ISSN 1330-3651 (Print), ISSN 1848-6339 (Online)
https://doi.org/10.17559/TV-20230426000575
Original scientific paper

Spatio-Temporal Analysis of Electric Cab Charging Behavior Based on Order Data

Junze YU*, Jiulin SONG, Jingya LIU, Xiangke CUI, Daqing GONG, Qiuyan ZHANG

Abstract: Due to the imperfect measures of overall construction and operation management of charging infrastructure in China, it leads to the problem of low overall utilization rate and difficult charging for users. By analyzing the charging characteristics of electric vehicles, this paper can derive their charging laws to guide the scientific and reasonable layout planning of charging piles; starting from the charging behavior data of electric cabs, the most common means of transportation in the field of public transportation, this paper carries out a study on the charging patterns and factors influencing charging behavior of electric cabs to improve the utilization rate of charging facilities on the whole. This paper analyzes the user-level demands reflected by the charging patterns of electric cabs, and the charging facility operators can make relevant improvements to the charging station operation system accordingly.

Keywords: charging behavior; charging facilities; cluster analysis

1 INTRODUCTION

Since 2015, China's new energy vehicles have formed a better industrial base, and electric vehicle sales have achieved steady growth. In recent years, China has made the development of new energy vehicles an important strategic initiative to address climate change and promote green development. Since 2020, a number of guidelines and financial subsidies have been introduced at the national and local levels to expand the scale of new energy vehicle applications in public services and to promote the development of the new energy vehicle market and the construction of infrastructure services. According to the "New Energy Vehicle Industry Development Plan (2020 - 2035)" issued by the State Council, public vehicles such as pure electric cabs will be fully electrified by 2035. According to the statistics of the China Automobile Association, from the perspective of new energy vehicle ownership, China's new energy vehicle ownership reaches 7.84 million in 2021, an increase of 2919800 compared to 2020, of which 6.4 million are pure electric vehicles. In terms of new energy vehicle sales, in 2020, China's new energy vehicle sales will be 1.367 million units, up to 10.9% year-on-year. In 2021, new energy vehicle sales will be 3.521 million units, up over 160% year-on-year, accounting for 13.4% of total vehicle sales, with passenger cars being the main model. It is expected that by 2030, the number of new energy vehicles in China can reach 80 million volume.

The transformation of traditional fuel vehicles to new energy vehicle industry has become an irreversible development direction. To promote the comprehensive electrification of public sector vehicles, the development and construction of charging infrastructure is an important part of the process. The core difference of pure electric vehicles compared to fuel vehicles is that the vehicle drive is entirely generated by the drive motor, while the energy of the whole vehicle is derived from the power battery. Pure electric vehicles need to consume longer time for power replenishment, so the charging infrastructure becomes a major limiting factor. Although the growth rate of China's charging infrastructure has gradually accelerated in recent years, the phenomenon of difficult charging is still prominent. Public data shows that the utilization rate of charging piles in China is only about 4%. The utilization rate of public charging piles in many cities is seriously polarized. On the one hand, some charging pile spaces are occupied by oil cars, some charging piles have to attach parking space fees to the basic cost, and some charging piles cannot be repaired even after the operation platform fails; on the other hand, in the early construction, some enterprises' site locations do not consider the density of passenger flow and utilization rate, and the layout is unbalanced. Under the fierce competition in the market, many enterprises began to layout non-vehicle business and put forward charging station ecological operation strategies, such as setting up automatic food sales, massage chairs, and charging treasure around charging stations to improve the service capacity of charging stations and increase the operation revenue.

Electric cabs are the most frequently used new energy vehicles in urban transportation. Taking Shanghai's electric cabs as an example, the main pure electric vehicle model put into use is the Rongwei Ei5, with a range of 400 km. A one-time full charge takes 0.67 hours for fast charging and 8.5 hours for slow charging. And with factors such as seasonal and temperature changes, the battery range will also experience seasonal changes, and when the weather is cold, the battery range drops to about 350 km. The utilization rate of the charging pile is divided into time utilization rate and power utilization rate. The time utilization rate is calculated as the charging time ratio of 24 hours. According to the municipal platform for public data collection and testing of charging and switching facilities in Shanghai, the average utilization rate of DC piles is only 7%, and the average utilization rate of AC piles is only about 1.5%. Although the current pile-to-vehicle ratio in Shanghai has reached 1.2:1, there is much room to improve the usage frequency of charging piles.

The low utilization rate of charging piles not only causes the waste of charging infrastructure construction resources, but also is directly linked to the operating income of the charging operation platform. The existence of the two extreme phenomena of difficult charging and idle charging piles is still a problem that needs urgent improvement in the development of electric vehicles in recent years. In order to promote the reasonable construction of charging infrastructure, greatly improve the utilization rate of charging piles, help improve the convenience of travel of new energy vehicle drivers, and increase the profitability of charging infrastructure, city managers should pay attention to and encourage the research related to charging behaviours laws, including the prediction of charging load and charging demand through statistical modelling methods, and the behaviours excavated from a large amount of real charging behaviours data. This will help city managers to make scientific and effective decisions on the rational layout of charging facilities.

2 LITERATURE RESEARCH

2.1 Charging Behavior Analysis

Understanding the behavioral characteristics and influencing factors of charging station users can guide the direction for the optimization of charging facilities and the development of effective charging preference policies to avoid the problems of difficult charging, or mismatch between the supply and demand of charging facilities. Early studies mainly relied on questionnaire surveys and empirical data to understand the charging preferences of electric vehicle owners [1]. Some scholars analysed the usage characteristics of new energy vehicles in terms of demographic attributes [2]. Some scholars collect real operational data from charging stations in cities to portray and model and analyse the usage of charging stations and the charging behavior of EVs. For example, the literature [3] analysed the utilization characteristics of charging facility operation and the temporal and spatial distribution characteristics of vehicle charging from the charging facility and vehicle side by extracting 2700 charging facility operation data and 5000 new energy trucks operation trajectory data, respectively, and in some areas where charging facilities are more abundant, the electric logistics vehicles start charging with low power, while the average power is around 40%. At the same time, the authors evaluate the utilization of public charging posts and obtain a utilization rate of only 11% for social public charging facilities in Beijing. Flammini et al. [4] investigate the information on EV charging transactions (charging time, idle time, connection time, power and energy) in the Netherlands and build a multimodal mixture model to represent the probability distribution of these multimodalities. The findings include that more than 50% of the recharge durations are less than four hours, the idle time that remains occupied after a full charge is approximately equal to four hours, and that the idle time is characterized differently with geographic location, with owners located in residential areas usually not unplugging immediately after a full charge, while the idle time after a full charge is more discrete in other geographic locations. The phenomenon of longer idle times in residential areas is also noted in another author's paper [5]. In their paper, Ucer et al. [6] simulate the possibility of different charging station setup scenarios, exploring "destination-based" and "community-based". The paper explores the charging demand, quality of service, and power of "destination" and "community" urban DC fast charging stations, derives three charging modes: "urban", "suburban", and "rural". The peak power of charging stations was found to be very dependent on the power acceptance of electric vehicles.

With the development of big data technology, data-driven approaches are becoming increasingly popular in domestic and international research on charging behaviour to solve charging infrastructure planning problems. First, regression models are a common approach for scholars to explore the factors influencing charging station operations and consumer charging behaviour. In a linear regression model, we assume that the dependent variable obeys a Gaussian distribution and that the input explanatory variables are continuous and have a linear relationship with the target variable. In logistic regression, we assume that the dependent variable obeys a Bernoulli distribution and maps the results of linear regression to (0, 1) by a Sigmoid function, which is commonly used to solve classification problems. From the perspectives of charging facility supply and charging demand, Kang et al. [7] proposed a method to estimate charging demand indicators (CDI) based on location-based service data (LBS) in their paper, which reveals the dynamic relationship between the spatial pattern of charging demand and the distribution of public charging stations through kernel density estimation, and establishes a regression model to confirm the spatial structure characteristics of cities and the facility distribution on charging behaviour. Using polynomial logistic regression techniques, Wolbertus et al. [8] identified the key factors explaining the heterogeneity of charging station charging, duration behaviour and investigated the relationship between charging stations, price of charging stations, power of charging posts and drivers' charging duration in different Dutch cities, midweek versus weekend, and time of day. Almaghrebi et al. [5] used the same approach for four charging modes by converting the problem of charging behaviour into a logistic regression classification problem of temporary, parking, work and home, where temporary charging behaviour is within 2 hours, while home charging usually lasts more than 12 hours. There are three main categories of input variables studied, the day of the week, the hours of the day, and the location of charging. Using the idea of geographically weighted regression, this paper by Li (2019) investigates the variability of the urban built environment in the spatial dimension on online taxi trips. Using online taxi trips in Chengdu as the study target, this gap in the past literature is filled by using geographically weighted regression (GWR) to examine the spatial heterogeneity of possible effects. The results show that the GWR model outperforms the global regression model (OLS) in terms of goodness-of-fit. It is demonstrated that the coefficients of the effects of different urban architectural and spatial features on online taxi trips vary spatially. In this paper, the authors consider geographic characteristics in urban space such as mixed land use, population density and road network density [9]. The paper by Montfort (2016) investigates the factors influencing the operation of public charging stations in The Hague, Netherlands, and their four hypotheses are confirmed; first, the higher kWh usage of electric vehicles with higher charging battery capacity; second, higher charging station density leads to higher kWh usage; third, higher address density leads to lower kWh usage; and fourth, higher vehicle density leads to lower kWh usage [10].

