

# Research on Behavior Characteristics of Charging User of Electric Vehicles

Xu Xin <sup>1</sup>, Fu Jun <sup>1</sup>, Sun Zhijie <sup>1</sup>, Li Xuemei <sup>1</sup>

<sup>1</sup>State Grid Jibei Electric Power Co., Ltd. Electric Power Research Institute, North China Electric Power Research Institute Co., Ltd., Beijing100045;

**Abstract.** It is difficult to find the characteristics of electric vehicle users through charging transaction data. However, these characteristics are very important for charging station operators to improve the service quality of their operation and maintenance. This paper uses the K-means clustering model to study the transaction data of the charging stations in northern Hebei from October 2016 to September 2017. It finds three types of users: high-speed primary users, high-speed secondary users, and urban users; Based on this, four types of users are further drilled out: play users, cross-city office users, pass users, and urban residents users. Based on these user characteristics, corresponding operational strategies, charging station planning and construction suggestions are proposed.

## 1 Introduction

The charging station serves as a charging infrastructure for electric vehicles and is a complementary product to electric vehicles. In the national "13th Five-Year Plan" on the charging infrastructure of electric vehicles, under the guidance of the construction of a moderately advanced, car pile, smart and efficient charging network system, the coverage area of electric vehicle charging stations is expanding, and the users of charging stations are also increasing. Correctly identifying the behavior characteristics of the charging station user or the electric vehicle user is of great significance for the operating company to improve the operational efficiency, enhance the operation and maintenance service quality, and guide the user's charging behavior. This article will use the relevant data of Jibei Company's running charging piles to study the behavior characteristics of electric vehicle charging users, and put forward relevant suggestions from the user's label, which will lay the foundation for the research in other fields.

## 2 Data Processing and Research Methods

The study of behavior characteristics of charging user mainly relies on the data extracted from the vehicle networking management platform, including the charging pile file, and the charging transaction data in the north of Hebei Province. These data generally only involve the user's charge card number, charge amount, transaction amount, transaction method, transaction time, transaction station number, which is short of accurate description of the user. How to grab user characteristics

from existing data depends on the choice of the right research method. Grabbing user characteristics is essentially the classification of users, and cluster analysis is the most effective classification tool in data analysis technology. Cluster analysis is a set of models and tools, which can divide the data objects into several categories. Due to the relatively large data of the car networking platform, we chose the K-means model for clustering.

The use of K-means model has certain requirements on the data structure, but the data structure extracted from the vehicle network platform is poor. Therefore, it is necessary to combine the business knowledge to organize and preprocess the original data, and to obtain the structural data that are related to the electric car users, unified in format, complete in data, and standardized.

### 2.1 Data processing

Data processing converts raw data from a business perspective into a format that can be used by the K-means model. This includes representative processing, business processing, and K-means processing.

Representative treatment: The main group of electric vehicle users' charging behavior is the public, and the bus charging is not representative. Therefore, the transaction data of bus charging stations in Jibei Company's transaction data from October 1, 2016 to September 30, 2017 will be removed.

Business processing: According to the operational needs of operating companies, the site is divided into seven categories of Beijing related high-speed, inter-city high-speed, Chengde City site, Langfang City site, Tangshan City site, Qinhuangdao City site, Zhangjiakou City site; the trading period is divided into 0-3:59, 4-7:59, 8-11:59, 12-15:59, 16-19:59, 20-23:59, and 6

<sup>a</sup> Corresponding author: 676269930@qq.com

periods; The transaction date is divided into three types: workday, weekend, and holiday.

K-means model clustering requires variables to be numeric. Therefore, in addition to the card number identifying the user as an observation unit, the charging transaction data includes high-speed, city station charging times, workdays, weekends, and holiday charging times, and the number of charging times in each day time period, and the number of charging times for each type of charging station are all numerical value type. See Table 1 for details of specific variables.

