Understanding taxi driving behaviors from movement data

Linfang Ding[1], Hongchao Fan[2], Liqiu Meng[1]

[1]Department of Cartography, Technical University of Munich, Germany

[2]Department of GIScience, Heidelberg University, Germany

Abstract. Understanding taxi mobility has significant social and economic impacts on the urban areas. The goal of this paper is to visualize and analyze the spatiotemporal driving patterns for two income-level groups, i.e. high-income and low-income taxis, when they are not occupied. Specifically, we differentiate the cruising and stationary states of non-occupied taxis and focus on the analysis of the mobility patterns of these two states. This work introduces an approach to detect the stationary spots from a large amount of non-occupied trajectory data. The visualization and analysis procedure comprises of mainly the visual analysis of the cruising trips and the stationary spots by integrating data mining and visualization techniques. Temporal patterns of the cruising trips and stationary spots of the two groups are compared based on the line charts and time graphs. A density-based spatial clustering approach is applied to cluster and aggregate the stationary spots. A variety of visualization methods, e.g. map, pie charts, and space-time cube views, are used to show the spatial and temporal distribution of the cruising centers and the clustered and aggregated stationary spots. The floating car data collected from about 2000 taxis in 47 days in Shanghai, China, is taken as the test dataset. The visual analytic results demonstrate that there are distinctive cruising and stationary driving behaviors between the high-income and low-income taxis.

Keywords. Taxi driving behavior, Mobility pattern, Movement data

1. Introduction

In modern society, advanced tracking technologies and devices, such as GPS receivers, mobile phones, have become increasingly pervasive, yielding massive movement data. This mobility data describes the changes of spatial positions of the mobile objects and provide the possibilities to investigate the urban dynamics. For instance, in intelligent transportation system, floating car data (FCD) collected at a relative low-cost can provide up-to-date high-quality traffic information (Huber et al. 1999). Some research works are dedicated to extracting geographic knowledge from movement data. Zheng and Zhou (2011) systematically investigated spatial trajectories from a wide spectrum of perspectives and disciplines, e.g. spatial database, mobile computing and data mining. Yuan et al. (2012) presented a recommender system for both taxi drivers and taker using the knowledge of both passengers' mobility patterns and taxi drivers' picking-up/dropping-off behaviors learned from the GPS trajectories of taxicabs. In human geography, Liu et al. (2012a, 2012b) and Yuan et al. (2012) examined large amounts of floating car data and mobile phone data respectively to understand human mobility patterns. In visual analytics, interactive visualization of movement data both at local scales focusing on individual trajectories (Guo et al. 2011, Tominski et al. 2012), or at large scales emphasizing on aggregated data (Andrienko and Andrienko 2011, Andrienko and Andrienko 2013) have been comprehensively studied to extract significant traffic mobility patterns.

The goal of this paper is to visualize and analyze the spatiotemporal mobility/driving patterns of non-occupied taxis of two income-level groups, i.e. high-income and low-income taxis, inferred from the average daily income that can be derived from floating car data. Intuitively, taxi driving behavior can be different when the taxi is vacant or occupied, or when the taxi driver is experienced or novice. When the taxi is occupied, the taxi driver usually finds out the fastest route to send passengers to a destination based on his knowledge (Yuan et al. 2013). On the contrary, when the taxi is vacant, the taxi driver has a large freedom to plan his/her routing to minimize his/her waiting time of the next trip. Generally speaking, experienced taxi drivers are more likely to pick up their next passengers quickly while novice drivers may cruise on the roads for a longer time. Therefore, different taxi driving behaviors may exhibit significant influence on the income of taxi drivers.

Previous studies on the driving behavior such as Liu et al. (2009) categorize drivers into top driver and ordinary driver by their average daily income and conduct spatiotemporal analysis of their operation behavior and skill (as measured by income) based on the operation zone. Our research aims to provide a deeper understanding of driving behaviors between low- and high-earning taxis utilizing visual and computational methods. We distinguish top and bottom taxis and not only look into their cruising traces but also take into consideration of their stationary spots. Furthermore, the time span of the spatiotemporal FCD is longer and the amount of taxis for analyzing are more than the existing approaches. In the presented approach we analyze the taxi-driving behaviors based on: (a) their incomes, (b) the spatiotemporal pattern of their cruising trips, and (c) the spatiotemporal pattern of their stationary spots. These analysis attempt to answer the questions such as what are the differences in terms of (1) the overall temporal patterns when they are cruising or stationary; (2) their spatial cruising distributions and (3) their stationary spatiotemporal characteristics.

The rest of this paper is organized as follows: in Section 2 the test dataset and preliminaries are described. In Section 3, we reconstruct the occupied trips and derive the taxi driver income, which is used to categorize taxis into high-income and low-income groups. Section 4 introduces an approach for the detection of the stationary spots. In section 5 and 6, we investigate in detail the spatiotemporal mobility patterns of the cruising trips and the stationary spots for the two income-level taxi groups. In section 7, we analyze and discuss the experiments of this paper. Finally section 8 concludes this paper.

