

Prediction of Electric Vehicle Energy Consumption by Combining Real Vehicle Data and Machine Learning Methods

Zhengqian Wu¹, Xiaobing Chen^{2*} and Yugang Jiang³

¹Hunan Mechanical & Electrical Polytechnic, Hunan Changsha, China

²Hunan Province Motor Vehicle Technician College, Hunan Shaoyang, China

³Hunan Biological and Electromechanical Polytechnic, Hunan Changsha, China

***Corresponding Author:** Xiaobing Chen, Hunan Province Motor Vehicle Technician College, Hunan Shaoyang, China.

Received: February 01, 2023; **Published:** February 22, 2023

DOI: 10.55162/MCET.04.115

Abstract

A highly nonlinear relationship exists between complex driving conditions, external influencing factors and vehicle energy consumption. Considering the spatiotemporal characteristics of vehicle operation, the significant feature parameters are extracted to improve the accuracy of vehicle energy consumption prediction. In this paper, an electric vehicle energy consumption prediction method that integrates real vehicle operation data and machine learning methods is proposed. Based on a large amount of real vehicle operation data, data cleaning and data integration methods are used to divide different kinematic segments. The key features of vehicle operation are extracted from the kinematic segment, and the correlation coefficient analysis is used to screen important feature values. Based on the XGBoost algorithm, the vehicle termination SOC prediction model is established to further obtain the results of the vehicle energy consumption. Through real vehicle operation data verification, the energy consumption prediction error within 0.04kWh, the results indicate that the proposed method gives out high accuracy.

Keywords: electric vehicle; energy consumption analysis; XGBoost algorithm; kinematics segment

Introduction

Climate change is the most significant non-traditional challenge to human development today, and most countries have reached agreements to reduce carbon emissions to address it. As a result, many investigations aim to peak carbon emissions and achieve carbon neutrality [1]. The transport sector is the third largest source of carbon emissions after the energy industry and the construction sector, so it is imperative that low-carbon transport policies be fully implemented [2]. In the transportation sector, more than 80% of carbon emissions come from vehicles, therefore, the electrification of transport is an effective path to reduce energy consumption and carbon emissions [3]. New energy vehicles play a key role in improving the development level of electrified transportation, Lorf C et al. [4] analyzed the energy consumption and emissions of 40 vehicles with different power sources, and the results reveal that the pure electric system equipped with regenerative braking system has the lowest emissions and energy consumption. Therefore, actively developing energy-saving and emission-reducing clean fuel vehicles is an effective way for China to reduce carbon emissions and address climate change.

Compared with traditional vehicles, electric vehicles have the advantages of high energy conversion efficiency and low power system cost, but their disadvantage is short driving range. At present, there are problems such as imperfect charging facilities and long charging time that lead to “range anxiety” among drivers. Therefore, a lot of research on energy consumption prediction of pure electric vehicles has been carried out. Based on the traditional method of calculating the future average energy consumption, Hao. [5] proposed an average energy consumption prediction method based on the support vector machine regression theory. Li. [6] mainly analyzed the influence of temperature, wind speed and vehicle type on the energy consumption of air conditioning, and then studied the influence on the energy consumption of electric vehicles. Based on the average speed predicted by LSTM, Cheng et al. [7] built an energy consumption model of pure electric vehicle considering air conditioning energy consumption, which indicated great robustness in the analysis of traffic network. Song. [8] mainly analyzed the influence of vehicle driving parameters on energy consumption based on different driving conditions. Lin et al. [9] proposed to use battery energy state (SOE) estimation to improve the driving range prediction accuracy of electric vehicles. Hong J et al. [10] proposed a hybrid modeling method combining physical equation based model and empirical data to improve the accuracy of energy consumption prediction model. Through the results of energy consumption prediction, drivers can make reasonable travel routes, solve “range anxiety” and promote the development and use of electric vehicles.

