International Journal of Innovative Computing, Information and Control
Volume **18**, Number **2**, April **2022**
ICIC International ©2022 ISSN 1349-4198
pp. 477–495

A SITING URBAN TAXI STATIONS MODEL BASED ON SPATIAL-TEMPORAL ORIGIN-DESTINATION DATA

Duo Tian[1], Jialing Lu[1] and Zhicheng Wei[1,2,*]

[1]College of Computer and Cyberspace Security
[2]Key Lab of Network and Information Security of Hebei Province
Hebei Normal University
No. 20, South Second Ring East Road, Yuhua District, Shijiazhuang 050024, P. R. China
{ tianduo; lujl }@stu.hebtu.edu.cn
*Corresponding author: weizhicheng@hebtu.edu.cn

Received October 2021; revised January 2022

Abstract. *To accommodate travel demands of residents and alleviate environmental impact of taxis, it is imperative to provide a plausible scheme for the construction of taxi stations in city. Moreover, it is more valuable to have stations in hotspots where there is more travel demand. The key to site taxi stations is stations location selection and parking spaces allocation in each station. The spatial-temporal origin-destination (OD) data are adopted, which can accurately locate passenger pick-up and drop-off locations, providing crucial data support for the establishment of taxi stations. Firstly, we extracted hotspot areas by calculating the number of trips and their economic value of each area to obtain a set of candidate areas. Given that site siting problem has the property of NP-Hard, we employ a heuristic genetic algorithm to derive an approximately feasible solution. We define the objective function by maximizing the total profit of stations and use queuing theory to determine the number of parking spaces in each station. Finally, the objective function is optimized through genetic operation in a finite number of iterations, with the optimal result being picked after several repetitions of the experiment. Our approach yielded promising performance on the Shijiazhuang dataset. We also analyzed the effect of the distances between demand area and station area, waiting times on station results. Our model was compared with a loss-based queueing theory model, and the results showed that our model performed significantly better in terms of objective function values. We also propose an improved adaptive genetic algorithm (AGA) to speed up the evolution of population. The findings indicate that our proposed methodology is well qualified to offer recommendations and reference for transport strategies and government decision-making.*
Keywords: Site selection for taxi stations, Adaptive genetic algorithm, Queueing theory, Traffic optimization

1. **Introduction.** Aiming to actively respond to climate change and promote low-carbon ecological environment, China strives to peak carbon dioxide emissions by 2030 and carbon neutrality by 2060. As an indispensable transport for daily travel, taxi is also a major contributor to carbon dioxide emissions. How to effectively reduce the carbon emissions caused by taxis is an important issue to be addressed. With the development of the Internet economy, people hail taxis using mobile apps. So there is no need for taxis without passenger to move around the streets. Where should these taxis be parked? The answer is taxi stations.

However, there are situations where the passenger who needs a taxi is located far away from a taxi station, causing driver dispatch costs and taxi carbon emissions increase. So taxi stations locations are the key problem. Through the proper deployment of taxi

DOI: 10.24507/ijicic.18.02.477

stations, the travel demand in each area can be accommodated more economically and conveniently.

The issue of siting stations in cities has been discussed by many researchers, for example, discussing taxi data analysis and taxi station hotspot detection [1-6], optimizing the location of taxi stations [7-9], establishing charging stations for electric taxis [10-19], and choosing sites for other public services [20-25]. Lyu et al. described the current status of big data research on traditional taxi and online car-hailing (TTOC) from six hot topics, providing an important reference for the transport department [1]. Zhang and Ukkusuri proposed a scalable taxi pooling clustering method. Through grouping similar taxi trips, potential taxi stand location can be obtained [2]. To accurately predict taxi demand, Liu et al. proposed three machine learning methods to predict the taxi demand located in hotspot areas [3]. Hu et al. used taxi OD data to construct a passenger evaluation model to identify high quality passengers and potentially high profit areas [4]. Riascos and Mateos used taxi OD data to analyze taxi trip datasets from New York City [5]. Shen et al. used the improved network kernel density estimation (imNKDE) method to reveal the spatial-temporal patterns of taxi trips [6]. All these studies provide important references for station siting.

There are few studies that consider the impact of the benefits generated by setting up the taxi station on the site's deployment. We have the following advantages in considering site selection combined with the benefits of establishing the station. On the one hand, setting taxi stations in areas with high travel benefits is better for serving more passengers, enabling station construction costs to be recovered in relative short time. On the other hand, station construction benefits involve taxi dispatching costs in corresponding coverage area, while a reasonable taxi station layout can reduce dispatching costs and allow passengers near station to call a taxi faster. Therefore, this paper focuses on the following key aspects: (1) Using taxi OD data, identify high demand and high economic value hotspots; (2) Design cost-effective, high-efficiency and high-value station construction solutions using genetic algorithms and queuing theory methods; (3) Analyze the impact of different constraints such as distance between sites, and queuing times, on the siting results.

The remainder of the paper is organized as follows. Section 2 provides an overview of the related works. Section 3 presents the definitions and the framework of our model. Section 4 describes the specific methods and mathematical descriptions. Section 5 shows the data and experiments. Conclusions and future work are in Section 6.

2. Related Works.

In the era of big data, taking taxis is one of the preferred ways to travel in a city. Using the massive volume of available data resources to optimize the distribution of taxis in cities is an important issue of urban transport. Qu et al. proposed a three-stage model to build taxi stations in a certain area, which uses a genetic algorithm to optimize the model based on taxi trajectory data [7]. Wang et al. proposed quantum annealing (QA) and brain-inspired clustering algorithm (QABICA) to solve the problem of urban taxi-stand locations [8]. Wang et al. compared the performance of three classical siting models in optimizing site selection [9]. These papers are mainly designed to solve the problem of setting up stations in a particular small area. Its data comes from taxi trajectory data. It neither accurately identifies residents' travel demands nor provides taxi services for the entire city.

2.1. Siting electric taxi charging stations.

With the rapid development of electric vehicles, the problem of siting electric vehicle charging stations is also a hot topic of current siting research. Asamer et al. proposed a method for placing charging stations in cities that focuses on finding areas where charging stations should be placed, rather than exact locations [10]. Tu et al. proposed a spatial-temporal demand coverage approach to optimize the spatial-temporal layout of electric taxi (ET) charging stations with the objective function that maximizes the level of ET service and the level of charging service [11]. Yang et al. proposed a model to provide chargers for battery electric vehicle (BEV). It uses integer linear program (ILP) and queuing theory to site charging stations, and its corresponding objective function is to minimize the investment in infrastructure [12]. Ghosh et al. proposed a hexagonal fuzzy multiple-criteria decision-making (MCDM) methodology for electric vehicle charging station site selection. The fuzzy technique for order of preference by similarity to ideal solution (FTOPSIS) and fuzzy complex proportional assessment (FCOPRAS) were used to rank the selected sites [13]. Kaya et al. proposed a method to solve the site selection problem of electric taxi charging station (ETCS) for Istanbul. They conducted their study using fuzzy analytic hierarchy process (fuzzy AHP) and technique for order preference by similarity to ideal solution (TOPSIS) methods [14]. Lin et al. also proposed a picture fuzzy multi-criteria decision making (MCDM) model to solve the car sharing stations siting problem [15]. The above study cares about selecting charging stations for electric taxis. It emphasizes solving charging problems rather than satisfying residents' travel needs.