2. Test dataset and preliminaries

The test dataset are temporally ordered position records collected from about 2000 GPS-enabled taxis within 47 days from 10th May to 30th June 2010, in Shanghai. The temporal resolution of the dataset is 10 seconds and thus theoretically around 8000 GPS points of each car would be recorded in one day (24 hours) given the GPS device effective. Each position record has nine attributes, i.e. car identification number, company name, current timestamp, current location (longitude, latitude), instantaneous velocity, and the GPS effectiveness. The detailed description of the fields is shown in Table 1.

In this work, two states of a taxi can be directly discerned based on the value of “car status”, i.e. occupied (O) with a “car status” value of 1, and non-occupied (N) of 0. Figure 1 illustrates the raw GPS points of a taxi with an identification number of 10003 on the 12th May 2010 with the occupied and unoccupied states color coded in blue and red. In addition, we consider two sub-states of the non-occupied state, namely cruising(C) and stationary (S) state. Stationary state of a taxi means a taxi is vacant and static and refers to the GPS points with car status of 0 and instantaneous velocity (denoted by the “Velocity” attribute) of 0, while cruising state means a taxi is moving without a passenger (see Table 2**).**

Table 1. Description of the fields of floating car data

Field	Example Field value	Field description
Date	20100517	8-digit number, yyyymmdd
Time	235903	6-digit number, HHMMSS
Company name	QS	2-digit letter
Car identifier	10003	5-digit number
Longitude	121.472038	Accurate to 6 decimal places, in degrees
Latitude	31.236135	Accurate to 6 decimal places, in degrees

Velocity	16.1	In km/h
Car status	1/0	1-occupied; 0-unoccupied
GPS effectiveness	1/0	1-GPS effective; 0-ineffective

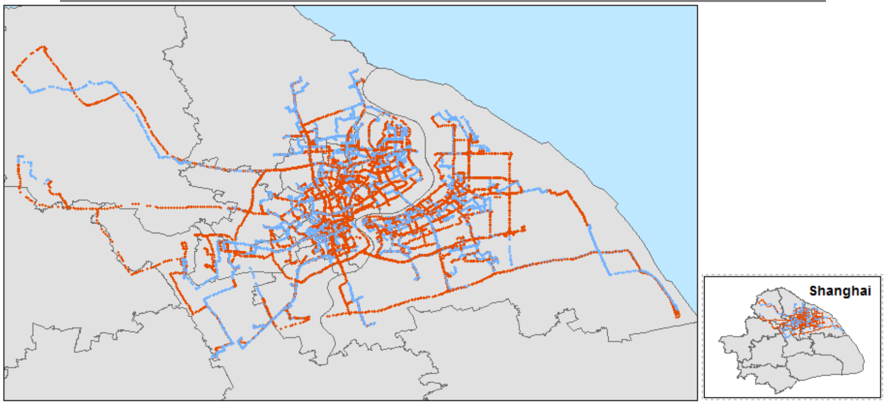


Fig. 1. GPS points of taxi with an ID of 10003 on the 12th May, 2010. Red and blue dots respective indicate GPS points of occupied and unoccupied state.

Table 2. Description of three states of a taxi

States	Sub States	Description
Occupied (O)		A taxi is occupied by a passenger
Non-occupied (N)	Cruising (C)	A taxi is moving without a passenger
	Stationary (S)	A taxi is static without a passenger

This work adopts similar definitions of taxi trajectory and trip as in Yuan et al. (2012) and defines a sequence of GPS points logged for a taxi as a taxi trajectory and a sub-trajectory with a single state as a taxi trip. According to the three states of a taxi defined in Table 2, this work introduces the occupied trips, cruising trips and stationary trips. Since a stationary trip comprises of a sequence of static GPS points at the same location and can be regarded as a stationary spot with associated stationary trip statistics, the following of this paper uses the term “stationary spot”.

3. Derivation of the income of taxi drivers

Occupied trips and non-occupied trips can be reconstructed by connecting the temporal sequences of GPS points with the “car status” attribute value of 1 and 0 respectively. A number of associated trip statistics, e.g., trip distance and duration, are also derived. Based on the occupied trip distance and the taxi fare system of Shanghai in 2010 shown in Table 3, the average daily income of each taxi driver can be calculated. Formula (1) is used to calculate the fare of an occupied trip.

Table 3. The taxi fare calculation system of Shanghai in 2010

	ITEM	Day timing (05:00am-23:00pm)	Night timing (23:00 pm -05:00am)
$\mathbf{P_0}$	Minimum fare for first 3Km	12 CNY	16 CNY
$\mathbf{P_3}$	Fare above minimum fare until 10Km	2.4 CNY/km	3.1 CNY/km
$\mathbf{P_{10}}$	Fare above 10Km	3.6 CNY/km	4.7 CNY/km

$$f(d) = P_0 + P_3 * Min(Max(d-3,0),7) + P_{10} * Max(d-10,0) \qquad (1)$$

where d is the distance of the occupied trip.

