A state of health estimation method for electric vehicle Li-ion batteries using GA-PSO-SVR

Yue Zhi^{1,3} · Heqi Wang² · Liang Wang⁴

Received: 27 August 2021 / Accepted: 10 December 2021 / Published online: 18 January 2022
© The Author(s) 2022

Abstract

State of health (SOH) is the ratio of the currently available maximum capacity of the battery to the rated capacity. It is an important index to describe the degradation state of a pure electric vehicle battery and has an important reference value in evaluating the health level of the retired battery and estimating the driving range. In this study, the random forest algorithm is first used to find the most important health factors to lithium-ion batteries based on the dataset released by National Aeronautics and Space Administration (NASA). Then the support vector regression (SVR) algorithm is developed to predict the SOH of a lithium-ion battery. The genetic algorithm-particle swarm optimization (GA-PSO) algorithm is brought forward to optimize the parameter values of the SVR, which could improve the estimation accuracy and convergence speed. The proposed SOH estimation method is applied to four batteries and gets a root mean square error (RMSE) of 0.40% and an average absolute percentage error (MAPE) of 0.56%. In addition, the method is also compared with genetic algorithm-support vector regression (GA-SVR) and particle swarm optimization-support vector regression (PSO-SVR), respectively. The results show that (i) compared with the PSO-SVR method, the proposed method can decrease the average RMSE by 0.10%, and the average MAPE by 0.17%; (ii) compared with the GA-PSO method, number of iterations under the proposed method can be reduced by 7 generations.

Keywords Pure electric vehicles · Lithium-ion battery · State of health · Estimation method · Support vector regression

Introduction

The development of pure electric vehicles (PEV) is of great significance for reducing dependence on fossil energy and reducing vehicle emissions [58–60]. The development of sustainable transportation system has become a common goal all over the world [46]. Therefore, many countries worldwide are vigorously promoting the deployment of electric vehicles [34], which may be used in public transit in the future [48].

The battery, which provides the power for the vehicles, is one of the most important parts of the PEV. Currently, lithium-ion batteries are gradually replacing lead-acid, nickel-metal hydride, and other traditional batteries with advantages such as high energy density, long cycle times, low self-discharge rate, and slight memory effect, and have become the preferred power solution for PEV [58–60]. As a result, the market share has increased exponentially in recent years [52].

During the use of the lithium-ion battery, changes in the internal micro-chemical structure will result in corresponding variations in their external characteristics, which are mainly manifested in the loss of capacity and the increase in impedance. In turn, the decline in the battery's performance will affect the power performance of the PEV and the user's driving experience. According to the recommendation of the American Electric Vehicle Association [10], to ensure the cruising range and driving safety, the service life of the power battery can be considered to reach the end when the capacity of the on-board power battery has decayed to 70–80% of the nominal capacity. Hence, it should be decommissioned and replaced with a new power battery pack in time.

✉ Liang Wang
13718230117@163.com

¹ School of Rehabilitation Medicine, Binzhou Medical University, No. 346, Guan Hailu, Laishan District, Yantai 264003, Shandong, China

² Department of Transport and Planning, Delft University of Technology, Stevinweg 1, Delft 2628 CN, The Netherlands

³ Institute of Rehabilitation Engineering, Binzhou Medical University, No. 346, Guan Hailu, Laishan District, Yantai 264003, Shandong, China

⁴ China Academy of Transportation Sciences, Ministry of Transportation China, Beijing 100029, China

Literature review

State of health (SOH) is generally expressed as the ratio of the currently available maximum capacity of the battery to the nominal capacity. SOH is an important parameter that characterizes the actual available capacity of the battery and its degradation state, which also has an important reference value in evaluating the health level of the retired battery and the estimation of the driving range [58–60]. The battery SOH estimation of PEV has always been a hot research topic. Existing researches mainly includes two categories: model-based methods and data-driven methods. According to different modeling mechanisms, model-based methods can then be divided into equivalent circuit models [5] and electrochemical models [6].

The equivalent circuit models ignore the complex physical and chemical processes inside the battery and simulate the output effect of the battery through basic electronic components. Tong et al. [42] proposed a new SOC (state of charge) and SOH relationship function based on the second-order equivalent circuit model, which can be used for the capacity update in SOC estimation. Cheng et al. [10] proposed a decoupling weighted recursive least square method to optimize the parameters of the equivalent circuit model to overcome the problem of low SOH estimation accuracy under different battery dynamic mechanisms. Based on the equivalent circuit model, these methods usually require less calculation, but the values of the model parameters are definite under fixed operation conditions, which limits its application conditions [13, 27, 37].

Based on the physical and chemical reactions inside the battery, the electrochemical model derives the battery performance degradation mechanism related to the material characteristics [4]. A pseudo-two-dimensional model (P2D model) based on porous electrode theory (Doyle, 1993) describes the internal dynamic mechanism of lithium-ion batteries through a series of partial differential equations (PDE). In this way, the SOH can be estimated more accurately, and the estimation method can be adapted to li-ion batteries with different electrode materials by adjusting parameters. However, solving the PDE is difficult [1]. To simplify the calculation and limit the complexity of the model, two model simplification schemes can be used: (i) applying a single particle model (SPM) [31]; (ii) reducing the order of PDE to ordinary differential equations. However, simplifying the model will reduce the model accuracy and the scope of application to varying degrees [3]. Therefore, the trade-off between the calculation and the accuracy requires further research [23, 63].

Using the data-driven method, the relationship between input characteristics and battery capacity can be established directly according to the historical data without analyzing the internal mechanism of the battery. In recent years, data-driven

methods have been widely used in battery state estimation. Data-driven methods can be used to establish a certain relationship between the input and output (Hussein et al. 2014). Wang et al. [45] proposes a blockchain-based framework to improve the efficiency of ship traffic in port. Wu et al. [49] proposed an online approach using feedforward neural network (FFNN) and importance sampling (IS) to estimate lithium-ion battery remaining useful life (RUL). Chaoui et al. [7] presents an application of dynamically driven recurrent networks (DDRNs) in online electric vehicle (EV) battery analysis. Long et al. [29] proposed an improved autoregressive model by particle swarm optimization using for prognostics of lithium-ion batteries. Song et al. [39] proposed a hybrid method of iterative nonlinear degradation autoregressive (IND-AR) model and particle filter (PF) algorithm are proposed in reference work to improve the long term prediction performance of lithium-ion battery. Song et al. [39] proposed an autoregressive and moving average (ARMA) model to solve the problem of online remaining useful lifetime prediction of fuel cells. Abdollahzadeh et al. [2] proposed African vultures optimization algorithm to solve global optimization problems and provided superior performance on the majority of engineering case studies.

Among these studies, the support vector regression (SVR) [20], which is developed based on statistical theory and the principle of structural risk minimization, has unique advantages for a small sample, nonlinear, and high-dimensional problems. Thus, SVR has received wide attention in SOH estimation. Although convex optimization problems [44] are able to solve large-scale practical engineering problems [9], random forest algorithm [53], not only has its unique merit in reducing the overfitting problems [55], but also has been widely used in feature screening [22]. In considering the nature and structure of the health of lithium-ion batteries, random forests are employed in this study.

Besides a better modeling method, the selection of health factors (HF) that can express the degree of battery degradation is also an important issue in SOH estimation. Liu and Chen [28] used three features as HFs during the constant current (CC) and constant voltage (CV) charging, which were the charging time interval of voltage varying from 3.9 to 4.2 V, the charging voltage varying from 3.9 V to the voltage after 500 s, the CV charging current drop between 1.5 A (the CC charge current) and the current after 1000 s, respectively. Xiong et al. [50] extracted the charging capacity during the constant current charging stage from 3.6 to 4.2 V as the HF of SOH. Guo et al. [18] extracted 8 kinds of HF from the voltage, current, and temperature curves to estimate the battery SOH. The reasonable selection of HF helps to accurately model the battery degradation. Compared with complex and changeable discharge curves, charging curves are relatively stable and easy to extract features. Notably, the features extracted based on the discharge curve are often only suitable for fixed

Fig. 1 Capacity degradation curves of two batteries

constant current discharge conditions. In addition, temperature changes also have a non-negligible impact on battery capacity degradation.

