Influencing factor analysis of car-sharing demand based on point of interest data

Huimin Dong1,a*, Xiaobao Yang1, Wencheng Wang2

1 MOT Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport, School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China
2 Beijing Municipal Institute of City Planning & Design, Beijing, China
* Corresponding author: *18120745@bjtu.edu.cn

Abstract—The rise of car-sharing can make full use of road space resources, alleviate traffic congestion and reduce traffic energy consumption. However, due to the failure to accurately predict the demand distribution of car-sharing, there are a lot of empty driving phenomena in vehicle scheduling, which leads to the economic losses of car-sharing operators. Existing studies show that the point of interest has a significant impact on urban residents' travel, but few of them study the quantitative relationship between interest points and shared car orders. Based on this, using the order data of car-sharing in Beijing, this paper establishes a negative binomial model of travel demand and density of different kind of points of interest, and analyzes their relationship. The analysis found that users tend to use car-sharing to go to leisure places. In areas where public transportation is underdeveloped, people will use car-sharing more. The methodology of this paper can provide a theoretical basis for sharing automobile enterprises to develop new operation areas and select reasonable car sharing stations.

1. INTRODUCTION

With the development of the automobile industry and the acceleration of urbanization, the number of private cars is increasing day by day. While cars provide more convenience to people, they also cause serious environmental pollution, energy consumption and road congestion, as well as various traffic accidents and parking difficulties caused by this. With the emergence of a series of problems and the continuous increase in the cost of private car purchases, car-sharing have emerged. Car-sharing provides a new option for urban travel. In China, car-sharing are in a stage of rapid development. Its operation not only reduces the burden of car buyers, but also satisfies the driving pleasure of drivers. At the same time, it also reduces the car empty rate and reduces the idle time in the parking lot. Therefore, it is very necessary to explore the factors that affect the demand for car-sharing, which will help the further promotion of car-sharing in China and provide theoretical decision-making reference for related enterprises.

At present, there are many researches on the impact of car-sharing travel, mainly focusing on the characteristics of car-sharing and the travel characteristics of individual users. For example, Marc found that the economy and convenience of car-sharing have a significant impact on the choice of travel mode. Paundra et al [1] studied whether differences in personal ownership psychology affect instrumental car attributes (price, parking convenience, and car type) and the effect of this difference on people's willingness to car-sharing. Jiang [2] studied the factors that affect consumers' choice of car-sharing,
among which family economic conditions, car usage-related expenses, car demand, personal characteristics and travel characteristics are the main influencing factors. Existing research, using point-of-interest (POI) data mainly focus on urban functional area division, spatial pattern analysis and hotspot area identification, etc. For example, Chi [3] have used the POI data to study the division of urban functional areas and related identification methods. Feng [4] analyzed the commercial spatial pattern of Shenzhen based on POI data. Ling [5] used point of interest data to conduct a hot spot analysis of the city. Existing studies have proved that the demand of car-sharing is partly drove by the distribution of POI in the operating area, but seldom studies analysis the relationship between POI and car-sharing orders. On this basis, this article reclassifies the POI data and applies it to the demand analysis of car-sharing. The article selects 1 km around the car sharing station as the affected area, and extracts the number of POI and calculates their density. We also consider the distance between the POI and the central site, and get the density sum of each type of POI. Finally, we analyze the impact of various POI on car-sharing orders in Beijing by establishing a negative binomial model which can provide a reference for operators to develop new operating areas or choose other operating sites.

2. STUDY AREA AND DATA

2.1. The study area
Based on the distribution of order data in different car sharing stations, this paper selects 11 areas in Beijing, Dongcheng District, Xicheng District, Chaoyang District, Fengtai District, Haidian District, Shijingshan District, Fangshan District, Tongzhou District, Shunyi District, Changping District and Daxing District as research area.

2.2. Study data
Study data mainly includes Beijing car-sharing order data and Beijing POI data.

2.2.1. Order data
The attribute fields of the original order data include order number, cost, pick up station, return station, use time, order length, return method, vehicle type, etc.