There are studies that combine charging demand with other factors providing a comprehensive approach to site charging stations. Wang et al. developed a fast-charging facility planning model based on consideration of battery degradation and vehicle heterogeneity in driving range [16]. Zhang et al. proposed a new two-stage location model based on k-means and barycentric method, considering the dynamic distribution of charging stations and charging demand together [17]. Meng et al. proposed a charging station construction method to take social factors into account. It can expand or downsize the scale of charging station construction when existing charging stations are not suitable for charging demand [18]. Zhang et al. proposed a hybrid charging management framework concerning the coexistence of plugin charging stations (CSs) and battery swap stations (BSSs) [19]. It enables electric taxis to be directed to suitable stations for charging or swapping depending on the urgency of demand. These papers consider the impact of station construction costs and suggest that charging stations can be altered to fit the usage of electric taxis. This provides thoughts on constructing taxi stations from an economic benefit perspective.

2.2. **Siting other public service facilities.** Additionally, there are also several site selections studies for other public service facilities. Zhao et al. proposed a method for site selection of "taxi canteens" to provide catering services to taxi drivers [20]. Its objective function is to minimize the total distance between the dining demand point and the "taxi canteen". Wei et al. proposed a multi-objective optimization model of site selection for bus gas stations [21]. Its objective function is to minimize the construction cost of the filling stations and the cost of refueling all buses. Nega et al. used survey research methods and geographic information system (GIS) tools to determine the toilet needs of residents in Debre Markos Town, Ethiopia and found suitable sites for public toilets [22]. Yu et al. proposed a multi-objective maximal covering location model using geospatial big data, which can accurately estimate fire rescue demand and travel costs [23]. Wang et al. proposed a digital signage site-selection model based on geographical location and multisource factor data using empirical location models and machine learning methods [24]. Alves et al. proposed an airport site selection methodology that combines GIS and analytic hierarchy process (AHP) through specific techniques that effectively reduce the subjectivity in airport siting decisions [25]. These papers use multiple datasets to site public facilities. The main objective is to reduce the distance of public facilities from the corresponding people in need or to increase the coverage of the facilities. However, their optimization models are mainly designed for specific target people. It cannot be fully implemented for constructing taxi stations.

We have found that the above site selection studies have some limitations. Most of the studies on site selection are based on taxi trajectory data, which cannot provide accurate information on passenger pick-up and drop-off. This in turn leads to an inaccurate analysis of passenger travel demand. The impact of the economic profit generated by taxis has rarely been considered in siting studies, resulting in station solutions that are not sensible enough for taxi drivers and travelling passengers. Therefore, we propose a method for siting taxi stations based on spatial-temporal OD data. Genetic algorithms can yield results in solving complex combinatorial optimization problems, which can lead to optimal results based on the fitness value [26,27]. Using genetic algorithms and queuing theory, we design a more reasonable objective function while considering economic benefits. Our model yields promising results on the Shijiazhuang dataset.

3. **Framework.** Imbalance in economic development leads to imbalance in the availability of public transport facilities in the city. As a result, different modes of transport are not equally attractive to residents with mixed travel needs. By capturing residents' demand for taxis and analyzing their community travel patterns, we propose a scheme for building taxi stands in cities. The genetic algorithms allow us to optimize distribution of sites in a reasonably efficient way. The queuing theory can allocate an appropriate number of taxi spaces to each station and significantly lower taxi energy waste.

3.1. **Definition.**

Definition 3.1. *In this paper, firstly we divide a city into 1 km $*$ 1 km grids and if a pick-up location of a trip falls on a grid, it belongs to that grid. Area traffic flow (ATF) is total number of trips in each grid during a period, i.e., one day. The value of ATF can be denoted as* $P = \{p_1, p_2, \ldots, p_n\}$*, where* n *denotes the number of grids and* p_i *is the ATF value of the* i*th grid.*

Definition 3.2. *Area average traffic flow (AATF) refers to the average traffic flow for a small time scale after dividing a long period equally into short periods. For example, by dividing one day into 24 hours, the average hour traffic flow is AATF. The value of AATF can be denoted as* $PA = \{pa_1, pa_2, \ldots, pa_n\}$*, where* n *denotes the number of grids and* pa_i *is the AATF value of the* i*th grid.*

Definition 3.3. *Further, travel time and distance of each trip during one day were counted. The cost of each trip is calculated according to the taxi charging rules given by the online taxi platform, as shown in Table 1. The trip costs for the same grid are accumulated to get the area traffic flow cost (ATFC), which can be expressed as* $PV = \{pv_1, pv_2, \ldots, pv_n\}$*, where* n *denotes the number of grids and* pv_i *is the ATFC value of the* i*th grid.*

3.2. **The model framework.** Figure 1 illustrates the overall framework of our model. First, we cleaned and pre-processed raw data based on trip OD data and taxi charging rules. Our model partitions a city into 1 km $*$ 1 km grids of equal size. Each grid is then considered as a potential area where taxi station can be built. The number of trips in each potential station area is counted using trip's pick-up point. Next, we count the number and cost of all trips per unit time in each grid, which means we calculate the value of ATF and ATFC for each area. By setting standards, we selected areas with high number of trips and high cost of travels as candidate station areas. Finally, genetic algorithm

Table 1. Taxi charging rules

Workdays			
	Flag-fall price (yuan)	Mileage fee (yuan/km)	Duration fee (yuan/min)
Normal time	8.00	1.40	0.35
00:00-05:00	9.00	2.40	0.65
05:00-09:00	9.00	1.60	0.65
17:00-19:00	9.00	1.53	0.65
23:00-00:00	9.00	2.40	0.65
Long-range price	10 km-20 km (yuan/km)	0.44	
	over 20 km (yuan/km)	0.82	
The flag-fall price includes mileage of 2.9 km and duration of 8 minutes.			

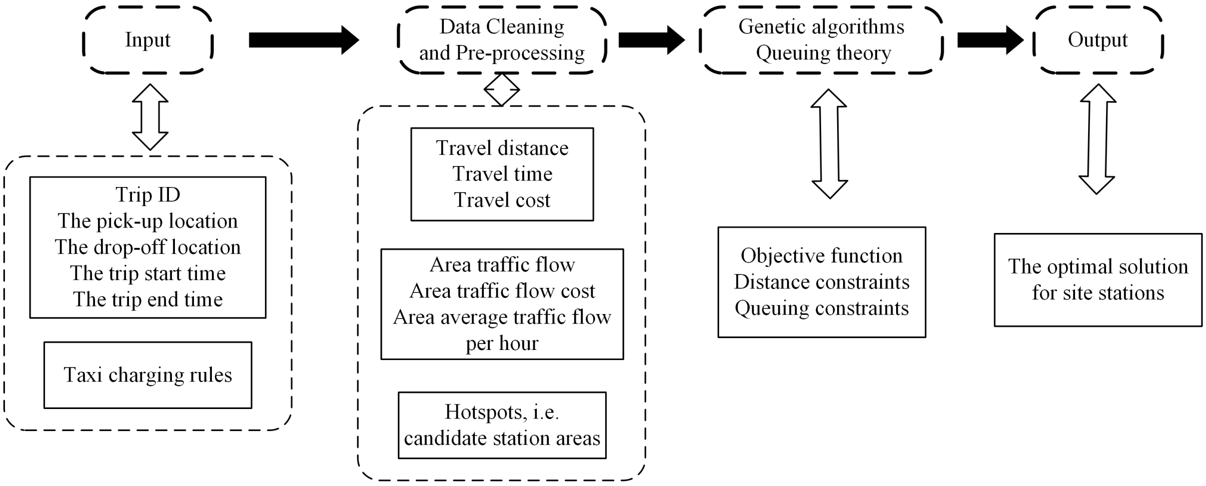


Figure 1. Model framework

was deployed to locate the ultimate station areas from candidate station areas. We also use the queuing theory to decide the number of spaces to be equipped with a station. After several rounds of iterative calculations, a superior and feasible taxi station solution is generated.