National Aeronautics and Space Administration (NASA) battery data set [36] is a data set measured by experimenters on battery charge and discharge cycles according to certain steps, which contains parameters such as voltage, current, temperature during charge and discharge as well as capacity after each charge and discharge cycle. Figure 1 is the battery capacity degradation curve of B0005 and B0007. According to the calculation method of SOH (i.e., the ratio of the currently available maximum capacity of the battery to the rated capacity), due to the fixed rated capacity, the decreasing tendency of SOH is just the same as the one of capacity. In fact, there exist side reactions that occur between the electrodes and the electrolyte of the battery. And this contributes to the degradation tendency of Li-ion battery available capacity as the number of charge and discharge cycles increases [19], which can be observed in Fig. 1 as well. While apart from that, another interesting phenomenon is presented in the figure that the batteries will regenerate the capacity in the process of capacity degradation [40], called capacity regeneration (CR). Actually, this is a common phenomenon. If the battery rests during the charging/discharging process, the remaining reaction products have a chance to dissipate, and in this way, the available capacity in the next cycle will immediately increase [30]. CR includes two stages, namely the capacity regeneration stage and the accelerated degradation stage. The available battery capacity exhibits an abnormal degradation tendency in these two stages, which will increase the difficulty of predicting the SOH of lithium batteries. Moreover, due to different influencing factors (e.g., charge and discharge current, voltage, time, etc.), the capacities of the two batteries become different after the same number of cycles.

Therefore, the number of charges and discharge cycles cannot be simply used as the only factor in battery SOH estimation.

Importance of this research

With the increase in the number of electric vehicles year by year, there will be a large number of retired or scrapped lithium-ion power batteries that need to be effectively recycled to achieve the best use of resources and avoid waste of resources and environmental pollution. Research by the National Renewable Energy Laboratory in the United States shows that lithium-ion power batteries retired from electric vehicles still have a certain value for recycling and reuse. According to different remaining capacities or health levels, these batteries can continue to play a role in public power grids, photovoltaic or wind power generation systems, and commercial residences [35].

Hence, it is essential to precisely estimate the SOH of lithium-ion power batteries: first, SOH is one of the important parameters for accurately assessing the energy status of power batteries and the driving range of vehicles; second, real-time monitoring of the SOH of power batteries can ensure that the potential safety hazards of the batteries during use are discovered in time, and can provide technical guarantee for the stable operation of the power battery packs. Finally, the SOH status information of vehicle-mounted power batteries is one of the important criteria for evaluating the health level of retired batteries, and has a very important reference value in the recycling of used lithium-ion power batteries. This article is based on NASA battery data set for SOH estimation.

On the basis of the statement in section “Literature Review”, previous studies did not take the problem of coupling between HF into account. Some use the Spearman

correlation coefficient and Pearson coefficient to describe the degree of correlation between SOH and HF, lacking more mature theoretical support. Besides, when establishing data-driven SOH estimation models for lithium-ion batteries, some scholars choose heuristic algorithms to optimize the parameters of SVR, such as genetic algorithm and particle swarm algorithm, without considering to improve the convergence speed and SOH estimation accuracy from the perspective of ameliorating the algorithm.

Contributions and objectives

This paper extracts the health factors that can describe battery capacity degradation from the monitoring parameters of the battery and establishes an SVR algorithm to estimate the battery SOH. The parameter values of the SVR algorithm pay considerable impacts on the estimation quality. Given the importance of the search cost, how to determine the optimal values for the parameters is a challenging task [32]. Particle swarm optimization (PSO) and genetic algorithm (GA) are both not suitable for parameter optimization because the former is not good at finding the global optimal results [51], and the latter is slow in convergence speed [11]. By combining PSO with GA algorithm and using the crossover and mutation operation of GA algorithm, the PSO particles increase the population diversity and make the particles fly to a new search area, which improves the convergence speed of the algorithm and avoids the local extremum. Hence, this paper uses the GA-PSO algorithm to find the optimal parameter values of the SVR, and a lithium-ion battery SOH estimation method based on GA-PSO-SVR is proposed.

The contributions of this paper are twofold:

- i. The HFs are extracted from the battery charging curves and temperature curves, and then a random forest algorithm is applied to find the optimal HFs to reduce data dimension.
- ii. An SVR algorithm is developed to estimate the SOH of lithium-ion batteries for PEV, and the GA-PSO algorithm is employed to optimize the parameters in the SVR algorithm. The proposed method in this paper has higher reliability, estimation accuracy, and faster convergence speed.

The outline of the paper is organized as follows. The next section presents the extraction of HFs from the battery charging current curves, charging voltage curves, and charging temperature curves. The subsequent section proposes a SOH estimation method based on SVR and uses GA-PSO for parameter optimization followed by which the SOH estimation results of lithium-ion batteries are discussed and the proposed method is compared with other methods. The final

section presents the conclusions from this study as well as some remarks.

Health factor extraction and selection

In this section, the random forest algorithm is used to select the optimal HFS from the parameters in the battery charging curve and temperature curve, to reduce the data dimension.

Influencing factors analysis of battery SOH

To estimate the battery SOH, it is first necessary to extract the health factors that have important influences on the degradation of the battery capacity. The battery capacity degradation data used in this study come from NASA's lithium battery cycle test data set. Table 1 is the basic information of the selected 4 battery parameters and experimental conditions. As mentioned earlier, this study extracts HFs from the current, voltage, and temperature curves during battery charging. The battery charging process includes two stages: constant current (CC) charging and constant voltage (CV) charging. First, the battery is charged with a constant current of 1.5A until the voltage rises to 4.2 V and then charged with a constant voltage until the current drops to 20 mA. During the CC charging phase, the battery temperature will continue to rise and reach a peak value and then begin to drop during the CV charging phase.

In the process of battery capacity decline, SOH can directly indicate the degree of decline, which is generally represented by the ratio of the current maximum available capacity to the initial capacity. The maximum available capacity is not easy to be measured, while with the HF extracted from the voltage, current, and temperature curve, the battery's capacity degradation can be indirectly reflected [10]. As an example, the current, voltage, and temperature curves of the B0005 battery in the 32nd cycle of the charging stage are shown in Fig. 2.

Widodo [47], Li [25, 26] mentioned that 12 HFs can be initially extracted from the charging current, voltage, and temperature curves, which can be divided into 4 categories: (i) the time and time ratio of different charging stages: CC charging phase time L1, CV charging phase time L2, their ratio L1/L2, and the total charging phase time L; (ii) the integral of the current curves of different charging stages over time: CT1 is for the current curve of the CC charging stage, CT2 is for the CV charging stage, and CT is for the total charging stage; (iii) the integral of the temperature curve of different charging stages over time: T1 is for the temperature curve of the CC charging stage, T2 is for the CV charging stage, and T is for the total charging stage; (iv) the maximum slope of the charging voltage curve K1 and the maximum slope of the charging current curve K2.

Table 1 Parameters and experimental conditions for selected batteries

Battery number	Discharge current/A	Cut-off voltage/V	Temperature/°C	Rated Capacity/Ah	End-of-life capacity/Ah
B0005	2	2.7	24	2	1.40
B0007	2	2.2	24	2	1.45
B0029	4	2.0	43	—	1.65
B0031	4	2.5	43	—	1.70

Fig. 2 Current, voltage, and temperature curves of B0005 in the charging stage **Fig. 3** Min–max normalized HF

Taking the B0005 battery as an example, the input HF are processed by the min–max normalization method, as shown in Fig. 3.