In recent years, benefiting from the rapid development of the Internet of Vehicles and intelligent transportation systems, real-time collection of vehicle operation data, environment and traffic information has become easy to achieve. Data-driven models do not have many parameters (e.g., automotive parameters and aerodynamic drag coefficients) and assumptions (e.g., distribution of parameters). Based on the actual driving data of electric vehicles, Bi et al. [11] proposed a robust nonlinear regression model, and discussed the relationship between speed and unit SOC driving distance. Jie Hu et al. [12] adopted machine learning algorithm to predict the SOC of electric vehicles by fusing macro and micro data and considering the impact of temperature on electric vehicles. Jun Bi et al. [13] explored the correlation between electric vehicle driving range and battery SOC through a data-driven way, so as to establish an electric vehicle driving range prediction model. Tan. [14] designed and studied the energy consumption prediction method based on machine learning algorithm by collecting the driving energy consumption data of electric vehicles. Yang et al. [15] proposed a real-time energy consumption prediction method for electric vehicles based on the GRU-NN model using the monitoring data of electric vehicles.

In summary, when using the vehicle model-driven approach, many vehicle parameters need to be obtained or assumed. However, in real driving conditions, it is difficult to obtain these vehicle parameters, especially when applied to a large number of models and large fleets. Based on a large amount of real vehicle operation data, considering the coupling effect of complex driving conditions and multiple influencing factors, this paper adopts machine learning to identify patterns and relationships by iterating on a large amount of data, and solves the existing highly nonlinear and complex spatiotemporal characteristics. Firstly, based on the real vehicle operation data, the kinematic fragment is divided into kinematic fragments of the vehicle operating state, the important feature values are extracted from the kinematic fragments. Then, the SOC of electric vehicles is predicted by XGBoost algorithm, so as to establish an electric vehicle energy consumption prediction model.

Data preprocessing

In this study, the data acquisition device was installed on 10 pure electric vehicles, and the experimental vehicles consisted of 7 taxis and 3 private cars. The data collection device is capable of acquiring real-time driving data of pure electric vehicles with a sampling time of 20Hz, and the collected vehicle data needs to be pre-processed, including data conversion and clean before analysis. The preliminary data processing process is shown in Figure 1. “Vehicle status” includes whether the vehicle is grid charged (2), in driving (1), or parked (0), and the primary processed data format is listed in Table 1.

Time	Speed	Current	Voltage	vehicle status	SOC
10:45:49.360	10.1	-0.4	363.2	1	71
10:45:49.410	10.1	-0.4	363.2	1	71
10:45:49.460	9.9	0.3	363	1	71
Λ	Λ	Λ	Λ	Λ	Λ
10:46:01.110	19.3	5.6	362.4	1	70.9

Table 1: Processed data formats.

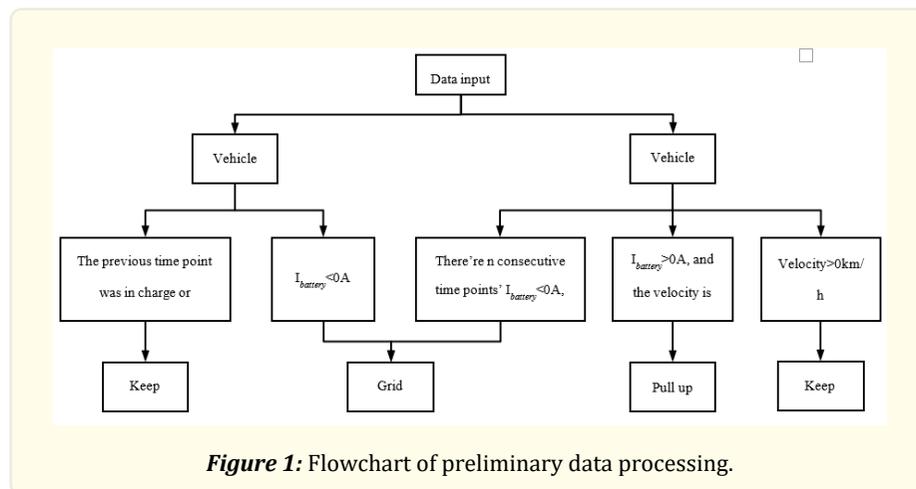


Figure 1: Flowchart of preliminary data processing.

