

Received August 21, 2020, accepted August 28, 2020, date of publication September 4, 2020, date of current version September 21, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3021745

# State-of-Charge Prediction of Battery Management System Based on Principal Component Analysis and Improved Support Vector Machine for Regression

LIANG XUAN<sup>1,2</sup>, LIJUN QIAN<sup>1</sup>, JIAN CHEN<sup>1</sup>, XIANXU BAI<sup>1</sup>, (Member, IEEE), AND BING WU<sup>1</sup>

<sup>1</sup>School of Automotive and Traffic Engineering, Hefei University of Technology, Hefei 230009, China

<sup>2</sup>School of Mechanical and Automotive Engineering, West Anhui University, Lu'an 237012, China

Corresponding author: Lijun Qian (qianlijun66@163.com)

This work was supported in part by the National Natural Science Foundation, China, under Grant 51875149, and in part by the Key Project of West Anhui University, China, under Grant WXZR201712.

**ABSTRACT** State-of-charge (SOC) prediction is an important part of the battery management system (BMS) in electric vehicles. Since external factors (voltage, current, temperature, arrangement of the battery, etc.) impact SOC prediction differently, the SOC is difficult to model. In this article, we apply principal component analysis (PCA) to analyze the contribution of various external factors and propose a new SOC prediction method based on an improved support vector machine for regression (SVR) with data classification and training set size optimization. Three groups of simulation experiments with different inputs based on the original SVR algorithm are conducted in the software ADVISOR, and the simulation results show that the input of three features of the battery (current, voltage and temperature) can satisfy the SOC prediction accuracy. The improved SVR algorithm is then applied to the simulation experiment of the three input features. The proposed method is demonstrated to be faster and more accurate than the original SVR algorithm through a comparison of the simulation results.

**INDEX TERMS** State-of-charge (SOC), principal component analysis (PCA), support vector machine for regression (SVR), battery management system (BMS), electric vehicles.

## I. INTRODUCTION

With the increase in global warming and pollution, countries have attached great importance to researching new energy vehicles during the past decade [1]. Electric vehicles as a new energy vehicle have also been vigorously promoted. Batteries, as one of the key components of electric vehicles, have been extensively studied [2]–[4]. Currently, research on batteries can be divided into two main categories. One is to establish accurate electrochemical models by studying the charging and discharging mechanism of batteries [5], [6], and the other is to develop a battery management system (BMS) to adapt batteries to different working environments and improve the service performance of batteries. However, regardless of the electrochemical model adopted, the application of the BMS is essential.

The associate editor coordinating the review of this manuscript and approving it for publication was Chao Yang .

State-of-charge (SOC) prediction is an important function of the BMS. The BMS prevents the overcharge and overdischarge of batteries by accurately estimating the SOC of batteries, reduces the misuse of batteries, prolongs the service life of batteries and reduces the cost of batteries [7]. At the same time, the accurate prediction of the SOC can realize the accurate estimation of the remaining driving mileage of electric vehicles, which can effectively alleviate driving anxiety of drivers. Researchers have made great efforts to improve the SOC prediction of electric vehicles.

The open circuit voltage (OCV) method is one of the commonly used methods for SOC prediction. Scientists have used experiments to find the mapping relationship between the battery OCV and SOC and have produced the OCV-SOC table [8]–[10]. The application of the OCV method is not complicated, and the real-time prediction of the battery SOC is realized. However, there is a major defect in the application of the OCV method in electric vehicles; that is, the exact

value of the OCV of batteries can be obtained only when the batteries are stationary for a long time, but there is no such condition during the operation of the vehicles. At the same time, the ideal mapping relationship between the OCV and SOC will also be affected by external interference factors, such as temperature. The Coulomb counting method is a common method for accumulating electricity. This method predicts the SOC value of the battery by integrating the currents of the battery charge and discharge and compensates for the SOC according to the temperature and discharge rate of the battery [11]–[13]. The Coulomb counting method is widely used in SOC prediction due to its very low computational complexity, but its main drawback is still its low accuracy. To improve the accuracy of the SOC prediction from Coulomb counting, the unscented particle filter (UPF) has been applied to suppress the model parameter disturbance in the battery working model [14]. The application results of the UPF show that learning algorithms can improve the accuracy of the SOC estimation, and their application in SOC prediction is of great significance.

The Kalman filter is an effective method for removing noise from signals. This filter can predict the SOC value of batteries based on incomplete and noisy data. The key to the Kalman optimal estimation of a power system is the minimum variance calculation [15]–[17]. The extended Kalman filter and unscented Kalman filter generated based on the Kalman filter algorithm are also used in battery SOC prediction [18]–[20]. The Kalman algorithm performs well in the estimation of the SOC and can be applied to any kind of battery, but its computational complexity leads to a high application cost.

Machine learning has recently become a hot topic for SOC prediction and is impartial to the battery type. Artificial neural networks (ANNs) and the support vector machine for regression (SVR) are the typical machine learning methods for SOC estimation. The ANN algorithm collects a large number of battery indicator data during the charging and discharging process; then, through neural network training, it obtains the relationship between the SOC and the corresponding indicators [21]–[23]. The ANN model is composed of an input layer, a hidden layer and an output layer. The selection of the activation function type and the number of hidden layers and nodes should be considered in the model building. These choices are mostly based on experience and continuous attempts, which take considerable time and may encounter overfitting problems. Compared with an ANN, SVR can further improve the accuracy of the SOC prediction [24]. Hansen *et al.* demonstrated that SVR exhibits good performance in estimating the SOC of lithium-ion and lithium-ion polymer battery packs [28], [29]. Similar to the ANN, the training set selection of SVR will greatly affect the time spent during the training process and the accuracy of the final estimation results. Therefore, it is necessary to select the input and optimize the size of the existing SVR training set. In [24], Hu *et al.* collected six features of a battery from ADVISOR to estimate the SOC, and the double-step

search method was proposed to optimize the parameters  $c$  and  $g$  of the SVR model. In [25], three features of the battery were collected, and  $\nu$ -SVR was applied to estimate the SOC with fixed parameters  $\nu$  and  $\gamma$ . Fuzzy information granulation (FIG) and SVR were used together to predict the battery SOC based on the battery voltage and current [26]. There is no uniform standard for selecting the battery features of the SVR training set in the existing literature [24]–[27]. To analyze the influence of different battery factors on battery SOC prediction, principal component analysis (PCA) is carried out. PCA has the function of feature extraction and data compression, which is widely used in image processing and data compression [30], [31].

In this study, we use PCA to design the battery SOC simulation experiment based on the cumulative contribution and propose a new SOC prediction algorithm based on an SVR model with data classification and training set size optimization. A simulation experiment with the cycle condition is conducted in ADVISOR to collect data samples of the features of the electric vehicle battery. Applying PCA analysis to the collected data samples, the corresponding contribution of each battery feature is obtained. Based on the cumulative contribution of the battery features, three groups of SOC prediction simulation tests are designed with different training set inputs. The simulation results show that when applying the original SVR to predict the battery SOC, using the three features of the battery as the input of the training set can meet the prediction accuracy requirements. The improved SVR algorithm is then applied to the simulation experiment of three input features. According to the driving conditions of the electric vehicle, the collected data samples are divided into two categories. The optimal sizes of the SVR training sets for the two types of samples are obtained by a novel algorithm. The improved SVR prediction models are established for the two types of samples. The improved SVR algorithm is demonstrated faster and more accurately than the original SVR algorithm through a comparison of the simulation results.

## II. PRINCIPLE AND METHODOLOGY

According to the definition of the American Advanced Battery Alliance, the SOC is the ratio of the residual capacity to the rated capacity. The range of the SOC varies from 0 to 100%. The expression is as follows:

$$SOC = C_r/C_n \times 100\% \quad (1)$$

where  $C_r$  and  $C_n$  are the residual capacity and the rated capacity of the battery, respectively. The rated capacity is measured under specified conditions and is given by the manufacturer.