Comparing the capacity decay curve of battery B0005 in Fig. 3 with Fig. 1, some HF are positively correlated with SOH, such as L1, L1/L2, CT1, etc.; some are negatively correlated with SOH, such as L2, CT, T2, etc., while some HF and SOH do not show significant correlation relationship, such as L, T, etc.

Optimal HF selection based on random forest

As shown in Fig. 3, the changing tendencies of various HF are different, and it is difficult to determine the degree of correlation between them and SOH. Due to the existence of autocorrelation among HF and the weak correlation between some HF and SOH, the selected 12 HF need to be further optimized to extract HF with a higher degree of correlation with SOH. Some scholars use Spearman correlation coefficient, gray correlation degree, Pearson correlation coefficient, etc. to evaluate the degree of correlation between SOH

and HFs [41]. The random forest (RF) model is considered in this study. The reason why random forest is selected for feature screening instead of the correlation coefficient method here is that, the correlation coefficient method only has a good performance in dealing with linear correlation problems and is very significantly affected by outliers, while there is no obvious linear relationship between battery SOH degradation and various features, and the data are concentrated with a large amount.

RF can be divided into two categories: regression tree model [14] and decision tree model [15]. The difference between them is mainly due to the different attributes of the nodes in the tree. The leaf node of the last layer in the classification problem (corresponding to the decision tree model) is the classification label, and the other nodes are all certain feature attributes of the problem. While in the regression problem (corresponding to the regression tree model), all leaf nodes are attributed to be regressed. Here the regression tree model is used.

The tree generation consists of two parts: training set construction and node splitting. The training set used to construct each tree is created by applying the Bootstrap sampling method on the original data set. Bootstrap is one of the data sampling methods with replacement. Specifically, it samples less than the total number of samples with replacement in the sample set. After the sampling is completed, the samples that are not drawn are called out of bag (OOB) samples. Decision trees constructed in this way have different structures. Hence, there will be no over-fitting problems of traditional decision trees. At the same time, the variance of the model can be reduced, making the model more stable. Node splitting includes algorithms such as Concept Learning System (CLS), Iterative Dichotomiser3 (ID3), C4.5, and Classification and Regression Trees (CART). Due to the page limits, they will not be elaborated here. The main idea of node splitting is to traverse a certain number of feature segmentation points in a certain way and then determine the selection criteria according to different problems. For regression problems, the mean square error minimization criterion is often selected. When a specified number of trees are generated in the forest, the final estimation result is the average of all tree estimation results, where each tree has an equal weight.

The RF feature selection relies on calculating the feature importance of each feature. Importance can be interpreted as the contribution of each feature to the regression result, which can be calculated in two ways: the permutation method and the Gini method. Since the Gini method is often implemented for classification problems, thus the permutation method is used in the regression problems in this study. In the permutation method, some features are replaced first and then re-measure the regression accuracy. By comparing the accuracies of the new data set and old data set due to feature

changes, to give the feature importance. The specific steps are as follows:

Step 1: For the original data set D , calculate the accuracy s in the model. The accuracy here refers to the fitting accuracy.

Step 2: For feature j in the 12 HFs extracted in section “Influencing factors analysis of battery SOH”, repeat Steps 2.1 and 2.2 B times. The purpose of calculating B times is to avoid the influence of the uncertainty of random replacement. The calculation steps are as follows:

Step 2.1: Replace a feature j , and record the newly generated data set as $\tilde{D}_{b,j}$ ($b = 1, 2, \dots, B$).

Step 2.2: Calculate the estimation accuracy $s_{b,j}$ ($b = 1, 2, \dots, B$) achieved by the new data set $\tilde{D}_{b,j}$ in the model.

Step 3: Calculate the feature importance i_j of feature j :

$$i_j = s - \frac{1}{B} \sum_{b=1}^B s_{b,j}. \quad (1)$$

Step 4: After calculating the importance of all features, normalize them to get the importance of each feature norm (i_j):

$$\text{norm}(i_j) = \frac{i_j}{\sum_{j \in D} i_j}. \quad (2)$$

In this study, 4 batteries with different parameters and operating conditions are taken as examples. Thus, the selected HFs of the 4 batteries using RF may be different. Taking battery B0005 as an example, the 12 HFs that have been preliminarily extracted are analyzed, and the feature importance calculated by RF is shown in Fig. 4. The HFs with higher feature importance are L1, CT1, CT, and T1. Comparing Fig. 3 and Fig. 1, it can be seen that these four HFs have basically the same downward tendency as the battery capacity, and there exists also a phenomenon similar to capacity regeneration. In addition, although CT2 and L1/L2 also have a downward tendency, due to the autocorrelation with the selected 4 HFs, their health importance is low; C2 shows an upward tendency, but its feature importance is also higher than T2, T, etc., which is caused by its changing tendency is opposite to the capacity decline tendency, proving that to a certain extent the RF algorithm can also reflect the negative correlation between HFs and SOH. Hence, the four most important HFs are selected, which are L1, CT1, CT, and T1. The calculated feature importance of the remaining HFs accounted for 0.15 in total, which is much lower than the selected HFs. Thus, they were excluded from the data set.

Fig. 4 Feature importance of HFs for B0005 battery

SOH estimation model and parameter optimization

This section proposes an SVR model to estimate the battery SOH based on the HFs selected in section “[Health factor extraction and selection](#)”. Battery capacity decline data are used to train the model, and GA-PSO is used to optimize model parameters to improve the accuracy of SOH estimation.

SOH estimation model

Support vector machine (SVM) [43] proposed by Vapnik et al., is a linear classifier that finds the separation hyperplane and maximizes the interval in the feature space [8]. At the same time, using SVM can also perform regression analysis on the data (Yu et al. 2017). When the input variable is classified data and the output variable is the label, support vector classification (SVC) can be implemented to classify the sampled data; when the input variables and output variables are both numerical data, SVR can be adopted to do regression. In addition, SVR has unique advantages for a small sample, nonlinear, and high-dimensional problems. By observing the sample data, it is found that the SOH estimation of lithium-ion batteries meets the application conditions of SVR.

SVR is a convex optimization problem of which the locally optimal solution is the globally optimal solution [43]. Hence, there is no problem of falling into local optimality. At the same time, the optimization objective of the SVR model is to minimize the structural risk, that is, the expected risk of the learning machine on the entire sample set could be controlled while ensuring accuracy. Therefore, the SVR model has a good capacity for generalization and applies to SOH estimation.

The Z health features extracted from section “[Health factor extraction and selection](#)” are stored in a vector $\mathbf{x} = (x_1, x_2, \dots, x_Z)$, and the actual SOH value is denoted by y . The original sample data can be expressed as $T = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_O, y_O)\}$. O is the sample size, that is, the total number of cycles. To estimate y through the SVR model, a function $f(\mathbf{x})$ needs to be constructed so that its value is as close as possible to y .

When performing linear regression, the constructed function $f(\mathbf{x})$ is

$$f(\mathbf{x}) = \mathbf{x}\mathbf{w}^T + w_0, \quad (3)$$

where $\mathbf{w} = (w_1, w_2, \dots, w_Z)$ is the row vector to be determined, and $\mathbf{w}^T$ is the transpose. It should be noted that its dimension is consistent with the dimension of $\mathbf{x}$. w_0 is the parameter to be determined.

