Revealing the Characteristics of Active Area in the City by Taxi GPS Data

A Study of Shenzhen, China

Lu Liu
School of Urban Planning and Management Harbin Institute of Technology, Shenzhen Graduate School
Center for Assessment and Development of Real Estate
Shenzhen, China
e-mail: luliu.wh@gmail.com

Lin Tang
Center for Assessment and Development of Real Estate
Shenzhen, China
e-mail: tanglin@139.com

***Abstract*—Taxi serves as a main transportation means for urban residents. The driving behavior recorded by Taxi GPS data can be a mirror to urban life. Therefore, this study adopts taxi GPS data to extract active area of the city and examine the weekday and weekend discrepancy. Moreover, the point of interests (POI) is combined to reveal the characteristics of the city. The method proposed in this study can be efficiently used to study the structure of the city and explore the living style of city residents, which provides valuable information for urban planners and policy makers to plan for public service, do spatial plan of urban land, and optimize urban function.**

Keywords-taxi GPS; point of interest; active area

I. INTRODUCTION

Taxi serves as a main transportation means for urban residents, People use taxis to commute to and from work in weekdays, and go leisure places in weekends and holidays. Therefore, the record of taxi traces can be an ideal reflection of urban activities. In recent years, GPS devices has been equipped on taxis in the city, and their traces are recorded in real time. These traces data has already been used in various applications to examine every aspects of urban life [1, 2]. Castro, Zhang, Chen, Li and Pan [3] has made a survey on the applications of adopting GPS equipped taxi data in studying social and community dynamics.

Especially, taxi GPS data can be adopted in many applications to delineate the function areas of the city [4-7]. In a metro city, the urban function in a region is usually mixed, for instance, a district in the city can be residential center and commercial center at the same time. Meanwhile, the function of a city is changing with time. In this case, the traditional data and method, such as the land use map, can neither give an accurate answer on urban function, nor monitor its change. In comparison, the taxi GPS data has obvious advantage in examining the urban function using real time data, and monitoring the changing information as urban evolve with time. Liu, Wang, Xiao and Gao [8] has used taxi GPS data of Shanghai in China to reveal intraurban land use variation, and successfully divided the city into six kinds of 'source-sink area'. Mazimpaka and Timpf [9] has combined taxi GPS data with Flickr data to discover the function of urban areas. Zhou, Fang, Thill, Li and Li [10] has further connected the urban function region with people's activities by using such taxi traces data.

The travel activities in weekdays are very different from that in weekends and holidays, and such discrepancy can reflect the unique characters of a city. However, in the existing studies, this information has not been thoroughly studied. Therefore, in this study, we intend to use taxi GPS data in examining the weekday and weekend discrepancy, and use the point of interests (POI) of a city as assistant data to reveal the characteristics of the city.

The paper is arranged as follows, first, the datasets and analyzing methods are introduced in section 2. Then the results are displayed and interpreted in section 3. A conclusion is given in section 4.

II. DATA AND METHODOLOGY

A. Data

In recent years, the emergence of various big data has changed every aspects of urban life. Big data is generally defined from three perspective. First, it has tremendous volume which cannot be processed by traditional methodology. Second, it has high diversity with numerous attributes and complex data structure. Third, the data is produced with high temporal frequency. Taxi GPS data is such a kind of big data. A key technique of utilizing big data is to process and simplify the datasets based on the application.

The original GPS data record every pick-up and drop-off behavior of 14000 taxi and the corresponding geographical location in Shenzhen of the year 2014. Based on the objective of this study, the community grid of Shenzhen is used to record the counts of these behavior. Correlation analysis of pick-up counts and drop-off counts shows that, they are highly related with the r square of 0.946, therefore, only the pick-up counts are used for each grid in the following analysis.

These pick-up counts are recorded daily. Fig. 1 shows the fluctuation of the counts all around the year for ten communities with largest yearly pick-up counts. Such fluctuation is mainly caused by the change of urban activities in weekdays, weekends and holidays, and the understanding of such change may provide valuable information on the characters of the city.

