Hotspot Identification and Facility Layout Perception Based on Taxi Data: A Case Study of Dezhou City

Yiwen Zhang, Xianghong Li*

Henan Polytechnic University of Energy Science and Engineering, Jiaozuo, Henan, 454003, China

(*lixianghong001@126.com)

Abstract. Based on the travel characteristics of urban residents in taxi data, this paper selects Dezhou as the research object, and uses DBSACN clustering and standardized ellipse to analyze the travel characteristics of urban residents in Dezhou for two weeks. The purpose is to deeply explore the travel characteristics of residents in Dezhou, analyze the characteristics of current residents' travel, and put forward targeted strategic suggestions based on this. The results show that the travel of residents in Dezhou is generally centered on the western part of Decheng District, the layout of urban facilities is relatively concentrated, its commercial facilities are relatively thin, and the level of urban commercialization needs to be improved.

Keywords: taxi data, residents' travel characteristics, cluster analysis, standardized ellipse

1 INTRODUCTION

With the acceleration of urbanization, urban traffic problems are becoming increasingly prominent. How to effectively manage and plan urban traffic, improve urban operation efficiency, and reduce traffic congestion has become an important topic in urban planning and management[1]. As an important part of the urban transportation system, taxis bear a large number of citizens' travel needs[2]. Their operation data reflects the travel characteristics and traffic flow of urban residents. Taxi GPS trajectory data combined with urban road network and other socio-economic attributes can be used to mine the spatiotemporal characteristics of residents' travel and reflect the current urban traffic status[345].

This paper analyzes the taxi trajectory data in Dezhou City, reveals the spatiotemporal dynamic characteristics of Dezhou City residents' travel, provides basic reference materials for alleviating urban traffic pressure, and has certain reference significance for Dezhou City's urban planning and construction.

G. Chen et al. (eds.), *Proceedings of the 2024 International Conference on Rail Transit and Transportation (ICRTT 2024)*, Advances in Engineering Research 254,
https://doi.org/10.2991/978-94-6463-610-9_13

2 RESEARCH METHODS

2.1 DBSCAN Clustering to Identify Travel Hotspots

DBSCAN is a density-based clustering method with excellent noise resistance. It can filter low-density data points when the data density is unevenly distributed, and divide the areas with high data point density into clusters, thereby identifying hotspot areas[6]. This method can identify clusters of any shape. At the same time, the distribution of taxi drop-off points is not constrained by a specific form, and hotspots of various shapes may be formed. The characteristics of the two are more consistent, so it is more appropriate to use this method to identify taxi travel hotspots[7].

There are two very important parameters in this algorithm, namely Eps neighborhood and MinPts. Eps neighborhood refers to the area with a radius of Eps (m) centered on a certain drop-off point p. MinPts refers to the minimum threshold required to determine whether a certain drop-off point is a core point. According to the experience of predecessors, Eps and MinPts are limited to a certain range[89]. Through repeated experiments, the number of clusters is observed as the number of Eps and MinPts changes with the combination of two of them. The results with a more reasonable number of clusters are selected for further image verification, and finally the best combination of Eps and MinPts is determined[7].

2.2 Construction of Standard Deviation Ellipse of Travel Hotspots

The spatial statistical method of standard deviation ellipse is used to quantitatively analyze the data distribution. To construct a standard deviation ellipse, the following parameters need to be calculated: the center of the ellipse, the rotation angle, and the length of the major and minor semi-axis.

The center of the data distribution, i.e. the center of the ellipse, is the arithmetic mean center of all data points $(\bar{X}, \bar{Y})$, and its calculation formula is shown in (1) and (2).

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad (1)$$

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} y_i \quad (2)$$

In the formula, $\bar{X}$, $\bar{Y}$ is the average value of the X-coordinates and Y-coordinates of all boarding points, x_i represents the X-axis coordinate value of the i-boarding point, y_i represents the Y-axis coordinate value of the i-boarding point, and n is the total number of boarding points.

