electronics

MDPI

Article

Charging Behavior Analysis Based on Operation Data of Private BEV Customers in Beijing

Hao Tian [1], Yujuan Sun [1,*], Fangfang Hu [2] and Jiuyu Du [3,*]

[1] Faculty of Materials and Manufacturing, Beijing University of Technology, Beijing 100124, China
[2] Beijing Products Quality Supervision and Inspection Research Institute/National Automotive Quality Inspection and Testing Center, Beijing 101300, China
[3] State Key Laboratory of Automotive Safety and Energy, Tsinghua University, Beijing 100084, China
* Correspondence: yjsun@bjut.edu.cn (Y.S.); dujiuyu@tsinghua.edu.cn (J.D.)

Abstract: Charging behavior is essential to understanding the real performance and evaluating the sustainability of battery electric vehicle (BEV) development and providing the basis for optimal infrastructure deployment. However, it is very hard to obtain the rules, due to lack of the data support, etc. In this research, analyzing the charging behavior of users with private charging piles (PCPs) is carried out based on the real vehicle data of 168 BEV users in Beijing, covering 8825 charging events for a one-year duration. In this study, the charging behaviors are defined by five indexes: the starting state of charge (SOC) of batteries, charging location selection, charging start time, driving distance, and duration between two charging events. To further find the influencing rules of the PCPs owning state, we setup a method to divide the data into two categories to process further analysis and comparison. Meanwhile, in order to better observe the impact of electric vehicle charging on the power grid, we use a Monte Carlo (MC) simulation to predict the charging load of different users based on the analysis. In addition, an agent-based trip chain model (ABTCM), a multinomial logistic regression (MLR), and a machine learning algorithm (MLA) approach are proposed to analyze the charging behavior. The results show that with 40% or lower charging start SOC, the proportion of users without PCPs (weekday: 55.9%; weekend: 59.9%) is larger than users with PCPs (weekday: 45.5%; weekend: 42.6%). Meanwhile, users without PCPs have a certain decrease in the range of 60–80% charging start SOC. The median charging time duration is 51.44 h for users with PCPs and is 17.25 h for users without PCPs. The charging peak effect is evident, and the two types of users have different power consumption distributions. Due to the existence of PCPs, users have lower mileage anxiety and more diverse charging time choices. The analysis results and method can provide a basis for optimal deployment and allocation of charging infrastructure, and to make suitable incentive policies for changing the charging behavior, targeting the carbon neutral objectives.

Keywords: private charging pile (PCP); battery electric vehicle (BEV); charging behavior; agent-based trip chain model (ABTCM); multinomial logistic regression (MLR); machine learning algorithm (MLA)

check for updates

Citation: Tian, H.; Sun, Y.; Hu, F.; Du, J. Charging Behavior Analysis Based on Operation Data of Private BEV Customers in Beijing. *Electronics* **2023**, *12*, 373. https://doi.org/10.3390/electronics12020373

Academic Editor: Fabio Corti

Received: 30 November 2022
Revised: 2 January 2023
Accepted: 6 January 2023
Published: 11 January 2023

1. Introduction

The increasing fossil fuel consumption and the resulting climate change have become an urgent global problem. Electric vehicles (EVs) are regarded as one of the most promising technologies in the transportation industry for their significant advantages in dealing with climate change and oil dependence [1]. As the awareness of environmental issues increases, the Chinese government regards the development of EVs as a strategic goal, and has introduced various subsidy policies [2]. By the end of 2021, the number of new energy vehicles in China reached 7.84 million, accounting for 2.60% of the total number of vehicles. Among them, battery electric vehicles (BEVs) were 6.4 million, accounting for 81.63% of the total new energy vehicles [3]. In the long run, BEV will play an important role in the future automobile market [4].

With the growth of BEV sales, the supporting infrastructure of charging pile construction is bound to usher in strong growth. According to the analysis report of the China EV Association (CEVA) [5], from January to September 2022, the national charging pile increments were 1.871 million units, of which public charging piles increased by 106.3% year-on-year, and private charging piles (PCPs) built with vehicles increased by 352.6% year-on-year. As of September 2022, the cumulative number of charging piles nationwide was 4.488 million units, with a year-on-year increase of 101.9%. At the same time, the use of PCPs may become an important factor affecting the willingness of people to purchase BEVs in the next few years [6]. Thus, PCP has amazing development potential.

However, with the substantial increase in PCPs, many problems have emerged. For example, the PCPs are mostly used at night and are idle during the day, resulting in low utilization of charging piles [7]. Although the number of charging piles continues to rise, it is still far from the ideal state (Car–pile ratio reaches 1:1). Increasing battery capacity [8,9] and establishing wider charging opportunities [10] are effective ways to reduce the peak demand for BEV charging [11]. In response to the prominent problem of difficult installation of residential charging piles, the government encourages the electrification of charging facilities into old residential renovation projects, to effectively alleviate the charging anxiety [12].

The charging pile is the carrier of charging behavior. Scientific analysis of the charging behavior of BEV users is the basis for improving the layout of charging piles. The current research on charging behavior is mainly based on public charging station data and vehicle travel data. For example, Jimenez et al. [13] used ML methods to characterize driver charging behavior at charging stations to predict and distinguish the energy consumption of EVs in different seasons. Xing et al. [14] established a new EV charging behavior model to characterize user selection and decision-making by modeling and mining the 'Didi' traffic travel data set. However, there are few studies on the differences in charging behavior between users with PCPs and users without PCPs. The possible charging modes of BEV users are very important for the construction of charging infrastructure and the promotion of BEVs.

As a result, as illustrated in Figure 1, based on actual charging and driving data from 168 BEVs in Beijing, this paper will distinguish users into high probability users with PCPs and high probability users without PCPs, deeply analyze the similarities and differences between the two types of users in charging and driving behavior, and clarify the impact of PCPs on the charging behavior of BEV users. The organization of this paper is as follows. Section 2 summarizes and processes the data sets used in this study to obtain charging events of all vehicles within 1 year. Section 3 classifies users according to the point of interest (POI) of the user's charging location, which is divided into high probability users with PCPs and high probability users without PCPs. Section 4 compares and analyzes two types of users from six aspects: charging location selection, charging start time, charging start state of charge (SOC), time/driving distance since last charging event, and charging energy consumption. Based on the analysis results, the charging load prediction of different users is realized, and the influence of the charging load on the power grid is evaluated. Section 5 proposes an agent-based trip chain model (ABTCM), a multinomial logistic regression (MLR), and a machine learning algorithm (MLA) approach to analyze the charging behavior of users with PCPs. At last, we conclude the paper and provide future prospects in Section 6.

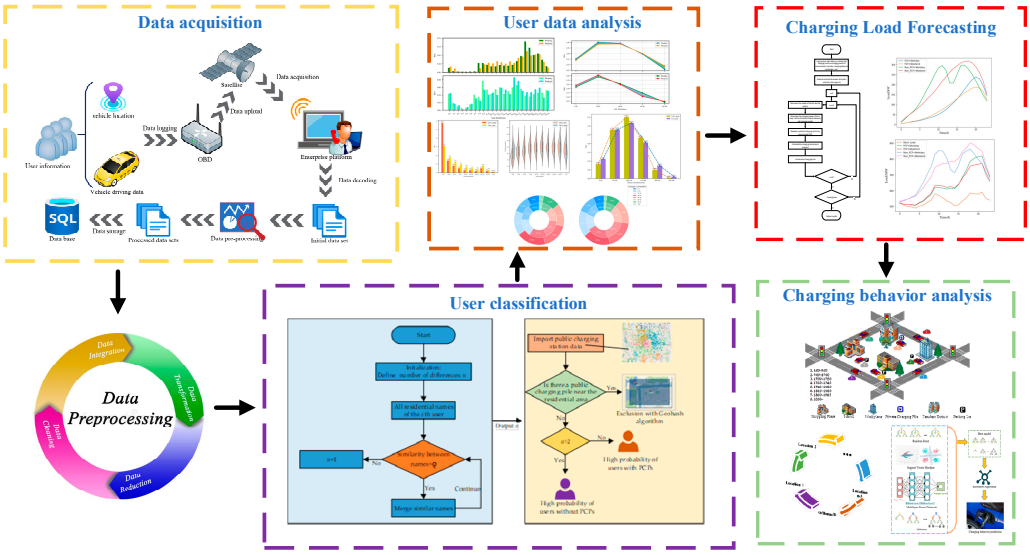


Figure 1. Analysis process of BEV charging mode for users with PCPs and users without PCPs in Beijing based on real data.

2. Data Collection and Processing

2.1. Data Sources

This paper uses the operation data of 168 BEVs in Beijing from January to November 2020, including vehicle data, drive motor data, and vehicle location data. When the vehicle starts, the enterprise platform monitors and collects the real-time operating state parameters of the BEV from the CAN-bus through the on-board diagnostic system (OBD). According to the transmission control protocol (TCP), the data is transmitted to the data center in the form of a data stream to generate the original data set, and the data set is saved through the database. The specific process is shown in Figure 2.

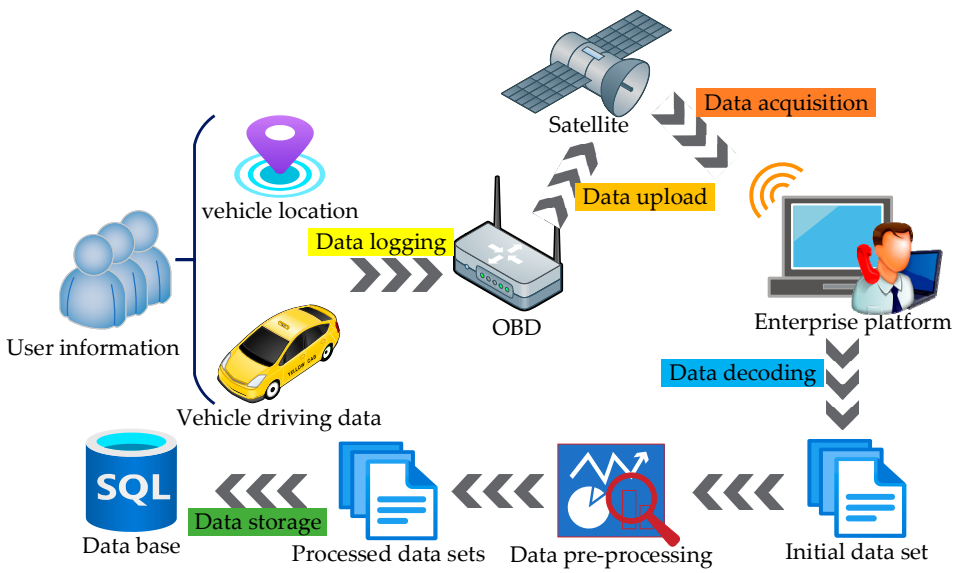


Figure 2. Data acquisition process.

The data sampling interval is 10 s. The data format is shown in Table 1, including vehicle status, charging status, vehicle speed, cumulative mileage, longitude coordinate, latitude coordinate and SOC value.

Table 1. The format of data collected.

Time	Vehicle ID	Vehicle State	Charging State	Speed (km/h)	Mileage (km)	Longitude	Latitude	SOC (%)
1 January 2020 00:00:06	98,341	2	1	0	34,279	116.3608	40.1142	68
1 January 2020 00:00:16	98,341	2	1	0	34,279	116.3608	40.1142	68
			...					
4 August 2020 07:12:47	98,341	1	3	21.8	41,911	116.2863	40.0129	43
4 August 2020 07:12:57	98,341	1	2	18.3	41,911	116.2857	40.0128	43
			...					
4 November 2020 02:17:28	98,341	2	1	0	45,970	116.3608	40.1142	100

Three types of BEVs are studied in this paper, and the main parameters are shown in Table 2, including battery capacity, battery type, fast charge power, slow charge power, driving range and maximum speed. In this paper, vehicle battery capacity is chosen as a measure of vehicle heterogeneity standards.

Table 2. Battery parameters of the vehicles involved in the data.