4. **Methodology.** We propose a genetic algorithm and queuing theory siting algorithm (GAQTSA). We firstly select candidate areas by setting values of ATF, ATFC. Then we use genetic algorithm to determine the distribution of ultimate station areas to cover the travel demand of entire candidate station area. It also incorporates the queuing theory to allocate the number of parking spaces for each station area. In the end, the best solution was output after several iterations.

4.1. **Stations location selection and parking spaces allocation.** The key to siting taxi stations is stations location selection and parking spaces in each station. In station assignment stage, we aim to establish a convenient, low cost and efficient solution for entire area with limited budget. Given $|D|$ candidate station areas and select $|A|$ areas from them as ultimate station areas. The goal of the objective function is to minimize construction cost, minimize the distance between station area and non-station area and maximize overall benefits.

In the taxi parking spaces allocation process, we target to allocate the appropriate number of spaces for each station area. We calculate the Manhattan distance d_{ij} for each non-station area from each station area. For each non-station area, the station area with the smallest d_{ij} is chosen as its taxi service provider and the travel demand generated by the non-station area is accumulated to the corresponding station area. We call this relationship as station area binding. Each station area and its binding non-station area are called the coverage area of that station area. In order to satisfy the travel demand of the coverage areas, we use queuing theory's M/M/s waiting system model to allocate spaces number in each station area. Thus, our algorithm can also be called wait-based genetic algorithm and queuing theory siting algorithm (Wait-based GAQTSA).

4.2. **Objective function and variables.** The objective function and constraints are as follows. The definitions of relevant variables and symbols are shown in Table 2.

The objective function (1) is to maximize net profitability of station construction, where the first term is the profit of each candidate station area, the second term is the cost of building and allocating spaces in each station area, and the third term is the cost of the

TABLE 2. Symbol description

Sets	D	The set of demand areas.
	A	The set of station areas.
Constants	$\|D\|$	The number of demand areas.
	$\|A\|$	The number of station areas.
Indexes	i	Index of the demand area.
	j	Index of the station area.
Parameters	MD	The furthest Manhattan distance allowed between the demand area and its corresponding station area.
	m_j	The number of spaces equipped in the station area.
	d_{ij}	Manhattan distance between the demand area i and its corresponding station area j.
	F_1	Basic cost of building a station.
	F_2	Cost of building a parking space.
	F_3	Taxi travel cost per km.
	V_i	The financial benefits generated by demand area i, namely ATFC.
	O_i	The average number of trips per hour generated in demand area i, namely AATF.
	Λ_j	Arrival rate. The number of trips per hour obtained in the station area j.
	μ	Service rate. The number of taxis that can be dispatched per hour per space (or the number of trips that can be handled), the reciprocal of which is the service time.
	ρ	Service strength per parking space.
	ρ_j	The service strength of the station area j.
	F	Total construction cost, i.e., the maximum cost of the station solution, including the cost of the basis of the station and the cost of the parking space.
	T	Maximum waiting time allowed for a trip.
Decision variables	X_j	1 if area j built the station.
	Y_{ij}	1 if the demand area i is closest to the station area j.

taxi from the station area to the binding non-station areas. The objective function (1)'s main goal is to get the maximum station construction benefits by optimizing the number of stations, the number of parking spaces and their distribution.

$$\text{Maximize: } \sum_{i\in D} V_i - \sum_{j\in A} X_j(F_1 + F_2 m_j) - \sum_{\substack{i\neq j\\ i\in D\\ j\in A}} Y_{ij} d_{ij} F_3 O_i \tag{1}$$

$$\text{Subject to: } \sum_{\forall j\in A} X_j(F_1 + F_2 m_j) \le F \tag{2}$$

$$Y_{ij} d_{ij} \le MD, \quad \forall i \in D,\ \forall j \in A \tag{3}$$

$$W_q \le T \tag{4}$$

$$\Lambda_j = \sum_{i\in D} O_i Y_{ij} \tag{5}$$

$$\rho = \frac{\Lambda_j}{\mu}, \quad \rho_j = \frac{\Lambda_j}{\mu \cdot m_j} < 1 \tag{6}$$

$$P_0 = \left[\sum_{n=0}^{m_j-1} \frac{\rho^n}{n!} + \frac{\rho^{m_j}}{m_j!(1-\rho_j)}\right]^{-1} \tag{7}$$

$$L_q = \frac{P_0 \rho^{m_j} \rho_j}{m_j!(1-\rho_j)^2} \tag{8}$$

$$W_q = \frac{L_q + \rho}{\Lambda_j} - \frac{1}{\mu} \tag{9}$$

Constraint (2) is the total station construction cost constraint, which ensures that the total cost of the station construction is not exceeded. Constraint (3) is the station distance constraint, ensuring that the radius of coverage in the station area is not less than MD. Constraint (4) is the queue time constraint, ensuring that the waiting time for each trip would not be too long. Equation (5) indicates that for each station area, the trips it needs to handle are traffic from coverage area of corresponding station area. Equation (6) represents the service strength for each parking space and the service strength for each station area. Equation (7) represents the probability that there is no travel demand for that station area. Equation (8) and Equation (9) represent the average queue length and average queue time for each trip in the system. The constraint functions (2)-(4) are to filter the generated chromosome sequences and eliminate the station solutions that do not meet the requirements, such as too expensive station building costs, too long distance between station and non-station areas, and too long waiting time for passengers.

4.3. **Adaptive genetic algorithms.** Genetic algorithms can produce better feasible solutions for relatively complex combinatorial optimization problems by simulating the genetic process of chromosomes. We consider each station area as a multi-server queuing system in which one parking space is one server. An appropriate number of parking spaces is needed for each station area. Consequently, we use the queuing theory of the M/M/s waiting system to allocate spaces to each station area.

4.3.1. *Coding scheme.* The coding example scheme of chromosome is shown in Figure 2. Figure 2(a) shows the representation of chromosomes, with 1 indicating a station and 0 indicating no station. Figure 2(b) presents the candidate station areas, i.e., selected areas with a high number of trips and high travel costs. We numbered it in zigzag order. Mapping chromosomes sequentially to candidate station areas yields five station areas

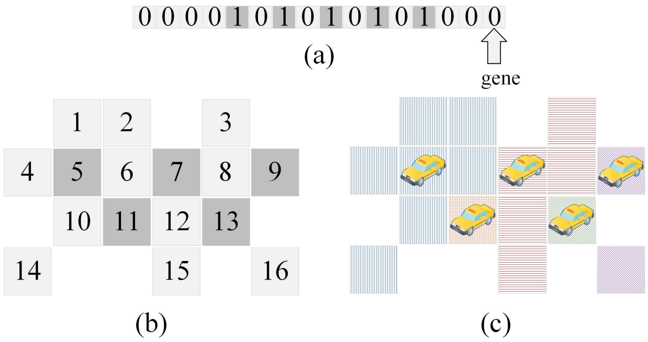


FIGURE 2. Coding example: (a) Chromosome coding representation; (b) numbered candidate station areas; (c) mapping chromosomes on geographical areas

numbered 5, 7, 9, 11, and 13. Next, the non-station area will be bound to the nearest station area. In particular, when the distances between a non-station area and several station areas are the same, the station area with the smaller number will be selected. The binding result is shown in Figure 2(c), and the areas with the same background texture belong to the same station area. Finally, according to the trips in each binding area, the number of their parking spaces is calculated separately using queuing theory. The detailed formula is in Section 4.2.

4.3.2. *Genetic operators.* The selection operation takes the ratio of individual's fitness value to the sum of fitness values of population as its selection probability, according to which the chromosome is randomly selected into a new population. Single-point crossover is adopted for crossover operation, and single-point mutation for mutation operation.