The data distribution direction, that is, the ellipse rotation angle, is the angle of the ellipse's major axis rotating clockwise from due north. Its calculation formula is shown in (3) to (6).

$$\theta = \arctan\left(\frac{A+B}{C}\right) \tag{3}$$

$$A = \sum_{i=1}^{n} (x_i - \bar{X})^2 - \sum_{i=1}^{n} (y_i - \bar{Y})^2 \tag{4}$$

$$B = \sqrt{\left(\sum_{i=1}^{n} (x_i - \bar{X})^2 - \sum_{i=1}^{n} (y_i - \bar{Y})^2\right)^2 + 4\left(\sum_{i=1}^{n} (x_i - \bar{X})(y_i - \bar{Y})\right)^2} \tag{5}$$

$$C = 2\sum_{i=1}^{n} (x_i - \bar{X})(y_i - \bar{Y}) \tag{6}$$

The core range of data distribution is measured by the elliptical area of a standard deviation ellipse. Formulas (7) and (8) are used to calculate the variance of the X-axis SDE_X and the variance of the Y-axis SDE_y of all boarding points in the original coordinate system to determine the values of the major and minor semi-axis. When $SDE_X > SDE_y$, the length of the major semi-axis is σ_x and the length of the minor semi-axis is σ_y; when $SDE_X < SDE_y$, the length of the major semi-axis is σ_y and the length of the minor semi-axis is σ_x.

$$SDE_X = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{X})^2}{n}} \tag{7}$$

$$SDE_y = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \bar{Y})^2}{n}} \tag{8}$$

After determining the center of the ellipse, the direction of rotation, the length of the major and minor axes, the standard deviation ellipse of the data set can be drawn and the area of the ellipse can be calculated. The calculation formula is shown in (9).

$$S = \pi\sigma_x\sigma_y \tag{9}$$

Each piece of travel data contains the latitude and longitude information of the starting and ending points of the trip. The mileage of each piece of data is calculated according to formula (10), and the average mileage is calculated, which is the average attraction distance of the hotspot.

$$\mathrm{d} = \mathrm{R}^{*}\arccos\left(\sin(\mathrm{l}at1)^{*}\sin(\mathrm{l}at2) + \cos(\mathrm{l}at1)^{*}\cos(\mathrm{l}at2)^{*}\cos(\mathrm{l}on2 - \mathrm{l}on1)\right) \tag{10}$$

3 EXAMPLE ANALYSIS

This paper takes Dezhou as the research object and analyzes its taxi data for 14 days from July 3, 2023 (Monday) to July 16, 2023 (Sunday). The data in this article comes from the Shandong Public Data Open Network. According to the surrounding environment of the generated travel hotspots, it is divided into 6 types of facilities, namely transportation hubs, medical facilities, commercial centers, residential areas, scenic parks, and educational service facilities.

3.1 Distribution Characteristics of Travel Hotspots

Since the drop-off hotspots to be identified in this paper are plots with a relatively single travel purpose, such as a hospital, railway station, or shopping mall, the drop-off points of such plots are mostly concentrated on the roads along the entrances and exits of buildings. Considering the distance between the parking spot and the facility, the initial Eps values are selected as 50, 75, 100, 120, and 150 meters. For the identification of travel hotspots during a certain period of time (two hours) in a day, the MinPts value should not be too large, and the initial MinPts value range of 50-120 is selected for discussion.

The data set during the morning rush hour on weekdays is used as the input set for cluster analysis. The change of the number of clusters with Eps under different MinPts is shown in Figure 1. When MinPts=50, the number of identified clusters is large. As the Eps value increases, the number of clusters cannot reach a stable state as it continues to increase. This is due to the small value of MinPts. When MinPts=75, there are 8 hotspots at Eps=50 and the number of clusters is relatively stable as Eps increases. When MinPts=85 and 100, the number of clusters changes unstably. Therefore, MinPts=75 is selected. Since the drop-off points in the residential area of the commercial center are relatively scattered, selecting a larger Eps can better identify this type of hotspot, so Eps=100 is selected.

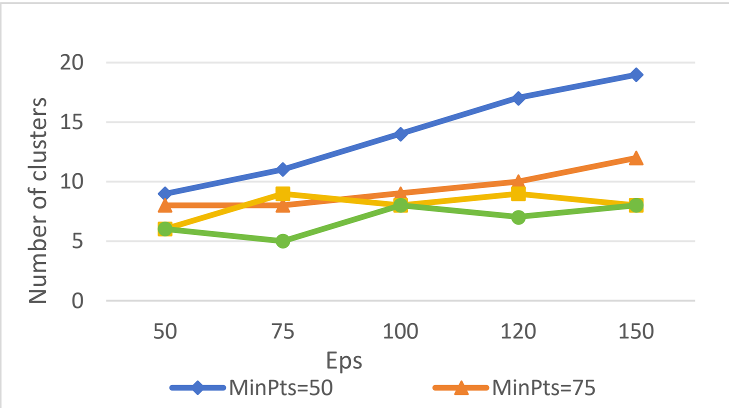
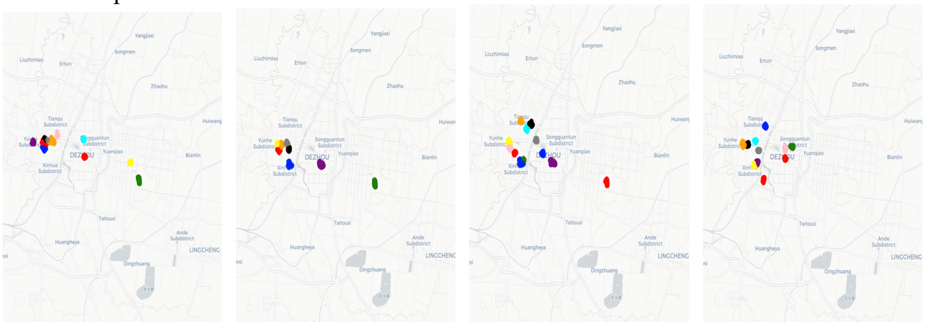