### A. SUPPORT VECTOR MACHINE REGRESSION

Support vector machines (SVMs) are designed to solve the classification problem at the beginning, including linear and nonlinear problems. When the sample data are linearly separable, the key of the SVM is to solve the problem of binary classification to find a classification line to divide the data

into two categories. When the data are two-dimensional or three-dimensional, the corresponding classification line is a straight line or plane, and when the data are multidimensional, the corresponding classification line is a linear hyperplane. When the data are linearly inseparable, the core of the SVM maps the original data from low-dimensional space to high dimensional space so that a linear hyperplane can be found to separate samples. The basic idea of the regression problem is similar to the basic idea of the classification problem; that is, they all need to build a bounded training set based on the sample data. The training set is a collection of training examples, which is also referred to as the training data. The training set is expressed as  $T = \{(x_1, y_1), \dots, (x_l, y_l)\} \subseteq (X \times Y)^l$ , where  $l$  is the number of examples,  $X$  denotes the input space,  $Y$  denotes the output domain, and  $T$  denotes the training set.

The difference between SVM regression and classification is the difference in the output. The output of the classification is only 1 or  $-1$ , and the output of the regression can be any real number. When we apply the SVM regression algorithm, we need to find a real-valued function  $f(\mathbf{x})$ , and the output can be inferred by  $y = f(\mathbf{x})$  with the corresponding input. For linear SVM regression, the real-valued function  $f(\mathbf{x})$  is mainly considered a linear function; that is,  $f(\mathbf{x}) = \langle \mathbf{w} \cdot \mathbf{x} \rangle + b$ , where  $(\mathbf{w}, \mathbf{x}) \in R^n \times R$ . For nonlinear SVM regression, the separable problem can be solved after mapping to a high or even infinitely high dimensional space by the kernel function.

Cortes and Vapnik [32] found the kernel function that under the Mercer condition in the reproducing kernel Hilbert space could approximate the dot product of two elements  $K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j)$ , where  $\phi(\mathbf{x}_j)$  is a kind of nonlinear function. The nonlinear regression estimation function can be expressed as the following form:

$$f(\mathbf{x}) = \mathbf{w}^T \cdot \phi(\mathbf{x}) + b \quad (2)$$

where  $\mathbf{x}$ ,  $\mathbf{w}$ , and  $b$  are the input data, the weight of the hyperplane and the intercept of the hyperplane, respectively.

Smola *et al.* discussed the concept of the insensitive loss function and gave the expression of the insensitive loss function in [33]. The expression is as follows:

$$\ell(f(\mathbf{x}), y) = \begin{cases} 0, & |y - f(\mathbf{x})| \leq \varepsilon \\ |y - f(\mathbf{x})| - \varepsilon, & |y - f(\mathbf{x})| > \varepsilon \end{cases} \quad (3)$$

where  $f(\mathbf{x})$  is the predicted value of  $\mathbf{x} \in R^n$  and  $\varepsilon$  is the error threshold between the predicted value  $f(\mathbf{x})$  and the true value  $y \in R^n$ . The selection of hyperplane parameters for SVR refers to reference [34]. The training process of the SVR algorithm is to find the hyperplane that minimizes the cumulative distance of all training samples from the hyperplane. The optimization problem can be represented as:

$$\begin{aligned} \min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2 \\ \text{s.t. } (\langle \mathbf{w} \cdot \mathbf{x}_i \rangle + b) - y_i \leq \varepsilon, \quad i = 1, \dots, l \end{aligned} \quad (4)$$

By introducing the Lagrange optimization method, the above problems can be transformed into a dual problem, and the nonlinear regression function can be obtained:

$$f(\mathbf{x}) = \sum (a_i - a_i^*) K(\mathbf{x}_i, \mathbf{x}) + b \quad (5)$$

Linear kernels, polynomial kernels, radial basis function kernels, and sigmoid kernels are the four common kernel functions. The radial basis function kernel is well known as a Gaussian kernel and works well in practical applications. The radial basis function kernel is chosen as the kernel function in this article and can be expressed as the following form:

$$K(\mathbf{x}_j, \mathbf{x}) = \exp \left( -\frac{\|\mathbf{x} - \mathbf{x}_j\|^2}{2\sigma^2} \right) \quad (6)$$

where  $\sigma$  denotes the standard deviation of the Gaussian kernel.

## B. THEORY OF PCA

PCA is a data dimension reduction method that chooses a few comprehensive variables by a linear transformation of multiple variables. This method is commonly used in pattern recognition, signal and image processing, machine learning and general exploratory data analysis. PCA provides an orthogonal projection onto a lower-dimensional subspace, referred to as the principal subspace, which captures most of the variance of the dataset. PCA dimension reduction is based on the idea that the first few principal components retain most of the variability of the original data. In many practical applications, PCA can be used to develop an understanding of the true dimensionality of the data as well as the contribution of each variable.

### 1) ESTABLISHING THE SAMPLE OBSERVATION MATRIX

Suppose there are  $m$  random vectors  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$ . Each vector contains  $n$  samples; that is,  $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{in})$ , and the  $m \times n$  sample observation matrix  $X$  is established.

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \quad (7)$$

where  $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m]^T$ . To solve the problem that the dimensions of different feature parameters cannot be compared, the original data are standardized.

$$x_{ij}^* = \frac{x_{ij} - \bar{x}_j}{\sqrt{\text{Var}(x_j)}} \quad (i = 1, 2, \dots, m; j = 1, 2, \dots, n) \quad (8)$$

where  $\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$ ,  $\text{Var}(x_j) = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2$ .

### 2) COVARIANCE MATRIX

$$C = \begin{bmatrix} \text{cov}(x_1, x_1) & \text{cov}(x_1, x_2) & \cdots & \text{cov}(x_1, x_m) \\ \text{cov}(x_2, x_1) & \text{cov}(x_2, x_2) & \cdots & \text{cov}(x_2, x_m) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(x_m, x_1) & \text{cov}(x_m, x_2) & \cdots & \text{cov}(x_m, x_m) \end{bmatrix} \quad (9)$$

where  $cov(x_i, x_j) = E[(x_i - E(x_i))(x_j - E(x_j))^T]$ . According to the principle of PCA, the characteristic equation of the covariance matrix  $\mathbf{C}$  is  $\mathbf{C}\alpha_i = \lambda_i\alpha_i$ ,  $\lambda_i$  is the eigenvalue of matrix  $\mathbf{C}$ , and  $\alpha_i$  is the eigenvector corresponding to  $\lambda_i$ .

### 3) CONTRIBUTION RATE OF M-DIMENSIONAL VECTOR PRINCIPAL COMPONENTS

The contribution rate of principal component  $i$  is  $L_i = \lambda_i/P(i = 1, 2, \dots, m; P = \sum_{i=1}^m \lambda_i)$ , and the cumulative contribution rate of the first  $l$  is  $L_l = \sum_{i=1}^l \lambda_i/P$ . The purpose of PCA is to reduce the number of variables. Generally, it does not take all principal components, but the former one, which can be determined according to the actual situation.

## III. SOC PREDICTION

### A. FEATURES IN THE DATASETS

The features used to predict the battery SOC value include the current of the battery, the voltage of the battery, the temperature of the battery, the requested power out of the energy storage system (ESS), the actual power loss for the ESS, and the heat removed from the battery by the cooling air, and usually some or all of these features are utilized for the SOC prediction. These features are commonly acquired by experiment or by simulation software. In this article, the features are obtained in the simulation software ADVISOR with the cycle condition New European Driving Cycle (NEDC), which is a comprehensive test condition formulated by the Ministry of Industry and Information Technology of China. When the SOC value of the battery decreases from 1 to 0, the features and the SOC value are collected. The samples of the features and the SOC value are shown in Fig. 1 and Fig. 2.

PCA analysis is carried out on the features of the battery collected in 3.1. The eigenvalue and contribution of each feature of the battery are obtained; subsequently, the cumulative contribution is obtained by accumulation.

In Fig. 3, the eigenvalues and the contributions of the battery voltage and current are much larger than those of other battery factors. Therefore, methods for predicting the battery SOC using only voltage or current are present. The corresponding methods are the OCV method and the Coulomb counting method. However, the direct use of the OCV method and the Coulomb counting method to predict the battery SOC value is not accurate, so there are many studies to improve it; that is, other battery factors are taken into account. The essence of the improvement is to increase the cumulative contribution of the factors used in the prediction.