Brand	GAC Thriving 14	BAIC EU260	SAIC Roewe ERX5
			
Capacity (kWh)	47.5	37.8	48.3
Battery Type	Lithium iron phosphate battery	Nickel–Cobalt–Manganese	Nickel–Cobalt–Manganese
Fast Charge Power (kW)	49.15	50.87	63.32
Slow Charge Power(kW)	3.05	6.13	6.42
Driving Range (km)	253	260	320
Maximum Speed (km/h)	150	140	135

2.2. Data Preprocessing

Due to the influence of sensor equipment failure, signal interference, and other factors, the data have the problems of wrong vehicle state, accumulated mileage, SOC, and latitude and longitude coordinates, so it is necessary to carry out data preprocessing, such as data cleaning and data filtering. The data cleaning is mainly about extracting abnormal data such as: (i) SOC is 0; (ii) the speed is negative; (iii) cumulative mileage is 0; (iv) abnormal vehicle operating status (charging status and vehicle status), filling the null value, zero value and missing value, and correcting the error value.

2.3. Extraction of Charging Segments

The running state of the vehicle can be determined by the vehicle state and charging state in Table 1 of Section 2.1. For the vehicle state, 1 indicates the start state, 2 indicates the power off state, and 3 indicates other states. For the charging state, 1 indicates parking charging, 2 indicates brake charging, 3 indicates uncharged state, and 4 indicates charging completed. The vehicle state can be used to determine whether the vehicle is in a parking state, and the charging state can be used to determine whether the vehicle is in a charging

state. Table 3 shows the different states of the vehicle by combining the vehicle state, charging state, and vehicle speed.

Table 3. List of possible states of the vehicle.

Vehicle State		Charging State		Speed		State
1	+	3	+	0	→	Status 1. Parking but power on
1	+	3	+	$\neq 0$	→	Status 2. Driving
2	+	3	+	0	→	Status 3. Parking and power off
2	+	1	+	0	→	Status 4. Parking and charging
2	+	4	+	0	→	Status 5. Charging completed but not unplugged

- Extracting fragment for the first time

In this paper, the charging start (end) time of the user is judged by the change of the charging state at two adjacent moments. For instance, if the charging state is 3 and the vehicle state is 1 at time t_1; while at next time t_2, the charging state becomes 1, the vehicle state becomes 2, and the vehicle speed is 0, then it can be considered that t_2 is the charging start time, and the corresponding SOC is the charging start SOC, that is, SOC_{start}. Similarly, if the charging state is 1 and the vehicle state is 2 at time t_3, while at next moment t_4, the charging state becomes 3 (if the charging is completed, the charging state becomes 4) and the vehicle state becomes 1 (or still 2, that is the charging is completed but the user does not start the car immediately), then it can be considered that t_3 is the end of the charging time, and the corresponding SOC is the end SOC of the charging, that is SOC_{end}.

We combine the records between t_2 and t_3 into a charging fragment and obtain 12,357 charging fragments.

- Extracting fragment for the second time

EV charging uses AC or DC power supply. At the same voltage level, DC charging is faster. The current DC fast charging (DCFC) technology can fully charge EVs in 20 min [15], while the AC charging may require 10–20 h charging time [16]. Based on the above research, we set the reasonable time duration of the charging segment between 10 min and 20 h. In addition, "mileage interval between adjacent charging event (Section 4.2.3) exceeds the driving range of the vehicle (Table 2)", "data acquisition error [17] caused by the duration between adjacent charging events is too small" and "variation of SOC ($SOC_{end} - SOC_{start}$) is less than 0" are also listed as outliers. After filtering, 8825 charging segments are obtained.

3. User Classification of BEVs

In order to study the impact of PCPs on the charging behavior of BEV users, the following four steps are needed to determine whether the 168 BEV users have PCPs. (i) Convert the geographic coordinates corresponding to the user's charging position to satisfy the coordinate system of Amap; (ii) Using the inverse geocoding function in the Amap open platform to convert coordinates into address information; (iii) Taking the POI type corresponding to the address as the first filtering standard; (iv) Excluding the possibility of charging at public charging station in a residential area, according to the public charging station data on the website.

3.1. Coordinate System Transformation

This paper mainly involves two common map coordinate systems: the world geodetic system (WGS-84) and the national survey bureau coordinate (GCJ-02). The positioning information obtained through the underlying interface usually belongs to the WGS-84 coordinate system, so the coordinate information in the data source uploaded by the BEV belongs to the WGS-84 coordinate system. For privacy protection, the GCJ-02 coordinate system uses the confusion algorithm by adding random offsets in latitude and longitude of

WGS-84 system. Most domestic internet map providers are using the GCJ-02 coordinate system, including Amap, Google Maps China, and so on. Since the inverse geocoding in Section 3.2 is based on the GCJ-02 coordinate system, the original coordinate system WGS-84 needs to be transformed into the GCJ-02 coordinate system. In this paper, a coordinate transformation tool Change Coordinate [18] is used to convert the vehicle positioning coordinates (WGS-84) into Amap's applicable coordinates (GCJ-02), as shown in Figure 3. The red dot is the actual charging position of the vehicle, and the purple dot is the converted vehicle position. It can be seen that the converted coordinates have a significant offset. If we put the coordinates of the purple point into the Amap, the visualized point is coinciding with the position of the red point on the Figure 3, which is located in the residential area.

Figure 3. Coordinate visualization.

3.2. Inverse Geocoding

Through the function of inverse geocoding in the Amap open platform [19], the transformed latitude and longitude coordinates of the BEVs at charging time are taken as input, and the corresponding information is returned as shown in Table 4. The Vehicle ID is the BEV number; Name represents the address name of the building closest to the charging location; District represents the area of the building belonging to; Type represents the POI type of the building; Minimum range represents the distance from the nearest building to the charging location (units m). The charging location here refers to the parking position of the electric vehicle in a charging state. Meanwhile, the reason for setting the item 'Minimum range' is that, in general, we can directly obtain the POI of the charging position through the address corresponding to the BEV coordinates, which is helpful for us to determine the POI type of charging position (such as residential area). However, some address information does not have a corresponding POI; in order to minimize data loss, we consider using the POI of nearby buildings to supplement based on the premise that the address name is similar. Position represents the longitude and latitude coordinates corresponding to the building. The accuracy of the method is verified by coordinate visualization as shown in the blue flag in Figure 4; the charging position is indeed within the range of the returned address.

Table 4. The format of data returned by inverse geocoding.

Vehicle ID	Name	District	Minimum Range (m)	Type	Position
98,363	Beijing University of Chinese Medicine	Chaoyang District	145.2680	School	116.428250, 39.971307

Figure 4. Verification of coordinate accuracy.

3.3. User Classification

Generally speaking, the charging locations of the BEV users with PCPs are relatively fixed in residential areas. Thus, whether a user has a PCP is determined by the POI type of the charging location in Section 3.2. If most of one user's charging locations are in a relatively fixed residential area, then it is considered that the user has a high probability of owning a PCP. When all of the user's charging positions are not in residential areas, i.e., the ratio of the user's residential charging r is 0, it is considered that the user has no PCP. By preliminary screening, we found a total of 24 BEV users without residential charging records, so we classify them as users without PCPs. We will further distinguish the remaining users in two steps, as follows.

3.3.1. First Screening

Figure 5 is the flow chart of the PCP user classification method. The blue block shows the method of obtaining how many different residential areas that the charging point is located for one user by traversing all the residential area names. If the similarity between two residential names exceeds a certain threshold ρ, the difference between the two names can be considered small. The similarity here is compared using Python's module-difflib. It is found that when the similarity exceeds 0.7, the two residential names are highly coincident, so let $\rho = 0.7$. When the traversal is completed, the number of different charging residential areas for each user is obtained.

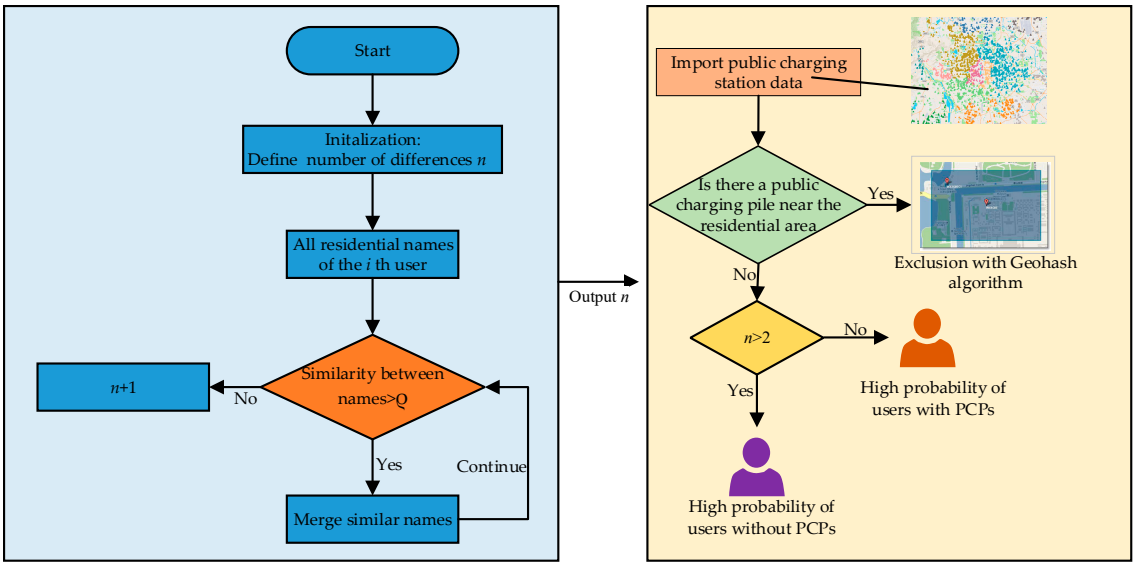


Figure 5. Flow chart of user classification method.

The yellow block shows how to classify users according to the number of different residential areas where the charging is located. Firstly, we exclude public charging piles near residential areas based on online public charging station data. Method details will be discussed in Section 3.3.2. At the same time, we believe that the residential location of users with PCPs general is relatively fixed, and not likely to appear in many different residential charging phenomena. Based on this principle, we set the judgment threshold to 2, that is, the number of different residential areas of the user cannot exceed 2 (certainly greater than 0). We believe that the user is a high probability PCPs user, and otherwise it is a high probability of a non-PCP user.

3.3.2. Second Screening

As shown in Figure 6, some public charging stations may locate in residential areas.

Figure 6. The public charging stations near residential areas.

Figure 7 shows the existing public charging station data on the website [20]. According to the above data and combining with Geohash algorithm, we determine whether there is a public charging station near the user's charging location, and then perform further screening.

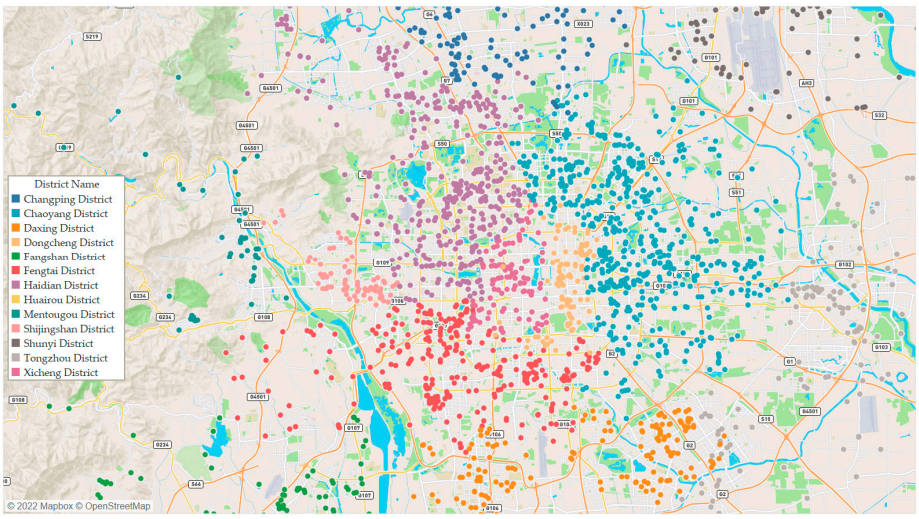


Figure 7. Distribution of public charging piles in Beijing.