2.2 Optimization of the Location and Size of Charging Stations

The location of charging stations has a significant impact on the experience of using new energy vehicles. Therefore, minimizing users' charging waiting time and grid load is also an important research aspect of charging station characteristics, location and scale optimization. Kong et al. [11] proposed a two-tier architecture for EV siting planning in their paper considering economic benefits, traffic congestion, users' evaluation, and grid load to systematically design and empirically prove the problem of charging station siting. Yang et al. [12] found 666 potentially suitable locations that can be used as charging stations by analysing and extracting the pattern of the time and place where cabs rest during the journey from real trajectory data of cab fleets, and established an integer linear programming approach with the objective function of minimizing the construction cost to study the trade-off between installing more chargers and providing more waiting space. Chen et al. [13] similarly investigated the problem of optimal location assignment of charging stations by incorporating station accessibility, local job and population density, and trip attributes into the variables of an integer linear programming that considers minimizing the cost of station access for EV users for determining the optimal location for installing a limited number of charging stations within 10 miles of downtown Seattle, Washington [12]. The literature simulated the distribution of EV charging based on the distribution of residential load, divided the planning area into zones, determined the weighting factors of station locations in terms of convenience, traffic flow, and land acquisition costs, and considered operating costs, network loss costs, and charger investments in the objective function to propose an optimal economic model for EV siting and capacity determination. A framework for deploying fast charging facilities suitable for electric cabs in large cities is proposed in the literature [13]. GPS trajectory data of conventional fuel vehicles are used to determine the charging availability level of the minimum feasible infrastructure by using DBSCAN clustering method and numerical optimization algorithm. The literature [14] proposes a mixed integer nonlinear programming model for location and scale optimization considering construction cost and grid load, and solves it using heuristic algorithms. Similarly, the literature [15, 16] proposes a heuristic model considering minimizing the number of sites. The literature proposes a siting model for charging facilities that do not exceed a given waiting time based on queuing theory and interception problem. The literature similarly considered the flow in road network planning and constructed an improved interceptor siting model to obtain an electric vehicle siting scheme in Taiyuan City. The literature proposed a charging probability calculation model based on vehicle travel records, and based on the *P*-median model and greedy algorithm, a parking lot that meets the demand is obtained from the alternative set with the goal of distance minimization.

2.3 Forecasting Study of Charging Demand

Taking a data-driven perspective, many other papers have gone beyond the analysis of user behaviour and attempted to predict a wide variety of charging patterns. These classes of studies generally model the driving characteristics and charging patterns of new energy vehicles to predict the charging demand at different times and locations, which can provide accurate information on the implementation of public charging stations in stages and multiple categories to help the infrastructure of public charging facilities to face the challenges that will come in the future. Through traditional statistical methods, Babrowski (2014) determined the charging behaviour of electric vehicles on weekdays and weekends by analysing multiple charging stations and interpreting travel data from six countries. Data from charging stations and travel data were used to predict the amount of electricity needed to charge the vehicle [17]. Many scholars used Monte Carlo simulation to predict the charging load of electric vehicles, for example, literature considered charging time, duration, electrical energy demand, and prediction of charging power for private cars in slow charging mode. Literature analysed the prediction of charging load for multiple vehicle types and charging modes; literature considered the effect of physical factors such as weather, space, on charging load prediction results. The literature constructed a charging demand prediction model from two perspectives: the influence of remaining power and willingness to charge and the relationship between actual parking duration. The authors of the literature [18] compared the charging demand predicted by two methods, Monte Carlo simulation and support vector machine, based on charging behaviour, considering arrival time and charging duration, and found that the support vector machine model is more sensitive to charging demand fluctuations and has good prediction results. In [19], the authors discuss the importance of accurate EV charging demand prediction in households, using a variety of widely used machine learning algorithms to predict the occurrence of household day EV charging and "no-charge" days, respectively. The performance of these algorithms is evaluated and compared. The literature [20-22] provides diverse methods such as kernel density estimation for charging demand prediction.

2.4 Literature Review

In summary, by reviewing the research results of EV charging infrastructure planning, we can divide the problems of charging facility planning into four major categories: portrayal of charging behaviour, optimization of charging station location and scale, prediction of charging demand, and operation and evaluation of charging facilities. The portrayal of charging behaviour includes the correlation analysis of charging behaviour with spatial characteristics in time, considering the influence of different seasons, dates, times, densities, and buildings in cities on charging demand, and is often regarded as a classification or regression problem, often using statistical modelling and machine learning techniques. Charging station layout is a typical planning problem that seeks the optimal configuration and size of charging stations considering economic factors, distribution network operating parameters, and the convenience of EV drivers. Charging demand prediction is a dynamic, multidimensional and multivariate problem to forecast the charging service demand at different times of the day and different locations. Charging operation model and evaluation method is a problem of charging infrastructure evaluation system which is a comprehensive consideration from society, users and enterprises. It can be seen that in the analysis and portrayal of charging behavior and charging demand, there are successive studies that take the geographic location of charging stations, charging time, and charging mode as explanatory variables, and in these studies, the characteristics of time and the spatial structure characteristics of cities are usually taken as explanatory variables. In the collection of geographic location information data, there is no specific division of the population or analysis of charging behaviour for a specific population, and the collection of geographic features is mainly through the data of points of information (POI) without considering the characteristics of road density, station density, etc. of the city. Based on these studies, this research will extend this part to mine and cluster the temporal and spatial features of charging behaviour of electric cabs. At the same time, this paper will combine some research methods of geospatial statistics with the attributes of geographical features of the study area to provide a multi-dimensional analysis for the research related to the influencing factors of the charging behaviour of electric cabs and provide fact-based suggestions for the future operation of charging platforms and optimization of charging facilities.

3 DATA ANALYSIS

3.1 Demand Situation

Since the first pilot operation of electric cabs in Shanghai in 2010, by the end of 2020, the number of licensed electric cabs in the city reached 6400 units. In April 2020, the Shanghai Municipal Development and Reform Commission released the "Shanghai Energy Conservation and Climate Change Reduction in 2020 Emission Reduction and Climate Change Response Key Work Arrangement". It is required to promote the electrification and cleanliness of urban transportation and vigorously promote pure electric cabs, and the proportion of electric cabs will reach 80% by 2022, about 36000 volumes. The main pure electric vehicle put into use in Shanghai is the Rongwei Ei5, with a battery capacity of 52 kWh and a theoretical range of 400 km. In actual driving, the range of electric vehicles will be compromised in the case of low ambient temperatures, with a range of about 380 km.

Table 1 Electric rental car ownership 2020 - 2024

Year	2020	2021	2022	2023	2024
Projected electric cab ownership	6690	16533	25619	31309	35763
Estimated annual charging capacity (million kWh)	16857	41658	64553	78891	90114

Based on the number of expiring oil vehicles and the rate of renewal estimated for 2020 - 2024, the ownership of electric cabs is shown in Tab. 1. From the data in Tab. 1, it can be seen that the number of electric cab holdings will continue to increase from 2020 - 2024, with the fastest growth in holdings from 2020 - 2021 until 2024, when the number of electric cabs will reach 35763.

3.2 Site Construction Status

Faced with the year-on-year growth in demand for new energy vehicle charging, Shanghai is also accelerating the layout of public charging stations. Since 2016, the Shanghai Municipal Development and Reform Commission has launched an official charging APP, "Shanghai Charging and Switching Facilities Public Data Collection and Monitoring Platform", or Lianlian Charging for short. Lianlian Charging went online in 2016 and has access to 138 charging and swapping facility operators by the end of 2020. Starting in 2020, the Shanghai Municipal Development and Reform Commission and Transportation Commission will launch 10 - 15 new energy cab charging demonstration stations each year. The new energy cab demonstration stations are charging stations that prioritize new energy cab services, are based on fast charging piles, have standardized service processes, are fully equipped with facilities, and actively play a demonstration role. In 2020, Shanghai launched the first batch of 21 charging demonstration stations for rental cars, and in 2021, 24 new demonstration stations were added. According to the Shanghai New Energy Vehicle Promotion Office, as of the end of March 2021, the municipal platform has access to a total of 112400 public and dedicated charging facilities. The city's vehicle-pile ratio is about 1.1:1, but the DC pile is only about 20%. The DC charging piles in Shanghai are mostly 30 - 60 kW, while the AC charging piles are mostly of 7 kW type and take 0.67 hours for fast charging and 8.5 hours for slow charging when fully charged at once. In the fast charging mode, the electric cab will reach 80% of power in half an hour, and the subsequent charging speed will slow down due to current protection. Literature [23] mentioned in the analysis about the current situation of electric cab charging market that there are only 375 charging stations and 3524 charging piles suitable for electric cabs in the city.

