

FANG ZONG, Ph.D.
E-mail: zongfang@jlu.edu.cn
Jilin University, College of Transportation
5988 Renmin Street, Changchun, Jilin, 130022, China
XIAO SUN, Prof.
Corresponding author
E-mail: sunxiao@jlu.edu.cn
Jilin University, Applied Technology College
5372 Nanhu Road, Changchun, Jilin, 130022, China
HUIYONG ZHANG, Ph.D.
E-mail: 15143173332@163.com
Jilin University, School of Management
5988 Renmin Street, Changchun, Jilin, 130022, China
XIUMEI ZHU, Ph.D.
E-mail: xmzhu@jlu.edu.cn, 327514076@qq.com
Jilin University, School of Management
5988 Renmin Street, Changchun, Jilin, 130022, China
WENTIAN QI, M.S.
E-mail: 492953906@qq.com
Suzhou Institute of Construction and Communications
Suzhou, Jiangsu, 215000, China

Traffic Management
Original Scientific Paper
Submitted: Sep. 25, 2014
Approved: Nov. 24, 2015

UNDERSTANDING TAXI DRIVERS' MULTI-DAY CRUISING PATTERNS

ABSTRACT

This study investigates taxi drivers' multi-day cruising behaviours with GPS data collected in Shenzhen, China. By calculating the inter-daily variability of taxi drivers' cruising behaviours, the multi-day cruising patterns are investigated. The impacts of learning feature and habitual feature on multi-day cruising behaviours are determined. The results prove that there is variability among taxis' day-to-day cruising behaviours, and the day-of-week pattern is that taxi drivers tend to cruise a larger area on Friday, and a rather focused area on Monday. The findings also indicate that the impacts of learning feature and habitual feature are more obvious between weekend days than among weekdays. Moreover, learning feature between two sequent weeks is found to be greater than that within one week, while the habitual feature shows recession over time. By revealing taxis' day-to-day cruising pattern and the factors influencing it, the study results provide us with crucial information in predicting taxis' multi-day cruising locations, which can be applied to simulate taxis' multi-day cruising behaviour as well as to determine the traffic volume derived from taxis' cruising behaviour. This can help us in planning of transportation facilities, such as stop stations or parking lots for taxis. Moreover, the findings can be also employed in predicting taxis' adjustments of multi-day cruising locations under the impact of traffic management strategies.

KEY WORDS

taxi; multi-day; cruising pattern; GPS;

1. INTRODUCTION

Taxi is one of the major transportation modes in urban transportation system. Yet, empty taxi cruising results in significant amount of wasted energy and emissions. As we know, up to 40 percent of taxis' daily mileage is spent with no passengers on board [1]. The average dead mileage for taxis in Beijing is 120 km per day, which accounts for 40% of total daily mileage. This means that 13 litres of fuel are used per day while driving with no passengers [2]. Therefore, it is important to identify ways to make empty taxis cruise more efficiently. This paper aims at learning taxis' multi-day cruising patterns and providing important information for taxi drivers and transportation managers to enhance taxis' cruising efficiency. By investigating the factors that affect taxi drivers' cruising decisions and by examining their multi-day cruising behaviours, we will try to find the underlying mechanisms of taxis' day-to-day cruising decisions.

In recent years, massive progress has been made in travel survey with the rapid development and application of Global Positioning System (GPS) technology. As an advanced computer-based technique, GPS offers significant advantages over traditional survey methods in collecting travel data, especially its ability to capture multi-day data [3]. Many researchers pre-

sented studies on taxi's driving behaviours by using GPS data [4]. This study will use multi-day GPS data collected from 2,536 taxi drivers in Shenzhen, China. As GPS-collected data are not ready for direct use in cruising behaviour analysis, some computer-based techniques and methods will be used, such as Quadrant Analysis and Entropy, in processing and analysing GPS data.

With wide application of GPS devices in multi-day travel survey, the analysis of multi-day travel behaviour has been one of the hot issues in transportation studies. However, none of the researchers analysed the multi-day operating patterns or modelled the process of day-to-day dynamics of taxis' cruising behaviours. In this paper an analysis of taxis' multi-day cruising patterns will be conducted. Two major features of taxis' day-to-day cruising behaviours, i.e., learning from previous pick-up experience (learning feature) vs. obeying cruising habits (habitual feature), will be investigated. In particular, we seek to answer three questions: 1) Is there variability between taxi drivers' day-to-day cruising behaviours? 2) What are the differences among multi-day cruising behaviours? 3) How do the two factors, i.e., learning and habits, affect taxi drivers' day-to-day cruising behaviours?

The remainder of this paper is organized as follows. Section 2 presents a review on the analysis of taxis' cruising behaviours in general. Section 3 presents a description of data and the study area. In Section 4 the spatial distributions of cruising traces and pick-up points are examined. In Section 5, the multi-day variability of cruising patterns is calculated, and the different impacts of learning feature and habitual feature in day-to-day cruising behaviours are investigated. The paper is concluded with Section 6, in which the findings are summarized and the directions for future research discussed.

2. LITERATURE REVIEW

The multi-day travel behaviour analysis has been one of the hot issues in transportation research. However, none of the studies analysed the multi-day operating pattern or modelled the day-to-day dynamics of taxis' cruising behaviours. There're only a few researchers who considered previous operating experience in modelling taxis' driving patterns. For example, Reference [5] modelled a taxi service system in urban areas, taking into account the taxi drivers' knowledge of the transportation network from their day-to-day experience. Reference [6] found that driver's travel experience as well as a renewal of all acquirable traffic information that can assist in confirming the reference points in road network have influence on the route choice of the next trip. Reference [7] presented an experiential approach to compute optimal paths, in which path-planning was supported by a flexible road

network hierarchy using the experience of taxi drivers. Other studies that investigated driving experience are Yu et al.'s work on modelling bus arrival times [8, 9] and vehicle's route choice [10], and Yao et al.'s study on accident detection [11], etc.

