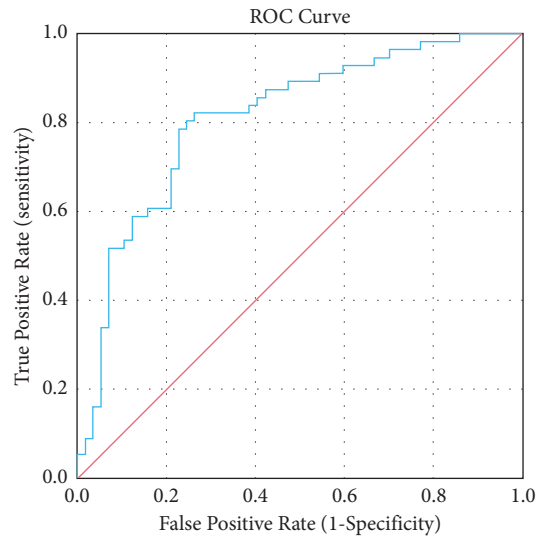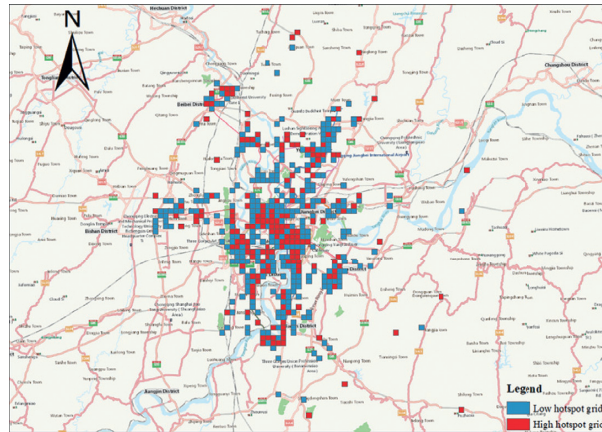
Figure 3: ROC curve.

Figure 4: Model prediction results.

hotspot unit grids (as shown in Figure 5(a)) have advantages for the operation of commuter CB routes. The high hotspot grids areas are considered as the service areas of commuter CB. And, the commuters in the high hotspot unit grids are the potential commuter CB travel demand crowd (as shown in Figure 5(b)).

4.2.4. Examples of Commuter CB Line Planning.
By analyzing the distribution of high hotspot grids and travel demand, we randomly chose one high hotspot unit grid each in Shapingba District, Beibei District, and Yubei District of Chongqing, China, as an example to plan commuter CB routes. The commuters in the high hotspot unit grids are considered as potential commuter CB travel demand. The lines information is shown in Table 3.

In this paper, the place of residence is considered as the pickup area and the place of work as the drop-off area. Three randomly selected residential grid areas are surveyed by random sampling to verify the accuracy of the model prediction results. And, in the chosen areas, conduct a questionnaire survey of the commuter CB SP for passengers. The purpose of the SP questionnaire is that the general travel intentions of people in the unit grid represent the travel intentions of potential commuters of CB travel in the unit grid.

One hundred questionnaires are distributed to each of the three chosen areas, for a total of 300 questionnaires, including 95 valid questionnaires for grid ID 4309, 98 valid questionnaires for grid ID 4342, and 94 valid questionnaires for residential grid ID 2654, for a total of 287 valid questionnaires. The results of the questionnaire survey show that the number of passengers in each grid who are inclined to choose commuter CB travel is greater than the predicted number of potential commuter CB travel demand people obtained from the model, which verifies the validity of the model prediction results.