The distance between one user's charging location and each public charging station can be calculated directly and then sorted to find the nearby public charging station. However, for the increasing number of charging stations, this method is obviously unrealistic with the

large amount of calculation. Geohash can solve this problem well. Geohash is a geospatial algorithm that can retrieve adjacent areas, so that the coordinate points in a certain area share a string, so we only need to filter the public charging stations in the user's charging area according to the string. Since there may be multiple charging stations in the area, the nearest public charging station can be judged according to Equation (1).

$$d = R \times \text{arcos}[\cos(y_1) \times \cos(y_2) \times \cos(x_1 - x_2) + \sin(y_1) \times \sin(y_2)] \tag{1}$$

In the formula, (x_1, y_1) represents the longitude and latitude coordinates of the user's charging location, and (x_2, y_2) represents the longitude and latitude coordinates of the public charging station; d is the distance between the two points; R is the radius of the earth, let R = 6371 km.

This article uses Python to implement Geohash encoding. The realization process is mainly divided into four steps: (i) The range represented by longitude and latitude is regarded as a two-dimensional plane rectangle; (ii) As shown in the Figure 8, the longitude and latitude coordinates are classified by the similar dichotomy. If the target longitude and latitude are in the division area, the assignment is 1, otherwise the assignment is 0, until the set accuracy requirements are met, and a binary code is obtained; (iii) Figure 9 shows the process of grouping codes based on the coding results. Each binary code is merged by the principle of even-digit longitude and odd-digit latitude; (iv) In order to facilitate the storage and usage, 32 letters in the Table 5 are used for Base32 encoding.

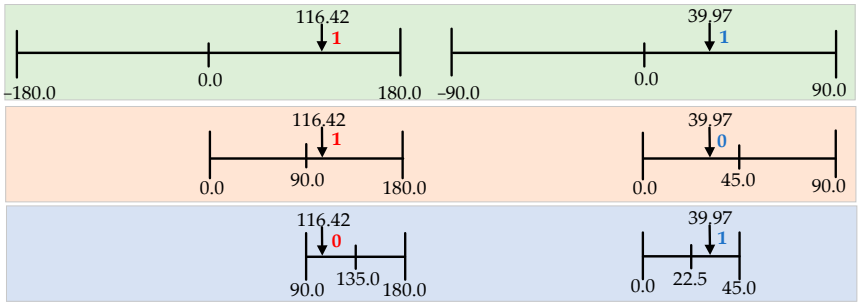


Figure 8. Binary coding of geographic latitude and longitude coordinates.

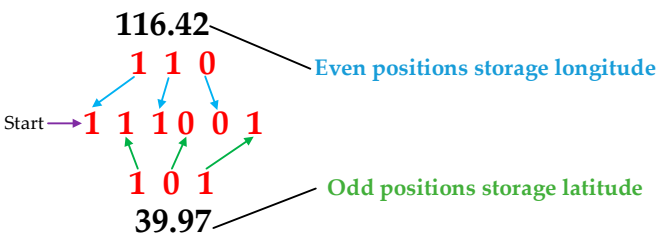


Figure 9. The process of Binary group coding.

Table 5. Base32 encoding/decoding.

Decimal	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
Base32	0	1	2	3	4	5	6	7	8	9	b	c	d	e	f	g	h	j	k	m	n	p	q	r	s	t	u	v	w	x	y	z

The string length corresponds to the accuracy of the position. The longer the string is, the smaller the divided area is, and the higher accuracy of the position is. Take the coordinates (116.42, 39.97) in Figure 8 as an example. If we require the length of the string to be 4, according to the above process, first, we can obtain the binary codes of longitude and latitude: 1101001011, 1011100011. Then we group the code according to the rules in Figure 9, and the result is 11100 11101 00100 01111. Finally, Base32 coding is performed according to Table 5, and the result is w x4 g.

After analysis and filtering, there are 135 users with PCPs, accounting for about 80.4%, and 33 users without PCPs. The classification results of all charging positions are marked on the map, as shown in Figure 10. In addition, for the case where two charging points are close, the distance is also marked in the figure.

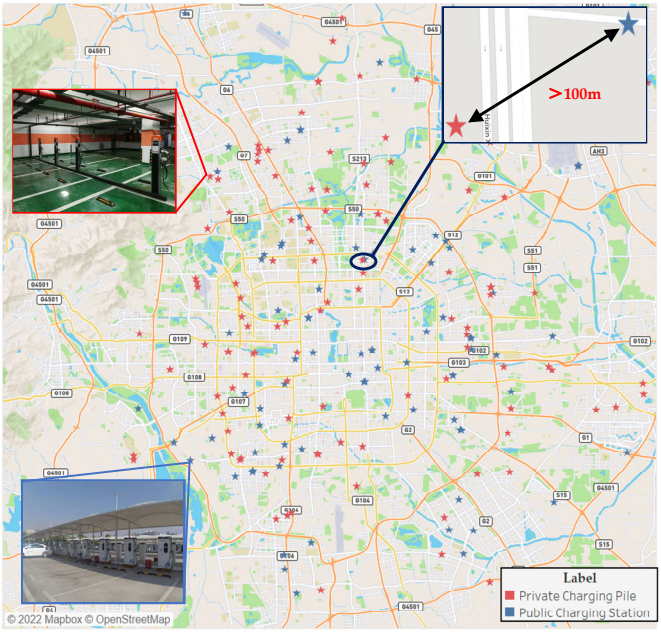


Figure 10. Distribution of users using private and public charging piles.

4. User Charging Behavior Analysis

4.1. Charging Location Selection

By mining the POI types of users' common charging points, it can provide assistance in the planning and layout of future charging piles construction [21]. Pagany et al. [22] distinguished categories based on user groups of different POI types, which constitutes the basic method for requirements calculation. Figures 11 and 12 show the POI distribution of charging points of users without PCPs and users with PCPs, respectively. The larger the font is, the larger the proportion of the POI type is. For POI of users without PCPs, a larger proportion is office buildings, industrial parks, public parking, and shopping related places. In general, office buildings or industrial parks can be regarded as the workplace, thus it can be concluded that the users without PCPs often charge in the company; followed by public places (shopping malls, parks, etc.) and public charging stations.

Figure 11. Charging location classification of users without PCPs.

Figure 12. Charging location classification of users with PCPs.

In addition to charging in fixed residential areas, users with PCPs also charge in public places, like users without PCPs. It can be found that users are not completely dependent on PCPs. When the SOC level is low or the user's willingness to charge is strong, the user with PCPs will also choose to charge in public places.

4.2. Charging Behavior

The current research on charging behavior mainly focuses on the analysis of the distribution of daily driving distance, charging start time, driving distance since last charge, and SOC before and after charging [23]. This paper mainly analyzes the difference in charging behavior between users with PCPs and users without PCPs from four aspects: charging start time [24], charging start SOC, driving distance since last charge, and time since last charge, and summarizes the possible reasons for the difference in charging behavior.

4.2.1. Charging Start Time

Figure 13a,b shows the charging start time distribution of users with PCPs and users without PCPs. On weekdays, the peak of the charging start time of users with PCPs appears at 18:00–19:00. The possible reason is that people return home from work, then charge in PCPs. However, the peak time of users without PCPs is between 8:00 and 9:00. During this period, consumers may charge in public charging piles near workplaces. Correspondingly, the proportion of users with PCPs choosing to charge in this period is small. The existence of PCPs greatly reduces the users' dependence on public charging piles.

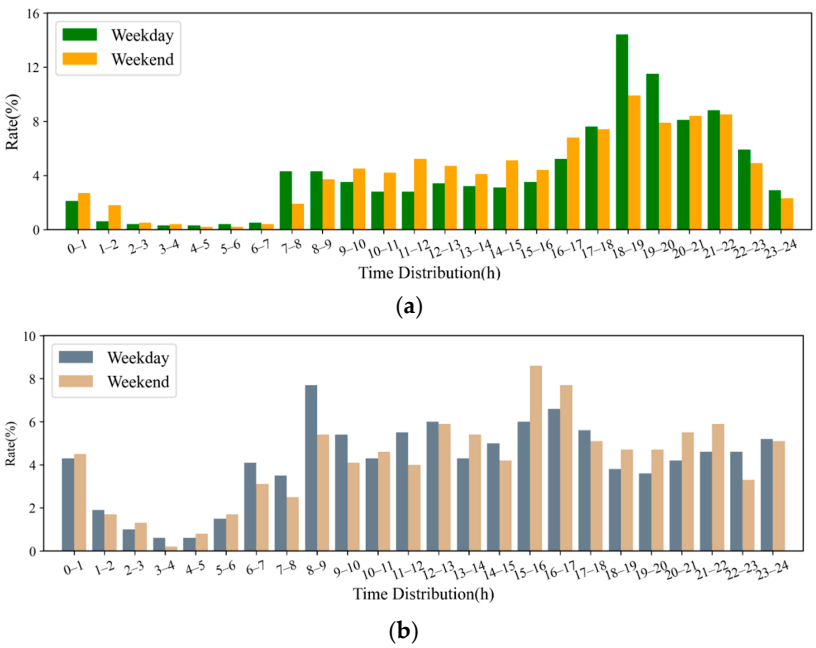


Figure 13. Comparison of user charging start time. (**a**) Users with PCPs; (**b**) Users without PCPs.

Whether on weekdays or weekends, the variation trend of charging start time of users with PCPs is roughly the same, explaining that the existence of PCPs ensures the convenience of charging and leads people to form a certain charging habit. The peak charging time of users without PCPs on weekends occurs between 15:00 and 16:00, and there could be many reasons for the difference between the peak charging time on weekdays and weekends, such as the availability of charging piles nearby or the SOC level at this time.

4.2.2. Charging Start SOC

Figure 14 shows the distribution of charging start SOC. Among charging events with 40% or lower start SOC, the proportion of users without PCPs (weekday: 55.9%; weekend: 59.9%) is larger than users with PCPs (weekday: 45.5%; weekend: 42.6%). Unreasonable charging station network layout and weak information interconnection between BEVs and charging piles could cause some problems, such as long charging waiting time and unbalanced utilization of charging stations [25]. Therefore, it can be inferred that the lack of available charging piles nearby is the main reason for the low charging start SOC level. Studies have also shown that the lower the remaining SOC of an BEV, the higher the driver's mileage anxiety [26,27], so that users without PCPs have higher mileage anxiety. Battery life is shortened in the case of over discharged batteries [28]. For users with PCPs, in order to extend the battery cycle life, it is recommended that people carry out shallow charging and shallow discharging, and to start charging at a relatively high SOC level.

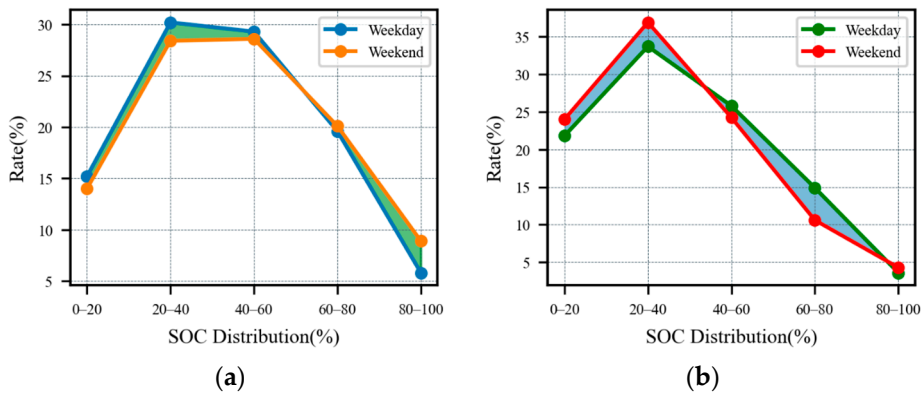


Figure 14. Comparison of charging start SOC for different users. (**a**) Users with PCPs; (**b**) Users without PCPs.

On the weekends, the start charging SOC of users with PCPs has a certain increase in the range of 80–100% SOC, mainly because they consider charging before the weekdays and pay no attention to the current SOC level. Meanwhile, users without PCPs have a certain decrease in the range of 60–80% start charging SOC. They do not choose to charge because of no available charging piles nearby, and the relatively high SOC level that the psychological pressure of users from cruising range anxiety is small.

4.2.3. Driving Distance since Last Charge

In order to ignore contingency, an interquartile range (IQR) analysis [29] was used to quantify the characteristics of the driving mileage distribution between two adjacent charging events, as shown in Table 6, where Q1, median, and Q3 represent the 25th, 50th and 75th percentiles of the mileage, respectively. In Section 4.2.4, IQR analysis will also be performed on the interval time since last charge and charging energy consumption.