Fig. 1. The number of travel hotspots changes with MinPts and Eps.

The arrival hotspot map and statistical results of each period on weekdays are shown in Figure 2 and Table 1 respectively. The travel hotspots from 7 to 9 in the morning are mainly transportation hubs and medical facilities; from 11 to 13 o'clock in the afternoon, commercial centers gradually become travel hotspots, and medical facilities travel hotspots decrease; from 17 to 21 o'clock in the evening, there are many types of travel. Among the 11 hotspots in this period, there are 4 commercial centers, 3 transportation hubs, 1 residential area, 2 scenic parks, and 1 educational service facility; from 21 to 23 o'clock at night, residential areas become the main travel hotspot.



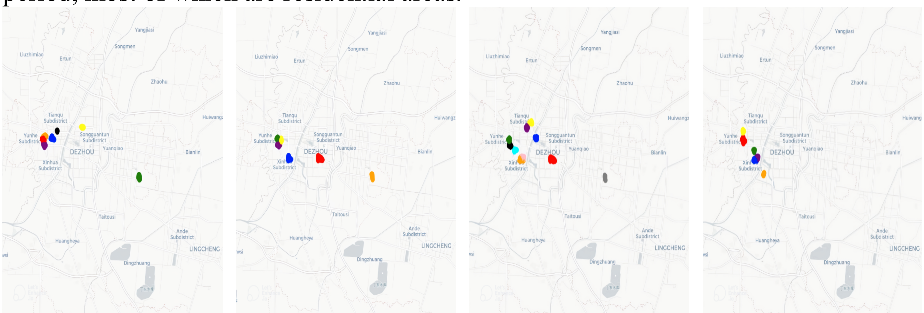
(a) 7：00-9：00 (b) 11：00-13：00 (c) 17：00-19：00 (d) 21：00-23：00

Fig. 2. Travel hotspots at different times on weekdays.

Table 1. Number of travel hotspots at different time periods on weekdays.

time Number of hot spots Facility Type	Transportation hub	Medical Facilities	Business Center	Residential area	Attraction Park	Educational service facilities
7：00-9：00	3	5	0	0	0	1
11：00-13：00	3	2	4	0	0	0
17：00-19：00	3	0	4	1	2	1
21：00-23：00	1	0	0	8	0	1

The arrival hotspot map and statistical results of each period on weekends are shown in Figure 3 and Table 2 respectively. The types of travel hotspots in each period on weekdays and weekends are similar. The main travel hotspots from 7 to 9 in the morning are still transportation hubs and medical facilities, but unlike weekdays, the travel appeal of medical care during this period has declined, and scenic parks and educational service facilities have become new travel hotspots; the travel hotspots from 11 to 13 in the afternoon are still transportation hubs and commercial centers; from 17 to 21 in the evening, the travel hotspots are still commercial centers and transportation hubs; from 21 to 23 at night, there are fewer travel hotspots during this period, most of which are residential areas.



(a) 7：00-9：00 (b) 11：00-13：00 (c) 17：00-19：00 (d) 21：00-23：00

Fig. 3. Travel hotspots at different times on weekends.

Table 2. Travel hotspots at different time periods on weekends.

