

Research Article

Exploring Urban Taxi Drivers' Activity Distribution Based on GPS Data

Xiaowei Hu, Shi An, and Jian Wang

School of Transportation Science and Engineering, Harbin Institute of Technology, Harbin 150090, China

Correspondence should be addressed to Xiaowei Hu; xiaowei_hu@hit.edu.cn

Received 10 April 2014; Revised 9 June 2014; Accepted 19 June 2014; Published 13 July 2014

Academic Editor: X. Zhang

Copyright © 2014 Xiaowei Hu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the rapid development of information communication technology and data mining technology, we can obtain taxi vehicle's real time operation status through the large-scale taxi GPS trajectories data and explore the drivers' activity distribution characteristics. Based on the 204 continuous hours of 3198 taxi vehicles' operation data of Shenzhen, China, this paper analyzed the urban taxi driver's activity distribution characteristics from different temporal and spatial levels. In the time level, we identified the difference with taxi daily operation pattern (weekday versus weekends), continuous time in one day, passengers in vehicle time, and taxi drivers' operation frequency; in the space level, we explored the taxi driver's searching pattern, including searching activity space distribution and the relationship between the pick-up locations and the drop-off locations. This research can be helpful for urban taxi drivers' operation and behavior pattern identification, as well as the contribution to the geographical activity space analysis.

1. Introduction

As an important component of the urban public passenger transportation system, taxi offers an all-weather, convenient, comfortable, and personalized travel services for the urban residents, as well as playing a key role in the urban passenger transportation development [1]. Taking the capital of China, Beijing, as an example, there were 66,600 registered taxi vehicles in the year 2009, and these taxis carried 680 million passengers annually (which means more than 2 million passengers per day), accounting for 9.4 percent of the total passenger volume of urban public transport (including the conventional bus, metro transit, and taxi service) [2].

Due to the 24 hours operation per day, taxis account for a higher proportion in the urban traffic volume and air emission pollution [3]; more researchers attempt to analyze taxi operation status and taxi driver's routing choice based on the GPS traces data, aiming to reduce taxi driver's vacant travel distance and cruising time, and also this can save energy resource. In the taxi operation status analysis aspect, researchers analyze the taxi daily average operation distance, taxi passenger demand distribution, daily average operation frequency, and different driver's operation patterns; the reader can be referred to Hu and Feng [4], Chen [5], Jiang

et al. [6], Wen et al. [7], Liu et al. [8], Hu et al. [9], and Zhang and He [10].

In the above studies, researchers have paid more attention to the basic taxi operation status and routing choice analysis, but they paid less attention to the taxi driver's temporal and spatial activity space analysis. Meanwhile, each taxi driver's activity space is quite different from the other drivers, and the analysis of taxi driver's travel activity space is a more interesting research field to explore. When the taxi is with a passenger, the driver will complete the trip in accordance with the passenger's travel will by adopting the shortest path from the origin to the destination [11]; this destination set can be called the taxi driver's drop-off locations; when the taxi is without passenger, the driver always wants to search the next potential passenger during the shortest possible time near or around his last drop-off passenger location [10, 12]; this set can be regarded as the taxi driver's pick-up locations. Although the taxi driver will not pick up his passengers at the same location each day, each taxi driver's pick-up passenger's location and drop-off passenger's location are close to each other; the relationship between taxi driver's pick-up locations and drop-off locations can be explored to describe the spatial distribution of taxi drivers.

Due to the randomness and lack of regularity of taxi operation characteristics, it is improper to adopt visit places' order or arrive at a fixed time to describe taxi driver's travel activity pattern. So how to describe the regularity of taxi driver's operation spaces and time period, is there any difference between weekends versus weekdays? In the urban road network, when a taxi completes a trip to arrive at the passenger's destination, then the taxi driver would be at the without passenger stage and search for the next passenger within a range of regions/areas. So taxi drivers may search and cruise around some region to find their next passenger; there may be a relationship between taxi driver's covering area and operation frequency. We assume that there is relationship between the two activity spaces, drop-off locations and pick-up locations. Also, each taxi driver's activity space dynamics can be explored as time evolves.

As a geographic concept, the activity space means "a composite of the locations where an individual conducts routine activities" [13–15]. A major application of activity theory is empirical measurement and analysis of space time activity (STA) data or records of where and when individuals conducted activities over a daily, weekly, or monthly cycle. Wang and Cheng [15] had summarized the basic components in the activity theory, which includes the activity, activity frequency, activity destination, trip, transport mode(s), and the activity space.

The special work characteristics of the taxi drivers, as a special group, make their daily activity space quite different from the commute workers or students' daily mobility variability, such as the regular arrival at the workplace at a relatively fixed time [16], visiting some shop or restaurant frequently [17], and visiting some places following a certain order [18]. Some researchers focus on taxi driver's cognitive space or travel time difference with the general public, which can reflect taxi driver's spatial-temporal cognitive of the urban space. The pioneer work can be traced to Giraud and Peruch [19], Peruch et al. [20], Spiers and Maguire [21], and Wakabayashi et al. [22]. Besides the traditional comparison of the cognitive maps method, Wakabayashi et al. [22] also adopted the SDE (standard deviational ellipse) to analyze the difference between the taxi drivers and the university students.

The aim of this paper is to provide measurements to analyze the taxi driver's activity space from the time level and individual level; in the time level, the taxi driver's daily (workdays versus weekends) pick-up locations and drop-off locations will be measured and analyzed, as well as the exploration of relationship between the two activity spaces; in the individual level, each taxi driver's activity space dynamics can be explored as time evolves. We will introduce the geographic concept into this research and attempt to analyze the taxi driver's activity space, which will reveal the relationship between the taxi driver's pick-up locations and drop-off locations and help drivers reduce cruising time and improve operation frequency. This research can also be helpful for saving taxi energy consumption and lowering air pollution emissions, to achieve a more sustainable and environmental development of urban taxi industry.

The paper is organized as follows. Section 2 provides an overview on taxi driver's behavior analysis and activity space measurements comparison. In Section 3, we present the taxi GPS traces data source and taxi driver's activity space measures in detail. Section 4 describes and compares the analysis results. Section 5 is the discussions. Finally, we conclude this paper in Section 6.

2. Literature Review

Around the world, research on taxi drivers' travel behavior analysis has been paid more attention from Pailhous [23] and Michon [24]. There are many influence factors on the taxi driver's travel choice behavior in the actual road network environment, such as the driver's road network familiarity, traffic control devices, road construction, and other urban traffic management and control measures. With the development of ICT (information communication technology), it provides a new way to explore taxi driver's route choice or way-finding behavior, which involves the shortest path optimal and driving destination or direction selection based on the huge number of taxi trajectories. Based on the taxi GPS traces and data mining technology, especially one can obtain the experienced/smart taxi driver's route choice behavior in real time and provide guides for other general public's shortest path optimal choice [11].

Another research field focuses on the taxi driver's practical travel behavior. Murakami and Wagner [25] adopted PDA (personal digital assistant) with a GPS receiver to obtain taxi vehicle's speed and daily travel information, which can be seen as the first trial in analysis of taxi driver's travel behavior by using GPS traces data, and this method can have higher data quality on travel start and end times, total trip time, and destination locations than daily travel diary and telephone retrieval methods. Liu et al. [8] described the searching behavior difference between the top taxi drivers and the ordinary divers in Shenzhen and found that the top drivers pay more attention on the less competitive and more profitable locations, rather than the central business district. Based on the taxi GPS trajectory data in Wuhan, Yue et al. [26] forecasted the taxi driver's pick-up locations on the basis of the selected drop-off locations. These researches also can help us to understand taxi driver's travel behavior, but they have paid less attention to the taxi driver's searching behavior dynamic pattern in their pick-up and drop-off activity space.