According to the characteristics of population mobility in Shanghai, the demand for electric cab charging varies greatly in each district. According to [23], the top three administrative districts in terms of cumulative charging volume and number of orders for electric cabs in 2020 are Minhang District, Pudong New District, and Baoshan District. The bottom three administrative districts are Hongkou District, Jinshan District, and Changning District. In order to meet the rapidly growing demand of cab charging volume, improve the service efficiency of existing charging stations and arrange reasonable charging stations in the future are the important directions that need to be focused on for supporting charging facilities nowadays.

3.3 Data Overview

This paper collects data on the charging behaviour of cab drivers at a charging station operating platform in Shanghai from September 2020 to February 2021. Among the 70000 order data collected, the charging stations covered 15 districts in Shanghai except Chongming District. The collected order data included 14 fields, including order number, station name, charging start time, end time, charging duration, power (kWh), charging cost, order amount, and unique user ID. The order number is the unique charging record generated with each charge. The unique user ID is associated with all charging records of a driver. The charging order data format is shown in Tab. 2. A total of 73931 charging messages were recorded during this period. For effective EV charging facility analysis, charging records with charging time less than 5 minutes were removed in the data collation, while records with charging power of 0 were also removed in the beginning stage. For EVs, it takes more than 20 minutes for a one-time full charge even with the most powerful charging post. Therefore, many orders with short charging times are due to charging posts not being able to connect to the network, or their own technical failures that prevent the car from charging. After filtering the data, a total of 68680 data were retained, with the retained valid data accounting for about 93% of the initial orders. This part of the data also indicates that there are many charging piles in the public charging facilities for new energy vehicles in Shanghai that have usage problems and need timely attention and follow-up maintenance from the operation platform.

A preliminary analysis of these 60000 data was conducted to calculate the power of each charging order, and those with power greater than 10 kWh were marked as fast charging (DC charging posts), and those with power less than or equal to 10 kWh were marked as slow charging (AC charging posts). Descriptive statistics of the main indicators can be obtained from the data including 152 charging stations, 4149 cab drivers, 65739 fast charging orders and 2941 slow charging orders. Statistical results can be obtained as in Tab. 3.

Table 2 Data field description

Field number	Field Description	Sample
1	Order No.	4250107650000000000031936042
2	Site Name	Shanghai Guangqi City Charging Station
3	Order Status	Payment completed
4	Starting time	2020/9/18 13:28:19
5	Ending time	2020/9/18 13:30:36
6	Charging time / sec	137
7	Electricity / degrees	0.62
8	Electricity	0.69
9	Service Charges	0.22
10	Parking fee waiver	2
11	Cab discount	0
12	Total order amount	0.91
13	Whether to print the small ticket	No
14	User ID	12000000000000120662

From the data statistics in Tab. 3, it can be seen that the main charging method of cabs is still mainly fast charging. Among all the orders collected, the number of orders using slow charging piles only accounts for less than about 5% of the total, and the average charging power through the slow charging method is only 7.66 degrees, while the loaded battery power of the Rongwei Ei5 is 52.5 degrees. The low number of orders and low average charging power are two characteristics that indicate that slow charging is not the main charging method for electric cabs, and when using slow charging, most electric cabs do not wait until the car is fully charged before they stop charging. The use of slow charging piles is very low among the electric cab driver community.

Table 3 Main field statistics

	Fast charge	Slow charge	Total
Number of orders / pcs	65739	2941	68680
Number of stations	-	-	152
Number of cab drivers / pcs	-	-	4149
Average charging time / sec	2947.10	5535.62	3057.94
Average charging power / degree	22.76	7.66	22.11
Average electricity charge / yuan	13.18	4.86	12.83
Average service charge / yuan	10.21	3.60	9.92
Average parking fee reduction / yuan	0.89	0.96	0.89
Average order amount / yuan	23.46	8.46	22.81

4 BASIC ANALYSIS OF CHARGING BEHAVIOUR

4.1 Charging Time Characteristics Analysis

An analysis of the characteristics of the start time and end time of charging of electric cabs will give us insight into the overall charging behaviour preferences of cab drivers. Charging start time is the key variable to focus on in the study of queuing phenomenon at charging stations during peak periods. Charging operation platforms can avoid the occurrence of congestion at charging stations during peak charging periods by focusing on the peak and low values of charging at different stations and formulating relevant policies. The distribution of charging start time of electric cab drivers is shown in Fig. 3 and Fig. 2. From the figure, it can be seen that charging orders start at 6:00 a.m. at the least, and the number of charges per hour increases from 6:00 a.m. 11 - 15:00 a.m. and 22 - 24:00 a.m. are the two most concentrated time periods for starting charging, and the number of charges reaches two peaks at 11:00 a.m. and 24:00 a.m. The reason for the concentrated charging of cabs during these two time periods may be that these two time periods are after the morning and evening rush hours of urban traffic, respectively. Taxis arrive indirectly during the rush hours with many passengers, travel long distances and consume a lot of power, therefore, electric vehicles need to be charged in time. Fig. 1 also shows the distribution of charging end time. The curve of charging end time and the curve of starting time fluctuate similarly, with a delay of about one hour in the number of changes, which is consistent with the average value of charging duration derived above. From the above, it is clear that the average value of the charging duration is around 3058 seconds, which is close to about an hour. Thus, the peak of the charging end time is one hour after the peak of the charging start time, at 12:00 and 1:00, respectively.

The above analysis is about the general characteristics of charging start time; in practice, the charging choice of cab drivers is influenced not only by time but also by geographical location information. Thus, it is necessary to analyse the charging behaviour corresponding to different types of public charging facilities. Charging stations are divided into six categories: enterprises and institutions, public institutions (parking lots, parks), office buildings, large buildings with parking lots (furniture cities, shopping centers), large cultural and sports facilities, residential areas and parks. Fig. 2 provides statistics on the distribution of charging start times for different types of charging stations, and the vertical axis indicates the ratio of the total number of charging times for a certain time period in a category. The statistical results show that at 11 - 12 o'clock, office buildings, enterprises and public institutions have the highest percentage of charging times in their respective categories, which overlaps with the charging peak time period in the previous section. However, at 23 - 24:00, the number of charging stations for all three types of charging stations shows a decreasing trend, which is opposite to the overall trend. Office buildings, enterprises and institutions, and large cultural and sports facilities have the lowest number of charging at 24:00. The peak of charging in residential areas is concentrated at 21 - 1 o'clock. It is worth noting that another peak in residential areas occurs at 4 a.m., while 4 a.m. is the low point of charging at all other types of charging stations. The main reason for a peak at 4:00 in residential areas is that for cab drivers, the work day can reach 13 hours. Whereas cab drivers usually start work at 4 or 5 a.m., a percentage of drivers tend to fully charge their cars before going to work each day, and this percentage of drivers will choose the closest place to their residence.

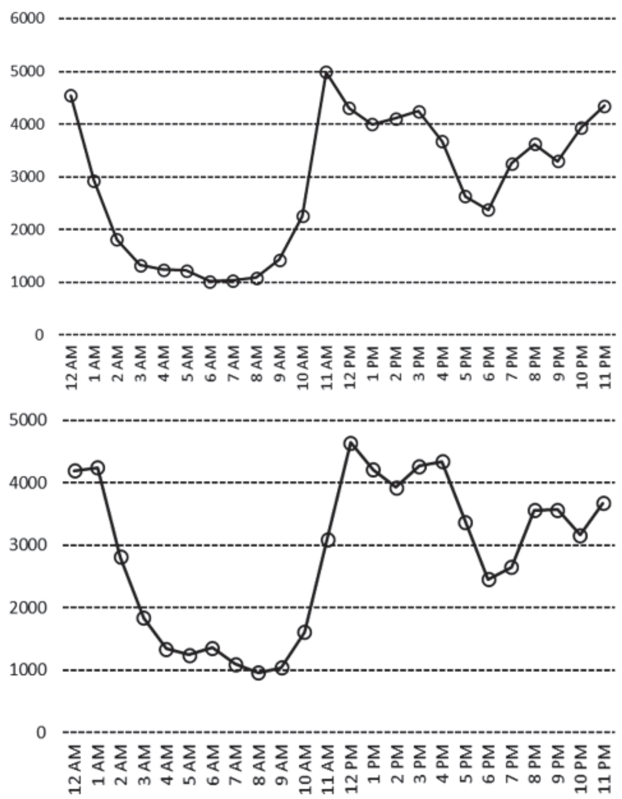

Figure 1 Charging start and end time

Based on the charging time of cab drivers and the duration of work, the charging time periods were divided into two categories, namely daytime charging (6:00 - 18:00) and evening charging (18:00 - 6:00). This yields 33,903 charging orders at night and 34777 charging orders during the day.

