Investigation on Taxi Trip Characteristics Based on Taxi GPS Trajectory

Weiyan Xu* and Yumei Huang
School of Mathematics and Statistics, Lanzhou University, Lanzhou, China

*E-mail: xuwy2017@lzu.edu.cn

Abstract. Knowledge of urban taxi trip characteristics is crucial at many levels of urban transportation planning, operations, and traffic management. In this paper, we investigate the characteristics of taxi trip through analyzing approximately 3,000 taxis spanning seven consecutive days in Lanzhou. The characteristics of taxi trip such as taxi trip speed, service time, vacation time, and spatio-temporal patterns of taxi trip are analyzed. The results show that the average speed of taxis varies between 17 km/h to 22 km/h from 7 am to 8 pm. Moreover, the average service time and vacation time are 20 minutes and 6 minutes, respectively. Although, the net flow ratio in heart of the city goes around zero, the flow direction of taxi trip is obviously different in the morning rush and the evening rush. Moreover, The rural-urban fringes have a strong positive net flow ratio which indicates that the strength of inflow is greater than that of outflow.

1. Introduction

With a large number of citizens tending to travel by taxis for convenience, taxis become one of the most significant traffic means. Taxi GPS trajectories data contain a large amount of spatial and temporal information, making it possible to understand the potential knowledge of human travel patterns and road network dynamics. Understanding the travel patterns of taxis could not only provide theory basis of urban traffic planning, design and management, but also guide the development of environment and economy [1-3].

Previous studies were mainly focused on understanding the travel patterns of taxi trip based on questionnaires of taxi drivers and passengers [4-6]. These traditional questionnaires are time-consuming and costly, limiting knowledge of taxi travel patterns. Recently, the technology of communication and computer has been developing faster and faster, the data from GPS known as Global Positioning System is being applied to all walks of life widely. GPS trajectories data could give abundant information of vehicles including time, position, running speed along with status, etc. This large-scale data gives an accurate description of taxis’ mobility and provides an approach to understand complex transportation system.

Based on taxi GPS data, ref [7] proposed an overview of mechanisms to analyze resident’s motilities, in which three dynamic characteristics containing operational dynamics, traffic dynamics and social dynamics are included. The mobility of human is explored by a generic architecture mPat designed in Ref [8], which uses the multi-source data including cellular phone call records, taxicab GPS data, bus GPS data, and smart card records for subway and bus in Shenzhen, China. They analyzed the spatial and temporal correlations between multi-source data. Ref [9] explored urban trip travel mobility, OD distribution, hotspots using large-scale taxi GPS data in Harbin city.
Ref [10] analyzed mobility patterns in the poor weather conditions by dint of data from GPS trajectories of taxis in New York City. What can be revealed is that travellers chose to travel for shorter periods during the snowstorm than during sunny days, and that trips of taxis in the period of working were relatively consistent with sunny days in spite of the heavy snowfall. Ref [11] established a spatial embedding network and analyzed the GPS data from taxis in Shanghai to study the structure of Shanghai and the mobility patterns. The city structure is analyzed by community detection method, and the centrality and flow characteristics of each layer were further analyzed. To predict the local customer-search movements of drivers in vacant taxis, a cell-based model has been given in Ref [12]. Ref [13] proposed a STTCLUS trajectory clustering algorithm based on spatio-temporal similarity. It mined the hot spots paths between hot spots during different periods, and analyzed the hot spots and hot spots paths of residents with the track data of 10,000 taxis in Chongqing. A new chord diagram plot was put forward for the research on spatio-temporal features and patterns about the mobility of urban citizens on the basis of taxi GPS trajectories data collected in a weekday in Beijing [14].

The analysis of taxi travel behavior has important application value in urban calculation, traffic flow control and planning, information recommendation and other aspects. As an important mobility data source, taxi GPS trajectories data has the characteristics of high quality, good continuity and wide distribution, which is suitable for travel behavior mining. In this paper, the taxi trajectory data is used to study the characterization of taxi travel behavior in a valley city Lanzhou.

The structure of this paper is as follows: in section 2, the data and how it is processed are explained. In section 3, the analysis of taxi trip behavior characteristics is showed. Finally, the conclusion is presented in Section 4.

