Received February 26, 2019, accepted May 7, 2019, date of publication May 13, 2019, date of current version May 24, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2916342

Location Optimization for Urban Taxi Stands Based on Taxi GPS Trajectory Big Data

ZHAOWEI QU, XIN WANG, XIANMIN SONG, ZHAOTIAN PAN, AND HAITAO LI
School of Transportation, Jilin University, Changchun 130022, China

Corresponding author: Xianmin Song (songxm@jlu.edu.cn)

This work was supported in part by the Jilin Province Natural Science Foundation of China under Grant 20180101063JC, and in part by the project of the Jilin Provincial Department of Education under Grant JJKH20190153KJ.

ABSTRACT A taxi stand can effectively regulate the behavior of taxi picking up passengers, reduce empty-run rate, and provide a convenient and orderly waiting environment for the public. However, the unreasonable setting of the existing taxi stands in most cities leads to an extremely low utilization rate and a waste of public space resources. This paper presents a novel three-stage strategy to address the taxi stands location problem (TSLP) incrementally. First, taxi demands hotspots are mined from a massive taxi Global Positioning System (GPS) data with GIS platform, and the optimal area for taxi stands siting in the following stages is determined. Then, the spatial interaction between taxi demands and taxi stands is explored to generate demand subsections and stand candidates along both the sides of the road. At last, a taxi stand location model (TSLM) is developed to minimize the total cost, which contains the access cost of passengers and the construction cost of taxi stands. The genetic algorithm-based procedure is adopted for TSLM optimization. A case study conducted in China verifies the effectiveness of the location strategy and investigate the impact of the maximum acceptable distance for passengers on TSLP. The experimental results describe the number and layout of taxi stand under a different demand coverage, which indicates that the proposed approach is beneficial to provide scientific reference for the municipal department in taxi stand site decisions and make a tradeoff between the interests of planners and users.

INDEX TERMS Taxi stand, location strategy, spatial-temporal demand, GPS big data, genetic algorithm.

I. INTRODUCTION

Taxi is an important part of urban transport, which has always been favored by people for its unique convenience and comfort since the 19th century [1]. In recent years, taxi services such as telephone booking and online booking have become more and more popular, but in most countries and regions, such as China, the main operating mode of taxis is still hailing along the roadside. In this traditional mode, taxis head to places with heavy traffic in searching for customers, which could aggravate traffic congestion and air pollution problems [2], [3]. In addition, the empirical cruise mode keeps passengers and drivers in a state of information isolation. There is always a mismatch between demand and supply for taxis in time and space, leading to high empty-run rate and low operating efficiency [4].

Taxi stand (TS) can provide taxis and passengers with a more secure and effective mode of interactive service, which has the characteristics of identifiable, orderly, efficient and quick [5]. Generally, it is not practical to set up taxi stands in every area where passengers may appear, and it will lead to the loss of flexibility and convenience of taxis. In practice, the location choice of TS mainly depends on the experience of the traffic manager or the feedback from some drivers and passengers. The lack of scientific and rigorous decision-making criteria leads to a low utilization, and some even exist in name only. Therefore, it is significant to solve the taxi stand location problem (TSLP) in a scientific way. Urban planners need to know where to locate TSs to maximize utilization and other considered objectives. To answer this question, two major problems need to be solved: taxi demand analysis and location optimization modeling.

The analysis of taxi demand with spatial-temporal information aims to mining the taxi travel hotspots, which is the premise of accurately identifying the optimal candidate

The associate editor coordinating the review of this manuscript and approving it for publication was Hua Fang.

location for taxi stands siting. Due to the limitations of data collection methods, early studies on the spatial distribution of urban travel hotspots are mainly based on land use type data, vehicle information collected by fixed detectors, and other survey data [6]. In recent years, a large number of taxis in urban areas have been equipped with Global Positioning System (GPS) to track and dispatch vehicles [7]. Compared with data from traditional surveys, mass GPS data with high geographical resolution are more accurate, objective, cost-effective, and accessible [8]. They provide a rich information source for describing residents' daily mobility, which has been widely used to mine information for the field of urban taxi transportation [9], [10], such as taxi service strategies uncovering [11], taxi driving directions recommend [12], travel purpose inferring [13], [14], etc. Moreira-Matias *et al.* [15] used GPS data of a taxi company in Porto, Portugal to predict the spatial and temporal distribution of taxi passenger demand in the short term. Ferreira *et al.* [16] proposed a model that can be used for users to visually query the taxi journey and the origin and destination location, and studied the mobility in the city. Bischoff *et al.* [17] analyzed the taxi travel behavior and taxi supply in Berlin on weekdays and weekends using a large amount of taxi GPS data. Tang *et al.* [18] extracted travel information from taxi GPS data in Harbin, China, analyzed the distribution of the origin and destination, and studied the searching behavior of drivers in pick-up locations.

Analysis of taxi trajectory data also provides a new perspective for researchers on improving taxi operation efficiency. Many researchers have applied real-time GPS data on exploring the behavior patterns of taxi drivers and passengers [19], [20], establishing the arrival probability model of passengers and empty taxis [21], and developing a recommendation system to provide taxi drivers with possible places and routes to pick up passengers quickly and suggest places where passengers can take a taxi rapidly [22]–[24]. Moreira-Matias *et al.* [25] combined the spatial-temporal distribution of taxi demand with the state of the road network, and recommended taxi drivers to taxi stands with the least waiting time in real time. Wong *et al.* [4] established a decision model for empty taxi drivers and analyzed their preferences in traveling towards taxi stands and waiting for customers at taxi stands. To some degree, previous research work can reduce the vacancy rate of taxis. However, the irrationality of TSs locations may lead to the invalid recommendation results.

Facility location is a reflection of human demands. TS is a kind of transportation service infrastructure, which reflects the taxi travel demand. In order to satisfy these demands of geographical distribution, a great deal of literature have adopted the location approaches to solve the problem of facility location [26]. In view of different practical problems, considering different demands and objectives, quite a few location models have been proposed such as the p-Median [27], p-Center [27], the maximum coverage location problem (MCLP) [28], the flow capture location model (FCLM) [29] and the flow refueling location model (FRLM) [30]. These models have been successfully employed in the siting optimization of transportation service infrastructure, their applications involve the siting of conventional and alternative refueling facilities [31]–[33], charging stations for electric vehicles [34]–[36], park-and-ride facilities [37]–[39], bike-sharing stations [40]–[42], etc.

In terms of taxi industry operation and service, there are few previous studies existing on the location research of TS. To date, the closest research is by Ocalir *et al.* [1], who develop a decision support system for taxi stand location decision. They assessed the existing taxi stands in parts of 99 traffic zones located in Ankara to decide whether give any more permission for new ones. In essence, the system evaluates the number of TSs in a certain region of the city on a macro level, rather than optimizes the specific location of TSs in a region.

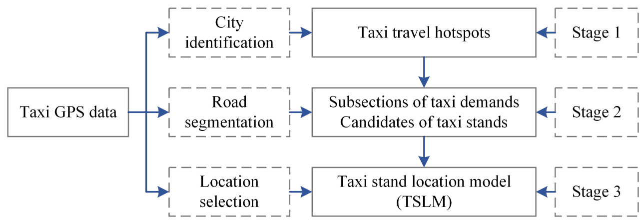


FIGURE 1. The flow chart of this paper.