Table 6. IQR analysis of different users regarding driving mileage since last charge event.

		Minimum	Q1	Median	Q3	Maximum
Have PCP?	Yes	1.0	79.00	117.00	159.00	288.00
	No	1.0	83.00	120.00	154.25	296.00

After data processing, the average driving distance between two adjacent charging events for users without PCPs is 120.25 km, with a median of 120.00 km and a maximum driving distance of 296.00 km. The average driving distance between two adjacent charging events for users with PCPs is 119.15 km, with a median of 117.00 km and a maximum of 288.00 km. It is further confirmed that, in most cases, the two types of users are not inclined to discharge the battery deeply, because the average driving distance after the last charging is less than half of their corresponding maximum driving range.

Figure 15 shows the driving distance distribution since last charge. The number of users who choose to charge in 100–150 km is the largest, accounting for 35.03% (users without PCPs) and 31.85% (users with PCPs), respectively. In addition, only 4.78% (users without PCPs) and 8.66% (users with PCPs) of charging events occurred after 200 km. Users with a mileage of 50–200 km account for 86.69% (users without PCPs) and 80.40% (users with PCPs). In view of the maximum driving range of the sample vehicle (Table 2), it can be found that people seem unwilling to make the BEV's mileage exceed this buffer range (50–200 km) before charging, which is consistent with the conclusion of Weldon et al. [30].

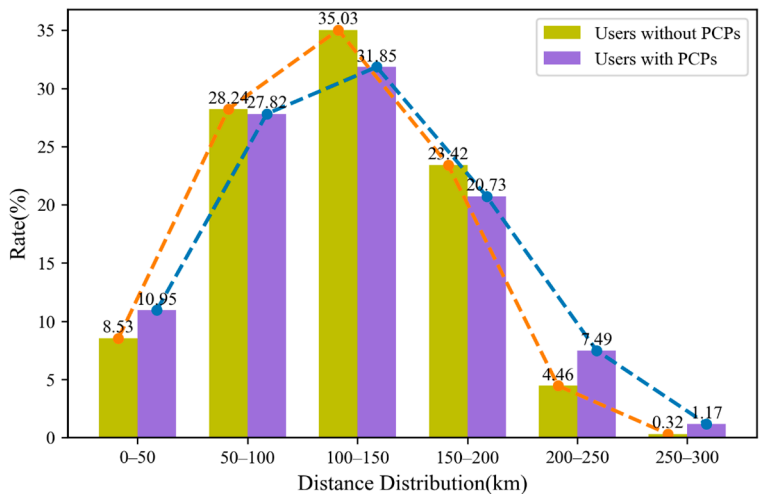


Figure 15. Driving mileage distribution between two adjacent charging events.

Distance parameters since the last charge can be used to analyze the charging habits of users relative to their driving range. At a relatively small mileage interval (0–50 km), the proportion of users with PCPs (10.95%) is larger than that of users without PCPs (8.53%). Due to the existence of PCPs, consumers have the habit of charging regularly, that is, charging at home after work every time. Therefore, it is possible to predict the charging behavior of users with PCPs [31], so as to provide a reference for optimizing the layout of charging infrastructure, PCP sharing, and realizing the intelligent dispatching of the power grid to some extent.

4.2.4. Time since Last Charge

The time parameters since last charge can quantify the frequency of charging requirements. Figure 16 shows the duration distribution between two adjacent charging events, and Table 7 is a quantitative result of the duration. The average duration of users with PCPs is 88.69 h (3–4 days), and the median is 51.44 h (2–3 days). The average duration for users without PCPs is 55.38 h (2–3 days), and the median is 17.25 h (0–1 day).

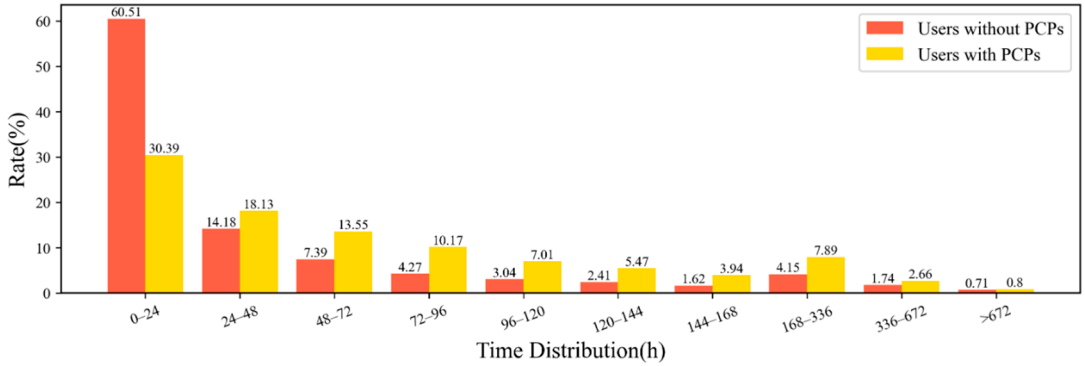


Figure 16. Duration distribution of two adjacent charging events.

Table 7. IQR analysis of different users regarding time since last charge event.

		Minimum	Q1	Median	Q3	Maximum
Have PCP?	Yes	0.27	20.74	51.44	109.66	2592.27
	No	0.26	7.58	17.25	48.61	3145.59

Due to the existence of extreme values, it is difficult to describe the overall level using the average value, so the median value is used. Obviously, the overall charging intervals of users without PCPs are shorter than that of users with PCPs, indicating that PCPs greatly decrease the charging frequency of users, and also shows that most users choose to charge 1–3 times a week [32].

In the range of 0–24 h duration, users without PCPs charged more frequently (60.51%) than users with PCPs (30.39%); due to the existence of mileage anxiety, users without PCPs are much more concerned about insufficient battery power during driving [15], which is consistent with the analysis in Section 4.2.2. For other durations, the proportion for users without PCPs is lower than users with PCPs. Therefore, compared with users without PCPs, due to the existence of PCPs, users with PCPs have lower mileage anxiety, more diversity of charging time selection, and wider charging time spans.

Meanwhile, combined with the results in Figure 17, it can be found that users without PCPs will have a long charging interval, but a large driving mileage. As shown in Figure 16, in the duration of 0–24 h since last charge, the driving mileage of users with PCPs is lower than that of users without PCPs. The possible reason is that the user buys a BEV car and installs a PCP just to meet the daily commuting requirements and charging convenience.

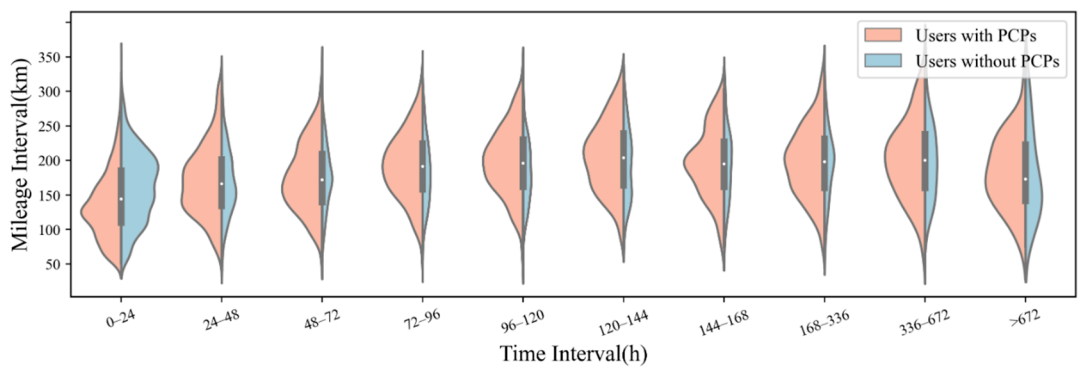


Figure 17. Relationship of duration and mileage between two adjacent charging events.

4.3. Charging Energy Consumption

The energy consumption E (kWh) during the charging process are calculated according to Equation (2) referring to the study of Siddique et al. [33]:

$$E = C \times \Delta\text{SOC} = C \times (\text{SOC}_{t_2} - \text{SOC}_{t_1}) \tag{2}$$

where C represents the battery capacity of BEVs, the range of SOC is between 0 and 1, ΔSOC represents the change of SOC in the i-th charging process, the start SOC and end SOC of the i-th charging data are recorded as SOC_{t_2} and SOC_{t_1}, respectively, t_1 represents the start time of the i-th charging, and t_2 represents the end time of the i-th charging.

Table 8 lists the IQR analysis results of charge energy consumption. It can be found that in most cases, the charging energy consumption of the two types of users is much lower than the rated capacity of the battery, and that the ratio of the median to the rated capacity is less than 0.6. One of the reasons might be that the battery is not fully charged (less than 100% SOC) at the end of the charging process. Another reasonable reason is that even at a high level of SOC, most drivers will charge the vehicles. It can also be verified from Figure 14 that users start to charge when the SOC is higher than 60%, accounting for about 20% of the total charging events.

Table 8. IQR analysis of different users about charging energy consumption.

			Minimum	Q1	Median	Q3	Maximum
Have PCP?	Yes	Total	0.47	14.08	22.33	31.40	47.82
		Weekday	0.47	14.73	22.80	31.86	47.82
		Weekend	0.48	11.59	20.90	29.95	47.03
	No	Total	0.41	14.49	22.36	30.64	46.08
		Weekday	0.41	14.25	22.33	30.43	46.08
		Weekend	0.83	16.56	23.44	31.45	45.60

In addition, Figure 18 shows the distribution of charging energy consumed by different users on weekdays (outer ring) and weekends (inner ring). The distribution trend of charging energy consumption of the two types of users is similar, whether on weekends or weekdays. On weekdays, the peak charging energy consumption of users without PCPs is in the range of 24–30 kWh, and for users with PCPs, it is in the range of 18–24 kWh, while on weekends, the peak charging energy consumption of both types of users is consistent in the range of 24–30 kWh. Because the distribution of charging energy consumption is approximately consistent with the normal distribution, it is meaningful to analyze the difference in the average value between two types of users. By analyzing the data, it is found that on weekdays, the average charging energy consumption of users with PCPs and users without PCPs are 22.91 kWh and 23.61 kWh, respectively. On weekends, the average charging energy consumption of users with PCPs and users without PCPs are 21.25 kWh and 22.32 kWh, respectively. There is no significant difference in charging energy consumption between the two types of users.

4.4. Prediction of Charging Load

Based on the above analysis results, we use a Monte Carlo simulation method to calculate the charging load of different types of users and explore the impact of charging load on the grid load. Figure 19 is the flow chart of BEV charging load simulation. The specific steps are as follows:

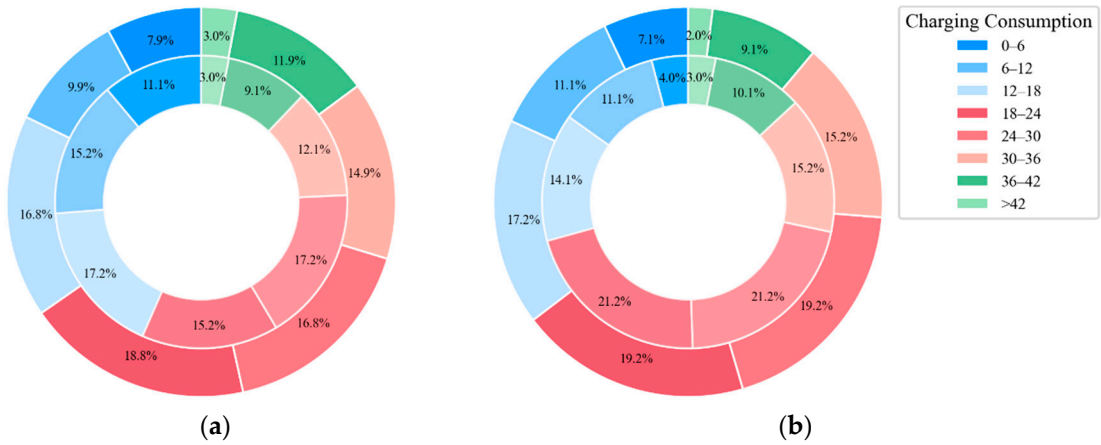


Figure 18. Comparison of charging energy consumption between weekdays (outer ring) and weekends (inner ring). (**a**) Users with PCPs. (**b**) Users without PCPs.