Due to the fact that the tested vehicle encounters bad weather, the data collection terminal is in the signal blind zone or fails, the collected data will have problems such as positioning drift and partial loss, resulting in errors and data redundancy in the location information of the electric vehicle obtained by the data collection device. To ensure the accuracy of the driving data collected by the experiment, the original data needs to be processed, including: (1) For the driving data that is stopped for a long time, it may be due to the charging or parking of the experimental vehicle, at this time the data collection terminal is still collecting data, such data is not a reflection of the real driving data of Tianjin city, and the collected data with a parking time of more than 180s needs to be removed. (2) Standardize and unify the data format and adjust it chronologically to correct the data order changes that may occur during data transmission. (3) Since the start of data collection terminal may be later than the start of the electric vehicle, the data of this discontinuous part needs to be removed. (4) Due to the influence of traffic environments and signals, the data collected by the data acquisition equipment will have unrealistic data points (such as obvious shifts in latitude and longitude or excessive speed, etc.) and need to be deleted. (5) In a congested environment, the vehicle is driving slowly, and it is necessary to remove the data sample under the non-free flow section to avoid data deviation.

Kinematic fragments are defined as velocity vs. time curves from the start of one idle state to the start of the next idle state [16]. Electric vehicle trip data is composed of multiple kinematic fragments, which are units of trip data that study the energy consumption. This is shown in Figure 2.

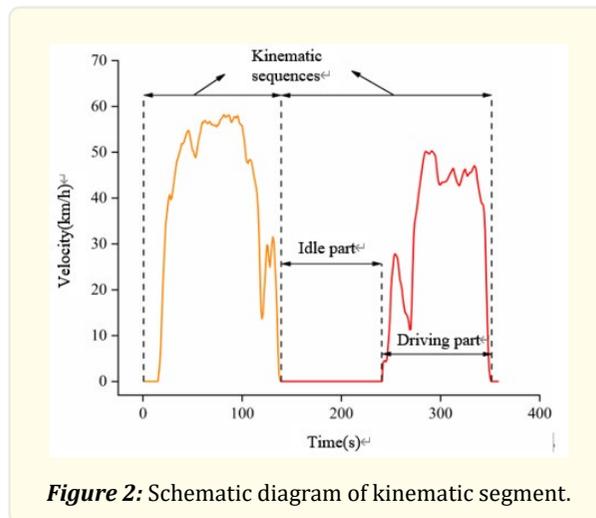


Figure 2: Schematic diagram of kinematic segment.

Data analysis

Kinematic fragment analysis

After data preprocessing, a total of 16052 effective kinematic fragments are obtained, and the following figure shows the travel time and distance information of the kinematic segments. It can be seen from Figure 3(a) that each part of the trip is within 10 minutes, the longest single stroke is 45 minutes, and Figure 3(b) shows that most of the travel distance is within 10 km.

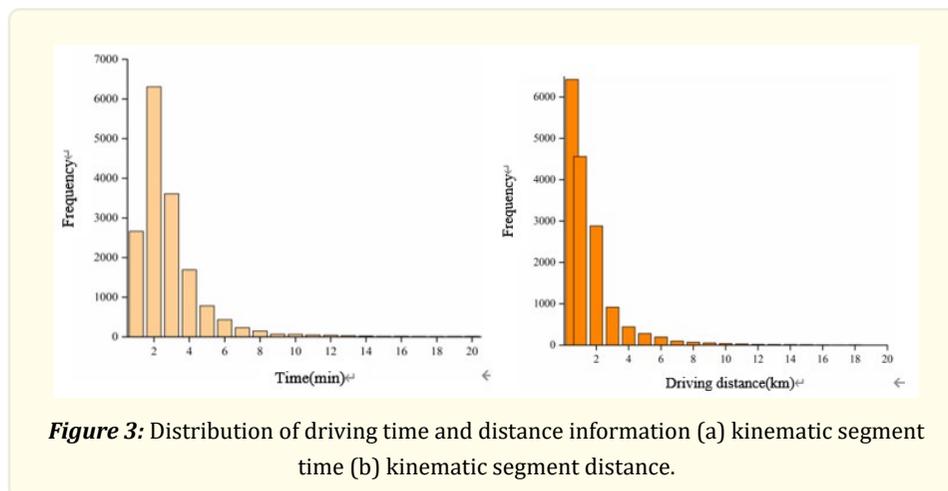


Figure 3: Distribution of driving time and distance information (a) kinematic segment time (b) kinematic segment distance.