This research intends to fill in the blank of the TSLP through a hierarchical methodology, which optimizes taxi stand not only in quantity but also in layout. In this paper, we study the location selection of taxi stands from the perspective of exploring the spatial-temporal dynamic attributes possessed in taxi demands and its spatial interaction with taxi stands. A three-stage location strategy is proposed using taxi GPS big data to solve the TSLP incrementally. The research flow chart is demonstrated in Fig. 1. The contributions of our work lie in the following aspects. First, in order to seek the appropriate setting environment of TSs, we extract the actual travel demands from taxi GPS big data and explore the distribution regularity in time and space dimension to identify the hotspots with high demands accurately. Second, we describe the spatial interaction between taxi demands and taxi stands. On this basis, travel demand subsections and candidate taxi stands are generated on both sides of the road in a staggered arrangement. Third, we design a taxi stand location model (TSLM) to solve the TSLP that achieve the objective of minimizing the access cost for passengers and the construction cost. Considering the heterogeneity of passenger behaviors, the concepts of demand coverage and maximum acceptable walking distance are introduced in the model to optimize the selection of the locations. To verify the validity of the proposed strategy and TSLM, a case study is conducted. The results show that the approach presented could be effectively applied to TSLP in metropolitan.

The remainder of this paper is organized as follows. Section II describes the research methods for TSLP, including three-stage location strategy and optimization algorithm. In section III, a case study is conducted to accomplish the acquisition of experimental parameters and demonstrate the effectiveness of our methodology. Section IV concludes the paper.

II. METHODOLOGY

In this section, a three-stage location strategy is proposed to explore taxi stands site step-by-step. In stage 1, we identified which areas of the city need to be set up for taxi stands. In stage 2, the streets in these areas are divided into taxi travel subsections, and a candidate taxi stand is generated in each subsection. In stage 3, the optimal setting location for taxi stands is selected from all the candidate points.

TABLE 1. Taxi demands data description.

Data field	Description
Taxi ID	A unique ID for a taxi
Timestamp	Date and time of GPS data recording
Longitude	The longitude of the vehicle's position when the GPS data is recorded
Latitude	The latitude of the vehicle's position when the GPS data is recorded
Status(0/1)	The passenger state of taxi: <0> means taxi vacant and <1> means taxi occupied

A. STAGE 1 – TAXI DEMANDS EXTRACTION AND TRAVEL HOTSPOTS IDENTIFICATION

There are several fields contained in GPS records, such as taxi id, timestamp (associated positioning time), taxi location (longitude, latitude), status (occupied/vacant), vehicle speed, et al. As shown in Table.1, only the taxi demand data, i.e., the information indicating taxis picking up or dropping off passengers is of interest in this work. To meet travel demands, taxis pick up passengers at the origin, then follow the driving route to the destination and drop off passengers. In this process, the field of status changes from $<0>$ to $<1>$ after the passenger gets on, and remains $<1>$ constant until the passenger gets off. Generally, the point in state $<1>$ is regarded as the pick-up location when the state shifts from $<0>$ to $<1>$, and the point in state $<0>$ is regarded as the drop-off location when the state shifts from $<1>$ to $<0>$. Accordingly, the pick-up and drop-off data sets which represent the actual travel demands of taxi can be expressed as $R^1(0 \to 1) = \{t, s, (x, y)\}$ and $R^0(1 \to 0) = \{t, s, (x, y)\}$, where t is timestamp, s is status, and (x, y) denotes the taxi location. After the processing and map matching of the raw taxi GPS data, these taxi demands with spatial–temporal information are extracted for the hotspots identifying.

Defining hotspots that reflect the intensive taxi travel is an essential foundation for TSs siting. In this research, a GIS platform is used to spatially analyze the potential setting areas of TSs, combining with the time-varying law of taxi demands and the geographical conditions of location factors. The kernel density analysis by GIS can reflect the distance decay effect in the geographical space distribution well, which is in line with the diffusive characteristics of the influence of urban facilities such as TSs on surrounding locations. The general form of the kernel density estimation (KDE) is expressed as:

$$\lambda(x, y) = \frac{1}{\pi r^2} \sum_{i=1}^{n} k\left(\frac{d_{is}}{r}\right) \tag{1}$$

where $\lambda(x, y)$ is the estimated density value at location $S(x, y)$, n is the total number of event points,r is the search bandwidth, d_{is} is the distance between event point i and location S, and k is the kernel function. In KDE, each event point i is covered by a smooth surface. As shown in Fig. 2, the kernel function value of each point is calculated according to the distance to the center point S. All the surfaces value superimposed on the reference point are summed to obtain the density estimate for the distribution of event points. The visualization operation system on GIS platform allows us to present the distribution of taxi travel hotspots on city map by an intuitive way, and help to make the mining process easier.

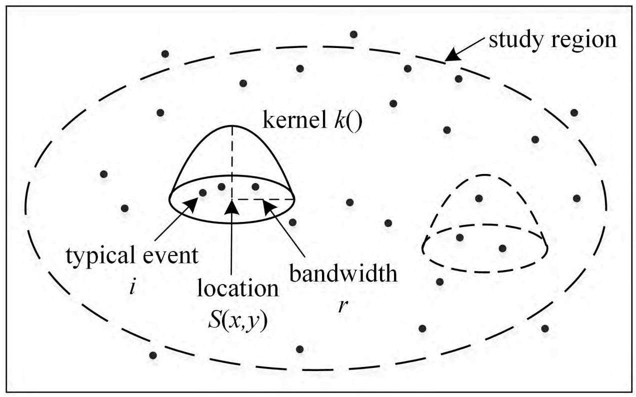


FIGURE 2. Diagram of kernel density method.

B. STAGE 2 – DEMAND SUBSECTIONS PARTITION AND CANDIDATE SITES GENERATION

Taxi stands provide serves for the public and taxi drivers. In reality, passengers appear randomly on the streets, which makes the location of roadside taxi stands in the road network special, i.e., the opposite and staggered distribution along both sides of the road. In order to determine the reasonable taxi stands candidate points, we cluster the taxi demands into travel subsections and generate the candidate points of TSs at the centroid of each subsection, which is illustrated in Fig. 3. In this paper, the direction to the east or south is defined as clockwise (CW), and to the west or north as counterclockwise (CCW).

Firstly, we determine a certain length of segmentation ($L = 2r$) to divide the roads in the setting area into adjacent grids (see Fig. 3a). Meanwhile, the taxi demands (Q_i) within the range of each grid, i.e. the number of taxi pick-up points, is extracted and calculated.

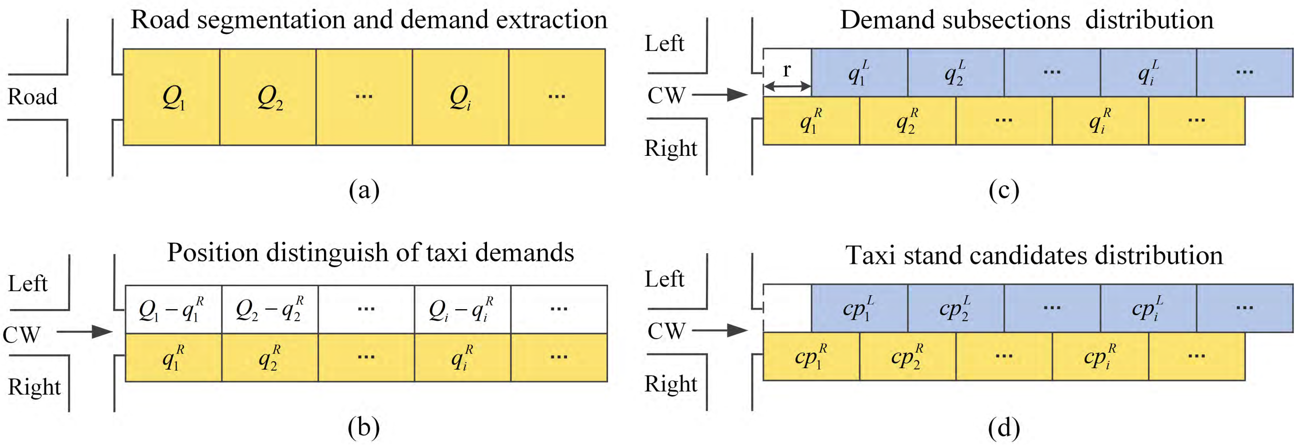


FIGURE 3. Generation of demand subsections and candidate points.