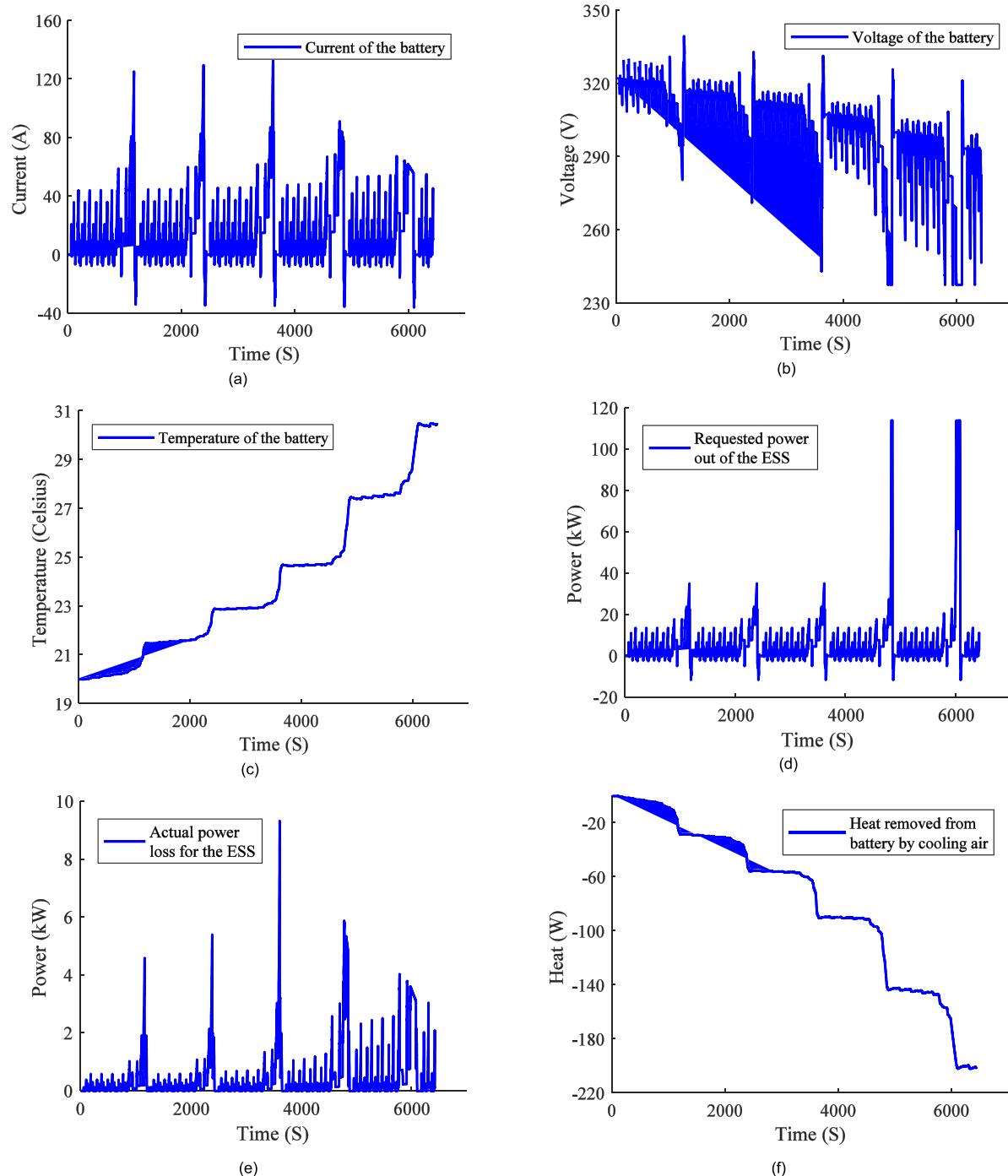
The current and voltage cumulative contribution of the battery is more than 90%, and the battery current, voltage, and temperature cumulative contribution are more than 95%. Therefore, the three groups of SOC simulation experiments are designed according to the cumulative contribution. The inputs of the training set features are the current and voltage of the battery, the current, voltage and temperature of the

battery, and the full features shown in Fig. 3. By comparing the simulation results of three groups of experiments, we can find the influence of feature selection on the SOC prediction.

### B. IMPROVED SVM REGRESSION

In this article, the improvement in the SVM for the regression algorithm mainly includes the establishment of a classification model for the sample set and the design of training set size optimization for accelerating the simulation process of the SVM. In practical applications, the classification accuracy of the SVM can reach more than 95% by using the algorithm and parameter optimization, while it is difficult to control the error of the regression prediction of the SVM within 5%. Therefore, the SVM has the characteristics of “easy classification, difficult regression”. The working characteristic of the battery in an electric vehicle is that braking energy recovery occurs during the driving process. Compared with the charge-discharge test of batteries in the laboratory, the battery SOC during the operation of the electric vehicle has a stronger nonlinearity. According to this characteristic, the sample set of the battery is classified, and the braking energy recovery process and other driving processes are distinguished.

When there are a large number of samples in the SVM object, the accuracy of the regression prediction results of the SVM is not necessarily proportional to the size of the training set when the SVM algorithm is applied directly. At the same time, the size of the training set directly affects the process of computer calculations, and the training of a large number of samples will take more time. In the case that the training accuracy is not proportional to the training size, directly selecting the training sample size according to experience cannot make full use of the sample data, which results in a waste of sample data and increases the training time. An optimization algorithm for the size of the training set is designed to obtain the best learning performance and reduce the training time. The design of the optimization algorithm is as follows: assuming the current sample size is  $m$ ,  $N$  is a constant. Then, a range of size changes is determined, using  $m/N, 2m/N, \dots, m$  sizes for training and testing, respectively. The ratio of the training set to the test set in each sample is set to 3:1, and the corresponding size samples are randomly sampled from the sample set for  $t$  time simulation. The training accuracy of the samples of different sizes can be obtained by averaging the value obtained from the  $t$  time simulation mean square error (MSE). The optimal training set size can be determined through the minimum MSE. The optimized size is selected as the training set size of the classified samples, and the corresponding regression prediction model is obtained by using SVM regression training. The prediction process of the SOC is as follows. The data are brought into the classification model, and the classification of SOC is determined. Then, this classification is brought into the corresponding regression model for regression prediction, and the prediction results of SOC are obtained. The flow chart of the simulation test design and the improved SVM regression to predict the SOC value step is shown in Fig. 4.



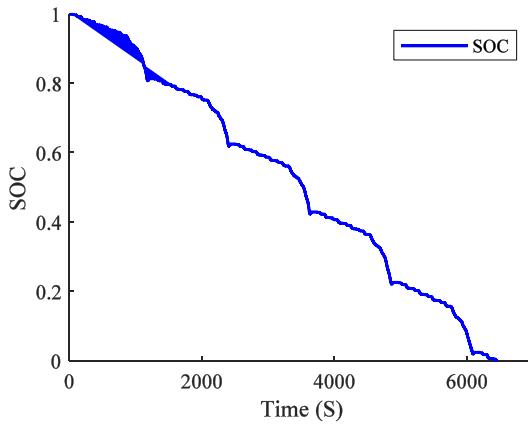
**FIGURE 1.** Features in the NEDC testing dataset. (a) Current of the battery. (b) Voltage of the battery. (c) Temperature of the battery. (d) Requested power out of the ESS. (e) Actual power loss for the ESS. (f) Heat removed from the battery by cooling air.

