

Effect Analysis of Recommended Parking Spots on Bike Sharing System Based on Clustering Algorithm

Shikun He ^a, Qiong Tian ^b, Jingtao Zhang ^c

Beihang University, Beijing 100191, China.

^abirimu@163.com, ^btianqiong@buaa.edu.cn, ^czhangjingtao921@163.com

Keywords: Clustering algorithm, Bike sharing, Effect analyze.

Abstract. Bike sharing has received a rapid development in recent China since 2016. However, the side effect from huge amount of sharing bikes poured into traffic networks has caused both side effect on traffic system and the restrictions from the government. In order to reduce the negative impact, recommended parking spots are made. In this paper, we build the time cost model and use clustering algorithm to compare the system performance before and after setting up the recommendation site by abstract model from real fact.

1. Introduction

As of February 2018, the number of bicycles in Beijing has soared to 2.2 million, which has exceeded 2.01 million, the ceiling estimated by Beijing traffic commission of the daily demand for sharing bikes [1]. Excessive cycles have seriously squeezed the road resources and have serious side effect on the whole traffic system.

In general, the process for a user to use a sharing bike is very simple: to find a bike first, then ride to the destination and lock the bike at last. The users can lock the bike near the destination at most with no limitation. This has brought the users much convenience. However, it has also brought big inconvenience at the same time [2]. We can often see a large number of shared bicycles crowded around CBDs, transit stations and pedestrian crossings, in no order. The disordered bicycles not only increase the operating cost and difficulty for the sharing bike company but also hinder the normal order of the urban traffic system. For example, the traffic speed can be quite slow where bikes are piled up, traffic accident can be more easily to happen, and riding in vehicle roads will interfere with normal driving of motor vehicle [3] s. In summary, every user in the transportation system has paid a price for bike sharing system,

In order to regulate sharing bicycle parking and maintain normal traffic order, the government and the bike sharing companies have taken certain measures [4]. For instance, the government has introduced policies to limit the increasing of sharing bikes and has marked the "sharing of bicycle parking spots" on some certain pedestrian crossings, the companies have also added slogans like "the current location is not suitable for parking, please put it in the recommended parking place, otherwise it will affect the personal credit" in pop-ups [5]. It seems that the recommended parking spots have been the most direct response to the pile-ups, but how effective is it? No certain quantify researches have been made in this field. So, for solving this problem, we build this model to evaluate the effect of recommended sites from several aspects such as total travel time, total travel cost, by using some operation data from the sharing cycling company.

2. Cluster Algorithm and K-Means Fast Clustering Method

Cluster analysis is a multivariate statistical method for quantitative analysis of multiple samples [6]. It is an unsupervised learning method, using distance or similar characteristics to classify samples [7]. According to the different analysis objects, it can be divided into q-type cluster analysis (for sample clustering) and r-type cluster analysis (for variable clustering). In this paper, the q-type clustering algorithm based on distance has been used. In addition, there are four main methods of cluster analysis: division method, hierarchy method, density-based method and network-based

method. Among them, the partitioning method is to divide each object based on distance into a certain group, which is very suitable for us to deal with the bike sharing data with latitude and longitude.

K-means clustering method is also called fast clustering method, which is a common method for clustering analysis [8]. It is often used for distance related data and is quite suitable for large samples. In this paper, K-Means clustering method is used to cluster the positional data of sharing bicycles, so as to get the virtual recommended parking spot for shared bicycles based on distance. The implementation process of KM algorithm is to calculate and iterate according to the given clustering number k , and then divide the data into k categories according to the specified distance calculating algorithm and objective function. The specific implementation steps are as follows:

Select the initial k centers for the k categories: $\mu_1 \mu_2 \mu_3 \dots \mu_k$

For each sample x_i , mark it with the category the same as the closest selected center

Update each category center to the mean of all samples belonging to the category.

Repeat steps 2 and 3 until the change of the category center is less than the specified threshold.

In this paper, we use the Euclidean distance. The formula of Euclidean distance is as follows:

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2} \quad (1)$$

The above expression represents the Euclidean distance of two points $A(x_1, y_1, z_1), B(x_2, y_2, z_2)$ in the three-dimensional space.