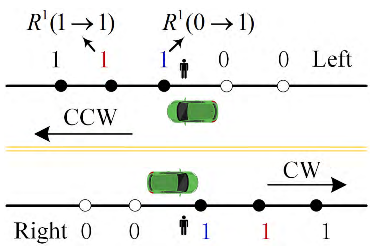


FIGURE 4. The judgment principle of taxi demand location based on the GPS trajectory.

Unless the vehicle happens to be parked at an intersection or other areas with turning conditions, the taxi will continue to drive in the same direction after picking up passengers at the side of the road, as shown in Fig. 4. Formally, we define the driving direction after a taxi demand is serviced as:

$$\begin{aligned} D_{td} &= [R^1(1 \to 1) - R^1(0 \to 1)] \\ &= [\{t_{i+1}, s_{i+1}, (x_{i+1}, y_{i+1})\} - \{t_i, s_i, (x_i, y_i)\}] \\ &= [\{\Delta t, \Delta s, (\Delta x, \Delta y)\}] \end{aligned} \tag{2}$$

where $R^1(1 \to 1)$ denotes the GPS points of taxi demand at time t_{i+1}, Δt is sampling interval, and $(\Delta x, \Delta y)$ is the difference between two tracing points. Therefore, the positive or negative values of Δx and Δy indicate the orientation of passengers, based on which we can obtain the exact location of each passenger on both sides of the road middle line.

Secondly, along the road middle line, we split the initial grids into left and right travel subsections (see Fig. 3b), and set up a candidate taxi stand at the centroid of each subsection. According to the driving direction (D_{td}), the taxi demands in the right travel subsections are calculated as q_i^R, while those in the left can be expressed as $Q_i - q_i^R$.

Thirdly, considering the setting pattern of TSs staggered along both sides of the road, we move the subsections on one side of the road (e.g., the left subsections in CW) from the original positions along the same direction (e.g., in CW) by a distance of half the splitting length (r). As shown in Fig. 3c, the taxi demands q_i^L are updated by:

$$q_i^L = \frac{1}{2}\left(Q_i - q_i^R\right) + \frac{1}{2}\left(Q_{i+1} - q_{i+1}^R\right) \quad (i = 1, 2, \cdots, N) \tag{3}$$

Finally, due to the displacement of the subsections, uncovered taxi demands will be allocated to the neighborhood nearest on the same side. Fig. 3d displays the final distribution of taxi stand candidates, where the right side is represented by cp_i^R and the left side is represented by cp_i^L.

The distance matrix of demand subsections and taxi stands can be represented as $D = [d_{ij}]$, where d_{ij} denotes the distance from the subsection i to the candidates j calculated based on three spatial location relationships between them.

(1) Same subsection: Assuming that taxi demands are uniformly distributed in a straight line on both sides of the candidate point, the distance d_{ij} is given by (4), where q_i is the taxi demands in subsection i.

$$d_{ij} = \frac{\frac{r}{q_i/2}\left(1 + 2 + \cdots \frac{q_i}{2}\right)}{q_i/2} = \frac{r}{q_i} + \frac{r}{2} \tag{4}$$

(2) Same side of the road, different subsections: The distance d_{ij} equals to the metropolitan metric between the centroids of subsections approximately, as given by:

$$d_{ij} = < x_i, y_i > - < x_j, y_j > \tag{5}$$

(3) Different sides of the road, different subsections: The distance d_{ij} is based on the addition of (5) and the street width d_s, which can be expressed as:

$$d_{ij} = < x_i, y_i > - < x_j, y_j > + d_s \tag{6}$$

C. STAGE 3 – LOCATION MODEL OF TAXI STAND

The TSLM aims to minimize both the total access costs of passengers and the total construction costs of taxi stands. The former sub-objective is measured by the total travel distances that passengers walk from their origins to the taxi stands. The concept of passengers' unit time value is introduced to transform travel distance or time into access cost for the consistency of the overall optimization objective. The shorter the travel distances, the less the access costs, and the higher is the travel convenience. To simplify the TSLP, the following reasonable hypothesis conditions are given in this paper:

- Due to complex actual situations such as building shielding, the distances from demand points to taxi stands is calculated according to the metropolitan metric and pavement width rather than the Euclidean metric.
- Regardless of the passenger's travel direction, passengers choose their destinations according to their distance from taxi stands.
- Taxis arrive at taxi stands to serve passengers continuously without interruption during rush hour.
- Passengers in the same travel demand point choose go to the same taxi stand.

In this paper, the parameters and notations to be used in the optimization model are summarized as follows:

I	Set of travel demand points
J	Set of candidate points for taxi stands ($J \subseteq I$)
T	Time buckets
p_j	Parking spaces at taxi stand j
P_j^{bb}	Maximum service capacity of parking space per hour in taxi stand j
DA	Maximum acceptable walking distance for passengers
d_{ij}	Distance of between demand point i and candidate point j
q_{it}	Taxi demand at travel demand point i in time t
c_j	Construction cost of candidate taxi stand j
c_p	Unit time value of passengers
v_p	Passenger walking speed
X_j	Binary variable, which equals 1 if a taxi stand is located in point j and 0 otherwise
Y_{ij}	Binary variable, which equals 1 if demand point i is served by taxi stand j and 0 otherwise

The taxi stand location model (TSLM) can be formulated as follows:

$$\text{Minimize: } C = \lambda \cdot c_p \cdot \sum_{t\in T}\sum_{i\in I}\sum_{j\in J} Y_{ij} \cdot d_{ij} \cdot q_{it} + \sum_{j\in J} c_j \cdot X_j \tag{7}$$

$$\text{Subject to: } \sum_{j\in J} Y_{ij} = 1 \quad \forall i \in I \tag{8}$$

$$Y_{ij} \le X_j \quad \forall i \in I, j \in J \tag{9}$$

$$X_j \le p_j \quad \forall j \in J \tag{10}$$

$$\sum_{i\in I} q_{it} Y_{ij} \le P_j^{bb} p_j \quad \forall j \in J, t \in T \tag{11}$$

$$\sum_{j\in J} d_{ij} Y_{ij} \le DA \quad \forall i \in I \tag{12}$$

$$X_j = \{0, 1\} j \in J \tag{13}$$

$$Y_{ij} = \{0, 1\} \quad \forall i \in I, j \in J \tag{14}$$

where λ is the inverse of pedestrian average walking speed v_p, which is introduced to transform walking distance into access time. The objective function (7) is to minimize the total cost containing the total access cost of passengers and the total construction cost of taxi stands. Constraint (8) guarantees all passengers can be covered by the service of taxi stands. Constraint (9) requires that passengers choose to take a taxi at point j only when a taxi stand is located at candidate point j. Constraint (10) specifies that a taxi stand is located only when there are passengers to take a taxi there. Constraint (11) indicates that the total number of passengers that choose to take a taxi in section j cannot exceed the maximum service capacity of taxi stand j within any time period. Constraint (12) represents that passengers at the demand point i choose to take a taxi at candidate stop j only when the distance between point i and point j cannot exceed the maximum acceptable walking distance for passengers. Constraint (13) and (14) impose integer conditions on binary decision variables X_j and Y_{ij}.