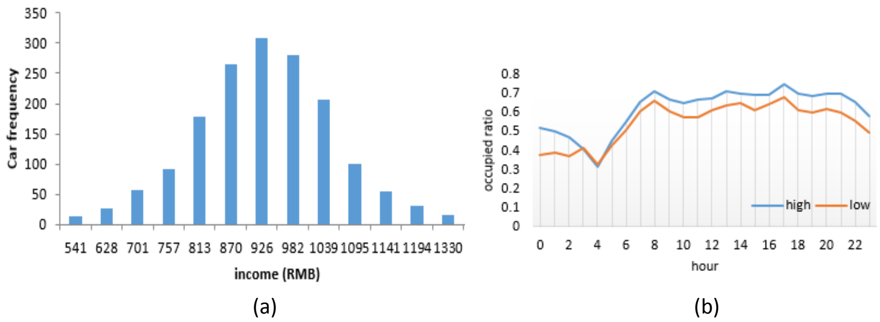
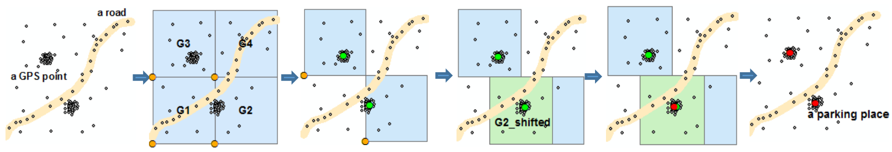


Fig. 2. (a) Average daily income distribution (b) Occupied ratio of high- and low-income taxis

In Figure 2a, the histogram represents the distribution of the average daily income of the taxi dataset (2000 taxis in 47 days). The normal distribution of the daily income indicates the income discrepancy among the taxis. This paper categorizes the 100 highest-income taxis as the top group and the 100 lowest-income taxis as the bottom group. The following sections of this paper discuss the spatiotemporal analysis based on these two groups. Moreover, we estimate the occupied ratios by utilizing the distance of the occupied and unoccupied trip with the formula $dist(trips_{occupied})/(dist(trips_{occupied}) + dist(trips_{unoccupied}))$. The line chart in Figure 2b illustrates the occupied ratios of the top and the bottom taxis. Basically, the occupied ratio of both groups has several peaks (e.g. at 8am) and valleys (e.g. at 4am) and the high-income taxi group generally has a much higher occupied ratio. However, in the early morning between 2am and 7am, the differences between the top and bottom taxi groups are much smaller than that in other time slots.

4. Detection of stationary spots and the cruising trips

This section details the process for detecting stationary spots from the non-occupied trips. A stationary spot is the location where a sequence of stop events occur and last for more than five minutes. A two-step approach is designed to detect the stationary spots. The specific workflow is illustrated in Figure 3 and described as follows:



(1a) (1b) (1c) (2a) (2b) (2c)

Fig. 3. The procedure of detecting the stationary spots. The black dots are stop events

First, we aggregate a sequence of static GPS points ($v = 0$) of a taxi to predefined grids (grid size is about 100m*100m) and calculate the centroids of the points inside each grid.

(1a) Extract raw GPS points from the dataset with cars of unoccupied state and with the instantaneous velocity of zero, and partition the study area into grids with the side length of l_g.

(1b) Assigning the extracted GPS points into the grids. In Figure 3 (1b), G1, G2, G3 and G4 are the grid divisions.

(1c) Calculate the duration and the count of the time-series GPS points inside each grid for each taxi. If a sequence of stop events in a grid lasts a long enough time period (e.g. more than 5mins) and are geographically close enough (e.g. $l_g/4$), we treat such sequence as a cluster. In Figure 3 (1c), the green dots in the grid G2 and G3 are two clusters of stop events.

Next, a refinement step is applied to the clusters, whose centroids are close to the boundary of the grid, by adjusting the surrounding grids to recalculate the clusters inside the new grid.

(2a) Move the grid by $l_g/2$ if the mean center inside a grid is close to its grid boundary. For instance, in Figure 3(2a) the green rectangle named $G2_{shifted}$ is a new grid.

(2b) Repeat step (1c) and get the cluster in the new grid. The red dot in Figure 3(2b) is the new mean center of $G2_{shifted}$.

(2c) Finally, the mean centers are identified as the stationary spots of the GPS points. In Figure 3 (2c), there are two stationary spots denoted by red dots.

The detection of stationary spots by the above procedure largely reduces the amount of GPS points. Note that the detected stationary spots are associated with temporal and thematic attributes (e.g. stationary duration, starting and ending timestamp) and they may refer to locations with different meanings such as parking lots, railway stations or hotels, or traffic congestion locations at road segments or intersections.

After the detection, the stationary spots can be excluded from the non-occupied trips and the rest of the GPS points are of the cruising state and can be used for reconstructing the cruising trips. Table 4 shows the numbers of the reconstructed occupied, cruising trips and the stationary spots analyzed in this paper.