To analyze and study the influencing factors of electric vehicle energy consumption, the unit energy consumption SEC is introduced, which is measured in kWh/km, and the energy consumption during the trip is calculated using the battery voltage and current [17]. Use the following equation to determine:

$$E_{trip} = \frac{1}{3600} \times \sum_{i=1}^n V_i \times \frac{I_i}{1000} \quad (1)$$

E_{trip} is the stroke energy consumption, kWh; V_i is the battery voltage V; I_i is the battery current measured at each time step, A; i is the time step; n is the total number of times.

Therefore, the unit energy consumption of the trip SEC can be calculated as follows, where d_{trip} is the travel distance in km:

$$SEC = \frac{E_{trip}}{d_{trip}} \quad (2)$$

$$d_{max} = \frac{C_{battery}}{SEC} \quad (3)$$

According to equations (2) and (3) above, it can be seen that the vehicle driving distance d_{trip} is closely related to battery capacity and terminal voltage, in which battery usage capacity can be analyzed through SOC changes. Therefore, in order to accurately predict electric vehicle energy consumption, it is necessary to realize accurate estimation of destination termination SOC based on driving data.

Calculation of Equivalent energy consumption rate

Considering that the running energy consumption of electric vehicles is not only related to the initial SOC and the terminal SOC, but also related to the battery terminal voltage of this trip, the calculation method of energy consumption rate is proposed in this paper, which is used as a reference for the energy consumption analysis of electric vehicles.

$$ECR = \frac{SOH \times C \times \int_{SOC_0}^{SOC_1} V \times d_{SOC}}{VS} \quad (4)$$

In the above equation, SOH represents the health state of the automobile power battery, and is the driving distance in a single trip.

To verify the effectiveness of this method, real vehicle data is used, as shown in Figure 4 (a). The relationship between SOC and terminal voltage of different trips is assumed to be linear fitting, and the relationship between equivalent energy consumption rate and real energy consumption rate can be obtained, as shown in Figure 4 (b).

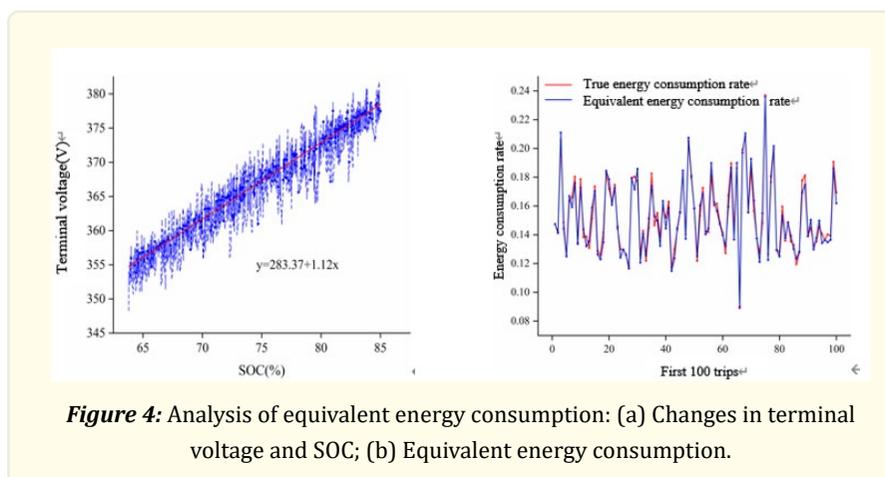


Figure 4: Analysis of equivalent energy consumption: (a) Changes in terminal voltage and SOC; (b) Equivalent energy consumption.