In the TSLP, the walking distance between taxi stands and demand points is an extremely crucial factor. Although the maximum acceptable walking distance for passengers is preset to ensure the travel service level and travel desire of passengers, in practice, however, different passengers have different acceptance for the same service level. In the above model, passengers identified as covered by the service are likely not to go to taxi stand, and we don't have to provide services for those passengers who are in low taxi demand, have high requirements for service level and service cost. Consequently, we introduce the concept of demand coverage D_c, which is estimated as the ratio of the travel demand covered by taxi stands to the total demand. Constraint (8) are replaced by the following constraint (16) and (17) to exclude a certain proportion of passengers from the coverage target, and avoid the model to solve the TSLP from a single perspective of full coverage.

A more general form of TSLM can be expressed as:

$$\text{Minimize: } C = \lambda \cdot c_p \cdot \sum_{t\in T}\sum_{i\in I}\sum_{j\in J} Y_{ij} \cdot d_{ij} \cdot q_{it} + \sum_{j\in J} c_j \cdot X_j \tag{15}$$

Subject to: (10) – (16),

$$\sum_{j\in J} Y_{ij} \le 1 \quad \forall i \in I \tag{16}$$

$$\sum_{t\in T}\sum_{i\in I}\sum_{j\in J} q_{it} Y_{ij} \ge D_c \sum_{t\in T}\sum_{i\in I} q_{it} \tag{17}$$

Here, constraints (16) means that passengers can choose not to go to the taxi stands, if they choose to go, they can only go to the same point. Constraints (17) ensures that passengers with a proportion of total demand can obtain service.

D. THE HEURISTIC ALGORITHM

Location problems are difficult to solve due to the inherent complexity of NP-hard [43]. In the past decades, many solution approaches have been developed for facility location problems, among which the heuristic algorithm is the most promising one. Genetic Algorithms (GA), originally formulated by Holland (1975) [44], is a most widely used heuristic algorithm and has been applied to many complex location problems successfully. In this paper, we use a genetic algorithm to solve the TSLP. GA is an optimization method for searching global optimal solution by simulating the biological evolution process of natural selection and genetic mechanism, completing the iterative search process of optimal solution through five basic steps of population generation, fitness evaluation, selection, crossover and mutation [45].

For the TSLP, the chromosome of the location scheme are encoded by binary coding, in which the code length of the genome is used to describe the number of candidate taxi stands, and the binary variable on each code position is used to represents whether the candidate location is selected to set up a taxi stand. An initial population is generated randomly to represent the initial location schemes. The fitness function of each individual is calculated by inverse transformation of the objective function of TSLM (7). The roulette wheel [45] is adopted as the selection strategy. We use the single-point crossover to generate new individuals and replace individual genes with a single-point mutation. If the termination condition is satisfied, the algorithm terminates; otherwise, continue to repeat the evolutionary process. The termination condition of evolution is that the loop reaches the maximum number of iterations or the objective function value has no improvement over a fixed number of iterations.

III. CASE STUDY

The taxi GPS data used in this paper are collected from about 7,200 taxis in a city, China. The data start from 0 a.m. to 24 p.m. in a week in June 2015. The data collection time interval is generally around 20s. Through a series of data preprocessing, more than 150 million valid data are obtained. The data sets representing travel demands and directions are extracted under the environment of the database Microsoft SQL Server 2008.

A. IDENTIFYING TAXI TRAVEL HOTSPOTS

Fig. 5 shows the daily variation pattern of taxi demands. Within a week, the highest and the lowest demands for taxi is 309,881 on Friday and 198,168 on Sunday respectively. With ArcGIS10.2, the spatial analysis was conducted on the 24-hour data set of Friday, and the visualization results were shown in Fig. 6.

Fig. 6a indicates the hotspots distribution of taxi pick-up points. There are 6 hotspots identified, of which the No.1-5 belong to the railway station or coach station, and the No.6 is located in the most prosperous business circle. Fig. 6b displays the hotspots distribution of taxi drop-off points. Compare with Fig. 5a, the number of the hotspots decreases relatively. Due to the disparity of the drop-off points, the No.3 and No.4 area shown in Fig. 6a no longer appear. However, the locations of 4 hotspots in Fig. 6b overlap with that of the pick-up hotspots.

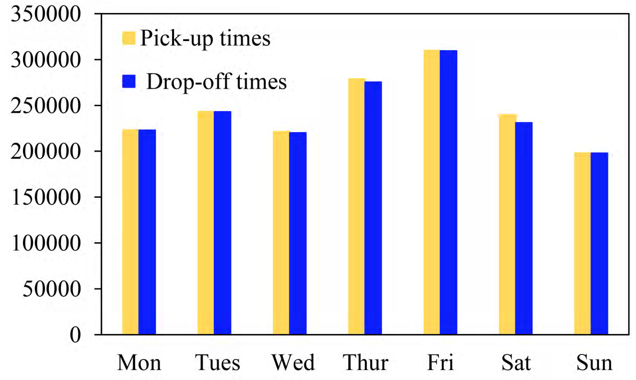


FIGURE 5. Daily variation of taxi demands.

The No.1-5 areas with large-scale people gathering and distributing are usually equipped with taxi stands to satisfy passengers' travel demands. These hotspots present a point radiation state, which does not meet the research conditions of roadside taxi stands. Inversely, the No.6 covers multiple streets with heavy traffic and exhibits a network state, which is suitable for the demand construction area of TSs. Hence, the No.6 hotspot is selected as the research area for taxi stands siting in following work.

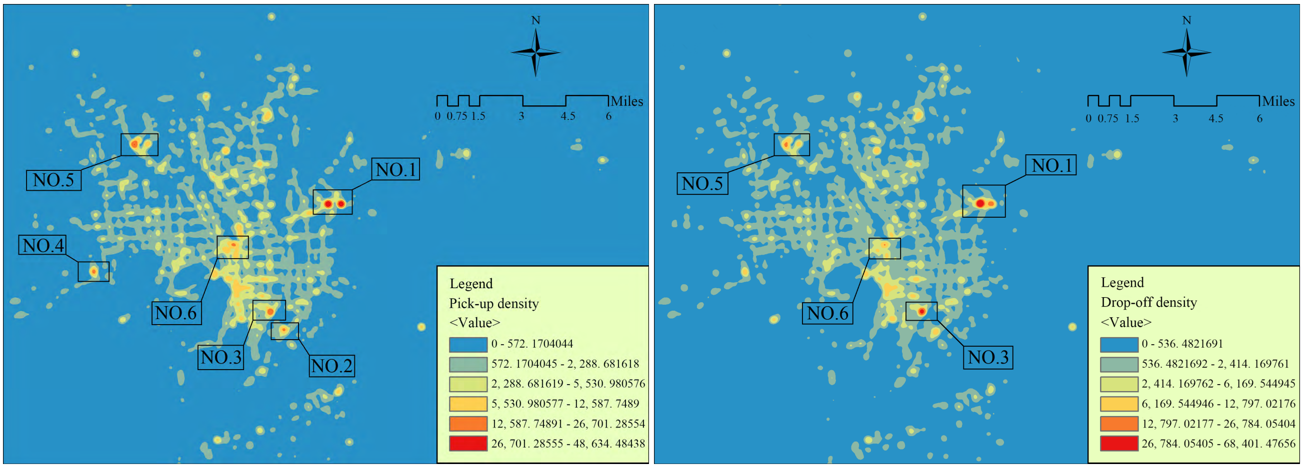
TABLE 2. Statistical indicators of the survey.