The activity theory has been adopted to analyze different kinds of traveler's spatial mobility dynamics. The special work characteristics of taxi drivers, as a special group, make their daily activity space quite different from the commute workers or students' daily mobility variability. Due to their professional relationship, the taxi drivers need to arrive at the destination according to the passenger's travel will or search the next potential passenger according to their individual characteristics (the driving experience and the familiarity of the road network or place); there is not a fixed or constant mobility route for the taxi driver to follow. Based on Wang and Cheng's [15] summary, we can identify taxi driver's basic components in the activity theory, which is shown in Table 1.

TABLE 1: Taxi driver's basic components of activity theory.

Component	Classification
Activity	Depending on the purpose, taxi driver's activity can be classified into pick-up passenger activity and drop-off passenger activity
Activity frequency	The number of times the pick-up passenger activity or drop-off passenger activity occurs during a given time period, in an hour or a day
Activity destination	The location where a pick-up passenger activity or a drop-off passenger occurs
Trip	Movement between the pick-up passenger activity and drop-off passenger activity destinations
Transport mode (s)	By taking taxi vehicle
Activity space	A composite of the locations where a taxi driver conducts pick-up passenger activity and drop-off passenger activity, so it can be divided into pick-up activity space (location) and drop-off activity space (location)

From Table 1, we can know that some basic components of taxi driver's activity have been determined, such as the activity, trip, and transport mode(s). So we pay more attention to the taxi driver's activity frequency, activity destination, and activity space.

For the taxi driver's activity frequency, we can adopt a taxi's daily number of operation service times to describe this component; this also can be divided into the pick-up passenger activity and the drop-off passenger activity. Due to the different passenger's destination demand, the taxi driver's drop-off locations have the irregularity characteristics and will be in accordance with the passenger's requirement; the taxi driver will search the next passenger near the last drop-off location, so the pick-up locations will have some relationship with the drop-off locations. The question will be focused on the pick-up locations and the drop-off location, which is also the taxi driver's activity space analysis.

In the research of Ge et al. [27], they had clustered the taxi drivers' pick-up locations during a certain time interval to obtain the centroids (mean centre) of these pick-up locations, which can be served as the recommended pick-up locations with a certain probability. This leverage can help a new taxi driver to search the potential passenger in a certain area; meanwhile it provides a reference for us to explore the dynamics in the activity spaces of taxi drivers over time and the relationship between each taxi driver's daily pick-up locations' area and drop-off locations' area. Susilo and Kitamura [16] had extended the action space to the second moment of the activity locations that it contains, and then they examined the day-to-day variation in the second moment. The second moment of activity locations has given us inspirer how to analyze each taxi driver's day-to-day variation in the pick-up and drop-off activity space (location).

Here we hypothesize that each taxi driver's activity space mean centre (centroid) has the relationship with the centroid of the whole taxi driver's space area, just as Susilo and Kitamura [16] analyzed the worker's daily activity locations relationship. We also analyze taxi driver's day-to-day variation on activity space and statistically analyze the variation of the second moments.

3. Taxi Driver's Activity Space Measures

There are various measurements to measure the activity space in the previous researches, such as the SDE (standard deviational ellipse), mean centre, the x/y axis ratio, SDC (standard deviational circle), and kernel density. Based on existing researches, we divided these measurements into two categories, the spatial distribution category and the extended second moments of activity locations measurement category.

3.1. The Spatial Distribution Category. The spatial distribution estimates the basic parameters about the distribution; they include mean centre, standard deviation of the X and Y coordinates, standard deviational ellipse, standard distance deviation, and convex hull.

The mean centre (MC) is the average location of taxi service events in the space (including the pick-up and drop-off event), which can be calculated by [28]

$$(\dot{x}_{mc}, \dot{y}_{mc}) = \left[\frac{\sum_{j=1}^n x_j}{n}, \frac{\sum_{j=1}^n y_j}{n} \right], \quad (1)$$

where $\dot{x}_{mc}, \dot{y}_{mc}$ are the coordinates of the mean centre, which can determine the space location of the MC, x_j, y_j are the coordinates of taxi service event j in the two-dimensional, respectively, and n is the total number of taxi service events.

Standard distance deviation (SDD) [28, 29] can describe the absolute dispersion degree for each taxi service event relative to the mean centre (MC); the formula can be expressed by

$$SDD = \sqrt{\frac{\sum (x_j - \dot{x}_{mc})^2 + \sum (y_j - \dot{y}_{mc})^2}{n}}, \quad (2)$$

where x_j, y_j are the coordinates of taxi service event j and $\dot{x}_{mc}, \dot{y}_{mc}$ are the coordinates of the mean centre.

Based on the mean centre and standard distance deviation, we can draw the standard deviational circle (SDC), which can express the dispersion of taxi service in all directions of space.

standard deviational ellipse (SDE) [28, 29] can determine the directional factors of the spatial distribution and find the main direction of the taxi service event in space. The calculation of SDE can be expressed by θ (the SDE y -axis in the clockwise rotation angle), where e_1 is the major axis of SDE and e_2 the minor axis of SDE. The detailed formulas are as follows:

$$\theta = \arctan \left(\left(\left[\sum (x_j - \dot{x}_{mc})^2 - \sum (y_j - \dot{y}_{mc})^2 \right] + \left(\left[\sum (x_j - \dot{x}_{mc})^2 - \sum (y_j - \dot{y}_{mc})^2 \right]^2 + 4 \left[\sum (x_j - \dot{x}_{mc})(y_j - \dot{y}_{mc}) \right]^2 \right)^{1/2} \right) \times (2 \sum (x_j - \dot{x}_{mc})(y_j - \dot{y}_{mc}))^{-1} \right),$$

$$e_1 = 2S_x = \sqrt{\frac{\sum [(x_j - \dot{x}_{mc}) \cos \theta - (y_j - \dot{y}_{mc}) \sin \theta]^2}{n-2}},$$

$$e_2 = 2S_y = \sqrt{\frac{\sum [(x_j - \dot{x}_{mc}) \sin \theta - (y_j - \dot{y}_{mc}) \cos \theta]^2}{n-2}},$$

$$A = \pi S_x S_y, \quad (3)$$

where A is area of SDE, S_x is the semimajor axis of SDE, S_y is the semiminor axis of SDE, \dot{x}_{mc} , \dot{y}_{mc} and x_j , y_j are consistent with formulas (1) and (2).

Figure 1 has shown a taxi operation example in Shenzhen from 0 am on April 18, 2011 to 12 am on April 26, 2011 (which consists of 204 continuous hours) in Shenzhen, China. The red points represent the taxi drop-off activity locations, while the cyan points represent the taxi pick-up activity locations. From Figure 1, we can find that the taxi's operation drop-off locations and pick-up locations' mean centre (centroid), standard deviational ellipse distribution, and the two activity spaces' standard deviation of the X and Y coordinates are quite different, and the y -axis is rotated clockwise through an angle. But the mean centres of the drop-off locations and pick-up locations are quite near to each other; we may use the extended second moments of activity locations measurement category to analyze this relationship and distribution.

From Figure 2, we can know that the taxi driver's daily operation has no fixed origin; this is quite different from other workers or students' origin-home location; meanwhile, there is no fixed operation lines or order for taxi driver to follow. But each taxi driver's operation has a mean centre that can be as a fixed location; based on this we can extend to measure taxi driver's daily operation relationship.