The linear regression problem can be rewritten as solving the following optimization problem:

$$\min_{\mathbf{w}, w_0} \frac{1}{2} \|\mathbf{w}\|^2, \quad (4)$$

$$s.t. |f(\mathbf{x}_o) - y_o| \leq \varepsilon, o = 1, 2, \dots, O, \quad (5)$$

where ε is the error threshold between $f(\mathbf{x})$ and y . Equation (5) is to control the error between the estimated function value and the true value within the error threshold. However, this does not always hold. Considering that there may be situations where Eq. (5) may not be true, slack variables need to be introduced, and the optimization problem is rewritten as follows:

$$\min_{\mathbf{w}, w_0, \xi_o, \hat{\xi}_o} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{o=1}^O (\xi_o + \hat{\xi}_o), \quad (6)$$

$$\begin{aligned}
s.t. \quad & f(\mathbf{x}_o) - y_o \leq \varepsilon + \xi_o, \\
& y_o - f(\mathbf{x}_o) \leq \varepsilon + \hat{\xi}_o, \\
& \xi_o \geq 0, \hat{\xi}_o \geq 0, o = 1, 2, \dots, O,
\end{aligned} \quad (7)$$

where C is the penalty coefficient; ξ_o and $\hat{\xi}_o$ are relaxation factors.

The first term in Eq. (6) makes the function flatter, thereby improving the generalization ability, and the purpose of the second term is to reduce the error. C is the weight coefficient used to adjust the two items. When C is too large, over-fitting is prone to occur, and when C is too small, under-fitting is prone to occur. To find the required function $f(\mathbf{x})$, it is necessary to obtain the extreme value of Eq. (6). For such problems with objective functions and constraints, Lagrangian multipliers are usually employed. Introducing Lagrange multiplier μ_o to the above formula, it can be got

$$\begin{aligned}
L(\mathbf{w}, w_0, \alpha, \hat{\alpha}, \xi, \hat{\xi}, \mu, \hat{\mu}) \\
= \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{o=1}^O (\xi_o + \hat{\xi}_o) - \sum_{o=1}^O \xi_o \mu_o - \sum_{o=1}^O \hat{\xi}_o \hat{\mu}_o \\
+ \sum_{o=1}^O \alpha_o (f(\mathbf{x}_o) - y_o - \varepsilon - \xi_o) + \sum_{o=1}^O \hat{\alpha}_o (f(\mathbf{x}_o) - y_o - \varepsilon - \hat{\xi}_o).
\end{aligned} \quad (8)$$

Let the partial derivative of $L(\mathbf{w}, w_0, \alpha, \hat{\alpha}, \xi, \hat{\xi}, \mu, \hat{\mu})$ with respect to $\mathbf{w}$, w_0 , ξ_o , $\hat{\xi}_o$ be zero:

$$\frac{\partial L}{\partial \mathbf{w}} = 0; \frac{\partial L}{\partial w_0} = 0; \frac{\partial L}{\partial \xi_o} = 0; \frac{\partial L}{\partial \hat{\xi}_o} = 0. \quad (9)$$

Then Eq. (9) can be solved:

$$\mathbf{w} = \sum_{o=1}^O (\hat{\alpha}_o - \alpha_o) \mathbf{x}_o, \quad (10)$$

$$0 = \sum_{o=1}^O (\hat{\alpha}_o - \alpha_o), \quad (11)$$

$$\begin{cases} C = \alpha_o + \mu_o \\ C = \hat{\alpha}_o + \hat{\mu}_o \end{cases}. \quad (12)$$

Any local extreme point of the convex optimization problem is the global extreme point, so the local optimal point is also the global optimal point. Transforming the linear regression problem into a convex optimization problem is conducive to obtaining the optimal estimated value of SOH. Although the SVR problem transformed by the Lagrangian multiplier method is not a convex optimization problem, its dual function is. Therefore, the original problem (the linear regression) is transformed into its dual problem [43]. Using

hadron duality, the problem now is to find the optimal solution to the dual problem. Substituting Eqs. (10)–(12) into Eq. (8), Eq. (6) can be obtained as follows:

$$\begin{aligned}
\max_{\mathbf{0}} \quad & \sum_{o=1}^O y_o (\hat{\alpha}_o - \alpha_o) - \varepsilon (\hat{\alpha}_o - \alpha_o) \\
& - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m (\hat{\alpha}_o - \alpha_o) (\hat{\alpha}_o - \alpha_o) x_{ti}^T x_{tj} s.t. \sum_{i=1}^m (\hat{\alpha}_o - \alpha_o) \\
& = 0, 0 \leq \alpha_o, \hat{\alpha}_o \leq C,
\end{aligned} \quad (13)$$

where x_{ti}, x_{tj} are different eigenvectors; $x_{ti}^T x_{tj}$ is the inner product value of eigenvectors x_{ti} and x_{tj} .

For convex optimization problems, the KKT (Karush–Kuhn–Tucker) condition is necessary for the existence of extreme points (also the global extreme points). Equations (8)–(13) needs to meet the KKT condition, that is, it needs to meet Eq. (14):

$$\begin{cases} \alpha_o (f(\mathbf{x}_o) - y_o - \varepsilon - \xi_o) = 0 \\ \hat{\alpha}_o (y_o - f(\mathbf{x}_o) - \varepsilon - \hat{\xi}_o) = 0 \\ \alpha_o \hat{\xi}_o = 0, \xi_o \hat{\alpha}_o = 0 \\ (C - \alpha_o) \xi_o = 0, (C - \hat{\alpha}_o) \hat{\xi}_o = 0 \end{cases} \quad (14)$$

Substituting Eq. (10) into Eq. (3), the final solution of SVR can be obtained as

$$f(\mathbf{x}) = \sum_{o=1}^O (\hat{\alpha}_o - \alpha_o) \mathbf{x}_o^T \mathbf{x} + w_0, \quad (15)$$

w_0 can be calculated by Eq. (16):

$$w_0 = y_o + \varepsilon - \sum_{o=1}^O (\hat{\alpha}_o - \alpha_o) \mathbf{x}_o^T \mathbf{x}, \alpha_o \in (0, C). \quad (16)$$

In this study, the input variables are multi-dimensional. For example, the number of HFs for battery B0005 is 4. In such as condition, the linear regression estimation is not suitable. Therefore, the kernel function method will be applied to directly use the function relationship between the inner product and the input variable to find the inner product value. The kernel function method can simplify the calculation process. Assuming that $\phi: x \rightarrow F$ is a mapping from the input space to a feature space, then

$$\mathbf{w} = \sum_{o=1}^O (\hat{\alpha}_o - \alpha_o) \phi(\mathbf{x}_o). \quad (17)$$

Replacing the feature mapping form, that is, substituting $\mathbf{x}$ with $\phi(\mathbf{x})$, then Eq. (3) can be expressed as

$$f(\mathbf{x}) = \sum_{o=1}^O (\hat{\alpha}_o - \alpha_o) \kappa(\mathbf{x}_o, \mathbf{x}) + w_0, \quad (18)$$

where $\kappa(\mathbf{x}_o, \mathbf{x}) = \phi(\mathbf{x}_o)^T \phi(\mathbf{x})$ is the kernel function. In this paper, the kernel function is a Radial Basis Function (RBF), which can map the objective function to infinite dimensions, and its specific form is:

$$\kappa(\mathbf{x}_o, \mathbf{x}) = \exp(-\gamma \|\mathbf{x}_o - \mathbf{x}\|^2). \quad (19)$$

Substituting Eq. (19) into Eq. (18), the final form of $f(\mathbf{x})$ can be obtained as

$$f(\mathbf{x}) = \sum_{o=1}^O (\hat{\alpha}_o - \alpha_o) \exp(-\gamma \|\mathbf{x}_o - \mathbf{x}\|^2) + w_0. \quad (20)$$

When constructing the SVR model, the parameters that need to be tuned include the penalty coefficient C and the parameter γ of the kernel function. The size of the penalty coefficient C will directly affect the degree of acceptance of noise and then affect the accuracy of the model. Besides, the parameter γ of the kernel function also determines the degree of acceptance of noise. When γ is larger, the model is less sensitive to noise, which leads to the loss of some sample data.