For further study the relationship between voltage and SOC, SOC is more detailed divided into 20 different intervals according to the 5% SOC scale. Linear fitting is performed in each interval to determine the corresponding slope and intercept, the concrete numerical value as shown in Table 2, the corresponding terminal voltage and SOC piecewise curve, as shown in Figure 5. Figure 5 (a) shows splicing curves of terminal voltage and SOC of 20 intervals. In addition, polynomial fitting of splicing curves is shown in Figure 5 (b).

<i>Period of SOC</i>	<i>Slope</i>	<i>Intercept</i>
95%~100%	1.032	291.386
90%~95%	1.223	273.001
85%~90%	1.033	290.096
80%~85%	0.962	296.115
75%~80%	1.003	292.762
70%~75%	1.053	289.134
65%~60%	1.050	289.074
60%~65%	0.787	306.040
55%~60%	0.584	318.197
50%~55%	0.462	324.851
45%~50%	0.352	330.425
40%~45%	0.245	335.169
35%~40%	0.266	334.030
30%~35%	0.524	325.333

Table 2: Piecewise fitting parameters of terminal voltage SOC.

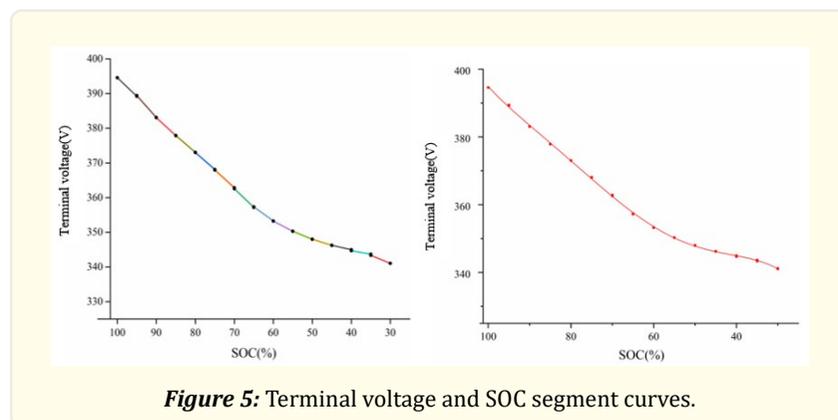


Figure 5: Terminal voltage and SOC segment curves.

SOC prediction feature value extraction

20 eigenvalues are selected in this paper for SOC prediction at the end of the trip and the eigenvalues are shown in Table 3. These feature parameters can be divided into four categories: vehicle speed feature parameters, variable speed feature parameters, driving state feature parameters, stroke feature parameters and battery state parameters. Vehicle speed characteristic parameters reflect vehicle speed, such as average speed, maximum speed, average driving speed; Variable speed characteristic parameters reflect the change of vehicle speed, such as average plus or minus speed, standard deviation of acceleration, standard deviation of speed, 95% quantile of acceleration, 5% quantile of deceleration; Driving state characteristic parameters reflect the proportion of various driving states in the whole trip, such as the proportion of five driving states; Trip characteristic parameters reflect the overall situation of the trip, such as driving distance and driving time; Battery status parameter indicates the battery energy consumption.

<i>Symbol</i>	<i>Defining</i>	<i>Unit</i>
V_a	The average velocity	km/h
V_m	Maximum speed	km/h
L	mileage	m
L_o	Cumulative miles traveled	km
P_s	Percentage of start time	%
P_a	Percentage of acceleration time	%
P_d	Percentage of deceleration time	%
P_c	Percentage of time at constant speed	%
P_i	Percentage of parking time	%
D_a	Average deceleration rate	m/s ²
A_m	Acceleration on average	m/s ²
$V_{a(r)}$	Travel speed and only non-zero speed	km/h
A_s	Standard deviation of acceleration	m/s ²
V_s	Standard deviation of velocity	km/h
$A_{0.95}$	The value of the acceleration at 95%	m/s ²
$A_{0.05}$	The value of the deceleration at 5%	m/s ²
T	Travel time	s
SOC_b	Before moving the SOC	%
V_b	The total voltage	V
I	The total current	I

Table 3: Eigenvalues of SOC prediction.