The clustering objective function we used is the sum of squares error (SSE), and to make the SSE the smallest. The formulation of SSE is as follows:

$$SSE = \sum_{i=1}^r \sum_{j=1}^{n_i} (X_{ij} - \bar{X})^2 \quad (2)$$

The formulation above expresses the degree of deviation of mean \bar{X} versus n_i samples X_{ij} in all r factors. The minimum SSE can represent the least variance between sample and the mean value. SSE is the most commonly used objective function.

3. Theoretical Model

We abstract the operation model of the bicycle sharing system as follows. We assume that in the sharing system, the total demand in the unit time is λ [9]. One user goes from starting point O to the destination D , and the distance between O and D is s . We mark the status with the recommended park point as A , and mark the status with no recommended park point as B . The process for users using sharing bikes in state A is: finding a shared bicycle at O , then unlocking the bike and riding to destination D , finally ending the using by locking the bike. While after setting up the "recommended parking spot", the user's sharing cycle in status B has changed: the user gets a bike from O and then rides to the recommended parking point D' near the destination, and finally walks to destination D on foot. The process in two statuses is expressed as the following Fig1.

The black line represents user behavior in status A , and the blue lines status B .

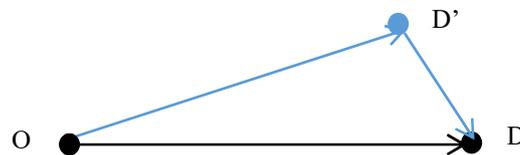


Fig. 1 User process in sharing bikes

We ignore the real gap between the truly final destination and the user's parking spot here because one can park the bike near the final destination as close as he can with no specific limitation in the current situation. In other words, D is the truly final destination of a user. Walking speed of the user is v_w , the unit time cost is α , riding speed is v_f , the unit time cost is β . Users think that riding is more efficient than walking, then $\alpha > \beta$. As riding speed is higher than walking speed, then $v_f > v_w$ [10]. In order to simplify the model, we do not consider the effect of the recommended parking spots on the behavior of the user before finding a bike which means that the starting point is always O .

The cost of each travel is divided into two parts: the time cost of cycling is C_t , and the ticket fair is F . We assume that the system is free of any parking constraint in state a , and all the travels are delayed because of the piled-up bikes. The delay rate is namely congestion coefficient, marked as f .

While in state B, for the sake of the recommended parking spots, parking is effectively controlled, so the users who accept the spots are no longer affected by the congestion.

We refer to the road congestion function BPR proposed by US highway administration here:

$$t = t_0 [1 + m (\frac{q}{c})^n] \tag{3}$$

t_0 Denotes the original travel time when the road congestion is not considered. m and n are undetermined parameters, q denotes the traffic flow of the path, and c represents the path capacity. $[1 + m (\frac{q}{c})^n]$ denotes congestion coefficient.

The US highway administration recommends $m = 0.15, n = 4$ [11]. In addition, for simplifying the model, we set q denotes the total amount of sharing bikes N , c denotes the maximum amount of sharing bikes N_{max} [12]. Then our congestion coefficient can be expressed as follows:

$$f(N) = [1 + 0.15 (\frac{N}{N_{max}})^4] \tag{4}$$

We assume that the user is completely rational, when the cost of a travel that the user adopts the recommended parking spot is higher, then the user will refuse the spot, and vice adopts it. The proportion of demands that accept the recommended spots in unite time is ρ . Other indicators in our model are as below [13].

Table 1. Main Formulas in This Model

Indicators \ status	Status A	Status B
Time for one trip	$T_{A0} = \frac{S}{v_f} f(N)$	$T_{B0} = \frac{S_1}{v_w} + \frac{S_2}{v_f} + \frac{S_3}{v_w}$
Cost for one trip	$C_{A0} = \frac{S}{v_f} \beta f(N) + F$	$C_{B0} = \frac{S_1}{v_w} \alpha + \frac{S_2}{v_f} \beta + \frac{S_3}{v_w} \alpha + F$
Total travel time in unite time	$T_A = \sum_{\lambda} \frac{S}{v_f} f(N)$	$T_B = \sum_{\lambda} (\frac{S_1}{v_w} + \frac{S_2}{v_f} + \frac{S_3}{v_w})$
Total cost in unite time	$C_A = \sum_{\lambda} \frac{S}{v_f} \beta f(N) + F$	$C_B = \sum_{\rho\lambda} (\frac{S_1}{v_w} \alpha + \frac{S_2}{v_f} \beta + \frac{S_3}{v_w} \alpha + F) + \sum_{(1-\rho)\lambda} (\frac{S}{v_f} \beta f(N) + F)$

4. Recommended Parking Spot Effect Analysis.

4.1 Sample Data.

We obtained the actual running data of shared bicycle companies, including the starting-ending longitude and latitude data. The thermodynamic diagram of the data is as shown below in Figure2.