2. GPS Trajectory and basic terminology

2.1. Data Description and Pre-processing

The GPS trajectories data used in this paper is acquired from the traffic system in Lanzhou, China. The data set spans seven consecutive days from March 6 to March 12, 2017. Information containing vehicles ID, state (vacancy or occupancy), driving direction and velocity, as well as longitude and latitude at each timestamp are recorded in detail. The location of taxis at each timestamp indicating recording time can be seen by the longitude and latitude. When there are no passengers in taxis, the state is marked by 0, or it is marked by 1, which indicates whether taxis are occupied or not. The velocity(km/h) of vehicles mentioned in this paper is the instantaneous velocity.

Because of the random and systematic errors, there are abnormal data in the raw GPS trajectories data generally. So the obvious errors must be removed from the raw data by the method of data cleaning for further analysis and application. First of all, the data from few taxis of which the speed is higher than 100 km/h that exceeds the limiting velocity in urban will be eliminated. Furthermore, the urban map of Lanzhou is obtained by the Minnesota Traffic Generator [15] and the abnormal value must be eliminated to realize the map-matching GPS data on the road network of Lanzhou. At last, the linear interpolation method is applied to supply the missing points between adjacent GPS points whose time granularity is larger than 30s.

2.2. Basic Terminology

**Definition 1** (Raw GPS trajectory). A raw GPS trajectory consists of a sequence of taxi locations that changes continuously over time, i.e. $\text{Tra}_i = \{P_1, P_2, P_3, ..., P_n\}$, where $P_i = (\text{lon}_i, \text{lat}_i, t_i, \text{flag}_i)$. $(\text{lon}_i, \text{lat}_i)$ is the longitude and latitude of taxi at time $t_i$, $\text{flag}_i = 1$ indicates if there are passengers in the taxi. If there are no passengers in the taxi, then it equals to 0.

**Definition 2** (Pick-up point and Drop-off point). Based on the field of $\text{flag}_i$ each taxi trajectory can be assembled as a sequence of 0 and 1. We define “pick-up point” as a location where the $\text{flag}_i$ shift from 0 to 1, and “drop-off point” as a location where the $\text{flag}_i$ shift from 1 to 0.

**Definition 3** (Passenger trajectory and vacant trajectory). A passenger trajectory denotes the trajectories when there are passengers inside the taxi, and a vacant trajectory indicates unoccupied
statuses. In other words, a passenger trajectory is defined as a point sequence beginning with pick-up point and ending with drop-off point. Similarly, we define the vacant trajectory begins with drop-off point and ends with pick-up point. In other words, we define a passenger trajectory as a point sequence which starts with the pick-up point and end up with the drop-off point. Likewise, the vacant trajectory is thought as a point sequence starting with the drop-off point. At the same time, it ends up with the pick-up point correspondingly.

**Definition 4** (Grid decomposition). The grid decomposition is defined as the division of urban areas into equal-sized grids. An \((p \times q)\)-grid decomposition consists of a \(p \times q\) matrix of grid cells, where the width of the grid cell is \(\frac{lon_{\text{max}} - lon_{\text{min}}}{q}\) and the length of the grid cell is \(\frac{lat_{\text{max}} - lat_{\text{min}}}{p}\), \(lon_{\text{max}} = \max_i\) \(lon_i\), \(lat_{\text{max}} = \max_i\) \(lat_i\), \(lon_{\text{min}} = \min_i\) \(lon_i\), \(lat_{\text{min}} = \min_i\) \(lat_i\).

**Definition 5** (Inflow, Outflow, and net flow ratio) Inflow of the grid \(i\) refers to the number of passenger trajectory from the other grid to the grid \(i\). Outflow of the grid \(i\) refers to the number of passenger trajectory from the grid \(i\) to the other grid. The net flow ratio is defined as the ratio (i.e. \([\text{inflow} - \text{outflow}] / [\text{inflow} + \text{outflow}]\)) of the inflow and outflow.

### 3. Taxi Trip Characteristics Analysis

#### 3.1. Average Trip Speed

The following formula is for the calculation of the average speed in which occupied taxis run:

\[
\bar{v} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{m} v_{ij}}{\sum_{i=1}^{n} \sum_{j=1}^{m} t_{ij}}
\]

where \(v_{ij}\) is the average speed of the \(j\)-th passenger trajectory of the \(i\)-th occupied taxi. \(m_i\) is the number of passenger trajectory for the \(i\)-th occupied taxi. \(n\) is the number of taxis. \(t_{ij}\) is the service time of the \(j\)-th passenger trajectory of the \(i\)-th occupied taxi. The average speed of vacant taxi can be obtained in a similar way.