XGBoost energy consumption prediction

XGBoost Algorithm

XGBoost is a boosting algorithm in ensemble learning methods. The basic idea of the algorithm is to continuously optimize and fit the deviation of the previous model with the new base model, so as to gradually reduce the overall model deviation. Compared to the decision tree algorithm (GBDT), XGBoost first regularization of objective function is simplified, which prevents fitting problems, the other by the missing value samples into left and right subtrees contrast method, reduces the pre-processor of lack of characteristic value and adopts the automatic defect eigenvalue processing strategy which improves the model generalization ability and performance.

XGBoost model is an additive model composed of k decision tree algorithms, and its objective function is composed of training loss function and regularization. Assuming that the dataset $D = \{(x_i, y_i)\}$ has n samples and m features, the specific objective function can be expressed as follows:

$$L(\theta) = \sum_l l(y_i, \hat{y}_i) + \sum_k \Omega(f_k) \quad (5)$$

In the above equation, l is the loss function used to measure the difference between the predicted value and the target value, and $\Omega(f_k)$ is the regularization term. The prediction function and regularization function are described as follows.

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), \quad f_k \in F \quad (6)$$

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (7)$$

\hat{y}_i represents the predicted value of the i th sample. Each tree function f_k corresponds to a structure q and leaf weight w , there are T leaves in total, and K is the number of trees. For ease of calculation, the loss function is approximated as the following function using Taylor expansion.

$$L \square \sum_{i=1}^n \left[l(y_i, \hat{y}_i^{-1}) + g_i f_i(x_i) + \frac{1}{2} h_i f_i^2(x_i) \right] + \Omega(f_i) + C \quad (8)$$

In conclusion, $f(x)$ is a tree function in function space $F = \{f(x) = wq(x)\}$ and the loss function at iteration t is,

$$L \square \sum_{i=1}^n \left[g_i f_i(x_i) + \frac{1}{2} h_i f_i^2(x_i) \right] + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (9)$$

$$L \square \sum_{j=1}^T \left[\left(\sum_{i \in I_j} g_i \right) w_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i + \lambda \right) w_j^2 \right] + \gamma T \quad (10)$$

The set of instances of leaf j is defined as $I_j = \{i \mid q(x_i) = j\}$, and we can obtain:

$$L \square \sum_{j=1}^T \left[\left(\sum_{i \in I_j} g_i \right) w_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i + \lambda \right) w_j^2 \right] + \gamma T \quad (11)$$

When we define $G_j = \sum_{i \in I_j} g_i$, $H_j = \sum_{i \in I_j} h_i$, we have the following simplification

$$L \square \sum_{j=1}^T \left[G_j w_j + \frac{1}{2} (H_j + \lambda) w_j^2 \right] + \gamma T \quad (12)$$

When the structure of the classification and regression tree has been determined, the optimal score of the optimal weight w_j^* of leaf j and the corresponding optimal value of the objective function can be obtained by solving the above equation:

$$w_j^* = -\frac{G_j}{H_j + \lambda} \quad (13)$$

$$Obj = -\frac{1}{2} \sum_{j=1}^T \frac{G_j^2}{H_j + \lambda} + \gamma T \quad (14)$$

As can be seen from Equation (14), when the value of the objective function is smaller, the classification and regression tree model is better. XGBoost algorithm selects the segmentation point with the smallest objective function value by traversing all the segmentation points. The payoff after segmentation will be maximized when selecting the best segmentation point.