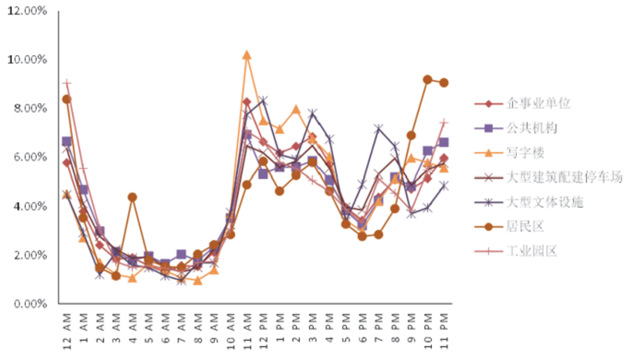

Figure 2 Percentage of time to start charging for different types of charging station locations

We propose the index I of cab charging time characteristics to describe the propensity of cabs to charge during daytime and night-time. It is calculated as:

$$I=\frac{x_{\text{day}}}{x_{\text{day}}+x_{\text{night}}} \tag{1}$$

where, x_{day} indicates the number of times drivers charge during the day, x_{night} indicates the number of times drivers charge at night, and I indicate the percentage of daytime charging, which describes the tendency of cab drivers to charge during the day. The index I was calculated for these 4149 drivers and preliminary statistics were performed to obtain the median and mean of the sample both at 0.58, as shown in Tab 4.

Table 4 Descriptive statistics of electric cab drivers' daytime charging preferences

Electric cab drivers tend to charge during the day	
Mean	0.58
Standard Error	0.01
Median	0.58
Count	4149
Confidence Level (95.0%)	0.01

The drivers in the sample were classified according to the different charging time propensities exhibited by the ratio of daytime charging times to night-time charging times. Drivers with daytime charging propensity I at [0.75, 1) were classified as daytime charging propensity drivers, at [0, 0.25) as night-time charging propensity drivers, and at [0.25, 0.75) as drivers with no significant time charging propensity. We can get 38.61% of the total number of daytime charging inclined drivers, 23.86% of the total number of night-time charging inclined drivers, and 37.53% of the total number of drivers with no obvious time charging inclination.

4.2 Analysis of Charging Time

Charging duration includes single charging duration analysis and interval charging duration analysis. Charging duration is a basic characteristic of cab charging behaviour and an important evaluation index of charging infrastructure. Single charge duration refers to the duration of one charge by the driver, and interval charge duration refers to the interval between two charges. The charging duration is related to the price of the charging station, the power of the equipment, and possibly to the personal habits of the cab driver. We first analyse the overall single-charge duration, using half-hour as the group spacing, and get the distribution of cab single-charge duration as shown in Fig. 3.

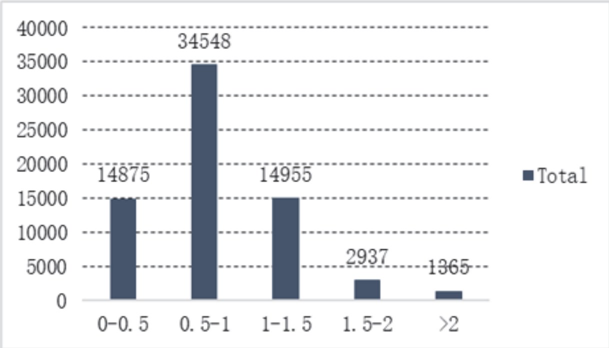

Figure 3 Electric cab single charge length distribution

As can be seen from the graph, the highest value of charging time frequency for electric cabs is between 0.5 - 1 hours, accounting for 50.30% of the total. The next most frequent interval of charging hours is 1 - 1.5 hours. The number of orders after one hour shows a decreasing trend. This data confirms the conclusion drawn above that the main charging method for cab drivers is fast charging, where cabs can have their batteries fully charged in 45 minutes to an hour and a half under fast charging piles. This indicates that the overall characteristics of the charging characteristics of electric cab drivers are predominantly fast-charging methods and short single charging times.

Taking cab drivers as the observation object and calculating the average charging time of each driver, we can get the user ID with the longest average charging time is 12000000000000144583 and the average single charging time is 9.15 hours.

The average charging interval length of a driver refers to the time difference between each charging order of a driver on a weekday. The average charging interval length reflects the driver's charging style habit, and the charging station operation platform can predict the future charging quantity based on the driver's usual charging style habit. We default the date when drivers do not charge as a non-working day and the date when they have charging orders as a working day. Firstly, the orders with different user IDs are filtered out and the first order start time T_1 and the last order start time T_n recorded in the time period of this study are extracted, and the different dates with orders are recorded as working days D_w and the number of charging is recorded as N. Then the average charging interval length of a single taxi driver is T_d:

$$T_d = \frac{T_n - T_1 - \left(dif\left(T_n - T_1 \right) + 1 - D_w \right)}{N - 1} \tag{2}$$

The role of the *dif* function is to calculate the number of days difference between T_n and T_1. Here is a specific example for analysis. For example, the driver with ID "12000000000000113148", the first recorded order time is 2020-09-28 18:42:39, the last recorded order time is 2021-02-28 14:51:00, then the number of days difference between the two orders is 152.84 days, the total difference between the last order and the first order is 153 days, and the number of days with orders recorded in the cycle is 105 days recorded as working days D_w, and there are 181 orders in these 105 days. Then according to the above formula, we can calculate the average length of time between charging for this driver is 0.63 days, which is 15.12 hours. Calculating the average charging interval length of all drivers in the order data, we can get the top five cab driver IDs with the largest interval length, as shown in Tab. 5. The descriptive statistics of the sample is also obtained, with an average charging interval length of 0.58 days (13.92 hours) and a median of 0.70 days (16.80 hours), as shown in Tab. 6.

Table 5 Top 5 average charging interval hours for electric taxi drivers

Cab DriverID	Interval / days
12000000000000124955	1.95
12000000000000132084	1.91
12000000000000102596	1.78
12000000000000119419	1.77
12000000000000157065	1.77

Table 6 Descriptive statistics of the average charging interval length of electric cabs

Average length of time between charges for electric cabs	
Mean	0.58
Standard Error	0.01
Median	0.70
Count	4149
Confidence Level (95.0%)	0.01

4.3 Concentration Analysis of Charging Station Locations

In addition to the different preferences in the average charging interval and charging time, different drivers may have different habitual characteristics in the selection of charging stations, some drivers may be accustomed to charging at a certain station, while some drivers are accustomed to charging near the pick-up and drop-off points. The former is more concentrated in the number of charging stations reflected in the order, and the latter is more scattered in the choice of charging stations. Therefore, it is necessary to make a suitable analysis of the concentration degree of the choice of stations for the group of electric cab drivers, so we propose an index describing the concentration degree of drivers' charging stations (Concentration Ratio), which is denoted as CR. CR is calculated as follows:

$$CR = \frac{S_1^2 + S_2^2 + \ldots + S_n^2}{\left(S_1^2 + S_2^2 + \ldots + S_n \right)^2} \tag{3}$$

S_1 is the number of orders initiated by the driver at the first charging station, and S_n is the number of orders initiated by the driver at the last station. There are n stations in total. The purpose of this calculation is to ignore the effect of the size of the number of charges on the concentration calculation, which regresses the concentration of all drivers' charging stations within the sample to a range between (0, 1]. The role of the numerator is to weight the number of charges at each charging station, and the role of the denominator is to reduce the concentration range to between 0, 1. The concentration is calculated for the drivers in the sample and the descriptive statistics are obtained as shown in Tab. 7.

The data shows that the average concentration of charging stations for drivers is around 0.70, and the median is around 0.84. This indicates that most cab drivers tend to charge at one fixed station, and only a small number of drivers choose charging stations that are more dispersed. Although this can indicate that some of the stations are well built and cab drivers will visit the station several times to charge, it can also indicate that cab drivers cannot find a suitable charging station when they need to charge, because the trajectory of cabs can reach various places in the city, and a good charging convenience should be reflected in the fact that cab drivers can find a suitable charging station at any time. The concentration of charging is high, one is the problem of the function of charging operation platform, which cannot do the intelligent recommendation of charging stations; the other is the construction of some stations is not perfect enough, and there is no good guidance.

Table 7 Descriptive statistics of the concentration of electric cab charging stations

Concentration of electric cab charging stations	
Mean	0.70
Standard Error	0.01
Median	0.84
Count	4149
Confidence Level (95.0%)	0.01

4.4 Analysis of Charging Distance

Charging station site concentration expresses the dispersion of drivers in their charging choices. It reflects the percentage of charging stations used by a driver to all stations used by that driver, and is used to calculate the dispersion in charging choices. It reflects the convenience of charging facilities to a certain extent, but ignores the physical distance between stations. If a cab is charged at a station that covers a larger area of the city, it indicates that the charging facility is more convenient, more accessible, and the charging operation platform can recommend the best nearby station for the cab driver. Thus, we propose the evaluation index of charging longest distance, and the meaning of charging longest distance is the maximum distance covered by the charging behaviour of electric cabs. The Euclidean distance is calculated here.