Notations	Signification	Annotation
T	Number of taxi trips	
p_i	Number of passengers per trip	$i = 1,2,\ldots\ldots,T$
n	Average number of passengers per trip	$n = \sum p_i / T$
t_{ai}	Pick-up time per trip(second)	$i = 1,2,\ldots\ldots,T$
t_{bi}	Drop-off time per trip(second)	$i = 1,2,\ldots\ldots,T$
t_{oci}	Time to open and close door per trip(second)	$i = 1,2,\ldots\ldots,T$

B. PARAMETERS CALCULATION

1) TAXI DEMANDS GENERATION

Fig. 7 depicts the road network structure of the hotspot No.6, covering 3,619 taxi trips on Friday. Generally, the distance between taxi stand and the signal intersection should be more than or equal to 50 meters. Thus, we take $r = 50$ to generate 48 demand subsections and candidate points, including 24 on the right and 18 on the left in CW. Note that taxi demand refers to the number of passengers rather than taxi trips in our work. To this end, an artificial survey was conducted in a week of June 2018 in Red-flag street business circle, Changchun city, China. The scale of the survey site selected is similar to the No.6 hotspot studied in this paper. The data collection time is 3 hours during the daily taxi travel peak, and the related statistical variables are shown in Table.2. According to the analysis results of the survey, the average passengers per taxi trip $n = 2$ is used in the calculation of taxi demands in each subsections. Then we have the taxi demand of each subsection $q_i = N_i^t \cdot n$, where N_i^t is the number of taxi trips occurred in each subsection i.



(a) (b)

FIGURE 6. **The distribution of taxi travel hotpots within the urban area. (a) Pick-up points. (b) Drop-off points.**

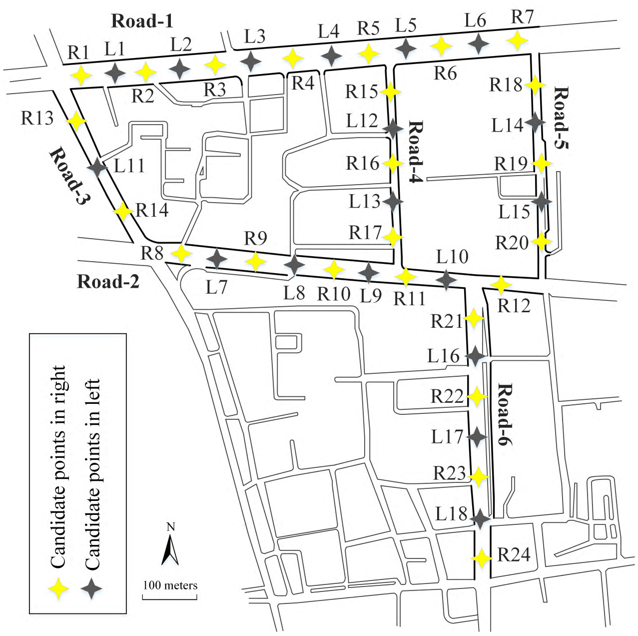


FIGURE 7. **The road network structure of hotspot NO.6.**

2) ECONOMIC INDEX DETERMINATION

Since different countries and regions hold different opinions on the number of parking spaces in taxi stands, we assume $p_j = 2$ for all candidate sites to simplify the TSLP. According to the actual investigation, the length of a taxi shelter is about 4 meters and the construction cost is about 1500 yuan per meter. In order to create a good waiting environment for the public and guide them to make effective use of taxi stands, we allocate corresponding taxi shelter for each parking space. Thus, $c_j = 12,000$ yuan is employed in the TSLM.

The free walking speed of pedestrians v_p is set as 1m/s in our research. The travelers' value of time (VOT) is related to their income: the VOT of taking a bus or car is estimated as 50% of a traveler' gross wage rate, and the walking and waiting time is 1.8 times of riding time [46]. Draw on the method of Zhu *et al.* [47], we assume that the average monthly gross wage rate of passengers is 20,000 yuan, the working time is 20 days a month and 8 hours a day, then we have $c_p = 112.5$ yuan per hour.

3) MAXIMUM SERVICE CAPACITY OF TAXI STAND

Assuming that the geometric size of TS is appropriate, the maximum number of taxis served per hour per parking space T_j^{bb} can be calculated by:

$$\mathrm{T}_j^{bb} = 3600\left(g/C\right)/\left[t_c + \left(g/C\right)t_d + z_a \cdot c_v \cdot t_\mathrm{d}\right] \quad (18)$$

where g/C is the effective green time in each signal cycle (the roadside taxi stand is 1.0), t_c is the time interval between two consecutive taxis (unit: second), Z_a is the unilateral test quantity corresponding to the probability of queuing at a taxi stand, and c_v is the deviation coefficient of residence time. The mean residence time t_d is given by:

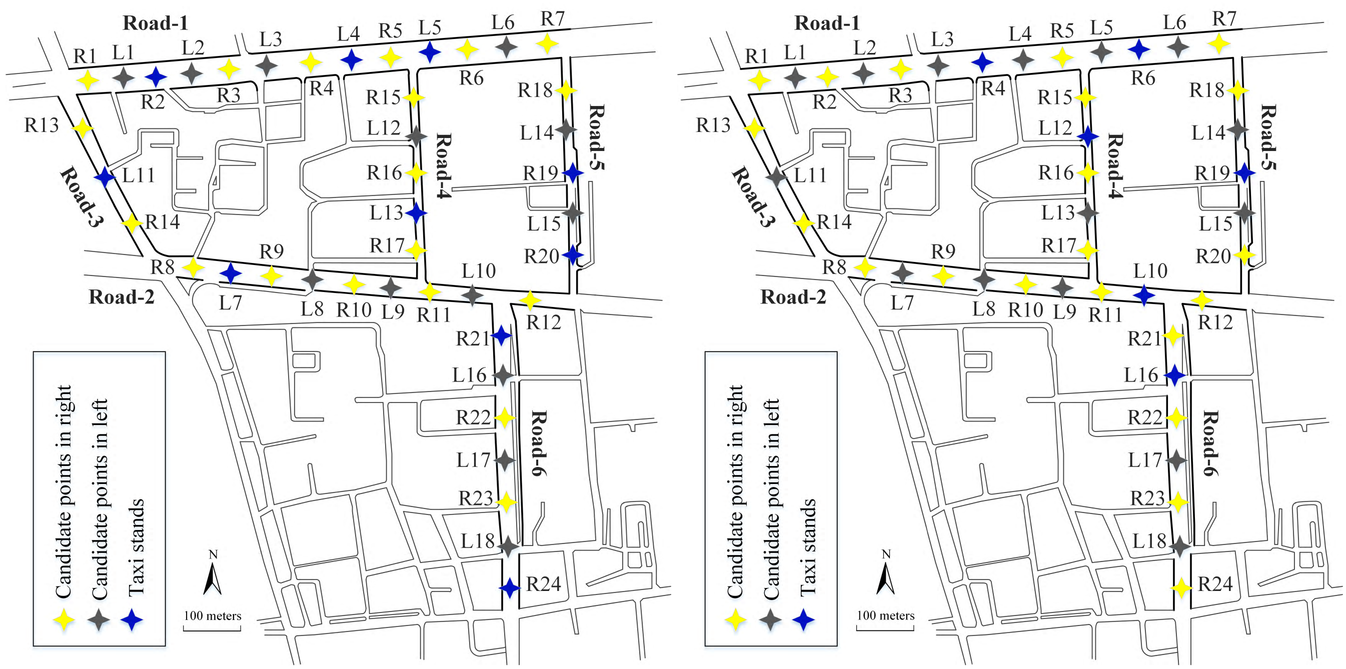
$$t_d = t_e + n \cdot t_a + n \cdot t_b + t_l + t_{oc} \quad (19)$$