978-1-5090-3619-6/17/$31.00 ©2017 IEEE

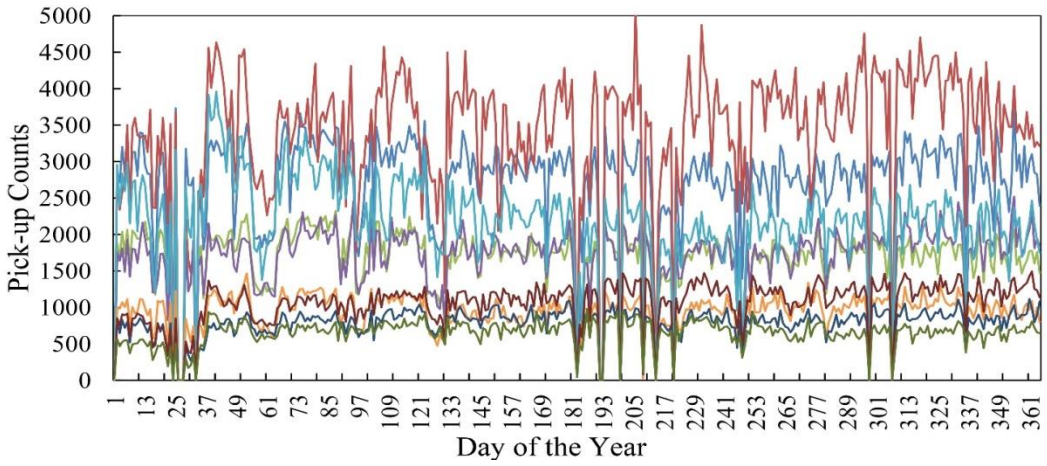

Figure 1. Flucuation of pick-up count over the year of 2014 for ten community grids with maximum yearly pick up counts.

To study travelling pattern of weekdays and weekends, the special days are first excluded. The holidays such as mid-autumn festival and dragon boat festival, and their related weekends are first eliminated. Then, for 20 communities with largest yearly pick-up counts, the outliers which are not within 85% value of the whole sample in a year are found for each community. The related dates are then excluded.

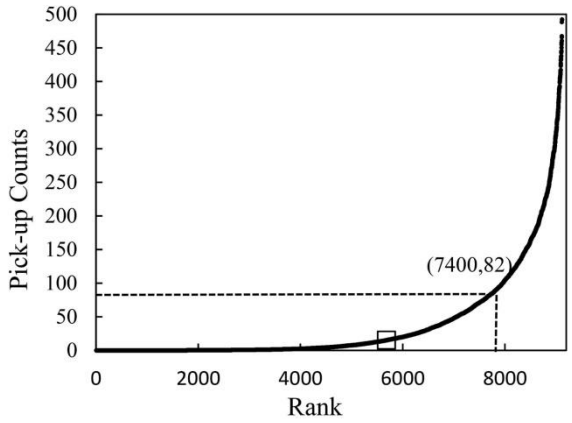

Figure 2. The relation of the pick-up counts and related rank for each community grid.

The remained data are divided into weekdays and weekends. The mean pick-up counts and mean drop-off counts are calculated individually for each community.

The POI used in the study is extracted from Baidu Map in the year 2014, the POI are classified into ten classes, and six of them are chosen to study the characters of the active area of the city, including recreation facility, education facility, physical facility, bus station, monetary facility and medical facility. The POI intensity of each class is calculated by dividing the POI counts in the grid by its area of lands for building.

B. Study Area

Shenzhen is a major city in Guangdong Province. It is located in the south China, and is neighboring to Hong Kong. In the past decades, Shenzhen has experienced fast economic development, and it becomes a main financial center in China. This city has a population of over 15 million in the year 2014, but the average age is only around 30. It is a young city not only because the low average age, but also because it starts its urbanization from 1980, after it was set as a special economic zone of China.

C. Methodology

Three steps are used to explore the characteristics of the active area of Shenzhen.

First, the active area is decided by the yearly accounts for each community. As shown in Fig. 2, a sudden change point can be found for the curve between the average pick-up volume and its related rank, which is decided by a sudden change of slope. The average pick-up volume of 82 is finally selected as the Threshold.

