World Scientific Research Journal (WSRJ) ISSN: 2472-3703
www.wsr-j.org

A Visual Analysis of User Behavior Based on the Trajectory Data of "DiDi" Orders

Dan Dou[1]

[1]School of Economics and Management, XiDian University, Xi'an, China

---

Abstract: Based on the trajectory data of "didi chuxing" orders, this paper makes a visual analysis of the trip rules of "didi" users. First, a descriptive statistical analysis of the trip timing characteristics of "didi" users was did; Secondly, the human segmentation model was proposed based on the data of point of interest, and DBSCAN was used to cluster the travel hot spots of different populations, so as to discover the differences in travel behavior. Finally, according to the clustering results of DBSCAN algorithm, visualization is implemented in ARCGIS software, so that the behavioral characteristics of "didi" users can be analyzed more intuitively, which is of certain theoretical and practical value to reveal the spatio-temporal interaction pattern of "didi" users and make relevant auxiliary decision in chengdu.
Keywords: User behavior; DBSCAN; visualization.

---

1. INTRODUCTION

The continuous development of mobile positioning technology has led to the widespread use of various smart mobile APPs that can record the real-time location of users, making it easier to obtain the resulting spatial trajectory data. The mining of spatial trajectory data has been widely used in urban planning, occupational and housing balance, and service provider location. Mining time and space analysis of user trajectory data has become a research hotspot of urban intelligent layout.

At present, many scholars have used the big data obtained by mobile positioning to conduct a large number of studies on the behaviour of human mobility. Zhong.G et al. extracted passenger travel information through mobile phone mobile data and analyzed the difference in population movements inside and outside the city center [1]. Wang Xianwen et al [2-4] extracted the starting and ending point data for the study of inter-regional population mobility. Some scholars based on various trajectory data, through different clustering algorithms to explore the popular routes and hotspots of residents [5-7]. A small number of scholars conducted gps tracking of the respondents through volunteers, and joined the research on travel destinations, so that the study of human travel patterns has entered the microscopic field [8-9]. In terms of commuting behavior, many scholars use the questionnaire, mobile phone signaling data or gps data to extract commuter trajectories and analyze the city's commuting behavior

and occupational relationship [10-11]. In terms of urban spatial layout, some scholars have explored the rationality of urban commercial land and urban functional areas [12-13]. Considering that the trajectory data only records the itinerary, but the semantic information is lost, the existing research has added interest point data, and carried out a more specific and in-depth study on user behavior.

The existing researches are all about the travel rules and hotspot functional studies of residents and commuters from a macro perspective. However, due to the low accuracy of data acquisition obtained by most studies, the travel destinations of travellers are less explored from a microscopic perspective. This study analyzes user travel timing characteristics with “DiDikchuxing” order data with high-precision positioning technology and closer to the user's destination. The user population is divided by the point of interest data, and the DBSCAN algorithm is used to cluster the hotspots of different groups of people, so as to analyze the behavior characteristics of different groups. The research results provide methods for the spatial planning of Chengdu, the location of service providers, and the decision-making of “DiDi”make orders.

2. DATA AND METHODS

2.1 Data sources

The data of this study comes from the driver order data of the “DiDi” platform in the Second Ring Road area of Chengdu and the point of interest data obtained from the Gaode Map API.The “DiDi” order data area includes the Jinniu District in the economic zone of the city center, the Qingyang District of the political and cultural center, and the prosperous areas of Chenghua District and Jinjiang District, which belong to the five central areas of Chengdu. This article obtained the “DiDi” full order data [14] in the above region in 2016 (November 1st to November 30th, total 30 days). The data includes information such as the order ID, the latitude and longitude of the board, and the boarding time. POI data is obtained from the official open API of Gaode Map according to the research scope, and mainly includes information such as the name, address, category, spatial coordinates, boundary coordinates and gate coordinates of the interest point. The original order data has an average daily order volume of about 236,000. The original data is weighted and deleted, and the remaining orders are about 200,000 per day.