Choosing the appropriate SVR parameters in practical applications is an arduous and important task because it affects the fitting accuracy and generalization performance of support vector regression. Therefore, it is better to develop an optimization method for determining the values of parameters in the SVR.

SVR parameter optimization

Various algorithms such as genetic algorithm (GA) and particle swarm optimization (PSO) can be applied to the parameter optimization in SVR. However, the genetic algorithm has a slow convergence rate due to the information sharing between the chromosomes, and the particle swarm algorithm is easy to fall into the local optimum due to the poor interaction between the particles. Therefore, in this paper, the improved GA-PSO algorithm is adopted for parameter selection. It is a product that combines the advantages of GA's global optimization effect and PSO's fast convergence speed.

Traditional PSO algorithm

The PSO algorithm is a swarm intelligence algorithm proposed by Eberhart and Kennedy [24], which is derived from the mathematical simulation of many biological activities in nature. The activities of an individual in a group are usually affected by its own and group experiences. PSO has the same characteristics of a swarm intelligence algorithm and the same advantages as evolutionary computing. It uses the mutual cooperation and competition between particles in the group to realize the search for the optimal solution to the problem.

Assuming a PSO is composed of M particles, the dimension of the search area is D , $\mathbf{l}_m = (l_{m1}, l_{m2}, \dots, l_{mD})$ refers to the position of particle m in the D -dimensional space, and $\mathbf{v}_m = (v_{m1}, v_{m2}, \dots, v_{mD})$ refers to the flying speed of particle m .

Fitness is the only indicator for judging the quality of PSO. After each iteration, the particle position information is selected as the parameter value of the SVR model, which is adopted for training and SOH estimation. Apart from that, the fitness function is used to calculate the fitness of the particle at the current position for subsequent flight speed calculation and position update. Therefore, the selection of fitness function will also affect the result of particle swarm optimization. The root mean square error (RMSE) is an effective indicator for judging the quality of a model. In this paper, the RMSE of the test set is taken as the fitness function. Therefore, the fitness function can be calculated as follows:

$$f_m = \frac{1}{O} \sqrt{\sum_{o=1}^O (y_o - \hat{y}_o)^2}, \quad (21)$$

where f_m is the fitness value of particle m , %. When f_m is smaller, the result of PSO is better; y_o is the actual SOH of the test set, %; $\hat{y}_o$ is the estimated SOH of the test set, %.

Define $f_{m,best}$ as the extreme value of the individual fitness of the particle, that is, the optimal solution found by the particle m in its own optimization history, and its position in the D -dimensional space is expressed as $\mathbf{p}_m = (p_{m1}, p_{m2}, \dots, p_{mD})$; define f_{best} as the minimum fitness found in the search history of the entire group, and its position in the D -dimensional search space is denoted as $\mathbf{pp} = (pp_1, pp_2, \dots, pp_D)$.

In the k th generation, the update methods of the optimal fitness function value of particle m and the entire population are as follows:

$$f_{m,best}^k = \begin{cases} f_1 & \text{if } m = 1 \\ \min(f_{m,best}^{k-1}, f_m^k) & \text{if } 2 \leq m \leq M' \end{cases} \quad (22)$$

$$f_{best}^k = \min(f_{1,best}^k, f_{2,best}^k, \dots, f_{M,best}^k), \quad (23)$$

where $\min()$ represents the minimizing function; the corresponding particle positions of $f_{m,\text{best}}^k, f_{\text{best}}^k$ are $\mathbf{p}_m^k = (p_{m1}^k, p_{m2}^k, \dots, p_{mD}^k)$ and $\mathbf{pp}^k = (pp_1^k, pp_1^k, \dots, pp_D^k)$, respectively.

For the $(k + 1)$ th iteration, the position of each particle in each dimension is updated according to the following two equations:

$$v_{md}^{k+1} = v_{md}^k + c_1 \text{rand}() (p_{md}^k - l_{md}^k) + c_2 \text{rand}() (pp_d^k - l_{md}^k), \quad (24)$$

$$l_{md}^{k+1} = l_{md}^k + v_{md}^{k+1}, \quad (25)$$

where v_{md}^{k+1} is the velocity of particle m in the $(k + 1)$ th generation and the d th dimension component; the value range of d is $[1, D]$, and the value range of m is $[1, M]$; l_{md}^{k+1} is the position of particle m in the $k + 1$ th generation and the d th dimension component; p_{md}^k is the individual optimal solution of particle m in the k th generation and the d th dimension component; c_1 and c_2 are learning factors, which adjust the maximum step length of particles flying in the direction of p_{md}^k and pp_d^k , respectively; $\text{rand}()$ is a random number with a value between $[0, 1]$.

Equation (24) is explained as follows: the first part of Eq. (24) represents the velocity of the particle in the previous generation; the second part is for the individual learning of the particle, and the individual optimal value represents the result of the particle's self-learning. The second part promotes the particles to perform a better global search and prevents the search process from falling into the local optimum. The third part is population learning, which embodies the mutual information sharing ability between particles and results of group learning within the entire population. Under the combined influence of these three parts, the change in the velocity of each particle determines the change in the position of each particle. Hence, during the continuous iteration of the entire particle population, the search area will develop in a better direction so that the particles can search for the best global position.

GA-PSO algorithm

In the traditional PSO algorithm, the particle velocity is updated according to the fixed method in Eqs. (24) and (25), and only the individual optimal particle and the overall optimal particle transmit information to other particles so that the flow of information is one-way, which makes it more likely to generate premature convergence and local optimum. In GA, the information sharing between chromosomes enables the entire race to move to the optimal region more evenly, and it is not easy to fall into the local optimum. However, at the same time, the disadvantage that accompanies this advantage is that the information sharing of chromosomes makes

the convergence too slow and also increases the complexity of the algorithm.

Therefore, the GA-PSO algorithm proposed in the literature [56] is adopted here, which adds the selection and crossover operations of the GA to the PSO to obtain the GA-PSO optimization algorithm. Adding selection operation to the PSO algorithm can control the evolution direction. Adding a crossover operation could help to reach a new search area and avoid the local optimum. At the same time, as the GA-PSO does not involve the encoding, decoding, mutation, and other operations in GA, the complexity of the new optimization algorithm has not been greatly increased, and the convergence speed is not reduced significantly.

The GA-PSO algorithm process is as follows:

Step 1: Initialize the population size, search dimension, number of iterations, learning factor, convergence accuracy, and other parameters of the PSO algorithm according to the optimization problem.

Step 2: Randomly assign a value to the position and velocity of each particle in the optimized space, and default the individual optimal position value of each particle at this moment to the position value of the current particle.

Step 3: Calculate the fitness value of each particle for each generation. Use roulette to select some particles. In the k th generation, the probability of the m th particle being selected is $1/(f_m^k + 1)$, according to the actual meaning of f_m^k , $0 < 1/(f_m^k + 1) \leq 1$. Randomly match all the selected particles into pairs and perform the next crossover operation (if the number of particles selected is an odd number, the extra particles after the pairing will not be crossed).

Step 4: For the paired particles selected in step 3, perform pairwise crossover operations according to Eqs. (26) and (27). In addition, calculate the fitness value of the particles before and after the crossover according to Eq. (21), keep two of the particles with higher fitness, and eliminate the two particles with lower fitness value to keep the particle population unchanged.

$$\mathbf{l}_{m'_1} = r\mathbf{l}_{m_1} + (1 - r)\mathbf{l}_{m_2}, \quad (26)$$

$$\mathbf{l}_{m'_2} = (1 - r)\mathbf{l}_{m_1} + r\mathbf{l}_{m_2}, \quad (27)$$

where $\mathbf{l}_{m_1}$ and $\mathbf{l}_{m_2}$ are the positions of the two particles that need to be performed crossover; $\mathbf{l}_{m'_1}$ and $\mathbf{l}_{m'_2}$ are the particle positions after the crossover operation; r is the crossover rate, which is a fixed value.

