Research on Taxi Drivers' Passenger Hotspot Selecting Patterns Based on GPS Data: A Case Study in Wuhan

Liu Liqun
Intelligent Transportation Systems Research Center, Wuhan University of Technology
Engineering Research Center of Transportation Safety, Ministry of Education
Wuhan, China, 430063
E-mail: liuliqun@whut.edu.cn

Wu Chaozhong
Intelligent Transportation Systems Research Center, Wuhan University of Technology
Engineering Research Center of Transportation Safety, Ministry of Education
Wuhan, China, 430063
Email: wucz@whut.edu.cn

Zhang Hui*
Intelligent Transportation Systems Research Center, Wuhan University of Technology
Engineering Research Center of Transportation Safety, Ministry of Education
Wuhan, China, 430063
Email: zhanghuiits@qq.com

Naji Hasan.A.H
Intelligent Transportation Systems Research Center, Wuhan University of Technology
Engineering Research Center of Transportation Safety, Ministry of Education
Wuhan, China, 430063
Email: hasanye1985@gmail.com

Chu Wenhui
Intelligent Transportation Systems Research Center, Wuhan University of Technology
Engineering Research Center of Transportation Safety, Ministry of Education
Wuhan, China, 430063
E-mail: chuwh@whut.edu.cn

Atombo Charles
Intelligent Transportation Systems Research Center, Wuhan University of Technology
Wuhan, China, 430063
Email: atombocharles@gmail.com
Department of Mechanical Engineering, Faculty of Engineering, Ho Polytechnic
Ghana.
Email: catombo@hp.edu.gh

Abstract-**Taxis are the most important components for public transit systems. The model of taxis is more flexible and convenient for passengers compared to other transit model like buses or railway. However, the dynamic of taxi locations will have influence on transit efficiency. Therefore, the hotspot selecting patterns when unoccupied with regards to different income level drivers were investigated in this study. Two month GPS data of 7200 taxis in Wuhan was used as the data samples. The Wuhan City was firstly Mapped Meshing and divided into 4623 grids. After preprocessing the taxi GPS data and taxi drivers' income level classification (top, ordinary and bottom), the pickup points were filtered and matched with all these grids. Heat and grid probability for passenger demand hotspot were proposed and analyzed in this study. The results showed that the correlation value between top drivers and heat are not always higher than ordinary taxi drivers but the correlation value between grid probability and top taxi drivers are always high at all time slots. The finding of this study reveals that the high income taxis drivers have the high ability for cruising to the closed hotspot having high grid probability compared with middle or low income drivers. It therefore recommended that, if such experience could be represented and learned by other taxi drivers, the efficiency of taxi transit systems will be improved. Furthermore, if the information of real time heat and grid probability of hotspot could be broadcasted to taxi drivers, it will be beneficial for taxi transit systems as well.**

Keywords: Taxi Transit Systems, GPS trace data, passenger demand, hotspot selection pattern, data mining, map meshing

I. INTRODUCTION

Taxis play an important role in the urban transportation system. A huge number of people take taxi for work or do other things around the world every day. However, according to statistics, in 2015, the occupied ratio of Wuhan was less than 50% while passengers always faces problem of difficulty in taking taxi.. The major reason of this problem is that, taxi drivers can't locate the right place to find their passengers. Thus, how to select their next destination for finding passengers is a very significant problem.

With the introduction of advance transportation technology in recent years, the development of intelligent transportation system is increasing rapidly, showing significant benefits to both drivers and passengers in major cities in the world. In domain of travelling behavior, this technology has minimized the problem of drivers locating passengers thereby reducing waste and maximizing profits. Based on this, many researchers in the field of transportation are beginning to give more attention to the taxi equipped with GPS device *(1)*, This GPS can collects the real-time operation information about the taxi, including it location and whether the taxi is occupied by a passenger or not. Some related works have been conducted and used large-scale GPS to extract geographic knowledge for improving taxis' efficiency and have achieved effective results by reducing the driving distance before carrying passengers in the real-world taxi trajectories.

The first aspect of this study would focus on unban human mobility by using taxi GPS data *(2-9)*, which discover the Spatiotemporal dynamic of passengers. Other researchers classify the existing work into three main categories: social dynamics, traffic dynamics and operational dynamics and explore the night bus route planning issue by using taxi GPS traces *(2-3)*. Shuo Ma *et al. (4)(9)* proposed a scheduling algorithm to efficiently serve real-time requests sent by taxi users and generates ridesharing schedules for reducing the total travel distance significantly. Li Meng *et al. (5)* conducted study and focused on how *to* locate the crowded area and also find the crowded period by pointing out the hot area where passengers flow is huge based on the taxi speed. Xiaowei Hu, Hartwig H, Tang *et al. (6-8)* analyzed the urban taxi driver's activity distribution characteristics from different temporal and spatial level on weekday and weekends. More so, others have researched on recommender system for both taxi drivers and passengers wanting to take taxi using the knowledge of both passengers' mobility pattern and taxi driver's operations behavior from the GPS trajectories of taxicabs

The second aspect is to recommends trajectory for taxi drivers to guide them how to cruise on roads in order to pick-up passengers more efficiently*(10-15)*H Hu *et al.(10)* Recommend suitable routes choice to taxi drivers for minimizing the travel distance of without carrying passengers. Jing Yuan *et al. (11-13)* present a recommender for taxi drivers and people who expects to take a taxi by using the knowledge of passengers' mobility patterns. Xin Wang and Leyi Song *et al. (14)(15)* present a recommendation system separately for taxi drivers and passengers. The primary purposes of all these studies were to recommend methods by which taxi drivers and passengers can locate each other quickly and conveniently without maximizing the travelling distance for taxi drivers. Given that, recommending suitable route choice would effectively raise taxi driver's incomes and reduced fuel consumptions especially when the drivers' numbers increases.

