Mining Spatiotemporal Characteristics of Car-sharing Demand

Jing-Jing Tian¹, Dong-Fan Xie¹ and Fu-Jun Ding²

¹MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, Beijing Jiaotong University, Beijing 100044, P.R. China.
²Gansuyixiangxing New Energy Developments Ltd, Lanzhou 730000, China

E-mail address: 17120756@bjtu.edu.cn (Tian), dfxie@bjtu.edu.cn (Xie)

Abstract. Since car-sharing demand plays an essential role on the management and service of car-sharing system, this paper attempts to analyze the temporal and spatial characteristics of the car-sharing demand, aiming to discover spatial and temporal patterns and association rules. Based on a clustering algorithm (i.e., DBSCAN), the spatiotemporal characteristics of car-sharing demand are studied. On the basis, the demand is divided into various clusters in space. Furthermore, the correlations between any two clusters are studied. The results show that car-sharing demand is high on Friday, Saturday and Sunday, and has strong correlation with time and space. It is expected the results can support the management of car-sharing system, and further promote the service level for passengers.

1. Introduction

With the intensification of traffic problems such as traffic congestion, car-sharing is considered to be one of the effective ways to relieve these problems. Under the strong support of government policies, car-sharing have achieved rapid development in China. Correspondingly, research on car-sharing is gradually becoming a hot topic. There are a lot of researches on the demand forecasting and location depot of car-sharing, while the analysis of spatial-temporal characteristics of demand is not enough.

As for studies based on the trajectory, most existing works focus on clustering and spatial-temporal association rules. The former based on segmentation of the moving trajectories, such as segment extraction and clustering of the moving trajectories (Ma X et al. 2013, Arthur et al. 2014, Almanna et al. 2018). The latter mainly taking into account the spatial and temporal constraints simultaneously, and focusing on the improvement of efficiency (XL Zhao et al. 2010, Y Xia et al. 2011, MQ Yan et al. 2017). There is also a method that main uses heat maps (Chang Yu et al., 2017) to analyze features of demand.

In order to have a deeper understanding of the car-sharing demand characteristics, this paper mainly explores the temporal and spatial characteristics of car-sharing demand through clustering and spatiotemporal association algorithm. The rest of the paper is organized as follows. In Section 2, we describe the dataset used in this paper and make some basic data analysis. In Section 3, we introduce the clustering algorithm named DBSCAN and analyze the results. In Section 4, we expand the apriori association algorithm and analyze the spatiotemporal relevance of car-sharing demand. In Section 5, we summarize the conclusion of this paper.
2. Data analysis
In this study, the data is obtained from a car-sharing company, including the rental data and the trajectory data of the car-sharing in Beijing. The rental data contains 81455 rental records from December 17, 2016 to October 16, 2017 (278 days). The car-trajectory data contains 123864309 records of trips from January 1, 2017 and October 31, 2017 (293 days). Preprocess raw data, and the datasets used in this study are given in tables 1 and 2 for rental records and GPS data, respectively.

Table 1 Car-sharing Rental Data

| Fields            | Description                                      |
|-------------------|--------------------------------------------------|
| Rental number     | Identity of an cars-haring use transaction       |
| Pick-up station   | Name of car picking-up station                   |
| Return station    | Name of car return station                       |
| Time              | Time of picking up car                           |
| Duration          | Total travel time of an transaction              |
| Distance          | Total mileage of an transaction                  |
| License Number    | Identity of a car                                |

Table 2 Trajectory Data of car-sharing

| Fields           | Description                                      |
|------------------|--------------------------------------------------|
| Car number       | Identity of a car                                |
| Time             | the time of recording car's position              |
| Longitude        | Longitude of car position                        |
| Latitude         | Latitude of car position                         |
| Speed            | Instantaneous speed of car                       |

2.1. Rental Data
Car-sharing system can be divided into three categories based on operating patterns: two-way (or round-trip) systems, one-way system and free floating. For the first, the car must be returned to the station where it was picked up. While for the second, it allows users to return a rented car to any of the designated stations. And the third allows the car to park in any place where it can be parked. It is clearly that the company applies the one-way car-sharing pattern, which is more prevalent. Therefore, station imbalance will occur.