K-means processing: Because the units of each variable are different, the variables need to be standardized; In order to prevent the large data from being dispersed, the distance is standardized by using the standard deviation within the class so that the class characteristics are more easily revealed.

**Table 1** variable details

Variable	Charging power	Charging times	High-speed station	City station	Beijing related
Name	Dl	Dc	Zg	Zs	Zb
Variable	Chengde City	Intercity speed	Langfang City	Qinhuangdao City	Tangshan City
Name	Zc	Zj	Zl	Zq	Zt
Variable	Zhangjiakou City	Workday	Weekend	Holiday	0-3:59
Name	Zz	Tg	Tz	Tj	T1
Variable	4-7:59	8-11:59	12-15:59	16-19:59	20-23:59
Name	T2	T3	T4	T5	T6

Through the above data processing, 20 variables are obtained to reflect the characteristics of the charging user. The data structure of these variables can be used for K-means model clustering calculation.

## 2.2 K-means model

The two basic steps of the K-means clustering model are: selecting similarity metrics, and selecting parallel algorithms. The similarity measure of K-means clustering is mainly based on the Euclidean distance, that is, the distance between the observation unit and the average value of each class needs to be continuously calculated. If the observation unit contains  $p$  variables, the formula for calculating the mean coordinates of each class is as follows:

$$\bar{x}_i = \frac{1}{n} \sum_{j=1}^n x_{ji} \quad (i = 1, 2, \dots, p) \quad (1)$$

Where  $n$  is the number of observation units included in this class,  $x_{ji}$  is the observation of the  $i$ -th variable of observation unit  $j$ . Therefore, the mean of this class is:

$$\bar{x} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_p) \quad (2)$$

The distance between the observation unit  $j$  and the mean of this class is the Euclidean distance of  $x_j =$

$(x_{j1}, x_{j2}, \dots, x_{jp})$  and  $\bar{x}$ , which also represents the distance between  $j$  and this class.

The parallel algorithm of the K-means clustering model is selected as follows:

Step 1: Divide the unit of observation into  $K$  initial classes.

Step 2: Traverse the list of observing units and assign each observing unit to one of its nearest  $K$  classes; recalculates the mean of the class that the class element changes.

Step 3: Repeat Step 2 until the target allocation reaches stability.

Step1 can be randomly divided into  $K$  classes. Step2 calculates the distance of each observation unit from each class, and then assigns it to the nearest class. If this kind of distribution occurs, it will inevitably lead to the reduction of one class element and the increase of another class element. Thus we need to recalculate the class mean for the changed class:

$$\bar{x}_{i,new} = \frac{n\bar{x}_i + x_{ji}}{n+1} \quad (3)$$

(if observing unit  $j$  is assigned to this class)

$$\bar{x}_{i,new} = \frac{n\bar{x}_i - x_{ji}}{n-1} \quad (4)$$

(if observing unit  $j$  is deleted from this class)

Where  $i = 1, 2, \dots, p$  is the  $i$ -th variable,  $\bar{x}_i$  is the mean of the original  $i$ -th variable of this class,  $\bar{x}_{i,new}$  is the new mean, and  $n$  is the number of observation units originally contained in this class. Step3 will repeat Step2 until  $K$  final classes are formed.

For the study of user characteristics, the observation unit is the charging user, or user ID, which is a total of 16012. Each user  $j$  has a total of  $p=20$  variables (see Table 1): D1, Dc, Zg, Zs, Zb, Zc, Zj, Zl, Zq, Zt, Zz, Tg, Tz, Tj, T1, T2, T3, T4, T5, and T6 correspond to  $x_{j1}, x_{j2}, \dots, x_{j20}$  in the K-means model.

The K-means clustering model's parallel algorithm can be implemented in most program languages or software packages.