Several models were developed to investigate the day-to-day dynamics of travel behaviours, such as reference [12] proposed a stochastic process approach to analyse day-to-day dynamics in a transportation network. A key assumption of the stochastic process is that the path choice probabilities are time homogeneous, which means that there are no time-dependent variations in drivers' perceptions. This approach was further extended by reference [13] to include day dynamics. Reference [14] described a simple route choice model over time whereby decisions are based on the weighted average of previous travel decisions' utilities. Based on this model, reference [15] proposed a myopic adjustment model by allowing the weights to vary across individuals. Reference [16] presented detailed discussions regarding the habitual and variable behaviour of individuals over time. They noted that some behaviours, when examined in a disjointed framework (say, a work trip examined in isolation of the overall daily activity-travel pattern), are repeated on a day-to-day or week-to-week basis. However, when the overall daily activity-travel pattern is examined in its entirety, they found that a one-day pattern is not representative of a person's routine travel.

Previous studies also proposed two major features of multi-day travel decisions: learning feature vs. habitual feature. With respect to learning feature, some researchers mentioned above proved that taxi drivers have the characteristics of learning, which means that they will update their traffic information from their day-to-day experience and use the information in route choice of the next trip [5, 6]. Besides taxi related studies there are also some researchers who modelled the learning process of other travel behaviours. For example, reference [17] proposed travellers' decision-making process based on Bayesian updating by considering day-to-day dynamics of travel behaviours. Reference [14] proposed a learning model for route choice selection. Reference [18] developed a learning model for mode choice.

The habitual feature has been also proved and utilized by many studies related to travel behaviour analysis. In many studies about mode choice modelling, the factor of habit was also taken as a major variable [19, 20]. Reference [21] demonstrated that when behaviours are well-practiced and repeatedly performed, frequency of past behaviour reflects habit strength and has direct effect on future performance. Reference [22] found that although most drivers thought that they had two or three alternative routes, the majority said they always stuck to the same one. This also demonstrated the crucial role of habits in travel decisions.

Though the above studies proved that travel behaviours have the learning features and habitual features, none of them conducted a quantitative comparison between them. This study will present a quantitative comparison between these two features, and measure the multi-day variability of taxi drivers' cruising patterns.

3. DATA AND STUDY AREA

The data used in this study are GPS traces collected from 2,536 taxis during a 9-day period (November 18th 2011 – November 26th 2011), in Shenzhen, China. The surveyed taxis account for about 17.14% of the total number of taxis and 76.64% of taxis with on-board GPS equipment in Shenzhen. Each record in the dataset represents a GPS signal that was captured consecutively at a 5-second interval (Shenzhen Transport Information Center, 2011). The basic information provided by the dataset contain taxis' traces in longitude and latitude form, vehicle identification, operation status (empty or occupied), timestamp, spot speed, and azimuth. During the 9-day period, 30,179,212 valid positions in total were recorded.

In order to learn the multi-day cruising patterns as well as the features of cruising behaviours, two types of points are selected from the GPS traces, i.e., cruising traces (without passengers) and pick-up points (locations whose operation status is from vacancy to occupancy), in which cruising traces represent taxi drivers' cruising patterns. The spatial correlations among different days' cruising traces and pick-up points are used to represent learning vs. habitual features.

The study area of this paper is the urban area of Shenzhen city, which is one of the four most developed cities in China. It had a residing population of 10.63 million in 2013 and 10 districts. The administrative division of Shenzhen is shown in *Figure 1*.

As GPS-collected data are not ready for direct use in cruising behaviour analysis, one of the widely used spatial analysis techniques – Quadrat Analysis – will be employed in data processing. According to the theory of Quadrat Analysis, the entire study area is divided into cells. The area of each cell ($Area_g$) can be calculated as:

$$Area_g = 2A/Q \tag{1}$$

where A is the entire study area and Q is the number of traces that will be analysed in the study area. The X and Y coordinates for the study area range from 167.53 to 255.81 and from 2,486.22 to 2,531.10, respectively. Based on these ranges, the total area A can be calculated as 3,962.20 square kilometres. Since this study is concerned with drivers' traces on one day, the average number of GPS traces of each driver per day, i.e., 9,083, is taken as the value of Q . Thus, the area of each cell is then calculated as 0.8724 square kilometres, with each side being equal to 0.9340 km. Thus, the study area is divided into $96 \times 50 = 4,800$ cells. For X coordinate, there are $(255.81 - 167.53) / 0.93 = 94.92$ segments, and for Y coordinate, there are $(2,531.10 - 2,486.22) / 0.93 = 48.26$ segments. In order to cover the four sides of the study area as well as to make the numbers of the segments integers, one cell is expanded for each side of the study area. This yields 96 and 50 segments for X coordinate and Y coordinate, respectively.