The third aspect is to analyzed hotspots about pickup/drop off points *(16-21)*, to find and predict where and when the hotspots of pickup/drop off points are. Han-wen Chang *et al. (17)(18)* proposed mining historical data to predict demand distributions with respect to contexts of time, weather, and taxi location. Xiaowei Hu *et al.(16)* search activity space distribution and the relationship between the pick-up locations and the drop-off locations. Luis Moreira-Matias *et al. (20)* predicted the spatial distribution of taxi–passengers for a short-term time horizon using streaming data. Raghu Ganti *et al. (21)* busy "party" times in the city, most traveled streets, and travel patterns on holidays.

The fourth aspect is to discover taxi driver behavior to find how taxi drivers run on roads occupied and unoccupied (22-27), Daqing Zhang *et al. (22)(23)*, discovered the efficient and inefficient taxi service strategies based on a large-scale GPS historical database. F. Zong *et al. (24)* analyzed the spatiotemporal driving patterns for two income-level groups, i.e. high-income and low-income taxis, when they are vacant. Liang Liu *et al. (25)* discover the spatial hotspot selecting pattern, context-aware spatio-temporal operation behavior, route choice behavior and operation tactics based on large-scale taxi GPS data. Enjian Yao *et al. (26)* discovered the influence of road network conditions and traffic status among taxi drivers' route choice behaviors.

The selection of pickup points of taxi drivers directly influences the amount of time and distance taxi is occupied/vacant, resulting different income. Good selection for pickup points not only improves efficiency of taxi service system but also leads to high income. Thus, discovering the hotspot selecting patterns of taxi drivers can beneficial to the taxi service system, taxi drivers and passengers. Although there are researches carried out on taxi driver behavior, however, for taxi driver behavior, there is no analysis about difference of taxi drivers' hotspot selecting pattern for different grid probability.

In general, the aim of the present study is to, analyze and visualize the spatiotemporal hotspot selecting patterns of vacant taxis of three kinds of income-level groups, i.e. high-income, middle-income and low-income taxis. In order to discover the relationship among three kinds of income-level groups for finding passengers in different space-time, we need to address the following three challenges. The first issue is to extract the GPS traces from large-scale dataset. The second issue is to calculate the grid probability of different areas and analyze the relationship between the pickup points and grid probability. The last issue is to analyze the taxi driver's hotspot selecting patterns with different time based on the distribution of pickup points and grid probability.

The rest of this paper is organized as follows: in Section 2, the data preprocessing and extracting are described, then we divide the grids of Wuhan urban areas and defined two measurements as heat and grid probability for next analysis. Thereafter, we design a procedure of analysis to discover selecting patterns. Furthermore, we analyze the income of each driver, which is used to classify them as top, ordinary and bottom drivers. More so, the spatiotemporal distinction between heat and grid probability of Wuhan was analyzed. Finally, we analyze the hotspot selecting pattern of different income levels to reveal the relationship between the hotspot selecting pattern of three kinds of drivers by analyzing correlation between heat, grid probability and different income levels. Section 3 shows the results of this research. Section 4 focused on discussion and conclusion of the study.

II. METHODOLOGY

A. Data Preprocessing

The taxi GPS data is introduce in this section of the research. We had a large-scale taxi GPS data of more than 7271 taxis served at Wuhan, China for two months (Sep 2013 and

Oct 2013). In the taxi dataset, each taxi was equipped with a GPS device for recording real-time taxi information at a frequency of 1-4 times per minute, Including its id, time, longitude/latitude, driving speed, the passenger state (i.e., “occupied/vacant/not working/invalid”) and driving direction. In this work, four states of a taxi can be directly be recognized based on the value of “car status”, i.e. occupied with a “car status” value of 3, vacant with a “car status” value of 1, not working with a “car status” value of 0, and invalid with a “car status” value of 2. An example of raw GPS data is shown in Table 1.

Table I. AN EXAMPLE OF RAW GPS DATA

Id	Time	Latitude	Longitude	Speed	Direction	State
3175694081	20130901000010	30.577666	114.188066	0	123.97	1
3175821825	20130901125913	30.527366	114.354716	77.8849992	289.75	1
3176084737	20130901000010	0	0	0	0	2
3175717377	20130901000010	30.57535	114.277	0	38.15	1

There are more than 220 million records in the taxi dataset. In order to reduce the amount of calculation, we chose 14 days in Sep 2013 as our research data, the area of longitude [113.920000, 114.639987] and latitude [30.300000, 30.899905]. First, we delete the not working records and invalid records caused by the device errors and noises. For example, if the taxi states are all the same at one day, we will discard the taxi data of that day.

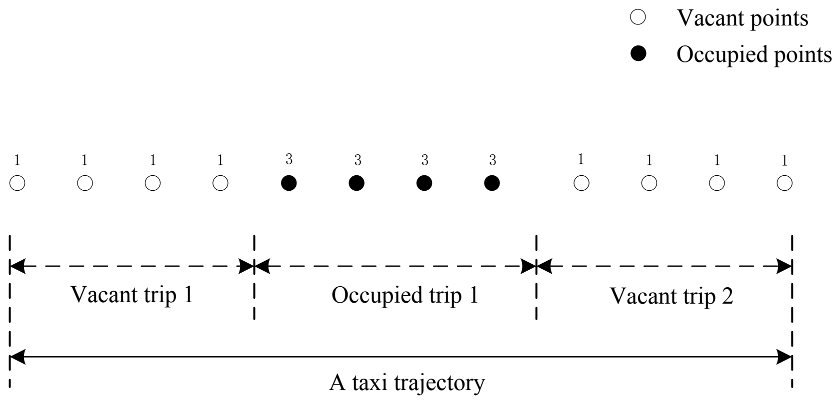


Fig.1. Taxi trajectory and taxi trip.