2.2 "Didi" user travel behavior analysis method

2.2.1 DBSCAN spatial clustering algorithm

The “DiDi” user hotspot area cluster analysis is based on the “DiDi” user's getting on and off point data, based on the distance, extracting the “DiDi” user's orders with a relatively large number of orders. In this paper, density-based spatial clustering algorithm (DBSCAN) is used for hotspot clustering. The algorithm can filter out low-density regions, and find spatial clusters of arbitrary shapes without being affected by noise points, which is in line with the characteristics of spatial hotspots. The traditional DBSCAN algorithm needs to determine the radius threshold eps and the density threshold minpts parameters autonomously. These two

parameters are sensitive to the clustering results. Scholars can only determine the parameters through continuous testing. In this paper, a DBSCAN algorithm for adaptively determining parameters is used to determine two parameters by k-distance and mathematical expectation [15], respectively, so that the clustering results are more accurate.

The specific steps of DBSCAN algorithm parameter optimization are as follows:

(1) Calculate the K-nearest distance matrix by Euclidean distance.

(2) Calculate eps. Extract the 4th closest distance of each object in the k-nearest distance matrix and arrange it in ascending order, and make the k-dist curve. The point where the curvature of the curve is abrupt may be considered as a more suitable eps value.

(3) Calculate the minpts value. According to the eps value calculated in second step, count the number of points in each eps range from the other boarding point, and use the mathematical expectation method to find the minpts threshold parameter. The calculation formula is like that:

$$\text{minpts} = \frac{1}{n}\sum_{i=1}^{n} p_i$$

Where pi is the number of points in the eps field. Due to the different distribution density of orders in different time periods, this paper uses different threshold parameters for cluster analysis in different time windows.

2.2.2 Crowd partitioning model

Since “DiDi” belongs to the network car, the passengers let the drip driver pick up the order by specifying the starting place, which makes the order data closer to the position of the passenger. Therefore, the basis of the division of crowd is the Drip travel positioning which accuracy and proximity.

This paper mainly studies the travel behaviors of college students, office workers and residents. Due to the large area of the school and residential area, the “DiDi” driver can directly enter the corresponding area to receive orders, so this article regards the 150-meter order in the school area and near the main entrance as the student order. Extract "DiDi" order data for residential areas. The division of office workers is more complicated and is discussed in several situations. For corporate and building interest points, this article takes orders from within 150 meters of the main entrance of the enterprise as orders for office workers at such points of interest; secondly, the interest points of industrial parks, due to the large area, the model is similar of Residential area interest points.

3. "DIDI" USER TRAVEL BEHAVIOR ANALYSIS

3.1 The Descriptive statistical analysis of “DiDi” user travel behavior

The statistical analysis of the “DiDi” user travel volume, travel time, and travel time Length ia conducted. The statistical analysis of the “DiDi” user travel volume, travel time, and travel time Length ia conducted .The Statistics on the distribution of weekly trips during the month shows that users have relatively few rides on Monday, and are gradually increasing on Tuesday,

Wednesday, Thursday, and the number of orders reached the maximum on Friday and Saturday. Secondly, the travel rules of the “DiDi” users in each time period were analyzed, and the comparison was found during the period from Monday to Friday. The order quantity is obviously more than other days. The regularity of travel on Friday is basically similar from other time points from 0:00 to 18:00, and the order peak period is reached at 8-10, 13-15, and 17-18 in the morning.After 18 o'clock, orders from Monday to Thursday gradually declined, while the number of users who took the "DiDi" on Friday dropped significantly between 18 and 19 o'clock, and then increased again, until the order volume began to expedite decline after 21 o'clock. In order to make it easier to compare the difference between working days and rest days, this time I chose Monday travel as a representative of working days. Comparing the folds on Mondays and weekends, you can see that orders from 0:00 to 4:00 have a certain increase over the working day. The comparison shows that it is different from Monday's change. The increase from 6 to 10 on weekends is relatively slow until 14 o'clock. When the maximum value is reached, it can be seen that due to more free time on weekends, people's travel time is delayed; as with the working day, the number of breaks at 17 o'clock has reached a late peak, and then the order volume is gradually reduced. Overall, the travel volume has increased during the period from 21 to -22 pm. Finally, the user's daily travel time is statistically known: the travel time is between 13-18 minutes. It is calculated that the average user's ride time of the "Drip" user is 22 minutes, which means that the users who travel by "Drips" generally travel for short distances, and the travel cost is not too high, which is in line with the general user's taxi behavior. Compared with other days, orders for Friday and Saturday with a time of 23 minutes or more have increased significantly. It may be due to the lack of leisure time on weekends, and people will not care too much about the length of the ride.