Step 5: In steps 3 and 4, the fitness values of all particles have been calculated. According to Eqs. (22) and (23), calculate the historical minimum fitness value of particle m and the minimum fitness value of the entire population, and update its optimal positions $\mathbf{p}_m^k$ and $\mathbf{pp}^k$;

Step 6: Update the position and velocity of each particle in the PSO algorithm using Eqs. (24) and (25), and re-assign

the values of particles that cross the boundary to keep the particle population unchanged.

Step 7: Check whether the end condition is met. If the number of iterations at the moment reaches the upper limit, or the optimization result meets the accuracy required by the convergence, the iteration is stopped, and the optimal result is output, otherwise return to step 3.

Compared with the traditional PSO, GA-PSO adds selection and crossover operations in steps 3 and 4 to achieve the purpose of controlling the evolution direction and searching for the global optimal solution.

The SVR estimation SOH process based on the GA-PSO optimization algorithm is shown in Fig. 5.

Estimation results and analysis

In this section, GA-PSO algorithm is applied to parameter optimization in SVR, and then SOH estimation accuracy and algorithm convergence speed results of GA-PSO-SVR are compared with other algorithms.

GA-PSO-SVR optimization results

Similar to many other studies [14, 16, 38, 61], Goa et al. 2021c), this paper uses the RMSE and average absolute percentage error (MAPE) as evaluation indicators to evaluate the effectiveness of the SOH estimation model. The calculation formulas for the 2 evaluation indicators are as follows:

$$\text{RMSE} = \frac{1}{O} \sqrt{\sum_{o=1}^O (y_o - \hat{y}_o)^2}, \quad (28)$$

$$\text{MAPE} = \frac{1}{O} \sum_{o=1}^O \left| \frac{y_o - \hat{y}_o}{y_o} \right| \times 100\%, \quad (29)$$

where y_o and $\hat{y}_o$ are the actual and estimated values of the sample, and O is the sample size.

In general, RMSE can well reflect the degree of deviation between the estimated value of the proposed model and the true value. However, in practice, if there are individual outliers with very large deviations, even a small number of outliers will make the evaluation effect of RMSE indicators poor. Therefore, this paper uses MAPE for auxiliary evaluation. The closer the MAPE is to 0, the better the model estimation is.

The first 60% of the total sample is set as the training set, and the last 40% as the test set, that is, for B0005 and B0007, the first 100 cycles are the training set, the last 68 cycles are the test set. And for B0029 and B0031, the first 23 cycles are training set, and the last 16 cycles are the test set. In GA-PSO-SVR, the population size of the algorithm is set to be

100, the crossover rate $r = 0.2$. Because optimal penalty coefficient C and parameter γ are the optimization objectives, the particle search dimension $D = 2$. The algorithm termination condition is that the maximum number of iterations is 40 or the fitness value decreases less than 1×10^{-6} within 30 consecutive generations.

Training is carried out according to the specific steps of GA-PSO-SVR. The iterative result of GA-PSO-SVR algorithm is shown in Fig. 6, and the results of the estimated values and the actual values are shown in Fig. 7.

In the figure, the number of iterations required for B0005, B0007, B0029, and B0031 to reach the optimal solution are 11, 11, 13, 12, respectively. Since the two sets of batteries (i.e., B0005 & B0007, B0029 & B0031) are tested under different conditions, when the algorithm converges, the fitness values within the groups are basically the same, but there are certain differences between the groups. However, it can be concluded that the algorithm basically reaches the optimal when it runs to about the 11th–13th generation.

The optimal parameter values of the estimation method and SOH estimation results of the selected 4 batteries are shown in Table 2:

It can be seen that the estimation accuracy of the two groups of batteries (B0005 & B0007, B0029 & B0031) are basically the same, as well as the optimal values of the parameters. The estimations between the two groups are quite different, which is mainly caused by the different experimental conditions for the two groups of batteries. Thus, on the other hand, the result also proves that the proposed method is applicable to lithium-ion batteries under two different experimental conditions.

Comparisons between GA-PSO-SVR and other algorithms

Evaluating the performance of SOH estimation algorithms is generally by comparing SOH estimation accuracy and algorithm convergence speed. Here, the evaluation indicators adopted for the estimation accuracy are still RMSE and MAPE, which can be found in Eqs. (28) and (29). The smaller the values of RMSE and MAPE, the higher the estimation accuracy. As GA, PSO, and GA-PSO all use population evolution to find the optimal results, the convergence speed is evaluated by the number of iterations required for the algorithm to find the optimal solution. The fewer the number of iterations, the higher the convergence speed.

In this section, the proposed GA-PSO-SVR is compared with the SVR based on particle swarm algorithm (PSO-SVR) and the support vector regression based on genetic algorithm (GA-SVR). These three algorithms use the same training set and test set for SOH estimation, and the final estimation accuracy is compared. Accuracy difference is analyzed.

Fig. 5 Flowchart of SOH estimation method based on GA-PSO-SVR

Fig. 6 The fitness value change with the iterations under GA-PSO-SVR algorithm

The parameter values (C and γ) obtained by PSO-SVR and GA-SVR algorithms are shown in Table 3.

The RMSE and MAPE under the three estimation algorithms are displayed in Tables 4 and 5.

As shown in Tables 4 and 5, the difference in estimation accuracy between GA-SVR and GA-PSO-SVR is small. The average RMSE under GA-SVR and GA-PSO-SVR is 0.41% and 0.40%, respectively. The difference is only 0.01%. In terms of MAPE, the values under the two algorithms are both 0.56%. The estimation error of PSO-SVR is relatively larger than that of the other two algorithms. Compared with GA-SVR and GA-PSO-SVR, the RMSE under PSO-SVR increases by 0.09% and 0.10%, respectively.

Though GA-SVR and GA-PSO-SVR are similar in accuracy, there is a large difference in the number of iterations to find the optimal solution. Taking battery B0005 as an example, the iterative diagrams under GA-SVR and GA-PSO-SVR to find the optimal parameters are displayed in Fig. 8.

Fig. 7 Comparison between the actual values and estimated values

Table 2 Optimal parameter values and SOH estimation results

Parameter	B0005	B0007	B0029	B0031
C	0.4150	0.3847	0.5420	0.5810
γ	1.3950	1.7854	0.8452	0.4984
RMSE (%)	0.28	0.31	0.49	0.52
MAPE (%)	0.41	0.47	0.65	0.72

It can be seen from Fig. 8 that GA-SVR reaches the optimal solution around the 18th generation, while GA-PSO-SVR reaches the optimal solution around the 11th generation. The results demonstrated that GA-PSO-SVR was able to generate faster and more accurate SVR models to obtain the optimal solution, as compared to classical GA-SVR and PSO-SVR.

Conclusions

The aim of this study is to develop a method to estimate the battery SOH. Based on NASA's lithium battery cycle test data set, 12 HFs from the charging current, voltage, and temperature curves are extracted. Then the random forest algorithm is used to find the four most important HFs. A SVR-based

method is established to estimate the SOH of electric vehicle batteries. The GA-PSO algorithm is adopted to optimize the parameters of SVR. The following conclusions can be obtained from this study.