where t_e and t_l respectively represent the time taken for a taxi to enter and leave the parking space, t_a and t_b respectively denote the alighting and the boarding time for a passenger, t_{oc} is the time to open and close door. In this paper, partial values were obtained through the investigation experiment in section III, and relevant parameter setting is given in Table.3.

TABLE 3. **The parameters setting.**

Notations	t_c	t_e	t_l	t_a	t_b	t_{oc}	Z_a
Value	3	2	5	4	4	3	0

Based upon (18) and (19), the maximum service capacity of taxi stand $P_j^{bb} = n \cdot T_j^{bb}$. If necessary, we recommend



(a) (b)

FIGURE 8. The spatial distribution of taxi stands with different demand coverage *Dc*. (a) *Dc* = 100%. (b) *Dc* = 80%.

a more appropriate numerical value for related parameter relying on the actual conditions in different cities and regions. However, this is not the focus of our research, and the current method using the parameters derived from specifications and actual survey is sufficient for us to obtain the appropriate results for solving TSLM.

TABLE 4. The optimal results of the TSLM with different demand coverage.

Optimal results	*Dc*=100%	*Dc*=80%
Number of taxi stands	10	6
Total construction costs of taxi stands (yuan)	120000	72000
Total walking distances for passengers (meter)	565118.58	624348.52
Total number of taxi demands covered by taxi service(person)	7238	5855
Average walking distance (meter)	78.08	106.64
Objective function value(yuan)	137518.68	91354.80

C. MODEL ESTIMATION AND ANALYSIS

According to relevant document [48], we set the maximum acceptable walking distance for passengers *DA* to 300m. The optimization results of the TSLM with different demand coverage *Dc* are summarized in Table.4. It indicates that a better solution can be obtained by excluding 20% of the taxi demands from the coverage target. As the demand coverage decreases from 100% to 80%, the total costs decreases from 137518.68 yuan with 10 taxi stands to 91354.80 yuan with 6 stands. Meanwhile, the total walking distances for passengers increase from 565118.58 m to 624348.52 m and the average walking distance also increase from 78.08 m to 106.64 m, which is due to the longer route to a stand with a fewer number of taxi stands. It should be noted that the decrease of taxi stands brings on an economy in the construction costs, the increase of walking distances leads to an addition in the access costs, while the total costs is actually reduced. This is mainly because the unit construction expenses of taxi stands is relatively large, which has a great impact on the objective function value. This also verifies that in order to minimize the total costs, taxi stands should be constructed as few as possible to reduce the construction costs on the premise of ensuring the demand coverage.

The spatial distribution of taxi stands is displayed in Fig. 8, the 10 construction points with $Dc = 100\%$ is located in subsection R2, L4, L5, L7, L11, L13, R19, R20 and R21 and R24, while the 6 construction points with $Dc = 80\%$ is located in subsection R4, R6, L10, L12, R19, L16.

Comparing Fig. 8a with Fig. 8b, we can find that the full coverage scenario requires an additional taxi stand on four roads of Road-1, 3, 5, 6 to serve the travel demands of all subsections. Fig. 9 depicts the coverage results of taxi demands in each subsections, which illustrates that passengers will take a taxi at their own subsections where taxi stands will be constructed. In 80% coverage scenario, taxi demands of subsections R8, L7, R13, L11, R14 and R24 can not be covered by taxi service, as the distances between their

Dc=100%	R2	R2	R2	R2	R2	L4	L4	L4	L4	L5	L5	L5	R 19	L7	L7	L7	L7	L 13	L 13	R 21	R 21
Subsection	R1	L1	R2	L2	R3	L3	R4	L4	R5	L5	R6	L6	R7	R8	L7	R9	L8	R 10	L9	R 11	L 10
Dc=80%	R4	R4	R4	R4	R4	R4	R4	R4	R4	R6	R6	R6	R6	×	×	L 10	L 10	L 10	L 10	L 10	L 10

Dc=100%	R 21	L 11	L 11	L 11	L5	L5	L 13	L 13	L 13	R 19	R 19	R 19	R 19	R 20	R 21	R 21	R 21	R 21	R 24	R 24	R 24
Subsection	R 12	R 13	L 11	R 14	R 15	L 12	R 16	L 13	R 17	R 18	L 14	R 19	L 15	R 20	R 21	L 16	R 22	L 17	R 23	L 18	R 24
Dc=80%	L 10	×	×	×	L 12	L 12	L 12	L 12	L 10	R 19	R 19	R 19	R 19	R 19	L 10	L 16	L 16	L 16	L 16	L 16	×

FIGURE 9. The coverage results of taxi demands in each subsections.

locations and the nearest taxi stands exceed the maximum acceptable range. In such cases, passengers can choose to take buses or subways, drive private cars or walk to areas without taxi parking restrictions and other alternative trips modes to achieve travel.

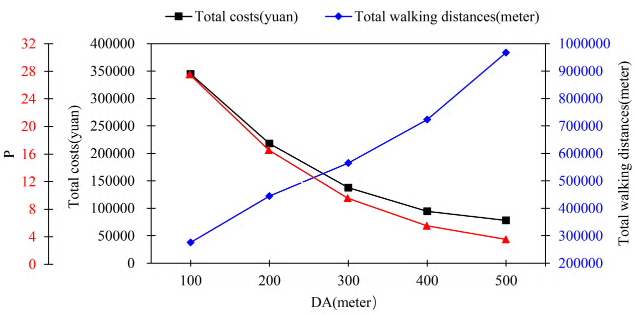


FIGURE 10. The impact of maximum acceptable walking distance DA for passengers on TSLP.

D. IMPACT OF MAXIMUM ACCEPTABLE WALKING DISTANCE

Furthermore, we also explore the impact of maximum acceptable walking distance DA for passengers on TSLP under the identical demand coverage $Dc = 100\%$, which is shown in Fig. 10. The relationship curve between DA and total cost shows a monotonic decreasing trend, suggesting that the larger the DA is, the better the objective value is. However, the monotone decline in changing rate of the curve indicates that the increase in the maximum acceptable walking distance reduces the marginal benefit. In the process of DA increasing from 100 m to 500 m, the total cost saved is 126768.68, 80252.01, 43098.14, and 16432.71 yuan correspondingly, and the reduced number of taxi stands is 11, 7, 4 and 2 respectively.

