WILEY | Hindawi

Research Article

Situation Awareness of Electric Vehicle Charging Load Based on Random Forest Algorithm

Wen Bo, Donglai Wang [ID], Yan Zhao, Quanzheng Li, and Zhen Zhang

Key Laboratory of Regional Multi-Energy System Integration and Control of Liaoning Province, Shenyang Institute of Engineering, Shenyang, Liaoning 110136, China

Correspondence should be addressed to Donglai Wang; donglai.wang@hotmail.com

Due to the important characteristics of energy saving and carbon reduction, electric vehicles have attracted worldwide attention. It can be predicted that the power grid will be faced with the access problem of large-scale electric vehicles. In order to master the user behavior characteristics of electric vehicle load, it is necessary to establish the model based on electric vehicle charging behavior. In this paper, combined with the electric vehicle charging demand and the situational awareness results of the dispatchable resources in the station area, the characteristic indicators of the electric vehicle load are quantitatively analyzed. Situational prediction of electric vehicle load based on random forest algorithm is proposed, and the sample set is divided and trained. A simulation example is used to verify the effectiveness of the method provided in load forecasting.

1. Introduction

Energy crisis and climate warming have become the key problems that restrict the sustainable development of human society. Electric vehicles have significant advantages in alleviating global energy shortages and environmental pollution [1–3]. At the same time, distributed power generation, which is mainly composed of photovoltaic and wind power, develops rapidly [4]. However, the access of large-scale electric vehicles and the intermittence of renewable energy will aggravate the peak-valley difference of power grid load, making the operation of power grid more difficult to control [5, 6]. On the one hand, as a flexible load, the electric vehicle's off-driving charging time is usually longer than the time required for its battery to be fully charged. Orderly regulating the charging behavior of a large number of electric vehicles can not only avoid the impact of large-scale random access of electric vehicles to the power grid but also realize the peak load cutting and valley load filling, which is very important for the safe operation of the power grid in the future. It is of great significance to improve the economic benefits of the power grid [7–10]. On the other hand, due to the strong randomness at the moment when electric vehicles connect and leave the power grid, the state of charge of the battery is uncertain, and the day-ahead scheduling process can not accurately predict the connection and departure of each electric vehicle in the future [11–13].

Short-term load forecasting is an important part of power system load forecasting, which mainly forecasts the load at any time in the future [14]. The factors that affect the accuracy of short-term load forecasting mainly include sudden weather changes, seasonal changes, dispatching plans, emergencies, large-scale social activities, and so on. Therefore, short-term load forecasting has randomness and uncertainty [15, 16]. However, in the process of random change, the load still has obvious periodicity in different periods, such as year, month, week, and day. Therefore, the short-term load variation comprehensively appears as a nonstationary random process in the time series [17, 18]. Accurate short-term load forecasting is of great significance to the economic dispatch of electric power dispatching departments, the control of unit allocation, and the current developing electric market [19].

The methods of early warning, decision making, and visualization in situational awareness technology play an important role in realizing the safe and stable operation of

smart grids. Reference [20] has designed a distribution network dispatching system based on situation awareness technology, which can flexibly detect and evaluate the running state of the distribution network. Reference [21] has already applied situational awareness to microgrid, and based on microgrid situational awareness, the situation leads to active decision. Reference [22] has proposed that the electric vehicle charging and discharging index is used for situational warning of the power supply capacity of the power distribution system. At present, the application of situational awareness in the distribution system is mainly aimed at the power supply and consumption early warning and monitoring of distribution network. However, the situation awareness of the distribution system to forecast and dispatch the charging load of electric vehicles has not been deeply developed.

In this paper, the basic user behavior characteristics of electric vehicle load are studied, and the charging behavior model of electric vehicle is established. The forecasting method of electric vehicle load based on random forest algorithm is put forward. The sample set is trained and compared with other methods in the field of short-term load forecasting. The validity and superiority of the random forest algorithm for electric vehicle load forecasting are verified, which can effectively improve the accuracy of electric vehicle charging load forecasting.

2. Modeling of Electric Vehicle Charging Behavior

2.1. Electric Vehicle Charging Behavior Analysis. After electric vehicles are connected to the power grid on a large scale, its charging load becomes an important factor that cannot be ignored in the power grid. Because the large-scale electric vehicle groups has certain aggregation characteristics, this paper takes the large-scale electric vehicle group as the research object and analyzes the charging demand of electric vehicles from the aspects of battery characteristics, charging time, mileage characteristics, and so on.

The characteristics of the battery mainly include battery capacity and the state of charge. The features include charging start time and charging duration. Driving features mainly include daily mileage and power consumption per kilometer of electric vehicles. The relationship between these 3 properties is as follows:

$$T_{ch} = \frac{Q_\omega \cdot d}{P_{EV}(k)} = \frac{Q_{EV} \cdot (1 - S_{SOC})}{P_{EV}(k)}, \tag{1}$$

where T_{ch} is the charging time in hours; Q_ω is the electricity consumption per kilometer of the electric vehicle, in kilowatt-hours per kilometer; d is the mileage of the electric vehicle, in kilometers; $P_{EV}(k)$ is the charging power in kilowatts; $P_{EV}(k)$ is the battery capacity in kWh; and S_{SOC} is the battery state of charge, which is the ratio of the remaining battery power to the fully charged state.