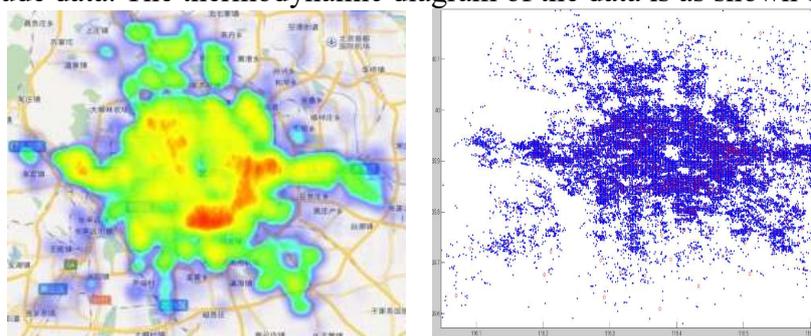


Fig. 2 Thermodynamic diagram and Clustering results of the Data

According to Figure2, we can find that there are relatively more demands in the center downtown area. The demand outside the five rings scattered more. In addition, more than 75 percent of the trips were within 1km. The average travel distance is 852m, and the longest travel distance in our data set is 34km, while the shortest is 116m. More than 70 percent of the orders were completed within five minutes, and the longest order took more than three hours. The average of rides per day per person is about 1.25 times.

4.2 Effect Analysis.

We set the time value parameter as $(\alpha, \beta) = (50 \text{ ¥/h}, 45 \text{ ¥/h})$, travel speed parameter as $(v_w, v_f) = (12 \text{ km/h}, 4 \text{ km/h})$, the fair cost is uniform as 1 ¥ [14]. We constantly enlarge the sample size, and calculate the system total cost, total travel time, adopt percentage of recommend

parking spot before and after perform the spots. The total data has been clustered based on the total number of N , shown in Figure2 on the right [15]. We find that the system has been improved both in the travel cost and in travel time.

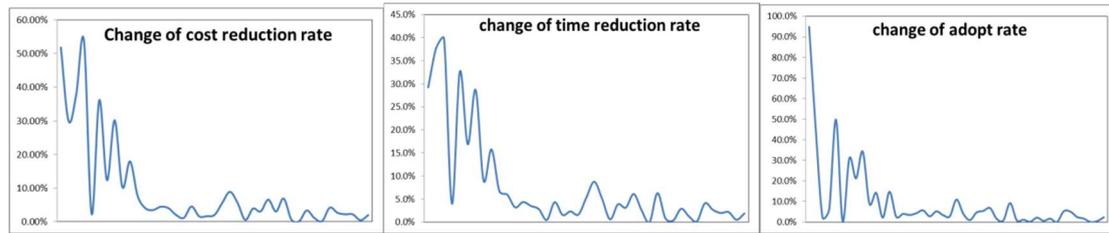


Fig. 3 Model Performance Results

Change of cost reduction rate. Since this model is based on rational user assumption, users must choose the way to make their travel costs not increase. Therefore, the total transportation cost of the system is decreased after the recommended parking spots are implemented compared with before. However, with the increase of sample size, the ratio of the system travel cost decrease rate after the execution of the recommended parking spots is reduced gradually to zero that means the cost reduction rate tend to a constant amount. According to our model data set, this rate tends to be stable at 2.7%. That is, the total cost of the system can be reduced by about 2.7% after the recommendation park site is executed.

Change of time reduction rate. The total cost of the system has also been reduced after executing the recommended parking spot. Although the total travel distance is increased, the total travel time of the system is also lower than before. It means that the positive effect of implementing recommendation site on travel time is more than negative. This is because the bike sharing system turn out to be in order after the recommendation park spots and it reduces the barriers for user travelling in the system. According to the figure in the middle above we can also find that the ratio of the system travel time decrease rate is also tend to be zero. Accordingly, the recommended parking spot can reduce nearly 3.6 percent of total travel time.