Table 8 Descriptive statistics of the longest distance charged by electric cabs

Electric cab charging distance	
Mean	30071.79
Median	30875.96
Minimum	0
25%	14402.98
50%	30875.96
75%	44707.24
Maximum	88487.33

Firstly, the latitude and longitude of the charging station points are converted into coordinates of the right angle coordinate system, then the Euclidean distance between two coordinates is calculated, and the final calculation result is converted into a symmetric matrix. This helps us to obtain the distance between any two points later. The charging range is obtained by taking the maximum value of the distance between any two stations out of all the stations passed by the driver in his charging behaviour. By calculation, we obtain the charging range of electric cab as shown in Tab. 8.

In the valid sample data, the maximum electric cab charging distance span can be 88.5 km, the mean of charging range is around 30.1 km and the median is around 30.9 km. By calculating all the charging stations collected, the maximum distance between any two stations can also be calculated to be 88.5 km. 25% of the electric cabs have a charging distance length of around 14.4 km and 75% of the electric cabs have a charging distance length of around 44.7 km. The total length from north to south in Shanghai is about 120 km. The behaviour of electric cabs in charging distance shows different behavioural habits, some electric cabs have charging trajectories at a longer distance, and some cabs are used to charging within a fixed range.

5 CLUSTERING ANALYSIS OF CHARGING BEHAVIOUR

Above, we proposed several evaluation metrics to characterize the charging behaviour of electric cab drivers, including the propensity to charge during daytime, the average interval charging time, the concentration of charging stations, and the charging distance. In addition, there are some easily observable characteristics of the charging behaviour of electric cab drivers, including the average charging duration and the average order amount. To explore the potential differences and relationships between each feature and the others, these five evaluation indicators were aggregated, and the units of average interval length were converted to hours to filter out the drivers and indicators of normal operation, and Tab. 8 shows a preview of the evaluation indicators of charging behaviour of some drivers. The clustering analysis of charging behaviour patterns was performed using the K-Means algorithm.

5.1 Introduction to K-Means Algorithm

K-Means is an unsupervised learning distance-based clustering algorithm with a large number of applications in classifying sample features to uncover the value information implied in the data. K-Means uses a distance based evaluation criterion, where two objects are considered to be more similar when they are closer together. The basic criterion of the algorithm of K-Means is the minimum error sum of squares, and its loss function is:

$$J(c, \mu) = \sum_{i=1}^{k} \left\| \chi^{(i)} - \mu_{c^{(i)}} \right\|^2 \tag{4}$$

In this equation, $\mu_{c^{(i)}}$ represents the mean of the ith cluster. The above equation portrays the closeness of the samples around the mean within each cluster of the sample, and the less the loss function, the more similar we consider the samples within the cluster. The specific process of the K-Means algorithm is as follows:

(1) In the input dataset, k data objects are randomly identified as the initial cluster centers of mass (clustering centers).

(2) For each point in the data set, calculate the distance from the point to the cluster center of mass, and assign each

point to the corresponding cluster closest to the center of mass.
(3) Calculate the average distance between clusters as the new cluster centers.
(4) Keep repeating steps (2) and (3) until the cluster centers are not updated.

The standard algorithm flow is shown in Tab. 9.

Table 9 Description of K-means algorithm

Algorithm：K-Means	
input：Sample set $D = \{x_1, x_2, ..., x_m\}$, Number of clusters：$k$. process：	
1:	k samples are randomly selected from D as the initial mean vector $\{\mu_1, \mu_2, ..., \mu_k\}$,
2:	**repeat**
3:	Let $C_i = ø(1 \le i \le k)$
4:	**for** $j = 1, 2, ..., m$ **do**
5:	Calculate the distance d_{ji} between sample x_j and each mean vector μ_i
6:	Determine the cluster labeling of x_j based on the closest mean vector. $\lambda_j = \arg\min_{i\in\{1, 2, ..., k\}} d_{ji}$
7:	Classify the sample x_j into the corresponding cluster C_i
8:	**end for**
9:	**for** $i = 1, 2, ..., k$ **do**
10:	Calculate the average of inter-cluster distance to get the new mean vector μ'_i
11:	**If** $\mu'_i \ne \mu_i$ **then**
12:	Update the current mean vector μ_i to μ'_i
13:	**else**
14:	Keep the current mean vector constant
15:	**end if**
16:	**end for**
17:	**until** Current mean vectors are not updated
output：Classification of clusters $C = \{C_1, C_2, ..., C_k\}$	

The K-Means algorithm is simple and efficient, but this algorithm has its disadvantages: first, the initial choice of the center of mass is randomized and influenced by the initial value, and the clustering result may not be globally optimal but locally optimal; second, the number of clusters k needs to be determined in advance, but in practice it is difficult for us to estimate the value of k for the optimal clusters; third, because each iteration needs to traverse the full amount of data, when the amount of data is large, the complexity of the algorithm will lead to a very slow convergence rate.

In the problem of how to choose a reasonable k value, the elbow method is generally used, that is, the horizontal coordinate is the k value and the vertical coordinate is the loss function calculated under each k value into a line graph, and the inflection point of the line graph is the optimal value of k. The numerical indicators of the elbow method are the squared error and SSE. The idea of the elbow method to determine the k value is that as the number of clusters increases, the data will be divided more finely and the degree of aggregation for each cluster will increase, and then the SSE will decrease. When the k value increases by a number from a certain value, the SSE decreases the most at this time, and the return from the degree of aggregation will be smaller when the k value is increased again, and the SSE decreases the least, and the k value at this time is the optimal number of clusters for the sample.

To address the first and third drawbacks, the final error of the classification results can be improved by improving the selection of the initial points. k-means++ clustering algorithm is an improved algorithm based on the K-means algorithm, which is improved by making the initial clustering centers as far away from each other as possible, i.e., in each search for the center k value of the cluster, the sample with the largest distance is selected as the new clustering center. By optimizing the initial values in this way, the closer the initial clustering centers are to the final convergence state, then the number of iterations is less, which reduces the time complexity of the algorithm. At the same time, due to the larger distance between the clustering centers of the initial state, the clustering results derived from this algorithm are more accurate, and the local optimum can be avoided. The specific steps of the algorithm are as follows:
(1) Randomly select a point from the data set as the center of the initial clustering.
(2) Calculate the distance between each sample point x in the sample and the initialized clustering center, and record the shortest distance as d.
(3) Select with probability the sample with the greatest distance as the new clustering center, and repeat the above process until all k clustering centers are identified.
(4) Repeat steps 2 and 3 until k clustering centers are selected.
(5) For these k initialized clustering centers to run the standard K-means algorithm.

From the above, it can be seen that the K-means++ algorithm is superior to the K-means algorithm in terms of result accuracy and number of iterations. Therefore, this experiment uses K-means++ to explore the potential patterns and dependencies of the charging behavior characteristics of electric cabs.

5.2 Experimental Data and Clustering Results

The charging characteristics of electric cabs consist of six characteristics: daytime charging tendency, average interval duration, charging station concentration, charging distance, average charging duration, and average order amount, and Tab. 10 shows the characteristic values of evaluation indexes for some users.

Table 10 Electric cab driver charging behavior evaluation index display

Driver ID	Charging distance / m	Daytime charging tendency	Average interval duration / h	Charging sites concentration	Average charging time / h	Average order amount / yuan
12000000000000140708	31429.66	0.52	0.59	0.38	1.40	24.59
12000000000000134799	28144.47	0.47	0.43	0.65	0.78	20.28
12000000000000139592	64691.14	0.48	0.34	0.09	0.87	16.10
12000000000000047126	9703.03	0.70	0.60	0.57	0.69	28.11
12000000000000124709	29836.28	0.85	0.71	0.48	1.48	20.07

After using the K-means++ algorithm for classification, the sum of squared errors was used as the evaluation criterion to obtain the best clustering when classified into four classes, and the relationship between *k* value and SSE is shown in Fig. 4. The clustering centers of these three classes of driver charging behavior indicators are shown in Tab. 11, respectively.

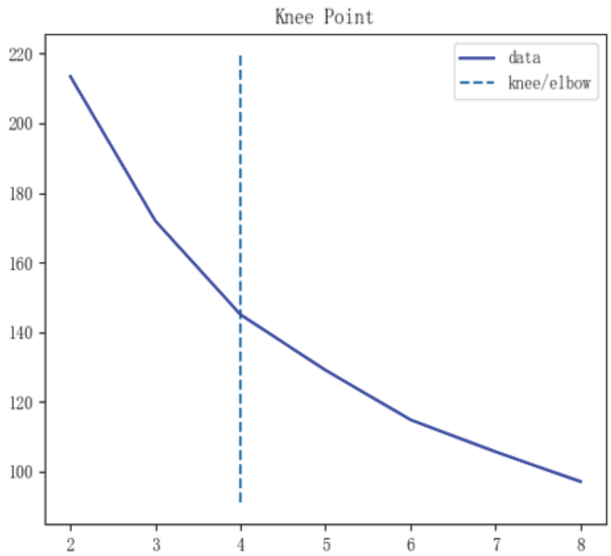

Figure 4 Sum of squared errors for different *k* values

The potential patterns of charging features can be obtained from Tab. 11. Based on the size characteristics of the clustering centers, we can summarize the four categories of charging characteristics types of cabs as follows:

(1) Long distance-daytime charging type: drivers in this category are characterized by long charging distance and low concentration of charging stations. About 78% of their charging is done during daytime, and the concentration of charging stations is only 0.29, indicating that they like to find charging stations along the road with their driving trajectory. At the same time, the average order amount of drivers of the long-distance - daytime charging type is also the highest at $ 24.51, as the cost of charging stations during the day can be higher.

