

SPATIO-TEMPORAL CHARACTERISTICS OF RESIDENT TRIP BASED ON POI AND OD DATA OF FLOAT CAR IN BEIJING

Naixia MOU^{a,b}, Jie Li^a, Lingxian ZHANG^a, Wenbao LIU^a, Yujing XU^a

^a College of Geomatics , Shandong University of Science and Technology, Qingdao 266590,China;

^b State Key Laboratory of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101,China

Commission VI, WG VI/4

KEY WORDS: Taxi Trajectory Data; Travel Mode; Spatio-Temporal Behavior; Spatial Clustering Analysis

ABSTRACT:

Due to the influence of the urban inherent regional functional distribution, the daily activities of the residents presented some spatio-temporal patterns (periodic patterns, gathering patterns, etc.). In order to further understand the spatial and temporal characteristics of urban residents, this paper research takes the taxi trajectory data of Beijing as a sample data and studies the spatio-temporal characteristics of the residents' activities on the weekdays. At first, according to the characteristics of the taxi trajectory data distributed along the road network, it takes the Voronoi generated by the road nodes as the research unit. This paper proposes a hybrid clustering method -- based on grid density, which is used to cluster the OD(origin and destination) data of taxi at different times. Then, combining with the POI data of Beijing, this research calculated the density of the POI data in the clustering results, and analyzed the relationship between the activities of residents in different periods and the functional types of the region. The final results showed that the residents were mainly commuting on weekdays. And it found that the distribution of travel density showed a concentric circle of the characteristics, focusing on residential areas and work areas. The results of cluster analysis and POI analysis showed that the residents' travel had experienced the process of " spatial relative dispersion - spatial aggregation - spatial relative dispersion " in one day.

1. INTRODUCTION

The widespread use of Information and Communication Technology (ICT) has accelerated the exchange of time and space for knowledge, technology, talent and capital. And that also has changed the way of people's life, residence, work and entertainment. With a wide range of sensors (GPS receivers, mobile phones, cameras, etc.), a large number of location-based data (float car data, cell phone location data, microblogging data, etc.) have been generated. As a class of sensors, taxi is an important means of daily travel. Compared with other public transport, it has relatively random nature and has larger travel density difference in different regions. Its travel behavior, time, up and down information are closely related with the individual travel activities, which can also reflect the residents of the time and space travel characteristics effectively. And the taxi's OD information (the customers' up and down information) are better to reflect the behavior of residents' travel^[1, 2].

Based on the Voronoi polygon which is generated by the main road network and the road node, this paper uses the clustering method of grid density to cluster the OD of the taxi at different times. In order to analyze the clustering results more intuitively, paper also combines the POI data of Beijing. Though analyzing

the density distribution in the clustering results, we can analyze the relationship between the active area of the residents and the functional types of the residents in different periods synthetically^[3,4].

2. DATA OVERVIEW

Taxi trajectory data is a set of time-stamped and time-ordered sequence of points, and its basic model can be expressed as

$$TR = \{(X_1, Y_1, T, v_1), (X_2, Y_2, T + \Delta t, v_2), \dots\} \quad (1)$$

Where, TR represents a spatio-temporal track data, and Δt represents the sampling time interval of taxi locus data. In addition to recording time and location information, v_i indicates other attribute information, including vehicle number, passenger status (0/1, empty / passenger), speed and other information.

This paper takes six districts of Beijing city (Chaoyang District, Haidian District, Fengtai District, Shi Jingshan District, Dongcheng and Xicheng) as the research areas. It uses the track data of more than 12,000 float cars in Beijing for seven days (March 1, 2012 - March 7, 2012) to study the travel time and space characteristics of residents. Opposing to the law of working day, the travel schedule of the weekend is not obvious.

The first author: Naixia MOU (1973-), male, associate professor, Pingdu, Shandong province, majored in trajectory data mining and other aspects of research. E-mail : mounaixia@163.com.