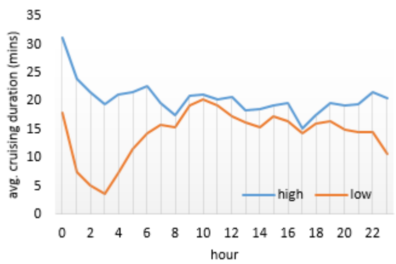
Table 4. The numbers of the occupied, cruising trips and the stationary spots.

	Top taxi group	Bottom taxi group
Occupied trips	217253	110753
Cruising trips	249600	141698
Stationary spots	30623	24263

5. Spatiotemporal patterns of the cruising trips

When a taxi driver is cruising, he or she plans to minimize his or her cruising time and intends to quickly pick up the next passenger. To investigate the temporal patterns of the cruising trips between top and bottom taxis, we calculate the average hourly cruising duration (shown in Figure 4) between the top and bottom taxi groups. The top taxis are normally cruising longer than the bottom taxis, especially in the time slots of midnight and early morning from 22pm to 6am. During the day, typically at 9am and 17pm, the differences of their cruising durations are relatively small.

To get an overview of the spatial distribution of the cruising trips, we examine the



distribution of the mean centers of the cruising trips.

Fig. 4. The average hourly cruising duration for top and bottom drivers.

Given a taxi trip $t = ((x_1, y_1), \dots, (x_n, y_n))$, its spatial mean center is

$$mc_{trip}\ (t) = (1/n \sum_{i=1,..n} x_i\, , 1/n \sum_{i=1,..n} y_i)$$

Consequently, given a set of taxi trips $T = (t_1, \dots\ t_m)$ in one day, the spatial mean center of the trips is

$$mc_{trip\_day} = \frac{1}{m} \sum_{t \in T} mc_{trip}(t)$$

Based on the above definitions, the average daily spatial mean centers of each car in the 47 days are calculated. Furthermore, to investigate the amount of variation or dispersion for each car, we also calculate the standard deviation of the daily spatial mean centers of each car.

Figure 5 shows the distribution of the average daily spatial mean centers (Figure 5(a, b)) of the cruising trips and their standard deviations (Figure 5(c, d)) of the bottom (Figure 5(a, c)) and top (Figure 5(b, d)) taxis. The rings in Figure 5 (a, b) are spatial partitions indicating the areas from the center of Shanghai to the peripheral. One can easily observe that the spatial distribution of the average daily cruise centers of the top taxi group is more compact and concentrated in the city center, while the spatial distribution of the daily cruise mean centers of the bottom taxi group is more dispersed. The orange ellipses in Figure 5 (c, d) indicate that the daily spatial mean center for the top taxi group is more centralized than those of the bottom taxi group. For the bottom taxis, the further they cruise from the city center, the more they spread from their daily cruising spatial mean centers.