Based on the number and distribution of commuter CB travel demands, the k-means clustering algorithm is used to spatially cluster the travel demand. Since the k-value has a

(a)   (b)
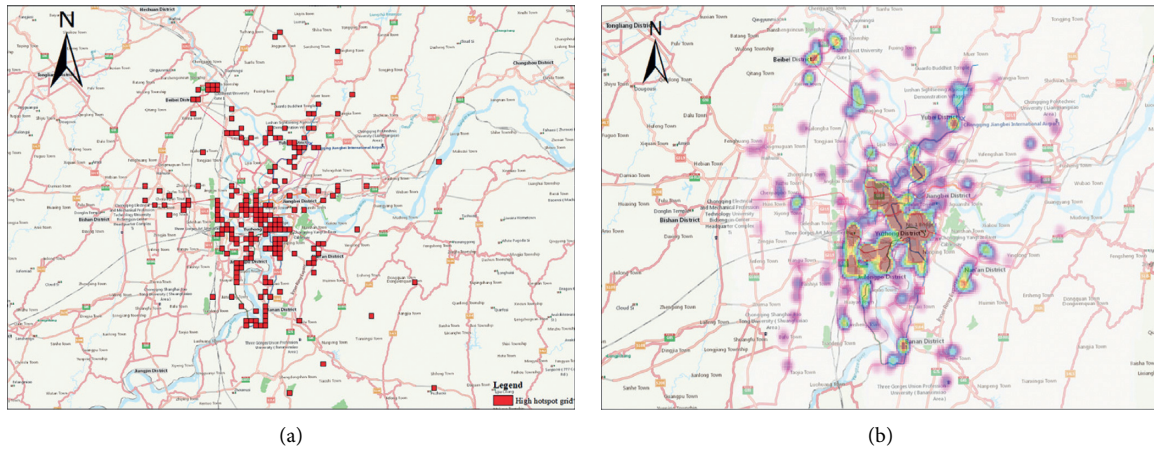
FIGURE 5: High hotspot grids and potential travel demand distribution. (a) High hotspot grids. (b) Potential travel demand distribution heat map.

TABLE 3: Line information of example.

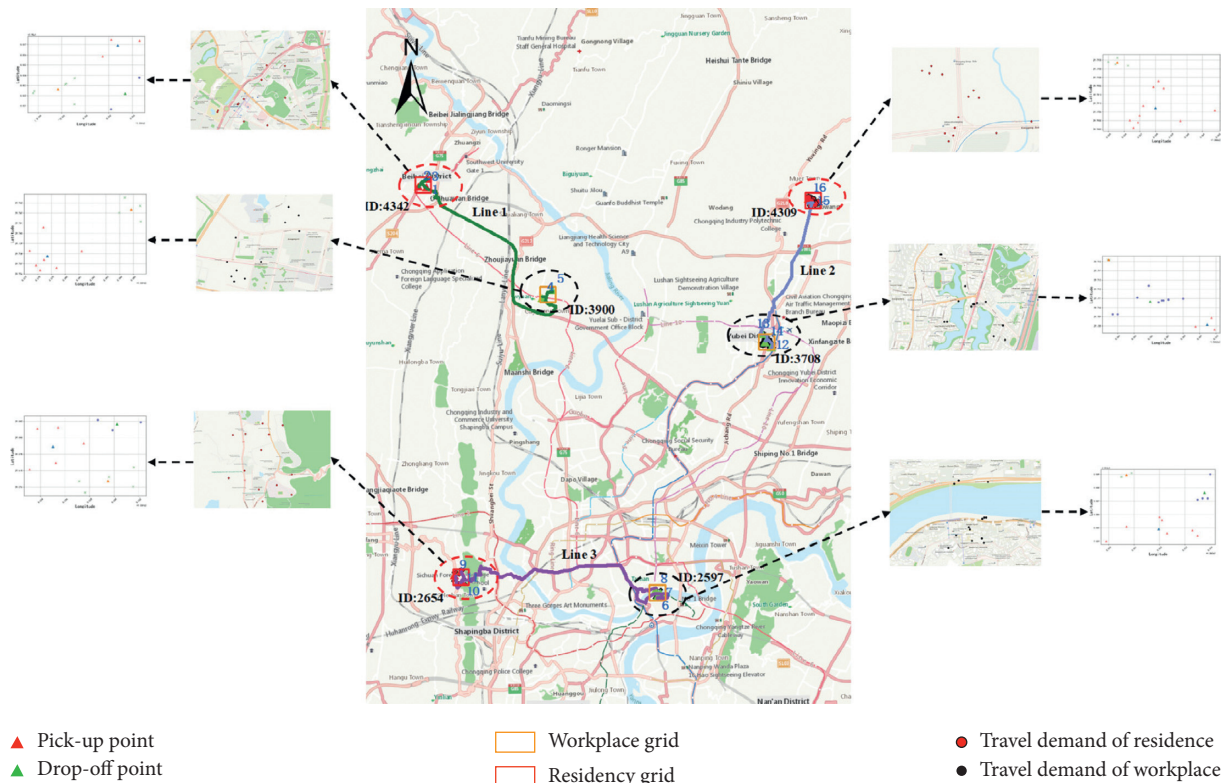| Line ID	Grid ID of residence	Grid ID of workplace	Distance (km)	Demand (person)
1	4309	3708	10.7	14
2	4342	3900	12	11
3	2654	2597	13	13