Supported by: National Natural Science Foundation of China(41771476), Natural Science Foundation of Shandong Province(ZR2016DM02), National key R & D plan(2016YFB0502104), Open fund of State Key Laboratory of Resources and Environment Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences

So the follow-up study of residents travel will have not a detailed study to the rest day.

The taxi trajectory data showed the following characteristics in space: (1)The distribution was extremely uneven in the global scope. The center distribution was strong. And the surrounding distribution is very sparse (2)In the local scope, it also appeared as non-uniform distribution, showing a linear clustering centered on urban roads (3)The data which was gathered at the intersection was more obvious. And the density of the taxi OD was proportional to the density of the network node in the space.

3. BASED ON GRID DENSITY CLUSTERING

Because of the obvious clustering of data at the intersection, this paper proposes a clustering method combining grid density of road network data to achieve better clustering effect and higher execution efficiency. The similarity measure of the algorithm includes two aspects: spatial similarity and attribute similarity. The spatial similarity lies in the fact that the clustered road grid must be associated with a common point or common edge adjacently. The similarity of the attribute is that the density of the grid cells between the clusters of the same cluster is not much different^[5-9]. The process and strategy of clustering method based on grid density are shown in Figure 1.

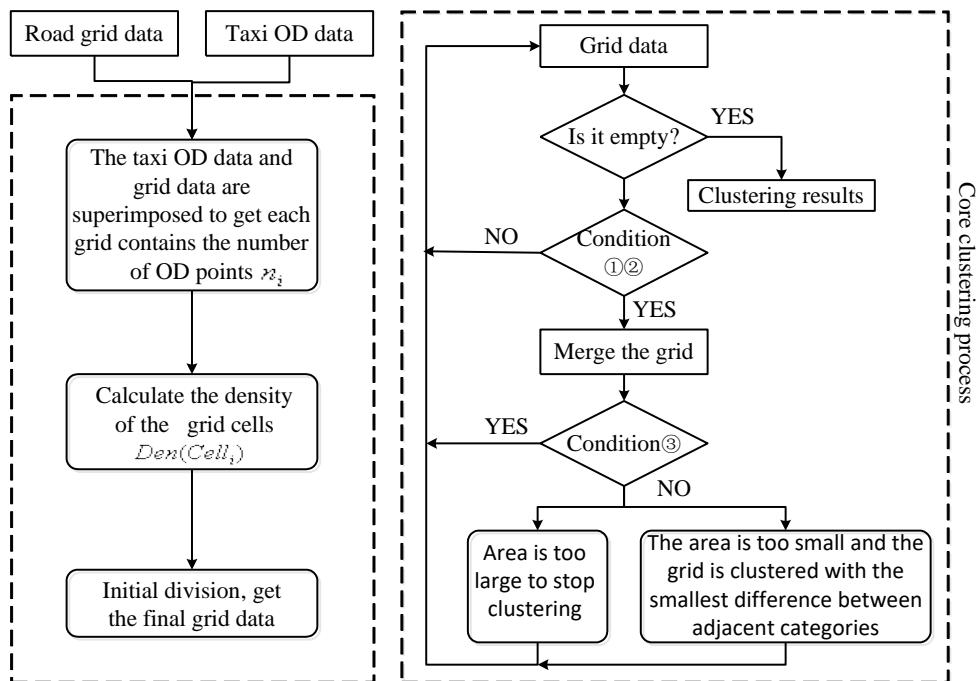


Figure 1. Flow chart based on grid clustering algorithm

Note: ① having adjacent grid ② the difference between the class is less than or equal to 3 ③ the area of cluster grid meet $0.01\text{km}^2 \leq S_h \leq 0.5\text{km}^2$

The definitions and algorithms involved in the flowchart are as follows:

Definition 1 Grid cell density The density of the grid ($Cell_i$) is the number of taxi OD points per unit area, which can be expressed as $Den(Cell_i)$,

$$Den(Cell_i) = \frac{n_i}{S_i} \quad (2)$$

Where, n_i represents the number of taxi OD points contained in the $Cell_i$, S_i represents the area of $Cell_i$.