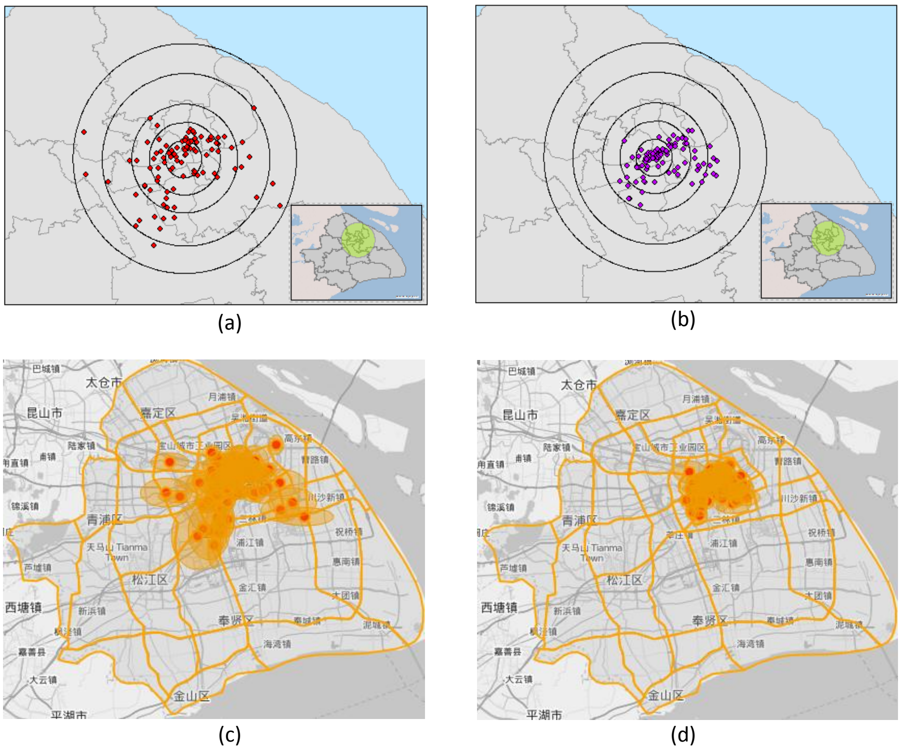


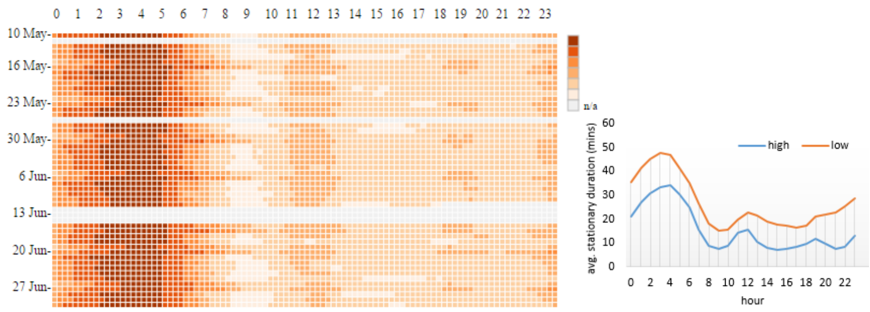
Fig. 5. The average daily spatial mean centers of the cruising trips (a, b) and their standard deviations (c, d) of the bottom (a, c) and top (b, d) taxi groups. In (a, b), dots represent the average daily spatial mean centers of each taxi and ellipses in (c, d) their standard deviations.

6. Spatiotemporal patterns of the stationary spots

Recall that in this paper, stationary spots are locations where taxis are static for more than five minutes and they may refer to parking lots, taxi stands or road segments where the traffic congestions happen. In this section, we investigate the spatiotemporal distributions of the stationary spots. Firstly Section 6.1 inspects the general temporal distribution of the stationary durations. Next Sections 6.2 and 6.3 study more detailed spatiotemporal patterns of stationary spots of long break and in off-peak hours.

6.1 General temporal patterns

To get a temporal overview of the stationary spots, we calculated the average stationary durations of all the 2000 taxis, divided into 15-minute time intervals. The average stationary duration is visualized in a time graph (see Figure 6 (a)). The rows represent the dates from 10th May to 30th June, and the columns represent every 15-minute in 24 hours. The color scheme is chosen from the ColorBrewer system. The darker the individual block is, the longer the taxis remain static. The white color means missing data (11th May, 4 hours on 11th June, and 12th – 14th June).



(a) (b)

Fig. 6. (a) The time graph of the stationary durations aggregated into 15-minute intervals (b) the average hourly duration at the stationary spots for top and bottom drivers.

The time graph in Figure 6a shows significant daily and weekly patterns. The daily pattern is illustrated in each row where the dark red colors occur at the early morning (1:00am - 6:00am), noon (especially from 11:30am - 13:00pm), and in a short period at evening (around 19:00 pm). These time slots can be interpreted as the sleeping, lunch and dinner periods respectively.

The weekly pattern is that for every seven rows the dark red color occurrences shifted about two or three hour's right, which means that stationary durations at weekends are about 1-2 hours longer and later than that at weekdays. Interestingly, one exceptional weekend pattern can be found on 15th and 16th June (Tuesday and Wednesday). By looking up the Chinese calendar of 2010, we found that 14th – 16th June (unfortunately the data on 14th June are missing) is the 3-day national holiday for the Duanwu Festival.

To study the difference of the top and bottom taxi groups, we also inspected the time graphs of the stationary duration for them and found similar daily and weekly patterns. Next, we created the line graph of Figure 6b which shows the average hourly stationary duraion of the top and bottom drivers. Unsurprisingly, the bottom taxi drivers stay at one places longer than the top taxi drivers.

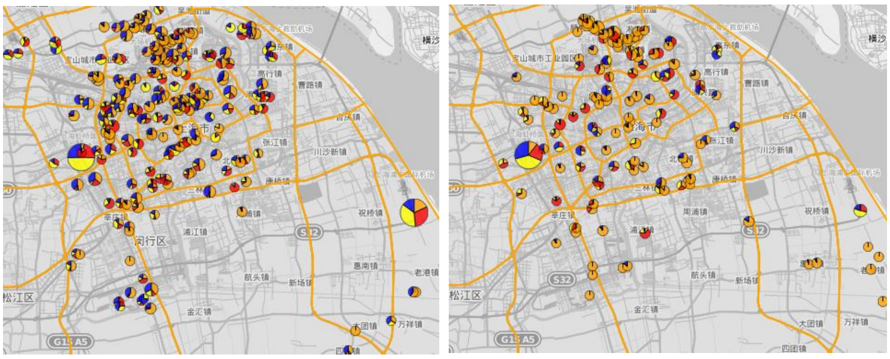
6.3 Spatiotemporal patterns of long break stationary spots

In this subsection, we investigate the spatiotemporal distribution of stationary spots with relatively long durations, which are normally breaks or long waiting periods. For instance, during the night, such stationary spots are mostly related to sleeping, and during the day, they may imply long waiting in a queue. It is important to know the spatiotemporal distribution of such events since they may have significant impacts on the taxis' overall income.