Some of the field properties are shown in the Table1.

2.2.2. POI data
POI mainly refers to the service facilities closely related to the daily life of residents, such as convenience stores, restaurants, hospitals, parks and libraries. POI data contains the location information and attribute information of those geographical entities, which are represented by Point elements in geospatial information data.

The traditional way to obtain POI is to conduct field investigation and collection, which requires a large amount of manpower and material resources. Moreover, the amount of data acquired at one time is limited and the updating speed is slow. With the development of geographic information technology, methods for geographic information data extraction from Websites such as Google Earth, Amap [6] and Baidu Map have been gradually rising. These websites have opened relatively perfect development interfaces, and accurate POI data can be obtained easily.

Table 1 Sample of car-sharing order data

| Field Name                  | The sample                      |
|-----------------------------|---------------------------------|
| Order number                | 179***124                       |
| The station to borrow the car| Huayuan Good world parking lot  |
| The station to return the car| Qiaozhuang shopping mall parking|
The POI data used in this article mainly is mainly obtained from Amap application programming interface platform (using GCJ - 20 coordinate system). As a kind of space data, POI data is a geographical entity in the physical world of reality of abstraction. Any physical entity, such as schools, hospitals, shopping malls, square, can be expressed as a kind of POI. It represents the spatial location of geographical entities, and also contains rich attribute information (latitude and longitude coordinates, name, classification, etc.). The detailed information of the data used in this paper is shown in Table 2.

2.3. Preprocessing of POI data
The original data has problems such as duplication, missing, disordered format, and coordinates of POI beyond the scope of the research area. To ensure the accuracy of the research results, the two types of data need to be preprocessed. After processing, the effective data contains 172 sites. This article selects the data from June to September 2017 for research.

2.3.1. Order data preprocessing
- Delete invalid order.
- Complete orders with missing values.
- Unified date and time format.

2.3.2. POI data preprocessing
On the basis of the original data and in combination with the content of this paper, the data are integrated and processed. 

**Deduplication.** There is some POI data with the same name and latitude and longitude coordinates, which need to be de-duplicated.

| Field Name | The meaning                      |
|------------|---------------------------------|
| Lon        | Point of interest longitude     |
| Lat        | Latitude of point of interest   |
| Name       | Name of interest point          |
| Address    | The address of the point of interest |
| Type1      | Belong to the category          |
| Type2      | Belong to the class             |
| Type3      | Belongs to a small class        |
| Id         | The ID of the point of interest  |

**Reclassification.** According to the original classification rule of Amap, some POI may belong to different categories at the same time. For example, cake shops belong to both catering service and shopping service. To avoid duplicate classification of POI and established a clearer classification...
standard, we adjust the original classification rules of Amap as follows:

Delete "Automobile Service", "Automobile Sales", "Automobile maintenance" and other related sites. Delete the "Motorcycle Service site".

The "sports and leisure service places" are divided into "sports venues", "entertainment venues", "vacation and recuperation venues", "cinemas and theaters" and other leisure venues, and analyzed respectively.

Industrial parks and buildings in "commercial residence" are classified as "companies and enterprises"

The category of "Science, Education and Cultural Services" selects three representative POI: "schools", "scientific research institutions" and "cultural venues".

In "Transportation facilities Service", the "port and wharf", "light rail station", "ferry station" and "cableway station" not involved in Beijing shall be deleted.

Delete "road ancillary facilities", "Geographical location and address information", "Event activities", "Access facilities", etc.

The reclassified data are shown in Table 3.

3. METHOD

3.1. Definition and selection of dependent variables
From the perspective of operators, order requirements and reasonable scheduling of vehicles are very important, which is related to the number of vehicles to be put on each station, whether to dispatch vehicles, when to dispatch vehicles, and how many vehicles to be dispatched, etc.

For users, the car-sharing station should ensure that there are enough vehicles for users to rent and enough parking space for users to return the car.

Therefore, based on the actual demand, this paper studies the daily order quantity of each car-sharing station as the dependent variable.