The daily mileage of electric vehicles obeys a normal distribution, and its probability density function is as follows:

$$f(d) = \frac{1}{d\delta_d \sqrt{2\pi}} \exp\left[-\frac{(\ln d - \mu_d)^2}{2\delta_d^2} \right], \tag{2}$$

where μ_d is the expected value of daily mileage, and δ_d is the standard deviation of daily mileage. According to [23], it is advisable to take $\mu_d = 3.2$ and $\delta_d = 0.88$.

The initial charging time of electric vehicles also has a certain normal distribution, and its probability density function is as follows:

$$f_{car}(t) = \frac{1}{t\delta_t \sqrt{2\pi}} \exp\left[-\frac{(\ln t - \mu_t)^2}{2\delta_t^2} \right], \tag{3}$$

where t is the initial charging time of each electric vehicle, and its value is in the range of [0, 1440] (24 h is divided into 1440 time periods, each of which is 1 minute); μ_t represents the expected value of charging time; and δ_t is the standard deviation of charging time. According to [23], it is advisable to take $\mu_t = 17.6$ and $\delta_t = 3.4$.

2.2. The Model of Single Electric Vehicle Charging Behavior. It shows that the charging power of each electric vehicle is constant and continuously adjustable in each optimization period. The mathematical model of the charging behavior of a single electric vehicle is shown in

$$
\begin{cases}
d_{EV}(t) = \eta \sum_{k=t_{in}}^{t} P_{EV}(k)\Delta t = d(t-1) + \eta P_{EV}(t)\Delta t, \\[2ex]
d_{min}(t_{out}) = d_{max}(t_{out}) = D_{expect}, \\[2ex]
d_{min}(t) \le d_{EV}(t) \le d_{max}(t), \\[2ex]
P_{max}(t) = \frac{\min(P_{EV,max}, d_{max}(t) - d(t-1))}{\eta \Delta t}, \\[2ex]
P_{max}(t) = \frac{\max(0, d_{min}(t) - d(t-1))}{\eta \Delta t}, \\[2ex]
P_{min}(t) \le P_{EV}(t) \le P_{max}(t),
\end{cases} \tag{4}
$$

where $d_{EV}(t)$ is the energy trajectory value of the electric vehicle at time t; $d_{min}(t)$ and $d_{max}(t)$ are the lower and upper bounds of the energy trajectory $d(t)$ at time t, respectively; η is the charging efficiency; $P_{EV}(k)$ is the electric vehicle charging power at time k and remains constant during the period between time k and time $k + 1$; Δt is the time interval of the scheduling period; t_{in} and t_{out} are the access and departure times of the electric vehicle, respectively; that is, the electric vehicle is connected at time t_{in} and leaves at time t_{out}; D_{expect} is the charging demand for electric vehicles; $P_{EV,max}$ is the upper limit of the rated charging power of the electric vehicle battery; $P_{EV}(t)$, $P_{max}(t)$, and $P_{min}(t)$ are the charging power of the electric vehicle at time t and the maximum and minimum charging power limited by the energy boundary constraints, respectively. The schematic

diagram of the energy boundary model of a single electric vehicle is shown in Figure 1.

In Figure 1, it is assumed that the electric vehicle is connected at time t_{in} and left at time t_{out}. The curve abd is the upper bound $d_{max}(t)$ of the energy boundary of the electric vehicle, which means that after the electric vehicle is connected to the power grid, it will be charged with the maximum charging power immediately until it reaches the user's expectation D_{expect}. The curve acd is the lower bound $d_{min}(t)$ of the energy boundary of the electric vehicle, which means that the electric vehicle delays charging after it is connected to the power grid, until the user's expectation D_{expect} is reached at the time of departure. The slopes of ab and cd represent the increase in battery power per unit time according to the maximum charging power of the electric vehicle, that is, $\eta P_{EV,max}$.
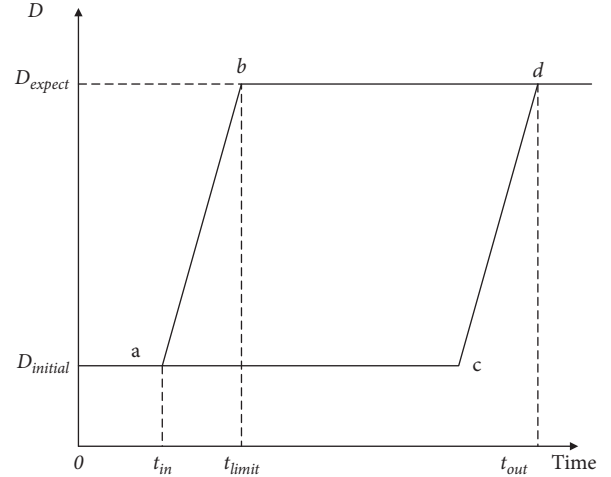
2.3. Cluster Equivalent Model for Electric Vehicles.