FIGURE 6: Example of line planning.

large impact on the result of the k-means clustering algorithm, the appropriate k-value is initially determined by applying the Silhouette Coefficient. Then, spatial clustering is carried out, respectively, for residential and workplace

travel demand, and line planning is performed for the area based on the clustering results. Through line planning, three vehicles are allocated to meet the passenger travel demand. From the perspective of enterprise operation, the company's

constant cost is 240 RMB, the variable cost is 25.23 RMB, and the enterprise's fare revenue is 304 RMB. Without considering the government subsidy, the total revenue is 38.77 RMB, which ensures that the operating enterprise is in a profitable state. The line planning results are shown in Figure 6.

5. Conclusion

Based on the current status of research by many scholars, this paper focuses on the current shortcomings and carries out an in-depth study on the issue of commuter CB travel demand and service areas. The main research contents of this paper are as follows:

(i) Firstly, based on the preprocessing of mobile phone signaling data and commuter OD identification, a commuter CB travel time-space potential model is proposed. Then, the study area is gridded, by designing an algorithm to solve the model.

(ii) Considering commuters who meet certain conditions, Logistic Regression model is applied to analyze the unit grid as the basic cell. We choose the objective factors that influence passengers' choice to take commuter CB as the output parameters of the model and deeply mine the potential population of commuter CB travel demand. We consider the high hotspot grids output of the model as the commuter CB service areas. Finally, using Chongqing, China, as a study case and three routes as examples, the results show that the operating companies are in a profitable state without government subsidies. The case results prove the effectiveness of the method proposed in this paper in practical applications.

In addition, some issues in this paper need to be further discussed:

(i) The data used in this paper are mobile phone signaling data based on COO cellular cell location technology, and there are certain defects in data accuracy. The article chooses to sort the samples of the upper 30% and the lower 30% of the grids, and other methods are also feasible, such as the upper 20% and the lower 20%.

(ii) The paper is not sufficient to justify the value of some model parameters, and it is expected that the parameters of the model can be further studied later to improve the accuracy of the model.

(iii) The operating company can combine the spatial and temporal distribution characteristics of the potential commuter CB travel demand obtained from this paper to introduce intentional routes to specific areas. This way can provide people with convenient travel services.

Data Availability

The data used to support the results of this study are not available because they contain user privacy.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] T. Liu and A. Ceder, "Analysis of a new public-transport-service concept: customized bus in China," *Transport Policy*, vol. 39, pp. 63–76, 2015.

[2] Y. Lyu, C. Chow, V. C. S. Lee, Y. Li, and J. Zeng, "T2CBS: mining taxi trajectories for customized bus systems," in *Proceedings of the 2016 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pp. 441–446, San Francisco, CA, USA, April 2016.

[3] S. Harms and B. Truffe, *The Emergence of a Nationwide Carsharing Co-operative in Switzerland*, Eidg, Anstalt fur Wasserversorgung und Gewasserschutz, Dübendorf, Switzerland, 1998.

[4] K. Tsubouchi, H. Yamato, and K. Hiekata, "Innovative on-demand bus system in Japan," *IET Intelligent Transport Systems*, vol. 4, no. 4, pp. 270–279, 2010.