Table 3 Reclassification of POI data

| Serial number | category       | Serial number | category       |
|---------------|----------------|---------------|----------------|
| 1             | Food and beverage | 12            | School         |
| 2             | Shopping        | 13            | Scientific research |
| 3             | Life service    | 14            | Residential    |
| 4             | Sports venue    | 15            | Airport        |
| 5             | Entertainment   | 16            | Railway station |
| 6             | Resort and spa  | 17            | Coach station  |
| 7             | Leisure place   | 18            | Subway station |
| 8             | Movie theater   | 19            | Bus stop       |
| 9             | Health care service | 20         | Financial      |
| 10            | Accommodation   | 21            | Company        |
| 11            | Scenic spot     | 22            | Public facility |
3.2. Definition and selection of explanatory variables

3.2.1. Definition of explanatory variables

Surveys of existing and potential customers show that the distribution of POI around residents’ travel destination will have an influence on people’s demand of car-sharing. Christoph, Konstantin [7] found that the usage demand of shared cars was indeed driven by POI in the operating area. Therefore, this paper considers the density of various POI within a certain range around the car-sharing station as the influencing factor of the number of car-sharing orders of that car-sharing station, and selects the POI with the most significant influence as the explanatory variables.

3.2.2. POI density calculation

In terms of the density of POI, this paper comprehensively considers the number of POI and the spatial distance. The distance between the center point and a certain type of POI can be obtained by using the semi-positive vector formula [8]:

\[ d(p_i, s_j) = 2R \arcsin \left( \sin^2 \frac{\phi_i - \phi_j}{2} + \cos(\phi_i) \cos(\phi_j) \sin^2 \frac{\lambda_i - \lambda_j}{2} \right) \]

\[ d(p_i, s_j) \mapsto \mathbb{R}_0^+ \]  

\[ d(p_i, s_j) \] -- the distance between the \(i\)th POI in the \(k\)th category and the \(j\)th car-sharing station;

\[ m(p_i, s_j) = \begin{cases} \cos \left( \frac{\pi}{2} d(p_i, s_j) \right) & \text{if } d(p_i, s_j) \leq 1 \\ 0 & \text{otherwise} \end{cases} \]

\[ m(p_i, s_j) \mapsto \mathbb{R} \in [0, 1] \]  

\[ \delta_{i,j,k} = \sum_{p_i \in \mathcal{P}_k} m(p_i, s_j) \]

\[ \delta_{i,j,k} \] -- the sum of the density values of \(i\) points of interest in the \(k\)th type of points of interest within 1 km around the \(j\)th site.

In this paper, we consider the influence distance of 1km as the central site, analyze whether each POI is within the influence range of the central site, and summarize the sum of the density values of each category of POI.

3.2.3. Selection of explanatory variables

3.2.3.1. Variance inflation factor analysis

In order to avoid the existence of multi-collinearity among explanatory variables, this research executed variance inflation factor (VIF) analysis for the explanatory variables, and excluded those explanatory variables with multi-collinearity according to the VIF results. The POI belonging to the life service place category were eliminated by using the gradual elimination method.

3.2.3.2. Feature selection algorithm

Feature selection is the process of selecting some of the most correlated features from a group of features to reduce the dimension of feature space, so as to optimize the specific index of the whole system. It can be used in this study to analyze the correlation between influencing factors and car-sharing orders, so as to improve the analysis accuracy relatively.

Selecting the all the feature sets related to the dependent variable, this research use Boruta algorithm to assess the importance of each feature. All the variables are marked as important or unimportant, and they are sorted according to their importance. This process is realized by R software programming, and
algorithm was executed 1000 iterations. The results of feature selection analysis are shown in Fig 1. From Fig 1 we can see that all the variables are identified as important variable, but the importance of airports and long-distance bus stations are significantly different from that of other variables. So we eliminated airports and long-distance bus stations.