After the electric vehicles are connected to the power grid, the electric vehicles with the same departure time can be grouped into the same subcluster according to the charging parameters input by the owner. By superimposing the charging models belonging to the same subcluster, the equivalent charging model of a single subcluster can be obtained as follows:

$$\begin{cases} d_{\min,m}(t) \leq d_m(t) \leq d_{\max,m}(t), \\[2mm] d_m(t) = \eta \sum_{k=1}^{t} P_m(k)\Delta t = d_m(t-1) + \eta P_m(t)\Delta t, \\[2mm] d_{\min,m}(t) = \sum_{l=1}^{n_m} d_{\min,m,l}(t), \\[2mm] d_{\max,m}(t) = \sum_{l=1}^{n_m} d_{\max,m,l}(t), \\[2mm] P_m(t) = \sum_{l=1}^{n_m} P_{EV,m,l}(t), \\[2mm] P_{\max,m}(t) = \sum_{l=1}^{n_m} P_{EV,\max,m,l}(t) \geq P_m(t), \end{cases} \tag{5}$$

where $d_m(t)$ is the energy trajectory of subcluster m at time t. $d_{\min,m}(t)$ and $d_{\max,m}(t)$ are the lower and upper bounds of the energy trajectory of subcluster m at time t, respectively. $P_m(t)$ is the total charging power of subcluster m at time t. $P_{\min,m}(t)$ and $P_{\max,m}(t)$ are the lower and upper limits of the charging power of subcluster m at time t, respectively. n_m is the number of electric vehicles belonging to subcluster m at time t. $d_{\min,m,l}(t)$, $d_{\max,m,l}(t)$ and $P_{\min,m,l}(t)$, $P_{\max,m,l}(t)$ are the lower and upper bounds of the energy trajectory and the lower and upper bounds of the charging power of the L-th electric vehicle of subcluster m at time t, respectively.

Because the electric vehicles belonging to the same subcluster will leave at the same time, the cluster charging model is equivalent to the single charging model; that is, the charging strategy of the subcluster meeting the cluster charging model must follow a certain energy allocation

FIGURE 1: Energy boundary model of single electric vehicle.

mode. For the cluster charging amount set $s = [s_1, s_2, \ldots, s_t, \ldots s_T]$ that satisfies the cluster charging model, there must be at least one charging power distribution method so that all electric vehicles in the cluster are fully charged before leaving under the condition that the single charging model is satisfied. According to the cluster charging model, the following inequality holds for $n = 1, 2, \ldots, T$.

$$\sum_{t=1}^{n} \sum_{l \in \{l|t_{out}=t\}} d_l \leq \sum_{t=1}^{n} s_t \leq \sum_{t=1}^{n} \sum_{l \in \{l|t_{in}=t\}} d_l, \tag{6}$$

where d_l is the charging power demand of the L-th vehicle.

First, when $n = 1$, the inequality is transformed into

$$\sum_{l \in \{l|t_{out}=1\}} d_l \leq s_1 \leq \sum_{l \in \{l|t_{in}=1\}} d_l. \tag{7}$$

Using the strategy of giving priority to charging the electric vehicles that leave first, the charging plan for all electric vehicles at departure time $t_{out} = 1$ can be set as

$$P_{l1} = d_l, \tag{8}$$

so that it can be filled before leaving. Therefore, when $n = 1$, the proposition holds.

It is assumed that, for all time points m, the electric vehicles that leave before time point m can be fully charged within the cluster by adopting a strategy of charging the electric vehicles that leave first. That is, there is at least one charging plan set that satisfies the following equation:

$$\sum_{t=t_{in}}^{t_{out}} P_{lt} = d_l, \quad \forall l \in \{l|t_{out} \leq m\}, \tag{9}$$

$$P_{lt} \geq 0, \quad \forall t = t_{in}, \ldots, t_{out}, \forall l \in \{l|t_{out} \leq m\}.$$

The result is as follows:

$$\sum_{t=1}^{m} \sum_{l \in \{l|t_{out}=t\}} d_l \leq \sum_{t=1}^{m} s_t, \tag{10}$$

TABLE 1: Classification rule of situation levels.

Situation levels	Situational range	Regional status	Detailed description
1	[0, 0.2)	Unobstructed	Resources can fully meet the charging needs of electric vehicles
2	[0.2, 0.4)	Basically unobstructed	The resources can basically meet the charging needs of electric vehicles, and occasionally, there is insufficient supply
3	[0.4, 0.6)	Light congestion	Under normal circumstances, the resources can meet the charging needs of electric vehicles, but sometimes there will be insufficient supply
4	[0.6, 0.8)	Moderate congestion	The resources can meet the charging needs of electric vehicles in most cases
5	[0.8, 1.0]	Serious congestion	Resources are basically unable to meet the needs of electric vehicles and other loads

where $\sum_{t=1}^{m} s_t - \sum_{t=1}^{m} \sum_{l \in \{t_{\text{out}}=1\}} d_l$ represents the total amount of charge of the electric vehicle whose departure time is after the time point m. Due to the strategy of charging the electric vehicle that leaves first, once the electric vehicle at the departure time point m has been fully charged, the electric vehicle at the departure time point $m + 1$ will be preferentially arranged to be charged.

When

$$\sum_{t=1}^{m} s_t - \sum_{t=1}^{m} \sum_{l \in \{l|t_{\text{out}}=t\}} d_1 \geq \sum_{t=1}^{m} \sum_{l \in \{l|t_{\text{out}}=m+1\}} d_l. \tag{11}$$

The electric vehicle whose departure time is $m + 1$ is charged, and the electric vehicle with departure time point $m + 2$ will be charged.

$$\sum_{t=1}^{m} s_t - \sum_{t=1}^{m} \sum_{l \in \{l|t_{\text{out}}=t\}} d_l < \sum_{l \in \{l|t_{\text{out}}=m+1\}} d_l. \tag{12}$$

Then, the electric vehicle with departure time point $m + 1$ will be charged, and the total amount of charging is $\sum_{t=1}^{m} s_t - \sum_{t=1}^{m} \sum_{l \in \{l|t_{\text{out}}=t\}} d_l$.