Definition 2 The initial division The "natural breaks" are used to divide the entire grid area by the density of the grid cells $Den(Cell_i)$, maximizing the differences between the categories. It minimizes the standard deviation within the category, and attaches the corresponding category attribute

values class for each grid unit.

Definition 3 Grid directly density reachable For grid cells $Cell_i$, if the absolute value of the difference of the Class attribute value of the adjacency grid cell $Cell_k$ is less than or equal to 3, then $Cell_i$ and $Cell_k$ are Directly Density Reachable, and there is symmetry between them.

Using the above algorithm, the clustering analysis of O-point and D-point can reflect the characteristics of residents' travel in several major hours of work day (7: 00 ~ 10: 00, 16: 00 ~ 19: 00, 22: 00 ~ 24: 00).The clustering results of different periods has been obtained. They are classified into 8 categories according to the travel density in the block, and are marked as C0, C1, C7 (as shown in Figure 2-4).

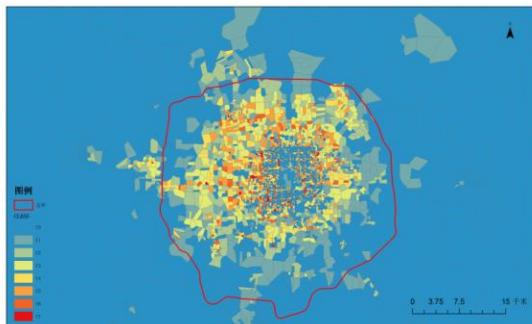


Figure 2(A). The clustering results of O-point

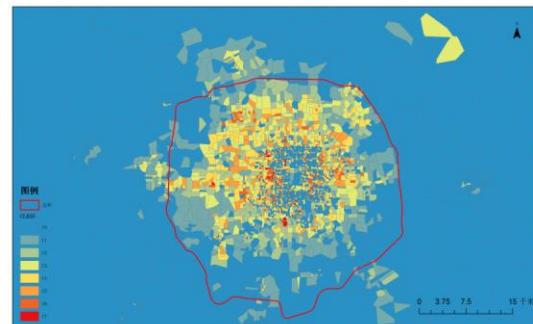


Figure 2. The clustering results of 07:00~10:00 on weekday

Figure 2(B). The clustering results of D-point

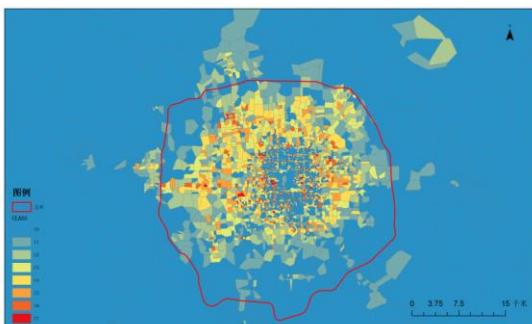


Figure 3 (A). The clustering results of O-point

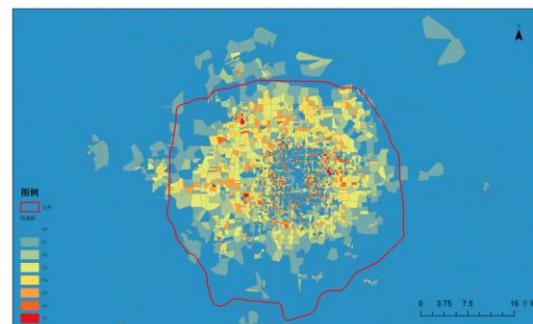


Figure 3. The clustering results of 16: 00 ~ 19: 00 on weekday

Figure 3(B). The clustering results of D-point

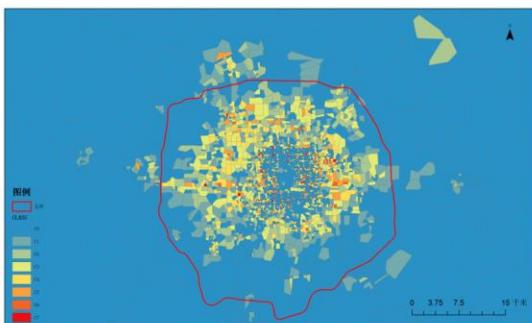


Figure 4(A). The clustering results of O-point

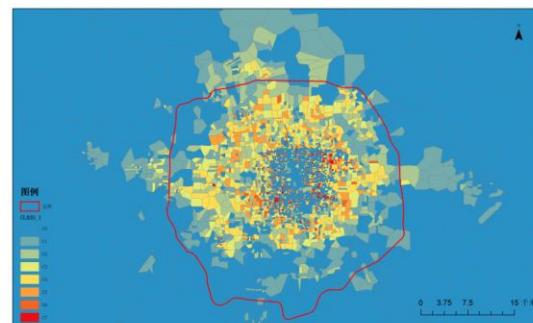


Figure 4. The clustering results of 22:00~24:00 on weekday

Figure 4(B). The clustering results of D-point

From the clustering results, it can be seen that most of the residents' travel within the five rings, and residents have less outside travel activities. There are some differences in the clustering results of each time period at O-point and D-point. But the distribution of concentric circles is shown in the overall space, the degree of aggregation was increased from suburban to urban areas gradually, and the clustering results are banded along the road distributed. C0 is a class which has the largest number of plots and the largest area of plots. Most of the outer area (the main urban area within a small part of this category, most of the park, the interior of the attractions) belong to this category. The number of O-point and D-point is very low in each period of time. So such land will not be analyzed. The clustering results serve as a basis for identifying the characteristics of the active mode in the region.