- (i) Some previous studies used Spearman's correlation coefficient or Pearson's correlation coefficient to analyze the correlation between SOH and health factors of lithium-ion batteries. However, in this study, the random forest algorithm is employed to extract the health factors, which can reduce the data dimension and improve the reliability of the estimation method.
- (ii) 4 battery data sets with different charge and discharge currents and different temperatures in the NASA data set are used for the test. The first 60% of battery data is used as the training set, and the last 40% is the test set. After implementing the estimation method proposed in this paper, the average RMSE is 0.40%, and the average MAPE is 0.56%, which proves the reliability of the proposed method.
- (iii) Compared with the PSO-SVR algorithm, the proposed estimation algorithm can decrease the average RMSE by 0.10% and the average MAPE by 0.17%. Compared with the GA-PSO algorithm, the number of iterations

Table 3 Optimal parameter values of PSO-SVR and GA-SVR algorithms (C , γ)

Algorithm	B0005	B0007	B0029	B0031
PSO-SVR	(0.4202, 1.4139)	(0.3601, 1.7611)	(0.5620, 0.8295)	(0.5699, 0.4695)
GA-SVR	(0.4119, 1.3928)	(0.3790, 1.7809)	(0.5339, 0.8348)	(0.5762, 0.4687)

Table 4 Comparison of RMSE of the three algorithms

Algorithm	B0005 (%)	B0007 (%)	B0029 (%)	B0031 (%)	Average (%)
PSO-SVR	0.37	0.38	0.59	0.66	0.50
GA-SVR	0.26	0.34	0.47	0.55	0.41
GA-PSO-SVR	0.28	0.31	0.49	0.52	0.40

Table 5 Comparison of MAPE of the three algorithms

Algorithm	B0005 (%)	B0007 (%)	B0029 (%)	B0031 (%)	Average (%)
PSO-SVR	0.61	0.53	0.84	0.93	0.73
GA-SVR	0.45	0.42	0.69	0.69	0.56
GA-PSO-SVR	0.41	0.47	0.65	0.72	0.56

Fig. 8 Comparison of the number of iterations under GA-SVR and GA-PSO-SVR

of the proposed algorithm can be reduced by 7 generations.

In this study, NASA's lithium battery cycle test data are used to train and evaluate the proposed method. However, there may be differences between the lab test data and real vehicle test data, which would affect the performance of the proposed method. In the next step, real vehicle test data should be adopted to fully evaluate the estimation method. Some foreseeable problems are that data duplication, data abnormalities, and missing data will also occur in the process of real vehicle data collection. Therefore, data preprocessing such as data cleaning and missing value filling will affect subsequent SOH calculations. There are two aspects of influence: (1) the accuracy loss caused by improper preprocessing of real vehicle data; (2) the poor real-time performance of

SOH estimation caused by the real-time data preprocessing. Therefore, although the GA-PSO-SVR proposed in this paper can take into account the convergence speed and estimation accuracy to a certain extent, future work may focus on specific types of real vehicle data. Besides, more attention will be paid to the development of the SOH estimation algorithm combined with the real vehicle data, as well as the improvement of the calculation efficiency of the SOH estimation. The results of this study can be used for dynamic public transit systems [33].

Declarations

Conflict of interest The authors declared that they have no conflicts of interest to this work.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abada S, Marlair G, Lecocq A, Petit M, Sauvart-Moynot V, Huet F (2016) Safety focused modeling of lithium-ion batteries: a review. *J Power Sources* 306:178–192
- Abdollahzadeh B, Gharehchopogh F, Mirjalili S (2021) African vultures optimization algorithm: a new nature-inspired metaheuristic algorithm for global optimization problems. *Comput Ind Eng* 158:107408. <https://doi.org/10.1016/j.cie.2021.107408>
- Ahmed R, El Sayed M, Arasaratnam I, Tjong J, Habibi S (2014) Reduced-order electrochemical model parameters identification and state of charge estimation for healthy and aged Li-ion batteries—Part II: aged battery model and state of charge estimation. *IEEE J Emerg Select Top Power Electron* 2(3):678–690
- Barré A, Deguilhem B, Grolleau S, Gérard M, Suard F, Riu D (2013) A review on lithium-ion battery ageing mechanisms and estimations for automotive applications. *J Power Sources* 241:680–689
- Bie Y, Hao M, Guo M (2021) Optimal electric bus scheduling based on the combination of all-stop and short-turning strategies. *Sustainability* 13:1827
- Bie Y, Ji J, Wang X, Qu X (2021b) Optimization of electric bus scheduling considering stochastic volatilities in trip travel time and energy consumption. *Comput-Aided Civil Infrastruct Eng*
- Chaoui H, Ibe-Ekeocha CC (2017) State of charge and state of health estimation for lithium batteries using recurrent neural networks. *IEEE Trans Veh Technol* 66(10):8773–8783
- Chapelle O, Vapnik V, Bousquet O, Mukherjee S (2002) Choosing multiple parameters for support vector machines. *Mach Learn* 46(1):131–159
- Chen X, Yin R, An Q et al (2021) Modeling a distance-based preferential fare scheme for park-and-ride services in a multimodal transport network[J]. *Sustainability* 13(5):2644
- Cheng AY, Wang Y, Cheng ZL, Cheng XS, Dei F (2018) November. State of charge estimation for batteries using recurrent neural networks. In 2018 Chinese Automation Congress (CAC). IEEE, pp 390–395
- Deng W, Zhao H, Zou L et al (2017) A novel collaborative optimization algorithm in solving complex optimization problems. *Soft Comput* 21:4387–4398. <https://doi.org/10.1007/s00500-016-2071-8>
- Doyle M, Fuller TF, Newman J (1993) Modeling of galvanostatic charge and discharge of the lithium/polymer/insertion cell. *J Electrochem Soc* 140(6):1526
- Fotouhi A, Auger DJ, Propp K, Longo S, Wild M (2016) A review on electric vehicle battery modelling: from Lithium-ion toward Lithium-Sulphur. *Renew Sustain Energy Rev* 56:1008–1021
- Gao K, Yang Y, Li A, Li J, Yu B (2021) Quantifying economic benefits from free-floating bike-sharing systems: a trip-level inference approach and city-scale analysis. *Transp Res Part A Policy Pract* 144:89–103
- Gao K, Yang Y, Qu X (2021) Examining nonlinear and interaction effects of multiple determinants on airline travel satisfaction. *Transp Res Part D Transp Environ* 97:102957
- Gao K, Yang Y, Sun L, Qu X (2020) Revealing psychological inertia in mode shift behavior and its quantitative influences on commuting trips. *Transp Res F Traffic Psychol Behav* 71:272–287
- Gao K, Yang Y, Zhang T, Li A, Qu X (2021) Extrapolation-enhanced model for travel decision making: an ensemble machine learning approach considering behavioral theory. *Knowl-Based Syst* 218:106882
- Guo P, Cheng Z, Yang L (2019) A data-driven remaining capacity estimation approach for lithium-ion batteries based on charging health feature extraction. *J Power Sources* 412:442–450
- Han X, Lu L, Zheng Y, Feng X, Li Z, Li J, Ouyang M (2019) A review on the key issues of the lithium ion battery degradation among the whole life cycle. *eTransportation* 1:100005
- Huang G, Song S, Wu C, You K (2012) Robust support vector regression for uncertain input and output data. *IEEE Trans Neural Netw Learn Syst* 23(11):1690–1700
- Hussein AA (2015) Capacity fade estimation in electric vehicle Li-ion batteries using artificial neural networks. *IEEE Trans Ind Appl* 51(3):2321–2330
- Jiao S, Chen W, Wang J et al (2018) Soil microbiomes with distinct assemblies through vertical soil profiles drive the cycling of multiple nutrients in reforested ecosystems. *Microbiome* 6(1):1–13
- Jokar A, Rajabloo B, Désilets M, Lacroix M (2016) Review of simplified pseudo-two-dimensional models of lithium-ion batteries. *J Power Sources* 327:44–55
- Kennedy J, Eberhart RC (1995) Particle swarm optimization. In: Proceedings of IEEE International Conference on Neural Networks, Perth, 27 November–1 December 1995, 1942–1948. <https://doi.org/10.1109/ICNN.1995.488968>
- Li J, Lyu C, Wang L, Zhang L, Li C (2014) Remaining capacity estimation of Li-ion batteries based on temperature sample entropy and particle filter. *J Power Sources* 268:895–903
- Liu D, Zhou J, Liao H, Peng Y, Peng X (2015) A health indicator extraction and optimization framework for lithium-ion battery degradation modeling and prognostics. *IEEE Trans Syst Man Cybern Syst* 45(6):915–928
- Liu H, Wei Z, He W, Zhao J (2017) Thermal issues about Li-ion batteries and recent progress in battery thermal management systems: a review. *Energy Convers Manag* 150:304–330
- Liu J, Chen Z (2019) Remaining useful life prediction of lithium-ion batteries based on health indicator and Gaussian process regression model. *Ieee Access* 7:39474–39484
- Long B, Xian W, Jiang L, Liu Z (2013) An improved autoregressive model by particle swarm optimization for prognostics of lithium-ion batteries. *Microelectron Reliab* 53(6):821–831
- Ma QH, Zheng Y, Yang WD, Zhang Y, Zhang H (2021) Remaining useful life prediction of lithium battery based on capacity regeneration point detection. *Energy* 234:121233
- Moura SJ, Argomedeo FB, Klein R, Mirtabatabaei A, Krstic M (2016) Battery state estimation for a single particle model with electrolyte dynamics. *IEEE Trans Control Syst Technol* 25(2):453–468
- Ortúzar J (2021) Future transportation: sustainability, complexity and individualization of choices. *Commun Transp Res* 1:100010. <https://doi.org/10.1016/j.commtr.2021.100010>
- Peled I, Lee K, Jiang Y, Dauwels J, Pereira FC (2021) On the quality requirements of demand prediction for dynamic public transport. *Commun Transp Res* 1:100008. <https://doi.org/10.1016/j.commtr.2021.100008>
- Qu X, Wang S (2021) Communications in transportation research: vision and scope. *Commun Transp Res* 1:100001. <https://doi.org/10.1016/j.commtr.2021.100001>