## 3 User behavior characteristics

The following analysis is based on the transaction data of Yubei Corporation from October 2016 to September 2017. After standardization of data processing, the clustering analysis was performed on 16012 users who used the charging pile in the northern Hebei region by using the K-means model. The setting of  $K$ , and the number of classes that should be produced finally is a problem for any clustering model. One possible method is to make experiments with a given set of  $K$  values so that the number of elements in the class is relatively balanced. The final number of classes is selected according to the expertise.

### 3.1. Preliminary K-means clustering results

In order to cover the potential number of clusters, we perform K-means cluster analysis on K=1 to 15, and finally select the clustering result of K=6, and further determine the final number of clusters. The clustering result of K=6 is shown in Table 2.

**Table 2** preliminary clustering categories and number of members

Category	1	2	3	4	5	6
Number of members	5497	4769	5744	1	1	0

From the preliminary clustering results, the number of members included in category 4, 5, and 6 is only one or none, and it is not necessary to pay attention to these three categories. Therefore, three relatively balanced categories 1, 2, and 3 were initially obtained. The following three categories were compared from high-speed use, charging amount, charging times, and charging time.

**Table 3** comparison of three types of users

Categories of clients	Charging power	Charging times	8:00-20:00	High speed use ratio
High-speed first level	81.50	9.92	83.07%	71.80%
High-speed second level	37.69	3.65	85.78%	87.46%
City	78.28	9.90	67.62%	3.98%

From Table 3, it can be seen that the first and second types of high-speed stations are used more frequently for charging, and the third type of users rarely use high-speed stations for charging. Therefore, these three categories are initially defined as high-speed first-class

customers, high-speed second-level customers, and urban customers. The annual per capita charge level of users of the first and third categories is much higher than that of the second users. However, the average charging capacity of the three types of users is relatively significant, and the charging time is mainly concentrated between 8:00 and 20:00, and we need further attention.

The first type of user is a high-speed first-level user, with a total of 5497 people, accounting for 34.33% of the users in North Hebei. Most users have a high rate of high-speed use and a high number of uses. The number dimension is per capita, and high-speed use is relatively high; the number of charging power and charging times is the highest among the three types of users and slightly higher than the third type of users.

The second type of users are high-speed second-level users, accounting for 29.78% of the users in North Hebei. Most of the users have a high rate of high-speed use, but the number of users is low, with a total of 4769 people. The number dimension is per capita, and the high-speed use ratio is high; the amount of charging power and the number of charging times are the lowest among the three types of users, and are far lower than those of other types of users.

The third type of users are urban users, which account for about 35.87% of the users in North Hebei. Most of the users have a low rate of high-speed use, but the number of users is high, with a total of 5,744 people. The number dimension is per capita. The high-speed use ratio is low and close to 0; the charge power and the number of charge times are similar to the first type of users, and slightly lower than the first type of users.

**Table 4** the comparison of holidays and site of three types of customer

Categories of clients	Working day	Weekend	Holiday	Related sites in Beijing	Chengde site	Intercity speed site	Langfang site	Qinhuangdao site	Tangshan site	Zhangjiakou site
High-speed first level	5.18	3.18	1.57	5.45	0.00	1.23	0.41	1.60	1.23	0.00
High-speed second level	1.91	1.22	0.52	1.96	0.98	0.24	0.31	0.06	0.03	0.07
City	6.36	2.93	0.60	0.51	0.00	0.35	8.04	0.00	0.00	0.98

From Table 4, it can be seen that the number of high-speed first-class user workdays: the number of weekends used: the number of holiday use  $\approx 3.3:2.0:1$ ; the number of use is high during 8-19:59; site usage, Beijing-related high speed are far higher than other station.

The number of high-speed second-class user workdays: the number of weekends used: the number of holiday use  $\approx 3.7:2.3:1$ ; the number of use is high during 8-19:59; site usage, Beijing-related high speed are far higher than other station.

Number of use of urban user during workday: the number of weekends used: the number of holiday use

$\approx 10.6:4.9:1$ ; the number of use is high during 12-19:59; site usage, Langfang City Site are far higher than other station.