4. CLUSTERING RESULTS ANALYSIS OF POI DATA

Through the clustering analysis of the taxi OD data, we get the different density grades. In order to analyze the relationship between the functional distribution characteristics and the residents' travel characteristics in Beijing city, the paper introduces the POI data of Beijing to express the semantic representation of the geographical position. The POI data is consisted by 653062 points, and the class 1 has 18 categories. In order to reflect the spatial and temporal characteristics of urban residents more clearly, 16 representative categories are selected from 18 POI categories and divided into 8 categories: government agencies, living services, working, shopping services, leisure and entertainment, residential quarters, research education and complex.

In order to analyze the relative distribution of POI data in various types of plots, the density of the different categories of POIs is handled by the maximum - minimum standardization method. First, the density values of each type of POI within the various types of plots in the clustering results are counted

$$Den_j^i = \frac{N_i}{\sum S_m} \quad (3)$$

Where, i represents the category of the POI ($i = 1, 2, \dots, 8$), and j represents the class of the plots ($j = 1, 2, \dots, 7$).

And then the process of maximum-minimum standardized for the various POI densities is shown below:

$$D_j^i = \frac{Den_j^i - Den_{\min}^i}{Den_{\max}^i - Den_{\min}^i} \quad (4)$$

Where, D_j^i is the normalized value of the i -th class POI in the j -th plot. Den_{\min}^i and Den_{\max}^i indicate the minimum and maximum values of the density of the same type of POI in the seven plots respectively.

4.1 Workday 07: 00 ~ 10: 00 Clustering Results Analysis

From the clustering results (Figure 2(A).) and the POI

	C1	C2	C3	C4	C5	C6	C7
Area (km ²) :	216.382	108.146	87.759	36.155	37.028	15.113	3.538
Government agencies	0	0.223307	0.555252	0.59808	0.678816	1	0.935136
Living services	0	0.103254	0.217009	0.338915	0.402939	0.605365	1
Working	0	0.135279	0.292956	0.464655	0.693823	1	0.676795
Shopping services	0	0.083029	0.224169	0.381676	0.465548	0.730611	1
Leisure and entertainment	0	0.111588	0.307731	0.424423	0.539697	0.72519	1
Residential quarters	0	0.409009	0.728658	0.941805	1	0.89039	0.98883
Research education	0	0.159798	0.628269	0.569953	0.798879	1	0.719329
Complex	0	0.122644	0.305501	0.407473	0.554837	0.766064	1