The standard genetic algorithm (SGA) uses same crossover rate (P_c) and mutation rate (P_m) for operating all individuals, which affects the efficiency of the algorithm. We found that modifying superior individuals yielded higher quality results faster. So, we propose an improved adaptive genetic algorithm (AGA) to accelerate convergence speed. We gave larger P_c and P_m to superior individuals, smaller P_c and P_m to individuals with fitness greater than and close to average fitness (f_{avg}), and maximum P_c and P_m to individuals smaller than f_{avg}, while preserving the optimal individual.

The AGA's P_c and P_m are formulated in Equations (10) and (11). f_{avg} denotes the average fitness value of population. $f_{\max}$ denotes the maximum fitness value of population. f^* denotes the higher fitness value when two individuals cross. f denotes the fitness value of a individual to mutate. k_1, k_2, k_3, k_4 take values between $[0,1]$, in this case, $k_1, k_2 = 1$, $k_3, k_4 = 0.5$.

$$P_c = \begin{cases} \dfrac{k_1(f^* - f_{avg})}{f_{\max} - f_{avg}}, & f^* \geq f_{avg} \\ k_2, & f^* \leq f_{avg} \end{cases} \tag{10}$$

$$P_m = \begin{cases} \dfrac{k_3(f - f_{avg})}{f_{\max} - f_{avg}}, & f \geq f_{avg} \\ k_4, & f \leq f_{avg} \end{cases} \tag{11}$$

4.3.3. *The algorithm process.* The process of the Wait-based GAQTSA is shown in Algorithm 1. First, initialize a given number of chromosome populations (step1), where the length of each chromosome is candidate station areas number, i.e., the number of demand areas $|D|$. The value of each gene point is randomly generated as a binary value, i.e., the value of X_j. During each population iteration, we first map each chromosome to the

Algorithm 1: The stations siting process

Input: the set of demand areas D, N_GENERATIONS, POP_SIZE.
Output: the set of station areas A, the number of spaces in each station area m_j.

```
1  Initialize the chromosome population;
2  for g = 1, 2, ..., N_GENERATIONS do
3      for pop = 1, 2, ..., POP_SIZE do
4          Mapping the chromosome to station areas;
5          for each station area do
6              Calculate dist(S0, Si); // station area S0, non-station area Si;
7              if dist(S0, Si) > MD then
8                  Regenerate this chromosome;
9              end
10         end
11         for each coverage area do
12             if W_q > T then
13                 Increase m_j;
14             end
15         end
16         Calculate the value of the objective function;
17     end
18     Obtain the fitness value from the objective function;
19     Preserve the chromosome with the highest fitness value;
20     Selection, crossover and mutation operations on populations;
21 end
22 Output the optimal site solution.
```

corresponding grid area (step4). Then we calculate the distance dist(S0, Si) between a station area and its coverage area, i.e., d_{ij}, to ensure its distance within MD; otherwise a chromosome is regenerated (step5-step10).

For each station area, the travel demand in its coverage area must be satisfied within the specified time (step11-step15). We then calculate the value of the objective function for each chromosome (step16) and perform selection, crossover, and mutation operations to generate new populations (step18-step20). Finally, the optimal site solution is output after a predetermined number of generations of iterations.

5. Case Studies.

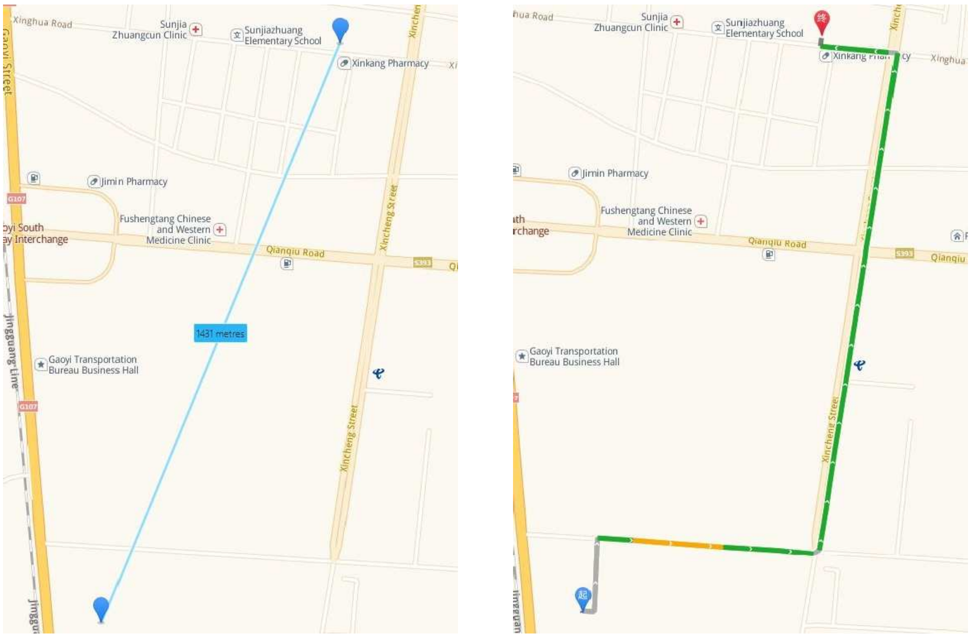
5.1. Data.

It is difficult to precisely know passenger's demands considering that most of the current studies on station selection are based on taxi trajectory data. We use passenger's trip OD data, which can accurately obtain the areas with high travel demand. The data is obtained from Shijiazhuang's all-day online taxi trip data on 1 March 2018, with 65,376 records in total. It mainly includes pick-up and drop-off locations of urban citizens' trips. The original OD data contains several fields, as shown in Table 3 including trip ID, pick-up location, drop-off location, trip start time and the trip end time.

Given the structure of urban roads, it is unlikely that people will travel in a straight line. To obtain more realistic trip distances, we called the Amap [28] API to extract the travel distance of each trip. An example of straight-line and travel distances derived using the Amap API is shown in Figure 3.

TABLE 3. OD data format

Data field	Description	Sample value
Trip ID	Anonymous trip ID	1
Pick-up location	Latitude and longitude of a departure	114.60062, 37.58946
Drop-off location	Latitude and longitude of a destination	114.60678, 37.60135
Trip start time	System time at the start of a trip	2018-03-01 10:42:09
Trip end time	System time at the end of a trip	2018-03-01 10:52:01

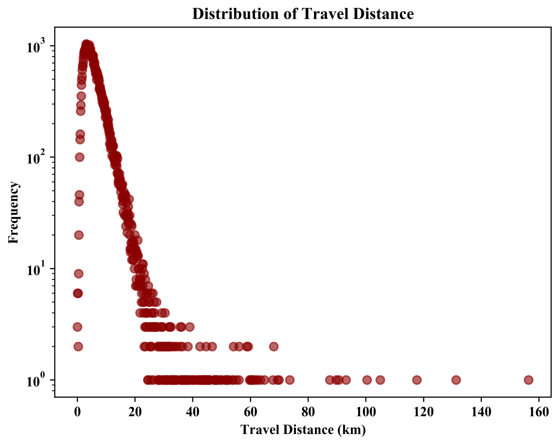


(a) (b)

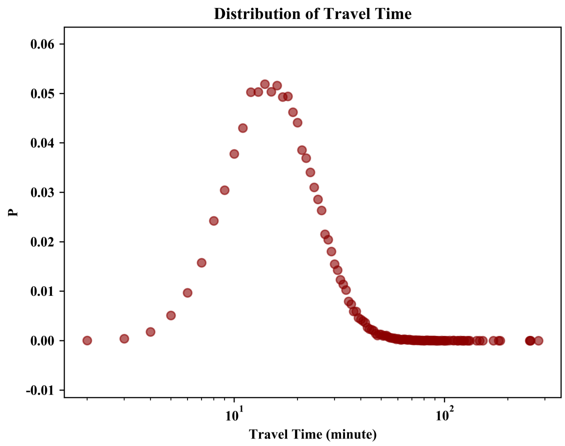
FIGURE 3. Straight-line distance and travel distance: (a) Straight-line distance, calculated directly from latitude and longitude; (b) travel distance, calculated by calling the Amap API