Second, a weekday-weekend (DE) index is designed to represents the discrepancy of travel activities in weekdays and weekends. The index can be calculated as follows:

$$DE^i = Counts^i_{weekend} / Counts^i_{weekday} \quad (1)$$

where i represents for each community grid, $Counts_{weekend}$ represents for the summed pick-up counts of the community grid, and $Counts_{weekday}$ represents for the summed pick-up counts of the community grid. The larger the value is, the more people are concentrated at the place during weekends, and the smaller the value is, the more people are concentrated at the place during weekdays. A value approaching 1 means that this place is popular both on weekdays and on weekends.

Third, for each level of DE value, the intensity of POI is calculated and analyzed.

III. Result and Intepretation

A. The Active Area of the City

As shown in Fig. 3, in 9192 community grids in Shenzhen, 1792 of them with are selected as the active area of the city. The large urban parks and beauty spot are

excluded based on the city map, since these places are long known to be hot in weekends with few or no other facilities, and the results may have bias if they are included in this research. Finally, 1492 community grids are used in the following analysis. A big daily pick-up volume represents that these areas are more active in the city than other places. In other words, these places bear more functions than other regions of the city, and are the main city space for people to live.

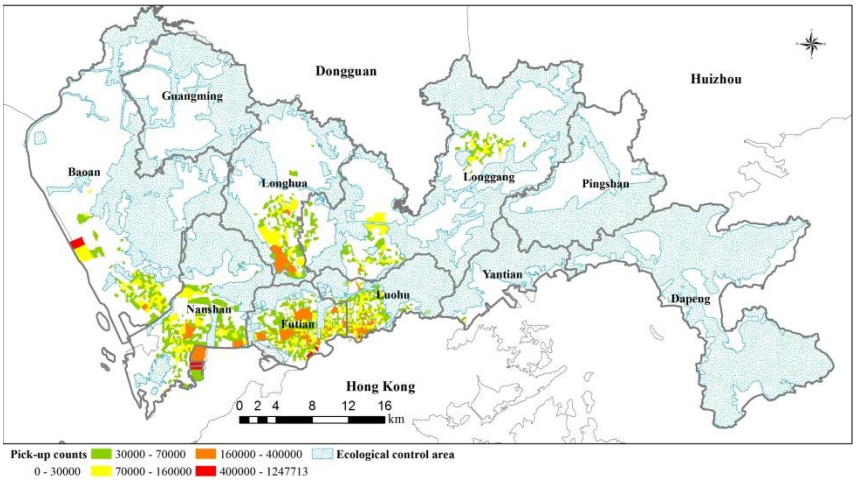

Figure 3. The active area of the city and its pick up counts.

B. *The DE Value and Level*

The DE value of each community grids are changing over space, which indicates that, the weekend-weekday travelling pattern are changing over spaces. This variation is caused by the discrepancy in urban function of each community. Studying the relationship between such driving pattern and the urban function can reveal the living style of the residents in a city.

To divide the data into different levels, the DE value of all samples in active areas are analyzed by normal distribution, as shown in Fig. 4. The mean value of DE is 1.1488, which indicates that generally, people travel more in weekends than in weekdays. By using the standard deviation, the active area of Shenzhen has been divided into three classes. When the DE value is smaller than 0.9493, the community grid is set to class 1, representing that people are more active in weekdays than in weekends. When the DE value is larger than 1.3483, the community grid is set to class 3, representing that people are more active in weekends than in weekdays. Other grids are set to class 2, representing that people are active there both on weekdays and weekends. Finally, 177 grids are classified to class 1, 1158 grids are classified to class 2, and 157 grids are classified to class 3.

C. *Characteristics of Active Area in the City*

For each class of the community grid, the intensity of POI is plotted by a boxing graph, as shown in Fig. 5. The results show that, generally, the intensity of facilities are highest for class 2. The high density of various facilities indicates that, the function of these places are very mixed used, and are more used than other places of the city. The activity of living, working and amuse are going on in the same grid.

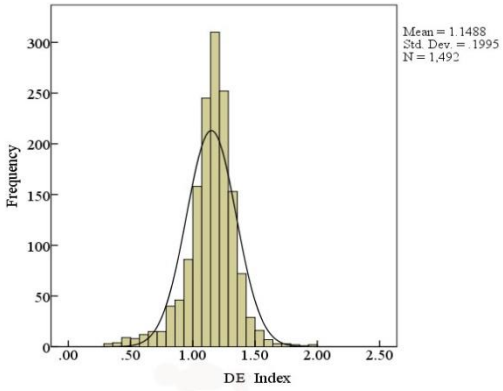

Figure 4. The normal distribution of DE value.