Fig 1. Order of importance

3.3. Model building
This paper selects a negative binomial regression model for demand impact analysis. The negative binomial regression model is an extension of the traditional Poisson regression model [9]. The principle of the traditional Poisson regression model is as follows:

\[
P(Y_i = y_i) = \frac{\lambda_i^y \exp(-\lambda_i)}{y_i!}
\]

\[
Y_i = P(\lambda_i)
\]

\[
\lambda_i = \exp(\beta_0 + \beta X_i)
\]

\[
\ln \lambda_i = \beta_0 + \beta X_i
\]

\[
X_i = (x_{i1}, x_{i2}, \ldots, x_{ik})
\]

\[
\beta = (\beta_1, \beta_2, \ldots, \beta_k)
\]

\[y_i\] -- order quantity of car-sharing station \(i\),

\[\lambda_i\] -- expectation of car-sharing order quantity;

\[x_{ik}\] -- the density value of the \(k\)th category of POI within the influence range of the \(i\)th car-sharing station;

\[\beta_k\] -- the coefficient of the corresponding variable;

\[\beta_0\] -- intercept.

Traditional Poisson regression model assumes that the sample mean and variance are equal, but this is not consistent with the actual situation. To overcome the limitation of Traditional Poisson regression model, negative binomial regression model adds a random effect term on the basis of the Poisson distribution. The mathematical expectation \(\lambda_i\) satisfies the following relationship:

\[
\lambda_i = \exp(\beta_0 + \beta X_i \exp(e_i))
\]

\[
\ln \lambda_i = \beta_0 + \beta X_i + e_i
\]

\[
P(Y_i = y_i | \exp(e_i)) = \frac{[\lambda_i \exp(e_i)]^y_i}{y_i!} \exp[-\lambda_i \exp(e_i)]
\]

\[
E(Y_i) = \lambda_i
\]
\begin{align*}
\text{Var}(Y_i) &= \lambda_i + \alpha \lambda_i,
\end{align*}

(16)

\text{exp}(\varepsilon_i) \text{-- random effects term;}

Unknown parameters $\beta_i$, $\beta$ and $\alpha$ are estimated by using maximum likelihood estimation method.

4. RESULTS & DISCUSSION

4.1. User profile analysis

The user generally rent the car or return the car from 6:00 am to 1:00 am the next day, and there are fewer rental and return behaviors from 1:00 to 6:00 at night, which is in line with people's daily routine.

From Monday to Friday on weekdays, the peak of user return occur during 8:00-9:00 in the morning, and the fluctuations in the return volume of other time periods are within the normal range. The user's rental volume reaches the peak between 17:00-18:00, and the car usage in other time periods fluctuates within the normal range. This may be because most users use car sharing for commuting, so they start to use car-sharing after get off work at night, and return the car when they go to work the next morning. It is worth noting that the number of car rental during Friday evening peak hour is higher than that during Monday-Thursday evening peak hours, which indicates that users start to use cars on Friday evening, possibly in preparation for weekend travel. Correspondingly, the number of cars returned during the morning peak hour on Monday is significantly higher than that during the morning peak hours from Tuesday to Friday. The reason of that phenomenon may be that the user returns the car to the station on Monday morning after completing the travel plan on the weekend.

Users' behavior of renting and returning a car during weekend is obviously different from that on working days. The user's car renting behavior mostly occurs in the morning and noon, and the car returning behavior mostly occurs in the evening and evening, which indicates that on non-working days, users may use car-sharing for leisure activities such as weekend trips.

4.2. Model result analysis

This paper use STATA to establish a negative binomial regression model. The output results are shown in Table 4.