Table 1. POI standardized values of various plots (07: 00 ~ 10: 00 O)

	C1	C2	C3	C4	C5	C6	C7
Area (km ²) :	219.266	98.719	102.603	42.227	32.228	12.801	4.780
Government agencies	0	0.192739	0.34092	0.642071	0.655885	0.627501	1
Living services	0	0.128601	0.214395	0.377204	0.648007	0.755776	1
Working	0	0.125456	0.233103	0.501224	1	0.962468	0.955739
Shopping services	0	0.139998	0.24002	0.372681	0.621744	0.823225	1
Leisure and entertainment	0	0.179204	0.301191	0.437344	0.773848	0.704213	1
Residential quarters	0	0.60294	0.662608	0.777792	1	0.501685	0.306313
Research education	0	0.269665	0.407341	0.484771	1	0.590816	0.413763
Complex	0	0.227262	0.373897	0.518444	0.901169	0.93313	1

Table 2. POI standardized values of various plots (07: 00 ~ 10: 00 D)

In summary, in the period of 7: 00 ~ 10: 00, the plots of C6 and C5 with high density are mostly residential area at O-points, and the C5 plot also has the function of work and education. C4 and C3 are mainly residential areas. For D-points, at this period, the area C6 with higher density is the

statistical results (Table 1.) of the O-points from 07:00 to 10:00, it can be seen that the block of C7 with the highest density is the smallest area. As the relatively high standard values of all POIs, C7 is a mixed area. The POI standard values of "government agency", "working", "residential quarter" and "research education" in C6 are 1, 1, 0.89039 and 1. So C6 may belong to the mixed area of work, school and residence. The standard value of "residential quarter" in C5, C4 and C3 are higher than the standard value of other categories obviously in the plots. So they may be typical residential areas. The standard value of "research education" in C5 is also higher. So the C5 plot has the function of housing and education.

From the clustering results (Figure 4(B).) and the POI statistical results (Table 2.) of this period, it can be seen that the function of the C7 plot is similar to that of the O-point, and it belongs to the mixed area too. The maximum POI standard value in the C6 block is "working", indicating the higher proportion of the work area in this plot. The POI standard value of "working", "residential quarter" and "research education" in C5 is 1, indicating that the block is a mixed area of work, residence, education. The maximum POI standard value of C4 and C3 plots is "residential quarter", indicating that such plots are residential area.

working area, and the C5 plot is a mixed area of work, residence, and education. C4 and C3 are mainly residential areas. It can be seen that the distribution of residential areas is relatively scattered, and the work area is more concentrated. It can be drawn a travel law of "from spatial relative

dispersion to spatial aggregation" from this period.

4.2 Workday 16: 00 ~ 19: 00 Clustering Results Analysis

From the clustering results (Figure 3(A).) and the POI statistical results (Table 3.) of the O-points from 16:00 to 19:00, it can be seen that the C7 has high POI standard values except "residential quarter" and "research education". It indicates that the region is a mixed area of non-residential education area. The standard value of the "working" type in C6 is the maximum and significantly higher than the other POI types of the plots, indicating that the area is the work area. The standard values of "working", "residential quarter" and "research education" in C5 is high, indicating that the area is a mixed area of work, residence and scientific research education. The standard value of "residential quarter" and "research education" is the maximum value of 1 in C4. It indicates that such plots

belong to the education and residential integrated land. The standard value of "residential area" in C3 is higher than that of other categories, indicating that C3 belongs to residential area.

From the clustering results (Figure 3(B).) and the POI statistical results (Table 4.) of this period, it can be seen that the function of the C7 block is similar to that of the O-points and belongs to the non-living mixed area. The "shopping service" and "complex" in C6 area have a higher POI standard value, indicating that the plot is a shopping area. The "residential quarter" POI standard values are higher than the other categories in C5, C4 and C3, indicating that the residential area occupies a high proportion.