35. Rahimi-Eichi H, Ojha U, Baronti F, Chow M (2013) Battery management system: an overview of its application in the smart grid and electric vehicles. *IEEE Ind Electron Mag* 7(2):4–16. <https://doi.org/10.1109/MIE.2013.2250351>
36. Saha B, Goebel K (2007) Battery data set. NASA Ames Prognostics Data Repository, NASA Ames Research Center, Moffett Field, CA. <http://ti.arc.nasa.gov/project/prognostic-data-repository>
37. Seaman A, Dao TS, McPhee J (2014) A survey of mathematics-based equivalent-circuit and electrochemical battery models for hybrid and electric vehicle simulation. *J Power Sources* 256:410–423
38. Shi X, Wang Z, Li X, Pei M (2021) The effect of ride experience on changing opinions toward autonomous vehicle safety. *Commun Transp Res* 1:100003
39. Song Y, Liu D, Yang C, Peng Y (2017) Data-driven hybrid remaining useful life estimation approach for spacecraft lithium-ion battery. *Microelectron Reliab* 75:142–153
40. Stroe A, Knap V, Stroe D (2018) Comparison of lithium-ion battery performance at beginning-of-life and end-of-life. *Microelectron Reliab* 88–90:1251–1255
41. Sui X, He S, Vilsen SB, Meng JH, Teodorescu R, Stroe DI (2021) A review of non-probabilistic machine learning-based state of health estimation techniques for Lithium-ion battery. *Appl Energy* 300:117346 (ISSN 0306-2619)
42. Tong S, Klein MP, Park JW (2015) On-line optimization of battery open circuit voltage for improved state-of-charge and state-of-health estimation. *J Power Sources* 293:416–428
43. Vapnik V (2013) The nature of statistical learning theory. Springer science & business media, New York
44. Vishnoi NK (2018) Geodesic Convex Optimization: Differentiation on Mani folds, Geodesics, and Convexity. CoRR, abs/1806.06373
45. Wang S, Zhen L, Xiao L, Alttard M (2020) Data-driven intelligent port management based on blockchain. *Asia-Pacific J Oper Res* 38:2040017
46. Wang Y, Wu J, Chen K, Liu P (2021) Are shared electric scooters energy efficient? *Commun Transp Res* 1:100022. <https://doi.org/10.1016/j.commtr.2021.100022>
47. Widodo A, Shim MC, Caesarendra W, Yang BS (2011) Intelligent prognostics for battery health monitoring based on sample entropy. *Expert Syst Appl* 38(9):11763–11769
48. Wu J, Kulcsár B, Selpi QuX (2021) A Modular, Adaptive, and Autonomous Transit System (MAATS): a in-motion transfer strategy and performance evaluation in urban grid transit networks. *Transp Res Part A Policy Pract* 151:81–98
49. Wu J, Zhang C, Chen Z (2016) An online method for lithium-ion battery remaining useful life estimation using importance sampling and neural networks. *Appl Energy* 173:134–140
50. Xiong R, Zhang Y, Wang J, He H, Peng S, Pecht M (2018) Lithium-ion battery health prognosis based on a real battery management system used in electric vehicles. *IEEE Trans Veh Technol* 68(5):4110–4121
51. Xu X, Rong H, Trovat M, Liptrott M, Bessi N (2018) CS-PSO: chaotic particle swarm optimization algorithm for solving combinatorial optimization problems. *Soft Comput* 22:783–795
52. Xu Y, Zheng Y, Yang Y (2021) On the movement simulations of electric vehicles: a behavioural model-based approach. *Appl Energy* 283:16356
53. Yan R, Wang S, Peng C (2021) An artificial intelligence model considering data imbalance for ship selection in port state control based on detention probabilities. *J Comput Sci* 48:101257. <https://doi.org/10.1016/j.jocs.2020.101257>
54. Yao B, Chen C, Cao Q, Jin L, Zhang M, Zhu H, Yu B (2017) Short-term traffic speed prediction for an urban corridor. *Comput-Aided Civil Infrastruct Eng* 32(2):154–169
55. Yu B, Wang H, Shan W, Yao B (2018) Prediction of bus travel time using random forests based on near neighbors. *Comput-Aided Civil Infrastruct Eng* 33(4):333–350
56. Yu S, Wei YM, Wang K (2012) A PSO–GA optimal model to estimate primary energy demand of China. *Energy Policy* 42:329–340
57. Zhang C, Allafi W, Dinh Q, Ascencio P, Marco J (2018) Online estimation of battery equivalent circuit model parameters and state of charge using decoupled least squares technique. *Energy* 142:678–688
58. Zhang L, Wang S, Qu X (2021) Optimal electric bus fleet scheduling considering battery degradation and non-linear charging profile. *Transp Res Part E Transp Logist Rev* 154:102445
59. Zhang L, Zeng Z, Qu X (2021) On the role of battery capacity fading mechanism in the lifecycle cost of electric bus fleet. *IEEE Trans Intell Transp Syst* 22(4):2371–2380
60. Zhang W, Zhao H, Xu M (2021) Optimal operating strategy of short turning lines for the battery electric bus system. *Commun Transp Res* 1:100023. <https://doi.org/10.1016/j.commtr.2021.100023>
61. Zheng (2021) Reasons, challenges, and some tools for doing reproducible Transportation Research. *Commun Transp Res* 1:100004
62. Zhou D, Al-Durra A, Zhang K, Ravey A, Gao F (2018) Online remaining useful lifetime prediction of proton exchange membrane fuel cells using a novel robust methodology. *J Power Sources* 399:314–328
63. Zou C, Zhang L, Hu X, Wang Z, Wik T, Pecht M (2018) A review of fractional-order techniques applied to lithium-ion batteries, lead-acid batteries, and supercapacitors. *J Power Sources* 390:286–296

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.