LOCAL MAXIMUM DENSITY APPROACH FOR SMALL-SCALE CLUSTERING OF URBAN TAXI STOPS

Han Wang¹, Xiao-Jian Chen², Ying Wang^{1,*}, Jie Shan^{3,*}

¹ School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China - (wanghanwhu, wangyingstill)@whu.edu.cn
² State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China - cxiaojian@whu.edu.cn
³ Lyles School of Civil Engineering, Purdue University, West Lafayette, IN 47907, USA - jshan@purdue.edu

KEY WORDS: Small-scale Hotspots, Local Maximum Density, Spatial-temporal Distribution, Taxi Stops

ABSTRACT:

Taxi trajectory data contains the detailed spatial and temporal traveling information of urban residents. By using a clustering algorithm, the hotspots' distributions of pick-up and drop-off points can be extracted to explore the patterns of taxi traveling behaviors and its relationship with urban environment. Comparing with traditional methods that determine hotspots at a relatively large scale, we propose an approach to detect small-scale hotspots, so called docking points, to represent the local clusters in both sparse and dense stops areas. In this method, we divide the research area into grids and extract the docking points by finding local maximums of a certain range. The extracted docking points are classified into five levels for the subsequent analysis. Finally, to uncover detail characteristics of taxi mobility patterns, we analyze the distributions of docking points from three aspects—the overall, by day of the week, and by time of the day.

1. INTRODUCTION

As an essential part of urban public transportation, taxis provide point-to-point services to urban residents. In recent years, with the rapid development of sensor technology, most taxis in China have been equipped with GPS track recording devices to track the route and space-time information of taxis (Mao et al., 2016; Yang et al., 2018). Taxi GPS trajectories data records abundant and detailed spatial-temporal traveling information of passengers, which reflects the traveling behaviors of urban residents to a certain degree (Wang et al., 2015; Zhang et al., 2015). Understanding the patterns is essential to alleviate traffic pressure and promote the healthy development of public transportation industry (Noulas et al., 2015; Tang et al., 2015; Zheng et al., 2016; Wang et al., 2017;).

In order to explore the patterns of taxi traveling behaviors and its relationship with urban environment, extracting the hotspots by clustering is one of the most widely used method (Li et al., 2012; Zhao et al., 2015; Zhou et al., 2016). The existing clustering methods aggregate taxi stops into a cluster if they are close, thus the clusters are usually very large in spatial extent. However, considering that taxis are point-to-point services, small-scale hotspots are actually in existence. Therefore, a small-scale hotspots detection method is needed to uncover abundant detailed patterns of the mobility behaviors of people.

This paper proposes a novel density-based approach to detect small-scale local hotspots (defined as 'docking points') for pickup and drop-off points of taxis. After dividing a city into grids, we first extract core grids according to the number of passengers the grids have. A neighborhood E is then used to determine whether a core grid is local maximum. Finally, the grids in neighborhood E are classified into corresponding local maximums. Based on these procedures of the proposed clustering method, it is named as local maximum density (LMD) approach. The remainder of the paper is organized as follows. In Section 2 we review related works. Section 3 describes data preliminary. In Section 4, the LMD method to detect docking points and the definition of different grads for docking points are described. In Section 5, we analyze the spatial distribution of docking points. Finally, concluding remarks and summary are given in Section 6.

2. RELATED WORK

According to similarity measures and cluster evaluation criteria, the current clustering algorithms for extracting hotspots based on trajectory data mainly include: partition-based methods, hierarchical-based methods, and density-based methods.

The core idea of partition-based method is to divide n objects into k (k < n) clusters by iterations according to certain criterion. The representative algorithms are K-Means (Kaufman L, 1987), K-Medoids (Huang Z, 1998), and K-Mode (Jia, 2014). There are also many works using partition-based method to extract hotspots of taxis' pick-up and drop-off points. For example, Lee et al. (2012) use K-means clustering algorithm to extract the clustering center of taxis of a 24-hour period based on the taxi trajectory data in Jeju area. Liu et al. (2009) use this algorithm to analyze the behavior patterns of taxi drivers and explore areas of interest for high-income drivers and general-income drivers. However, the algorithm needs to define the number of classes in advance, which is hard to determine (Bianchi et al., 2014; Yang et al., 2017). If k is too small, the clusters would be very large. One the other hand, if k is too large, the results would be too dispersive to show the clustering structure.