As shown in Figure 1, we extract the first continuous “3” as the pickup points, and the last continuous “3” as the drop off points. And to avoid taxi drivers’ errors behavior, we filter the pickup events if the number of continuous “3” is less than three. Besides, we also estimate the time slot for every record, dividing one day into 24 fields separately represented from 0 to 23. Finally, 4408 taxis are extracted, which is an example of information extracted from raw GPS records for pickup/drop off points as presented in Table 2.

Table II. AN EXAMPLE OF PICKUP/DROP OFF RECORDS

Taxi	Time	Pickup Points		Drop Off Points		Time Slot	Distance
		Longitude	*Latitude*	*Longitude*	*Latitude*		
2623745	20130907083358	30.56036	114.2187	30.58517	114.2182	8	3.8
2631681	20130907201432	30.60321	114.3013	30.58867	114.2895	20	6.5
3336449	20130907004941	30.54084	114.2954	30.55072	114.3287	0	2.1
3339009	20130907070748	30.72961	114.2488	30.71382	114.2712	7	10.8

B. *Map meshing and measurements*

As displayed in Figure 2, the Center Area of Wuhan was divided (longitude [113.920000, 114.639987], latitude [30.300000, 30.899905]) into 69×67 grids with equal intervals. Each grid cell was given a number (1-4623) for the later analysis. The grid cells were also connected to the places of Wuhan (i.e. Airport, Wuhan Railway Station and Han Kou Railway Station).

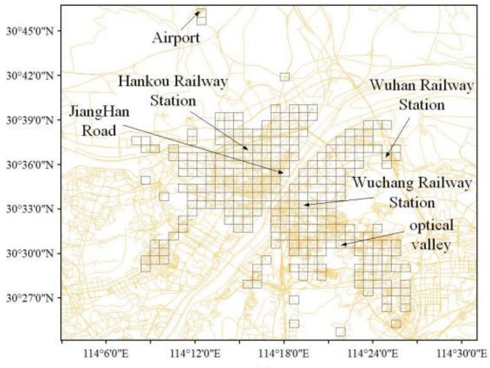


Fig.2. The schematic of grids connected with important regions

The heat in each grid cell will be defined as follows. The heat H of each grid cell i during a given time period $[t_1,t_2]$ is defined as Equation (1):

$$H(i,t_1,t_2)=\sum_{t_1}^{t_2} p \tag{1}$$

Where p is the number of pickup points in grid cell i at time slot t.

The grid probability of taking a passenger in a grid cell during a time period will be defined as follows. The grid probability of taking a passenger in a grid cell i during a given time period $[t_1,t_2]$ is defined as Equation (2):

$$r(i,t_1,t_2)=\begin{cases}\dfrac{p}{E}, & E\neq 0\\ 0, & E=0\end{cases} \tag{2}$$

Where p is the number of pickup points in grid i during the time period $[t_1,t_2]$ and E is the total number of taxis entering grid i except it taking passenger all the time through the grid i.

III. Analysis procedures

A. Correlation between heat and grid probability of hotspot

First, the means and the standard deviations of all grids' average heat and grid probability were calculated. And then, we estimate the heat and the grid probability during the extracted 14 days in each grid and selected the top 30 grids of pickup points and top 30 grids of grid probability for this research. We use the correlation coefficient between the pickup points P and grid probability R as

$$corr(P,R)=\frac{\sum_{i=1}^{n}(p_i-\bar{p})(r_i-\bar{r})}{\sqrt{\sum_{i=1}^{n}(p_i-\bar{p})^2}\sqrt{\sum_{i=1}^{n}(r_i-\bar{r})^2}} \tag{3}$$

Where p_i is the number of pickup points of grid i during given time period and r_i is the grid probability of grid i during given time period, n is the number of grids.

B. Taxi drivers' income level classification

The focus of this section is to discover the difference among different income level of drivers. Because there are usually three shifts in Wuhan, and not all the taxis work with the same time every day, so we consider analyzing the hourly income as the drivers' performance. Occupied trips can be built by the extracted pickup points and drop off points with temporal sequences of GPS points. We can calculate the hourly income of each taxi driver based on the distance of each occupied trip and the accounting rules of Wuhan in 2013. The income of each occupied trip can be calculated as

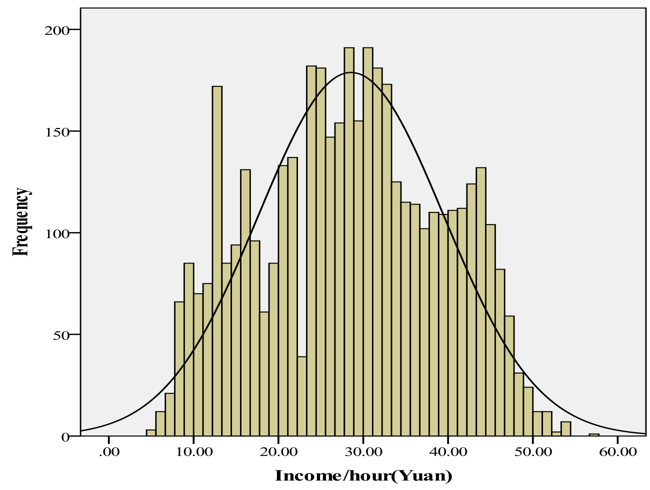
$$I=\begin{cases}8 & d\le 2\\ 8+1.5\times d & 2<d\le 7\\ 15.5+1.7\times d & d>7\end{cases} \tag{4}$$

Where I is the income of occupied trip, and d is the distance of occupied trip.