The detailed explanations of the flow chart in Fig. 4 are as follows:

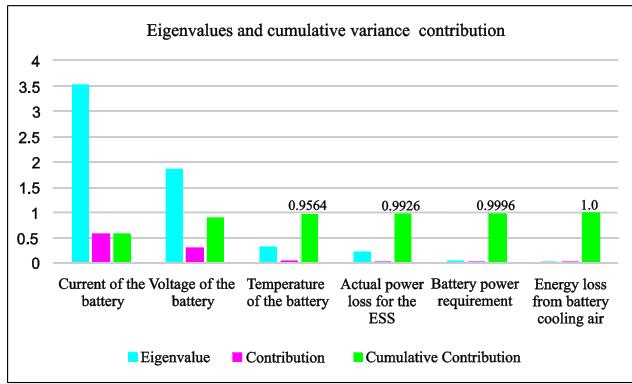
- 1) The original training samples for the simulation experiments are obtained from ADVISOR;
- 2) PCA analysis of the features of the battery obtains the cumulative contribution;
- 3) According to the cumulative contribution, three groups of SVM regression simulation experiments with different training set inputs are carried out;

4) The improved SVM regression algorithm is applied to the simulation experiment; the training set inputs are the current, voltage and temperature of the battery. With prior knowledge, the features SVM classification model is established, and then the SVM regression model is established for the classified samples with the optimal training set size;

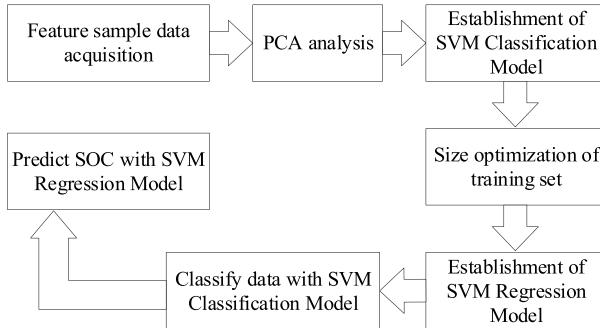
5) To deal with the data previously collected, the classification model obtained in step (4) is used to classify the data.



**FIGURE 2.** The actual SOC.



**FIGURE 3.** Eigenvalues and cumulative variance contribution.



**FIGURE 4.** The SOC prediction flow chart.

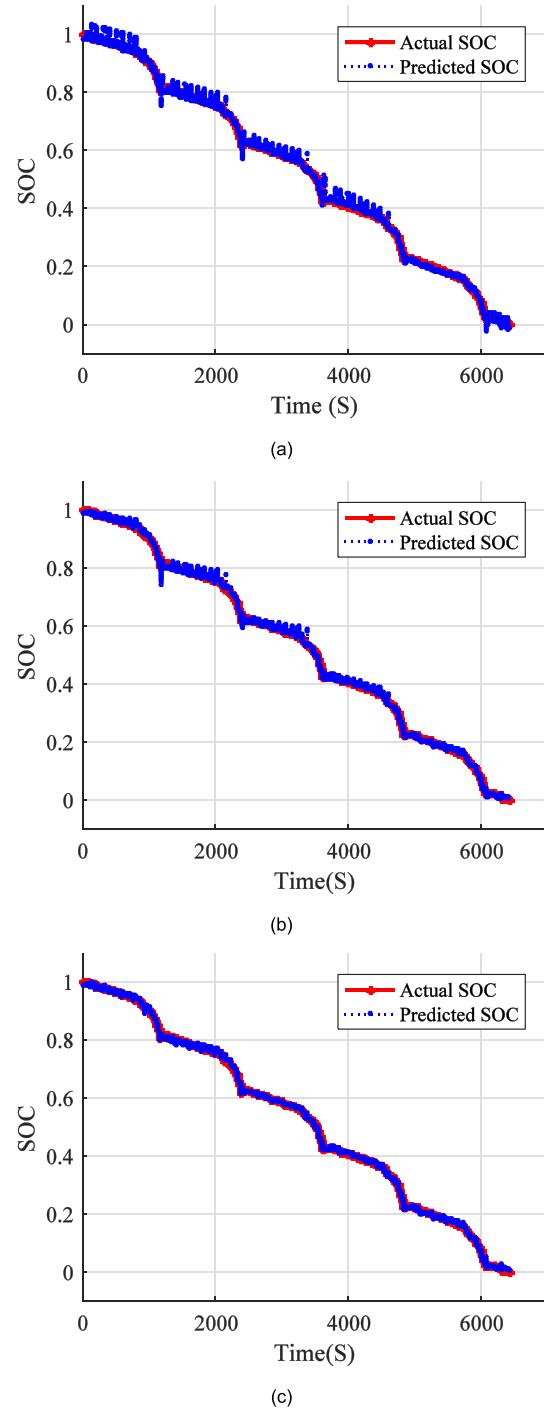
After classifying the data, the regression model in step (4) is used to obtain the prediction results.

## IV. RESULTS AND DISCUSSION

### A. ORIGINAL SVM REGRESSION PREDICTION

According to the cumulative contribution obtained from PCA analysis, three groups of battery SOC simulation experiments with different inputs are designed, as described in Subsection 3.2. In this section, the SOC value is predicted by SVM regression without improvement, and the parameters  $c$  and  $g$  of the SVM regression are optimized by a grid search and

cross-validation. The training set is randomly selected in the sample set, and the size of the training set is 0.75 of the sample set. A comparison between the predicted and actual values of the SOC obtained by simulation is shown in Fig. 5.



**FIGURE 5.** Comparison of the predicted and actual SOC: (a) Simulation results of two feature parameters. (b) Simulation results of three feature parameters. (c) Simulation results of all the feature parameters.

The MSE evaluation results of the three groups of experiments are shown in Table 1. In the table, C, V, T, L, P, and E represent the current of the battery, the voltage of the

**TABLE 1.** Original SVR prediction MSE.

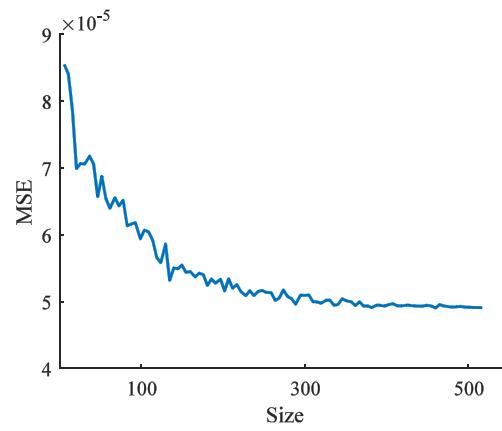
Input of the training set	MSE
C V	0.0005836
C V T	0.00026151
C V T L P E	0.00010273

battery, the temperature of the battery, the actual power loss for the ESS, the battery power requirement, and the energy loss from the battery cooling air, respectively. When the input of the training set is C and V, the predicted value of SOC has exceeded the actual range of the SOC, as shown in Fig. 5 (a), and the MSE of the SOC is the highest in the three groups of experiments. The predicted value of the SOC with the input of the training set three features (C, V, T) and all features (C, V, T, L, P, E) is in the actual range, as shown in Fig. 5 (b) and Fig. 5 (c).

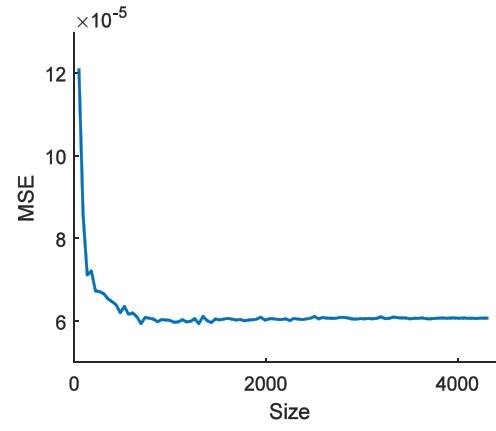
When the training set inputs are C and V, although the predicted value of the SOC exceeds the SOC limit at this time, the MSE obtained by the simulation experiment is not very different from that of the other two sets of experiments. Therefore, some studies use the battery current and voltage to predict the SOC value when the required accuracy is not high. However, under the NEDC cycle condition, the vehicle accelerates and decelerates frequently. The training set inputs C and V change drastically, resulting in the prediction result exceeding the limit value at the initial and end of the prediction process. This behavior is also related to the characteristics of the battery itself. When the battery begins to discharge, the electrochemical reaction inside the battery has just begun, and the internal temperature slowly rises, failing to reach the optimal temperature of the electrochemical reaction. When the battery is about to be exhausted, the polarization resistance inside the battery increases rapidly, causing the electrochemical reaction to be greatly affected. The internal electrochemical reaction of the battery corresponding to the discharge phases of the above two batteries changes drastically, which makes it difficult to accurately predict the SOC value. When T is added to the training set input, the cumulative contribution of the training set input exceeds 95%, and the trend of the temperature change is relatively fixed, with fewer mutations, which has a limiting effect on the change in the SOC value, so the SOC prediction value of Fig. 5 (b) and Fig. 5 (c) do not exceed the limit. Therefore, when the accuracy of the SOC prediction requirements is high, we still need to consider the impact of battery temperature in the SOC prediction process. In addition, when three battery features and all battery features as the input of the training set are used in the simulation experiments, similar prediction results are obtained. However, the cost of collecting all of the battery features is high, and the acquisition method is complicated. Therefore, when using SVR to predict the battery SOC value, it is sufficient to use the three battery features (C, V, T) as the input of the training set.