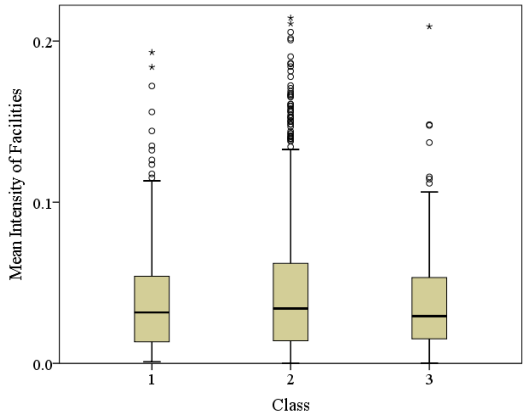

Figure 5. The mean intensity of POI for three classes of community grid.

Patterns for each class of POI point are then further explored, as shown in Fig. 6.

For community grid with class 1, the intensity of recreation facilities is low, but the intensity of monetary facility is high. This result first proved that Shenzhen is a very young city. Since young people who counts for most of the city population are likely to commuting working places and home on weekdays, the weekday active places have less recreation facilities. On the other hand, this result reveals the important role of bank business in the city. The monetary facilities includes bank, insurance, and every kinds of finance company, people are doing their job in these institutions.

For community grid with class 2, the result shows that except monetary facilities, all other facilities have the highest density among all three classes. As discussed before, these regions are very mixed used in the perspective of urban functions.

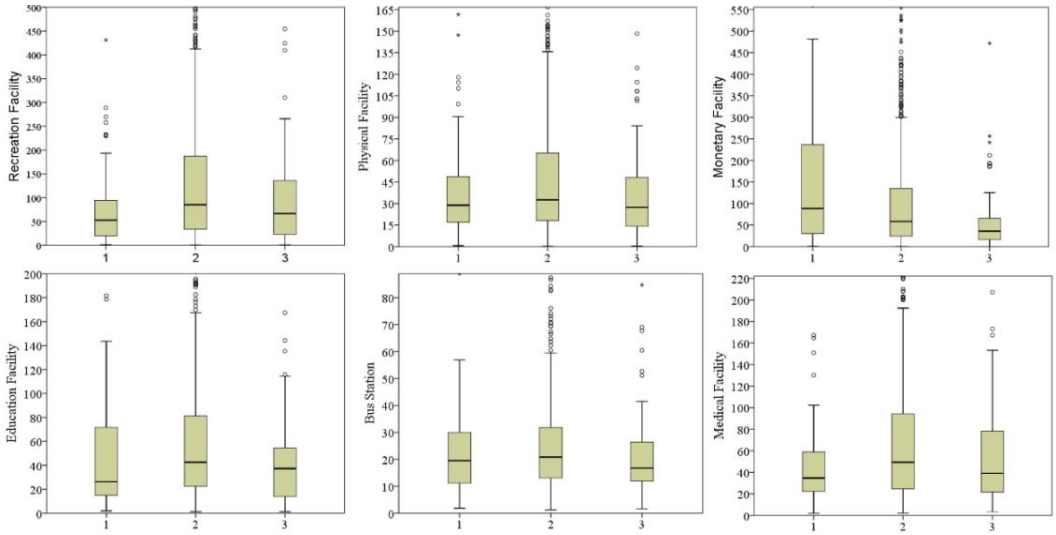

Figure 6. The density of POI for 3 classes of community grid.

For community grid with class 3, it could be seen that, the density of recreation facilities are higher than those in class 1, however, the density of physical facilities are almost equal to that of class 1. This result indicates that, the city residents go to recreation sites more in weekend than in weekdays, but they choose to do sports in weekend as much as in weekdays. For density of medical facilities, the class 1 are almost equal to class 3, indicating that, people choose to see doctor both in weekday and in weekend. Moreover, it could be seen that, community grid with class 3 has obviously higher intensity of education facilities. The underlying cause of this phenomenon may be that, children in the city go to various education institutions on weekends for extracurricular courses by taking a taxi, and in weekdays, they are more likely to walk to a school near their home.