According to the start and end time of an OD data, we obtained the travel time. Further we calculate the travel cost of each trip according to the taxi charging rules. The distributions of travel distance, travel time and travel cost are shown in Figure 4. The frequency distribution of travel distance is shown in Figure 4(a), whose horizontal axis is travel distance and the vertical axis is the number of trips. It can be found that most of the trips are within 20 km, meaning that most people travel by taxi for short and medium range trips. The probability distribution of travel time is shown in Figure 4(b), with the horizontal axis representing travel time and the vertical axis representing the probability of travel time. It can be noticed that most of trips took 10-20 minutes to travel, which is in accordance with the travel distance of short and medium range trips. It also reflects that passengers are not inclined to take taxis to travel for excessively short or long range trips, as their travel costs may be much higher than buses and subways. The frequency distribution of the travel cost is shown in Figure 4(c). The horizontal axis indicates the cost spent on each trip and the vertical axis denotes the number of trips corresponding to that cost. Trips frequency above 100 times have travel costs between 8 yuan and 50 yuan, which is in accordance with the travel distance for short and medium range trips.

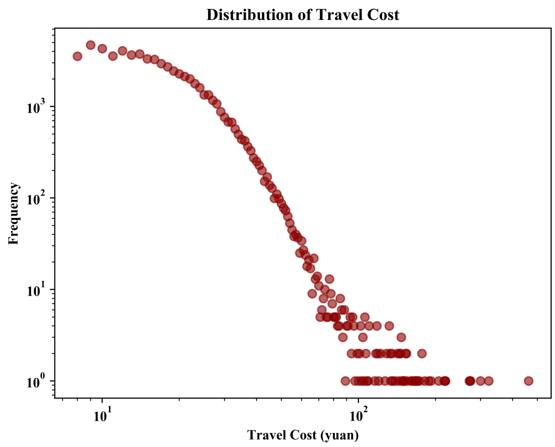
Then, we selected the central part of Shijiazhuang for our study, with its latitude and longitude in the range of 37.85-38.25°N, 114.3-114.7°E. We divided the study area into



(a)



(b)



(c)

FIGURE 4. Frequency distribution of travel distance, probability distribution of travel time, and frequency distribution of travel cost: (a) Travel distance, obtained by calling the Amap API based on the pick-up and drop-off location from the raw data; (b) travel time, obtained from the trip start time and the trip end time; (c) travel cost, based on travel distance and travel time, calculated according to taxi charging rules

1 km $*$ 1 km grids and counted the number of pick-ups, drop-offs and travel costs for each area and plotted them in the heat map, as shown in Figure 5. The heat map of the pick-up locations is shown in Figure 5(a), where the values in each grid area are their ATF values. The heat map of the drop-off locations is shown in Figure 5(b). It can be seen that the heat map distribution of the drop-off locations is more dispersed than the pick-up locations, which indicates that most passengers do not travel to a concentrated destination. The travel cost heat map is shown in Figure 5(c), with the value of each grid area as its ATFC value.

Further, the grid areas with ATF values over 300 times and ATFC values over 5000 yuan were considered as hotspot areas and were treated as candidate station areas. So 69 candidate areas were selected, as shown in Figure 6.

A detailed statistical analysis was conducted for the 69 high value candidate station areas. Figure 7(a) shows the trip flow box diagram for these areas, with the different colors representing the different areas. It can be seen that the number of trips from 0:00

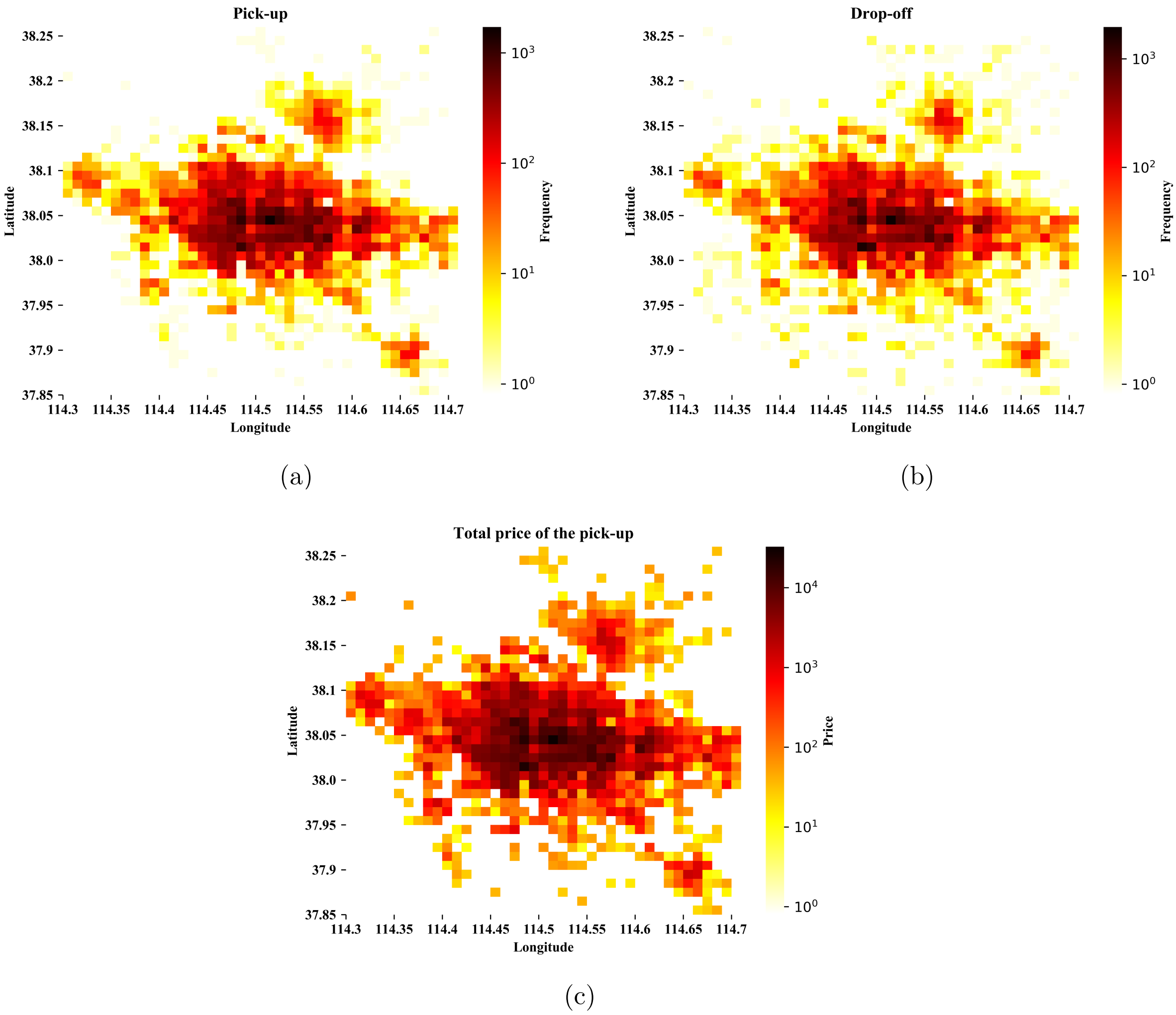


Figure 5. Heat map of the pick-up location, drop-off location and travel cost: (a) Pick-up locations, i.e., the ATF value; (b) drop-off locations; (c) travel cost, i.e., the ATFC value