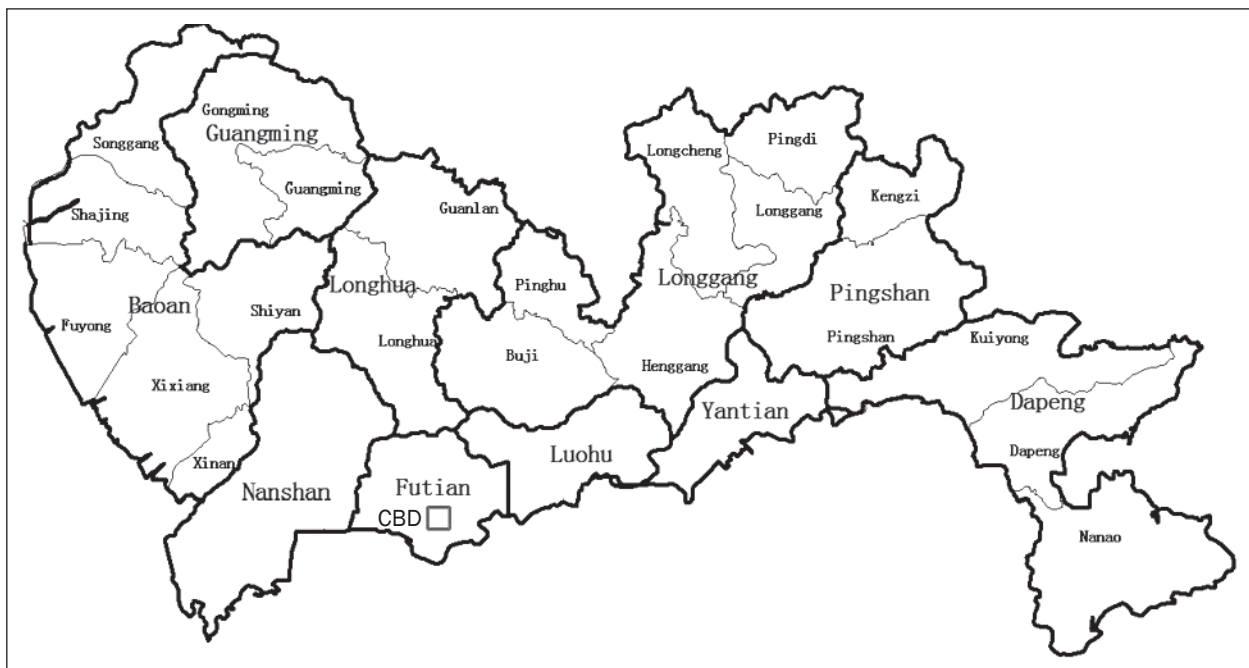


Figure 1 – Administrative division of Shenzhen city

4. SPATIAL DISTRIBUTIONS OF CRUISING TRACES AND PICK-UP POINTS

4.1 Entropy

By employing a parameter of entropy, which is a measure of the uncertainty in a random variable [23], the basic characteristics of spatial distribution of cruising traces and pick-up points can be examined. If we define a discrete random variable Z as the event that one point locates in a zone, then the entropy of Z , i.e. $H(Z)$, refers to the unpredictability of event Z :

$$H(Z) = -\sum_{i=1}^{25} \frac{n_i}{N} \ln\left(\frac{n_i}{N}\right) \quad (2)$$

in which, n_i is the number of points in zone i with $i=1, 2, \dots, 25$ (the study area is divided into 25 zones as shown in *Figure 1*), N is the total number of points, that is $N = \sum_{i=1}^{25} n_i$, then $\frac{n_i}{N}$ means the probability that one point is located in zone i , and $\sum_{i=1}^{25} \frac{n_i}{N} = 1$.

When all the points are located in one zone, we can get the minimum value of $H(Z)$, that is 0, which means that event Z is most predictable, or points are most clustered concerning spatial distribution. On the contrary, when all the points are located in 25 zones evenly, we can get the maximum value of entropy, defined as H_m , which means that event Z is most unpredictable, or points are most dispersed concerning spatial distribution. In this study, $H_m = \ln 25$, which is obtained from probability $\frac{n_i}{N} = \frac{1}{25}$ for each zone i . Then C can

be introduced to represent the cluster degree of points:

$$C = 1 - H(Z)/H_m \quad (3)$$

As we know, C is a value between 0 and 1. The higher the cluster degree C is, the more clustered the spatial distribution of points is. According to reference [24]' study, cluster degree >0.45 means that the data points are very clustered, and $0.4 < C \leq 0.45$ denotes that the data points are clustered, while $C \leq 0.4$ means that the points are dispersed.

4.2 Analysing spatial distribution with entropy

The mean value of cluster degree C of cruising traces and pick-up points per person per day is calculated to be 0.6080 and 0.6936, respectively. This indicates that both cruising traces and pick-up points exhibit a very clustered spatial distribution. As the cluster degree C of pick-up points is bigger than that of cruising traces, we can infer that pick-up points are more clustered than cruising traces in spatial distribution. One potential reason is that the taxis' pick-up points are also the start points of taxi trips, and the spatial

distribution of the start points is related to the spatial distribution of land use and population in the study area [25]. In the study area of this paper, i.e., Shenzhen city, the high-density area of land use and population is the area covered by Luohu and Futian district (shown in *Figure 1*). It is the clustered spatial distribution of land use and population that leads to the aggregated spatial distribution of taxis' pick-up points. On the contrary, the spatial distribution of cruising traces depends on taxi driver's individual cruising behaviour, which is much more a subjective factor, comparing with land use and population distribution which are objective factors. Being the results of subjective decision behaviours, cruising traces are then more disaggregated than pick-up points in spatial distribution.

4.3 Variability in multi-day cluster degrees

In order to identify the potential variability in multi-day cruising patterns, we will measure the changes among multi-day cluster degrees of cruising traces, and compare them with those of the pick-up points. As the GPS traces recorded the cruising traces of different days and different taxis, the total variability in GPS records contain both inter-daily variability and inter-personal variability. Therefore, in order to identify the variability among multi-day cruising behaviours, a distinction should be made between inter-daily variability and inter-personal variability. Reference [26] and [27] provided the theory of TSS (total sum of squares) to measure and quantify the inter-personal variability and intra-personal variability in a travel survey data set. *Figure 2* shows the framework of TSS as adopted from their work.