This paper divides the 600 highest-income drivers as top drivers, the 600 lowest-income drivers as bottom drivers, and others as ordinary drivers. We will analyze the difference of three kinds of drivers in the following procedure. Moreover, we compared the three kinds of drivers' average hourly income in the different time periods.

C. Income level classification

In Figure 3, the histogram represents the drivers' average hourly income of the 4408 taxis. The distribution of the hourly income indicates the income differences among the taxis. Most drivers' income are around 15-45 Yuan (approximately 2.5-7.5 US$), some are very high around 50 Yuan (approximately 8 US$), and some are very low around 10 Yuan (approximately 1.8 US$).



(a)

Fig.3. . The distribution of average hourly income of all drivers.

D. Spatial Correlation analysis between income level and hotspots

This part will compare the hotspot selecting patterns of three kinds of taxi drivers to discover the relationship between taxi drivers and the hotspots. We analyzed the hotspot selecting patterns in two parts. These are as follows: separate correlation between heat distribution and different income levels and the correlation between the grid probability and different income levels. First, we divided one day into five time slots according to the number of the pickup points of one day. Thereafter, we calculated the correlation values between heat and different income levels and the correlation between the grid probability and different income levels during the five time slots by using Pearson correlation.

IV. RESULTS

A. Spatiotemporal variation of heat and probability for hotspot

The heat and the grid probability hourly change are shown in Figure 4. The results indicated that, the peak of the pickup points happen at about 21:00pm and the bottom is about 4:00am. During 0:00-4:00h, the heat decreases at all time, because people rarely go out at that time. From the 4:00am to 7:00am, the heat increases rapidly because people begin to get up. From 7:00am to 9:00am, the heat keeps a value smoothly with morning peak. From noon, (12:00-14:00), the heat reaches the peak again. The result further shows that, during evening (17:00-21:00h), the heat reaches peak. With the grid probability, the changes are more smoothly than heat, which keep between 0.5-0.65 at all time slots. But at 4:00am and23:00 pm, reduction of the grid probability is the same as curve of heat, because less people travel at that time, and begin to increase at about 0.62 at 9:00 am and 18:00 am at the time more people begin to travel.

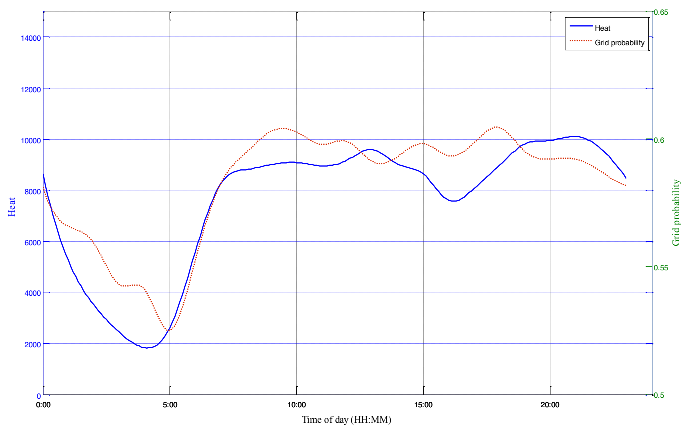
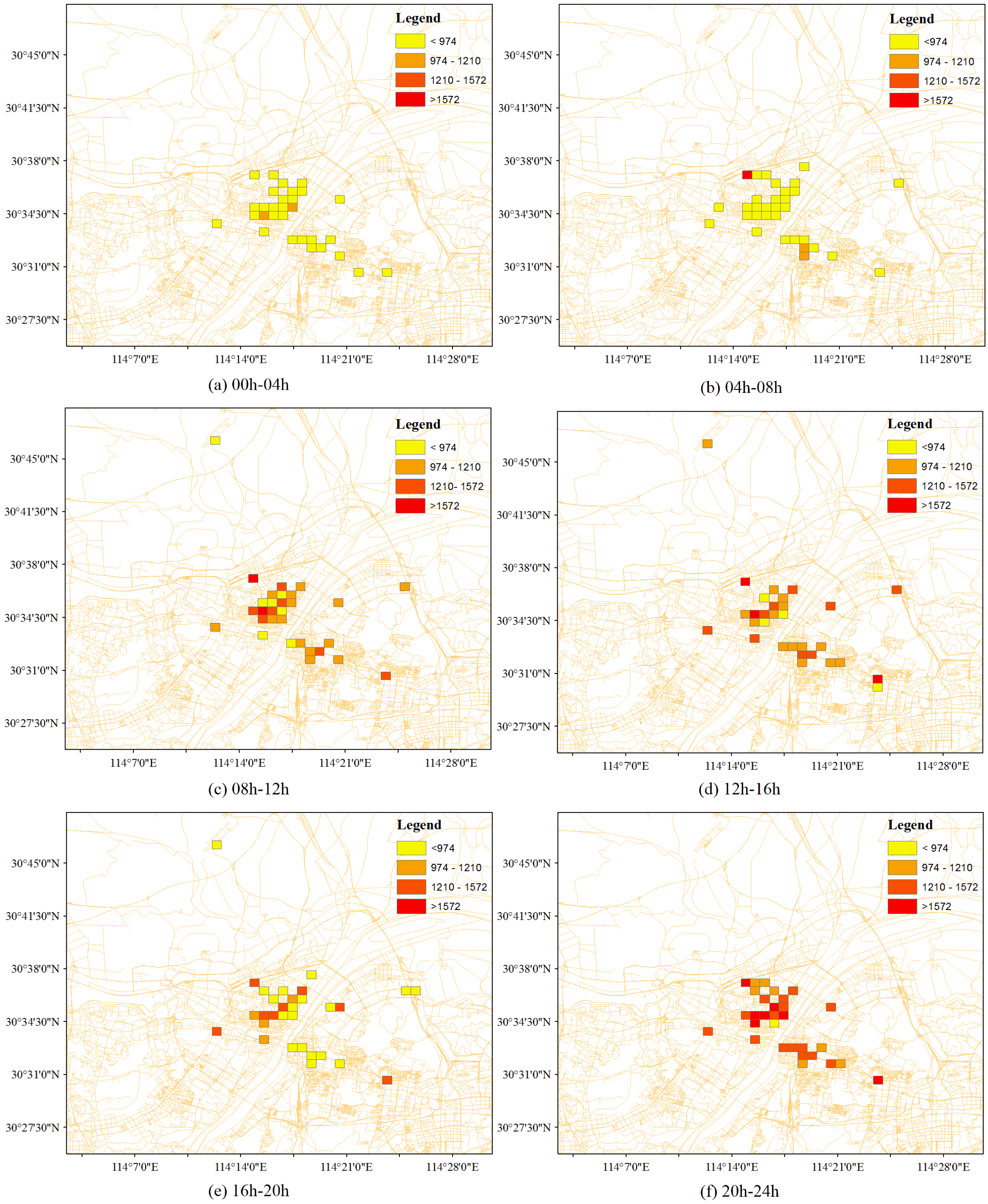