3.2 "DiDi" user travel hotspot area visualization

3.2.1 The Visualization of hot spots in college students

First we analyze the travel situation of colleges. Analysis results show that the number of boarding per day in colleges is always less than the number of Departure. The number of orders for getting on and off from Monday to Thursday has not changed much. The number of orders on the Friday and Saturday is higher. The number of getting off orders on Saturdays and Sundays every week is much higher than the number of getting on orders placed. The possible explanation for the above phenomenon is that college students need to take classes during the week, and the amount of travel orders is relatively small. On Friday, college students begin to be active, and college students need to go out to relax or go home. Back home from school on Saturday and Sunday, or returning from the trip, led to a sharp increase in the number of passengers.

As can be seen from the Analysis results, there is no fixed time period for college trips. From 8:00 to 22:00, the order volume has changed little and the order volume is at a high level. the order for getting off the train has a certain regularity. At 9 13-14, 17-18, the overall order volume reached a peak. Orders for returning to school from 19:00 pm to 2:00 pm on Friday have been more frequent, especially at 21:00 and 22:00, orders are significantly higher than the

other four days. On Saturdays and Sundays, orders at 8:00, 14:00, and 17:00 were the largest. Overall, orders from 12 to 17 were at a relatively high level, and orders at 19-22 were relatively high.

According to the travel time analysis results, the time zones of 0, 6-9 and 22 are selected to cluster and visualize the hotspots in colleges.

The picture below shows the order clustering of the getting off the trains from the college at 0. Combined with the data of points of interest, there are many hotspots showing that during the 0-1 period, college students mainly go home from three types of leisure and entertainment places, such as Internet cafes, KTV and bars. Some users will also go from restaurants and studios, shopping malls and other places; through the clustering of orders from colleges and universities on the 0 point, most users are destined for residential areas, and some are hotels. Therefore, the "DiDi" driver can pick up the order closer to his place of residence after 0 o'clock.

The most important places for college students to travel at 6-9 points are clusters in the city center (such as Chengdu 339, Chunxi Road, Yipin Tianxia, Hummer City, etc.), passenger stations and major hospitals in the city. This shows that if the “DiDi” driver is close to the university, he or she can go to the university to take the order without going directly to the city center.

Get off at 0 o'clock Get on at 6-9 o'clock Get off at 22 o'clock

Figure 1. Distribution of hotspots at 0, 6-9, and 22 in colleges

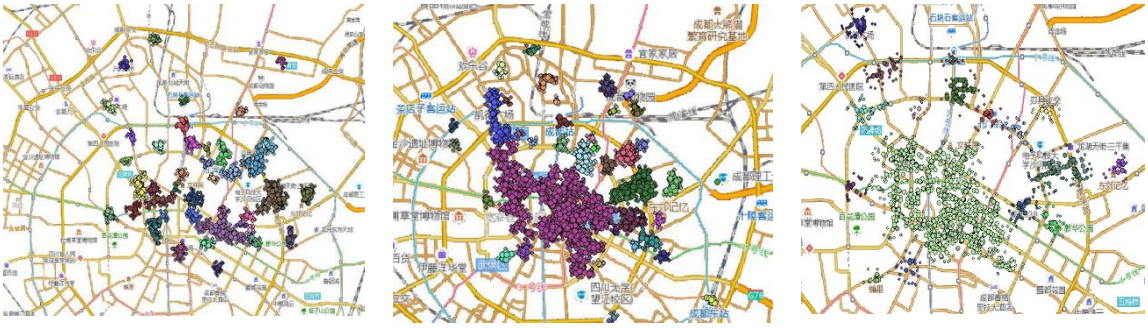
College students at 22 o'clock travel destinations are mainly subway, commercial and nearby residential areas. Most students are ready to go home and a few choose to go shopping. At 22 o'clock, colleges and universities mainly return to major shopping malls and commercial streets, and a small number of them return to schools from residential areas, passenger stations and scenic spots. It can be seen that college students prefer to go shopping and enjoy nightlife.