	C1	C2	C3	C4	C5	C6	C7
Area (km ²) :	97.357	63.777	31.325	34.099	17.402	7.086	1.244
Government agencies	0	0.275385	0.565319	0.689529	0.796066	0.774223	1
Living services	0	0.058254	0.205598	0.331319	0.52961	0.690165	1
Working	0	0.095972	0.28497	0.606932	0.88065	1	0.86868
Shopping services	0	0.027303	0.119753	0.183479	0.310246	0.384945	1
Leisure and entertainment	0	0.056449	0.194429	0.296595	0.456512	0.538271	1
Residential quarters	0.05390	0.326844	0.913074	1	0.963394	0.716557	0
Research education	0	0.295909	0.515811	1	0.961198	0.738964	0.489602
Complex	0	0.115704	0.347457	0.478759	0.734846	0.717997	1

Table 3. POI standardized values of various plots (16: 00 ~ 19: 00 O)

	C1	C2	C3	C4	C5	C6	C7
Area (km ²) :	95.455	58.628	47.136	21.166	13.955	3.718	0.761
Government agencies	0	0.111415	0.315219	0.498779	0.426913	0.732829	1
Living services	0	0.062949	0.206548	0.285965	0.405228	0.660554	1
Working	0	0.100195	0.271854	0.452167	0.663979	0.762472	1
Shopping services	0	0.057439	0.190274	0.267246	0.382683	1	0.947979
Leisure and entertainment	0	0.067464	0.190335	0.308735	0.352858	0.687765	1
Residential quarters	0	0.336246	0.728603	1	0.796762	0.708218	0.547254
Research education	0	0.073941	0.576323	0.598658	0.488693	0.636752	1
Complex	0	0.13719	0.381177	0.547527	0.620196	1	0.876196

Table 4. POI standardized values of various plots (16: 00 ~ 19: 00 D)

In summary, in the period of 16: 00 ~ 19: 00, the block of C6 and C5 with high density is mostly working area at O-points. At the same time, it is found that the C6 and C5 plots in the clustering results of this period at O-points coincide with the C6 and C5 plots in the clustering results of 7: 00 ~ 10: 00 at D-points. And for D-points, the C6 area with high density is shopping integrated area. C5, C4 and C3 block are mainly residential area. It can be seen that the shopping area is more gathered than the working area, and residents moved from the working area to the shopping area or residential area during this period. It can be concluded a trip law of "space to further gather" or "space from spatial aggregation to spatial dispersion".

4.3 Working days 22: 00 ~ 24: 00 Clustering results

From the clustering results (Figure 4(A).) and the POI statistical results (Table 5.) of the vehicle from 22:00 to 24:00, it can be seen that the travel volume of the period is significantly lower than that of the above two periods. The scope of the gathering area of this period has been reduced, and this period of travel is not particularly obvious regularity. The C7 block which has many types of urban functions is located in the center of the city. The standard value of the C1 ~ C6 block is not much different. So it is difficult to determine the function of the plot.

From the clustering results of this time (Figure 4(B).) and the POI statistical results (Table 6.), it can be seen that the arrival of the late-night residents still have a certain law. The standard value of the "residential quarters" and "research education" in

C4, C5 and C6 blocks is significantly higher than other categories, indicating that the plot belongs to the residential education of the integrated area. C7 is located in the central area. It is a multi-functional integrated area.

	C1	C2	C3	C4	C5	C6	C7
Area (km ²) :	75.870	48.991	40.232	17.909	9.902	2.993	0.949
Government agencies	0	0.066687	0.196737	0.190194	0.494904	1	0.667456
Living services	0	0.042595	0.106197	0.157111	0.418275	0.85906	1
Working	0	0.046128	0.110617	0.164917	0.509779	1	0.496041
Shopping services	0	0.020757	0.087876	0.145399	0.41433	1	0.855654
Leisure and entertainment	0	0.035619	0.140438	0.180174	0.538232	1	0.963896
Residential quarters	0	0.079054	0.191647	0.22526	0.726331	1	0.716297
Research education	0	0.024544	0.199465	0.152132	0.568217	1	0.514673
Complex	0	0.04248	0.136125	0.16785	0.526354	1	0.921145