Change of adopt rate. The decrease ratio of the system adopted recommended parking spots rate also reduced with the increase of the sample amount as the two before and can be shown as the left figure above. The growth of the adopt rate is gradually tending to zero, which can tell that the rate of the adopted recommended parking spots tends to be constant value. According to the calculating, the ratio in our data turns out to be 19.5%. In other words, nearly one fifth of demand accept the recommended parking spots

In summary, the system with recommended parking spots turns out to be cheaper and faster, and we can tell the rate of the system improvement by given the data. Although the certain ratio can be varied with the factors such as number of shared bikes in the system, value of time, ticket fare, clustering algorithm, the overall trend can be still.

5. Summary

This paper has established a theoretical model to analyze the effect of bike sharing system after executing the recommended parking spots by using clustering algorithm to predict the spots based on the real data. There are several findings according to the data we got as follows:

After the implementation of the recommended parking strategy, the total cost of the system decreased by about 2.7%.

The total travel time of the system can have decreased by about 3.6%.

The system is willing to adopt the recommended parking spots, which accounts for about 19.5% of the total demand.

It can be said that the recommended parking spots for sharing bikes can effectively improve the whole system.

There are also some shortcomings in this model. Firstly, different clustering algorithms may perform kind of different when predicting the recommended parking spots, and we can in further consider more precise and advanced algorithm like density clustering algorithm. Secondly, the

numeric conclusion in this paper may be different by using different parameter, but the general trend will not change with varies situation. Thirdly, as limited by the data we can get, we use uniform congestion coefficient, which may cause deviation from the real fact. But with the bike sharing companies becoming more open, we can expect more actual data and get a better understanding of this newly famous transit mode.

References

- [1]. Information on: www.techweb.com.cn/internet/2018-02-13/2638900.shtml
- [2]. Tej Kumar Karki,Liu Tao. How accessible and convenient are the public bicycle sharing programs in China? Experiences from Suzhou city Habitat International, (2016),53. p. 188-194
- [3]. Xu-HuaYang, ZhiCheng, GuangChen, LeiWang, etl al. The impact of a public bicycle-sharing system on urban public transport networks. *Transportation Research Part A: Policy and Practice*. Vol 107 (2018), p. 246-256
- [4]. Zhang Yan,Zhao Min. Discussion on the Eficiency and Policy Orientation of Public Bicycle Sharing System in China. *Shanghai Urban Planning Review*,2014(06) p. 117-123.
- [5]. Information on: www.ahyouth.com/news/20170427/1294689.shtml
- [6]. YunlongFeng, Roberta Costa Affonso, MarcZolghadri. Analysis of bike sharing system by clustering: the Vélip' case. 20th IFAC World Congress. Toulous France, 2017.7.9-14, p. 12422-12427
- [7]. Sun Jigui,Liu Jie,Zhao Lianyu. Clustering Algorithms Research. *Journal of Software*, Vol. 1 (2008) No. 1, p. 48-61.
- [8]. Zhexue Huang,Michael K. Ng. A Note on K-modes Clustering[J] *Journal of Classification*. 2003 (2)
- [9]. Inês Frade,Anabela Ribeiro. Bicycle Sharing Systems Demand. *Procedia Social and Behavioral Sciences*, (2014),111. p. 518-527
- [10]. Aritra Pal, Yu Zhang. Free-floating bike sharing: Solving real-life large-scale static rebalancing problems. *Transportation Research Part C*, 2017 80, p. 92-116
- [11]. Wang Yuanqing,Zhou Wei,Lu Lian`en. Theory and Application Study of the Road Traffic Impedance Function [J]. *Journal of Highway and Transportation Research and Development*. 2004(09), p.82-85
- [12]. Sakari Jäppinen,Tuuli Toivonen,Maria Salonen. Modelling the potential effect of shared bicycles on public transport travel times in Greater Helsinki: An open data approach[J]. *Applied Geography*,2013,43. p. 13-24
- [13]. Yasuo Tomita,Akihiko Nakayama. Demand and cost structure analyses on Japanese successful bicycle sharing system called “Ekirin-kun” to install cycle ports at railway stations. *Transportation Research Procedia*, (2017)25. p. 3412-3420
- [14]. Borgnat, P. Abry, P. Flandrin. Shared bicycles in a city: A signal processing and data analysis perspective. *Advances in Complex Systems*, Vol. 14 (2011), No. 03, p. 415-438
- [15]. Shi Xiaoying,Yu Zhenhai,Xu Haitao. Clustering the Stations of Bicycle Sharing System. *Journal of Donghua University (English Edition)*, Vol 33(06) (2018). p. 968-972.