**TABLE 2.** Classification results.

V (V)	C (A)	T (°C)	Category	Classification results
322.2401	0.0310	20.0000	1	1
322.2400	0.0310	20.0000	1	1
320.2179	6.6611	20.0000	1	1
318.8191	11.2407	20.0000	1	1
317.2227	16.4647	20.0000	1	1
⋮	⋮	⋮	⋮	⋮
297.2327	-5.6726	30.4949	-1	-1
296.4973	-4.7353	30.4952	-1	-1
295.7856	-3.8285	30.4951	-1	-1
295.0857	-2.9371	30.4949	-1	-1
293.6287	-1.0837	30.4945	-1	-1
⋮	⋮	⋮	⋮	⋮



(a)

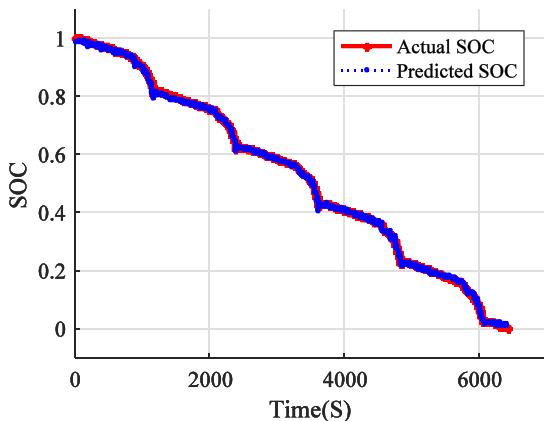


(b)

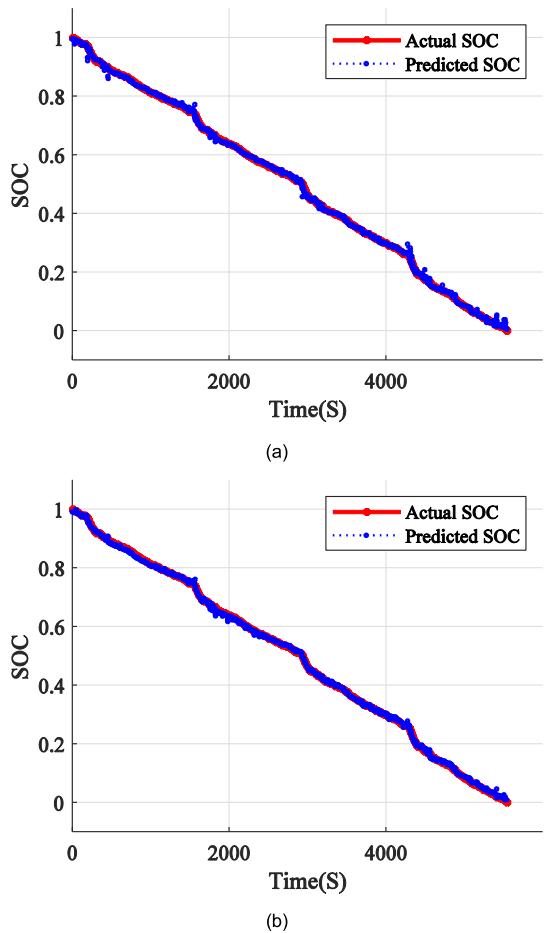
**FIGURE 6.** Simulation results for training set size optimization: (a) The optimal size of the braking energy recovery process. (b) The optimal size of the other driving processes.

### B. IMPROVED SVM REGRESSION PREDICTION

To speed up the simulation operation and improve the accuracy of the prediction results, the improved SVM regression algorithm is applied to the SOC prediction experiment in



**FIGURE 7.** Comparison of the predicted and actual SOC with improved SVR.

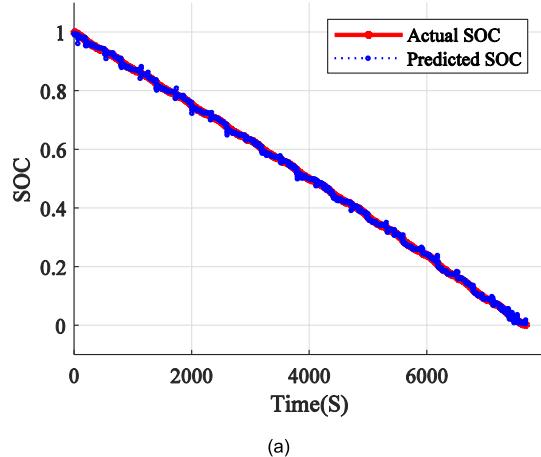


**FIGURE 8.** Comparison of the predicted and actual SOC in the UDDS driving condition: (a) original SVR (b) improved SVR.

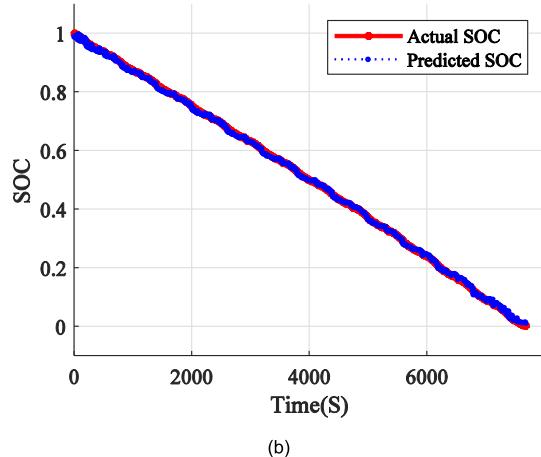
which the input of the training set is three features ( $C$ ,  $V$ ,  $T$ ). The SVM classification model is used to classify the samples obtained from the software ADVISOR. The samples that have no brake energy recovery are taken as the first category and are recorded as 1. The samples that have brake energy recovery are taken as the second category and are recorded as -1. There are 5754 sets of data in the first category and 689 sets of data in the second category. The test samples are

**TABLE 3.** Comparison of the original and improved SVR prediction.

Classification	Time-consuming (s)	MSE
Original SVR	4396	$2.6151 \times 10^{-4}$
Improved SVR	1081	$5.816 \times 10^{-5}$



(a)



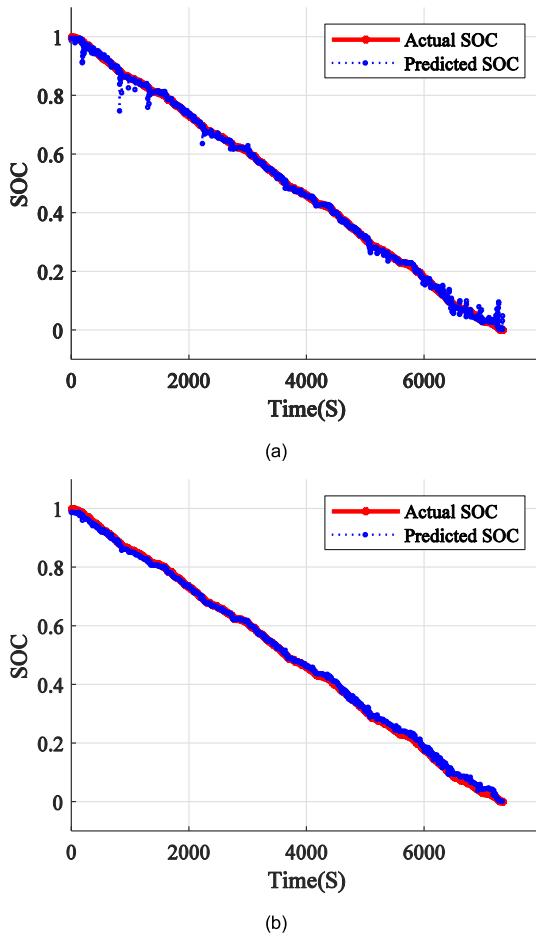
(b)

**FIGURE 9.** Comparison of the predicted and actual SOC in the NYCC driving condition: (a) original SVR (b) improved SVR.