3.2.2 The Visualization of residents' travel hotspots

First of all, this article makes descriptive statistics on residents' travel orders. From the results, it is known that the distribution of orders on the daily basis of residents is similar to that of the overall users, but the 9-point, 13-point, and 17-point travels are larger than other time periods. The order quantity for getting off has gradually increased from 9:00 to 18:00, and the order

volume has reached the maximum until 17:00 to 18:00. It may be the result of a resident returning home by car after a day's work.

As shown in Figure 3.2.2.1, the residents of the “DiDi” users mainly go to hospitals and vehicles Stations, subway stations, office buildings, and various business districts at 9:00. The 13-point resident travel destination is roughly the same as 9:00. At 17 o'clock, residents will go to various squares, snack street, clothing market and other shopping malls, and some residents will return to school, commercial and residential companies. Residents at 9 o'clock mainly from the hospital, office buildings (may be night shifts) back to their place of residence. At 13 o'clock, residents mainly returned to residential areas from shopping malls, hospitals, office buildings and schools. At 17 o'clock, residents returned to the residential area from hospitals, scenic spots, business districts, and commercial streets. At this time,the number of people returning from the hospital was the highest. The “DiDi” driver could take orders near the hospital at this time.



9点下车　　13点上车　　17点下车

Figure 2. Distribution of hotspots at 9, 13 and 17 inhabitants

3.2.3 The Visualization of office workers' travel hotspots

First, the amount of orders for each time period of the office workers is counted. According to the statistical results, the number of commuters getting on the train is slightly more than the number of people getting off the bus, and the order volume is the highest on Friday. For the company's traffic volume at various time periods, the amount of travel at 9:00 and 11:00 in the week is slightly higher, and the amount of travel is higher at 13:00 to 14:00, 17:00 and 21:00. The amount of travel on Saturday reached a maximum at 17 o'clock. Orders on Sunday increased slowly by 14 points to reach the maximum and then slowly declined without much ups and downs. The distribution of vehicles in each period of the company during the week is similar to the overall law. Compared with other days, the arrival amount at 19:00 on Friday is slightly higher.

Get on the car at 9:00 Get off the train at 8:00 Get on the train at 17:00

Figure 3.The Distribution of office workers'hotspots at 8, 9 and 17

The clustering results show that employees may go to business hotels, subway stations, stations and other places to pick up and drop off customers at 9:00 in the morning. Some trade unions go to the raw material procurement mall, the tax bureau, the Geological Survey Bureau, etc. Some employees return to residential areas and schools. The reason may be that they are on irregular classes. Most of the 8 o'clock workers return to the company from the residential area, and some are from the hotel. At 17 o'clock in the afternoon, employees may need to rest for a day, so most of the employees returned to their places of residence, while the downtown areas such as Chengdu 339, Kaide Plaza and Chunxi Road are still hot spots, especially the snack city and hot pot restaurant. Such dining places are hot spots for employees to choose.

This study visually analyzes the spatio-temporal behavior of 'DIDI' users by combining specific actual data. According to the cluster analysis of hotspots in three groups of people, the following conclusions can be drawn:

(1) College students are free to act. From 9:00 to 2:00 pm, there are always have orders. The "DiDi" driver can take orders at a university closer to him, without having to go to the city center to take orders.

(2) Hospitals and shopping malls are hot spots for residents to choose, so hospitals and famous commercial streets are the must-have areas for "DiDi" drivers.