Kamruzzaman and Hine [30] analyzed the correlations between the different measures of activity space size; this method can help us to understand different activity space

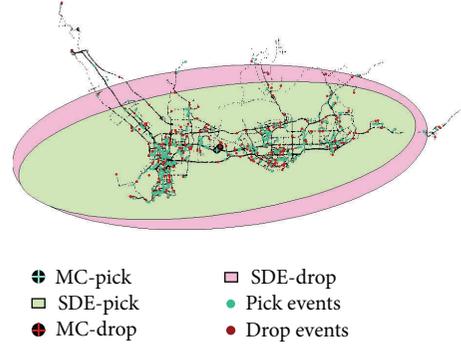


FIGURE 1: Drop-off locations, pick-up locations, SDE, and mean centre figure of one taxi vehicle operation.

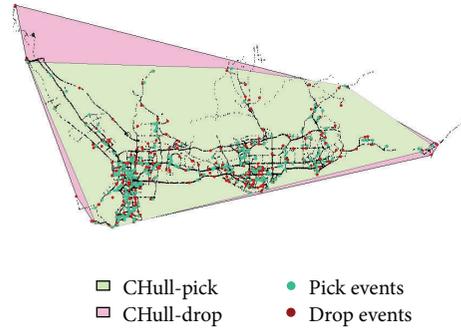


FIGURE 2: The convex hull figure of one taxi vehicle operation.

measures relationship of each taxi driver's operation. We will analyze these relationships:

- (1) area size of SDE and area size of SDD of pick-up locations,
- (2) area size of SDE and area size of SDD of drop-off locations,
- (3) area size of SDE comparison between the pick-up and drop-off locations,
- (4) area size of SDD comparison between the pick-up and drop-off locations.

3.2. The Extended Second Moments of Activity Locations Measurement Category. The researches of Susilo and Kitamura [16] and Ge et al. [27] have given us inspirer how to analyze each taxi driver's day-to-day variation in the pick-up and drop-off activity space. Each taxi driver's daily activity space area mean centre may have the relationship with the centroid of the whole taxi drivers' activity space, just as Susilo and Kitamura [16] analyzed the worker's daily activity locations relationship. We can analyze taxi driver's day-to-day variation on activity space and statistically analyze the second moments of activity locations:

- (1) the Great-circle distance between the mean centre of each taxi driver's drop-off location and pick-up location, with calculating the statistics distribution,

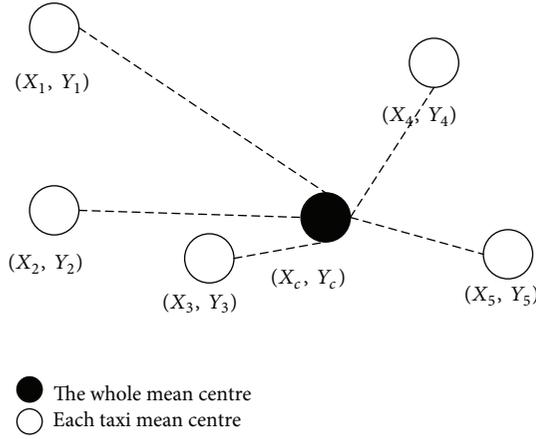


FIGURE 3: An illustration of the locations mean centre of each taxi driver and the all taxi drivers.

which can be referred to in Figure 1 of the MC pick and MC drop (by using (1)),

- (2) the Great-circle distance between the drop-off (pick-up) locations mean centre of each taxi driver and all taxi drivers, with calculating the statistics distribution, which is shown in Figure 3,
- (3) the fullness of activity spaces to describe the individual SDE feature class.

The next part is how to calculate the Great-circle distance between two points on the globe surface. Here we adopt the simple spherical law of cosines formula to calculate it [31, 32], assuming the Earth as a spherical earth (ignoring the ellipsoidal effects). The spherical law of cosines formula gives well-conditioned results down to distances as small as around 1 meter. Given two points' latitude and longitude, the formula of the spherical law of cosines is as follows:

$$\begin{aligned}
 D &= R * \arccos [\sin (\text{lat}_1) * \sin (\text{lat}_2) + \cos (\text{lat}_1) \\
 &\quad * \cos (\text{lat}_2) * \cos (\text{long}_2 - \text{long}_1)] \\
 &= 2R * \arcsin \left(\left(\sin^2 \left(\frac{\text{lat}_2 - \text{lat}_1}{2} \right) + \cos (\text{lat}_1) \right. \right. \\
 &\quad \left. \left. * \cos (\text{lat}_2) \sin^2 \left(\frac{\text{long}_2 - \text{long}_1}{2} \right) \right)^{1/2} \right), \quad (4)
 \end{aligned}$$

where R is the Earth radius; here we use the average radius of the earth, which is equal to 6371.004 kilometers in average; lat_i , long_i represent the i point's latitude and longitude (in radian), respectively; D is the Great-circle distance between the two points.

Kamruzzaman and Hine [33] had proposed the fullness of activity spaces to describe the individual SDE feature class. The formula can be expressed as

$$F = \begin{cases} \frac{l_{Y,\text{axis}}}{l_{X,\text{axis}}}, & \text{if } l_{Y,\text{axis}} > l_{X,\text{axis}}, \\ \frac{l_{X,\text{axis}}}{l_{Y,\text{axis}}}, & \text{else if,} \end{cases} \quad (5)$$

where F is the fullness of activity spaces; $l_{Y,\text{axis}}$, $l_{X,\text{axis}}$ represent the length of the activity spaces locations' SDE y -axis and X -axis, respectively.

4. Results Analysis

4.1. Data Sources and Processing. In this research, we use the Taxi GPS traces data, which contains 13798 taxi records of nine consecutive days, from April 18, 2011 to the noon of April 26, 2011, and the total hours are 204 hours in Shenzhen, China. Among the 13798 taxis equipped with GPS, only 8845 taxi records can be used for analysis, and the other 4953 taxi GPS traces records have wrong or no record. After the data qualification, there are 3198 taxi vehicle's GPS traces covering 204 continuous hours. For each taxi equipped with GPS, the records information including taxi ID, date, time, location (longitude, latitude), velocity, driving direction, and taxi operation status (having passengers or not) is recorded every 5 or 10 seconds for the driving environment effects; under the effect of the traffic environment, the GPS record collection interval time period is not constant.

In the taxi GPS traces data processing, we adopt ArcGIS 9.3, R statistics software and Crime Stat software to do data mining and statistics work. Figure 4 shows the GPS data mining process.

We have divided the GPS data mining into three levels: the time level analysis, personal (each taxi driver) level, and the top 1% and last 1% taxi driver's comparison. From the time level we want to find taxi daily operation pattern (weekdays versus weekends), continuous time in one day, passenger's in vehicle time, and taxi's operation service frequency. This can be used in taxi operation planning, taxi pricing model, and taxi management policy implementation. From the personal level, we want to identify taxi driver's searching pattern, including searching activity space distribution and the relationship between the pick-up locations and the drop-off locations, which can be used in taxi dispatching and setting taxi stop location. From the comparison of the top and last 1% taxi drivers, the difference can be identified, which will be helpful for reducing passenger's waiting time and increasing driver's daily operation service frequency.

4.2. Time Level Analysis. Our first analysis has taken all the 3198 taxi vehicles operations as a whole to explore their operation temporal pattern per hour, which includes the pick-up passengers' locations and drop-off passengers' locations analysis.

Figure 5 shows all the taxi operation service frequency analysis; we can find these useful pieces of information.

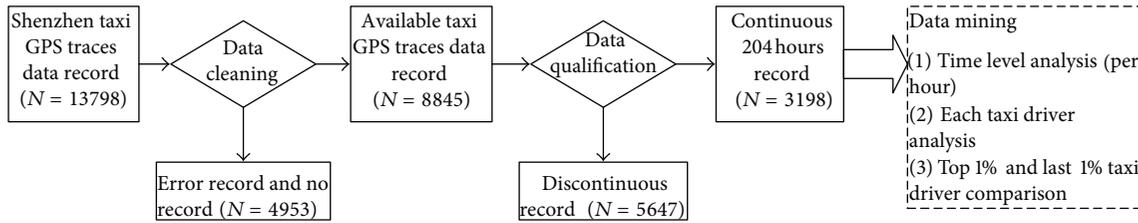


FIGURE 4: Taxi GPS data mining process.