Energy consumption prediction

In the process of machine learning modeling, the multi-dimensional input feature parameters will increase the computational complexity. Therefore, in order to reduce the computational burden and improve the robustness of the model, the correlation analysis of the proposed 20 driving features can be carried out, and the training data dimension can be determined according to the correlation. In this paper, Pearson correlation coefficient (rp) is used to screen important driving feature parameters and select feature parameters

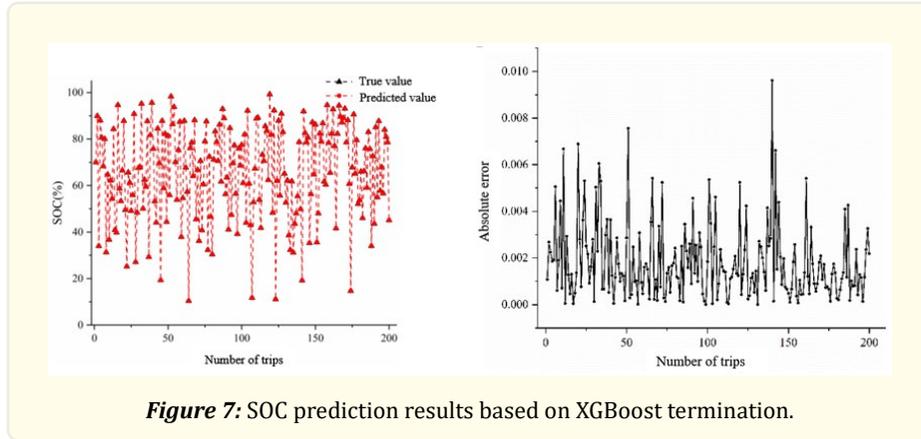


Figure 7: SOC prediction results based on XGBoost termination.

For quantitative analysis of the electric car energy consumption prediction error, using four kinds of error evaluation method, respectively, the root means square error, mean absolute error decision coefficient, mean absolute percentage error calculation method for the following type (17) to (20), as shown in SOC prediction model based on XGBoost algorithm error results as shown in table 3:

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \bar{y}_i^*)^2}{\sum_{i=1}^N (y_i - \bar{y}_i)^2} \quad (17)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \bar{y}_i^*)^2} \quad (18)$$

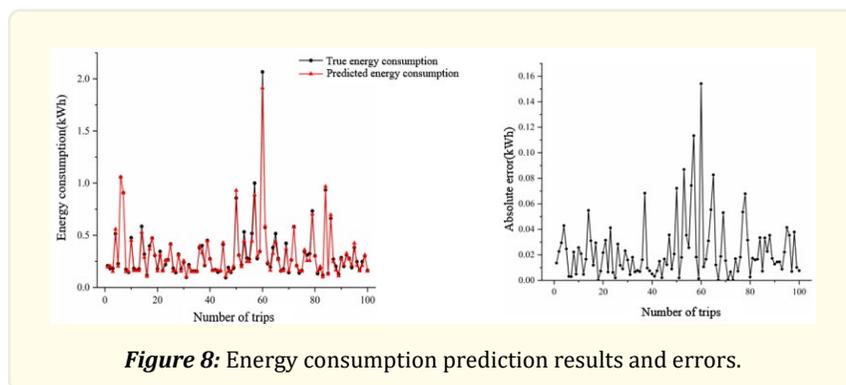
$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \bar{y}_i^*| \quad (19)$$

$$MAPE = \frac{100\%}{n} \sum_{i=1}^N \left| \frac{\bar{y}_i^* - y_i}{y_i} \right| \quad (20)$$

<i>R</i> ²	<i>RMSE</i>	<i>MAE</i>	<i>MAPE</i>
0.999	0.016	0.0029	0.0045

Table 4: Error result analysis.

According to the above verification of the real vehicle data, it can be seen from the error analysis that the average error of terminal SOC prediction is 0.0029, and the root mean square error is 0.016, which are listed in Table 4. The results show that using XGBoost algorithm combined with eigenvalue screening for terminal SOC prediction can obtain high precision terminal SOC prediction results. According to the SOC at the initial time of the trip and the calculated SOC, the energy consumption of the trip can be calculated. The energy consumption prediction result and its absolute error of the last 100 trips are analyzed, as shown in Figure 8. It can be seen from the figure that the maximum error of energy consumption prediction is 0.16kWh, and most of the prediction errors are within 0.04kWh. The proposed method based on the final SOC predicted energy consumption analysis has high prediction accuracy.