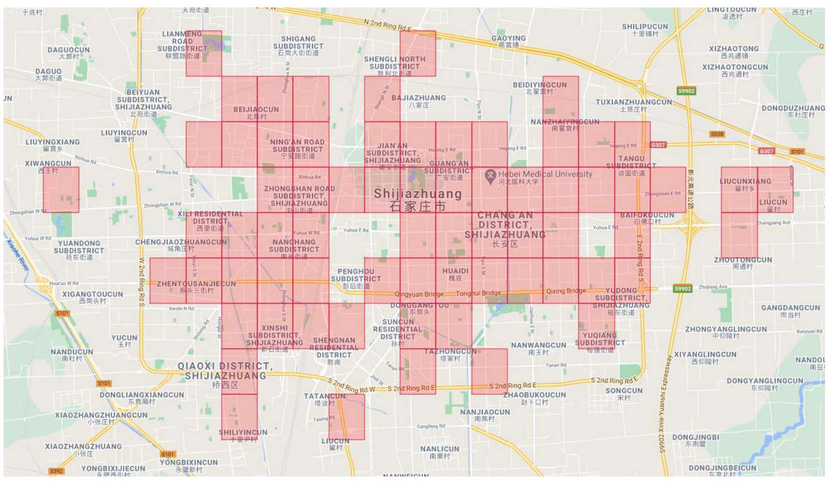
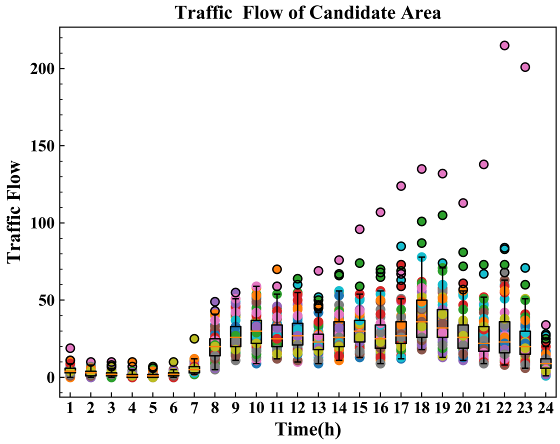
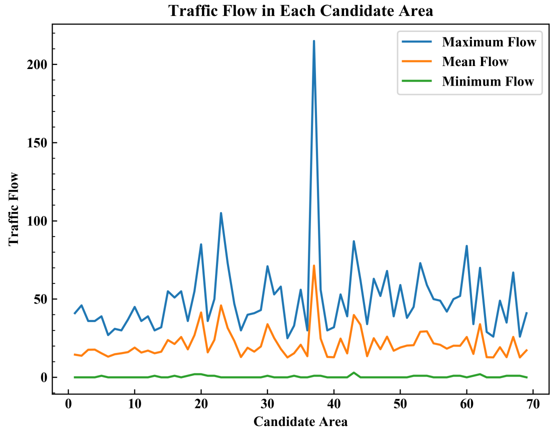


Figure 6. The 69 candidate station areas



(a)



(b)

FIGURE 7. (color online) Traffic flow and traffic flow statistical for the 69 candidate station areas: (a) The 24-hour traffic flow graph; (b) the maximum, minimum and average values of traffic flow in a day

to 7:00 is not high for all areas as this period is sleep hours. Furthermore, the number of trips in each area is not constant throughout the day; for example, some areas may have a high number of trips at a particular period and a low number of trips at other times. We also measured the maximum, minimum and average number of trips per hour of the day for each area, as shown in Figure 7(b). The top line represents maximum value, the bottom line represents minimum value and the middle line represents average value. It can be found that for most areas the minimum number of trips is 0 and the maximum is no more than 100, but there are a few areas where the peak is above 200. For convenience and reasonability of calculation, average value is obtained by dividing the number of trips per hour for each area, namely the AATF value.

5.2. **Experiments.** After the candidate station area has been identified, we use the AATF value of each area as its travel demand value. The objective function is subsequently solved using AGA and queuing theory following Algorithm 1.

The parameters in the experiment were set as follows: N_GENERATIONS = 300, POP_SIZE = 600, $|D| = 69$, $F_1 = 50000$, $F_2 = 5000$, $F_3 = 1.87$, $F = 10000000$. We have also designated the areas with the highest travel traffic as compulsory station areas to ensure the availability of taxi station services. We used T and MD as variables in multiple sets of contrast experiments, the results of which are shown in Table 4.

It can be seen that as the MD increases, total number of stations gradually decreases, the cost of stations gradually decreases, and the objective function value gradually increases. This is because when MD increases, the distance between the station area and the non-station areas in its coverage area becomes larger, resulting in fewer stations created. Let the MD be constant, as waiting time T gradually increases, the total number of spaces gradually decreases and the objective function value gradually increases. When T keeps increasing, the number of spaces does not continue to decrease, this is because there are no more passengers to serve.

Then we visualize the station construction results at $T = 10$ minutes, as shown in Figure 8. It can be noticed that the stations are scattered over the city. When the MD increases, the number of station areas decreases. We found that the highest travel traffic and the highest economic value area is located right in city center. In addition, when

TABLE 4. Experimental results based on different MD and T

No.	MD (km)	T (minutes)	Objective function value (yuan)	Cost of stations (yuan)	Total number of stations	Total number of spaces
1	4	1	−255149	905000	8	101
2	4	3	−165535	815000	8	83
3	4	5	−155615	805000	8	81
4	4	10	−140294	790000	8	78
5	5	1	−116209	765000	6	93
6	5	3	−51786	700000	6	80
7	5	5	−46455	695000	6	79
8	5	10	−36531	685000	6	77
9	6	1	12339	635000	4	87
10	6	3	56851	590000	4	78
11	6	5	62551	585000	4	77
12	6	10	67655	580000	4	76
13	7	1	12906	635000	4	87
14	7	3	58024	590000	4	78
15	7	5	63062	585000	4	77
16	7	10	67968	580000	4	76
17	8	1	77384	570000	3	84
18	8	3	111904	535000	3	77
19	8	5	117315	530000	3	76
20	8	10	122124	525000	3	75

(a) (b) (c) (d)

FIGURE 8. Visualization of station build results with $T = 10$ minutes: (a) $MD = 4$ km; (b) $MD = 5$ km; (c) $MD = 6$ km or $MD = 7$ km; (d) $MD = 8$ km

the coverage distance of stations is limited, the stations are uniformly distributed. As the coverage distance becomes wider, stations are almost located inside the 2nd Ring Road. Specifically, when $MD = 6$ km or $MD = 7$ km, there is little difference in the results of the station construction.

We use the loss-based genetic algorithm and queueing theory siting algorithm (Loss-based GAQTSA) as a comparison of the Wait-based GAQTSA. In this case, Loss-based GAQTSA only considers the trips currently handleable and does not allow passengers to wait. The comparison results are shown in Table 5, where r denotes the loss rate and the values in the table are the objective function values. The visualization results of Table 5 are shown in Figure 9. It can be seen that as the loss rate r increases, the objective function value of the loss-based model gradually increases. This is caused by the fact that fewer trips need to be served, which leads to a reduction in the total number of spaces and the cost of stations. Compared with the loss-based model, the average performance of the wait-based model is better. When the values of $MD = 4, 5, 6, 7$, and 8 km, the average value of the wait-based model is higher than the loss-based model by 131775.5, 130252.25, 113967.75, 108277.75, and 93991.75, respectively. The objective function value of the loss-based model with $r = 0.15$ and $MD = 4$ km is higher than the wait-based model with $T = 1$ min, $MD = 4$ km. However, when T gradually increases, the objective function value of the wait-based model is larger. This shows that the Wait-based GAQTSA can serve more passengers by adding less waiting time.