When $n = m + 1$, because of

$$\sum_{t=1}^{m+1} \sum_{l \in \{l|t_{\text{out}}=t\}} d_l \leq \sum_{t=1}^{m+1} S_t, \tag{13}$$

it can be obtained as follows:

$$s_{m+1} \geq \sum_{l \in \{l|t_{\text{out}}=m+1\}} d_l - \left(\sum_{t=1}^{m} s_t - \sum_{t=1}^{m} \sum_{l \in \{l|t_{\text{out}}=t\}} d_l \right). \tag{14}$$

The above equations show that the charging plan s_{m+1} can meet the charging demand of the electric vehicle at the departure time point $m + 1$.

$$\sum_{t=t_{\text{in}}}^{t_{\text{out}}} P_{lt} = d_l, \quad \forall l \in \{l|t_{\text{out}} \leq m+1\},$$

$$P_{lt} \geq 0, \quad \forall t = t_{\text{in}}, \ldots, t_{\text{out}}, \forall l \in \{l|t_{\text{out}} \leq m+1\}. \tag{15}$$

Because the upper and lower limits of the cluster charging power in the cluster charging mode will not exceed the sum of the upper and lower limits of all electric vehicles in the cluster, all charging strategies that meet the cluster charging mode must have at least one power allocation method so that the charging amount of the cluster at the start and the single charging mode can meet.

3. Situational Awareness of Power Supply Resources for Electric Vehicles

3.1. Situational Awareness Model of Electric Vehicle Power Supply Resources. In the process of grid power resource scheduling, it is necessary to extract the relevant factors of electric vehicle demand change. Then, the situation is understood, the obtained information is integrated and classified, and finally, the situation is predicted, and the development trend is predicted and judged. In this way, the charging demand and power supply resources of electric vehicles can be accurately grasped, and the corresponding situation classification rules can be formulated so that the power supply resources can be reasonably allocated according to the real-time load status of the station area, as shown in Table 1.

The process of situational awareness mainly includes three parts: (1) extracting the characteristic parameters of situational awareness, preprocessing, and providing data support for situational understanding and prediction; (2) situational understanding, integrating, and classifying the obtained information; and (3) making predictions and judgments about the situation.

Based on the charging characteristics of the electric vehicle, the characteristic parameters of the situational awareness model are extracted, including the current charging state of electric vehicle, the charging start time, the charging time, and the charging pile usage interval time. We quantify the indicators and then get the situational awareness value, so as to reasonably dispatch the station resources. According to the mined indicators, the original data are collected, and the features are extracted to generate sample sets. If the total number of samples is n and the number of feature indicators is m, the sample matrix will be an $m \times n$ matrix. Then, the training sample set is taken as the input, and the value corresponding to the sample matrix is taken as the training output, so as to train the integrated model.

3.2. Prediction of Electric Vehicle Load Situation Based on Random Forest Algorithm. In this section, a situation prediction model of electric vehicle power supply resources based on stochastic forest algorithm is proposed, and the
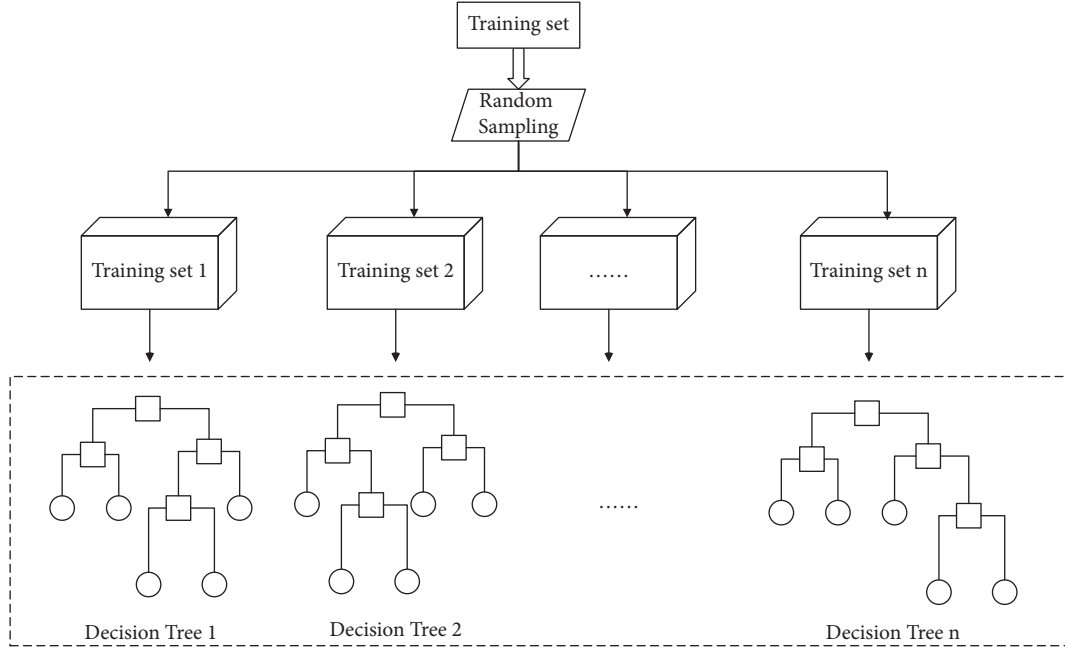
FIGURE 2: Random forest training process.

situation prediction of electric vehicle charging demand and regional power supply resources in the power supply station is made.