(2) Short distance - daytime charging type: Drivers in this category have the characteristics of short charging distance, longest average charging interval, long average charging time, and the largest number of daytime charging. About 82% of their charging is done during the daytime, and the concentration of charging stations reaches 89%, indicating that almost all their charging behaviors are concentrated in one charging station. At the same time, their charging distance is the shortest among the four categories. It indicates that not only the charging concentration is high, they also prefer adjacent stations in the charging station selection. Their charging interval is the longest, about 21 hours, and they consume the most power due to the longer charging interval, and they also have the longest one-time charging time, about 52 minutes.

(3) Long distance - night charging type: This category of drivers has the characteristics of long charging distance, the shortest average charging interval, the most dispersed charging stations and the shortest average charging time. 60% of their charging is done at night, and the concentration of charging stations is only 0.21, indicating that they work at night and prefer to find charging stations near their destinations or starting places with their work driving trajectories. The longest charging distance of the clustering center can reach 45 km. Due to the short duration of each charging, they also have the lowest average amount of orders at only $ 21.02.

(4) Short distance-night charging type: Drivers in this category have the characteristics of short charging distance, high concentration of charging stations, and the longest average charging time. Most of their working hours are at night, and the charging time can reach about one hour each time. The concentration degree of their charging stations is 0.77 and the charging distance is 19 km, indicating that although they are more concentrated in the selection of stations, they will also consider some charging stations that may be used as targets on their driving routes, and when the charging operation platform recommends high-quality stations to them, they may consider charging in the vicinity of their driving routes.

Table 11 Clustering center for charging behavior indicators

Clustering	Charging distance / m	Daytime charging tendency	Average interval duration / h	Charging sites concentration	Average charging Time / h	Average order amount / yuan
1	34244.97	0.78	0.84	0.29	0.79	24.51
2	7344.90	0.82	0.87	0.89	0.89	24.47
3	45374.52	0.40	0.69	0.21	0.78	21.02
4	19074.25	0.31	0.73	0.77	0.94	23.75

In summary, we obtained the behavioral characteristics of electric cabs in six dimensions: charging distance, charging time, average interval duration, charging station concentration, average charging duration and average order amount. For the different preferences embodied by each classified user, the charging station operation platform should focus on the user needs reflected by the analysis of these charging behavior characteristics. Thus, this paper puts forward suggestions for improvement in the following aspects:

(1) Promote the intelligent development of the charging facility operation platform and strengthen the functions of station recommendation and station search near cab pick-up and drop-off locations to meet the charging needs of electric cabs traveling long distances to multiple destinations and improve the utilization efficiency of charging facilities.

(2) Strengthen the real-time monitoring of data on the charging operation platform, including improving the real-time and accuracy of displaying occupied and idle pile data in the charging APP, and at the same time, cooperate with the recommendation of other stations near the network to achieve attraction. For drivers who have short charging distance and tend to use fixed stations, the probability of encountering charging peaks is greater, and the scientific scheduling through the operation platform can reduce the time consumption caused by the crowded charging spaces in stations. Time consumption can be reduced by scientific scheduling of operation platform.

(3) Promote open sharing of charging facility operation platform and map platform, and realize the interconnection

of open map platform and charging operation data, so that navigation, addressing, and detection function of charging pile occupancy can be carried out in the same platform at the same time, so as to enhance the convenience of drivers in finding charging stations and improve the efficiency of using charging facilities.

5.3 Hot Spot Analysis of Charging Behavior

Hot spot analysis is a method to identify local spatial autocorrelation phenomena by evaluating and analyzing the overall data using the average value of all attributes, and hot spot analysis can identify cold and hot spots in space. In the hot spot analysis spatial aggregation model, higher values of local aggregation become hot spots and lower values of local aggregation become cold spots. To be a hot spot with significant statistical need, an element should have high values and be surrounded by other elements that also have high values. The local sum of an element and its neighboring elements will be compared to the sum of all elements; when the local sum is so different from the expected local sum that it cannot be a randomly generated result, a statistically significant z-score is generated. The formula for the hot spot analysis is:

$$G_i^* = \frac{\sum_{j=1}^{n} w_{i,j} x_j - \overline{X} \sum_{j=1}^{n} w_{i,j}}{S \sqrt{\frac{n \sum_{j=1}^{n} w_{i,j}^2 - \left(\sum_{j=1}^{n} w_{i,j} \right)^2}{n-1}}} \tag{5}$$

Inside the formula i represents the central element, j is all elements in the neighborhood, x_j represents the attribute value of the jth element in the neighborhood, $w_{i,j}$ represents the spatial distance between i and j. S is the standard deviation, and n is the total number of samples in the neighborhood.

Hotspot analysis provides a method to identify the aggregation pattern of a certain event in space. Through hotspot analysis, we can identify the hotspot areas and coldspot areas where the charging behavior of electric cabs in Shanghai occurs. Charging hotspots and cold spots have the practical significance that the charging demand in the hotspot area is high and the charging demand in the surrounding neighborhoods is also high. In contrast, cold spots not only have low charging demand themselves, but also have low charging demand in the neighboring areas around the cold spots. The charging operation platform can adjust the number and layout of charging stations accordingly based on the analyzed hot spots and cold spots.

The number of charging times reflects the magnitude of charging demand. Taking the number of charging times of charging stations in these nearly six months as the element value of hot spot analysis, the boundary of Shanghai is divided into 3 km × 3 km grids, and the number of charging behaviors in each grid is connected to the attribute value of the grid, and the hot spot analysis is carried out with each grid as the research object, and the results obtained are shown in Fig. 5.

The information in Fig. 8 shows that the hot spot areas of charging behavior are concentrated in Huangpu, Xuhui, Jing'an, Minhang, and Lujiazui areas of Pudong New District in Shanghai. There are only two charging cold spot areas with statistical significance, one is near Jiasong North Road in Jiading District and the other is near Baihe Town in Qingpu District.

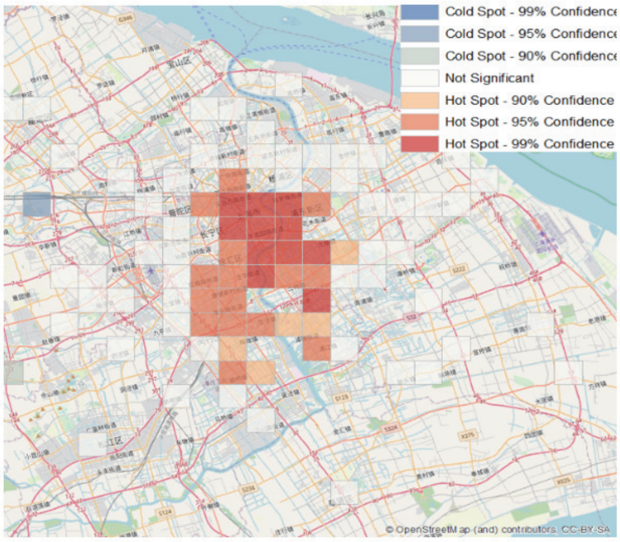

Figure 5 Hot and cold spots of electric cab charging behavior

Through the hot spot analysis, we can clearly see the concern of charging times with the distance from the city center. The closer the location to the city center, the higher the probability of charging hotspots occurring; the location far from the city center, the higher the probability of charging cold spots occurring. The hotspot areas derived in this study are only analyzed by the actual charging orders in the sample and have some degree of limitation; the results of hotspot analysis will be more accurate when the sample size is larger.

6 CONCLUSIONS

As the transition from traditional fuel vehicles to new energy vehicles accelerates, the reasonable layout and planning of charging infrastructure is becoming more and more important in order to meet the growing charging demand of new energy vehicles in cities. Based on the order data of real charging platforms, this paper proposes charging facility evaluation indexes based on charging behaviors, conducts a cluster analysis of electric cab charging behaviors, and explores the dependency relationships among these charging characteristics. Firstly, the sample data was pre-processed, and a large number of invalid orders were found at charging sites, presumably due to charging pile failures, reflecting the shortcomings of charging outlets for poor maintenance and monitoring of piles, so that the management level of charging facilities in China needs to be further improved. Secondly, by proposing four charging evaluation indexes, the characteristics of cab charging behavior in time and space are reflected. And through the K-means++ algorithm, these characteristics are clustered and analyzed to derive four types of charging patterns: long-distance-daytime, long-distance-nighttime, short-distance-daytime, and short-distance-nighttime, and relevant policy suggestions are made for these characteristics of charging user groups. Finally, the hot and cold spots where charging behavior occurs are explored, reflecting the phenomenon of high value aggregation and low value of effective utilization of charging equipment in urban areas, and operators can use the research results as a reference to remove or reduce the charging piles configured at the operating stations in the cold spot areas and increase the charging piles in the charging hot spot areas in an appropriate amount, so as to improve the utilization efficiency of charging piles.