Fig.4. Average hourly heat and grid probability.

The top 30 grids heat and grids probability are shown in Figure 5 and Figure 6 respectively. The Figure 5 represents heat distributions of Wuhan unban area. As indicated, most top hotspots of heat are surrounding the Jianghan road, because it's one of the most active areas of economic. The Wuchang and Hankou train station are among top 30 at all time slots while the Wuhan train station is not in the top 30 during 00:00-04:00h and 20:00-24:00h. The possible reason to this finding could be that, because the Wuhan train station is of high-speed rail, people chose better time to get off when traveling. The airport isn't among top 30 grids during the 00:00-08:00h and 20:00-24:00h, the reason could be that, passengers are more like to travel through the airport in daytime since they can get home conveniently.



(a) 00h-04h

(b) 04h-08h

(c) 08h-12h

(d) 12h-16h

(e) 16h-20h

(f) 20h-24h

Fig.5. The distribution of heat based grids

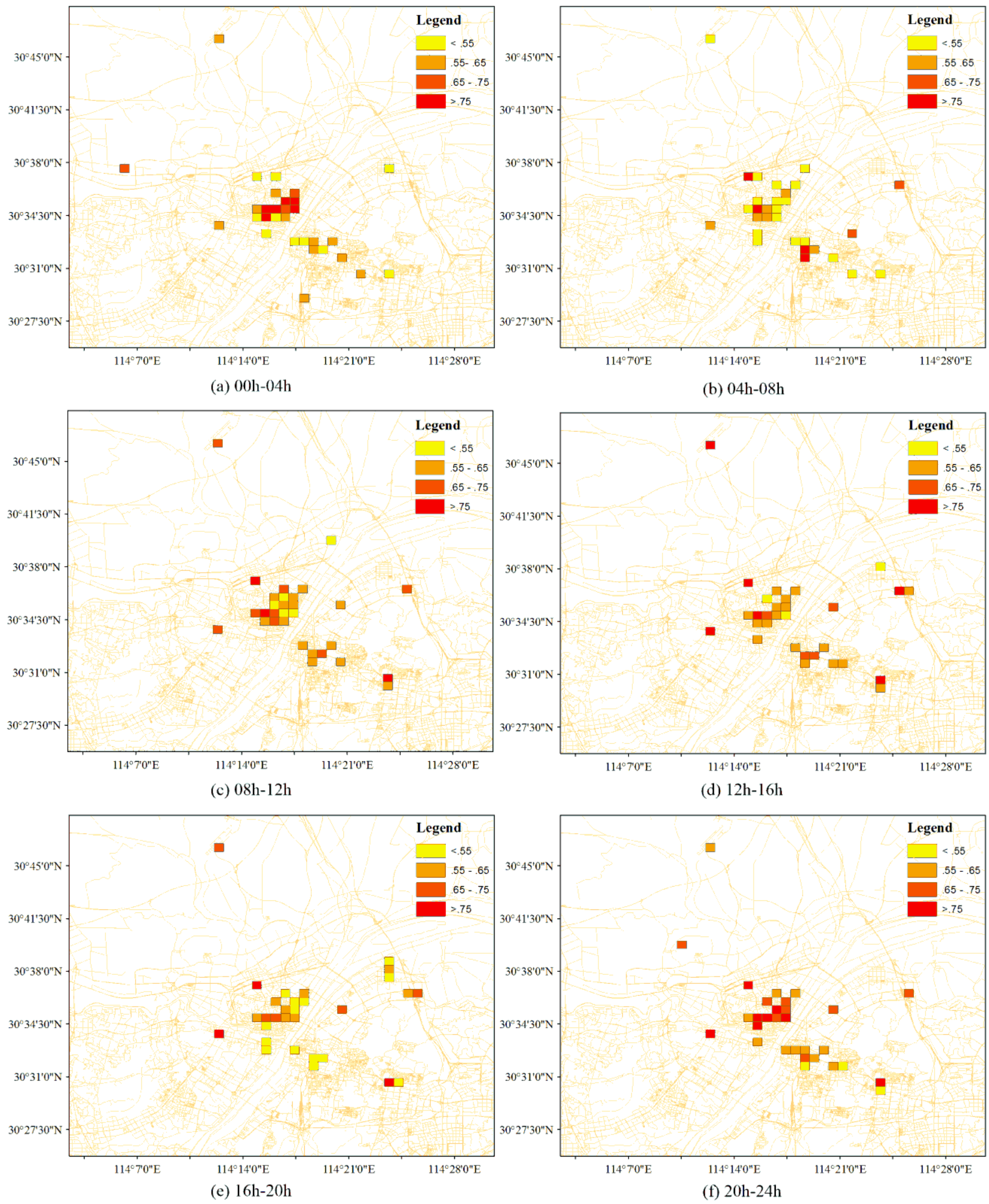


Fig.6. The distribution of grid probability based grids.

As shown in Figure 6, grid probability distribution demonstrated that, the top hotspots of grid probability is also around Jianghan road, Wuchang, Hankou, Wuhan train stations and airport are all among the top 30 grids during all day. The grid probability distribution is more dispersed than heat distribution. Based on these findings, we can say that the hotspots areas of heat are not always areas of high grid probability areas. The results of means, standard deviation and correlation values between distribution of heat and