The decision tree of the random forest algorithm generally adopts the Classification and Regression Tree, which can effectively deal with large data samples and solve nonlinear problems. For classification problems, CART (Classification and Regression Tree) uses the Gini index as the attribute measure. The smaller the Gini coefficient is, the more accurate the classification effect will be. The Gini coefficient is defined as shown in

$$\text{Gini} = 1 - \sum_{i=1}^{c} [p(i|t)]^2, \tag{16}$$

where $p(i|t)$ is the probability that the test variable t belongs to the sample of class i, and c is the number of samples. When Gini = 0, all samples belong to the same class.

$$\text{Gini}_A(C) = \frac{|C_1|}{C} \text{Gini}(C_1) + \frac{|C_2|}{C} \text{Gini}(C_2). \tag{17}$$

If the attribute satisfies a certain purity, the decision tree generation algorithm divides the sample into the left subtree; otherwise, it divides the sample into the right subtree. The CART decision tree generation algorithm selects the split attribute rules according to the principle of the smallest Gini index. Assuming that the attribute A in the training set C divides C into C_1 and C_2, the Gini index of the given partition C is shown in

$$\text{Gini}_A(C) = \frac{|C_1|}{|C|} \text{Gini}(C_1) + \frac{|C_2|}{|C|} \text{Gini}(C_2). \tag{18}$$

For the regression problem, CART selects the optimal bisection cut point based on the sum of squared errors and
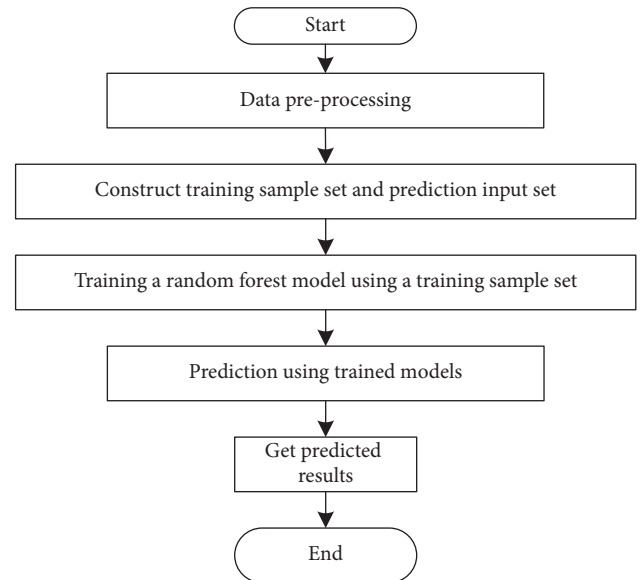

FIGURE 3: Forecasting process.

selects the optimal division attribute A and cut point c. The way to get it is as follows:

$$\min_{A,s} \left[\min \sum_{x_i \in R(A,s)} (y_i - c_1)^2 + \min \sum_{x_i \in R_2(A,s)} (y_i - c_2)^2 \right]. \tag{19}$$

As an integrated algorithm based on decision trees, in the process of building a random forest model, different training sets are constructed to train each decision tree, thus increasing the difference between the classifiers and making the classification performance of random forest algorithm

surpass that of a single decision tree algorithm. Figure 2 shows the random forest training process. To reflect the randomness of the random forest model, the construction of the training set includes the following two key processes.

The random forest algorithm performs random sampling with replacement on the original training dataset. We construct a subdataset, the sample size of which is consistent with the original dataset. Samples in different subdatasets can be repeated, and samples in the same subdataset can be repeated. We generate a decision tree for each subset of data.

The splitting process of each decision tree in the random forest model only uses some of all the candidate features. The random forest algorithm first randomly selects a certain number of features from all the features to be selected and then uses the decision tree generation algorithm to select the optimal feature for splitting among the randomly selected features.

The specific implementation steps of the random forest algorithm are as follows:

(a) Preprocess the data required for forecasting. And the missing data are filled in by linear interpolation. The training sample set is divided according to the needs of the prediction algorithm.

(b) Training the random forest algorithm by using the training sample set. According to the parameter settings of the random forest algorithm, several decision trees are built, and the prediction model of the random forest is completed.

(c) Use the characteristic data of the forecast samples to make the forecast. In the model, the characteristic data of the predicted samples will be followed by multiple decision tree prediction processes. The random forest algorithm will summarize and output the results to obtain the prediction results.

The flowchart of short-term electric vehicle charging load prediction based on random forest algorithm is shown in Figure 3.

At present, with the continuous improvement of electric vehicle data and other external data collection and storage level, there are many functions to choose from. Therefore, it is impossible to find all the features to participate in the fitting of the model, which may lead to problems such as overfitting of the model and reduce the accuracy of predictions. Therefore, considering the current research status of load forecasting, it is necessary to analyze the factors that affect the trip behavior of electric vehicles. This paper summarizes the following factors involved in the construction of feature engineering:

3.2.1. Historical Data of Electric Vehicle Charging Load.
The charging load of electric vehicles in the group also has the continuous characteristics of other conventional loads, and the historical load data closer to the predicted time can better reflect the load change trend. Of all the features, the historical data of load often have the greatest influence on the accuracy of the load model. The load history data of electric vehicle are selected as the input of the model, and the

built feature set determines the accuracy of the load forecasting model to a great extent.