Inter-personal variability refers to the differences in the behaviour among different individuals on the same days. Behavioural differences among taxi drivers may be explained partially by differences in the characteristics of individuals. By incorporating such characteristics into a model, one can account for systematic differences in behaviour among individuals. The portion of inter-personal variability that can be explained systematically through differences in socio-economic characteristics is referred to as explained variability. The remainder is referred to as unexplained variability.

Similarly, intra-personal variability (inter-daily variability) may also be considered as having two components. The first component is called systematic day-of-week variability. This refers to the portion of intra-personal variability that may be attributed to systematic day-of-week effects. Intra-personal variability that cannot be explained by day-of-week effects is random and is referred to as residual intra-personal variability.

In this study, the total intra-personal (inter-daily) variability as well as the total inter-personal variability will be quantified by employing the TSS theory. The to-

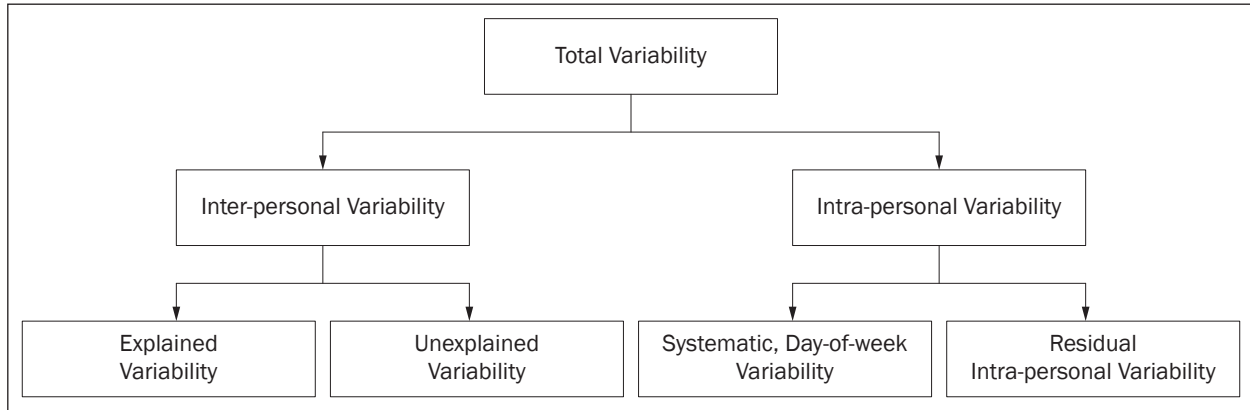


Figure 2 – Methodological framework for measuring inter-personal/intra-personal variability

tal variability in various measures of cruising patterns is split into two components which are represented by appropriate sums of squares. In this representation, the total variability is represented by the total sum of squares (TSS), as follows:

$$TSS = \sum_i \sum_j (C_{ij} - \bar{C})^2 \tag{4}$$

where C_{ij} is the cluster degree C of person i on day j , and \bar{C} is defined as the overall sample mean of C per person per day.

The inter-personal sum of squares (IPSS) is given by:

$$IPSS = \sum_i J_i (\bar{C}_i - \bar{C})^2 \tag{5}$$

where J_i is the number of days for which individual i reported GPS cruising trajectory, and \bar{C}_i means the cluster degree C of person i per day.

The inter-daily sum of squares (IDSS) is given by:

$$IDSS = \sum_i \sum_j (C_{ij} - \bar{C}_i)^2 \tag{6}$$

Also, we have:

$$\bar{C}_i = \frac{\sum_j C_{ij}}{J_i} \tag{7}$$

and

$$\bar{C} = \frac{\sum_i J_i \bar{C}_i}{\sum_i J_i} = \frac{\sum_i \sum_j C_{ij}}{\sum_i J_i} \tag{8}$$

It can be readily seen that,

$$TSS = IPSS + IDSS \tag{9}$$

Therefore, the ratio $IDSS/TSS$ provides a measure of the proportion of total variability in cruising patterns that may be attributed to inter-daily variability. Similarly, $IPSS/TSS$ provides a measure of the proportion of total variability in cruising patterns that may be attributed to inter-personal variability.

The calculating results of TSS parameters for cluster degrees of cruising points and pick-up points are shown in Table 1. The results indicate that, for both cruising traces and pick-up points, the value of IPSS is greater than that of IDSS: the inter-personal variability and inter-daily variability can explain 66.27% and 33.73% of the total variability in cluster degree of the cruising traces, respectively; Similarly, the inter-personal variability and inter-daily variability can explain 64.53% and 35.47% of the total variability in cluster degree of the pick-up points, respectively. Concerning the results of cruising traces, because IPSS accounts for 66.27% of the total variability it indicates that there is large variability between different taxi drivers. In other words, different types of taxi drivers may have different strategies for cruising location choice: some of them may cruise in a large area, while some of them may choose a relatively small area.

In order to explain the inter-daily variability in cruising traces and pick-up points, the mean values of cluster degree of different days per person are calculated. The results are shown in Table 2.