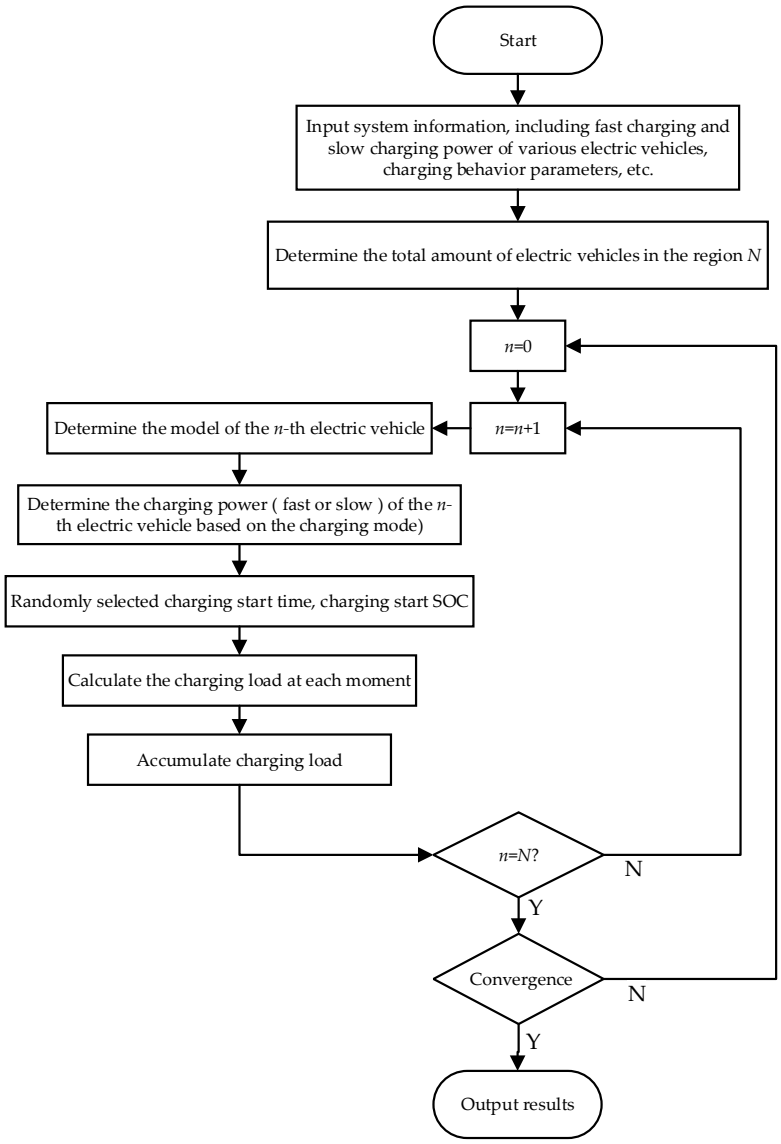


Figure 19. Charging load forecasting method flow chart.

(1). Determine the total number N of BEVs in the region; this paper set N to 50,000.
(2). Determine the model of BEV. Since the three types of vehicles in Table 2 occupy a large proportion of the BEV market, these three types of vehicles are taken as the main object. The possibilities of extracting these three types of vehicles are set to be the same.
(3). Determine the charging mode of the electric vehicle, namely fast charging or slow charging.
(4). According to the previous analysis results, the charging start time and the charging start SOC are randomly selected. The charging duration T(h) is calculated according to Equation (3).

$$T = (SOC_{targ} - SOC_{start}) \times C/(\eta \times P) \tag{3}$$

In the formula, SOC_{targ} is the target SOC, SOC_{start} is the SOC at the beginning of charging, C is the battery capacity, η is the charging efficiency, and P is the charging power. In this simulation, the SOC_{targ} of each charging of the vehicle is set to be 100%. At the same time, the vehicle battery capacity C and charging power P are shown in Table 2.

(5). The charging load of each time period is calculated in hours. Then the charging load curve of one vehicle is generated.
(6). Repeat the above process, accumulate the charging load curve generated each time, and finally obtain the charging load curve of all vehicles.
(7). Variance coefficient is used to judge whether the algorithm converges in this paper:

$$\beta_i = \frac{\sigma_i(\overline{P})}{\sqrt{k} \cdot \overline{P_i}} \tag{4}$$

where β_i is the variance coefficient of charging load at time i; $\sigma_i(\overline{P})$ is the standard deviation of charging load at time i; $\overline{P_i}$ is the expectation of charging load at time i; and k is the number of calculations. We set $\max(\beta_i) \leq 0.05\%$.

In comparing the charging load differences in different types of users, as shown in Figure 20, we calculated the charging load distribution of users with or without PCPs on weekdays and weekends. It can be found that: (i). On weekdays, users without PCPs have obvious load peaks around 10:00 and 17:00; while on weekends, the peak time of the charging load appears at about 19:00; (ii). The peak time of the charging load curve of users with PCPs on weekdays and weekends are not much different, which is concentrated at about 20:00; (iii). Due to the existence of PCPs, whether on weekdays or on weekends, the charging load of users with PCPs are lower than the users without PCPs.

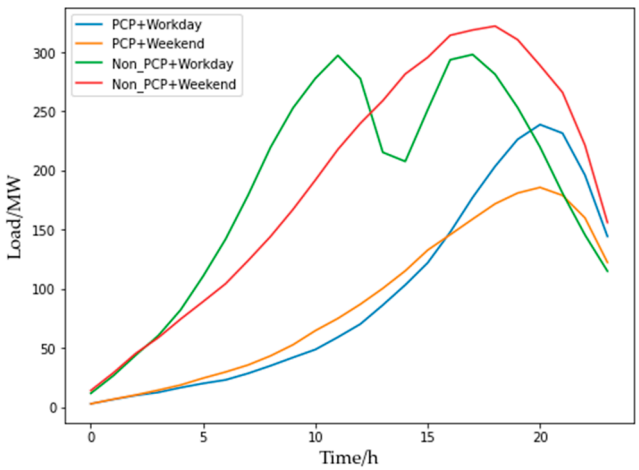


Figure 20. Charging load distribution of different types of users.

In addition, in terms of the impact of charging load on the power grid, as shown in Figure 21, we superimpose the charging load of different types of BEV users and base load. It can be found that the electricity demand is the largest in the period of 19:00–21:00. At the same time, there is an obvious phenomenon of peak addition, which increases the pressure of the power grid to a certain extent. Therefore, it is necessary to take orderly charging or optimize power equipment to meet greater power demand.

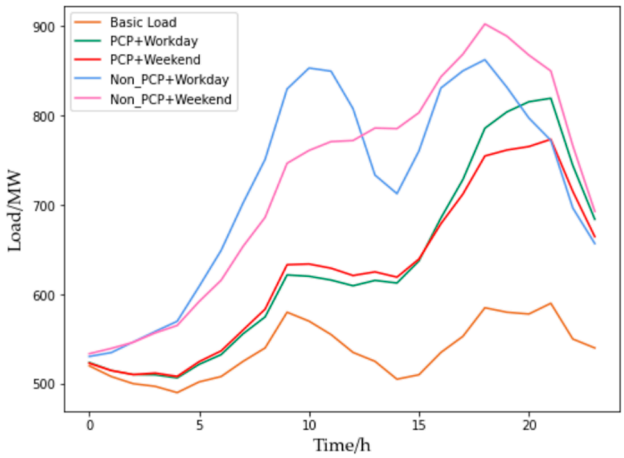


Figure 21. Superposition distribution of charging load and power load.

5. Influence of Trip Chain

Based on the discretization of the user's charging time, the utilization rate R of the PCP is analyzed. The utilization rate R_i in the i-th time period is equal to the ratio of the number of users being charged during this time period Num_i to the total number of users Num. Figure 22 shows the utilization rate of PCPs on 6 January. It can be found that the using peaks of PCPs are at 6:00–7:00 and 19:00. In addition, the utilization rate of PCPs is very low, or even 0, in some time periods. PCP sharing is an effective way to improve the utilization rate of PCPs.

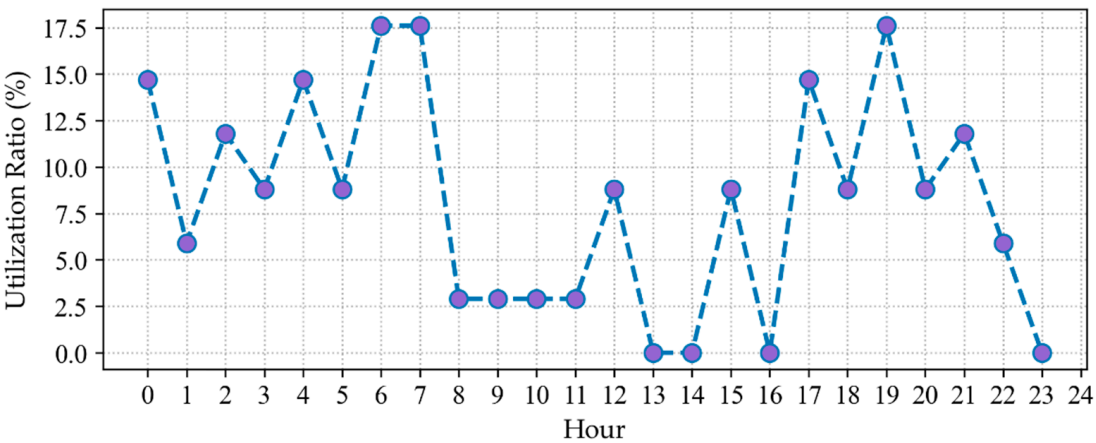


Figure 22. PCP utilization rate on a certain day.

In Beijing, there have been many enterprises exploring the sharing mode of charging piles as shown in Figure 23, but it still needs to be improved. This section will analyze the charging behavior of users with PCPs and summarize the influence mechanism of the user's charging mode. Finally, the MLA is used to analyze the charging behavior, which provides a reference for the future realization of intelligent sharing of PCPs.

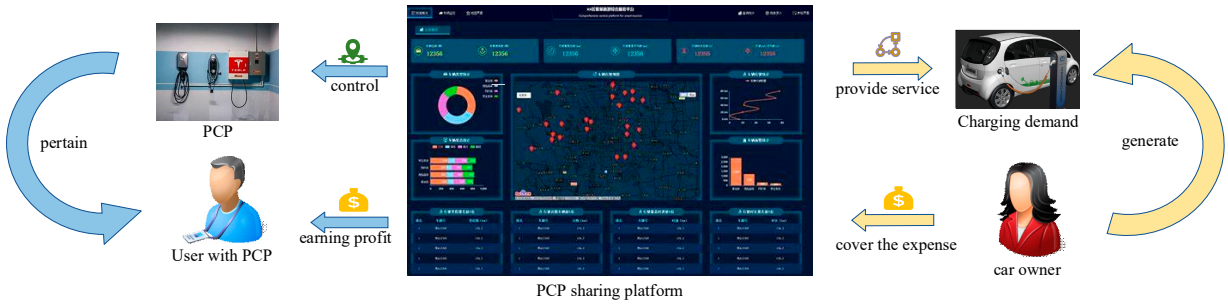


Figure 23. PCP sharing platform.

5.1. Trip Chain Generation

According to the results obtained in Part 3, the data of users with PCPs are screened from all users. According to the Table 3, the "Parking and power off" is used as the main criterion for distinguishing trips. When the "Parking and power off" state lasts for more than 3 min [34], the trip is considered to be over. When the state is switched to another state, the corresponding data point is the beginning of the next trip. Similarly, the charging state is mainly judged according to state 4. When the "Parking and Charging" state lasts more than 5 min, it is considered that the current vehicle is in charging.

A preliminary trip chain is generated for each trip in chronological order. It is necessary to test the consistency of the starting position and the ending position, since the trip chain should be a closed loop. Considering that there will be a certain deviation in each parking location, the start point and the end point of the trip chain may differ by a certain distance. Referring to the research of Zhao et al. [35], this paper sets the threshold to 1000 feet. When the distance between the start points and the end point is less than 1000 feet, the travel chain is retained. Otherwise, the travel chain is discarded.