Summarizes

Using the real driving data collected from electric vehicles in a certain region, this study constructed an energy consumption prediction model for electric vehicles. The proposed energy consumption prediction method only needs the initial state information at the departure time of the trip, and the training and prediction based on XGBoost model requires relatively short time, which can provide more accurate energy consumption prediction results. The energy consumption calculation method proposed in this paper can calculate the SOC value at the end of the trip, and adopt the relationship between terminal voltage and SOC to calculate the energy consumption directly by using the trapezoidal integral formula. In addition, this paper analyzes the input of the model combined with feature screening, and the whole model has high accuracy and strong interpretability.

Reference

1. Liu Z., et al. "Challenges and opportunities for carbon neutrality in China". *Nature Reviews Earth & Environment* 3.2 (2022): 141-155.
2. Jun Wu and Yunpeng Cai. "Thinking on Countermeasures of transportation transformation and development under the goal of "carbon peak and carbon neutrality". *Transportation Energy Conservation and Environmental Protection* 17.05 (2021): 33-36. (In Chinese)
3. Bellocchi S., et al. "On the role of electric vehicles towards low-carbon energy systems: Italy and Germany in comparison". *Applied energy* 255 (2019): 113848.
4. Lorf C., et al. "Comparative analysis of the energy consumption and CO2 emissions of 40 electric, plug-in hybrid electric, hybrid electric and internal combustion engine vehicles". *Transportation Research Part D Transport & Environment* 23 (2013): 12-19.
5. Hao XB. "Driving range estimation of electric vehicles based on SOC state and future energy consumption prediction". Jilin University (2016). (In Chinese)
6. Li Weiwei. "Modeling and Energy Consumption Ratio Analysis of Pure Electric Vehicle Air Conditioning System". Jilin University (2020). (In Chinese)
7. Cheng Jiangzhou., et al. "Electric vehicle Energy consumption prediction considering Multi-influence in urban road Network". *Electric Measurement & Instrumentation* 57.20 (2020): 8. (In Chinese)
8. Song YY. "Research on Energy Consumption Modeling and Driving range Estimation of pure electric vehicles based on driving conditions". Beijing Jiaotong University (2014). (In Chinese)
9. Shili Lin., et al. "Prediction of Driving Range of electric Vehicles based on Battery SOE". *Battery* 47.03 (2017): 137-139. (In Chinese)
10. Hong J, Park S and Chang N. "Accurate remaining range estimation for Electric vehicles". 2016 21st Asia and South Pacific Design Automation Conference (ASP-DAC). IEEE (2016).
11. Bi J., et al. "Estimating remaining driving range of battery electric vehicles based on real-world data: A case study of Beijing,

- China". Energy 169 (2019): 833-843.
12. Hu Jie and Gao Zhiwen. "Data-driven SOC Prediction of Electric Vehicle Power Battery". Automotive Engineering 43.01 (2021): 1-9. (In Chinese)
 13. Bi Jun., et al. "Correlation Analysis and Modeling of Electric Vehicle Driving Range and Battery SOC". Transportation Systems Engineering and Information 15.01 (2015): 49-54. (In Chinese)
 14. [Tan Shanmao. "Energy Consumption Prediction Method for Pure Electric Vehicles Based on Machine Learning Algorithm". Automotive Test Report 8 (2022): 60-62. (In Chinese)
 15. Yang Xiaodong and Lu Yelin, Permission. "Real-time Energy Consumption Prediction Method for Electric Vehicles based on GRU-NN Model". Transportation Energy Conservation and Environmental Protection 18.4 (2022): 59-65. (In Chinese)
 16. Jiachen Guo., et al. "Construction Method of Driving Conditions for Urban Road Vehicles". Journal of Traffic and Transportation Engineering 20.06 (2020): 197-209. (In Chinese)
 17. Guangming Liu., et al. "Research on Electric Vehicle Driving Range Estimation Method Based on Battery Energy State Estimation and Vehicle Energy Consumption Prediction". Automotive Engineering 36.11 (2014): 1302-1309. (In Chinese)

Volume 4 Issue 3 March 2023

© All rights are reserved by Xiaobing Chen., et al.