[5] F. Qiu, W. Li, and J. Zhang, "A dynamic station strategy to improve the performance of flex-route transit services," *Transportation Research Part C: Emerging Technologies*, vol. 48, pp. 229–240, 2014.

[6] L. Scott, M. Lee-Gosselin, A. Sivakumar, and J. Polak, "A new approach to predict the market and impacts of round-trip and point-to-point carsharing systems: case study of London," *Transportation Research Part D: Transport and Environment*, vol. 32, pp. 218–229, 2014.

[7] K. An and H. K. Lo, "Two-phase stochastic program for transit network design under demand uncertainty," *Transportation Research Part B: Methodological*, vol. 84, pp. 157–181, 2016.

[8] T. Liu, A. Ceder, A. Ceder, R. Bologna, and B. Cabantous, "Commuting by customized bus: a comparative analysis with private car and conventional public transport in two cities," *Journal of Public Transportation*, vol. 19, no. 2, pp. 55–74, 2016.

[9] Y. Lyu, C.-Y. Chow, V. C. S. Lee, J. K. Y. Ng, Y. Li, and J. Zeng, "CB-Planner: a bus line planning framework for customized bus systems," *Transportation Research Part C: Emerging Technologies*, vol. 101, pp. 233–253, 2019.

[10] J. Zhong, H. Bai, and Z. Wu, "Division method of urban customized bus passenger flow distribution area for commuter- oriented demand," in *Proceedings of the 2018 World Transport Conference*, Geneva, Switzerland, August 2018.

[11] Y. Cheng, A. Huang, G. Qi, and B. Zhang, "Mining customized bus demand spots based on smart card data: a case study of the beijing public transit system," *IEEE Access*, vol. 7, pp. 181626–181647, 2019.

[12] Q. Yu, W. Li, and H. Zhang, "Mobile phone data in urban customized bus: a network-based hierarchical location selection method with an application to system layout design in the urban agglomeration," *Sustainability*, vol. 12, 2020.

[13] Q. Liu, Q. Li, and C. Tang, "A visual analytics approach to scheduling customized shuttle buses via perceiving passengers' travel demands," 2020, https://arxiv.org/abs/2009.02538.

[14] J. Zhang, D. Z. Wang, and M. Meng, "Analyzing customized bus service ona multimodal travel corridor: an analytical

modeling approach," *Journal of Transportation Engineering*, vol. 143, no. 11, 2017.

[15] L. Tong, L. Zhou, J. Liu, and X. Zhou, "Customized bus service design for jointly optimizing passenger-to-vehicle assignment and vehicle routing," *Transportation Research Part C: Emerging Technologies*, vol. 85, pp. 451–475, 2017.

[16] W. Huang, W. Jin, J. Huang, and B. Han, "Pricing problem of customized bus under different market strategies," *Journal of Guangxi Normal University*, vol. 2, no. 2, 2018.

[17] J. Ma, Y. Yang, and W. Guan, "Large scale demand driven design of a customized bus network: a methodological framework and beijing case study," *Journal of Advanced Transportation*, vol. 2017, Article ID 3865701, 14 pages, 2017.

[18] D. Li, X. Ye, and J. Ma, "Empirical analysis of factors influencing potential demand of customized buses in Shanghai, China," *Journal of Urban Planning and Development*, vol. 145, no. 2, 2019.

[19] X. Y. Tang, T. Zhou, B. C. Lu, and Z. G. Gao, "A commuting OD matrix training method based on mobile phone data," *Transportation System Engineering and Information*, vol. 16, no. 5, pp. 64–70, 2016, in Chinese.

[20] E. H. Geng, *Analysis of Ride-Sharing Potential Based on Bus Data*, University of Electronic Science and Technology, Chengdu, China, 2017, in Chinese.

[21] C. Yu, C. Xu, X. Ding, and L. Zeng, "Optimizing location of car-sharing stations based on potential travel demand and present operation characteristics: the case of chengdu," *Journal of Advanced Transportation*, vol. 2019, Article ID 7546303, 13 pages, 2019.