7 REFERENCES

[1] Smart, J. & Schey, S. (2012). Battery electric vehicle driving and charging behavior observed early in the EV project. *SAE International Journal of Alternative Powertrains*, *1*(1), 27-33. https://doi.org/10.4271/2012-01-0199

[2] Liu, X., Roberts, M. C., & Sioshansi, R. (2017). Spatial effects on hybrid electric vehicle adoption. *Transportation Research Part D: Transport and Environment*, *52*, 85-97. https://doi.org/10.1016/j.trd.2017.02.014

[3] Tan, Z., Yang, Y., Wang, P., & Li, Y. (2021). Charging behavior analysis of new energy vehicles. *Sustainability*, *13*(9), 4837. https://doi.org/10.3390/su13094837

[4] Flammini, M. G., Prettico, G., Julea, A., Fulli, G., Mazza, A., & Chicco, G. (2019). Statistical characterisation of the real transaction data gathered from electric vehicle charging stations. *Electric Power Systems Research*, *166*, 136-150. https://doi.org/10.1016/j.epsr.2018.09.022

[5] Almaghrebi, A., Shom, S., Al Juheshi, F., James, K., & Alahmad, M. (2019). Analysis of user charging behavior at public charging stations. *IEEE Transportation Electrification Conference and Expo (ITEC)*, 1-6. https://doi.org/10.1109/ITEC.2019.8790534

[6] Ucer, E., Koyuncu, I., Kisacikoglu, M. C., Yavuz, M., Meintz, A., & Rames, C. (2019). Modeling and analysis of a fast charging station and evaluation of service quality for electric vehicles. *IEEE Transactions on Transportation Electrification*, *5*(1), 215-225. https://doi.org/10.1109/TTE.2019.2897088

[7] Kang, J., Kong, H., Lin, Z., & Dang, A. (2022). Mapping the dynamics of electric vehicle charging demand within Beijing's spatial structure. *Sustainable Cities and Society*, *76*, 103507. https://doi.org/10.1016/j.scs.2021.103507

[8] Wolbertus, R., Kroesen, M., van den Hoed, R., & Chorus, C. G. (2018). Policy effects on charging behaviour of electric vehicle owners and on purchase intentions of prospective owners: Natural and stated choice experiments. *Transportation Research Part D: Transport and Environment*, *62*, 283-297. https://doi.org/10.1016/j.trd.2018.03.012

[9] Li, T., Jing, P., Li, L. C., Sun, D. Z., & Yan, W. B. (2019). Revealing the varying impact of urban built environment on online car-hailing travel in spatio-temporal dimension: an exploratory analysis in Chengdu, China. *Sustainability*, *11*(5), 1336. https://doi.org/10.3390/su11051336

[10] Van Montfort, K., Kooi, M., Van der Poel, G., & van den Hoed, R. (2016). Which factors influence the success of public charging stations of electric vehicles? *6th Hybrid and Electric Vehicles Conference (HEVC 2016)*, 1-5. https://doi.org/10.1049/cp.2016.0971

[11] Kong, W., Luo, Y., Feng, G., Li, K., & Peng, H. (2019). Optimal location planning method of fast charging station for electric vehicles considering operators, drivers, vehicles, traffic flow and power grid. *Energy*, *186*, 115826. https://doi.org/10.1016/j.energy.2019.07.156

[12] Yang, J., Dong, J., & Hu, L. (2017). A data-driven optimization-based approach for siting and sizing of electric taxi charging stations. *Transportation Research Part C: Emerging Technologies*, *77*, 462-477. https://doi.org/10.1016/j.trc.2017.02.014

[13] Chen, T. D., Kockelman, K. M., & Khan, M. (2013). Locating electric vehicle charging stations: Parking-based assignment method for Seattle, Washington. *Transportation research record*, *2385*(1), 28-36. https://doi.org/10.3141/2385-04

[14] Cilio, L. & Babacan, O. (2021). Allocation optimisation of rapid charging stations in large urban areas to support fully electric taxi fleets. *Applied Energy*, *295*, 117072. https://doi.org/10.1016/j.apenergy.2021.117072

[15] Sadeghi-Barzani, P., Rajabi-Ghahnavieh, A., & Kazemi-Karegar, H. (2014). Optimal fast charging station placing and sizing. *Applied Energy*, *125*, 289-299. https://doi.org/10.1016/j.apenergy.2014.03.077

[16] Tu, W., Tu, W., Li, Q., Fang, Z., Shaw, S. L., Zhou, B., & Chang, X. (2016). Optimizing the locations of electric taxi charging stations: A spatial temporal demand coverage approach. *Transportation Research Part C: Emerging Technologies*, *65*, 172-189. https://doi.org/10.1016/j.trc.2015.10.004

[17] Han, D., Ahn, Y., Park, S., & Yeo, H. (2016). Trajectory-interception based method for electric vehicle taxi charging station problem with real taxi data. *International Journal of Sustainable Transportation*, *10*(8), 671-682. https://doi.org/10.1080/15568318.2015.1104565

[18] Babrowski, S., Heinrichs, H., Jochem, P., & Fichtner, W. (2014). Load shift potential of electric vehicles in Europe. *Journal of Power Sources*, *255*, 283-293. https://doi.org/10.1016/j.jpowsour.2014.01.019

[19] Xydas, E. S., Marmaras, C. E., Cipcigan, L. M., Hassan, A. S., & Jenkins, N. (2013). Forecasting electric vehicle charging demand using support vector machines. *48th International Universities' Power Engineering Conference (UPEC)*, 1-6. https://doi.org/10.1109/UPEC.2013.6714942

[20] Ai, S., Chakravorty, A., & Rong, C. (2018). Household EV charging demand prediction using machine and ensemble learning. *IEEE International Conference on Energy Internet (ICEI)*, 163- 168. https://doi.org/10.1109/ICEI.2018.00037

[21] Li, C., Dong, Z., Chen, G., Zhou, B., Zhang, J., & Yu, X. (2021). Data-driven planning of electric vehicle charging infrastructure: a case study of Sydney, Australia. *IEEE Transactions on Smart Grid*, *12*(4), 3289-3304. https://doi.org/10.1109/TSG.2021.3054763

[22] Shahriar, S., Al-Ali, A. R., Osman, A. H., Dhou, S., & Nijim, M. (2021). Prediction of EV charging behavior using machine learning. *IEEE Access*, *9*, 111576-111586. https://doi.org/10.1109/ACCESS.2021.3103119

[23] Chen, L., Huang, X., & Zhang, H. (2020). Modeling the charging behaviors for electric vehicles based on ternary symmetric kernel density estimation. *Energies*, *13*(7), 1551. https://doi.org/10.3390/en13071551

Contact information:

Junze YU
(Corresponding author)
School of Science, Qingdao University of Technology,
Qingdao 266520, China
E-mail: 1265535425@qq.com

Jiulin SONG
Northeastern Universit, Boston, America
E-mail: song.jiu@northeastern.edu

Jingya LIU
School of Economics and Management, Beijing Jiaotong University,
Beijing 100044, China
E-mail: 19120615@bjtu.edu.cn

Xiangke CUI
School of Economics and Management, Beijing Jiaotong University,
Beijing 100044, China
E-mail: 19113026@bjtu.edu.cn

Daqing GONG
School of Economics and Management, Beijing Jiaotong University,
Beijing 100044, China
E-mail: gongdq@bjtu.edu.cn

Qiuyan ZHANG
School of Traffic and Transportation, Beijing Jiaotong University,
Beijing 100044, China
E-mail: qyzhang1@bjtu.edu.cn