Table 5. Comparison results of Loss-based GAQTSA and Wait-based GAQTSA

MD (km)	Loss-based GAQTSA				
	r				
	0.01	0.05	0.1	0.15	avg
4	−411647	−318198	−290661	−223189	−310923.75
5	−267688	−188591	−162144	−153567	−192997.5
6	−124029	−60022	−37583	−34841	−64118.75
7	−114344	−54587	−32944	−29276	−57787.75
8	−31041	18565	31446	33790	13190.0
MD (km)	Wait-based GAQTSA				
	T (minutes)				
	1	3	5	10	avg
4	−255149	−165535	−155615	−140294	−179148.25
5	−116209	−51786	−46455	−36531	−62745.25
6	12339	56851	62551	67655	49849.0
7	12906	58024	63062	67968	50490.0
8	77384	111904	117315	122124	107181.75

Based on the Wait-based GAQTSA, the comparison results of SGA and AGA are shown in Figure 10, where P_m of all SGA is 0.2. The objective function optimization process of the two algorithms is shown in Figure 10(a), where AGA converges around 150 generations and the fastest SGA with $P_c = 0.3$ converges around 200 generations. All algorithms result in 3 stations and 75 parking spaces. The difference lies in the dispatch cost, i.e., the third part in the objective function. In the early stage of population, there are a lot of station areas in the chromosome with lower dispatch cost. After a period of optimization, the number of stations is gradually settled, and it is time to adjust the location of station area to get a lower dispatch cost. The AGA algorithm and the SGA algorithm with P_c of 0.6 and 0.3 have the smallest dispatch costs. The dispatch cost

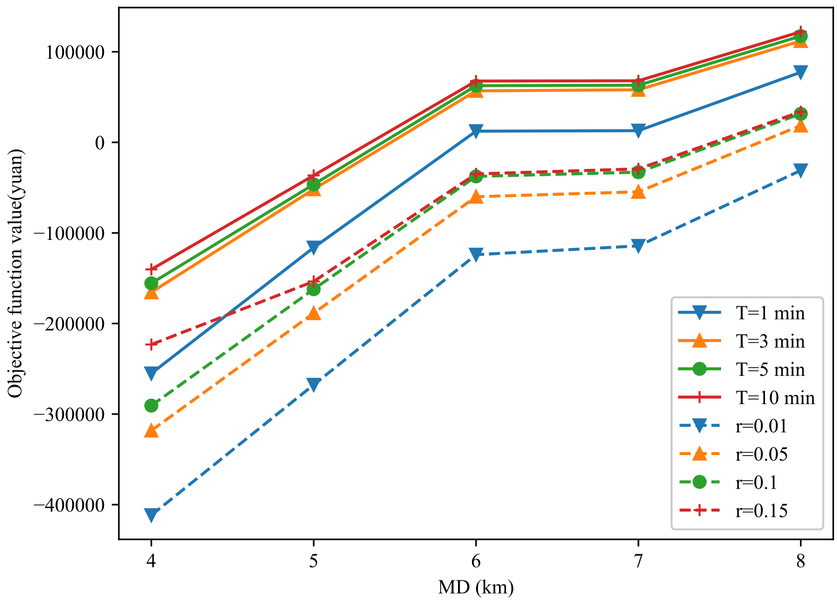
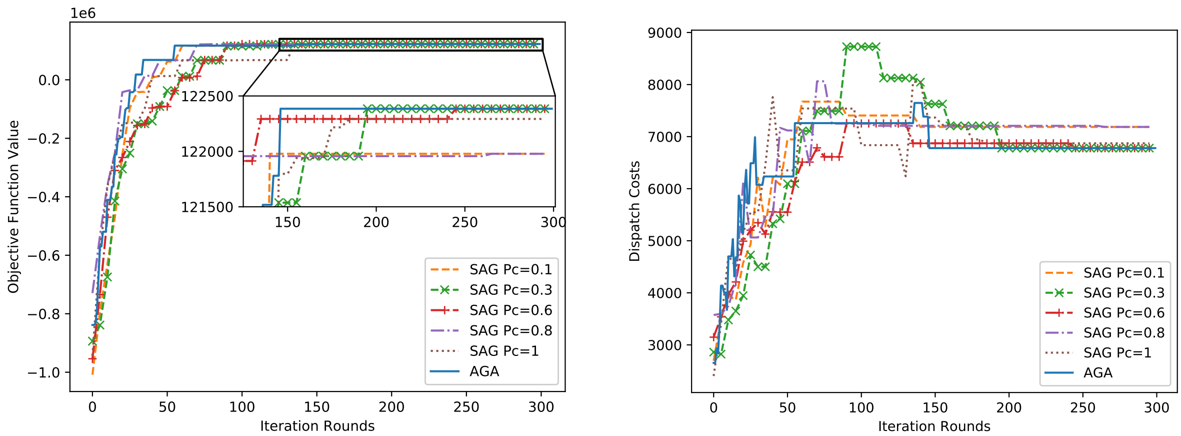


FIGURE 9. Comparison of the objective function values of the Wait-based GAQTSA and the Loss-based GAQTSA, where T represents the waiting time of the waiting system model and r denotes the loss rate of the losing system model



(a) (b)

FIGURE 10. Comparison of SGA and AGA iteration process, where P_c denotes the crossover rate: (a) The objective function value iteration process. The subfigure magnifies the process from 125 to 300 rounds; (b) the dispatch costs evolution process

optimization process is shown in Figure 10(b). The dispatch cost of the AGA algorithm is able to converge quickly to a small value, while the SGA algorithm converges slower. It indicates the improved AGA is able to get optimal solution faster without specifying the crossover rate and mutation rate.

6. Conclusion.

To reduce carbon emissions in the city caused by stochastic movement of taxis, it is necessary to site urban taxi stations. This paper proposes a model for siting taxi stations in a city. This model provides the optimal location of taxi stations in urban hotspots to serve the travel demands of residents. We construct objective functions with constraints, which are solved using genetic algorithms and queuing theory by maximizing the profits of station construction. The model was validated on Shijiazhuang dataset. We have analyzed the relevant experimental results and demonstrated that our model is feasible and effective. We also compare the Wait-based GAQTSA and the Loss-based GAQTSA and the results show that the Wait-based GAQTSA has a better average performance and can serve more passengers in the allowed time. We further proposed an improved adaptive genetic algorithm (AGA) to accelerate the evolution of populations without fixing the crossover rate and mutation rate. Therefore, this study can provide some reference for taxi station planning and government transport decisions.

Based on the generated station results, we recommend the following practical suggestions for site implementation. Firstly, construct stations in areas with high travel traffic and high economic value, so that the travel demand in these areas can be satisfied in time with higher benefits. In addition, when the station budget is sufficient, a uniform distribution can be applied to construct stations in the urban hotspots, then allocating parking spaces according to the demand of each area. When the station budget is scarce, it is preferred to install stations in central urban areas, keeping the station area relatively centrally distributed while considering the coverage distance of the station area. This can effectively reduce taxi dispatching distance considering the high travel traffic in city center area.

This paper considers taxi stations allocation from an economic value perspective, which can achieve high economic payback while meeting the travel demands of passengers in each area. However, it is noted that this model assumes that the average travel demand for each area is the travel demand for each period, ignoring the effect of the regularity of people's travel on this issue. Hence, it is a significant part of our work in the future to integrate taxi stations with the travel patterns of residents.