The statistical results of trip chains are shown in Table 9. A total of 13,341 trip chains are extracted from 168 BEV users. It can be found that in the trip chains involving charging behavior, the proportion of PCP charging is 17.73%, and non-PCP charging also accounts for a certain proportion of 11.59%. Figure 12 also shows that users with PCPs have many other charging options besides charging in residential areas. At the same time, it is noted that the above two charging trip chains are far less than the no-charging trip chain (70.68%), which is indicating that most BEVs users do not choose to charge during their daily trips.

Table 9. Overview of trip chain summary results.

Type	Non-PCP Charging	PCP Charging	No-Charging	Total
Number of chains	1547	2365	9429	13,341
Rate (%)	11.59	17.73	70.68	100

The ABTCM method [36] is adopted to extract the charging and driving behavior of users with PCPs from vehicle operation data. This section focuses on charging in PCPs for users with PCPs, who start and end at home among their trips and may pass many places in a day, as depicted in Figure 24. There are many factors that affect a user's charging behavior, such as season, weekday, charging location, charging fees, etc. [37,38]. This study systematically extracted the corresponding trip information involving vehicle state and external environment from ABTCM, as shown in Table 10.

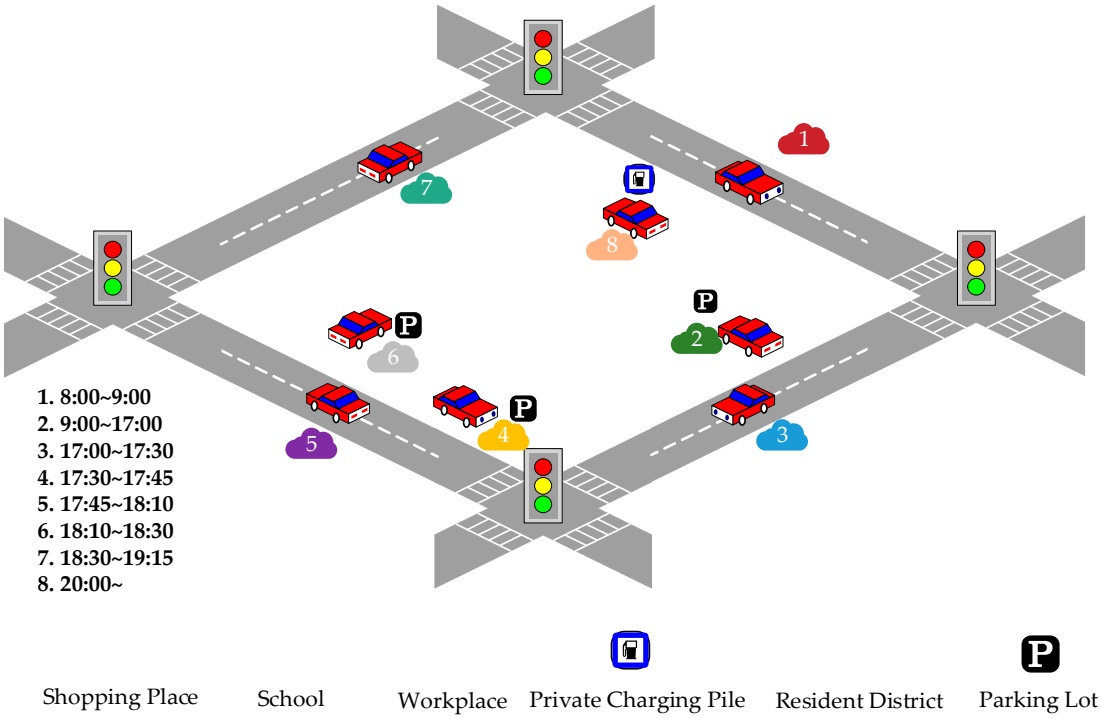


Figure 24. Possible travel trajectories for users with PCPs.

Table 10. List of affecting variables.

Type	Variables	Description
Vehicle State	Start SOC	SOC at the beginning of the journey
	SOC decline	SOC declined in the trip
	Start Time	The start time of the trip
	Travel time	Travel time (excluding parking)
	Mileage	Distance in trip
	Average Speed	Average speed in a trip
External environment	Average Temperature	Temperature of the day of travel
	Month	Represents the change of seasons
	Week	Whether or not is weekday

Figure 25 is a schematic diagram of the BEV trip chain. Assuming that the number of trips on the user's i-th day is n, the user's entire travel chain is $\{Trip_1, Trip_2, \ldots, Trip_n\}$.

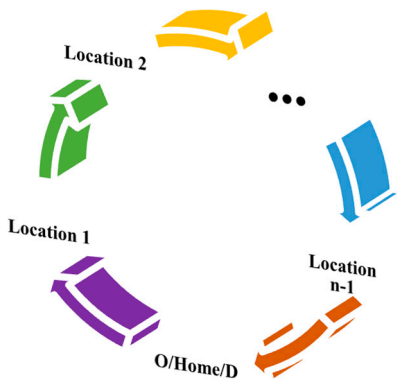
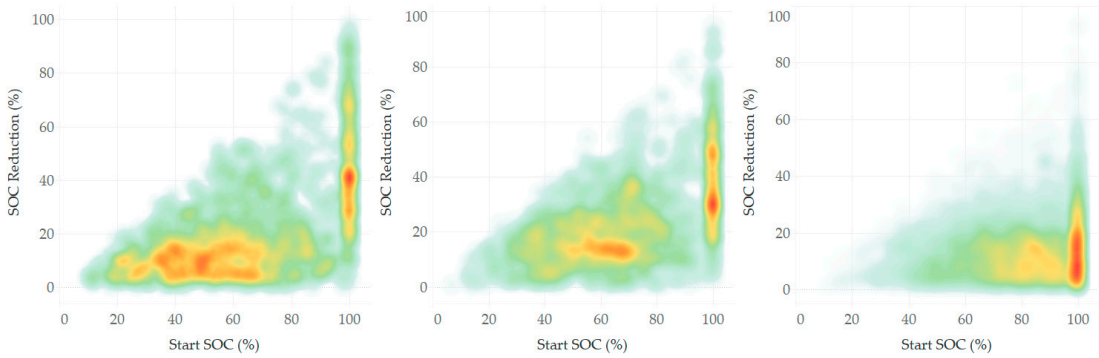


Figure 25. BEV trip chain diagram. O: Origin; D: Destination.

5.2. Key Factors Analysis

Figure 26 shows the comparison of the factors affecting the choice of different charging methods. In the three cases of PCP charging, Non-PCP charging and No-charging, the initial SOC is partially concentrated at 90–100%. Because the users with PCPs choose to travel at a higher SOC level in many cases, no matter whether it is charged or not in the day. Through comparison, it is found that the possibility of users charging during travel will increase significantly with the decrease in SOC at the beginning. In Figure 26a,b, for the Non-PCP charging, the SOC decline by 0–20% is more obvious. The SOC decline of Non-PCP charging and PCP charging is obviously concentrated between 20% and 60%. It can be found that the larger the SOC decline, the greater the willingness of the users to choose to charge, but the selected charging method is not fixed. Meanwhile, as shown in Figure 26c, people are less likely to charge at higher SOC levels.



(**a**) (**b**) (**c**)

Figure 26. Variable comparison under different charging modes: (**a**) Non-PCP charging; (**b**) PCP charging; (**c**) No charging.

It can also be found that for short-distance travel (0–50 km), the proportion of users No charging (76.8%) is significantly higher than other types (PCP charging: 13.2%; Non-PCP charging: 10.0%). In the case of mileage greater than 50 km, the proportion of users No charging only accounted for 55.1%. It is worth mentioning that the user might turn on the automotive air conditioner in lower or higher temperatures, which will accelerate the SOC decline and stimulate the user's charging behavior.

Because it is a multi-classification problem, the MLR model [39] is used to quantify the impact of relevant variables on the user's charging mode, assuming that there are p independent variables, k response variables, and n samples. In order to construct logic in the case of polynomials, one of the categories will be considered as a reference, and all other logit are constructed relative to it. Without loss of generality, category 1 is set as the reference. Let p_j denote the probability that the target value belongs to class j, and the multivariate logistic regression model illustrates the relationship between the probability p_j and the p explanatory variables $x_1, x_2, \ldots, x_p$, according to Equation (5).

$$log(\frac{p_j(x_i)}{p_1(x_i)}) = a_i + \beta_{1j}x_{1i} + \beta_{2j}x_{2i} + \ldots + \beta_{pj}x_{pi},\ i = 1, 2, \ldots, n,\ j = 2, \ldots, k \quad (5)$$

Taking the No-charging data as a reference, the model parameters are estimated by the maximum likelihood method. The model results are shown in Table 11. These results are generally consistent with our intuitive understanding of BEV charging behavior. The fact is that users decide to charge or not based on whether the remaining energy is sufficient to support their remaining trip, and the remaining energy is related to the SOC at the time of departure and the SOC reduction during the trip. Obviously, Start SOC and SOC decline have significant negative and positive effects, respectively. In the case of lower start

SOC and higher SOC decline during driving, users will choose to charge. The charging probability also increases with the increasing travel distance.

Table 11. Logistic Regression outputs.

Variables	Coefficient	Standard Error	Odds Ratio	*p*	[95% Conf. Interval]	
			PCP charging			
Start SOC (%)	−4.513	0.161	0.011	<0.001	0.009	0.014
SOC Decline (%)	3.574	0.129	35.651	<0.001	15.940	79.738
Mileage (km)	4.134	0.411	62.413	<0.001	21.387	182.135
Average Speed (km/h)	1.294	0.546	3.649	<0.001	2.269	5.867
Week	0.023	0.242	1.023	0.777	0.872	1.200
Start Time (h)	−1.435	0.081	0.238	<0.001	0.151	0.374
Travel Time (h)	−0.190	0.231	0.827	0.202	0.617	1.107
Month	−0.026	0.149	0.974	0.820	0.777	1.221
Average Temperature (°C)	−0.661	0.115	0.516	<0.001	0.401	0.665
			Non-PCP charging			
Start SOC (%)	−4.897	0.151	0.007	<0.001	0.006	0.010
SOC Decline (%)	5.360	0.497	212.682	<0.001	80.303	563.285
Mileage (km)	−0.135	0.686	0.873	0.843	0.228	3.351
Average Speed (km/h)	3.588	0.273	36.157	<0.001	21.159	61.789
Week	−0.020	0.096	0.980	0.833	0.812	1.182
Start Time (h)	−2.535	0.279	0.079	<0.001	0.046	0.137
Travel Time (h)	−1.086	0.186	0.338	<0.001	0.235	0.486
Month	0.411	0.133	1.509	0.002	1.163	1.957
Average Temperature (°C)	−0.682	0.149	0.506	<0.001	0.377	0.678

Reference Category: No charging. Pseudo R-squared: 0.321.

In addition, by comparing the differences in the charging attributes of Non-PCP charging and PCP charging, it can be found that: (i) The choice of PCPs charging is more sensitive to distance than that of Non-PCP charging. For PCP charging, the driving distance shows a positive impact, which indicates that the longer the driving distance, the greater possibility of choosing PCP charging. (ii) When the user chooses to charge in public charging piles, the average speed in the trip has a more obvious impact, which is indicating that during the trip, the faster the user's speed, the greater possibility of charging in PCPs due to mileage anxiety.

5.3. Influencing Analysis

In order to better understand the user's charging behavior and make a foundation for the future intelligent sharing of PCPs, in this section, 13,341 trip chain data extracted by ABTCM are used to analyze the user's charging behavior, and the importance of influencing factors are sorted. The model has three categorical variables, including No-charging, Non-PCP charging and PCP charging. All potential impact variables (independent variables) considered to be related to the charging mode are input into the model. For the selection of training data and verification data, random sampling is used to ensure that training data and verification data obey the same distribution condition. Due to the high imbalance of different types of data, the training effect of the model is poor. Therefore, the method of oversampling is adopted in this paper to increase the number of charging samples.

At present, many studies have used machine learning methods to predict charging behavior [40–42], and the prediction effect is good. Therefore, Random forest (RF), Deep neural network (DNN), XGBoost, and Support vector machine (SVM) are selected in this paper to predict the charging behavior of BEV users, and the prediction results of different models are compared.