APPENDIX

Program Code

Code1: EXCEL data generation shp format file for GIS analysis

```
import os
from osgeo import ogr
import pandas as pd
from osgeo import osr

ogr.UseExceptions()
shp_path = r'C:\Users\elish\lpthw\station'

csv_filename = 'Monthcsv.csv' # Read in the csv file information and
set the field properties of the point geometry
csv_df = pd.read_csv(csv_filename)# Create a dot shp file using
a .csv file # Get the driver
driver = ogr.GetDriverByName('ESRI Shapefile')# Creating a Data
Source
shp_filename = 'Monthcsv' + '.shp' # Check if the data source already
exists

if os.path.exists(os.path.join(shp_path, shp_filename)):
      driver.DeleteDataSource(os.path.join(shp_path, shp_filename))
ds = driver.CreateDataSource(os.path.join(shp_path, shp_filename))#
Layer Name

layer_name = os.path.basename(csv_filename)[:-4] # Defining the
coordinate system object
sr = osr.SpatialReference() # Using the WGS84 geographic
coordinate system
sr.ImportFromEPSG(4326) # Create a point layer, and set the
coordinate system
out_lyr = ds.CreateLayer(layer_name, srs=sr,
geom_type=ogr.wkbPoint)

   # Create layer definition
   # Create 4 attribute fields using csv files that have four fields
   # station Field
station_fld = ogr.FieldDefn('Station', ogr.OFTString)
station_fld.SetWidth(6)
out_lyr.CreateField(station_fld)
   # Latitude Field
lat_fld = ogr.FieldDefn('Lat', ogr.OFTReal)
lat_fld.SetWidth(9)
lat_fld.SetPrecision(5)
out_lyr.CreateField(lat_fld)
   # Longitude Field
lon_fld = ogr.FieldDefn('Lon', ogr.OFTReal)
lon_fld.SetWidth(9)
lon_fld.SetPrecision(5)
out_lyr.CreateField(lon_fld)

count_fld = ogr.FieldDefn('Count', ogr.OFTReal)
count_fld.SetWidth(9)
count_fld.SetPrecision(2)
out_lyr.CreateField(count_fld)
   # Read the corresponding FEATURE type from the LAYER and
create the FEATURE
featureDefn = out_lyr.GetLayerDefn()
feature = ogr.Feature(featureDefn)
   # Set geometry
point = ogr.Geometry(ogr.wkbPoint)
   # Read in the csv file information and set the field properties of the
point geometry
for i in range(len(csv_df)):
      # Set property value section
      # Station Id
   feature.SetField('Station', str(csv_df.iloc[i, 0]))
   feature.SetField('Lon', float(csv_df.iloc[i, 1]))
   feature.SetField('Lat', float(csv_df.iloc[i, 2]))
   feature.SetField('Count', float(csv_df.iloc[i, 3]))
      # Set geometry information section
      # Create points using latitude and longitude, X is the longitude
and Y is the latitude
   point.AddPoint(float(csv_df.iloc[i, 1]), float(csv_df.iloc[i, 2]))
   feature.SetGeometry(point)
      # write feature to layer
   out_lyr.CreateFeature(feature)
```

```
    # Clear ds from memory and write data to disk
ds.Destroy()
```

Code2: Crawl poi & get downtown distance

```
import requests
import json

list=[]
with open("location.txt",'r',encoding='UTF-8') as txt_file: # Open the file
    for each_line in txt_file:
        new_line = each_line.replace('\n','')
        list.append(new_line)
    #print(list)
    txt_file.close()

def getjson():
    for i in list:
        pa = {
            'key': '5110174cd3fe50a34739d9cc4115ec23',  # Request from the console
            'location': i,
            'types': '060000',# Mall    #Location id，Recorded in the Gaode map file
            'radius': '500'
        }
        response = requests.get('https://restapi.amap.com/v3/place/around?parameters', params = pa)
        #(response.text)
        decodejson = json.loads(response.text)
        total = decodejson['count']
        print(total)

# Get Downtown Distance
def get_cbd():
    for i in list:
        #print(i)
        pa = {
            'key': '5110174cd3fe50a34739d9cc4115ec23',
            'origins': i,
            'destination': '',
            #'121.473667,31.230525',
            'type': '0'
        }
        response = requests.get('https://restapi.amap.com/v3/distance?parameters', params=pa)
        decodejson = json.loads(response.text)
        distance = decodejson['results'][0]['distance']
        print(distance)

if __name__ == '__main__':
        getjson()
        getcbd()
```

Code3:Coordinate system conversion

```
import math
import xlrd

x_pi = float(3.14159265358979324 * 3000.0 / 180.0)
# //pai
pi = float(3.1415926535897932384626)
# // Centrifugal rate
ee = float(0.00669342162296594323)
# // Long half shaft
a = float(6378245.0)
file = 'attempt_5.xls'

# // Baidu to National Bureau of Measurement
def bd09togcj02(bd_lon, bd_lat):
    x = (bd_lon - 0.0065)
    y = (bd_lat - 0.006)
    z = (math.sqrt(x * x + y * y)) - (0.00002 * math.sin(y * x_pi))
    theta = math.atan2(y, x) - 0.000003 * math.cos(x * x_pi)
    gg_lng = z * math.cos(theta)
    gg_lat = z * math.sin(theta)
    print(gg_lng, gg_lat)

def read_excel():
    wb = xlrd.open_workbook(filename=file) # Open the file
    sheet1 = wb.sheet_by_index(0)
    #i = sheet1.nrows
    #print(i)
    for i in range (1, 151):
        x = float(sheet1.row_values(i, start_colx=1, end_colx=2)[0])
        y = float(sheet1.row_values(i, start_colx=2, end_colx=3)[0])
        #bd09togcj02(x, y)
        gcj02towgs84(x, y)

if __name__ == '__main__':
    read_excel()
```

Code4:Calculate the actual number of days a cab works

```
import pandas as pd

session = pd.read_excel("Timespend.xlsx",sheet_name = 'Spend')# Charging session data
df = pd.read_excel("Timespend.xlsx",sheet_name = 'Sheet3') # Retrieve all the different driver ids
userid = df['Driver'].tolist() #Dataframe Convert to list

# Calculate the actual number of days worked
def diffdays(list):
    diffday = 1
    for i in range(1,len(list)):
        if list[i] != list[i-1]:
            diffday += 1
    print(diffday)

# Query each id separately in the session data
def findrecord():
    id = 12000000000000000476
    #print(session)
    for uid in userid:
        list1 = []
        session2 = session[session[' User ID '] == uid]
        session3 = session2.loc[:, " Start time "]
        for i in session3:
            #i = datetime.datetime.strptime(i, "%Y-%m-%d %H:%M:%S").date()
            str_i = str(i.date()) # Convert timestamp data to date format
            list1.append(str_i) # Write the date of each charge record to the list1
        diffdays(list1) # Calculate the actual number of days worked

if __name__ == '__main__':
    #diffdays(days)
    findrecord()
```

Code 6: Calculate the average charge interval

```
import pandas as pd

df = pd.read_excel("Timespend.xlsx",sheet_name = 'Spend')# Charging session data
df2 = pd.read_excel("Timespend.xlsx",sheet_name= 'Sheet6')
user_id = df2['Driver'].tolist()
id = 12000000000000113148

for id in user_id:
    session = df[df['User ID'] == id].loc[:, "Start time"]
    work_days = df2[df2['Driver'] == id].loc[:, " Number of working days "].iloc[0]
    number = df2[df2['Driver'] == id].loc[:, " Charge times "].iloc[0]
    if number == 1:
        print(0)
    else:
        T1 = session.iloc[0]
        Tn = session.iloc[len(session) - 1]
        T_days = (Tn - T1).days
        T_seconds = (Tn - T1).seconds / 86400
        T_day = T_days + T_seconds
        T_abstract = (Tn.date() - T1.date()).days + 1
        interval = (T_day - (T_abstract - work_days)) / (number - 1)
        return interval
```

Code 7:Select the optimal k value

```
from kneed import KneeLocator
from sklearn.metrics import silhouette_score
import matplotlib.pyplot as plt
df = pd.read_excel("Total.xlsx",sheet_name='Sheet4')
df2 = df.iloc[:, 1:5]
#print(df2.head(5))
# 2/ Instantiate a converter class
transfer = MinMaxScaler()
    # 3/transfer for use fit_transform

data_new = transfer.fit_transform(df2)
Scores = []
for k in range(2, 9):
    cluster = KMeans(n_clusters=k, random_state=0).fit(data_new)
    result = cluster.inertia_
    Scores.append(result)
X = range(2, 9)
kneedle = KneeLocator(X, Scores, S=1.0, curve='convex',
direction='decreasing')
plt.rcParams['font.sans-serif'] = ['SimSun']
print(The x-axis where the f' inflection point is located is:
{kneedle.elbow}')
kneedle.plot_knee()
plt.show()
```

Code 8: Charging distance analysis

```
import pandas as pd
import numpy as np
from scipy.spatial.distance import pdist, squareform

driver = pd.read_excel("Total.xlsx",
sheet_name='Sheet1',usecols=[0], converters={'ID':str}) # Retrieve
all driver user ids
join = pd.read_excel("Total.xlsx", sheet_name=' Site ', usecols=[0,1])
# retrieve site
station = pd.read_excel("Total.xlsx", sheet_name=' Site ',
usecols=[4,5], names=None) # Retrieve all site coordinates
session = pd.read_excel("Timespend.xlsx", sheet_name='Spend',
usecols=[1, 2, 14], converters={'User ID':str}) # Retrieving order
data


df = station.values.tolist() #xyCoordinate to list conversion
X = np.array(df) # List to array
Y = squareform(pdist(X, 'euclidean')) # Matrix Calculation

def user_distance():
    for i in user_id:
        large = 0  # Longest distance
        place = session[session['User ID'] == i].loc[:, ' Site Name ']  #
Query the stations each user passes through
        visit = pd.merge(place, join, how='inner', on=' Site Name ',
sort=False)  # Connection site and site id
        l1 = visit['id'].unique().tolist()
        for j in range(0, len(l1) - 1):
            n1 = l1[j]
            for k in range(1, len(l1)):
                n2 = l1[k]
                distance = Y[n1 - 1, n2 - 1]
                if distance > large:
                    large = distance
        print(large)

def largest_distance():
    large = 0
    for j in range(0, 149):
        for k in range(1, 150):
            distance = Y[j, k]
            if distance > large:
                large = distance
    print(large)

if __name__ == '__main__':
    #user_distance()
    largest_distance()
```