3.2.2. Meteorological Factor Data.
Meteorological factor data are temperature, humidity, and weather conditions. The influence of temperature and humidity on the behavior of electric vehicles is mainly reflected in the use of onboard air conditioners of electric vehicles, which will increase the power consumption of vehicles. Therefore, the influence reflected in the EV charging load is a delay in the falling time of the charging curve, which takes longer time to completely charge the EV battery. The weather often affects road conditions and driving behavior of electric vehicles. Vehicles that tend to drive in bad weather also consume more energy.

3.2.3. Date Type.
The types of appointments are mainly divided into working days and weekends. In general, traffic congestion will be caused during peak hours on weekdays, which will affect the driving behavior of electric vehicles, increase the electricity consumption, and also affect the charging time of electric vehicles, thus affecting overall performance.

3.2.4. Time Type.
The charging load of group electric vehicle, like the conventional load, has certain periodicity. Therefore, the difference of load capacity in this type at a certain time of day will not be particularly great. This is mainly due to the fact that the charging characteristics of group electric vehicles show a certain regularity. Therefore, the type of torque can also be used as an optional input feature.

4. Study Case and Simulation Results

In this paper, the charging piles in a certain area are selected as the simulation object. We assume that the number of electric vehicles connected to the charging station in this area is 300 and that the connected vehicles are only pure electric vehicles. The electric vehicle parameters use the data in Table 2. We assume an EV charging efficiency of 99%. The charging pile is an AC charging pile, and the power selection is set to 3.5 kW and 7 kW; the two are randomly generated with equal probability.

First of all, it is necessary to divide training data and prediction data. This paper selects the historical electric vehicle charging load data, meteorological data, date type data, and time type data of Shenyang City from November 15, 2021, to November 21, 2021, to construct the training sample set, forecast the load from November 22, 2021, to November 28, 2021, and compare the forecast results with the actual values to verify the effectiveness of the method.

The model input includes the following: load data at the moment before the time to be predicted, load data at the previous two moments, and load data at the same moment last week, temperature, humidity, wind direction, wind speed, weather condition information, date type, and time type.

TABLE 2: Pure electric power electric vehicle parameter model.

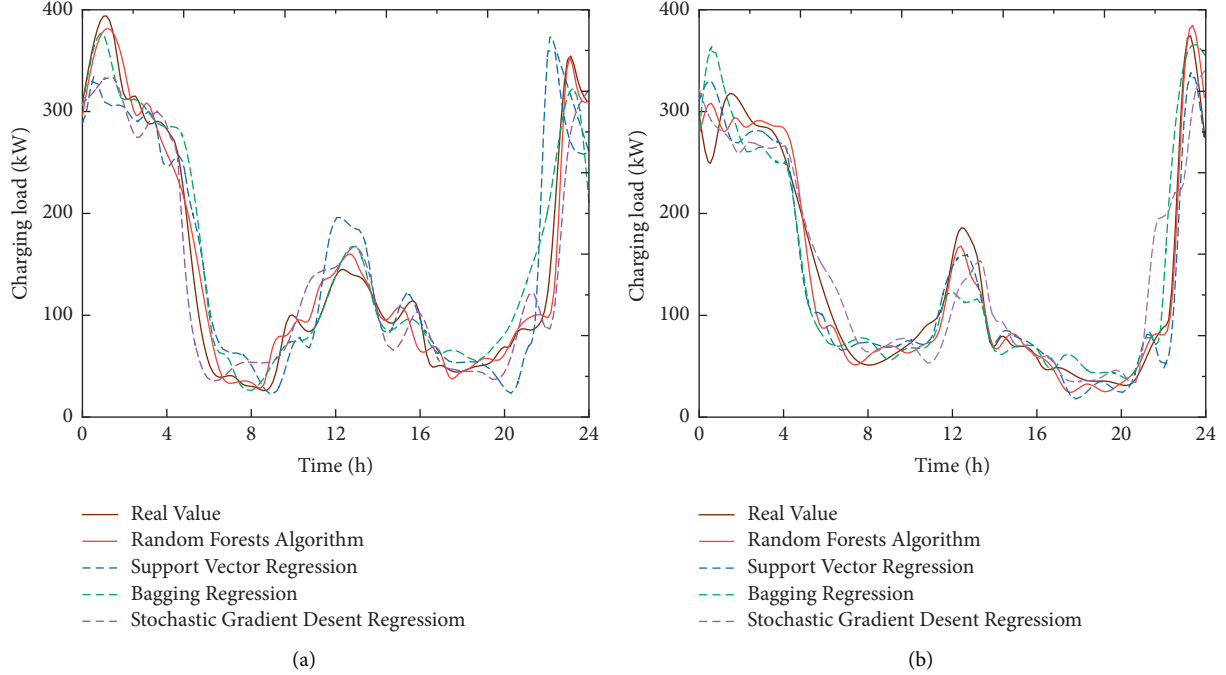| No.	Model name	Calibration mileage (km)	Battery capacity (kWh)	Power consumption per 100 kilometers (kWh/100 km)
1	BYD E9	506	64.8	15.8
2	Xiaopeng P5	460	55.9	14.6
3	WULING MINIEV	170	13.9	10.5
4	BJEV EC5	403	48	14.5
5	GAC AION S	410	58.8	13.9

FIGURE 4: Forecasts for different time types. (a) Weekdays. (b) Weekends.