Table 1 – TSS parameters for cluster degrees

Item	Cluster degree of cruising traces		Cluster degree of pick-up points	
	Value	Ratio	Value	Ratio
TSS	73.81	-	52.65	-
IPSS	48.91	IPSS/TSS: 66.27%	33.98	IPSS/TSS :64.53%
IDSS	24.90	IDSS/TSS: 33.73%	18.67	IDSS/TSS: 35.47%

Table 2 – Multi-day cluster degrees

Day	Cluster degree of cruising traces		Cluster degree of pick-up points	
	Mean value \bar{C}_j	Percent of deviation $\frac{\bar{C}_j - \bar{C}}{\bar{C}} \times 100\%$	Mean value \bar{C}_j	Percent of deviation $\frac{\bar{C}_j - \bar{C}}{\bar{C}} \times 100\%$
April 18 th	0.6126	0.71%	0.6989	0.76%
April 19 th	0.6051	-0.53%	0.6896	-0.58%
April 20 th	0.6106	0.38%	0.6939	0.04%
April 21 st	0.6060	-0.38%	0.6907	-0.42%
April 22 nd	0.6020	-1.04%	0.6881	-0.79%
April 23 rd	0.6091	0.13%	0.6951	0.22%
April 24 th	0.6098	0.25%	0.7002	0.95%
April 25 th	0.6089	0.10%	0.6920	-0.23%
April 26 th	0.6102	0.31%	0.6939	0.04%
Mean value \bar{C}	0.6083	-	0.6936	-

The results indicate that there are two similar characteristics for the multi-day variability of both cruising traces and pick-up points' cluster degree: (1) the multi-day variability is small, i.e., the cluster degrees of different days are similar to each other; (2) there is a same day-to-day changing pattern, i.e., the cluster degree reaches the minimum on April 22nd (Friday). One of the reasons may be that on Friday passengers made taxi trips not only for the working purpose like other weekdays, but also for other trip purposes, such as entertainment and maintenance, like on weekend. This relative complicated composition of trip purposes on Friday leads to the disaggregated distribution of both cruising traces and pick-up points.

Concerning the maximum of cluster degree, the value on April 18th (Monday) is the maximal value in the multi-day cluster degrees of cruising traces. A potential reason is that most of the taxi trips on Monday are working trips, and taxi drivers may choose a "Monday cruising mode" according to their own cruising habits or land use characteristics of the city. It may be the relatively regular "Monday cruising mode" that makes the distribution of cruising traces more clustered than other days of the week. As for the pick-up points, the maximum of cluster degree is the value on April 24th (Sunday). A potential reason is that most of the taxi trips on Sunday are trips for entertainment or maintenance purposes starting from a high-density area, such as the Central Business District. Therefore, the pick-up points concentrate on the high-density areas and present an aggregated distribution.

5. LEARNING AND HABITUAL FEATURES OF CRUISING BEHAVIOUR

The learning feature means that taxi drivers remember their previous cruising experience, learn from it, infer potential pick-up locations from it, and apply

them in future cruising location choices. In this paper, the spatial correlation between current cruising traces and previous pick-up points $t_j \& p_{j-}$ is chosen to represent the learning feature. The footprint j denotes current day and $j-$ refers to previous day, t and p represent cruising traces and pick-up points, respectively.

Habits denote one's customary ways of behaving [21, 28]. In this paper, the habitual feature presents that taxi drivers will accumulate some cruising habits from day-to-day cruising experience and follow the cruising habits consciously or unconsciously in future cruising decisions. The similarity in spatial distribution between current cruising traces and previous cruising traces ($t_j \& p_{j-}$) is taken as the habitual feature in cruising pattern. A diagram of the learning feature and habitual feature is shown in Figure 3.

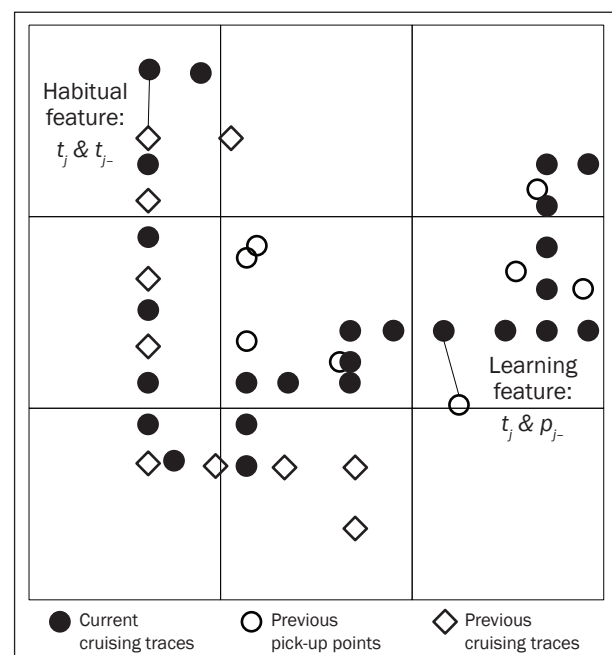


Figure 3 – A diagram of learning feature and habitual feature

As Figure 3 shows, the relationship between t_j and p_{j-1} , such as the distance between current cruising trace and previous pick-up point, is selected to represent the learning feature of cruising behaviour. Similarly, the relationship between t_j and t_{j-1} can be seen as taxi driver's habitual feature in cruising decision.

5.1 Analysing the distance between hot spots

One of the most direct approaches to measure spatial correlation between spatial points is to calculate the distance between them. To do this, we first identify the hot spots of cruising traces and pick-up points for each driver and each day, and then take the hot spots as representations of cruising traces and pick-up points, respectively, for distance calculation. According to the theory of hot spot analysis, we calculate the number of cruising traces and pick-up points for each taxi and each day in each cell of the study area. The centroid of the cell with the largest amount of cruising traces or pick-up points is noted as the hot spot of cruising traces or pick-up points, respectively. Then one taxi's one day's cruising traces and pick-up points can be represented by their calculated hot spot.