5.3.1. Support Vector Machine

SVM has the advantages of global optimization and simple structure, and it still has strong classification in the case of small samples and high dimensions [43].

According to SVM theory, there is a hyperplane for the samples to be classified, so that the two types of samples are completely separated. The hyperplane is:

$$w^T x + b = 0 \tag{6}$$

In Equation (6): w represents the normal vector of the hyperplane and determines the direction of the hyperplane; b represents the distance between the hyperplane and the origin:

$$miny = \frac{\| w \|^2}{2} + C\left(\sum_{i=1}^{n} \xi_i\right) \tag{7}$$

$$s.t.\ \ y_i\left(x_i \cdot w^T + b\right) \geq 1 - \xi_i, i = 1, 2, \ldots, n \tag{8}$$

where C is the penalty factor, which controls the penalty degree of the misclassified samples; ξ_i is a slack variable, and each sample has a corresponding slack variable that characterizes the extent to which the sample does not satisfy the constraint.

By optimizing Equations (7) and (8), an optimal hyperplane can be found that separates different types of data and separates the farthest.

5.3.2. XGBoost

XGBoost is to continuously add a new weak estimator to fit the error generated by the previous weak estimator training, so that the residual between the true value and the predicted value is continuously reduced. After iteration ends, the predicted results on each estimator are weighted and combined to obtain the predicted results. In XGBoost, its prediction function is shown in Equation (9):

$$y_i^* = \sum_{j=1}^{N} f_j(x_i) \tag{9}$$

In the formula: x_i is the input of the i-th sample; N is the number of decision trees; $f_j(x_i)$ is the predicted value of the i-th sample on the j-th tree; and y_i^* is the predicted value of the i-th sample.

The objective function of XGBoost consists of two parts: loss function and regularization. The regularization objective function is shown in Equations (10) and (11):

$$Obj^{(t)} = \sum_{i}^{n} l(y_i, y_i^*) + \sum_{i=1}^{t} \Omega(f_i) \tag{10}$$

$$\Omega(f_i) = \gamma T + \frac{1}{2}\alpha \sum_{j=1}^{T} w_j^2 \tag{11}$$

In Equation (10) $Obj^{(t)}$ denotes the objective function after t iterations; y_i represents the actual value of the i-th sample; $\Omega(f_i)$ is the penalty term of the t-th iteration model, which can reduce the overfitting of the model. In Equation (11), w_j is the weight of leaf node; T is the number of leaf nodes; γ is the penalty coefficient; and α is the regularization coefficient.

5.3.3. Random Forest

Different from other machine learning models, random forest samples and random sampling of features reduce the sensitivity to data noise and outliers in the classification process, and effectively avoid overfitting.

The basic classifier of random forest algorithm is Classification and Regression Tree (CART). The algorithm constructs a decision tree based on information entropy and specific criteria. The Gini index minimum criterion is used to select the feature attributes when the node is split. The feature attributes can be selected by the Gini index and the purity of the sample can be reflected by the Gini value.

The purity of the data set δ is defined as:

$$Gini(\delta) = \sum_{i=1}^{I} \sum_{i' \neq i} p_i p_{i'} = 1 - \sum_{i=1}^{I} p_i^2 \tag{12}$$

In the formula: p_i is the probability that the sample point belongs to the class i; I represents the category of the data set δ.

The Gini index is defined as:

$$Gini(\delta, \alpha) = \sum_{k=1}^{I} \frac{\delta^k}{\delta} \cdot Gini\left(\delta^k\right) \tag{13}$$

In the formula: α is the characteristic condition; k represents the class in the data set δ that satisfies the characteristic condition α.

5.3.4. Deep Neural Network

The most important functions of DNN are nonlinear mapping function and strong generalization ability. This paper implements the DNN model through the TensorFlow deep learning framework. The DNN structure is shown in the Figure 27. The input variable of the network is $x = (x_1, x_2, \ldots, x_n)^T$, the output vector is $y = (y_1, y_2, \ldots, y_k)^T$, and the weight matrices w and v in the hidden layer are $(w_1, w_2, \ldots, w_m)$ and $(v_1, v_2, \ldots, v_l)$, respectively.

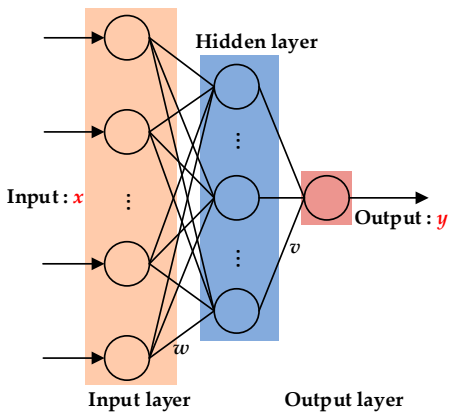


Figure 27. The structure of neural network.

For hidden layers:

$$o_j = f^1\left(\sum_{i=0}^{N} w_{ij} y_i\right) \quad i = 1, 2, \ldots, l \tag{14}$$

For output layer:

$$y_j = f^2\left(\sum_{i=0}^{N} v_{ij} x_i\right) \quad i = 1, 2, \ldots, n \tag{15}$$

The function f^1 corresponding to the hidden layer is the relu function, and the function f^2 corresponding to the output layer is the SoftMax function.

As shown in Figure 28, the RF uses the Bagging method to generate an independent identically distributed training sample set for each decision tree, and the final classification result depends on the voting of all decision trees. In this paper, the grid search method

is used to determine the hyperparameters of SVM, XGBoost, and RF models. For RF: the number of trees is 250 and the depth of the tree is 25; for SVM: Gaussian kernel function is selected, the kernel function coefficient is 0.01, and the penalty coefficient is 1000; for XGBoost: the learning rate is 0.3, the maximum depth of the tree is 15, and the number of trees is 300.

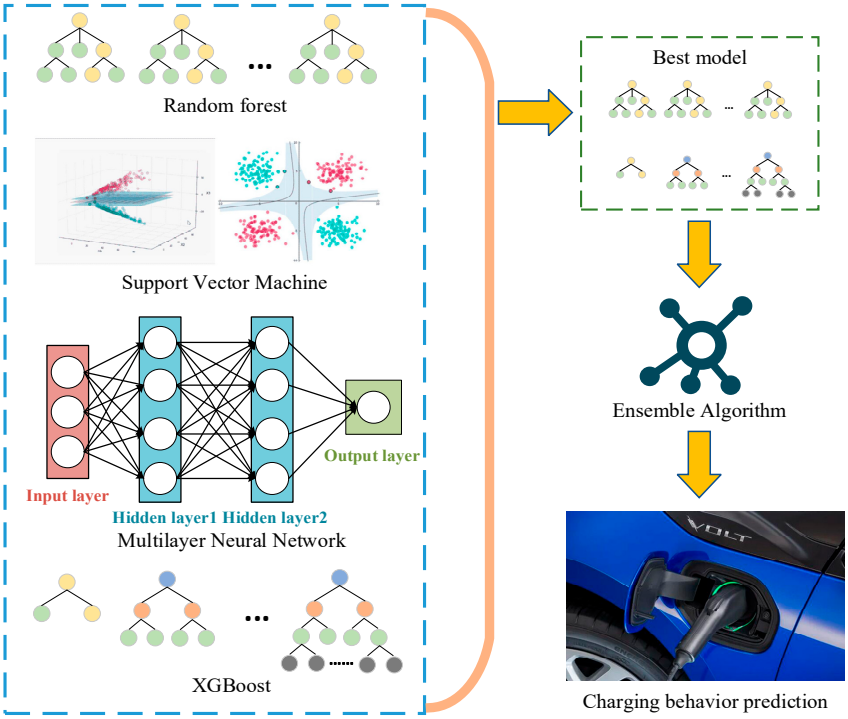


Figure 28. Flow chart of prediction method for charging behavior.

For ANN training, when the number of hidden layers of the network is three (the numbers of neurons in each layer are 128, 256 and 128, respectively), the structure is the most suitable, and the dropout method is used to prevent overfitting during the training process. The Adam algorithm is used to optimize the model, with Relu as the activation function of the hidden layer and softmax as the activation function of the output layer. In addition, the learning rate is set to 0.001, the batch size is 32, and the number of iterations is 15 epochs.

Table 12 shows the output of the model's test set. A confusion matrix is used, and three test metrics are adopted to evaluate the prediction effect of the models, which are Precision, Recall, and F1-score. It can be found that the Precision of RF and XGBoost is slightly higher, which indicates that the prediction results of the RF model are often closer to the actual results.

Table 12. Results of different models.

Type	Precision			Recall			F1-score		
	PCPC	NPCPC	NC	PCPC	NPCPC	NC	PCPC	NPCPC	NC
ANN	0.67	0.68	0.77	0.68	0.76	0.68	0.67	0.71	0.72
RF	0.84	0.88	0.85	0.87	0.91	0.79	0.86	0.90	0.82
XGBoost	0.84	0.89	0.87	0.89	0.91	0.81	0.87	0.90	0.84
SVM	0.54	0.56	0.78	0.29	0.13	0.96	0.38	0.21	0.86
Stacking Ensemble	0.86	0.86	0.91	0.89	0.90	0.83	0.87	0.90	0.85
Voting Ensemble	0.85	0.90	0.87	0.89	0.91	0.81	0.87	0.90	0.84

PCPC: PCP Charging; NPCPC: Non-PCP Charging; NC: No Charging.

To further improve the accuracy of the model, the above two best performing models are then integrated as the Ensemble model. We use two variants of ensemble superposition, namely voting classifier and stacking classifier. In the voting classifier, multiple base classifiers are trained on the entire training set, and the average prediction made by the base model is taken as the final prediction. The stacking classifier is based on the concept of stacked generalization, where prediction from the underlying model is used as inputs to the final estimator, which is trained using cross-validation to generate prediction [44]. It can be found that the best prediction results are obtained using the integrated algorithm with a maximum accuracy of 91%.

6. Conclusions

Based on the on-board data of 168 private BEV users in Beijing, deep insight on the charging behaviors are carried out to find the rules and relationship between the influencing factors, such as driving cycle, temperature, owning private charging piles or not, etc. The appearance of the PCPs can affect the users' charging behaviors to a certain extent. The main conclusions are summarized as follows:

The charging peak effect is evident, no matter on the weekday or weekend. The different charging patterns for weekdays and weekends among users with PCPs and users without PCPs were identified. For users with PCPs, the charging peak appears at night from 20:00 to 21:00. While for the users without PCPs, on the weekday, there are two charging peaks in public charging piles; one is at 10:00–11:00 and the other is at 17:00–18:00. However, at the weekend, the peak appears at 19:00–20:00. In addition to comparing the charging differences in different user groups, we superimposed the charging load curve of the BEVs with the base load curve and found a very obvious peak addition situation. This will lead to potential high negative influence on the power load to grid. Therefore, the orderly charging based on peak–valley price differences will be the key factor to change the user's charging behavior.

The charging behaviors of Beijing BEV users are influenced by many factors; however, range anxiety plays the most important role among them for users without PCPs. Due to the presence of PCPs, users with PCPs have lower mileage anxiety, more diverse charging time options, and a wider charging time duration. With 40% or lower start SOC, the proportion of users without PCPs (weekday: 55.9%; weekend: 59.9%) is larger than users with PCPs (weekday: 45.5%; weekend: 42.6%). The lack of available charging piles nearby is the main reason for the low charging start SOC level. Meanwhile, users without PCPs have a certain decrease in the range of 60–80% start charging SOC. The median charging time duration of users with PCPs is 51.44 h (2–3 days). In addition, the median charging time duration is 17.25 h (0–1 day) for users without PCPs. Obviously, the overall charging interval of users without PCPs are shorter than that of users with PCPs, indicating that PCPs greatly decreases the charging frequency of users.

The rules of charging behaviors are very important to obtain an optimal infrastructure deployment scheme and draft more feasible incentive policies. However, in the next step, still some important factors should be included in our research. Future research will consider the driving behavior of users and establish the usage patterns of BEVs for different users, taking the influence of the performance evolution of BEV into consideration, as well to complete the model, so as to better evaluate the impact of BEV charging load on the electricity grid, optimize the layout of charging facilities, and realize the construction of a PCP sharing platform.