The total walking distance increases from 275463.40 m to 967349.04 m with the increase of DA, which is consistent with our intuition. The smaller DA states that passengers are reluctant to travel a long way to take a taxi. Therefore, it is necessary to set more TSs close to the taxi demands, which increases the construction costs. The larger DA indicates that passengers can accept a longer distance to take a ride. Therefore, there are more choices to decide where to locate TSs. In order to minimize the total cost, fewer TSs should be set up, which induces an increase of total walking distances. These comparison results verify that maximum acceptable walking distance has a significant impact on the location selection and demand coverage of TSs.

IV. CONCLUSION

This paper proposed a three-stage location strategy to address the taxi stands location problem (TSLP) in urban, which can determine appropriate construction locations, rather than merely evaluate a rational quantity. In stage1, we preprocessed massive taxi GPS data and extracted "0" and "1" data sets representing the pick-up and drop-off location. With the spatial-temporal analysis of demand data, the taxi travel hotspots were identified as the potential settings areas for taxi stand (TS). In stage2, subsections representing travel demands were staggered on the road network based on the spatial interaction between taxi demands and taxi stands. The center of each subsection serves as the candidates for TSs. In stage3, a taxi stand location model (TSLM) was established to minimize the construction cost and the access cost for passengers. The case analysis results in China proves the validity of the location strategy we proposed. The optimal construction scheme under different demand coverage can be obtained by solving the TSLM, which can effectively balance the interests of the government and the public and support the site decision of TS in urban. In addition, the impact of the maximum walking distance passengers can accept on the location selection of TSs is also discussed.

The contribution of our study lies in the three-stage location strategy for the TSLP. Prior to this study, almost no research on the exact location selection of TSs were present. This paper attempts to develop a scientific basis for decision makers to evaluate the location choices of TSs in cities with the help of hotspot analysis and location modeling. Nevertheless, there are two main limitations in our research. Firstly, passengers in the same subsection are restricted to go to the same TS. In practice, under the influence of travel time, travel distance, travel direction and personal attributes of the traveler, passengers in the same point may choose to go to different TSs. Hence, it is necessary to model the passenger's choice behavior for TSs in the future. Secondly, there is no clear specification and standard to provide reasonable suggestions for the setting separation distance of TSs at present. Therefore, the length range we choose in the division of subsections remains to be further discussed in future research work.

REFERENCES

[1] E. V. Ocalir, O. Y. Ercoskun, and R. Tur, "An integrated model of GIS and fuzzy logic (FMOTS) for location decisions of taxicab stands," *Expert Syst. Appl.*, vol. 37, no. 7, pp. 4892–4901, Jul. 2010.

[2] H.-W. Chang, Y.-C. Tai, and Y.-J. Hsu, "Context-aware taxi demand hotspots prediction," *Int. J. Bus. Intell. Data Mining*, vol. 5, no. 1, pp. 3–18, 2010.

[3] L. Moreira-Matias, J. M endes-Moreira, M. Ferreira, J. Gama, and L. Damas, "An Online learning framework for predicting the taxi stand's profitability," in *Proc. 17th Int. IEEE Conf. Conf. Intell. Transp. Syst.*, Qingdao, China, Oct. 2014, pp. 2009–2014.

[4] R. C. P. Wong, W. Y. Szeto, and S. C. Wong, "Bi-level decisions of vacant taxi drivers traveling towards taxi stands in customer-search: Modeling methodology and policy implications," *Transp. Policy*, vol. 33, no. 2, pp. 73–81, May 2014.

[5] R. W. Giuliani, J. B. Rose, and I. Weinshall, "Taxi stands in times square and the theater district a technical memorandum for the midtown manhattan pedestrian network development project," Dept. City Planning, Dept. Transp., New York, NY, USA, Tech. Rep., 2001. [Online]. Available: http://www.docin.com/p-947506856.html

[6] Z. Xia and J. Yan, "Kernel Density estimation of traffic accidents in a network space," *Comput. Environ. Urban Syst.*, vol. 32, no. 5, pp. 396–406, Sep. 2008.

[7] N. J. Yuan, Y. Zheng, L. Zhang, and X. Xie, "T-finder: A recommender system for finding passengers and vacant taxis," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 10, pp. 2390–2403, Oct. 2013.

[8] X. Liu, L. Gong, Y. Liu, and Y. Gong, "Revealing travel patterns and city structure with taxi trip data," *J. Transp. Geogr.*, vol. 43, pp. 78–90, Feb. 2015.

[9] X. Qian and S. V. Ukkusuri, "Spatial variation of the urban taxi ridership using GPS data," *Appl. Geogr.*, vol. 59, pp. 31–42, May 2015.

[10] P. S. Castro, D. Zhang, C. Chen, S. Li, and G. Pan, "From taxi GPS traces to social and community dynamics: A survey," *ACM Comput. Surv.*, vol. 46, no. 2, p. 17, Nov. 2013.

[11] D. Zhang *et al.*, "Understanding taxi service strategies from taxi GPS traces," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 1, pp. 123–135, Feb. 2015.

[12] J. Yuan, Y. Zheng, X. Xie, and G. Sun, "T-drive: Enhancing driving directions with taxi drivers' intelligence," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 1, pp. 220–232, Jan. 2013.

[13] C. Chen, S. Jiao, S. Zhang, W. Liu, L. Feng, and Y. Wang, "TripImputor: Real-time imputing taxi trip purpose leveraging multi-sourced urban data," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 10, pp. 3292–3304, Oct. 2018.

[14] L. Gong, X. Liu, L. Wu, and Y. Liu, "Inferring trip purposes and uncovering travel patterns from taxi trajectory data," *Cartogr. Geogr. Inf. Sci.*, vol. 43, no. 2, pp. 103–114, Mar. 2016.

[15] L. Moreira-Matias, J. Gama, M. Ferreira, J. Mendes-Moreira, and L. Damas, "Predicting taxi–passenger demand using streaming data," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 3, pp. 1393–1402, Sep. 2013.

[16] N. Ferreira, J. Poco, H. T. Vo, J. Freire, and C. T. Silva, "Visual exploration of big spatio-temporal urban data: A study of New York city taxi trips," *IEEE Trans. Vis. Comput. Graphics*, vol. 19, no. 12, pp. 2149–2158, Dec. 2013.

[17] J. Bischoff, M. Maciejewski, and A. Sohr, "Analysis of Berlin's taxi services by exploring GPS traces," in *Proc. Int. Conf. Models Technol. Intell. Transp. Syst.*, Jun. 2015, pp. 209–215.

[18] J. Tang, F. Liu, Y. Wang, and H. Wang, "Uncovering urban human mobility from large scale taxi GPS data," *Phys. A, Stat. Mech. Appl.*, vol. 438, pp. 140–153, Nov. 2015.

[19] J. Lee, I. Shin, and G.-L. Park, "Analysis of the passenger pick-up pattern for taxi location recommendation," in *Proc. 4th Int. Conf. Netw. Comput. Adv. Inf. Manage.*, vol. 11, Sep. 2008, pp. 199–204.

[20] G. Qi *et al.*, "How long a passenger waits for a vacant taxi–large-scale taxi trace mining for smart cities," in *Proc. IEEE Int. Conf. Green Comput. Commun., IEEE Internet Things, IEEE Cyber, Phys. Social Comput.*, Aug. 2013, pp. 1029–1036.

[21] J. Tang, J. Liang, S. Zhang, H. Huang, and F. Liu, "Inferring driving trajectories based on probabilistic model from large scale taxi GPS data," *Phys. A, Stat. Mech. Appl.*, vol. 506, pp. 566–577, Sep. 2018.