The distance between the hot spots of current cruising traces and previous pick-up points can be defined as $D(t_j, p_{j-1})$, and the distance between the hot spots of current cruising traces and previous cruising traces is noted as $D(t_j, t_{j-1})$. The mean value of $D(t_j, p_{j-1})$ and $D(t_j, t_{j-1})$ per driver is calculated as 6.94 km and 6.20 km, respectively. $j-1$ and j denote sequential two days in the 9-day period. This indicates that the distance between today's cruising traces and yesterday's cruising traces is smaller than that between today's cruising traces and yesterday's pick-up points. This reveals that, concerning the distance between hot spots,

habitual feature is a more important factor than learning feature in taxi drivers' cruising location choice.

In order to measure multi-day variability of learning feature and habitual feature, the theory of TSS mentioned above is adopted. In equations (4) to (8), the cluster degree related parameters can then be switched to distance related parameters which present distances among hot spots of multi-day cruising traces and pick-up points. The results of TSS analysis for distance between hot spots are shown in Table 3.

The IDSS and IPSS value of $D(t_j, p_{j-1})$ reveals that there are both inter-personal and inter-daily variabilities concerning taxi drivers' learning feature. However, the inter-personal variability is a little bit greater than the inter-daily variability. This reveals that different types of taxi drivers have different cruising behaviours concerning the learning feature. In other words, different taxi drivers are found to have different levels of ability to learn from previous pick-up experiences. Therefore, when measuring taxi drivers' learning feature, not only multi-day variability but also variability among different types of drivers should be considered.

The ratios IDSS/TSS and IPSS/TSS for $D(t_j, t_{j-1})$ show that the inter-daily variability accounts for 63.27% of total variability and the inter-personal variability covers 36.73% of total variability. This shows a larger variability of multi-day habitual feature than that of inter-personal habitual feature. In other words, it proves that a taxi driver has different cruising habits on different days of the week. The multi-day changes of cruising habits should be considered in cruising patterns analysis.

Therefore, further analysis of multi-day variability is conducted by calculating multi-day distance between t_j & t_{j-1} and t_j & p_{j-1} hot spots. The results are shown in Table 4.

Table 3 - TSS parameters for distances between hot spots

Item	$D(t_j, p_{j-1})$		$D(t_j, t_{j-1})$	
	Value	Ratio	Value	Ratio
TSS	3,081,451	-	2,832,022	-
IPSS	1,586,149	IPSS/TSS: 51.47%	1,040,226	IPSS/TSS: 36.73%
IDSS	1,495,303	IDSS/TSS: 48.53%	1,791,797	IDSS/TSS: 63.27%

Table 4 - Statistics of distance between multi-day hot spots

Time	$D(t_j, p_{j-1})$		$D(t_j, t_{j-1})$		
	Value (km)	Standard deviation	Value (km)	Standard deviation	
Mean value per day	6.94	10.79	6.20	10.34	
Day of week	Between two consecutive weekdays	7.45	8.65	6.50	7.64
	Sunday and Saturday	6.59	10.31	6.08	9.95
	Saturday and Friday	6.92	10.80	6.40	10.72
Between days of two consecutive weeks	Between Mondays	6.95	10.97	6.95	10.97
	Between Tuesdays	5.69	9.57	6.56	10.55

The smaller the distance, the larger is the correlation between the two types of spatial points. Concerning habitual feature, the results show that, for two consecutive days, the habitual feature between Sunday and Saturday is the largest, while that between two weekdays is the smallest. This indicates that the impacts of habits are more obvious between weekend days than among weekdays. The distances between t_j & t_{j-1} hot spots of two consecutive weeks reveal that there are also cruising habits between two consecutive weeks. However, the impacts of cruising habits of two weeks is smaller than those within one week.

The results reveal a similar pattern of inter-daily learning feature in one week as the one of the habitual feature, that is, the learning degree between Sunday and Saturday is the largest among learning degrees between two consecutive days in one week, while that between two weekdays is the smallest. This indicates that the impacts of learning feature are more obvious between weekend days than among weekdays. In other words, taxi drivers consider more yesterday's pick-up experience on Sunday than on other days of the week. The distances between t_j & p_{j-1} hot spots of two sequent weeks reveal that taxi drivers also accumulate cruising knowledge from pick-up points of last week. Moreover, the learning degrees between two consecutive weeks are greater than those within one week. This reveals that learning feature does not decay over time as habitual feature does.

5.2 Analysing frequency of the hot spots in the same cell

Besides distance between hot spots, the probability of t_j & p_{j-1} hot spots vs. t_j & t_{j-1} hot spots being in the same cell is also adopted to measure multi-day learning feature vs. habitual feature of cruising behaviour, respectively. In detail, the probability that the hot spots of current cruising traces and previous pick-up points are in the same cell is defined as $P(t_j, p_{j-1})$, and is taken as a parameter to measure the learning feature of the cruising behaviour. Similarly, the probability that the hot spots of current cruising traces and previous cruising traces are in the same cell is defined as $P(t_j, t_{j-1})$ in order to be adopted in examining habitual feature of cruising behaviour. $P(t_j, p_{j-1})$ and $P(t_j, t_{j-1})$ are given by:

$$P(t_j, p_{j-1}) = \frac{\sum_{i=1}^I \delta_i^{sp}}{I} \tag{10}$$

where δ_i^{sp} is an indicator parameter, $\delta_i^{sp}=1$ if driver i 's hot spots of cruising traces on day j (current) and pick-up points on day $j-$ (previous) are in a same cell, and $\delta_i^{sp}=0$ if their hot spots of cruising traces on day j and pick-up points on day $j-$ are not in the same cell. I is the total number of taxi drivers.

$$P(t_j, t_{j-1}) = \frac{\sum_{i=1}^I \delta_i^{tt}}{I} \tag{11}$$

where δ_i^{tt} is another indicator parameter, $\delta_i^{tt}=1$ if driver i 's hot spots of cruising traces on day j and cruising traces on day $j-$ are in the same cell, and $\delta_i^{tt}=0$ if the hot spots of cruising traces on day j and cruising traces on day $j-$ are not in the same cell.