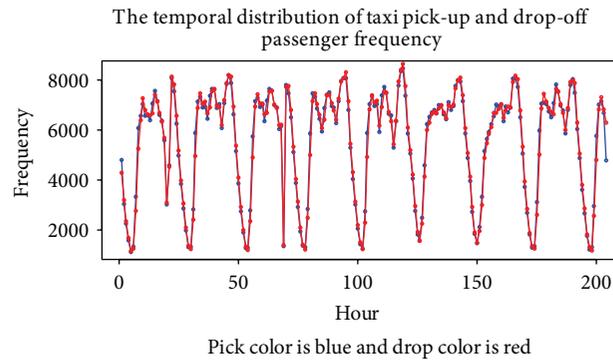


FIGURE 5: Taxi operation service frequency analysis.

- (1) In weekdays, there are three peak times in one day, which are 9 am to 10 am, 3 pm to 4 pm, and 10 pm to 11 pm, and the highest peak time interval is during 10 pm to 11 pm.
- (2) At weekends, there are only two peak times, which are 11 am to 12 pm on Saturday and 2 pm to 3 pm on Sunday, as well as the 10 pm to 11 pm at night. Meanwhile, the beginning time of trip increasing at weekdays is an hour later than weekdays.
- (3) The trend of pick-up and drop-off changes with the time is almost the same from Monday to Friday, and the lowest service time is from 5 am to 6 am.
- (4) When the service frequency increases, the number of pick-up locations will be higher than the drop-off locations; on the contrary, when the service frequency decreases, the number of pick-up locations will be lower than the drop-off locations.

Here we adopt the service frequency divided by the total taxi numbers ($N = 3198$) and obtain each taxi daily operation service frequency situation. The results are shown in Table 2; we can see that the lowest taxi service frequency is less than 0.4 per hour, and the highest is almost 2.7 per hour. The average number of each taxi operating per day is between 39 and 44 times, which is higher than other cities statistics, such as in Harbin where there are 32 taxi service times for each taxi operating per day in 2008 [3], as well as 31 to 33 taxi service times per day in Beijing in 2007 [34]. During 2006 to 2007, Chen [5] also calculated the average operation time for each taxi per day, which is between 35 and 40 times per taxi per day, which is also lower than that in 2011.

According to the researches of Kamruzzaman and Hine [30, 33], we analyze the correlations between the different measures of activity space area, which are shown in Table 3 and Figure 6.

From Figure 6 and Table 3 we can find that the drop-off locations cover area range bigger than the pick-up locations, which can also reflect the disorder and discrete characteristic of the passenger's destination distribution. Meanwhile, the area of location's SDD is bigger than the SDE, regardless of the pick-up or drop-off locations.

4.3. The Personal Level. In this part, we have calculated the distance between the mean centre of each taxi driver's pick-up location and drop-off location and then analyzed the distribution of these distances and the relationship between the pick-up locations and drop-off locations.

Table 4 shows the statistical information of the entire 3198 taxi vehicle's service frequency and the distance between the mean centre of each taxi driver's pick-up locations and drop-off locations. With the combination of Figure 7, more than 80 percent of the taxi vehicles operation service frequency is between 300 and 500 times during the 204 hours.

Figure 8 shows the distance distribution and the frequency distribution curve of taxi pick-up and drop-off passenger locations' mean centre, respectively. The distance between each taxi vehicle's drop-off and pick-up locations' mean centre is below 5000 meters; however, the mean centre of pick-up (or drop-off) locations' distance between each taxi vehicle and the whole operation is changing from 500 meters to 30000 meters, which means each taxi vehicle's operation discrete and randomness. The density line color is blue, and the normal distribution line color is red.

TABLE 2: Each taxi driver’s daily average operation service frequency statistics.

Time periods	April 18		April 19		April 20		April 21		April 22		April 23		April 24		April 25		April 26	
	Monday		Tuesday		Wednesday		Thursday		Friday		Saturday		Sunday		Monday		Tuesday	
	Drop	Pick																
0:00–0:59	1.35	1.50	1.62	1.56	1.68	1.61	1.67	1.60	1.70	1.65	1.91	1.83	1.90	1.83	1.67	1.59	1.58	1.53
1:00–1:59	1.00	0.95	1.24	1.20	1.28	1.21	1.26	1.21	1.35	1.27	1.63	1.58	1.59	1.53	1.24	1.16	1.19	1.13
2:00–2:59	0.74	0.71	0.96	0.90	0.91	0.86	0.98	0.92	0.99	0.95	1.32	1.27	1.30	1.24	0.89	0.84	0.90	0.83
3:00–3:59	0.53	0.50	0.66	0.62	0.63	0.59	0.66	0.61	0.69	0.64	0.93	0.87	0.91	0.86	0.59	0.57	0.57	0.55
4:00–4:59	0.36	0.35	0.43	0.41	0.42	0.40	0.44	0.43	0.47	0.45	0.60	0.57	0.60	0.58	0.42	0.41	0.41	0.38
5:00–5:59	0.39	0.42	0.39	0.42	0.37	0.41	0.38	0.42	0.39	0.40	0.49	0.49	0.47	0.46	0.39	0.42	0.37	0.41
6:00–6:59	0.86	1.04	0.75	0.89	0.74	0.87	0.78	0.89	0.72	0.86	0.70	0.78	0.62	0.67	0.82	0.97	0.81	0.93
7:00–7:59	1.64	1.90	1.55	1.84	1.54	1.80	1.56	1.83	1.54	1.84	1.29	1.43	0.94	1.05	1.57	1.87	1.50	1.80
8:00–8:59	2.00	2.05	2.15	2.23	2.17	2.24	2.24	2.34	2.14	2.20	1.88	1.95	1.53	1.61	2.15	2.21	2.14	2.19
9:00–9:59	2.27	2.20	2.34	2.30	2.33	2.28	2.34	2.29	2.31	2.30	1.97	2.04	1.71	1.77	2.33	2.22	2.29	2.24
10:00–10:59	2.13	2.06	2.20	2.16	2.21	2.16	2.20	2.15	2.22	2.18	2.08	2.10	1.83	1.85	2.23	2.20	2.13	2.09
10:00–11:59	2.07	2.06	2.23	2.17	2.21	2.20	2.07	2.01	2.24	2.20	2.11	2.13	1.91	1.95	2.15	2.11	1.97	1.50
12:00–12:59	2.06	2.00	2.08	2.01	2.07	1.99	1.90	1.86	1.93	1.85	2.09	2.09	2.09	2.05	2.10	2.04		
13:00–13:59	2.09	2.21	2.16	2.31	2.09	2.25	2.01	2.15	2.17	2.31	2.14	2.19	2.09	2.19	2.07	2.21		
14:00–14:59	2.32	2.37	2.36	2.39	2.36	2.39	2.31	2.32	2.36	2.42	2.19	2.19	2.14	2.17	2.39	2.45		
15:00–15:59	2.23	2.24	2.39	2.39	2.36	2.37	2.32	2.35	2.34	2.34	2.09	2.08	2.20	2.19	2.36	2.35		
16:00–16:59	2.09	2.07	2.18	2.15	2.19	2.19	2.21	2.17	2.11	2.08	2.02	2.01	2.03	1.99	2.20	2.19		
17:00–17:59	1.98	1.99	2.16	2.20	2.15	2.16	2.12	2.16	1.99	2.06	2.19	2.22	2.12	2.17	2.09	2.12		
18:00–18:59	1.78	1.75	1.94	1.90	1.94	1.89	2.00	1.96	1.70	1.65	2.17	2.13	2.17	2.10	1.88	1.84		
19:00–19:59	0.97	0.95	2.25	2.22	1.94	1.92	2.27	2.24	2.00	1.98	2.19	2.18	2.16	2.16	2.16	2.13		
20:00–20:59	1.42	1.44	2.47	2.45	0.44	0.42	2.48	2.49	2.48	2.43	2.40	2.43	2.47	2.52	2.48	2.44		
21:00–21:59	2.54	2.52	2.56	2.56	2.41	2.44	2.53	2.53	2.63	2.61	2.50	2.50	2.56	2.52	2.50	2.52		
22:00–22:59	2.45	2.36	2.54	2.46	2.43	2.33	2.60	2.53	2.70	2.65	2.53	2.48	2.51	2.42	2.46	2.34		
23:00–23:59	2.06	1.96	2.17	2.07	2.13	2.03	2.24	2.13	2.43	2.30	2.31	2.24	2.12	2.04	1.97	1.89		
Total	39.33	39.6	43.78	43.81	41	41.01	43.57	43.59	43.6	43.62	43.73	43.78	41.97	41.92	43.11	43.09	15.86	15.58