In this study, the datasets contain 174 car-sharing stations and 960 cars. Based on the dataset of 278 days, the difference between the total numbers of borrowed and returned cars is used to indicate the degree of imbalance in the stations, as shown in Figure 1. The greater the absolute value of the difference, the higher the imbalance of the station. It can also be found that the station with larger amount of rental volume has a higher degree of imbalance. Few stations are in balance states.

Fig 1 Difference between pick-up and return

Fig 2 Car usage frequency
In addition, the car-sharing were sorted in terms of the rental amount, as shown in Figure 2. One can see that there are two turning points on the curve. The turning point divides the cars into three categories according to rental amount of each car.

For the type of cars with a high rental amount, it is found that almost all the cars are picked up and returned at the same site, and these stations are more balanced. We can infer that even if there is not much rental data at these sites, the utilization rate of cars is higher.

2.2. Trajectory Data
The trajectory data has total 123,864,309 track points, including 1071 cars. The trajectory of a car consists of multiple trips by many users, and the trajectory data implies the departure point and destination of the user, which should be split for the following analysis. In this study, for the sake of simplicity, we refer to the related literature, the point where the stay time exceeds ten minutes is used as the travel stop point, and the stop point is used as the end point of the last trip and the starting point of the next trip. The OD demand is extracted from the trajectory data, concluding 194 working days, and 99 non-working days. The resulting dataset is shown in table 1.

| Field  | Description                        |
|--------|------------------------------------|
| Otime  | Time of origin                     |
| Olon   | Longitude of origin                |
| Olat   | Latitude of origin                 |
| Dtime  | Time of destination                |
| Dlon   | Longitude of destination           |
| Dlat   | Latitude of destination            |
| Distance | Distance between origin and destination |
| Time   | Duration between origin and destination |

3. Clustering analysis of car-sharing demand
Clustering analysis is a branch of data mining. Based on similarity, there is more similarity between patterns in one cluster than patterns in different clusters. The algorithm of cluster analysis can be divided into five classes: Partitioning Methods, Hierarchical Methods, density-based methods, grid-based methods and Model-Based Methods.

3.1. Clustering Algorithm
This study applied DBSCAN algorithm for clustering. DBSCAN (Ester et al. 1996) is a density-based method with two parameters: eps and minpts, corresponding to the maximum radius of the neighborhood and the minimum number of neighbors for a core point. A cluster is a maximal set of density-connected points.

3.2. Clustering Analysis
After OD trips were extracted from trajectory data, the DBSCAN clustering algorithm is used to analysis the start and end points of the OD trip. By continuously adjusting the parameters, the clustering results are given in Figures 3 and 4.

It can be seen that the spatial distribution of each cluster of origin and destination is consistent. Therefore, only the characteristics of the origin point will be studied in the following. Figure 5 presents the demand of all clusters. The number refers to the proportion of non-working day (including holiday). It is clear that cluster 5 has the most demand, and cluster 6 has the most demand of non-working day.

Figures 6, 7 and 8 reveal the demand from month, hour, and day of the week, respectively. The demand of cluster 5 has increased significantly in August, while other clusters have seen a significant downward trend. The distribution of each cluster is basically the same in the day, and there are three peak hours in a day: 8-10am, 12-14pm, and 16-18pm. Figure 8 show the distribution of each cluster on
the day of the week. For all clusters, the proportion of Friday, Saturday and Sunday is larger, and the difference between other working days is small.

The extracted origin-destination trips also include duration and distance. In order to analyze their characteristics, we make their own cumulative distribution curves, as shown in Figures 9 and 10. We can see that 90% of the origin-destination travel distance is within 30km, and 95% of the duration is within 90 minutes.
4. Association analysis of car-sharing demand

The study of association rules is another topic of data mining. It refers to discover the connections between things from the data. The association rules were originally proposed for the Market Basket Analysis problem in order to discover the correlation between different commodities in the transaction database.