The results (shown in Table 5) indicate that 41.69% of the taxi drivers chose the same cruising area (hot spot) according to previous day's cruising area, while only 13.75% of taxi drivers chose the same cruising area as previous day's pick-up hot spot. This indicates that taxi drivers prefer to obey cruising habits more than to learn from previous pick-up knowledge. This finding is similar to the result obtained from the hot spot analysis presented above.

The results also show similar multi-day variability of learning feature and habitual feature: the impacts of learning feature and habitual feature between Sunday and Saturday is the largest among the values between two consecutive days in one week, respectively, while that between two weekdays is the smallest. Moreover, the difference between learning feature and habitual feature is that learning feature between two consecutive weeks is greater than that within one week, while the habitual feature presents recession over time.

6. CONCLUSION

This study investigates taxi drivers' multi-day cruising behaviours by using GPS data collected in Shenzhen. Our results show that the inter-daily variability can account for 33.73% of the total variability in cluster degree of the cruising traces. This answers the

Table 5 – Probabilities of different hot spots being in the same cell

Time		$P(t_j, p_{j-1})$	$P(t_j, t_{j-1})$
Mean value per day		13.75	41.69
Day of week	Between two consecutive weekdays	13.34	40.44
	Sunday and Saturday	14.29	43.55
	Saturday and Friday	13.75	41.15
Between days of two consecutive weeks	Between Mondays	13.30	36.11
	Between Tuesdays	14.05	36.11

first question that we proposed in Section 1: there is variability between taxi drivers' day-to-day cruising behaviours.

By investigating the TSS statistics of cluster degrees of multi-day cruising traces, the second question proposed in Section 1 is answered. Our findings reveal that there are minor differences among cluster degrees of multi-day cruising traces. The cluster degree reaches the minimum on April 22nd (Friday) and the maximum on April 18th (Monday), which reveals that taxi drivers tend to cruise a larger area on Friday but a focused area on Monday.

To answer question 3, we conducted a comparison of the impacts of learning and habitual feature on day-to-day cruising traces. The results show that different types of taxi drivers obey the previous pick-up experience in different degrees. At the same time, a taxi driver has different degrees of cruising habits on different days of a week. This suggested that we should pay attention to the variability among different types of drivers concerning learning feature, while considering the day-of-week characteristic for the habitual feature of cruising behaviours. The results also show similar multi-day variability of learning feature and habitual feature: the impacts of the two features are more obvious between weekend days than among weekdays. However, learning feature between two consecutive weeks is greater than that within one week, while the habitual feature presents recession over time.

By revealing taxis' day-to-day cruising pattern and the factors influencing it, the study results provide crucial information in predicting taxis' multi-day cruising locations. In detail, the probability can be calculated that a driver will choose a location as today's cruising location by considering their cruising locations and pick-up points in the previous days. This finding can be applied to simulate taxis' multi-day cruising behaviour. Moreover, when making a traffic management strategy, we can also employ the study results in predicting taxis' adjustments of multi-day cruising locations under the impact of the strategy. This work also makes a contribution in determining the traffic volume derived from taxis' cruising behaviour in the study area. This finding is essential for planning of transportation facilities, like stop stations or parking lots for taxis.

It should be pointed out that only learning feature and habitual feature are considered in this paper. In reality, there may be other factors, such as last drop-off location, land use characteristics [29] and traffic condition [28, 30, 31], that have potential effects on taxis' cruising behaviour. Further study should be conducted to model the impacts of these factors. Besides, in order to get more information concerning multi-day cruising patterns, the GPS data of more days should be used in the future study.

ACKNOWLEDGEMENTS

The research is funded by the National Natural Science Foundation of China (50908099), the Humanity and Social Science Youth foundation of Ministry of Education (14YJC630225), the China Postdoctoral Science Special Foundation (2014M551191) and Jilin University outstanding youth fund (2013JQ007). Special thanks to the anonymous reviewers whose valuable comments helped tremendously to improve this paper.

出租车的多日巡航行为研究

宗芳¹, 孙晓², 张慧永³, 朱秀梅³, 祁文田⁴

¹吉林大学交通学院 长春 中国

²吉林大学应用技术学院 长春 中国

³吉林大学管理学院 长春 中国

⁴苏州建设交通高等职业技术学校 苏州 中国

本文应用深圳市的出租车车载GPS数据分析了出租车的多日的空载巡航行为。通过计算出租车各日巡航点空间分布特征值间的变化, 分析总结了巡航行为的多日变化规律。对比分析了司机的学习性和习惯性对巡航行为的影响。研究结果表明出租车逐日的巡航行为间具有一定差异: 他们倾向于在周五巡航更大的区域范围, 而周一的巡航区域比较集中; 另外, 周六和周日之间学习性和习惯性的影响较每周工作日之间的表现更明显; 同时, 两周之间的学习性比一周内的影响更大, 而习惯性的影响具有随时间增长而衰减的特征。本文的研究成果体现了出租车的多日巡航行为决策机理, 对于提出有效措施提高出租车的巡航效率具有一定借鉴意义。