Acknowledgment. This work is partially supported by the Natural Science Foundation of Hebei Province (No. F2018205102), partially by the Key Project of Hebei Education Department (No. ZD2021063), partially by the Hebei Normal University 2021 Postgraduate Innovation Funding Project (No. CXZZSS2021072 and No. CXZZSS2021071). The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] T. Lyu, P. S. Wang, Y. Gao and Y. Wang, Research on the big data of traditional taxi and online car-hailing: A systematic review, *Journal of Traffic and Transportation Engineering (English Edition)*, 2021.

[2] W. Zhang and S. V. Ukkusuri, Share-a-Cab: Scalable clustering taxi group ride stand from huge geolocation data, *IEEE Access*, vol.9, pp.9771-9776, 2021.

[3] Z. Liu, H. Chen, Y. Li and Q. Zhang, Taxi demand prediction based on a combination forecasting model in hotspots, *Journal of Advanced Transportation*, 2020.

[4] B. Hu, X. Xia, X. Shen and X. Dong, Analyses of the imbalance of urban taxis' high-quality customers based on Didi trajectory data, *Journal of Advanced Transportation*, 2019.

[5] A. P. Riascos and J. L. Mateos, Networks and long-range mobility in cities: A study of more than one billion taxi trips in New York City, *Scientific Reports*, vol.10, no.1, pp.1-14, 2020.

[6] B. Shen, X. Xu, J. Li, A. Plaza and Q. Huang, Unfolding spatial-temporal patterns of taxi trip based on an improved network kernel density estimation, *ISPRS International Journal of Geo-Information*, vol.9, no.11, p.683, 2020.

[7] Z. Qu, X. Wang, X. Song, Z. Pan and H. Li, Location optimization for urban taxi stands based on taxi GPS trajectory big data, *IEEE Access*, vol.7, pp.62273-62283, 2019.

[8] N. Wang, G. Guo, B. Wang and C. Wang, Traffic clustering algorithm of urban data brain based on a hybrid-augmented architecture of quantum annealing and brain-inspired cognitive computing, *Tsinghua Science and Technology*, vol.25, no.6, pp.813-825, 2020.
[9] X. Wang, Z. Qu, X. Song, H. Li and Z. Pan, Using taxi GPS trajectory data to optimize the spatial layout of urban taxi stands, *Transportation Research Record*, vol.2675, no.3, pp.301-312, 2021.
[10] J. Asamer, M. Reinthaler, M. Ruthmair, M. Straub and J. Puchinger, Optimizing charging station locations for urban taxi providers, *Transportation Research Part A: Policy and Practice*, vol.85, pp.233-246, 2016.
[11] 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, *Transportation Research Part C: Emerging Technologies*, vol.65, pp.172-189, 2016.
[12] J. Yang, J. Dong and L. Hu, A data-driven optimization-based approach for siting and sizing of electric taxi charging stations, *Transportation Research Part C: Emerging Technologies*, vol.77, pp.462-477, 2017.
[13] N. Ghosh, Ghorui, S. P. Mondal, S. Kumari, B. K. Mondal, A. Das and M. S. Gupta, Application of hexagonal fuzzy MCDM methodology for site selection of electric vehicle charging station, *Mathematics*, vol.9, no.4, p.393, 2021.
[14] Ö. Kaya, K. D. Alemdar and M. Y. Çodur, A novel two stage approach for electric taxis charging station site selection, *Sustainable Cities and Society*, vol.62, 2020.
[15] M. Lin, C. Huang and Z. Xu, MULTIMOORA based MCDM model for site selection of car sharing station under picture fuzzy environment, *Sustainable Cities and Society*, vol.53, 2020.
[16] H. Wang, D. Zhao, Y. Cai, Q. Meng and G. P. Ong, Taxi trajectory data based fast-charging facility planning for urban electric taxi systems, *Applied Energy*, vol.286, 2021.
[17] S. Zhang, H. Wang, Y. F. Zhang and Y. Z. Li, A novel two-stage location model of charging station considering dynamic distribution of electric taxis, *Sustainable Cities and Society*, vol.51, 2019.
[18] X. Meng, W. Zhang, Y. Bao, Y. Yan, R. Yuan, Z. Chen and J. Li, Sequential construction planning of electric taxi charging stations considering the development of charging demand, *Journal of Cleaner Production*, vol.259, 2020.
[19] X. Zhang, L. Peng, Y. Cao, S. Liu, H. Zhou and K. Huang, Towards holistic charging management for urban electric taxi via a hybrid deployment of battery charging and swap stations, *Renewable Energy*, vol.155, pp.703-716, 2020.
[20] P. Zhao, X. Liu, M. P. Kwan and W. Shi, Unveiling cabdrivers' dining behavior patterns for site selection of 'taxi canteen' using taxi trajectory data, *Transportmetrica A: Transport Science*, vol.16, no.1, pp.137-160, 2020.
[21] M. Wei, B. Sun and R. Sun, Expected value model of bus gas station site layout problem with fuzzy demand in supplementary fuel using genetic algorithm, *Cluster Computing*, vol.22, no.1, pp.809-818, 2019.
[22] W. Nega, Y. Hunie, M. Tenaw, T. Dires, S. Kassaw and R. Mansberger, Demand-driven suitable sites for public toilets: A case study for GIS-based site selection in Debre Markos Town, Ethiopia, *GeoJournal*, pp.1-14, 2021.
[23] W. Yu, M. Guan and Y. Chen, Fire stations siting with multiple objectives and geospatial big data, *Earth Science Informatics*, vol.14, no.1, pp.141-160, 2021.
[24] Y. Wang, S. Li, X. Zhang, D. Jiang, M. Hao and R. Zhou, Site selection of digital signage in Beijing: A combination of machine learning and an empirical approach, *ISPRS International Journal of Geo-Information*, vol.9, no.4, p.217, 2020.
[25] C. J. P. Alves, E. J. da Silva, C. Müller, G. M. R. Borille, M. X. Guterres, E. M. Arraut and R. J. dos Santos, Towards an objective decision-making framework for regional airport site selection, *Journal of Air Transport Management*, vol.89, 2020.
[26] Meiliana, L. C. Dewi and A. Chandra, Optimization of test case generation from UML activity diagram and sequence diagram by using genetic algorithm, *ICIC Express Letters*, vol.13, no.7, pp.585-591, 2019.
[27] M. Zhou, Y. Long, W. Zhang, Q. Pu, Y. Wang, W. Nie and W. He, Adaptive genetic algorithm-aided neural network with channel state information tensor decomposition for indoor localization, *IEEE Transactions on Evolutionary Computation*, 2021.
[28] *Amap*, https://lbs.amap.com/, Accessed on 30 June 2021.

Author Biography

Duo Tian received the B.Eng. degree in the College of Information Technology from Hebei Normal University, China in 2019.

Ms. Tian is currently a postgraduate student at the College of Computer and Cyber Security, Hebei Normal University, China. She is hosting a Hebei Normal University 2021 Postgraduate Innovation Grant Project. Her main research interests include the computational social science, genetic algorithms and graph neural networks.

Jialing Lu received the B.Eng. degree in the College of Information Technology from Hebei Normal University, China in 2019.

Ms. Lu is currently a postgraduate student at the College of Computer and Cyber Security, Hebei Normal University, China. She is hosting a Hebei Normal University 2021 Postgraduate Innovation Funding Project. Her main research interests include the computational social science, and complex networks.

Zhicheng Wei obtained his Ph.D. degree in Electronic and Information Engineering College from Tianjin University, China in 2007; he worked at the Department of Electrical&Computer Engineering, Ryerson Univeristy of Canada as the postdoctoral from May 2009 to May 2010.

Prof. Wei is currently a full-time professor at the College of Computer and Cyber Security, Hebei Normal University, China. He has published over 30 papers in journals and conferences. His main research interests include the computational social science, complex networks and graph networks.