The linear interpolation method is used to preprocess the numerical anomaly data of time series. It is assumed that a few continuous data points show a linear variation. It mainly adopts the mean value of the nonmissing time series data before and after as the repair value of the missing data. It is suitable for scenes with high data acquisition quality and few missing values.

There are different processing methods for missing time series data in different positions, mainly including beginning and end missing and the middle missing. For the missing data in the first and last numbers, the nonmissing value closest to the first and last numbers will be used to fill in. The calculation formula is as follows:

$$a_1 = a_2 = \cdots = a_s,$$
$$a_N = a_{N-1} = \cdots = a_e,$$
(20)

where a_s is the nonmissing value closest to the first position in the daily loading curve, and a_e is the nonmissing value closest to the last position. N is the data dimension of a single load data.

For the abnormal data of time series, this paper adopts the nearest-neighbor filling method. In the case of missing intermediate data, if single data are missing and the data of its front and rear load point are known, the corresponding data can be filled by line interpolation; that is, the average value of the load value before and after the time is used for filling. If there are missing points, the linear expressions are calculated for the front and back nonmissing data points, and all the missing data points are calculated proportionally. The calculation formula is as follows:

$$a_x = a_m + \frac{a_n - a_m}{n - m}(x - m),$$
(21)

where a_x is the missing data, and a_m and a_n are the nearest nonmissing values before and after consecutive missing data.

Figure 4 shows the prediction results based on the random forest algorithm on weekdays and weekends. The trend in the figure can also see the travel habits of electric vehicles in this area. After 8:00 p.m., it gradually entered the charging peak, reached its peak around 11:00 p.m., and maintained a high level at around 5:00 a.m. Some vehicles have been charged, and the load begins to drop. It reaches a valley value around 8:00 a.m. and then continues to climb, with a small peak appearing at around 12:00 noon and then dropping.

In order to verify the effectiveness and superiority of the random forest algorithm for electric vehicle load forecasting,

TABLE 3: Forecasting error of each algorithm.

	Random forests algorithm	Support vector regression	Bagging regression	Stochastic gradient descend regression
MAPE	18.78%	29.92%	22.54%	30.85%
MSE	25.37	33.32	30.56	46.17

the other three algorithms are selected for forecasting, and the forecasting effect was compared with that of the random forest algorithm. They are support vector regression machine, Bagging regression machine, and stochastic gradient descent regression.

The prediction performance of the algorithm was evaluated using mean absolute percentage error (MAPE) and mean square error (MSE). The calculation formulas of MAPE and MSE are shown in following equations:

$$\delta = \frac{1}{N} \sum_{1}^{N} \left| \frac{P_i' - P_i}{P_i} \right| \times 100\%, \quad (i = 1, 2, \ldots, N), \tag{22}$$

$$\varepsilon = \frac{1}{N} \sum_{i=1}^{N} \left(P_i' - P_i \right)^2, \quad (i = 1, 2, 3, \ldots, N), \tag{23}$$

where P_i' is the predicted load value at the time i predicted by the algorithm. P_i is the actual load value at the time i moment of the day. The smaller δ and ε are, the more accurate the prediction effect of the algorithm is.

Table 3 shows the statistics of load forecasting error of each forecasting algorithm. Among all experimental algorithms, the random forest algorithm has the smallest prediction error. It shows the effective load forecasting ability of electric vehicle. At the same time, horizontal comparison can be made, which shows that the method adopted in this paper can effectively improve the prediction accuracy of electric vehicles.

5. Conclusions

This paper analyzes the basic user behavior characteristics of electric vehicle load and establishes a single electric vehicle charging energy boundary model and a clustering classification model. A random forest algorithm-based electric vehicle load situation prediction method is proposed. The situational awareness results of the dispatch able resources in the power supply area are divided into sample sets and trained. Compared with other methods in the field of short-term load forecasting, the validity and superiority of the random forest algorithm for electric vehicle load forecasting are verified, and the charging performance and load forecasting accuracy of electric vehicles can be effectively improved. The following points can be explained by the simulation example.

(a) According to the usage habits of electric vehicle users, the travel model of electric vehicles is established, and then, the mathematical model of charging behavior of electric vehicles clusters is established, and its correctness is proved. Aiming at the adverse impact of electric vehicle disorderly access on power grid dispatching, and considering the changing characteristics of both the supply and demand sides, a situation awareness method is proposed to reasonably evaluate and forecast the power supply situation.

(b) From a data-driven point of view, a method of electric vehicle load forecasting based on random forest algorithm is proposed, and the training set is constructed, which improves the accuracy of electric vehicle charging load forecasting. Moreover, the actual bearing capacity of the power supply area can also be taken into account while meeting the needs of users to the maximum extent.

Data Availability

The data presented in this study are available on request from the corresponding author. The data are not publicly available due to the data use agreement.

Conflicts of Interest

The authors declared that they have no conflicts of interest.