Hierarchical-based method aims at building a hierarchical tree for the objects. The strategies for hierarchical clustering generally fall into 'bottom-up' and 'top-down' types (Wigman et al., 2015; Moseley et al., 2017). Between them, the 'bottomup' methods represented by single-connection are often used to extract taxis' hotspots. For example, Ma et al. (2014) use singleconnection method to extract the hotspots of taxis in different time periods, and find time-dependent areas of interest. However, the hotspots extracted by this algorithm are often connected to multiple roads, and as such the extraction results are large in scale.

Density-based clustering methods consider the spatial clusters as a series of high-density regions separated by low-density regions. The classical algorithm is DBSCAN algorithm (Density-based Spatial Clustering of Application with Noise) (Ester et al., 1996). As for extracting hotspots for taxi, Pan et al. (2013) propose an improved DBSCAN algorithm to extract hotspots from the taxi data in Hangzhou in different periods of time within one week to identify urban functional areas. However, DBSCAN is difficult to extract hotspots with uneven distribution if global parameters are used. The clusters in the dense areas are usually very large.

In summary, the extraction results of existing algorithms usually focus on hotspots in dense area. The extraction scale of the hotspots is more or less too large to uncover the detail structures of the distributions. Therefore, we propose the concept of docking points to represent the local, small-scale hotspots in both sparse and dense stops areas.

3. DATA PREPARATION

In this section, we introduce the taxi GPS dataset used in this paper, and describe the city decomposition method for later work.

3.1 Taxi Dataset

We use a dataset of 6,857 taxis in Wuhan, China for 14 days (from March 1 to March 14, 2015). The GPS sampling intervals are 10s to 60s. After removing erroneous trajectories, a total of 3,098,896 taxi trajectories are extracted for this study.

3.2 City Gridding

The research area lies between 30.29° N to 30.79° N latitude and 113.99° E to 114.59° E longitude, and covers 13 districts of Wuhan. We select 30m as the grid size in this paper and divide the study area into 1851*1910 grids. Then we count the number of passengers in each grid and filter out the grids with no passengers. The division result is shown in Figure 1.



Figure 1. Gird division of the study area in Wuhan, China

4. DOCKING POINTS AND THEIR EXTRACTION

In this section, we describe local maximum density (LMD) method to detect small-scale hotspots of taxi stops. First, the concept of docking points is introduced. Then the details of LMD algorithm are described. Third, the definition of different levels of docking points is given. Finally, the comparison between LMD method and DBSCAN is conducted to demonstrate the performance of our method.

4.1 Analysis Based on Grids

Of the 3,533,410 grids in the study area, only 6.3% (224,186) of the grids have pick-up or drop-off points. Further statistics on the cumulative distribution found that the top 0.9% (32,926) grids cover 80% of the getting on and off the taxi. It means that most of pick-up and drop-off behaviors take place in a small percentage of grids.

Further, we observe the grids exist many local maximums. Different colors are used to indicate the intensity of the passengers as shown in Figure 2 (take a business circle and an ordinary street as an example). Although the number of passengers in different areas is not uniform, there exists local maximums in the consecutive grids with a higher number of passengers in the neighborhood. The small-scale area around each local maximum is the small-scale local hotspot we aims at.



Figure 2. Detail density map of the pick-up/drop-off points

4.2 Docking Points

In previous works, the extracted hotspot areas usually have a large range and tend to concentrate on dense stops areas, which often ignore the local hotspots in relatively sparse stops regions and overlook the detailed structures of the distributions. Therefore, we proposed the concept of 'docking points'.

A 'docking point' is defined as a small-scale area around a local maximum grid. Docking points are essentially hotspots for taxi pick-up/drop-off points, but they have their own unique qualities. It is mainly reflected in the following two aspects. First, in a relatively sparse stops area, there is a specific small-scale area where the grids surround around a local maximum grid (Fig. 2b). Second, a large hotspot area includes some sub-areas which exist local maximums near the center. Thus, compared with the hotspot 'region', the concept of 'docking point' is finer and has a small-scale characteristic.

Based on the analysis above, we draw the conclusion that the docking point we are to extract should meet the following three properties: ①it is a small-scale area; ②the number of pick-up/drop-off passengers in this area is large enough; ③there exists a local maximum grid near the center of this area.

4.3 Docking Points Extraction

According to the concept of 'docking points' in Section 4.2, this paper proposes LMD method based on grids to extract the docking points. It mainly includes the following four stages (the detail procedure is given in Algorithm below):

(1) Two thresholds, $MinPts_0$ and MinPts, are determined to respectively filter out the grids and extract docking points with a few passengers.