grid probability in all time slots are shown in Table 3 indicating that the correlation of 0:00-4:00h and 20:00-24:00h are more than 0.7, showing that the correlation of 0:00-4:00h and 20:00-24:00h are higher than other time slots. Comparing with Figure 4, the results suggest that the correlation is high when the average hourly heat is low. The reason may be that the traffic congestion of traveling peak decreases the grid probability of region of high heat.

Table III. CORRELATION BETWEEN HEAT AND GRID PROBABILITY

Time slot		0-4	4-8	8-12	12-16	16-20	20-24
Mean	Heat	13.153	14.436	29.392	29.172	27.715	30.554
	Grid probability	0.034	0.03732	0.0468	0.0467	0.044	0.044
Standard Deviation	Heat	70.000	68.263	129.510	132.252	117.132	143.739
	Grid probability	0.097	0.101	0.117	0.117	0.108	0.112
Corr		0.710	0.679	0.647	0.661	0.693	0.704

B. Spatiotemporal correlation analysis between income level and hotspots

Time slots of one day were divided into 0:00-6:59h, 7:00-10:59h, 11:00-12:59h, 13:00-16:59h, and 20:00-23:59h. The correlation value between income levels and the heat are shown in Fig 7(a), and the correlation value between income levels and the grid probability shown in Fig 7(b). The blue column represents the top drivers, the green column as ordinary drivers and the pink representing bottom drivers. The X axis of the histogram represents time and the Y axis of the histogram represents correlation value.

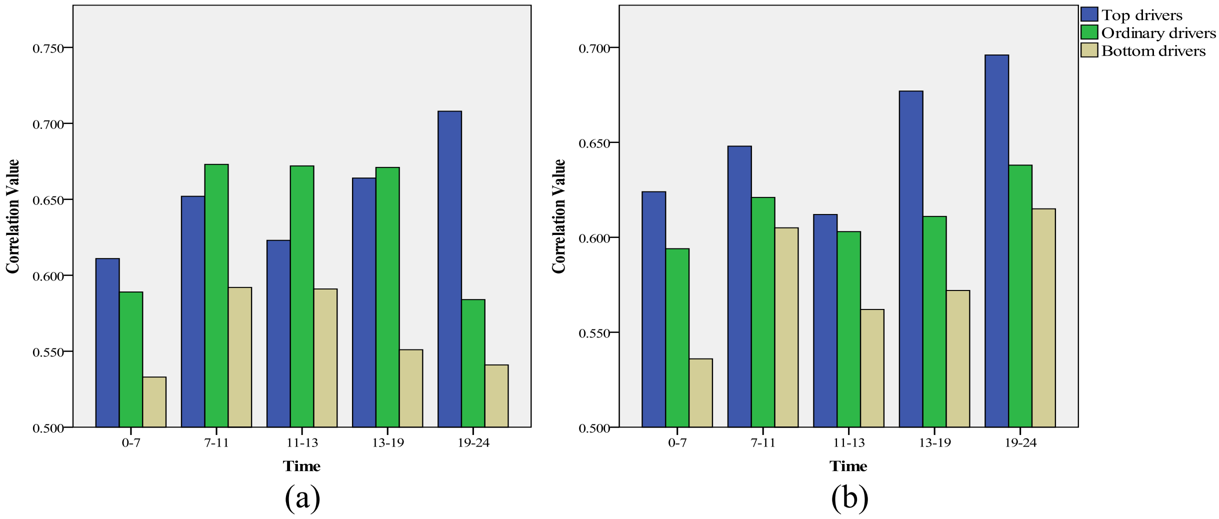


Fig.7. (a). Correlation between income levels and heat at five time slots. (b). Correlation between income levels and grid probability at five time slots.