Table 5. POI standardized values of various plots (22: 00 ~ 24: 00 O)

	C1	C2	C3	C4	C5	C6	C7
Area (km ²) :	75.870	48.991	40.232	17.909	9.902	2.993	0.949
Government agencies	0	0.209301	0.537429	0.626122	0.890794	0.642117	0.869839
Living services	0	0.089645	0.193724	0.334905	0.548699	0.566036	1
Working	0	0.089444	0.395361	0.546714	0.980768	0.625164	0.789868
Shopping services	0	0.042952	0.131946	0.218408	0.359974	0.453591	1
Leisure and entertainment	0	0.091546	0.208719	0.372931	0.591534	0.567712	1
Residential quarters	0	0.320324	0.420079	0.529114	1	0.862154	0.964378
Research education	0	0.400066	0.525364	0.626915	1	0.892140	0.686485
Complex	0	0.116204	0.275931	0.448856	0.740951	0.797993	1

Table 6. POI standardized values of various plots (22: 00 ~ 24: 00 D)

In summary, travel law at 22: 00 ~ 24: 00 time period is not obvious, especially on the area of O-point is mostly integrated area. However, according to the clustering situation of O-point area in this period, it can be found that C6 is expressed as residential area, and C5 is represented as a mixed area of residence, education and work. Therefore, this period of travel is from the integrated area to the residential area. The travel law of residents is "from the relative space aggregation to spatial dispersion".

5. CONCLUSION

Combined with POI data, this paper research analyzed the clustering results of important travel time during the day synthetically. And the following results are obtained:

The law of the residents' travel in the working day can be reflected as follows: (1) In 7: 00 ~ 9: 00, travel density is higher. Residents move from the residential area to the work area. And a large number of areas belong to the mixed areas, especially the mixed area of residence, work and education; (2) It is mainly from the work to the place of residence or shopping recreation area in 16: 00 ~ 19: 00; (3) In 22: 00 ~ 24: 00, travel density was significantly lower than the previous two periods. Travel behavior is mainly from the integrated area to the residence area.

From this paper research, it can be summed that resident residence is relatively scattered, and the work area is relatively gathered. The law of the residents' travel in the working day can be reflected as follows: living - work - shopping and leisure - living, that is to say, the process is expressed as "spatial relative dispersion - spatial aggregation - spatial relative dispersion".

REFERENCES

MOU, N X., 2015. A Review on the Application Research of Trajectory Data Mining in Urban Cities. *Journal of Geo-Information Science*, 17(10), pp. 1136-1142.

LU, F., 2014. Big Data and Generalized GIS. *Geomatics and Information Science of Wuhan University*, 39(6), pp. 645-654.

Zheng, Y., 2013a. U-Air: when urban air quality inference meets big data. *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, USA, pp.1436-1444.

Zheng, Y., 2014a. Diagnosing New York city's noises with ubiquitous data. *ACM International Joint Conference on Pervasive and Ubiquitous Computing*, New York, USA, pp. 715-725.

ZHAO, H., 2010. Grid-Based Clustering Algorithm. *Computer Technology and Development*, 20(09), pp. 83-85.

ZHOU, T., 2012. Clustering algorithm research advances on data mining. *Computer Engineering and Applications*, 48(12), pp. 100-111.

Edla, D R., 2013a. A grid clustering algorithm using cluster boundaries: *Information and Communication Technologies*, Trivandrum, India 15 (3) , pp.254-259.

Wang, W., 1997a. STING: A Statistical Information Grid Approach to Spatial Data Mining: *Vldb'97, Proceedings of International Conference on Very Large Data Bases*, Athens, Greece.

Agrawal, R., 1998a. Automatic subspace clustering of high dimensional data for data mining applications. *ACM SIGMOD Record*. ACM, 27(2), pp.94-105.