The average speed of the taxi can directly reflect the traffic condition of the city. The Traffic congestion evaluation indicators are established according to the traffic conditions of various countries. A quantitative evaluation indicators system based on average speed has been published in 2002 by the Chinese Ministry of Public Security to promote the management of urban traffic [16]. The specific levels are shown in Table 1.

| Traffic status  | Vehicle speeds (km/h) | Congestion level |
|-----------------|-----------------------|-----------------|
| Unblocked       | \(v>30\)              | 0               |
| Amble           | \(20< v \leq 30\)     | 1               |
| Congestion      | \(10< v \leq 20\)     | 2               |
| Severe congestion| \(v \leq 10\)        | 3               |

Figure 1 describes the average speed of the occupied taxis in each hour. It varies significantly at different times of the day. Between 7 a.m. to 8 p.m. in a day, the average speed of taxis varies from 17 km/h to 22 km/h. Except work time, the average speed during the other time mostly maintain at about 30 km/h. The average driving speed maintains almost at 30 km/h except taxis run in working time. The average speed is taxis run in the average speed less than 20 km/h in rush hours consisted of the periods...
of 7:00-9:00 and 17:00-19:00. Compare to Table 1, the urban road network exhibit congestion during the rush hours. The urban traffic conditions of Lanzhou city

![Graph showing average speed for occupied taxis during the 24 hours day period]

**Figure 1.** The average speed for the occupied taxis during the 24 hours day period

### 3.2. Service Time and Vacation time

The taxi service time is an important metric to reflect the status of a taxi. The status of a taxi can be reflected by an important metric as service time well. Figure 2 gives the average service time for each vehicle during the 24 hours day period. As can be seen from the figure, on weekdays and weekends, the trend of service time is obviously different. The longer service time occurs in morning rush hour in weekdays. The longer service time in weekends appears from noon to 5 pm. The service time from midnight to 6:00 am is relatively short, generally about 11 minutes. Figure 3 gives the average vacation time for each vehicle during the 24 hours day period. From 6 a.m. to 11 p.m., the average vacation time is about 9 minutes.

![Graph showing average service time during the 24 hours day period](image1)

**Figure 2.** The average service time during the 24 hours day period

![Graph showing average vacation time during the 24 hours day period](image2)

**Figure 3.** The average vacation time during the 24 hours day period

We also provide the statistical characteristics of the service time and give the empirical cumulative distribution function (CDF) plots for service time in Figure 4. We observe that about 78% of the taxi trips have service time less than 20 minutes. Figure 5 plots the PDF for service time. The statistical validation is conducted using the Kolmogorov–Smirnov (KS) statistic. It is found that service time distributions fit better to log-normal distribution for this continuous case. What’s shown in this continuous case is that the distributions of the service time can fit better to the log-normal distribution.

### 3.3. Spatio-Temporal Patterns of Taxi Trip

A day is divided into 12 continuous periods, each of which lasts 2 hours, for the research of the temporal patterns of taxi trips. We obtain the inflow, outflow, and net flow ratio of each grid.
The rush hour in the morning (7:00-9:00) and evening (17:00-19:00) indicate the heavy-traffic in a day. The net flow ratios in rush hour on Monday morning and evening are shown respectively in Figure 6 and Figure 7. As can be seen from two figures, some areas represented by the most warming color of them all have a strong positive net flow ratio. In other words, the inflow in these areas is much larger than the outflow. Such areas are mostly located in rural-urban fringe. There are fewer taxis waiting for customers. The most taxis load the passengers from other areas and unload them in these areas.

Comparing to evening rush hour, the outflows of three districts (Aning district, Xigu district, and Qilihe district) in morning rush hour are larger than that of inflows. This result shows that there are relatively more residential areas in these three areas. The net flow ratio in heart of the city (Chengguan district) vary between -0.25 and 0.25. Although the inflow and outflow of taxi passengers are relatively close to each area, the direction of passenger flow is obviously different. This result can be used to identify the residential and non-residential areas in the area. We also conclude that the Chengguan district is a mix area which includes commercial, administrative and residential areas.