Here, we extract and examine the stationary spots with the duration more than 30 minutes. A density-based spatial clustering approach is applied to cluster these stationary spots. Specifically, DBSCAN (density-based spatial clustering of applications with noise) algorithm (Ester et.al 1996) is applied to detect clusters of arbitray shapes, with the distance parameter set to 20m and the minum point parameter to 10. The number of the extracted spatial clusters of the bottom and top taxi groups are 216 and 137 respectively.

Furthurmore, we divided one day into four time intervals, namely midnight (12am - 06am), morning (06am -12pm), afternoon (12pm - 18pm) and evenning (18pm - 12am) and calculate in each cluster the number of cluster elements (stationary spots) in each interval. Pie chart maps are used to visualize the spatial and temporal distribution of the clusters. The location of each pie chart represents a cluster centoid; the size of the pie chart is proporional to the total number of elements in each cluster; and the colors (i.e. orange, red, yellow and blue) of the pie chart sectors correspond to the four time intervals and these sectors are proportionally sized to the respective number of the cluster elements in each time interval.

The two screeshots in Figure 7 show the spatiotemporal distribution of the cluster centroids of stationary spots with the duration more than 30 mins. There are two relativly large pie charts in both graphs, which means that both top and bottom taxis often stay at these two places. These two pie charts have large proportions of afternoon and evenning. Actually, they are located at the two transport hubs of Hongqiao (left) and Pudong international airport (right). We also notice that the cluster at Pudong for top taxi groups is much smaller than that for the bottom taxi group. For both top and bottom taxi groups, the relative small pie charts are mostly in orange and correspond to the midnight while for the bottom taxi groups there are more blue portions (Figure 7a) and refers to more stationary spots during the evening interval (18pm - 24pm). We can interpret that the small circles with a large midnight and evening portions may correspond to their long-break and familiar places, e.g. the locations of the taxi companies or taxi drivers' home.



(a) (b)

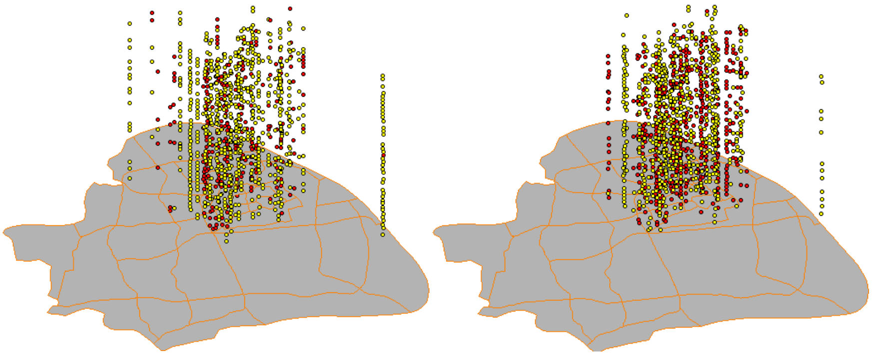
Fig. 7. Comparison of the spatiotemporal distribution of the cluster centroids of the stationary spots between the bottom (a) and top (b) taxi drivers. The predefined four time slots are color coded in orange, red, yellow and blue; the number of stationary spots in the each cluster is proportionally sized to the size of each pie.

6.3 Spatiotemporal patterns of stationary spots in off-peak hours

We noticed in Figure 2b there is a large discrepcany of the occupied ratios between the top and bottom driver groups from 10am to 12pm. This off-peak-hour occupied ratio difference might result from distinct driving behaviors and is interesting to get an insight.

To understand the spatiotemporal patterns of the stationary spots from 10am to 12pm, we extract from the detected stationary spots about 1100 and 1500 stationary spots during this time interval for the bottom and top taxi groups respectively.

Until now, we did not differentiate the meanings of the stationary spots. Here, we classify the stationary spots from 10am to12pm to traffic congestion and parking places. To identify the traffic congestion clusters, we utilize the road networks to calculate the distance from the clustered stationary spots to their nearest roads. A threshold (e.g. 20 m) is adopted between a traffic congestion place (<20m) and a parking place (>=20m). Figure 8 shows the space-time cube of the stationary spots of the bottom and top taxi groups from 10am to12pm in the 47 days. Red dots represent the traffic congestion places and yellow dots the parking places. From Figure 8, we can see that there are obviously more parking events indicated by the yellow dots from the bottom taxi group (Figure 8a), while there are more congestion events denoted by the red dots from the top taxi group (Figure 8b).