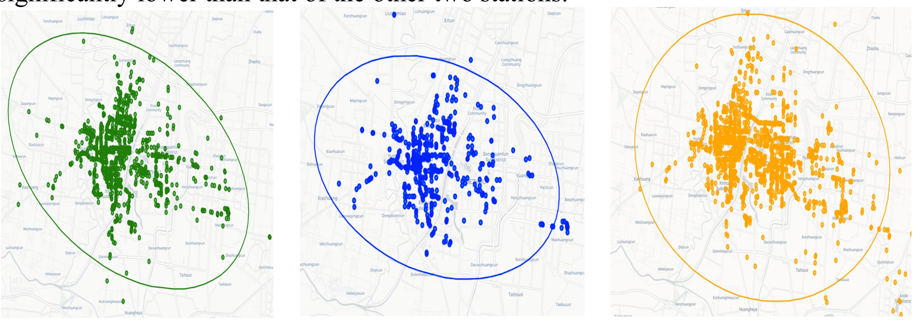
time Number of hot spots Facility Type	Trans-portation hub	Medical Facili-ties	Busi-ness Center	Residen-tial area	Attrac-tion Park	Education-al service facilities
7：00-9：00	3	3	0	0	1	2
11：00-13：00	2	0	3	0	0	0
17：00-19：00	3	0	3	1	1	1
21：00-23：00	1	0	0	4	0	0

3.2 Analysis of the Attraction Range of Travel Hotspots

Through the distribution of travel hotspots, we can see that transportation hubs are continuous hotspots, and commercial centers have relatively good travel attraction on both weekdays and weekends. Therefore, this article will analyze the attraction range of these two facilities.

(1) Transportation hubs

As shown in Figure 4, the travel sources of the three stations are mainly from the western part of Decheng District, and the attraction range of Dezhou Bus Terminal is significantly lower than that of the other two stations.



(a)Dezhou Station (b) Dezhou Bus Terminal (c) Dezhou East Station.

Fig. 4. Comparison of the attraction range of transportation facilities.

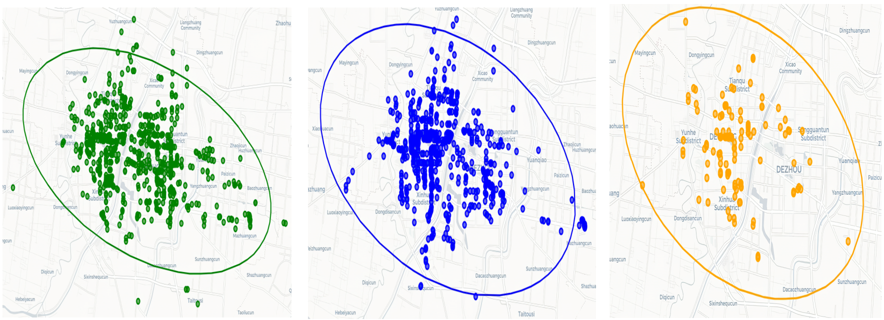
As shown in Table 3, from the perspective of the distance between the long and short axes and the area of the ellipse, the boarding points at Dezhou Station are distributed over a wide range, and there is an obvious elongation trend along the long axis. The boarding points at Dezhou Bus Terminal are relatively concentrated. The boarding points at East Station have the largest distribution range and are relatively evenly distributed, with less influence from a single direction. From the perspective of average travel distance, the average attraction distance between Dezhou Station and Dezhou Bus Terminal is significantly lower than that of Dezhou East Station. This is because Dezhou East Station is located in the east of Decheng District, and most of its travel sources are concentrated in the west of Decheng District, resulting in Its average attraction distance is longer.

Table 3. Parameters of attraction range and average attraction distance of transportation and commercial facilities.

Hotspot Name	Hot period	Minor axis distance (km)	Long axis distance (km)	Rotation angle (degrees)	Ellipse area (km²)	Average attraction distance (km)
Texas Station	7：00-9：00	8.18	14.96	-21.17	384.32	4.7
Dezhou Bus Terminal	7：00-9：00	7.07	12.24	-13.24	271.84	3.6
Dezhou East Station	7：00-9：00	9.97	14.82	-7.79	464.21	10.9
Audley Shopping Center	11：00-13：00	5.53	11.88	-17.11	206.59	4.2
Wanda square	11：00-13：00	6.32	11.46	-17.72	227.49	4.0
Chenghuang Temple Pedestrian Street	11：00-13：00	6.06	10.07	-21.30	191.73	3.3

(2) Commercial Center

From Figure 5 and Table 3, from the perspective of the long and short axis distance and the ellipse area, the boarding points of the Audela Shopping Center are widely distributed in the long axis direction, more concentrated in the short axis direction, and the overall distribution is elongated, with a moderate distribution space range. The boarding points of Wanda Plaza are distributed in a moderate range in the long and short axis directions, and the overall distribution is also elongated, but slightly more dispersed than the Audela Shopping Center, and the spatial range is slightly larger. The boarding points of Chenghuangmiao Pedestrian Street have a relatively small distribution range in the long and short axis directions, and the overall distribution trend is more concentrated, with the smallest spatial range; from the perspective of average attraction distance, Chenghuangmiao Pedestrian Street has the smallest attraction distance, which is due to the small commercial scale of Chenghuangmiao Pedestrian Street and its location in the western part of Decheng District, where residents have a short travel distance.