Author Contributions: Conceptualization, Y.S.; supervision, J.D.; data curation, F.H.; Writing—original draft, H.T.; writing—review and editing, H.T., Y.S., J.D. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the General Program of Science and Technology Development Project of Beijing Municipal Education Commission (No. KM202010005019) and Beijing Natural Science Foundation (3192016).

Data Availability Statement: The authors do not have permission to share data.

Conflicts of Interest: The authors declare no conflict of interest.

Nomenclature

PCP	Private charging pile
EV	Electric vehicle
BEV	Battery electric vehicle
MC	Monte Carlo
ABTCM	Agent-based trip chain model
MLR	Multinomial logistic regression
MLA	Machine learning algorithm
RF	Random forest
DNN	Deep neural network
SVM	Support vector machine
SOC	State of charge
POI	Point of interest
CART	Classification and regression tree

References

1. Hu, X.; Chen, N.; Wu, N.; Yin, B. The Potential Impacts of Electric Vehicles on Urban Air Quality in Shanghai City. *Sustainability* **2021**, *13*, 496. [CrossRef]
2. Song, M.; Fisher, R.; Kwoh, Y. Technological Challenges of Green Innovation and Sustainable Resource Management with Large Scale Data. *Technol. Forecast. Soc. Chang.* **2019**, *144*, 361–368. [CrossRef]
3. By the End of 2021, China's New Energy Vehicle Ownership Reached 7.84 Million—Rolling News—Chinese Government Network. Available online: http://www.gov.cn/xinwen/2022-01/12/content_5667734.htm (accessed on 28 October 2022).
4. Jian, L.; Yongqiang, Z.; Larsen, G.N.S.; Snartum, A. Implications of Road Transport Electrification: A Long-Term Scenario-Dependent Analysis in China. *eTransportation* **2020**, *6*, 100072. [CrossRef]
5. China Charging Alliance: Number of Public Charging Piles Increased by 56.6% in September—China Electric Vehicle Association. Available online: http://www.ceva.org.cn/cn/viewnews/20221012/2022101210056.htm (accessed on 3 November 2022).
6. Patt, A.; Aplyn, D.; Weyrich, P.; van Vliet, O. Availability of Private Charging Infrastructure Influences Readiness to Buy Electric Cars. *Transp. Res. A Policy Pract.* **2019**, *125*, 1–7. [CrossRef]
7. Tan, Z.; Yang, Y.; Wang, P.; Li, Y. Charging Behavior Analysis of New Energy Vehicles. *Sustainability* **2021**, *13*, 4837. [CrossRef]
8. Chen, Y.; Xia, D. Systematic Density Functional Theory Investigations on Cubic Lithium-Rich Iron-Based Li_2FeO_3: A Multiple Electrons Cationic and Anionic Redox Cathode Material. *eTransportation* **2021**, *10*, 100141. [CrossRef]
9. Liu, T.; Yang, X.-G.; Ge, S.; Leng, Y.; Wang, C.-Y. Ultrafast Charging of Energy-Dense Lithium-Ion Batteries for Urban Air Mobility. *eTransportation* **2021**, *7*, 100103. [CrossRef]
10. Hecht, C.; Victor, K.; Zurmühlen, S.; Sauer, D.U. Electric Vehicle Route Planning Using Real-World Charging Infrastructure in Germany. *eTransportation* **2021**, *10*, 100143. [CrossRef]
11. Dixon, J.; Bell, K. Electric Vehicles: Battery Capacity, Charger Power, Access to Charging and the Impacts on Distribution Networks. *eTransportation* **2020**, *4*, 100059. [CrossRef]
12. '14th Five-Year' Beijing Electric Vehicle Charging Pile Will Reach 700,000—Local Government—Chinese Government Network. Available online: http://www.gov.cn/xinwen/2022-04/13/content_5684980.htm (accessed on 28 October 2022).
13. Dominguez-Jimenez, J.A.; Campillo, J.E.; Montoya, O.D.; Delahoz, E.; Hernández, J.C. Seasonality Effect Analysis and Recognition of Charging Behaviors of Electric Vehicles: A Data Science Approach. *Sustainability* **2020**, *12*, 7769. [CrossRef]
14. Xing, Q.; Chen, Z.; Zhang, Z.; Wang, R.; Zhang, T. Modelling Driving and Charging Behaviours of Electric Vehicles Using a Data-Driven Approach Combined with Behavioural Economics Theory. *J. Clean. Prod.* **2021**, *324*, 129243. [CrossRef]
15. Nicholas, M.A.; Tal, G.; Davies, J.; Woodjack, J. DC Fast as the Only Public Charging Option? Scenario Testing from GPS-Tracked Vehicles. In Proceedings of the Transportation Research Board Conference, Washington, DC, USA, 22–26 January 2012; Available online: https://trid.trb.org/view/1130070 (accessed on 29 November 2022).
16. He, Y.; Kockelman, K.M.; Perrine, K.A. Optimal Locations of U.S. Fast Charging Stations for Long-Distance Trip Completion by Battery Electric Vehicles. *J. Clean. Prod.* **2019**, *214*, 452–461. [CrossRef]
17. Xu, Z.; Wang, J.; Lund, P.D.; Zhang, Y. Estimation and Prediction of State of Health of Electric Vehicle Batteries Using Discrete Incremental Capacity Analysis Based on Real Driving Data. *Energy* **2021**, *225*, 120160. [CrossRef]
18. GaryLea ChangeCoordinate 2022. Available online: https://github.com/GaryLea/ChangeCoordinate (accessed on 21 November 2022).

19. Geo/Inverse Geocoding—API Documentation—Development Guide—Web Services API—Amap API. Available online: https://developer.amap.com/api/webservice/guide/api/georegeo (accessed on 4 November 2022).
20. Electric Pile Home—Electric Life. Available online: https://www.powerlife.com.cn/pile/ops/map/default.html (accessed on 22 November 2022).
21. Jordán, J.; Palanca, J.; Del Val, E.; Julian, V.; Botti, V. A Multi-Agent System for the Dynamic Emplacement of Electric Vehicle Charging Stations. *Appl. Sci.* **2018**, *8*, 313. [CrossRef]
22. Pagany, R.; Marquardt, A.; Zink, R. Electric Charging Demand Location Model—A User- and Destination-Based Locating Approach for Electric Vehicle Charging Stations. *Sustainability* **2019**, *11*, 2301. [CrossRef]
23. Lu, Z.; Zhang, Q.; Yuan, Y.; Tong, W. Optimal Driving Range for Battery Electric Vehicles Based on Modeling Users' Driving and Charging Behavior. *J. Adv. Transp.* **2020**, *2020*, 8813137. [CrossRef]
24. Quirós-Tortós, J.; Ochoa, L.F.; Lees, B. A Statistical Analysis of EV Charging Behavior in the UK. In Proceedings of the IEEE/PES Innovative Smart Grid Technologies ISGT Latin America, Montevideo, Uruguay, 5–7 October 2015; pp. 445–449.
25. Canizes, B.; Soares, J.; Vale, Z.; Corchado, J.M. Optimal Distribution Grid Operation Using DLMP-Based Pricing for Electric Vehicle Charging Infrastructure in a Smart City. *Energies* **2019**, *12*, 686. [CrossRef]
26. Xu, M.; Yang, H.; Wang, S. Mitigate the Range Anxiety: Siting Battery Charging Stations for Electric Vehicle Drivers. *Transp. Res. C Emerg. Technol.* **2020**, *114*, 164–188. [CrossRef]
27. Yang, Y.; Yao, E.; Yang, Z.; Zhang, R. Modeling the Charging and Route Choice Behavior of BEV Drivers. *Transp. Res. C Emerg. Technol.* **2016**, *65*, 190–204. [CrossRef]
28. Williard, N.; He, W.; Osterman, M.; Pecht, M. Comparative Analysis of Features for Determining State of Health in Lithium-Ion Batteries. *IJPHM* **2020**, *4*, 1–7. [CrossRef]
29. Wan, X.; Wang, W.; Liu, J.; Tong, T. Estimating the Sample Mean and Standard Deviation from the Sample Size, Median, Range and/or Interquartile Range. *BMC Med. Res. Methodol.* **2014**, *14*, 135. [CrossRef] [PubMed]
30. Weldon, P.; Morrissey, P.; Brady, J.; O'Mahony, M. An Investigation into Usage Patterns of Electric Vehicles in Ireland. *Transp. Res. D Transp. Environ.* **2016**, *43*, 207–225. [CrossRef]
31. Sun, D.; Ou, Q.; Yao, X.; Gao, S.; Wang, Z.; Ma, W.; Li, W. Integrated Human-Machine Intelligence for EV Charging Prediction in 5G Smart Grid. *EURASIP J. Wirel. Commun. Netw.* **2020**, *2020*, 139. [CrossRef]
32. Gonzalez Venegas, F.; Petit, M.; Perez, Y. Plug-in Behavior of Electric Vehicles Users: Insights from a Large-Scale Trial and Impacts for Grid Integration Studies. *eTransportation* **2021**, *10*, 100131. [CrossRef]
33. Siddique, C.; Afifah, F.; Guo, Z.; Zhou, Y. Data Mining of Plug-in Electric Vehicles Charging Behavior Using Supply-Side Data. *Energy Policy* **2022**, *161*, 112710. [CrossRef]
34. Ren, Y.; Lan, Z.; Yu, H.; Jiao, G. Analysis and Prediction of Charging Behaviors for Private Battery Electric Vehicles with Regular Commuting: A Case Study in Beijing. *Energy* **2022**, *253*, 124160. [CrossRef]
35. Zhao, P.; Guan, H.; Wei, H.; Liu, S. Mathematical Modeling and Heuristic Approaches to Optimize Shared Parking Resources: A Case Study of Beijing, China. *Transp. Res. Interdiscip. Perspect.* **2021**, *9*, 100317. [CrossRef]
36. Lin, H.; Fu, K.; Wang, Y.; Sun, Q.; Li, H.; Hu, Y.; Sun, B.; Wennersten, R. Characteristics of Electric Vehicle Charging Demand at Multiple Types of Location—Application of an Agent-Based Trip Chain Model. *Energy* **2019**, *188*, 116122. [CrossRef]
37. Almaghrebi, A.; Aljuheshi, F.; Rafaie, M.; James, K.; Alahmad, M. Data-Driven Charging Demand Prediction at Public Charging Stations Using Supervised Machine Learning Regression Methods. *Energies* **2020**, *13*, 4231. [CrossRef]
38. Frendo, O.; Gaertner, N.; Stuckenschmidt, H. Improving Smart Charging Prioritization by Predicting Electric Vehicle Departure Time. *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 6646–6653. [CrossRef]
39. Chung, Y.-W.; Khaki, B.; Li, T.; Chu, C.; Gadh, R. Ensemble Machine Learning-Based Algorithm for Electric Vehicle User Behavior Prediction. *Appl. Energy* **2019**, *254*, 113732. [CrossRef]
40. Ge, X.; Shi, L.; Fu, Y.; Muyeen, S.M.; Zhang, Z.; He, H. Data-Driven Spatial-Temporal Prediction of Electric Vehicle Load Profile Considering Charging Behavior. *Electr. Power Syst. Res.* **2020**, *187*, 106469. [CrossRef]
41. Shahriar, S.; Al-Ali, A.R.; Osman, A.H.; Dhou, S.; Nijim, M. Prediction of EV Charging Behavior Using Machine Learning. *IEEE Access* **2021**, *9*, 111576–111586. [CrossRef]
42. Khwaja, A.S.; Venkatesh, B.; Anpalagan, A. Short-Term Individual Electric Vehicle Charging Behavior Prediction Using Long Short-Term Memory Networks. In Proceedings of the 2020 IEEE 25th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD), Pisa, Italy, 14–16 September 2020; pp. 1–7.
43. Li, Z.; Feng, X.; Wu, Z.; Yang, C.; Bai, B.; Yang, Q. Classification of Atrial Fibrillation Recurrence Based on a Convolution Neural Network With SVM Architecture. *IEEE Access* **2019**, *7*, 77849–77856. [CrossRef]
44. Naimi, A.I.; Balzer, L.B. Stacked Generalization: An Introduction to Super Learning. *Eur. J. Epidemiol.* **2018**, *33*, 459–464. [CrossRef] [PubMed]