Since most of the grids only have a few passengers, we set a threshold $MinPts_0$ to filter out the grids with fewer passengers before determining the local maximums. MinPts represents the minimum number of the passengers for a docking point should have. Therefore, an appropriate MinPts can not only extract the docking points in the dense area, but also in the relatively sparse area.

In order to determine the two thresholds according to the distribution of the pick-up/drop-off points of the input grids, we use the following formula.

$$P_{v_t} = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} value_{ij}(value_{ij} \ge v_t)}{Sum(grid_{ij})}$$

where *m* and *n* are the numbers of grids in latitude and longitude correspondingly, $value_{ij}$ is the number of the pick-up/drop-off passengers in the $grid_{ij}$, $grid_{value} = \{v_0, ..., v_t, ..., v_m\}$ ($v_0 < \cdots < v_t < \cdots < v_m$) is the set of all different $value_{ij}$. Therefore, P_{v_t} represents the percentage of the total number of passengers of the grids with the value exceeds or equals to v_t .

When the value of P_{v_t} is less than 0.8 for the first time, the corresponding v_t is regarded as $MinPts_0$. Similarly, when the value of P_{v_t} is less than 0.5 for the first time, the corresponding value is regarded as MinPts. This purpose is to ensure that the extracted docking points reflect the positions of most taxi passengers as much as possible.

(2) Check whether a core grid is local maximum.

We extract grids with the number of passengers higher than $MinPts_0$. Then we check the neighborhood E of this grid to see whether this core grid is a local maximum grid.

(3) Classify grids in the neighborhood E into local maximums. If the grids have been classified into other local maximums before, reclassify them according to the smaller distances to the local maximums.

(4) Calculate the total number of taxi pick-up/drop-off points of every cluster. If the sum exceeds the threshold *MinPts*, the cluster is docking point. Set the weighted center for each docking point.

4.4 Definition of Docking Point Levels

So far we have extracted the small-scale hotspots area (docking points) to describe the distributions of pick-up/drop-off locations. We now distinguish the different levels of the docking points for further analysis of distribution patterns.

We define five levels for docking points according to the number of people pick-up (drop-off) there. We sorted the docking points by the number of passengers from high to low. Level 5 represents the docking points that contain the top 3% of pick-up (drop-off) behaviors. Level 4 represents 3%-10%. Level 3 represents 10%-25%. Level 2 represents 25%-50%. And Level 1 represents the remaining docking points.

4.5 Validation and Evaluation

To evaluate the validity of proposed LMD method, we check the spatial distributions of docking points and compare the extraction results with DBSCAN method. In this paper, we set 7*7 grids as the neighborhood *E*. Fig 3 shows the spatial distributions of pick-up and drop-off docking points. As expected, the docking-points are very dense around the city center and sparse in remote area. Also, as shown in Table 1, we detect 2,381 pick-up docking points and 4,271 drop-off docking points which cover 75.3% of the pick-up behaviors and 79.2% of the drop-off behaviors respectively. Considering the numbers of grids with pick-up and drop-off behaviors are 130,493 and 202,475 respectively, this shows that we have extracted a small number of hotspots covering most of data from a large number of grids.



(a) Pick-up Points



(b) Drop-off Points Figure 3. Spatial distributions of docking points

Algorithm: LMD method for Taxi Docking Points Extraction						
Input: The grid set $Grid = \{grid_{ij}\}, grid_{ij} = \{i, j, value_{ij}, lng_{ij}, lat_{ij}\}$, and the neighborhood E						
Output: Docking points set $R = \{id, Grid_{id}, lng_{id}, lat_{id}, count_{id}\}$						
Algorithm steps:						
1. Calculate Sum(grid _{ij});						
2. Calculate <i>MinPts</i> ₀ , <i>MinPts</i> ;						
3. Add attribute columns to $grid_{ij}$, $id = 0$ and $if_center = 0$;						
4. Initialize $id_value = 0$;						
5. Traverse the $grid_{ij}$ in sequence						
1) if $(value_{ij} > MinPts_0)$						
2) Find the maximum value in the neighborhood E, MaxE_value;						
3) if $(value_{ij} = MaxE_value)$						
4) Update the current grid, $if_center = 1$, $id = id_value$;						
5) if (there exist grids with $id > 0$ in the neighborhood E)						
Compare the distance to the current center and the original one. If the former is closer,						
update the <i>id</i> ;						
6) Update all the grids in the neighborhood with $id = 0$, $id = id_value$;						
7) $id_value = id_value + 1;$						
Until $grid_{ij}$ traverse is complete;						
6. for $(i = 1: id_value)$						
1) $Grid_i$ is the set of all the grids with $id = i$						
2) calculate the sum_i , values of all the grids with $id = i$						
3) if $(sum_i > MinPts)$						
4) calculate lng_i and lat_i the weighted average of all grid center coordinates with $id = i$;						