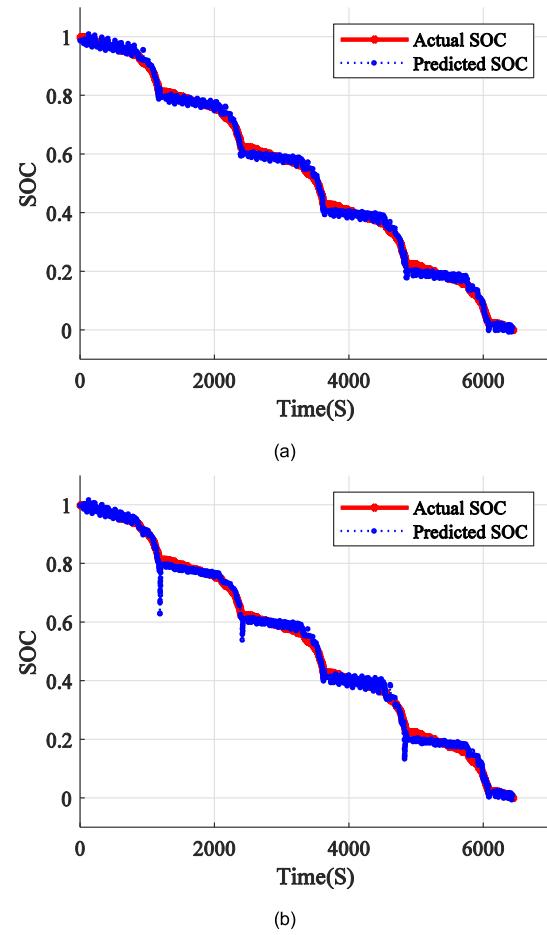
classified, and the results are shown in Table 2. The accuracy of the classification results is 100%.

The classified sample is optimized using the optimization algorithm designed in Section 3.3 to obtain the optimal training set size. The results of training set size optimization are shown in Fig. 6.

The optimal training set sizes for the braking energy recovery process and the other driving processes are 490 and 680, respectively. The optimal training set sizes are used to establish the SVM regression training sets. The penalty parameters  $c$  and kernel parameter  $g$  of SVM regression are optimized by using a grid search and cross-validation, and SVM regression models for the braking energy recovery process and other driving processes are established. A comparison of the



**FIGURE 10.** Comparison of the predicted and actual SOC in the WVUCITY driving condition: (a) original SVR (b) improved SVR.



**FIGURE 11.** Comparison of the predicted and actual SOC in the NEDC driving condition: (a) BPNN (b) Elman network.

predicted improved SVM regression results and actual SOC values is shown in Fig. 7. In this article, an ASUS computer is used in the simulation, the CPU is an I7-6500U, the memory is 12 GB, the solid-state hard disk is 250 GB, and the system is WIN10.

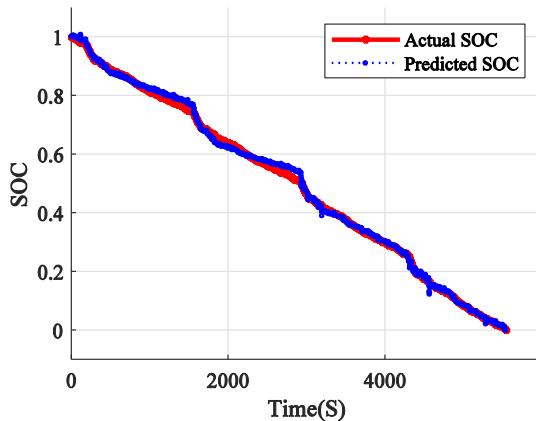
Comparing Fig. 7 and Fig. 5 (b), it can be seen that the predicted SOC results in Fig. 7 are significantly better than those in Fig. 5 (b). It can be seen from the results in Table 3 that when the improved SVR algorithm is used for the three features input of the training set to predict the battery SOC, the accuracy of the obtained simulation results is not only higher than that of the original SVR algorithm but is also higher than that of the original SVR algorithm using all features as the input of the training set. Compared with the original SVR algorithm, the improved SVR algorithm reduces the MSE by 4.496 times. This result indicates that the preclassification of the sample set and the optimization of the size of the training set are beneficial for the improvement in the accuracy of the estimation results. It can be seen from the comparison of the time spent on the simulation that the time spent on the improved SVR algorithm simulation is 75.4% lower than the time spent on the original SVR algorithm.

because the size of the training set used by the improved SVR algorithm is much smaller than that used by the original SVR algorithm.

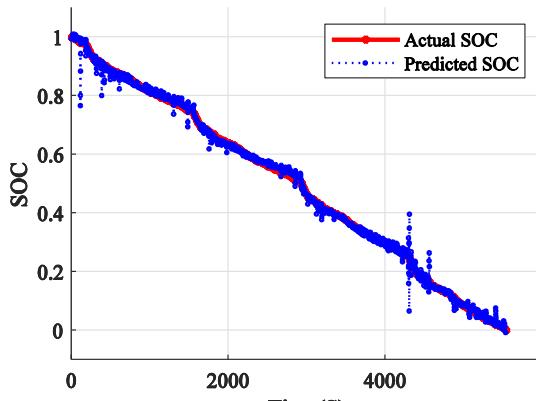
### C. DIFFERENT DRIVING CONDITIONS AND OTHER BASIC SCHEMES

The simulation results in the NEDC driving condition show that the improved SVR algorithm has a better prediction effect than the original SVR algorithm. To verify whether the improved SVR algorithm has the same promotion effect on other driving conditions, the Urban Dynamometer Driving Schedule (UDDS), New York City Council (NYCC), and the West Virginia University City (WVUCITY) driving conditions are selected for the simulation experiments. The input of the training set comprises three features of the battery (C, V, and T), and the output of the training set is the battery SOC. The comparison between the predicted SOC by the original SVR algorithm and the improved SVR algorithm and the actual SOC is shown in Fig. 8 to Fig. 10. The simulation results are shown in Table 4.

In Fig. 8, Fig. 9, and Fig. 10, the battery SOC takes different amounts of times from one to zero due to different

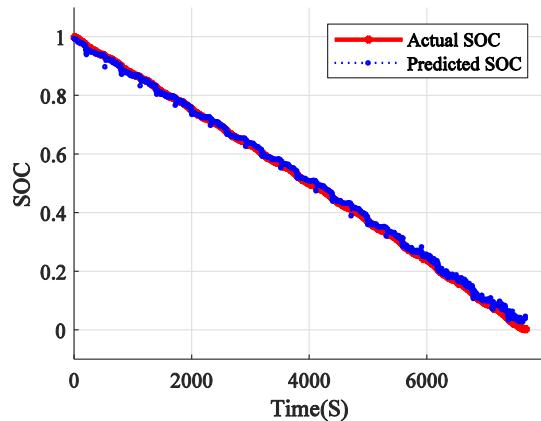


(a)

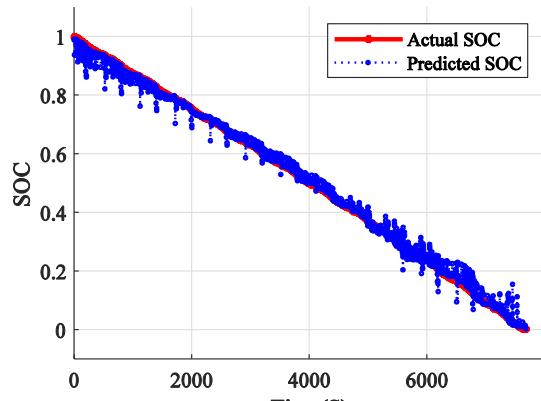


(b)

**FIGURE 12.** Comparison of the predicted and actual SOC in the UDDS driving condition: (a) BPNN (b) Elman network.



(a)



(b)

**FIGURE 13.** Comparison of the predicted and actual SOC in the NYCC driving condition: (a) BPNN (b) Elman network.

driving conditions, and the prediction effect of the improved SVR algorithm is obviously better than that of the original SVR algorithm. Table 4 shows that the time consumption of the improved SVR algorithm is lower than that of the original SVR algorithm. The time spent on the improved SVR algorithm simulation is 72%, 76.3%, and 78% lower than the time spent on the original SVR algorithm in the UDDS, NYCC, and WVUCITY driving conditions, respectively. The simulation results show that the improved SVR algorithm can adapt to different driving conditions.