(3) From the starting point and the destination hotspot area of the office workers, it can be seen that most of the office workers' work places and residences are in the second ring, but a small number of employees work out the second ring. Therefore, urban planners can better achieve the balance of occupation and residence in Chengdu by developing the suburban economy.

4. CONCLUSION

Although this research has certain theoretical contributions and practical significance, it also has some shortcomings.

First of all, in terms of data, because the "DiDi" platform only provides the order trajectory data of the local area of the second ring of Chengdu, the behavior of different groups of people outside the second ring can not be distinguished from the behavior of the bustling city center users, and cannot be provide good idear of the second ring of the city'construction. Secondly,

due to the irregularity of the boundary of interest points, the order extraction of students and residents contains a small amount of order data that does not belong to this group of people, but because of the destination cluster analysis, a small part of the order data will not have a large impact on the results. Future research can make a more detailed exploration of the travel rules of “DiDi” users in different points of interest, such as catering, scenic spots, entertainment places, etc., and provide more accurate theoretical basis and practical significance for the development of these industries.

REFERENCES

[1] Zhong G, Yin T, Zhang J, et al. Characteristics analysis for travel behavior of transportation hub passengers using mobile phone data[J]. Transportation, 2018(3):1-24.

[2] Wang Xianwen, Wang Hongyin. The Study on short-term population flow of Beijing-Tianjin-Hebei urban agglomeration based on geographic location data [J].Journal of Dalian University of Technology (Social Sciences Edition), 2017, 38 (2):105-113.

[3] Jin C, Chen Y, Yang M. A Multi-scale Visualization Method for the Trajectory Origin-Destination Data [J]. Journal of Geo-Information Science, 2017.

[4] Zhang WS, Li SJ, Pan G. Mining the semantics of origin-destination flows using taxi traces[C]// ACM SIGKDD Conference on Ubiquitous Computing. ACM, 2012: 943-949.

[5] Xianyu J, Rasouli S, Timmermans H. Analysis of variability in multi-day GPS imputed activity-travel diaries using multi-dimensional sequence alignment and panel effects regression models[J]. Transportation, 2017, 44(3):533-553.

[6] Zhou M, Wang D. Investigating inter-generational changes in activity-travel behavior: a disaggregate approach [J]. Transportation, 2018:1-45.

[7] Sun GD, Zhang B, Liu YQ，eta1．Taxi Hot Area Function Discovery Based on Passenger Data [J]. Computer Engineering, 2017, 43(5): 16-22.

[8] Krueger R, Thom D, Ertl T. Visual Analysis of Movement Behavior Using Web Data for Context Enrichment[C]// Visualization Symposium. IEEE, 2014:193-200.

[9] Jia Xiaoxia. The Visualization of User Behavior Analysis Based on Electric Bicycle Data [D]. Zhejiang University, 2017.

[10] Fu Xin, Sun Maopeng, Sun Wei. A Analysis of Time and Space Characteristics of Taxi Commuting Recognition Based on GPS Data [J]. China Journal of Highway, 2017, 30(7): 134-143.

[11] Tian Jinling, Wang De, Xie Dongcan, et al. A Analysis of commuting characteristics and model summary of typical employment areas in Shanghai —— Comparison of cases between Zhangjiang, Jinqiao and Lujiazui [J]. GEOGRAPHICAL RESEARCH, 2017, 36(1): 134-148.

[12] Yuan J, Zheng Y, Xie X. Discovering regions of different functions in a city using human mobility and POIs[C]// ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2012:186-194.

[13] Pan G, Qi G, Wu Z, et al. Land-Use Classification Using Taxi GPS Traces [J]. IEEE

Transactions on Intelligent Transportation Systems, 2013, 14(1):113-123.
[14] Data source: https://gaia.didichuxing.com.
[15] Zhou Hongfang, Wang Peng. The Research on Parameter Adaptive Determination Method in DBSCAN Algorithm [J]. Journal of Xi'an University of Technology, 2012, 28(3): 289-292.