- 5) $count_i = sum_i$;
- 6) Save (*i*, $Grid_i$, lng_i , lat_i , $count_i$);

Level	Proportion Range	#Passengers' Range	#Total Passengers	Proportion	#Docking Points
5	0-3%	16306-7023	93462	3%	9
4	3%-10%	6893-3708	219420	7%	46
3	10%-25%	3727-2194	462111	15%	166
2	25%-50%	2182-1046	774773	25%	513
1	50%-75.3%	1045-97	782958	25.3%	1647
1-5	0-75.3%	16306-97	2332724	75.3%	2381

Table 1. Extraction results of pick-up docking points

Level	Proportion Range	#Passengers' Range	#Total Passengers	Proportion	#Docking Points
5	0-3%	27341-11766	96401	3%	5
4	3%-10%	11159-3519	215290	7%	42
3	10%-25%	3457-1647	463951	15%	209
2	25%-50%	1645-713	774080	25%	735
1	50%-79.2%	713-70	903953	29.2%	3280
1-5	0-79.2%	27341-70	2453675	79.2%	4271

Table 2. Extraction results of drop-off docking points

To compare DBSCAN method with LMD method, two parameters *EPS* and *MinPts*_{DBSCAN} are required for DBSCAN. *EPS* is the parameter specifying the radius of a neighborhood from some points, and *MinPts*_{DBSCAN} is the threshold determining whether a point is a core point. More specifically, for a point, if the number of other points within its *EPS* radius is more than $MinPts_{DBSCAN}$, it would be a core point. Therefore, *EPS* and $MinPts_{DBSCAN}$ are corresponding to the radius of neighborhood *E* and the minimum count of a cluster in LMD method. Since the neighborhood *E* we set is 7*7 grids, the *EPS* is set as 105 meters. As for *MinPts*_{DBSCAN}, we check the extraction result of level 3-5 drop-off docking points as an example here, thus the value is set as 1647 (Table 2).

The main difference between DBSCAN and LMD method is that DBSCAN detects clusters by connecting core points and corresponding *EPS* areas, while our method aims at extracting a small E range area around a local maximum grid. Therefore, the detected hot areas by DBSCAN are much larger than the small area extracted by our method. As shown in Fig. 4, the level 3 to 5 drop-off docking points are consistent with the hotspots extracted by DBSCAN method (Fig 4). But LMD method decomposes a big cluster into several relative smaller docking points (Fig 4 Inset).





(b) Result of level 3-5 by LMD Figure 4. Drop-off hotspots extractions by DBSCAN method and LMD method. (Inset) Zoom-in to one of clusters.

5. RESULT AND ANALYSES

In this section, we are going to analyze the spatial distributions of pick-up and drop-off docking points to uncover people's mobility patterns. They are analyzed from three aspects—the overall, by day of the week, and by time of the day.

5.1 Overall Analysis

First, we check all 14 days of taxi data to see the characteristics of pick-up and drop-off docking points in general. The overall distribution characteristics can be concluded as follows.

(1) The number of drop-off docking points is much higher than the number of pick-up docking points. As shown in Table 1 and 2, a total of 2,381 pick-up points and 4,271 drop-off points are extracted. For level 1 to 3, the numbers of drop-off points are much higher than the pick-up points, especially for level 1 with 3280 drop-off points to 1647 pick-up points. The numbers of docking points at level 4 and 5 are not much different. The results show that the spatial distribution of drop-off points is more dispersed than that of pick-up points.

(2) Level 5 docking points are usually surrounded by lower-level docking points. Taking the drop-off docking points as an example (Figure 5). There are five drop-off docking points at level 5, which are around Wuhan Tianhe Airport, Hankou Railway Station, Wuhan Railway Station, Zhongshan Avenue Wusheng Road Overpass and Wuchang Railway Station. There are several docking points around level 5 points. This indicates that there are also a certain number of passengers near the dense stops areas.