TABLE 3: The statistics of the different measures of activity space area.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Pick-up location SDD area (Km ²)	1227	1410	1477	1483	1547	1796
Pick-up location SDE area (Km ²)	1075	1200	1257	1266	1318	1532
Drop-off location SDD area (Km ²)	1344	1481	1544	1580	1655	2727
Drop-off location SDE area (Km ²)	1129	1271	1336	1361	1427	2173

TABLE 4: The statistics of the different measures of each driver’s operation activity space.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Service frequency (times)	150	302	369	356	413	597
Distance between the two mean centres (meters)	0.0	252.1	392.5	432.4	565.5	5121.8
Pick-up location SDC area (Km ²)	34.55	477.32	629.26	623.68	782.41	1947.05
Pick-up location SDE area (Km ²)	20.66	345.98	451.03	452.95	555.66	2806.99
Drop-off location SDC area (Km ²)	35.4	563.6	727.2	727.1	889.5	6279.4
Drop-off location SDE area (Km ²)	18.56	425.63	542.24	549.60	663.40	2921.57

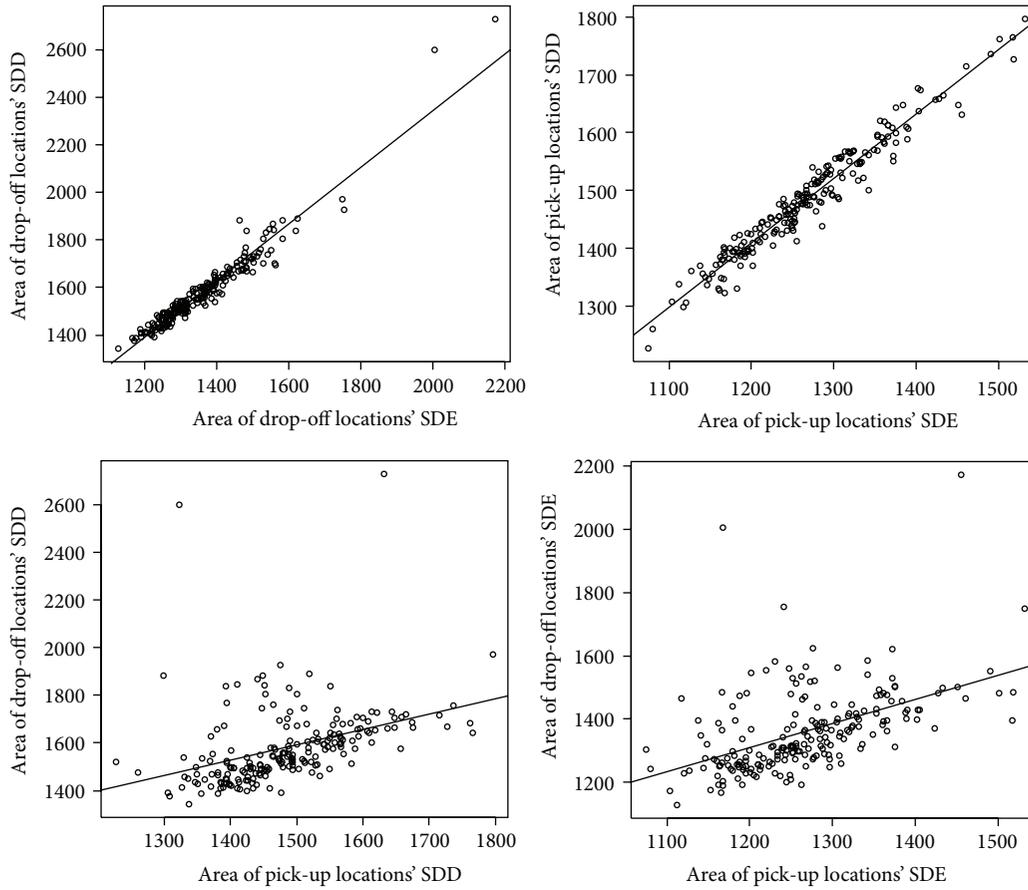


FIGURE 6: The correlations between the different measures of activity space area.

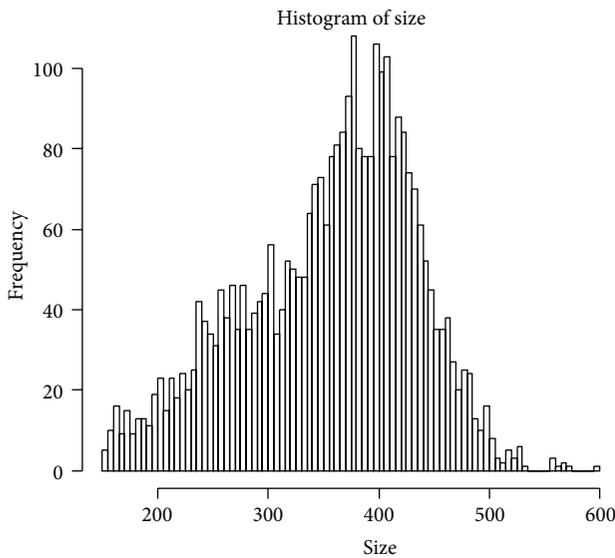


FIGURE 7: Taxi's service frequency distribution graph.

From Figure 8, we can also find that the frequency distribution of taxi pick-up and drop-off passenger locations' mean centre has some aggregate character, so we adopt

nearest neighbor hierarchical clustering method to calculate each taxi vehicles' pick-up and drop-off location, which are shown in Figures 9 and 10, respectively.

The nearest neighbor hierarchical clustering method can identify small geographical environments where there are concentrated pick-up or drop-off activities. And the linkages between several small clusters can be seen through the second- and third-order clusters; there are different scales to the clustering of points' different geographical levels. Also the analysis can show the number of points found per cluster and the density of points found per cluster.

From Figure 11 we can find that the drop-off locations cover area range bigger than the pick-up locations, which can also reflect the disorder of the passenger's destination distribution.

Kamruzzaman and Hine [33] had proposed the fullness of activity spaces to describe the individual SDE feature class. This can be seen in Figure 12.