关键词

出租车; 多日的; 巡航行为; GPS

REFERENCES

- [1] Empty cabs waste fuel, cause pollution. Shanghai Daily 2011 Feb 16; Available from: http://www.china.org.cn/environment/2011-02/16/content_21932450.htm
- [2] Li JZ. Optimal incentive for transportation management under a symmetric information. Journal of Chongqing Jiao Tong University (Natural Science). 2007;26(5):117-121.
- [3] Castro PS, Zhang D, Li S. Urban traffic modelling and prediction using large scale taxi GPS traces. Proceedings of the 10th International Conference on Pervasive Computing; 2012 June 18-22; Newcastle, UK. Springer-Verlag Berlin Heidelberg; 2012. p. 57-72.
- [4] Liu L, Andris C, Biderman A, Ratti C. Uncovering taxi driver's mobility intelligence through his trace. Seventh Annual IEEE International Conference on Pervasive Computing and Communications; 2009 March 9-13; Galveston, Texas.
- [5] Kim H, Oh J-S, Jayakrishnan R. Effect of Taxi Information System on Efficiency and Quality of Taxi Services. Transportation Research Record: Journal of the Transportation Research Board. 2005;1903:96-104.

- [6] Xu HL, Zhou J, Xu W. A decision-making rule for modeling travelers' route choice behavior based on cumulative prospect theory. *Transportation Research Part C*. 2011;19(2):218-228.
- [7] Li Q, Zeng Z, Zhang T, Li J, Wu Z. Path-finding through flexible hierarchical road networks: An experiential approach using taxi trajectory data. *International Journal of Applied Earth Observation and Geoinformation*. 2011;13(1):110-119.
- [8] Yu B, Yang ZZ, Chen K. Hybrid model for prediction of bus arrival times at next station. *Journal of Advanced Transportation*. 2010;44(3):193-204.
- [9] Yu B, Yang ZZ, Li S. Real-Time partway deadheading strategy based on transit service reliability assessment. *Transportation Research Part A*. 2012;46(8):1265-1279.
- [10] Yu B, Yang ZZ, Yao BZ. A hybrid algorithm for vehicle routing problem with time windows. *Expert Systems with Applications*. 2011;38(1):435-441.
- [11] Yao BZ, Hu P, Lu XH, Gao JJ, Zhang MH. Transit network design based on travel time reliability. *Transportation Research Part C*. 2014;43:233-248.
- [12] Cascetta E. A stochastic process approach to the analysis of temporal dynamics in transportation networks. *Transportation Research B*. 1989;23(1):1-17.
- [13] Cascetta E, Cantarella GE. A day-to-day and within day dynamic stochastic assignment model. *Transportation Research A*. 1991;25(5):277-291.
- [14] Horowitz JL. The stability of stochastic equilibrium in a two-link transportation network. *Transportation Research B*. 1984;18(1):13-28.
- [15] Mahmassani H, Chang G, Herman R. Individual decisions and collective effects in a simulated traffic systems. *Transportation Science*. 1986;21(2):258-271.
- [16] Hanson S, Huff JO. Assessing day-to-day variability in complex travel patterns. *Transportation Research Record: Journal of the Transportation Research Board*. 1982;891:18-24.
- [17] Jha M, Madanat S, Peeta S. Perception updating and day-to-day travel choice dynamics in trac networks with information provision. *Transportation Research Part C*, 1998;6:189-212.
- [18] Lerman S, Manski C. A model of the effect of information diffusion on travel. *Transportation science*. 1982;16(2):171-199.
- [19] Aarts H, Verplanken B, van Knippenberg A. Predicting behavior from actions in the past: Repeated decision-making or a matter of habit? *Journal of Applied Social Psychology*. 1998;28(15):1355-1374.
- [20] Bamberg S, Ajzen I, Schmidt P. Choice of travel mode in the theory of planned behavior: The roles of past behavior, habit, and reasoned action. *Basic and Applied Social Psychology*. 2003;25(3):175-187.
- [21] Ouellette JA, Wood W. Habit and intention in everyday life: the multiple processes by which past behavior predicts future behavior. *Psychological Bulletin*. 1998;124(1):54-74.
- [22] Benschhoof VA. Characteristics of drivers' route selection behaviour. *Traffic Engineering and Control*. 1970;11(12):604-606.
- [23] Duda RO. *Pattern Classification*, 2nd ed. American: John Wiley; 2003.
- [24] Ban X. Research on mutual relationship of land use structure and industrial structure [Master thesis]. China University of Geosciences; 2012.
- [25] Liu Y, Wang F, Xiao Y, Gao S. Urban land uses and traffic "source-sink areas": Evidence from GPS-enabled taxi data in Shanghai. *Landscape and Urban Planning*. 2012;106(1):73-87.
- [26] Pas EI. Intra-personal variability and model goodness-of-fit. *Transportation Research A*. 1987;21(6):431-438.
- [27] Pas EI, Sundar S. Intra-personal variability in daily urban travel behavior: Some additional evidence. *Transportation*. 1995;22(2):135-150.
- [28] Castro PS, Zhang D, Chen C, Li S, Pan G. From taxi GPS traces to social and community dynamics: A survey. *ACM Computing Surveys*. 2013;46(2):17-17.
- [29] Zong F. Understanding taxi driver's cruising pattern with GPS data. *Journal of Central South University of Technology*. 2014;21(8):3404-3410.
- [30] Salanova JM, Estrada MA, Mitsakis E, Stamos I. Agent based modeling for simulating taxi services. *Journal of Traffic and Logistics Engineering*. 2013;1(2):159-163.
- [31] Song ZQ, Tong CO. A simulation based dynamic model of taxi service. *Proceedings of the First International Symposium on Dynamic Traffic Assignment (DTA2006)*; 2006 June 21-23; Leeds, UK: Institute for Transport Studies; 2006. p. 355-360.