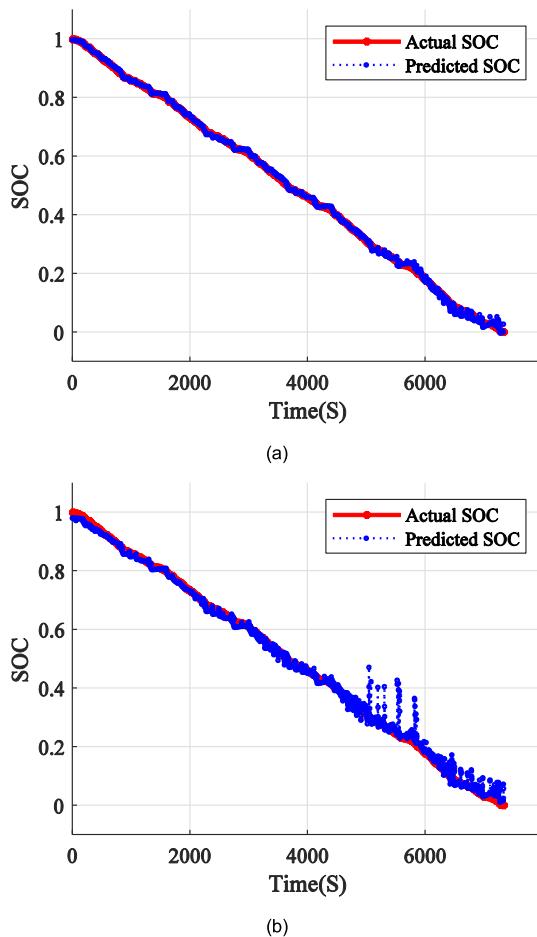
Two basic schemes of the back propagation neural network (BPNN) and Elman network are used to predict the battery SOC under four driving conditions: the NEDC, UDDS, NYCC, and WVUCITY. The BPNN includes 10 hidden neurons using the log-sigmoid transfer function in the hidden layer, and one output is produced using the pure linear output function in the output layer. The Elman network includes 5 hidden neurons using the hyperbolic tangent sigmoid transfer function in the first hidden layer and 3 hidden neurons in the second hidden layer, and one output is produced using the linear output function in the output layer. The input and output of the BPNN and Elman are three features of the battery

**TABLE 4.** Comparison of the original and improved SVR prediction under different driving conditions.

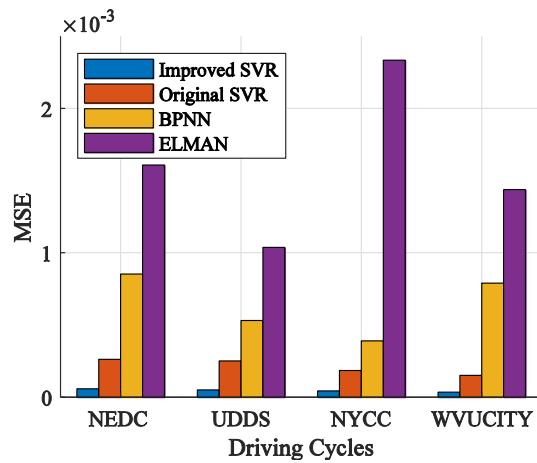
Driving Cycle	Classification	Time-consuming (s)	MSE
UDDS	Original SVR	3522	$2.503 \times 10^{-4}$
	Improved SVR	951	$4.9501 \times 10^{-5}$
NYCC	Original SVR	5250	$1.8469 \times 10^{-4}$
	Improved SVR	1397	$4.3169 \times 10^{-5}$
WVUCITY	Original SVR	5330	$1.5077 \times 10^{-4}$
	Improved SVR	1415	$3.4458 \times 10^{-5}$

(C, V, and T) and the battery SOC. The comparison between the predicted SOC by the BPNN and Elman network with the actual SOC is shown in Fig. 11 to Fig. 14. The predicted curves from the Elman network fluctuate severely in the running process and the predicted results from the BPNN are obviously better than those of the Elman network.

The MSE evaluations for the four driving conditions are shown in Fig. 15. It can be seen that the MSE of the



**FIGURE 14.** Comparison of the predicted and actual SOC in the WVUCITY driving condition: (a) BPNN (b) Elman network.



**FIGURE 15.** Comparison of the MSE from the improved SVR, original SVR, BPNN, and Elman network under different driving conditions.

improved SVR method is much lower than that of the other estimations.

## V. CONCLUSION

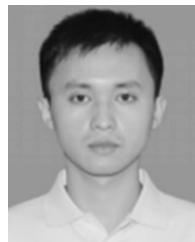
The prediction of the battery SOC is an important function of the BMS. The accurate prediction of the battery SOC is

a goal that researchers have been pursuing. In this article, an improved SVR model, which is obtained by classifying the collected feature samples of the battery and optimizing the training set sizes, is used to predict the battery SOC of electric vehicles. Multiparameter grid search and cross-validation methods are used to optimize the SVR parameters. The simulation experiments are designed by PCA analysis of battery features, and the results of the experiments show that the three feature inputs of the training set can meet the SOC prediction requirements. The improved SVR algorithm is applied to the SOC prediction in which the input of the training set is three features, and the results show that the simulation time drops dramatically and the prediction accuracy is improved. When it is evaluated using the driving conditions in ADVISOR, the improved SVR algorithm in this article behaves more accurately than other estimations. Future work for the SOC prediction of electric vehicle batteries using SVR will be carried out to evaluate complicated driving conditions.

## REFERENCES

- [1] K. Palmer, J. E. Tate, Z. Wadud, and J. Nellthorp, "Total cost of ownership and market share for hybrid and electric vehicles in the UK, US and Japan," *Appl. Energy*, vol. 209, pp. 108–119, Jan. 2018.
- [2] M. Faisal, M. A. Hannan, P. J. Ker, A. Hussain, M. B. Mansor, and F. Blaabjerg, "Review of energy storage system technologies in microgrid applications: Issues and challenges," *IEEE Access*, vol. 6, pp. 35143–35164, 2018.
- [3] R. Xiong, J. Cao, Q. Yu, H. He, and F. Sun, "Critical review on the battery state of charge estimation methods for electric vehicles," *IEEE Access*, vol. 6, pp. 1832–1843, 2018.
- [4] M. A. Hannan, M. M. Hoque, A. Hussain, Y. Yusof, and P. J. Ker, "State-of-the-art and energy management system of lithium-ion batteries in electric vehicle applications: Issues and recommendations," *IEEE Access*, vol. 6, pp. 19362–19378, 2018.
- [5] F. Xia, H. Chen, L. Chen, and X. Qin, "A hierarchical navigation strategy of EV fast charging based on dynamic scene," *IEEE Access*, vol. 7, pp. 29173–29184, 2019.
- [6] P.-Y. Kong and G. K. Karagiannidis, "Charging schemes for plug-in hybrid electric vehicles in smart grid: A survey," *IEEE Access*, vol. 4, pp. 6846–6875, 2016.
- [7] R. Xiong, L. Li, Z. Li, Q. Yu, and H. Mu, "An electrochemical model based degradation state identification method of lithium-ion battery for all-climate electric vehicles application," *Appl. Energy*, vol. 219, pp. 264–275, Jun. 2018.
- [8] C. Lin, Q. Yu, R. Xiong, and L. Y. Wang, "A study on the impact of open circuit voltage tests on state of charge estimation for lithium-ion batteries," *Appl. Energy*, vol. 205, pp. 892–902, Nov. 2017.
- [9] G. Dong, J. Wei, C. Zhang, and Z. Chen, "Online state of charge estimation and open circuit voltage hysteresis modeling of LiFePO<sub>4</sub> battery using invariant imbedding method," *Appl. Energy*, vol. 162, pp. 163–171, Jan. 2016.
- [10] F. Zheng, Y. Xing, J. Jiang, B. Sun, J. Kim, and M. Pecht, "Influence of different open circuit voltage tests on state of charge online estimation for lithium-ion batteries," *Appl. Energy*, vol. 183, pp. 513–525, Dec. 2016.
- [11] K. S. Ng, C.-S. Moo, Y.-P. Chen, and Y.-C. Hsieh, "Enhanced Coulomb counting method for estimating state-of-charge and state-of-health of lithium-ion batteries," *Appl. Energy*, vol. 86, no. 9, pp. 1506–1511, Sep. 2009.
- [12] C. Truchot, M. Dubarry, and B. Y. Liaw, "State-of-charge estimation and uncertainty for lithium-ion battery strings," *Appl. Energy*, vol. 119, pp. 218–227, Apr. 2014.
- [13] Y. Wang, C. Zhang, and Z. Chen, "A method for state-of-charge estimation of li-ion batteries based on multi-model switching strategy," *Appl. Energy*, vol. 137, pp. 427–434, Jan. 2015.
- [14] Y. He, X. Liu, C. Zhang, and Z. Chen, "A new model for state-of-charge (SOC) estimation for high-power li-ion batteries," *Appl. Energy*, vol. 101, pp. 808–814, Jan. 2013.