4.4. *The Comparison between the Top and Last 1% Taxi Drivers' Level.* From Figure 13 and Table 5 we can find the difference between the top 1% and last 1% taxi searching time/operation time. The top 1% taxi drivers have shorter time both in searching and in operation time than the last 1% taxi drivers. There is more than 22 percent of the last 1% taxi driver's

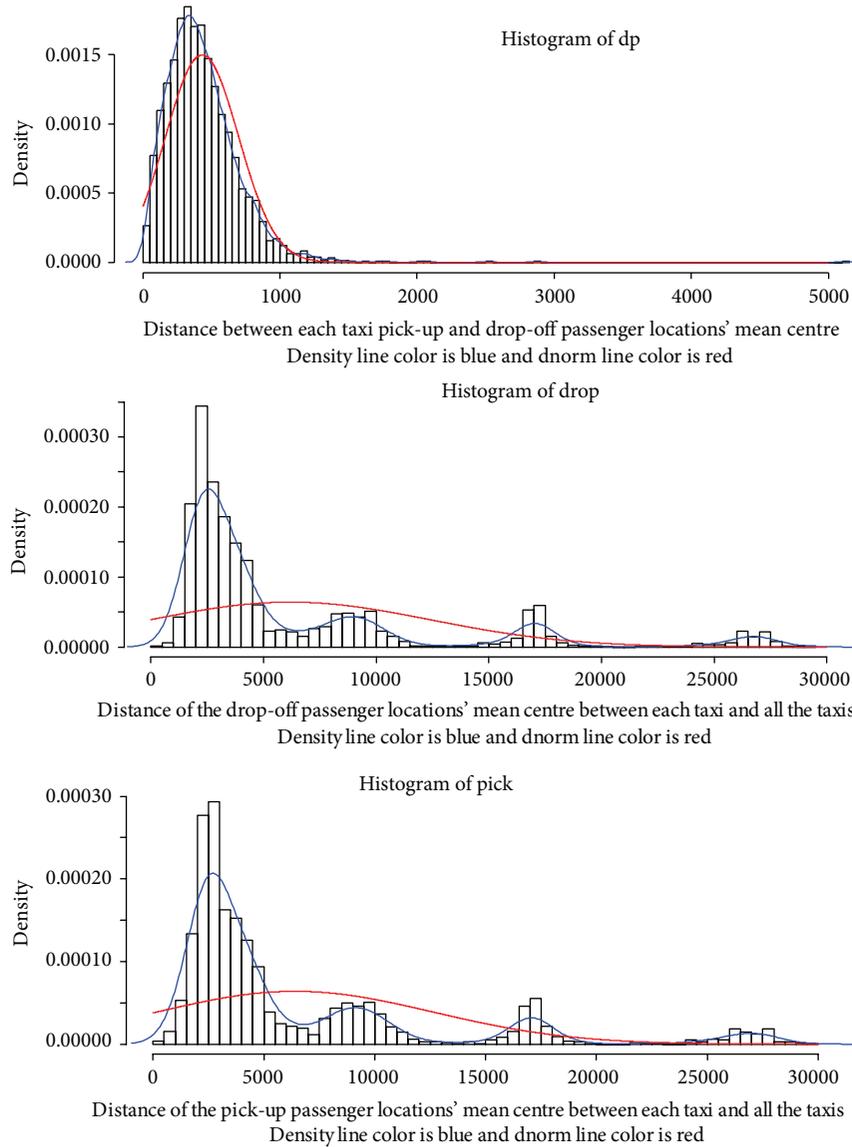


FIGURE 8: The frequency distribution curve of taxi pick-up and drop-off passenger locations' mean centre.

TABLE 5: The statistics of the top 1% and last 1% taxi searching time/operation time.

Percent	Searching time (minutes)		Operation time (minutes)	
	Top 1%	Last 1%	Top 1%	Last 1%
25%	3	7	6	7
50%	6	20	9	12
75%	14	42	15	24
85%	23	>60	20	35

searching time which is above 60 minutes; furthermore, the top 1% taxi drivers are only 5 percent above 60 minutes.

Referring to the concept of 85 percent vehicle speed, here we calculate the 85 percent searching time and operation time of the top 1% and last 1% taxi drivers, which are shown in

Table 5. The top 1% taxi drivers' 85 percent operation and searching time are at the interval from 20 to 23 minutes; however, the last 1% taxi drivers' 85 percent operation and searching time is 35 and more than 60 minutes, respectively. That is why the top 1% taxi drivers can operate more service time than the last 1% taxi drivers.

5. Discussions

In the time level, we have found that taxi's operations in weekdays and weekends have different operation status; in weekdays there are three peak hours during one day and in weekends there are only two peak hours. Meanwhile, the time level of taxis operation reflects its difference with the normal traffic demand peak hour distribution. There are always two peak hours in the normal traffic demand (as the dual hump curve), which corresponds to going to work and getting off



FIGURE 9: (a) First-, second-, and third-order pick-up locations' mean centre "hot spots": ellipses. (b) First-, second-, and third-order pick-up locations' mean centre "hot spots": convex hull.

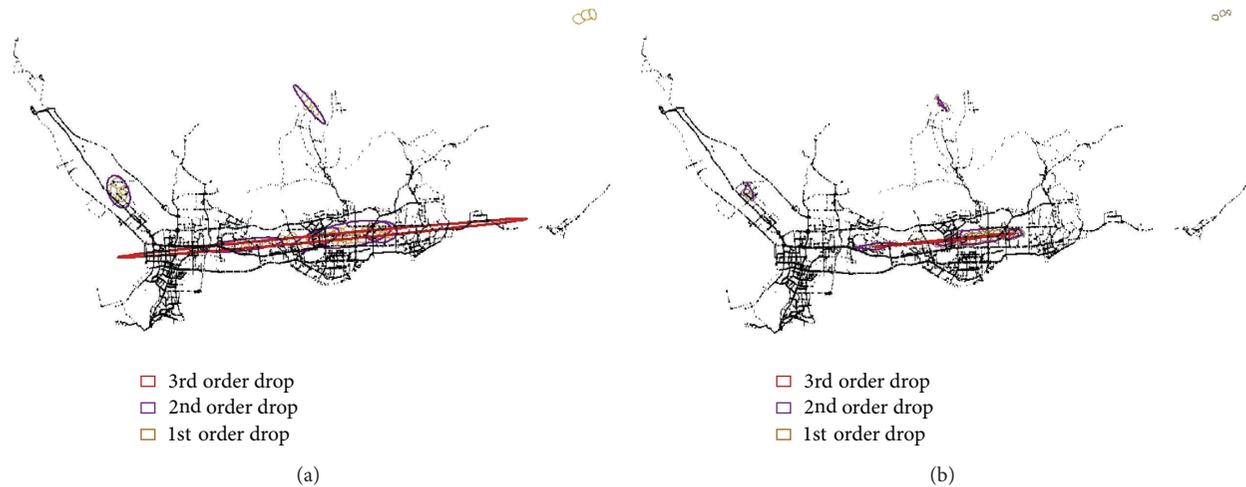


FIGURE 10: (a) First-, second-, and third-order drop-off locations' mean centre "hot spots": ellipses. (b) First-, second-, and third-order drop-off locations' mean centre "hot spots": convex hull.

work; in addition to this, the taxi operation demand has the third peak hour at night in weekdays, which is also the biggest peak hour. So in the traffic demand forecasting and planning process, we should pay more attention to this new phenomenon and take the taxi demand into account separately.

In the study field of taxi service model, researchers usually adopt virtual customer origin-destination demand pattern to analyze the model, such as Arnott [35], Yang and Wong [36], and Wong et al. [37]. While with the taxi GPS traces data, we can obtain the customers real-time origin and destination demand, which can assist researchers to validate the taxi service model.

On the personal (each taxi driver) level, the distance between each taxi driver's pick-up locations' mean centre and drop-off locations' mean centre is almost around 1000 meters, which reflects the taxi driver's cruise and search for

next passenger' pick-up location will be around the previous drop-off location. The cluster analysis also identifies the small geographical environments where there are concentrated pick-up and drop-off activities. This finding can be useful in the taxi dispatching and setting taxi stop location chosen.