4. Conclusions
In our work, we investigated the characteristics of taxi trip through studying approximately 3,000 taxis trip extracted from seven consecutive days in Lanzhou. We focused on the statistics: taxi trip speed, service time, and vacation time. The distributions of pick-up point and along with the drop-off point were analyzed. The results show that the average speed of taxis varies between 17 km/h to 22 km/h from 7 am to 8 pm. Moreover, the average service time and vacation time are 20 minutes and 6 minutes, respectively. The distribution of service time was investigated. The results of Kolmogorov–Smirnov (K–S) test indicate that the service time distributions could be approximated by log-normal distributions. We also investigated the spatio-temporal patterns of taxi trip. The results showed that the net flow ratio in heart of the city goes around zero, but the flow direction of taxi trip is obviously
different in the morning rush and the evening rush. Moreover, the rural-urban fringes have a strong positive net flow ratio which indicates there are fewer taxis waiting for customers in rural-urban fringes areas.

Acknowledgment
I thank Fengshan Bai and Zhenjuan Yang for their assistance in GPS data pre-processing. This work is partially supported by the National Natural Science Foundation of China under Grant 11571156 and 71761031.

References
[1] Feng H, Bai F, Xu Y. Identification of critical roads in urban transportation network based on GPS trajectory data. *Physica A: Statistical Mechanics and its Applications*, 2019, **535**:122337.
[2] Berlingerio M, Calabrese F, Lorenzo G Di, et al. Allaboard: a system for exploring urban mobility and optimizing public transport using cellphone data. In: *Machine learning and knowledge discovery in databases*. Springer, 2013, pp. 663–666.
[3] Cui J, Liu F, Janssens D, et al. Detecting urban road network accessibility problems using taxi GPS data. *J Transp Geogr*, 2016, **51**:147–157
[4] Song Y, Ikeda T, A study on present Shanghai urban inhabitants’ leisure activities and sites, *Journal of Asian Architecture and Building Engineering*, 2005, **4**(2), 301-306.
[5] Shi J, Tao L, Li X Y, et al., (2014). A survey of taxi drivers’ aberrant driving behavior in Beijing. *Journal of Transportation Safety & Security*, 2014, **6**(1): 34-43.
[6] Wong R C P, Szeto W. Y and Wong S C. Behavior of taxi customers in hailing vacant taxis: A nested logit model for policy analysis. *Journal of Advanced Transportation*, 2015, **49**: 867-883.
[7] Castro P S, Zhang D, Chen C, et al. From taxi GPS traces to social and community dynamics: a survey. *ACM Comput. Surv.* 2013, **46**(2):17
[8] Zhang D, Huang J and Li Y. Exploring human mobility with multi-source data at extremely large metropolitan scales, *Proceedings of the MobiCom’14*, Maui, HI, USA, September 7-11, 2014.
[9] Wong R C P, Szeto W Y, Wong S C, et al. Modeling multi-period customer searching behavior of taxi drivers. *Transportmetrica B*, 2014, **2**(1):40–49.
[10] Qing C, Parfenov S and Kim L J. Identifying travel patterns during extreme weather using taxi GPS data. *Transportation Research Board 94th Annual Meeting*, Washington DC, United States, 2015.
[11] Liu X, Gong L, Gong Y, Liu Y. Revealing travel patterns and city structure with taxi trip data, *Journal of Transport Geography*, 2015, **43**: 78–90.
[12] Tang J, Jiang H, Li Z, et al. A two-layer model for taxi customer searching behaviors using GPS trajectory data. IEEE Trans. Intell. Transp. Syst. 2016, **17**(11): 3318–3324.
[13] Zheng L, Xia D, Zhao X, et al. Spatial–temporal travel pattern mining using massive taxi trajectory data, *Physica A*, 2018, **501**: 24-41.
[14] Wang H, Huang H, Ni X and Zeng W. Revealing spatial-temporal characteristics and patterns of urban travel: a large-scale analysis and visualization study with taxi GPS data. *International Journal of Geo-Information*, 2019, **8**:257.
[15] Mokbel M F, Alarabi L, Bao J, et al. MNTG: An extensible web-based traffic generator, Advances in Spatial and Temporal Databases. SSTD 2013. *Lecture Notes in Computer Science*, Springer, Berlin, Heidelberg. 2013