IV. Discussion and Conclusion

In this study, taxi GPS data is used to examine the weekday and weekend discrepancy. Meanwhile the POIs of a city have been used to reveal the characteristics of the city.

First, the active areas of the city are selected from the datasets. An index DE is then designed to study the discrepancy of weekday and weekend. Thereafter, the DE values of the community grids of active area are classified into three classes, the intensity of POI for grids in each classes are obtained and compared to reveal the characters of the city. The results indicates that, as a financial center, people in Shenzhen are active in places with monetary facilities in weekdays. The results also reveals the living style of Shenzhen, people are more prone to go leisure places in weekends, but they choose to do sports and see doctor both in weekdays and weekends. Moreover, the places with high density of various facilities are used in a more balanced style, since people are active in these places both on weekdays and weekends.

This study provides a method to study the structure of the city, and explore the living style of city residents, which provide valuable information for urban planners and policy makers to plan for public service, do spatial plan of urban land and so on.

In the future work, the temporal change of the DE value can analyzed and compared with the change of POI in the city, thus to reveal the evolution of a city. Meanwhile, the POI can be classified more specifically to describe the urban function.

References

[1] Tang, J., Liu, F., Wang, Y., and Wang, H.: 'Uncovering urban human mobility from large scale taxi GPS data', Physica A: Statistical Mechanics and its Applications, 2015, 438, pp. 140-153

[2] Gao, M., Zhu, T., Wan, X., and Wang, Q.: 'Analysis of Travel Time Patterns in Urban Using Taxi GPS Data', in Editor (Ed.)^(Eds.): 'Book Analysis of Travel Time Patterns in Urban Using Taxi GPS Data' (2013, edn.), pp. 512-517

[3] Castro, P.S., Zhang, D., Chen, C., Li, S., and Pan, G.: 'From taxi GPS traces to social and community dynamics: A survey', ACM Comput. Surv., 2013, 46, (2), pp. 1-34

[4] Castro, P. S., Zhang, D., and Li, S.: 'Urban Traffic Modelling and Prediction Using Large Scale Taxi GPS Traces', in Kay, J., Lukowicz, P., Tokuda, H., Olivier, P., and Krüger, A. (Eds.): 'Pervasive Computing: 10th International Conference, Pervasive 2012, Newcastle, UK, June 18-22, 2012. Proceedings' (Springer Berlin Heidelberg, 2012), pp. 57-72

[5] Pan, G., Qi, G., Wu, Z., Zhang, D., and Li, S.: 'Land-Use Classification Using Taxi GPS Traces', IEEE Transactions on Intelligent Transportation Systems, 2013, 14, (1), pp. 113-123

[6] Guande, Q., Xiaolong, L., Shijian, L., Gang, P., Zonghui, W., and Daqing, Z.: 'Measuring social functions of city regions from large-scale taxi behaviors', in Editor (Ed.)^(Eds.): 'Book Measuring social functions of city regions from large-scale taxi behaviors' (2011, edn.), pp. 384-388

[7] Veloso, M., Phithakkitnukoon, S., and Bento, C.: 'Urban mobility study using taxi traces'. Proc. Proceedings of the 2011 international workshop on Trajectory data mining and analysis, Beijing, China2011 pp. Pages

[8] Liu, Y., Wang, F., Xiao, Y., and Gao, S.: 'Urban land uses and traffic 'source-sink areas': Evidence from GPS-enabled taxi data in Shanghai', Landscape and Urban Planning, 2012, 106, (1), pp. 73-87

[9] Mazimpaka, J.D., and Timpf, S.: 'Exploring the Potential of Combining Taxi GPS and Flickr Data for Discovering Functional Regions', in Bacao, F., Santos, M.Y., and Painho, M. (Eds.): 'AGILE 2015: Geographic Information Science as an Enabler of Smarter Cities and Communities' (Springer International Publishing, 2015), pp. 3-18

[10] Zhou, Y., Fang, Z., Thill, J.-C., Li, Q., and Li, Y.: 'Functionally critical locations in an urban transportation network: Identification and space–time analysis using taxi trajectories', Computers, Environment and Urban Systems, 2015, 52, pp. 34-47