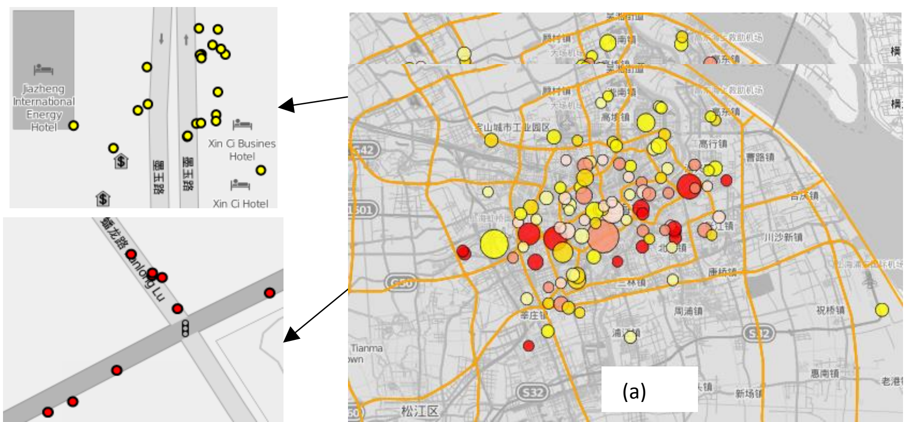


(a) (b)

Fig. 8. Comparison of the spatiotemporal distribution of the stationary spots with space-time cube between the bottom (a) and top (b) taxi groups from 10am to12pm in the 47 days. Red dots represent the traffic congestion places and yellow dots the parking places.

To compensate the occlusion effect the space-time cube, a pie chart map of the cluster centroids are inspected (shown in Figure 9). Similarly, we apply the DBSCAN clustering method and get 95 and 102 clusters after the clustering method. Each pie chart corresponds to a cluster centroid. If more than half of the elements in a cluster are traffic congestion, the cluster centroid is regarded as a traffic congestion place and coded in red; otherwise it is a parking place and coded in yellow. The color values correspond to the average hourly stationary duration in the cluster. The size of the pie charts is proportional to the number of elements in clusters. There are about 33 cluster centroids representing traffic congestions for the bottom taxis and 47 for the top taxis. We can see from Figure 9 that bottom taxi group are more likely go eastward to the airport “Pudong” at this time interval with a longer parking time.

Figure 10 shows the average stationary duration (in minutes) spent on the traffic congestion and the parking places (waiting for passengers) for top and bottom taxi groups. The average time spend on traffic congestion is more or less the same, while the parking time from 10am to12pm for the bottom taxi group are much longer than the top taxi group.



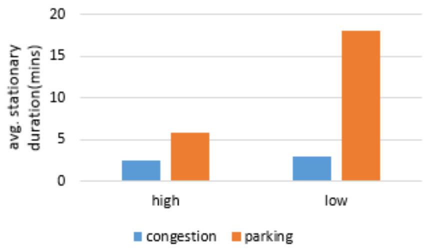


Fig. 9. Comparison of the spatiotemporal distribution of the parking and traffic congestion places between the bottom (a) and top (b) taxi groups. The parking places are in yellow and the congestion places in red. The color values correspond to the average duration at the places. The size of the circle is proportional to the count of the cluster elements of parking or traffic congestion.

Fig. 10. The average stationary duration spent on the traffic congestion and parking places (waiting for passengers) for top and bottom taxi groups.

(b)

7. Analysis and discussion

The above sections studied in detail the spatial and temporal distribution of the trajectory traces of top and bottom taxi groups, which reflects their distinctive driving behaviors. Now we try to answer the questions proposed in the introduction for top and bottom taxi groups.

(1) Differences of the overall temporal patterns when they are cruising or stationary. The trend of the occupied ratios between the top and bottom taxi groups (Figure 2b) reveals that the top taxi group drives longer distance with passengers than the bottom group. In spite of their longer occupied trips, in terms of cruising durations, the top taxi group normally cruises longer than the bottom taxi group especially in the evening and during the midnight (see Figure 4), which might reduce the profit via the cost of gas consumptions. However, in terms of stationary durations, top taxis have constantly shorter average durations (see Figure 6b), which might in turn compensate the loss of longer cruising. In addition, we also found the daily and weekly routines of all the taxis by investigating their stationary durations (see Figure 6a). The daily patterns show that long breaks often occur in the

midnight and early morning, lunch or dinnertime. The weekly pattern shows obvious differences between weekdays and weekends.

(2) Differences of their spatial cruising distributions. In terms of spatial driving behaviors, the cruising patterns of the top taxi group are more compact and concentrated in the city center, as shown in Figure 5, while the spatial distribution of the bottom taxi group is more dispersed and the further they cruise from the city center, the more they spread. Moreover, comparing the cruising patterns in Figure 5 with the spatial distribution of the long break stationary spots, we can observe that there are no obvious relationships of the cruising mean centers and the long-parking places around midnight, and thus can interpret that taxis might not cruise around their long-parking or long-break places.

(3) Differences of their stationary spatiotemporal characteristics. We observe from the occupied ratio plot (see Figure 2b) that there are relatively large differences during off-peak hours, which might be the main reason that result in the income differences. Thus we study in detail the spatial and temporal distributions during the off-peak time intervals, especially from 10am to 12 pm, and found that the top and bottom taxi group spend similar quantity of time on the road congestions but significantly distinct quantity of time on the parking places (see Figure 10). One reason might be that bottom drivers wait much longer in some places, for instance, in Figure 9, we can easily see that there are a large number of bottom taxis (Figure 9b) in the Pudong airport waiting for a relative long time period.