### 3.2 User Characteristics

Through the above research, the following results can be generally obtained according to the operation and maintenance business experience:

Most of high-speed customers are cross-city office users. The main feature of its charging behavior is that the charging power is larger and the number of charging times is higher in the workday in Beijing-Tianjin-Hebei

area, and the charging stations use frequent people, and they are often charged at relevant sites and intercity high-speed stations in Beijing.

Most of high-speed second-level customers are pass users and play users. It is characterized by a small amount of charge and a small number of charges, but the site of the charge is relatively evenly distributed.

Urban customers are basically urban residents and office workers in the same city. It is characterized by a relatively large number of charging days on the working day, but the charging station is not even.

## 4 The proposal

Based on the above research, we propose operational strategies for different types of users.

High-speed first-level customers: launch segment pricing and package services.

This strategy can be used to provide customers with a certain degree of concessions through various combinations of sub-time, sub-division, or weekday holidays, which attracts customers and improves service quality better.

High-speed second-level customers: Promote nearby scenic information, peripheral services (catering, etc.), and charging stations nearby.

This strategy starts from the point of interest of customers and can not only effectively serve customers, but also cooperate with surrounding scenic spots and other service industries to discover new service models.

City customers: Push the nearby low-frequency stations and give them discounts and push new sites.

This strategy can guide customers to use low-frequency sites at preferential prices, especially recommending new sites nearby. This not only reduces waiting time of customer but also makes full use of charging network resources.

Build city sites: Add charging stations around the city sites that charge high frequencies.

In the entire charging network system, the frequency of use of each site is very different. For example, the number of sites in the Langfang area is much higher than other sites. This requires the addition of new sites near high-frequency sites in the region, making the network layout of site more reasonable.

## 5 Conclusion

By sorting out the first-hand data of charging transactions in the northern Hebei region and using the K-means clustering of big data technology, we discovered some of the main characteristics of charging user behavior in the northern Hebei region. According to the characteristics of different users, recommendations are made for operation and maintenance management, planning and construction. These suggestions have practical significance and application value for the development of the electric vehicle charging network in northern Hebei and Beijing-Tianjin-Hebei. With the constant improvement of transaction data and the further

improvement of big data technology, there are more valuable customer information that we need to dig.

## References

1. Li Yue. Demand Prediction of Electric Vehicle Charging Station Based on User Classification[J]. *Electrical Generation*, 2013(12):98-100.
2. Wang Feilong, Sun Kaihang, Li Yanqing. Demand prediction of electric vehicle charging station based on user classification[J]. *Heilongjiang Electric Power*, 2013(02):132-134.
3. Xing Long, Zhang Peichao. Micro-grid operation optimization based on user demand classification for electric vehicle participation[J]. *Renewable Energy*, 2013(5):46-50.
4. Chen Lidan, Nie Yongquan, Zhong Qing. Prediction Model of Electric Vehicle Charging Load Based on Trip Chain[J]. *Transactions of China Electrotechnical Society*, 2015(4):216-225.
5. Chen Jingpeng, Ai Wei, Xiao Fei. Electric vehicle charging station planning based on user travel demand[J]. *Electric Power Automation Equipment*, 2016(6):34-39.
6. S. Speidel, T. Bräunl. Driving and charging patterns of electric vehicles for energy usage[J]. *Renewable and Sustainable Energy Reviews*, 2014(40):97-110.
7. R. Bi, J. Xiao, V. Viswanathan, A. Knoll. Influence of charging behaviour given charging infrastructure specification: A case study of Singapore[J]. *Journal of Computational Science*, 2017(20): 118-128.
8. J. B. MacQueen. Some Methods for Classification and Analysis of Multivariate Observations[J]. *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, CA: University of California Press, 1967: 281-297.
9. R. A. Johnson, D. W. Wichern. *Applied Multivariate Statistical Analysis* [M]. Pearson Prentice Hall, 2007.