4.1. Apriori association Algorithm

The Apriori algorithm (Agrawal, 1994) is the most commonly used algorithm for mining association rules. The main idea of the algorithm is to discover frequent item-sets gradually by increasing the number of elements in the item-set.

Before mining association rules, we need to divide data of each attributes into different levels according to certain rules. In this paper, we divide the study area into grids, and a day into eight time periods. Then, we count the demand for each time period of every grid. In order to make the results more accurate and reliable, this study divides the dataset into two categories: workday (excluding holidays) and non-working days (including holidays). After the initial dataset is established, set minimum support and confidence and get strong association rules.

4.2. Apriori Analysis Results

Perform the correlation analysis according to the method in 4.1, and visualize the obtained strong association rules. Fig 11 and 12 show the association rules for weekdays and non-working day, respectively. Blue, green and red represent the three levels of low demand, medium demand and high demand. The flat formed by the horizontal grid number and the vertical grid number represents the geographic plane of the study areas, and the vertical axis represents the time period.
The association rules can be described as: area A in the period B ⇒ demand (sparse, medium and dense). Whether it is a working day or a non-working day, the area covered by the association rules is similar, indicating that this part has a distinctive feature. And it’s clear that the demand on non-working day is higher than weekdays.

On working days, it can be seen that the area covered by blue is concentrated at 0-6 points, that is, the demand is low, and the part with medium demand is wide distributed. On non-working days, the distribution of low demand is relatively uniform, and the spatiotemporal range covered by the high demand part is widely distributed.

5. Conclusion
With the development of car-sharing, it is increasingly necessary to study its mechanism. In order to conduct more reliable researches on the location depot and car relocation of car-sharing, it is important to conduct a further analysis of the car-sharing demand.

Through the clustering analysis and spatiotemporal correlation analysis, the following effective conclusions can be obtained. These conclusions can be applied to subsequent demand forecasting and site depot researches:

1. The total amount of the site with high degree of imbalance is higher, while the site with high balance is lower, but the utilization rate of the car is quite high.
2. Users mainly for short-term travel and short or medium distance.
3. Different clusters have different characteristics, and should be considered separately when forecasting car-sharing demand.
4. The strong association rules are mainly reflected in the relationship between spatiotemporal and demand, and the gap between working days and non-working days is large.

Acknowledgments
This work is supported by the project from Gansuyixiangxing New Energy Developments Ltd, Lanzhou, China.

References
[1] Ma X, Wu Y J, Wang Y, et al. 2013 Mining smart card data for transit riders’ travel patterns Transportation Research Part C Emerging Technologies. 36 1-12
[2] Arthur, Shaw and Gopalan 2014 Finding frequent trajectories by clustering and sequential pattern mining. Journal of Traffic and Transportation Engineering. 1 393-403
[3] Almannaa M, Elhenawy M and Rakha H 2018 A Novel Clustering Algorithm for Traffic Operational Analysis Transportation Research Board
[4] Zhao X L and Xu W X 2010 Mining spatio-temporal association rules in bus IC card databases International Conference on Power Electronics and Intelligent Transportation System. 125-128
[5] Xia Y and Chengdu 2011 Spatio-temporal Association Rule Mining Algorithm and its Application in Intelligent Transportation System Computer Science. 38 173-176
[6] Yan M Q, Guo Z Y and Ren Z H 2017 Spatio-temporal analysis of bus pickpocketing using association rules based on clustering Journal of East China Normal University
[7] Yu C and He Z C 2017 Analysing the spatial-temporal characteristics of bus travel demand using the heat map Journal of Transport Geography. 58 247-255
[8] Ester M, Kriegel H P and Xu X 1996 A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise International Conference on Knowledge Discovery and Data Mining. 226-231
[9] Agrawal R and Srikant R 1994 Fast Algorithms for Mining Association Rules in Large Databases International Conference on Very Large Data Bases. 487-499