Through the comparison between the top and last 1% taxi drivers' operation status, the difference between the searching time and operation time statistics can help the taxi operation company to make strategy for increasing the taxi driver's daily operation service frequency, which will also reduce the passenger's waiting time and increase their satisfaction.

6. Conclusions

This paper first took the taxi driver's pick-up and drop-off passengers' location as the objective, based on large-scale taxi GPS trace data analyzing the urban taxi driver's temporal

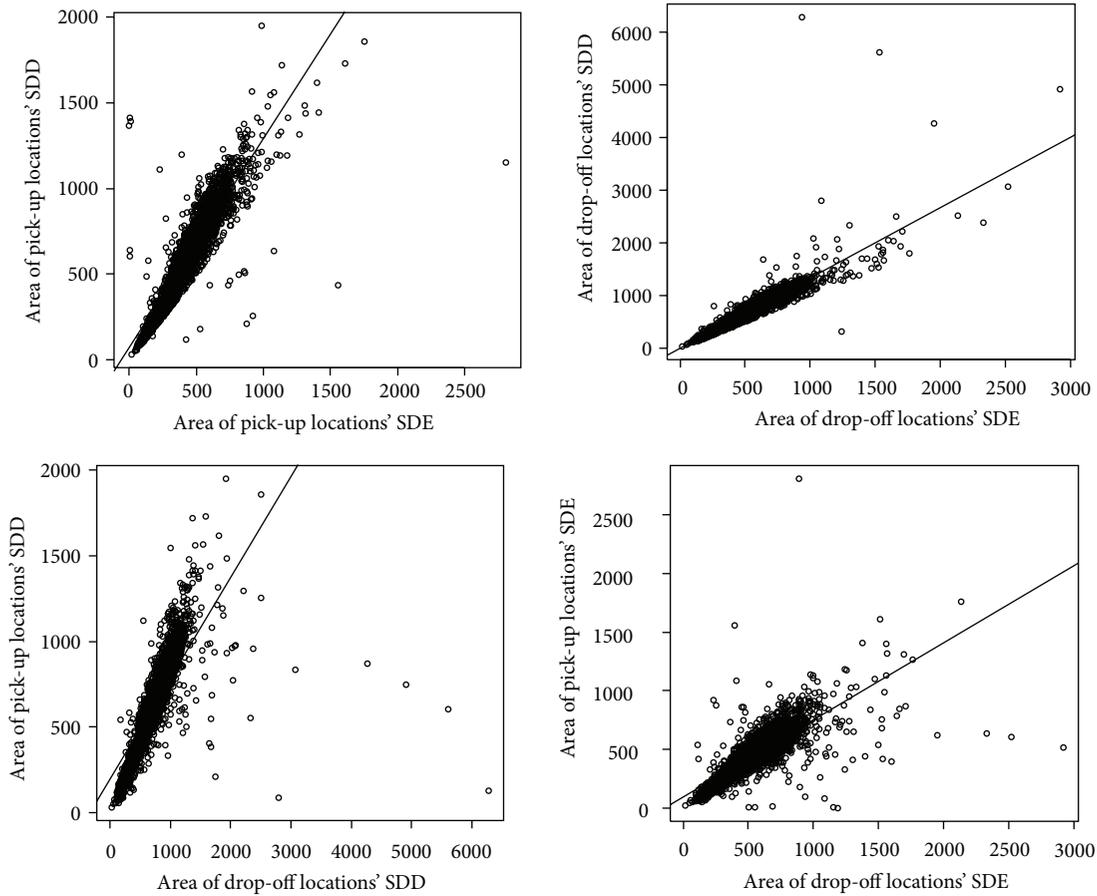


FIGURE 11: The correlations between the different measures of activity space area.

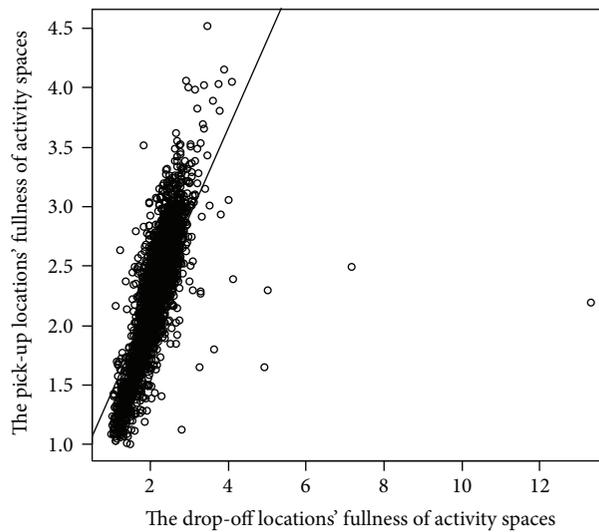


FIGURE 12: The fullness of activity spaces of each taxi driver's pick-up location and drop-off location.

and spatial distribution. From the time level, we found the difference between taxi daily operation pattern (weekday and weekends), continuous time in one day, passenger's in-vehicle time, and taxi's operation frequency. From the personal level

(each taxi vehicle), we have identified the taxi driver's searching pattern, including searching activity space distribution and the relationship between the pick-up locations and the drop-off locations. Through the comparison between the

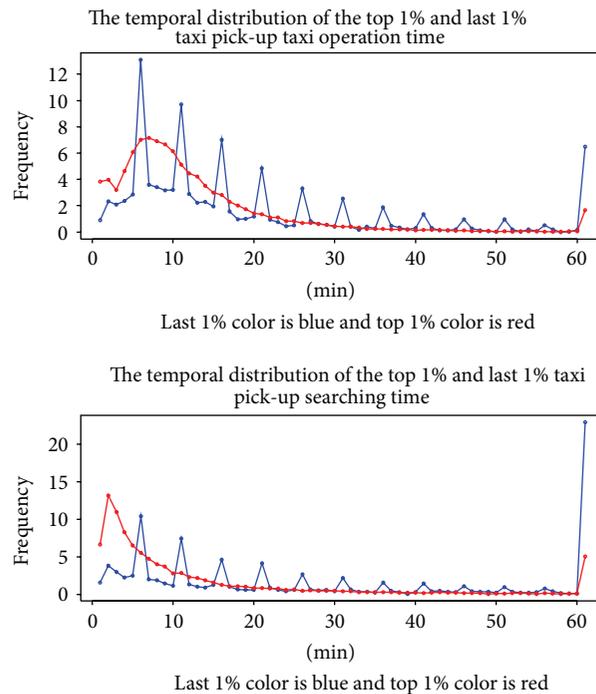


FIGURE 13: The temporal distribution of the top 1% and last 1% taxi searching time/operation time.

top and last 1% taxi drivers difference on searching time, operation time, and service frequency, which can obtain the pattern of top and last 1% taxi drivers operation pattern. This paper can be helpful for the urban taxi management authority to identify hot spots and develop policies that reduce the taxi drivers' searching time, road congestion, and taxi air pollution emission. The measurements of the taxi driver's activity space analysis also make some contribution to the geographical dynamics analysis.

The next step is to identify the relationship between the taxi driver's pick-up activity and the land-use attribute, as well as each taxi driver's temporal distribution, which will help us to understand taxi drivers' searching behavior pattern.

Conflict of Interests

The authors declare that they have no conflict of interests.

Acknowledgments

This research was supported by the National High-Tech Research and Development Program of China (863 Program, no. 2012AA112310), the China Postdoctoral Science Foundation funded project (no. 2013M540299), the Fundamental Research Funds for the Central Universities (no. HIT.NSRIF.2015075), and the Heilongjiang Postdoctoral Fund (no. LBH-Z13111). The authors would like to express their sincere thanks to Professor Yanfeng Ouyang (University of Illinois at Urbana-Champaign) and the anonymous reviewers for their helpful comments and valuable suggestions on this paper.