From Figure 7(a), it clearly shown that, the correlation value between top drivers and heat are not always higher than ordinary taxi drivers. During 0:00-7:00h and 19:00-24:00h, the correlation value of top drivers is higher than other two kinds of drivers. While during 7:00-11:00h, 11:00-13:00h, and 13:00-19:00h the correlation value of top drivers are lower than ordinary drivers. It illustrates that the top drivers can select the region having more pickup points and more accurate than other drivers during the time slots of 0:00-7:00h and 19:00-24:00h, whiles selection for region having more pickup points is less accurate than ordinary drivers at 7:00-11:00h, 11:00-13:00h, and 13:00-19:00h.

Correlation value between grid probability and income level is shown in Figure 7(b). It shows that, the correlation value between grid probability and top taxi drivers are always highest at all time slots, and the ordinary drivers are always higher than bottom drivers at all time slots. It's clear that the ability of catching the high grid probability of top drivers is highest at all the time slots, the second is ordinary drivers, and the last is bottom drivers. According to Table 2, Figure 7(a) and Figure 7(b), we can conclude that the top drivers will consider the high heat hotspots when the traffic condition is good (i.e. 0:00-7:00h and 19:00-24:00h), and they will be prepared to select the high grid probability region when the traffic condition is pretty good.

V. CONCLUSIONS

Taxi GPS traces are valuable resources to discover taxi drivers' hotspot selecting patterns. In this paper we have investigated the taxi drivers' hotspot selecting patterns by using thousands of taxi drivers' trips extracted from 7271 taxis in Wuhan, aiming to discover the taxi drivers' hotspot selecting pattern. To understand the taxi drivers' patterns, we first extract the GPS trips of taxi drivers based on the taxis' status (vacant and occupied) and calculated the distance of each occupied trip. Second, we map meshing Wuhan urban areas and separate define heat and grid probability as two hotspot measurements for analyzing hotspots distribution. We further calculated the heat and the grid probability of each grid in different time slots, which can be used to analyze the relationship during all the time slots. Third, we calculated the hourly income of each taxi according to the distance of each occupied trip and the accounting rules, and divided all taxi drivers into three groups (high income, middle income and bottom income). Finally, we present the correlation between different income levels and heat, different income levels and grid probability to discover the hotspot selecting pattern of three kinds of taxi drivers.

In this paper, heat and grid probability are defined for building the different hotspot measurements about hotspots distribution, which is aim to discover the different selecting patterns between different income level of taxi drivers. The results indicate that, there are obvious differences between them. Comparing time based variation of heat and probability for hotspot, they are all get the lowest at about 4:00am and 5:00am, and heat reaching the peak at 21:00pm while grid probability decrease to peak at 9:00am and 18:00. On the other hand, for the spatial distribution, the grid probability distribution is more dispersed than heat distribution. Comparing the correlation, the results reveal that the correlation is high when the average hourly heat is low. The reason may be that the traffic congestion of traveling peak decreases the grid probability of region of high heat.

In order to discover the selecting patterns among different income levels, the correlation between income level and hotspots were analyzed separately. Results show that there are significant differences of taxi drivers' hotspot selecting pattern. The ability of catching the high grid probability of high income level taxi drivers are highest at all the time slots, the second is middle income taxi drivers, and the last is low income level taxi drivers. This suggests that, the high income level taxi drivers will consider the high heat region when the traffic condition is good (i.e. 0:00-7:00h and 19:00-24:00h), and they will be prepared to select the high grid probability region when the traffic condition is pretty good.

Clearly, the problem of analyzing the selecting patterns of taxi drivers is a very complex problem, impacted by a lot of parameters such as the last drop off point, the weather, the festival, weekends etc. In this research, we have used large-scale taxis GPS data; still, the data are just two months that are not comprehensive because the selecting patterns are different from summer to winter. This research also didn't consider the effects of weekends and weather. One contribution of this research is that the high-income selecting patterns are discovered, which could guide taxi drivers how to select the next destination and decrease time of vacant taxis traveling on the roads. In addition if such experience could be represented and learned by other taxi drivers, the efficiency of taxi transit systems will be improved. Furthermore, if the information of real time heat and grid probability of hotspot could be broadcasted to taxi drivers, it will be beneficial for taxi transit systems as well.

In the future we plan to broaden and deepen this research in three directions. First, we plan to analyze the hotspot selecting pattern during different weather (i.e. sunshine, shower, raining). Second, we plan to investigate the differences of hotspot selecting patterns between weekdays and weekends. Third, we plan to build a model to calculate the driver performance of each grid cell.

ACKNOWLEDGEMENT

This paper is supported by the National Nature Science Foundation of China (Analysis of Time-Based Fatigue Level Prediction Model for Full Driving Cycle Considering the Effect of Circadian Rhythm and Driving Workload), National Key Projects in the Science & Technology of China (2014BAG01B03, 2014BAG01B05), the Academic Leader Project of Wuhan (2012711304457) and the Fundamental Research Funds for the Central Universities (WUT: 2014-IV-137).

REFERENCE

[1] Huber, W., M, Lädke and R Ogger. Extended floating-car data for the acquisition of traffic information. *Plos One*, Vol. 5, 1999.

[2] Pablo, S. C., D. Zhang and C. Chen. From taxi GPS traces to social and community dynamics: A survey. *Acm Computing Surveys,* Vol. 46, No. 2, 2014, pp. 1167-1182.