- [15] G. L. Plett, "Sigma-point Kalman filtering for battery management systems of LiPB-based HEV battery packs. Part 1: Introduction and state estimation," *J. Power Sources*, vol. 161, no. 2, pp. 1356–1368, 2006.
- [16] G. L. Plett, "Sigma-point Kalman filtering for battery management systems of LiPB-based HEV battery packs. Part 2: Simultaneous state and parameter estimation," *J. Power Sources*, vol. 161, no. 2, pp. 1356–1368, 2006.
- [17] G. L. Plett, "Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs - Part 3. State and parameter estimation," *J. Power Sources*, vol. 134, no. 2, pp. 277–292, 2004.
- [18] W. Wang and J. Mu, "State of charge estimation for lithium-ion battery in electric vehicle based on Kalman filter considering model error," *IEEE Access*, vol. 7, pp. 29223–29235, 2019.
- [19] W. Wang, X. Wang, C. Xiang, C. Wei, and Y. Zhao, "Unscented Kalman filter-based battery SOC estimation and peak power prediction method for power distribution of hybrid electric vehicles," *IEEE Access*, vol. 6, pp. 35957–35965, 2018.
- [20] X. Wu, X. Li, and J. Du, "State of charge estimation of lithium-ion batteries over wide temperature range using unscented Kalman filter," *IEEE Access*, vol. 6, pp. 41993–42003, 2018.
- [21] F. Yang, X. Song, F. Xu, and K.-L. Tsui, "State-of-charge estimation of lithium-ion batteries via long short-term memory network," *IEEE Access*, vol. 7, pp. 53792–53799, 2019.
- [22] M. A. Hannan, M. S. H. Lipu, A. Hussain, M. H. Saad, and A. Ayob, "Neural network approach for estimating state of charge of lithium-ion battery using backtracking search algorithm," *IEEE Access*, vol. 6, pp. 10069–10079, 2018.
- [23] X. Chen, W. Shen, M. Dai, Z. Cao, J. Jin, and A. Kapoor, "Robust adaptive sliding-mode observer using RBF neural network for lithium-ion battery state of charge estimation in electric vehicles," *IEEE Trans. Veh. Technol.*, vol. 65, no. 4, pp. 1936–1947, Apr. 2016.
- [24] J. N. Hu, J. J. Hu, H. B. Lin, X. P. Li, C. L. Jiang, X. H. Qiu, and W. S. Li, "State-of-charge estimation for battery management system using optimized support vector machine for regression," *J. Power Sources*, vol. 269, pp. 682–693, 2014.
- [25] J. C. Antón, P. J. Nieto, F. J. de Cos Juez, F. Lasheras, M. Vega, and M. N. Gutiérrez, "Battery state-of-charge estimator using the SVM technique," *Appl. Math. Model.*, vol. 37, no. 9, pp. 6244–6253, May 2013.
- [26] W. Huang, Y. Zhao, and Q. Huangpeng, "SOC prediction of lithium battery based on fuzzy information granulation and support vector regression," in *Proc. 4th Int. Conf. Electr. Electron. Eng. (ICEEE)*, Ankara, Turkey, Apr. 2017, pp. 177–180.
- [27] T. Qin, S. Zeng, and J. Guo, "Robust prognostics for state of health estimation of lithium-ion batteries based on an improved PSO-SVR model," *Microelectron. Rel.*, vol. 55, nos. 9–10, pp. 1280–1284, Aug. 2015.
- [28] T. Hansen and C.-J. Wang, "Support vector based battery state of charge estimator," *J. Power Sources*, vol. 141, no. 2, pp. 351–358, Mar. 2005.
- [29] D. Andre, C. Appel, T. Soczka-Guth, and D. U. Sauer, "Advanced mathematical methods of SOC and SOH estimation for lithium-ion batteries," *J. Power Sources*, vol. 224, pp. 20–27, Feb. 2013.
- [30] N. Gaitani, C. Lehmann, M. Santamouris, G. Mihalakakou, and P. Patargias, "Using principal component and cluster analysis in the heating evaluation of the school building sector," *Appl. Energy*, vol. 87, no. 6, pp. 2079–2086, Jun. 2010.
- [31] J. Ma, Z. Wang, C. Zhu, Y. Xu, and Z. Wu, "Electrogenesis reduces the combustion efficiency of sewage sludge," *Appl. Energy*, vol. 114, pp. 283–289, Feb. 2014.
- [32] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, pp. 273–297, Sep. 1995.
- [33] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statist. Comput.*, vol. 14, no. 3, pp. 199–222, Aug. 2004.
- [34] V. Cherkassky and Y. Ma, "Practical selection of SVM parameters and noise estimation for SVM regression," *Neural Netw.*, vol. 17, no. 1, pp. 113–126, 2004.



**LIANG XUAN** was born in Lu'an, Anhui, China, in 1991. He received the M.S. degree in automotive engineering from the School of Automotive and Traffic Engineering, Hefei University of Technology, China, in 2014, where he is currently pursuing the Ph.D. degree. His research interests include electric vehicle (EV) charging strategies and EV control strategy with connected and intelligent environment.

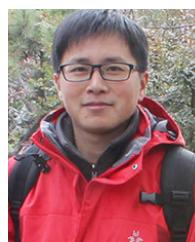


**LIJUN QIAN** was born in Anhui, China, in 1962. He received the M.S. degree from the Academy of Armored Force Engineering, Beijing, China, in 1990, and the Ph.D. degree from the Hefei University of Technology, China, in 2004, all in vehicle engineering.

He is currently a Professor with the School of Automotive and Traffic Engineering, Hefei University of Technology. He has published more than 200 articles and holds 40 patents. His research interests include connected and intelligent vehicles, intelligent transportation, new energy vehicles, and vehicle dynamics and control.



**JIAN CHEN** received the B.E. degree in vehicle engineering from the Hefei University of Technology, Anhui, China, in 2017, where he is currently pursuing the master's degree. His research interests include vehicle dynamics and control technology.



**XIANXU BAI** (Member, IEEE) was born in Lu'an, Anhui, China, in 1984. He received the M.S. and Ph.D. degrees in optoelectronic engineering from Chongqing University, Chongqing, China, in 2013.

He is currently a Professor with the School of Automotive and Traffic Engineering, Hefei University of Technology, Hefei, China. His research interests mainly include intelligent structure systems, vehicle system dynamics and control, intelligent vehicle active braking, and intelligent suspension/mounting direction.



**BING WU** was born in Anhui, China, in 1986. He received the M.S. degree in automotive engineering from the School of Automotive and Traffic Engineering, Hefei University of Technology, China, in 2012, where he is currently pursuing the Ph.D. degree. His research interests include vehicle automation, motion planning, real-time trajectory generation, and advanced driver assistance systems (ADAS).

• • •