References

- [1] S. S. Han, "Managing motorization in sustainable transport planning: the Singapore experience," *Journal of Transport Geography*, vol. 18, no. 2, pp. 314–321, 2010.
- [2] BTRC, "Beijing Transport," Annual Report, Beijing Transportation Research Center, Beijing, China, 2010, (Chinese).
- [3] S. An, X. Hu, and J. Wang, "Urban taxis and air pollution: a case study in Harbin, China," *Journal of Transport Geography*, vol. 19, no. 4, pp. 960–967, 2011.
- [4] X. Hu and J. Feng, "Research on characteristics of taxi traffic based on GPS data," *Urban Transport of China*, vol. 5, no. 2, pp. 91–95, 2007 (Chinese).
- [5] L.-H. Chen, *Study of taxi operation trait based on GPS float data [M.S. thesis]*, Tongji University, 2008 (Chinese).
- [6] B. Jiang, J. Yin, and S. Zhao, "Characterizing the human mobility pattern in a large street network," *Physical Review E*, vol. 80, no. 2, Article ID 021136, 2009.
- [7] J. Weng, W. Liu, Z. Chen, and J. Rong, "Research on floating car data based taxi operation and management," *Journal of Beijing University of Technology*, vol. 36, no. 6, pp. 779–784, 2010 (Chinese).
- [8] L. Liu, C. Andris, and C. Ratti, "Uncovering cabdrivers' behavior patterns from their digital traces," *Computers, Environment and Urban Systems*, vol. 34, no. 6, pp. 541–548, 2010.
- [9] J. Hu, H. Xie, and G. Zhong, "A method for determining taxi reasonable scale based on floating car data," in *ICTIS 2011: Multimodal Approach to Sustained Transportation System Development: Information, Technology, Implementation, Wuhan, China*, pp. 1255–1262, 2011.
- [10] Z. Zhang and X. He, "Analysis and application of spatial distribution of taxi service in city subareas based on taxi GPS data," in *Proceedings of the 11th International Conference*

- of Chinese Transportation Professionals: Towards Sustainable Transportation Systems (ICCTP '11), pp. 1232–1243, Nanjing, China, August 2011.
- [11] Y. Zheng, J. Yuan, W. Xie, X. Xie, and G. Sun, “Drive smartly as a taxi driver,” in *Proceedings of the Symposia and Workshops on Ubiquitous, Autonomic and Trusted Computing (UIC-ATC '10)*, pp. 484–486, IEEE Computer Society, Washington, Wash, USA, 2010.
- [12] D. Luo, *Urban mixed traffic network equilibrium analysis under the influence of taxi services*, Central South University, 2009 (Chinese).
- [13] R. G. Golledge and R. J. Stimson, *Spatial Behavior*, The Guilford Press, New York, NY, USA, 1997.
- [14] K. W. Axhausen, “Travel diaries: an annotated catalogue,” Working Paper, Imperial College, London, UK, 1994.
- [15] D. Wang and T. Cheng, “A spatio-temporal data model for activity-based transport demand modelling,” *International Journal of Geographical Information Science*, vol. 15, no. 6, pp. 561–585, 2001.
- [16] Y. O. Susilo and R. Kitamura, “Analysis of the day-to-day variability in the individual’s action space: an exploration of the six-week Mobidrive travel diary data,” *Transportation Research Record*, no. 1902, pp. 124–133, 2005.
- [17] M. C. González, C. Hidalgo, and A.-L. Barabási, “Understanding individual human mobility patterns,” *Nature*, vol. 453, pp. 779–782, 2008.
- [18] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási, “Limits of predictability in human mobility,” *Science*, vol. 327, no. 5968, pp. 1018–1021, 2010.
- [19] M. Giraud and P. Peruch, “Spatio-temporal aspects of the mental representation of urban space,” *Journal of Environmental Psychology*, vol. 8, no. 1, pp. 9–17, 1988.
- [20] P. Peruch, M.-D. Giraud, and T. Garling, “Distance cognition by taxi drivers and the general public,” *Journal of Environmental Psychology*, vol. 9, no. 3, pp. 233–239, 1989.
- [21] H. J. Spiers and E. A. Maguire, “The dynamic nature of cognition during wayfinding,” *Journal of Environmental Psychology*, vol. 28, no. 3, pp. 232–249, 2008.
- [22] Y. Wakabayashi, S. Itoh, and Y. Nagami, “The use of geospatial information and spatial cognition of taxi drivers in Tokyo,” *Procedia Social and Behavioral Sciences*, vol. 21, pp. 353–361, 2011.
- [23] J. Pailhous, *La Representation de l’Espace Urbain*, Presses Universitaires de France, Paris, France, 1970.
- [24] J. A. Michon, “Processing of temporal information and the cognitive theory of time experience,” in *The Study of Time*, J. T. Fraser, F. C. Haber, and C. H. Muller, Eds., pp. 242–257, Springer, Heidelberg, Germany, 1972.
- [25] E. Murakami and D. P. Wagner, “Can using global positioning system (GPS) improve trip reporting?” *Transportation Research Part C: Emerging Technologies*, vol. 7, no. 2-3, pp. 149–165, 1999.
- [26] Y. Yue, H. Wang, B. Hu, Q. Li, Y. Li, and A. G. O. Yeh, “Exploratory calibration of a spatial interaction model using taxi GPS trajectories,” *Computers, Environment and Urban Systems*, vol. 36, no. 2, pp. 140–153, 2012.
- [27] Y. Ge, H. Xiong, A. Tuzhilin, K. Xiao, M. Gruteser, and M. J. Pazzani, “An energy-efficient mobile recommender system,” in *Proceeding of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '10)*, pp. 899–908, New York, NY, USA, July 2010.
- [28] J. Lee and D. W. S. Wong, *Statistical Analysis with Arcview GIS*, John Wiley & Sons, New York, NY, USA, 2001.
- [29] Ned Levine & Associates, *CrimeStat: A Spatial Statistics Program for the Analysis of Crime Incident Locations (v 3.0)*, Ned Levine & Associates, Houston, Tex, USA; National Institute of Justice, Washington, DC, USA, 2004.
- [30] M. Kamruzzaman and J. Hine, “Participation index: a measure to identify rural transport disadvantage?” *Journal of Transport Geography*, vol. 19, no. 4, pp. 882–899, 2011.
- [31] R. W. Sinnott, “Virtues of the Haversine,” *Sky and Telescope*, vol. 68, no. 2, article 159, 1984.
- [32] K. Gade, “A non-singular horizontal position representation,” *Journal of Navigation*, vol. 63, no. 3, pp. 395–417, 2010.
- [33] M. Kamruzzaman and J. Hine, “Analysis of rural activity spaces and transport disadvantage using a multi-method approach,” *Transport Policy*, vol. 19, no. 1, pp. 105–120, 2012.
- [34] L. Song, D. Zhang, J. Chen, J. An, and Y. Su, “Analysis of taxi operation characteristics with traffic control,” *Journal of Transportation Systems Engineering and Information Technology*, vol. 8, no. 6, pp. 127–121, 2008 (Chinese).
- [35] R. Arnott, “Taxi travel should be subsidized,” *Journal of Urban Economics*, vol. 40, no. 3, pp. 316–333, 1996.
- [36] H. Yang and S. C. Wong, “A network model of urban taxi services,” *Transportation Research B: Methodological*, vol. 32, no. 4, pp. 235–246, 1998.
- [37] K. I. Wong, S. C. Wong, and H. Yang, “Modeling urban taxi services in congested road networks with elastic demand,” *Transportation Research B: Methodological*, vol. 35, no. 9, pp. 819–842, 2001.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