Zhongshan Avenue Wusheng Road Overpas

Wuchang Railway Station

Figure 5. Extraction results of intensive pick-up areas

(3) Level 4 and 5 docking points are mainly distributed around city center and along the subway lines (Figure 6). City center has the most intensive business circles, transport hub ports as well as large hospitals and universities. Thus, the regions around the city center are the main active places for people. What's more, it is surprising that the spatial distributions of level 4 and 5 docking points are consistent with subway lines. This indicates that the design of subway lines in Wuhan is fitted well with the popular places in city.

5.2 Analysis by Day of the Week

In order to analyze the different patterns of resident trips on different types of days, 14 days are divided into weekdays and weekends, and then extracting the pick-up and drop-off docking points.



Figure 6. Spatial distribution of the level 4 and 5 docking points.

Since the high-level docking points can best reflect the hotspots, we focus on level 4 and 5 docking points. The docking point distributions on the weekdays and weekends are respectively shown in Figure 7. We find that the pick-up and drop-off docking points are almost the same on the weekdays and the weekends, which are dominated by transportation hub ports and commercial districts at the city center. However, on the weekends, there are several new docking points (red circle in Figure 7). These places are ① Okuyama Century Plaza, Wushang Zhongyuan Square, ② Wanda Plaza (Wuhan Economic Development Store), ③Hankou River Beach and ④ Wangjiawan Subway Station. The results show that there are more people and more demand for taxis in business circles and leisure places during the weekends than on weekdays.



(c)Weekday pick-up points (d)Weekday drop-off points Figure 7. Distribution of level 4-5 docking points on weekdays and weekends

5.3 Analysis by Time of the Day

In order to identify the hotspots at different time periods, this paper further extracts the pick-up/drop-off docking points in different time slots, with three taxi peaks at 7:00-9:00, 12:00-14:00, and 22:00-24:00 respectively. The same as section 5.2, we mainly focus on level 4 and 5 docking points. By comparing with the spatial distributions of pick-up and drop-off docking points in multi time periods, we found three patterns in these peak times.



Figure 8. Distributions of level 4 and 5 docking points in different periods.

(1) During the morning rush hour from 7:00 to 9:00 (Figure 8(a), (b)), the number of high-level pick-up docking points is much higher than that of drop-off docking points. It means that the taxi pick-up point is relatively dispersed during this period, and the drop-off point is more concentrated, indicating people take taxis from a large number of areas to a small number of places. This is because people go to work from their homes. And the residences usually disperse in the city, while the work places are concentrated around the city center.

(2) During the taxi peak at 12:00-14:00 (Figure 8(c), (d)), the spatial distributions of pick-up and drop-off docking points are similar, which is different from the phenomenon observed from 7:00 to 9:00. This indicates the interactions among different areas in city are balanced and active at noon.

(3) During the night time from 22:00-24:00 (Figure 8(e), (f)), the taxi flows exhibits opposite trend as the one in the morning. The pick-up points are concentrated while the drop-off points are scattered. This is because the popular night-life places usually lie in commercial areas at the city center, and 22:00-24:00 is the time people take taxis from these places to homes.

6. CONCLUSION AND FUTURE WORK

This paper proposes LMD method for docking points extraction based on grids, and analyzes the characteristics of mobility patterns from the change of the spatial-temporal distributions of the docking points. Compared with existing algorithms, this method could extract small-scale hotspots in both sparse and dense stops regions by checking local maximum density. Further, the resident mobility patterns are analyzed from three aspects the overall pattern, different day of the week, and different time of the day. Specifically, we extract the pick-up/drop-off docking points separately, and take high-level docking points with relatively more passengers as key analysis subject. By comparing their spatial distributions, the more detailed patterns of urban taxis are observed and understood.

However, it should be pointed out that our method is based on grids. The influence of different sizes of grids is still an open question. Also, the definition of neighborhood E is fixed, which influences the shape of docking points. A dynamic growth of a neighborhood E is expected. Finally, the experiment in this paper only uses taxi trajectory to analyze the spatial-temporal patterns of urban hotspots. In the future work, buses, subways and shared bicycles for resident trips should be combined to conduct a more comprehensive analysis.

REFERENCES

Bianchi, F., Livi, L., Rizzi, A., 2014. Two density-based kmeans initialization algorithms for non-metric data clustering. *Pattern Analysis and Applications*, 1-19.