[3] Chen, C., D. Zhang, N. Li and Z. H. Zhou. B-planner: planning bidirectional night bus routes using large-scale taxi gps traces. *IEEE Transactions on Intelligent Transportation Systems,* Vol. 15, No. 4, 2014, pp. 1451-1465.

[4] Ma, S., Y. Zheng, and O. Wolfson. T-share: A large-scale dynamic taxi ridesharing service. *Data Engineering, IEEE International Conference on IEEE*, 2013, pp. 410-421.

[5] Meng, L., R. T. Li, and X. Young. Analysis of Urban Traffic Based on Taxi GPS Data. *Advances in Computer Science and its Applications*, 2014, pp. 1007-1015.

[6] Hu, X. W., S. An, and J. Wang. Exploring Urban Taxi Drivers' Activity Distribution Based on GPS Data. *Mathematical Problems in Engineering,* No. 2, 2014, pp. 1-13.

[7] Hochmair, H. H. Spatio-Temporal Pattern Analysis of Taxi Trips in New York City. *In Data Applications in the Taxi Industry*. CD-ROM. Transportation Research Board of the National Academies, Washington, D.C., 2016, pp. 1-20.

[8] Tang, J. J., L. Fang, Z. B. Li. Exploring Intra-urban Travel Mobility from Taxi GPS Data. *In Recent Developments in Paratransit and Taxi Research*. CD-ROM. Transportation Research Board of the National Academies, Washington, D.C., 2016, pp. 1-10.

[9] Tao, C. C. Dynamic taxi-sharing service using intelligent transportation system technologies. *2007 International Conference on Wireless Communications, Networking and Mobile Computing*, 2007, pp. 3209-3212.

[10] Hu, H. R., Z. Wu, B. Mao, Y. Zhuang, J. Cao and J. Pan. Pick-Up Tree Based Route Recommendation from Taxi Trajectories. *Web-Age Information Management,* 2012, pp. 471-483.

[11] Yuan, N. J., Y. Zheng, L. Zhang and X. Xie. T-Finder: A Recommender System for Finding Passengers and Vacant Taxis. *IEEE Transactions on Knowledge & Data Engineering,* Vol. 25, No. 10, 2013, pp. 2390-2403.

[12] Yuan, J. Z., Y. Zhang, L. H. Xie and G. Z. Sun. Where to find my next passenger. *International Conference on Ubiquitous Computing*, 2011, pp. 109-118.

[13] Yuan, Yuan., J. Zheng, Y. C. Zhang, W. Xie, X. Xie and G. Sun. T-drive: driving directions based on taxi trajectories. *Sigspatial International Conference on Advances in Geographic Information Systems,* 2010, pp. 99-108.

[14] Yang, W., X. Wang, S. M. Rahimi and J. Luo. Recommending Profitable Taxi Travel Routes Based on Big Taxi Trajectories Data. *Advances in Knowledge Discovery and Data Mining. Springer International Publishing*, 2015, pp. 370-382.

[15] Song, L. TaxiHailer: A Situation-Specific Taxi Pick-Up Points Recommendation System. *International Conference on Database Systems for Advanced Applications,* 2014, pp. 523-526.

[16] Hu, X. W., S. An and J. Wang. Exploring Urban Taxi Drivers' Activity Distribution Based on GPS Data. *Mathematical Problems in Engineering*, 2014, No. 2, pp. 1-13.

[17] Chang, H. W., Y. C. Tai and Y. J. Hsu. Context-aware taxi demand hotspots prediction. *International Journal of Business Intelligence & Data Mining*, Vol. 5, No. 1, 2010, pp. 3-18.

[18] Chang, H. W., W. Tai and J. Hsu. iTaxi: Context-Aware Taxi Demand Hotspots Prediction Using Ontology and Data Mining Approaches. *International Journal of Business Intelligence and Data Mining,* Vol. 5, No. 1, 2009, pp. 3-18.

[19] Krisp, J. M., S. Peters, K. Polous and L. Meng. Getting in and out of a taxi: spatio-temporal hotspot analysis for floating taxi data in Shanghai. *Proceedings of Networks for Mobility*, 2012.

[20] Moreira-Matias, L., L. Gama and L. Damas. Predicting Taxi–Passenger Demand Using Streaming Data. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 14, No. 3, 2013, pp. 1393-1402.

[21] Ganti, R., I. Mohomed, R. Raghavenda and A. Ranganathan. Analysis of Data from a Taxi Cab Participatory Sensor Network. *Mobile and Ubiquitous Systems: Computing, Networking, and Services.* 2012.

[22] Zhang, D. Q., L. Sun, B. Li, C. Chen, G. Pan and S. Li. Understanding Taxi Service Strategies from Taxi GPS Traces. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 16, No. 1, 2015, pp.123-135.

[23] Ding, L. Fang, H. Fan and L. Meng. Understanding Taxi Driving Behaviors from Movement Data. *Agile,* 2015, pp. 219-234.

[24] Zong, F., X. Sun, H. Zhang, X. Zhu and W. Qi. Understanding Taxi Drivers' Multi-day Cruising Patterns. *Promet - Traffic – Traffico*, Vol. 27, No. 6, 2015.

[25] Liu, L., C. Andris and C. Ratti. Uncovering cabdrivers' behavior patterns from their digital traces. *Computers Environment & Urban Systems*, Vol. 34, No. 6, 2010, pp.541-548.

[26] Yao, E. J., L. Pan, Y. Yang and Y. Zhang. Taxi Driver's Route Choice Behavior Analysis Based on Floating Car Data. *Applied Mechanics & Materials,* Vol. 361, 2013, pp. 2036-2039.