| Parameter          | Coef.   | Std. Err. | z     | P>|z|   |
|--------------------|---------|-----------|-------|--------|
| Food and beverage  | -0.012  | 0.0028795 | -4.19 | <0.001 |
| Shopping           | -0.004  | 0.0019225 | -2.21 | 0.027  |
| Sports venue       | 0.033   | 0.0056863 | 5.78  | <0.001 |
| Entertainment      | 0.005   | 0.0055635 | 0.95  | 0.341  |
| Resort and spa     | 0.116   | 0.0405669 | 2.85  | 0.004  |
| Movie theater      | 0.077   | 0.0171142 | 4.51  | <0.001 |
| Leisure place      | 0.037   | 0.0106102 | 3.48  | <0.001 |
| Health care service| 0.024   | 0.0024999 | 9.62  | <0.001 |
| Accommodation      | 0.005   | 0.001671  | 2.85  | 0.004  |
| Scenic spot        | 0.020   | 0.0027589 | 7.22  | <0.001 |
| Residential        | -0.021  | 0.0038093 | -5.42 | <0.001 |
| School             | -0.003  | 0.0051862 | -0.58 | 0.561  |
| Scientific research| 0.052   | 0.0125283 | 4.13  | <0.001 |
The results of negative binomial regression show that:

- Catering service, shopping service, leisure place (sports place, resort and recuperation place, cinema and theater), medical care place, accommodation service place, scenic spot, residential area, scientific research institution, subway, railway station, bus station, financial service, company and enterprise have significant influence on the station demand.

- The regression coefficient of catering service and shopping service is negative. As the public transportation around catering and shopping places is more developed and it is more convenient for users to take public transportation, they are less likely to choose car-sharing. Residential areas also show significant negative correlation with car-sharing order. Residential areas are mainly composed of residential buildings and other living facilities, while large and medium-sized parking lots are relatively few. The stations studied in this paper are all set in over-ground or underground parking lots, which may lead to a certain distance between the stations and residential areas.

- The regression coefficients of leisure place, including sports venue, resort and recuperation place, cinema and theater, etc. are all positive, indicating that users are more likely to use car sharing for leisure activities. Similarly, scenic spots are also destinations of leisure and entertainment for users, and scenic spots attract some out-of-town tourists who may use car-sharing as a temporary means of transportation.

- In terms of transportation facility, railway and subway are used in combination with car-sharing, while public transport has a competitive relationship with car sharing.

5. CONCLUSIONS

Based on car sharing order data and POI data in Beijing, China, this research reclassifies the POI, and analyzes the relationship between the density of various POI and the demand of car sharing. The analysis found that people will use car-sharing more when the travel destination is related places such as leisure and vacation. Car-sharing are mostly used for commuting during working days and leisure and entertainment activities on non-working days. As a mode of transportation, car sharing has a competitive relationship with public transport, while have a cooperative relationship with subway and railway. So car-sharing appear more often in areas where public transportation is underdeveloped.

The research results in this article can provide a certain reference for operators to choose new car-sharing stations. There are also certain limitations in the selection of POI, model construction, and station location determination in this paper. These issues need to be considered in future research when more relevant factors are analyzed in depth.

ACKNOWLEDGMENTS

This research was supported by the National Natural Science Foundation of China (Grants Nos. 91746201 and 71621001).

REFERENCES

[1] J. Paundra, (2017) “Preferences for car sharing services: Effects of instrumental attributes and psychological ownership,” Journal of Environmental Psychology , 121-130.
[2] Jiang. Y Q, (2015) “Influencing factors and development countermeasures of urban car sharing based on consumer choice,” Public Management, 157-161.
[3] Chi. J, (2016) “Quantitative Identification and Visualization of Functional Area Based on POI Data,” Journal of Geomatics, Vol.41 No.2, 68-73.
[4] Feng, D C, (2018) “Commercial spatial pattern analysis in Shenzhen based on POI data,” Wuhan: Wuhan University.

[5] Ling, T, (2018) “Analysis of urban hot spots based on POI data,” Kunming: Kunming University of Science and Technology.

[6] Gaode Map Open platform, https://lbs.amap.com/.

[7] W. Christoph, K. Konstantin, and B. Tobias, (2017) “Moving in time and space: Location intelligence for carsharing decision support,” Decision Support Systems 99:75-85.

[8] W. Sebastian, B. Tobias, and N. Dirk, (2016) “In free float: Developing Business Analytics support for carsharing providers,” Omega 59:4-14.

[9] Qin W, (2017) “Neural network Crash prediction Model of freeway based on Negative Binomial regression analysis,” Harbin: Harbin Institute of Technology.