Ester, M., Kriegel, H., Sander, J., Xu, X., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. *KDD-96*, AAAI Press, 226-231.

Huang, Z., 1998. Extensions to the k-means algorithm for clustering large data sets with categorical values. Data mining and knowledge discovery, 2(3), 283-304.

Jia, A., 2014. Review of clustering algorithms based on partitioning. *Electronic Design Engineering*, 22(23), 38-41.

Kaufman, L., 1987. Clustering by means of medoids. *Statistical data analysis based on the L*₁-norm and related methods, 405-416.

Lee, J., Shin, I., Park, G., 2008. Analysis of the Passenger Pick-Up Pattern for Taxi Location Recommendation. *Fourth International Conference on Networked Computing and Advanced Information Management. IEEE Computer Society.*

Li, X., Pan, G., Wu, Z., Qi, G., Li, S., Zhang, D., Zhang, W., Wang Z., 2012. Prediction of urban human mobility using large-scale taxi traces and its applications. *Frontiers of Computer Science*, 6(1), 111-121.

Liu, L., Andris, C., Bidderman, A., 2009. Revealing taxi drivers mobility intelligence through his trace. *IEEE Pervasive*

Computing, 160, 1-17.

Ma, Y., 2014. Research on hotspot area and space-time characteristics of residents based on taxi track points--taking Kunshan City as an example. *Nanjing Normal University*.

Mao, F., Ji, M., Liu, T., 2016. Mining spatiotemporal patterns of urban dwellers from taxi trajectory data. *Frontiers of Earth Science*, 10(2).

Moseley, B., Wang, J., 2017. Approximation Bounds for Hierarchical Clustering: Average Linkage, Bisecting K-means, and Local Search. *Advances in Neural Information Processing Systems 30*, 3094–3103.

Noulas, A., Salnikov, V., Lambiotte, R., Mascolo, C., 2015. Mining open datasets for transparency in taxi transport in metropolitan environments. *EPJ Data Science*, 4(1): 23.

Pan, G., Qi, G., Wu, Z., Zhang, D., Li, S., 2013. Land-use classification using taxi GPS traces. *IEEE Transactions on Intelligent Transportation Systems*, 14(1), 113-123.

Tang, J., Liu, F., Wang, Y., Wang, H., 2015. Uncovering Urban Human Mobility From Large Scale Taxi Gps Data. *Physica A: Statistical Mechanics and its Applications*, 438, 140-153.

Wang, W., Pan, L., Yuan, N., Zhang, S., Liu, D., 2015. A Comparative Analysis of Intra-City Human Mobility by Taxi. *Physica A: Statistical Mechanics and its Applications*, 420, 134-147.

Wang, Y., Yang, W., Xu, Z., 2017. Hotspots Mining Based on Taxi Tracks. *China Computer and Communication*, 16.

Wigman, J., Van Os, J., Borsboom, D., Wardenaar, K., Epskamp, S., Klippel, A., Viechtbauer, W., Myin-Germeys, I., Wichers, M., 2015. Exploring the underlying structure of mental disorders: cross-diagnostic differences and similarities from a network perspective using both a top-down and a bottom-up approach. *Psychological Medicine*, 45(11), 2375-2387.

Yang, S., Dong, X., Sun, L., Zhou, Y., 2017. A Data-driven Process Recommender Framework. *The 23rd ACM SIGKDD International Conference*. ACM.

Yang, Y., He, Z., Song, Z., Fu, X., Wang, C., 2018. Investigation on structural and spatial characteristics of taxi trip trajectory network in Xi'an, China. *Physica A: Statistical Mechanics and its Applications*, 506.

Zhang, J., Wu, F., Zhang, H., 2015. Urban Residents Travel Characteristics Mining Utilizing Taxi Trajectory Data. *Geography and Geo-Information Science*, 31(6).

Zhao, P., Qin, K., Zhou, Q., Li, Z., 2015. Detecting Hotspots from Taxi Trajectory Data Using Spatial Cluster Analysis. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2(4): 131-135.

Zheng, Z., Rasouli, S., Harry Timmermans, 2016. Two-Regime Pattern in Human Mobility: Evidence From Gps Taxi Trajectory Data. *Geographical Analysis*, 48(2), 157-175.

Zhou, Q., Qin, K., Chen, Y., Li, Z., 2016. Hotspots Detection from Taxi Trajectory Data Based on Data Field Clustering. Geography and Geo-Information Science, 32(6): 51-56.