8. Conclusion

In this paper, we investigated the spatio-temporal patterns of the driving behaviors between low- and high-income taxis from large amount of floating car data in Shanghai. This work firstly differentiates two income-level driver groups derived from the GPS-enabled taxi trajectory data. An approach is designed to detect the stationary spots from massive trajectory data. To investigate the cruising patterns, we calculate and visualize the average daily mean centers of the cruising trips as well as the variation of the daily cruising mean centers. With regard to the stationary spots, we design a time graph to show the temporal patterns based on the aggregated durations of the stationary spots. Based on the clustering result and a distance method, parking places and traffic congestion are distinguished. A space-time cube is used to show the clustered spatiotemporal pattern of the traffic congestions and the parking places and a pie chart map to reveal the aggregated cluster centroid.

Preliminary results show that there are obvious driving behavior differences between the low-income and high-income taxis. The top taxi group mostly cruises in the city center while the bottom taxi group tends to cruise in the peripheral areas of Shanghai. The evidences can be found by the spatial distribution of their respective cruising mean centers and the spatial variation of the daily cruising mean centers. With regard to the stationary state, both top and bottom taxi groups exhibit a similar daily and weekly temporal pattern. However, compared to the top taxi group, the bottom taxi group generally has a much longer waiting time.

This paper utilizes various visualization and computational methods to discover knowledge hidden in the massive floating car data. For instance, a grid-based approach is designed for the detection of the stationary places and a density-based clustering approach for clustering the detected stationary places. Line charts and time graphs are used to reveal the temporal patterns; pie chart maps and space-time cubes are applied to reveal the spatial patterns. In the future, further visualization options will be experimented and tested with target users.

Acknowledgements. This work is partially funded by the China Scholarship Council. We would like to thank Prof. Chun Liu for sharing with us the Shanghai Taxi FCD dataset.

References

Andrienko, N., Andrienko, G. (2011) Spatial Generalization and Aggregation of Massive Movement Data, IEEE Transactions on Visualization and Computer Graphics (TVCG), 2011, v.17 (2), pp.205-219

Andrienko, G., Andrienko, N., Hurter, C., Rinzivillo, S., Wrobel, S. (2013) Scalable Analysis of Movement Data for Extracting and Exploring Significant Places , *IEEE Transactions on Visualization and Computer Graphics*, 2013, v. 19(7), pp.1078-1094

Ester, M., Kriegel, H., Sander, J., Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In Simoudis, Evangelos; Han, Jiawei; Fayyad, Usama M. *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*. AAAI Press. pp. 226–231. ISBN 1-57735-004-9.

Guo, H., Wang, Z., Yu, B., Zhao, H., Yuan, X. (2011) TripVista: Triple Perspective Visual Trajectory Analytics and Its Application on Microscopic Traffic Data at a Road Intersection. *In Proceedings of IEEE Pacific Visualization Symposium (PacificVis 2011),* pages 163-170, Hong Kong, March. 1-4, 2011.

Huber, W., Lädke, M., & Ogger, R. (1999). Extended floating-car data for the acquisition of traffic information. In 6th World congress on intelligent transport systems.

Liu, L., Andris, C., Biderman, A., Ratti C. (2009) Uncovering Taxi Driver's Mobility Intelligence through His Trace, *IEEE Pervasive Computing*.

Liu, Y., Kang, C., Gao, S., Xiao, Y., Tian, Y. (2012a) Understanding intra-urban trip patterns from taxi trajectory data. Journal of Geographical Systems. 2012, 14(4), 463-483. (doi:10.1007/s10109 -012-0166-z)

Liu, Y., Wang, F., Xiao, Y., Gao, S. (2012b) 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. (doi: 10.1016/j.landurbplan.2012.02.012).

Tominski, C., Schumann, H., Andrienko, G. & Andrienko, N. (2012) Stacking-Based Visualization of Trajectory Attribute Data , IEEE Transactions on Visualization and Computer Graphics, Vol. 18, No. 12, 2012.

Yuan, J., Zheng, Y., Zhang, L., and Xie, X. (2012), T-Finder: A Recommender System for Finding Passengers and Vacant Taxis, *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, September, 2012

Yuan, J., Zheng, Y., Sun, G., and Xie, X. (2013), T-Drive: Enhancing Driving Directions with Taxi Drivers' Intelligence, *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 25(1): 220-232, 2013.

Yuan, Y., Raubal, M., and Liu, Y. (2012) Correlating mobile phone usage and travel behavior – a case study of Harbin, China. Computers, *Environment and Urban Systems*, 2012, 36(2): 118-130.

Zheng, Y., and Zhou, X. (2011) *Computing with Spatial Trajectories*. Springer 2011. ISBN:978-1-4614-1628-9