(a)Audley Shopping Center (b) Wanda square (c) Chenghuang Temple Pedestrian Street

Fig. 5. Comparison of the attraction range of commercial centers

4 CONCLUSIONS AND RECOMMENDATIONS

4.1 Conclusion

This paper uses clustering and standard deviation ellipse to analyze the taxi data of Dezhou City for two weeks, in order to explore the travel characteristics of Dezhou residents and reflect the urban structure layout. The main conclusions are as follows:

(1) From the perspective of travel hotspots, Dezhou City has formed a travel center centered on Decheng District, with a large proportion of travel to transportation facilities. Dezhou Station, Dezhou Bus Terminal and Dezhou East Station all have good travel attractions; but its residents' travel is relatively monotonous, and their taxi destinations are mostly concentrated in transportation facilities, especially in commercial facilities. There are fewer trips, and the evening business is less developed. The business is mainly based on Wanda Plaza and Aodele Shopping Plaza, and there are fewer commercial centers.

(2) From the perspective of hotspot attraction range, most travel sources are located in the western part of Decheng District, and there are obvious differences between the east and the west. This may be caused by reasons such as urban layout and economic conditions. The average attraction distance of each hotspot is small, reflecting that the layout of the main urban facilities is relatively concentrated.

4.2 Recommendations

Based on the above conclusions, the following suggestions are put forward with the goal of improving the future development of Dezhou City:

(1) From the perspective of geographical location, Decheng District is currently the central urban area for travel. It is necessary to give full play to its leading role, radiate and drive the economic development of other districts and counties, and form a spatial pattern with Decheng District as the core and point-to-surface. There may be a phenomenon of "rich in the west and poor in the east" in Decheng District. It is necessary to consider balancing the economic gap between the two sides to avoid an obvious "one-sided" situation.

(2) From the perspective of facilities, for transportation facilities, Dezhou Station and Dezhou East Station need more extensive traffic management measures and resource investment to ensure that they meet the needs of most boarding points. Dezhou Bus Terminal can adopt a centralized management method to meet the transportation needs of local areas; from the perspective of commercial facilities, commercial facilities are relatively scarce, especially at night. Most residents travel to residential areas. Compared with more developed cities, there is a lack of large-scale comprehensive commercial centers.

REFERENCES

1. Wu Jiayi. Research on urban spatial structure information mining based on floating vehicle data and travel activity perspective[D]. Nanjing University,2018
2. Zhang Xianyuan. Analysis of spatiotemporal characteristics of urban residents' travel based on taxi data[D]. Wuhan University, 2020.
3. Wei Haitao, Xiao Tiancong, Hu Baosheng, et al. Analysis of residents' travel behavior and urban spatial structure based on taxi GPS data—taking Beijing as an example[J]. Shanghai Urban Planning, 2020(1): 69-76
4. Liu Yu, Zhan Zhaohui, Zhu Di, et al. Integrating multi-source geographic big data to perceive urban spatial heterogeneity[J]. Journal of Wuhan University (Information Science Edition), 2018, 43(3): 327-335
5. Chen X J, Wang Y, Xie J, et al. Urban hotspots detection of taxi stops with local maximum density[J]. Computers, Environment and Urban Systems, 2021, 89: 101661.
6. Zheng W, Cheng Y, Li N, et al. An Improved DBSCAN Algorithm to Analyze Taxi Pick-Up Hotspots[C]//International Conference on 5G for Future Wireless Networks. Cham: Springer Nature Switzerland, 2022: 180-195.
7. Xia Jiaqi. Analysis of spatiotemporal characteristics of urban residents' travel and mining of spatial structure information based on taxi data[D]. Qingdao University of Technology, 2022. DOI: 10.27263/d.cnki.gqudc.2022.000543.
8. Liu Y, Singleton A, Arribas-Bel D, et al. Identifying and understanding road-constrained areas of interest (AOIs) through spatiotemporal taxi GPS data: A case study in New York City[J]. Computers, Environment and Urban Systems, 2021, 86: 101592.
9. Angkhawey U, Muangsin V. Detecting points of interest in a city from taxi gps with adaptive dbscan[C]//2018 Seventh ICT International Student Project Conference (ICT-ISPC). IEEE, 2018: 1-6.