[22] R.-H. Hwang, Y.-L. Hsueh, and Y.-T. Chen, "An effective taxi recommender system based on a spatio-temporal factor analysis model," *Inf. Sci.*, vol. 314, pp. 28–40, Sep. 2015.

[23] Z. Luo *et al.*, "Dynamic taxi service planning by minimizing cruising distance without passengers," *IEEE Access*, vol. 6, pp. 70005–70016, 2018.

[24] X. Wang, H. Zhang, L. Wang, and Z. Ning, "A demand-supply oriented taxi recommendation system for vehicular social networks," *IEEE Access*, vol. 6, pp. 41529–41538, 2018.

[25] L. Moreira-Matias, R. Fernandes, J. Gama, M. Ferreira, J. Mendes-Moreira, and L. Damas, "An online recommendation system for the taxi stand choice problem (Poster)," in *Proc. Veh. Netw. Conf.*, Nov. 2013, pp. 173–180.

[26] R. L. Church, "Geographical information systems and location science," *Comput. Oper. Res.*, vol. 29, no. 6, pp. 541–562, May 2002.

[27] S. L. Hakimi, "Optimum locations of switching centers and the absolute centers and medians of a graph," *Oper. Res.*, vol. 12, no. 3, pp. 450–459, Jun. 1964.

[28] R. Church and C. ReVelle, "The maximal covering location problem," *Papers Regional Sci. Assoc.*, vol. 32, no. 1, pp. 101–118, Dec. 1974.

[29] M. J. Hodgson, "A flow-capturing location-allocation model," *Geogr. Anal.*, vol. 22, no. 3, pp. 270–279, Jul. 1990.

[30] M. Kuby and S. Lim, "The flow-refueling location problem for alternative-fuel vehicles," *Socio-Econ. Planning Sci.*, vol. 39, no. 2, pp. 125–145, Jun. 2005.

[31] M. A. Nicholas, S. Handy, and D. Sperling, "Using geographic information systems to evaluate siting and networks of hydrogen stations," *Transp. Res. Rec.*, vol. 1880, pp. 126–134. Jan. 2004.

[32] M. Kuby and S. Lim, "Location of alternative-fuel stations using the flow-refueling location model and dispersion of candidate sites on arcs," *Netw. Spatial Econ.*, vol. 7, no. 2, pp. 129–152, Jun. 2006.

[33] C. Upchurch, M. Kuby, and S. Lim, "A model for location of capacitated alternative-fuel stations," *Geogr. Anal.*, vol. 41, no. 1, pp. 85–106, Jan. 2009.

[34] X. Xi, R. Sioshansi, and V. Marano, "Simulation–optimization model for location of a public electric vehicle charging infrastructure," *Transp. Res. D, Transp. Environ.*, vol. 22, pp. 60–69, Jul. 2013.

[35] N. Shahraki, H. Cai, M. Turkay, and M. Xu, "Optimal locations of electric public charging stations using real world vehicle travel patterns," *Transp. Res. D, Transp. Environ.*, vol. 41, pp. 165–176, Dec. 2015.

[36] J. Yang, J. Dong, and L. Hu, "A data-driven optimization-based approach for siting and sizing of electric taxi charging stations," *Transp. Res. C, Emerg. Technol.*, vol. 77, pp. 462–477, Apr. 2017.

[37] B. Farhana and A. T. Murray, "Siting park-and-ride facilities using a multi-objective spatial optimization model," *Comput. Oper. Res.*, vol. 35, no. 2, pp. 445–456, Feb. 2008.

[38] A. Khakbaz, A. S. Nookabadi, and S. N. Shetab-bushehri, "A model for locating park-and-ride facilities on urban networks based on maximizing flow capture: A case study of Isfahan, Iran," *Netw. Spatial Econ.*, vol. 13, no. 1, pp. 43–66, Mar. 2013.

[39] F. Aros-Vera, V. Marianov, and J. E. Mitchell, "*p*-hub approach for the optimal park-and-ride facility location problem," *Eur. J. Oper. Res.*, vol. 226, no. 2, pp. 277–285, Apr. 2013.

[40] J. C. García-Palomares, J. Gutiérrez, and M. Latorre, "Optimizing the location of stations in bike-sharing programs: A GIS approach," *Appl. Geogr.*, vol. 35, nos. 1–2, pp. 235–246, Nov. 2012.

[41] I. Frade and A. Ribeiro, "Bike-sharing stations: A maximal covering location approach," *Transp. Res. A, Policy Pract.*, vol. 82, pp. 216–227, Dec. 2015.

[42] D. Zhang, C. Yu, J. Desai, H. Y. K. Lau, and S. Srivathsan, "A time-space network flow approach to dynamic repositioning in bicycle sharing systems," *Transp. Res. B, Methodol.*, vol. 103, pp. 188–207, Sep. 2017.

[43] W. Tu, Q. Li, Z. Fang, S.-L. Shaw, B. Zhou, and X. Chang, ''Optimizing the locations of electric taxi charging stations: A spatial–temporal demand coverage approach,'' *Transp. Res. C, Emerg. Technol.*, vol. 65, pp. 172–189, Apr. 2016.

[44] J. H. Holland, *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. Ann Arbor, MI, USA: Univ. of Michigan Press, 1975.

[45] K. F. Man, K. S. Tang, and S. Kwong, ''Genetic algorithms: Concepts and applications,'' *IEEE Trans. Ind. Electron.*, vol. IE-43, no. 5, pp. 519–534, Oct. 1996.

[46] R. Wang, ''Autos, transit and bicycles: Comparing the costs in large Chinese cities,'' *Transp. Policy*, vol. 18, no. 1, pp. 139–146, Jan. 2011.

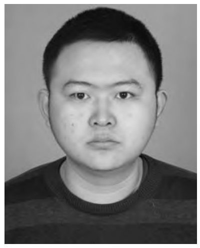
[47] Z.-H. Zhu, Z.-Y. Gao, J.-F. Zheng, and H.-M. Du, ''Charging station location problem of plug-in electric vehicles,'' *J. Transp. Geogr.*, vol. 52, pp. 11–22, Apr. 2016.

[48] *Code for Transport Planning on Urban Road*, China National Standard GB 50220-95, 1995.

ZHAOWEI QU received the Ph.D. degree from the School of Transportation, Jilin University, Changchun, China, where he is currently a Professor and a Ph.D. Supervisor. His current research interests mainly include traffic control theory, traffic organization, traffic big data, and intelligent transportation systems.

XIN WANG received the B.Sc. degree from the School of Transportation, Jilin University, Changchun, China, in 2016, where she is currently pursuing the Ph.D. degree in traffic engineering. Her research interests focus on traffic organization, traffic data analysis, and its applications related approach.

XIANMIN SONG received the M.Sc. and Ph.D. degrees in traffic engineering from Jilin University, Changchun, China, in 2005 and 2008, respectively, where she is currently a Professor with the School of Transportation. Her research interests include intelligent traffic control, traffic flow theory, and traffic big data.



ZHAOTIAN PAN received the B.Sc. degree from the School of Transportation, Jilin University, Changchun, China, in 2016, where he is currently pursuing the Ph.D. degree in traffic engineering. His research interests include traffic flow theory and traffic information processing.

HAITAO LI received the B.Sc. degree from the School of Transportation, Jilin University, Changchun, China, in 2016, where he is currently pursuing the Ph.D. degree in traffic engineering. His research interests include pattern recognition and traffic information processing.