Acknowledgments

References

[1] S. R. Sinsel, R. L. Riemke, and V. H. Hoffmann, "Challenges and solution technologies for the integration of variable renewable energy sources-a review," *Renewable Energy*, vol. 145, pp. 2271–2285, 2020.

[2] H.-I. Su and A. El Gamal, "Modeling and analysis of the role of energy storage for renewable integration: power balancing," *IEEE Transactions on Power Systems*, vol. 28, pp. 4109–4117, 2013.

[3] L. Cozzi, T. Gould, S. Bouckaert et al., *World Energy Outlook 2020*, International Energy Agency, Paris, Frace, 2020.

[4] S. Kim and J. Hur, "A probabilistic modeling based on Monte Carlo simulation of wind powered EV charging stations for steady-states security analysis," *Energies*, vol. 13, 2020.

[5] M. P. Anand, B. Bagen, and A. Rajapakse, "Probabilistic reliability evaluation of distribution systems considering the spatial and temporal distribution of electric vehicles," *International Journal of Electrical Power & Energy Systems*, vol. 117, 2020.

[6] M. Aziz, T. Oda, T. Mitani, Y. Watanabe, and T. Kashiwagi, "Utilization of electric vehicles and their used batteries for peak-load shifting," *Energies*, vol. 8, pp. 3720–3738, 2015.

[7] Y. Wang and D. Infield, "Markov chain Monte Carlo simulation of electric vehicle use for network integration studies," *International Journal of Electrical Power & Energy Systems*, vol. 99, pp. 85–94, 2018.

[8] Q. Dai, S. Duan, T. Cai, T. Chen, and L. Huang, "Cost-benefit model and its sensitivity analysis for battery charging and swapping station for electric vehicles," *Automation of Electric Power Systems*, vol. 38, pp. 41–47, 2014.

[9] A. Colmenar-Santos, A. M. Muñoz-Gómez, E. Rosales-Asensio, and A. Lopez-Rey, "Electric vehicle charging strategy to support renewable energy sources in Europe 2050 low-carbon scenario," *Energy*, vol. 183, pp. 61–74, 2019.

[10] R. Wang, Q. Sun, D. Ma, and Z. Liu, "The small-signal stability analysis of the droop-controlled converter in electromagnetic timescale," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 3, pp. 1459–1469, 2019.

[11] S. Deilami, A. S. Masoum, P. S. Moses, and M. A. S. Masoum, "Real-time coordination of plug-in electric vehicle charging in smart grids to minimize power losses and improve voltage profile," *IEEE Transactions on Smart Grid*, vol. 2, pp. 456–467, 2011.

[12] K. Valentine, W. G. Temple, and K. M. Zhang, "Intelligent electric vehicle charging: rethinking the valley-fill," *Journal of Power Sources*, vol. 196, pp. 10717–10726, 2011.

[13] R. Wang, Q. Sun, W. Hu, Y. Li, D. Ma, and P. Wang, "SoC-based droop coefficients stability region analysis of the battery for stand-alone supply systems with constant power loads," *IEEE Transactions on Power Electronics*, vol. 36, no. 7, pp. 7866–7879, 2021.

[14] K. Wan and R. Liu, "Application of interval time-series vector autoregressive model in short-term load forecasting," *Power System Technology*, vol. 36, pp. 77–81, 2012.

[15] H. Wang, J. Zhou, and K. Liu, "Summary of research on the short-term load forecasting method of the electric power system," *Electrical Automation*, vol. 37, pp. 1–3, 2015.

[16] R. Wang, Q. Sun, P. Tu, J. Xiao, Y. Gui, and P. Wang, "Reduced-order aggregate model for large-scale converters with inhomogeneous initial conditions in DC microgrids," *IEEE Transactions on Energy Conversion*, vol. 36, no. 3, pp. 2473–2484, 2021.

[17] N. Liao, Z. Hu, and Y. Ma, "Review of the short-term load forecasting methods of electric power system," *Power System Protection and Control*, vol. 39, pp. 147–152, 2011.

[18] X. Huang, J. Chen, and Y. Chen, "Load forecasting method for electric vehicle charging station based on big data," *Automation of Electric Power Systems*, vol. 40, pp. 68–74, 2016.

[19] W. Rui, S. Qiuye, Z. Pinjia, G. Yonghao, Q. Dehao, and W. Peng, "Reduced-order transfer function model of the droop-controlled inverter via Jordan continued-fraction expansion," *IEEE Transactions on Energy Conversion*, vol. 35, no. 3, pp. 1585–1595, 2020.

[20] X. Zhou, M. Farivar, Z. Liu, L. Chen, and S. H. Low, "Reverse and forward engineering of local voltage control in distribution networks," *IEEE Transactions on Automatic Control*, vol. 66, pp. 1116–1128, 2021.

[21] N. Eghtedarpour and E. Farjah, "Power control and management in a hybrid AC/DC microgrid," *IEEE Transactions on Smart Grid*, vol. 5, pp. 1494–1505, 2014.

[22] W. Huang, Q. Liu, and S. Yang, "Security situation awareness based on power-supply ability model of active distribution system," *Electric Power Automation Equipment*, vol. 37, pp. 74–80, 2017.

[23] H. Su, G. Wang, and L. Liu, "Ordered charging of electric vehicles based on time-space complementary," *Electrical Measurement & Instrumentation*, vol. 54, pp. 99–104, 2017.