

Received May 2, 2020, accepted May 16, 2020, date of publication May 20, 2020, date of current version June 3, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2995899

# Novel Lithium-Ion Battery State-of-Health Estimation Method Using a Genetic Programming Model

HANG YAO<sup>1</sup>, XIANG JIA<sup>1</sup>, QIAN ZHAO<sup>2</sup>, ZHI-JUN CHENG<sup>1</sup>, AND BO GUO<sup>1</sup>

<sup>1</sup>College of Systems Engineering, National University of Defense Technology, Changsha 410073, China

<sup>2</sup>College of Information and Communication, National University of Defense Technology, Xi'an 710106, China

Corresponding author: Xiang Jia (jiaxiang09@sina.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 71801219, and in part by the Natural Science Foundation of Hunan Province under Grant 2019JJ50730.

**ABSTRACT** State-of-health (SOH) is a health index (HI) that directly reflects the performance degradation of lithium-ion batteries in engineering, but the SOH of Li-ion batteries is difficult to measure directly. In this paper, a novel data-driven method is proposed to estimate the SOH of Li-ion batteries accurately and explore the relationship-like mechanism. First, the features of the battery should be extracted from the performance data. Next, by using the evolution of genetic programming to reflect the change in SOH, a mathematical model describing the relationship between the features and the SOH is constructed based on the data. Additionally, it has strong randomness in the formula model, which can cover most of the structural space of SOH and features. An illustrative example is presented to evaluate the SOH of the two batches of Li-ion batteries from the NASA database using the proposed method. One batch of batteries was used for testing and comparison, and another was chosen to verify the test results. Through experimental comparison and verification, it is demonstrated that the proposed method is rather useful and accurate.

**INDEX TERMS** Genetic programming, Li-ion battery, state-of-health (SOH), prognostic and health management.

## I. INTRODUCTION

Currently, prognostic and health management (PHM) has been widely studied in industrial management to improve industrial production efficiency and ensure production safety [1], [2]. As an energy storage system, Li-ion batteries provide the most reliable power for various kinds of important equipment in different fields, such as satellite, aerospace and electrified vehicles. Therefore, it is extremely significant to consider the real-time accurate health status of Li-ion batteries to ensure the reliability of equipment [3].

With the development of monitoring technology, using Li-ion battery monitoring data to study PHM has become one of the most feasible and effective approaches. The SOH reflects the current capability of a battery to store and supply energy relative to that at the beginning of its life, so it is an indicator to evaluate the degradation level of batteries [4]. Unfortunately, achieving an accurate estimate of the SOH

is a challenging task. Due to the limitations of monitoring technology, it is difficult to directly measure the SOH of Li-ion batteries. To solve this problem and realize further management of Li-ion batteries, researchers have made many contributions. In general, existing methods mainly include the model-based methods and data-driven methods [5]. The model-based methods attempt to establish physical models or use mathematical representations to describe the SOH degradation of Li-ion batteries. Although experimental methods require a large number of experiments to analyse battery ageing behaviour, they can be used to study ageing mechanisms, providing a theoretical basis for model-based methods such as the extended Kalman filter (EKF) [6]–[9], multiscale EKF [10], wiener process [11], and particle filter (PF) [12]–[15]. Although Li-ion battery life can be accurately predicted by these methods, it is difficult to track the SOH degradation process of the battery in details.

One of the most widely used data-driven techniques is incremental capacity/differential voltage (IC/DV) analysis. IC/DV analysis has been proven to be a powerful tool for

The associate editor coordinating the review of this manuscript and approving it for publication was Jiajie Fan<sup>1</sup>.

battery capacity estimation [16]. Based on this method, the voltage plateaus on charging/discharging curves can be transformed into clearly identifiable peaks on IC/DV curves. Each peak of the curve represents a specific electrochemical process taking place in the cell and can be characterized by features such as the intensity and position. These peak features are closely related to battery capacity fade and can be used as indicators for SOH estimation. Weng *et al.* [17] estimated the battery SOH by relating it to the peak intensity of IC curves. Li *et al.* [18] established a linear regression relationship between battery capacity and the peak position on IC curves. However, IC/DV curves are sensitive to measurement noise inherent in battery systems [18], [19]. Accordingly, proper smoothing methods have to be proposed for obtaining smooth curves that facilitate the identification and evaluation of IC/DV curve features.

The fuzzy logic model is also a data-driven method that used to describe the performance degradation and calculate the SOH of Li-ion batteries [20]. In this model, the number of cycles should be considered as the feature parameter of Li-ion batteries, and the relationship between the feature parameter and SOH of Li-ion batteries can be described as

$$y_{fit} = a_0 + a_1 e^{-\left(\frac{x}{a_1}\right)^{\beta_1}} + a_2 e^{-\left(\frac{x}{a_2}\right)^{\beta_2}}, \quad (1)$$

where  $x$  represents the number of cycles and  $y$  is the value of SOH. The fuzzy logic computes the HI using a fitting curve with an error between 5% and 10%. Because many Li-ion battery parameters change as the capacity decreases, it is difficult to reflect the SOH of Li-ion batteries using the aforementioned performance degradation model. Simultaneously, only the state of the overall trend of the Li-ion battery can be given, resulting in the inaccuracy of the model.

To solve the problem of estimation accuracy, many machine learning methods have been used to estimate the SOH of Li-ion batteries and have achieved good results, such as neural networks (NNs) [21], support vector machine (SVM) [17], [22], and Gaussian process regression (GPR) [23], [24]. The artificial neural network is a data-driven method widely used in the research of battery performance degradation. Hussein [25] used an artificial neural network to achieve capacity fade in Li-ion batteries for electric vehicles. You *et al.* [26] designed a recurrent neural network-based model to estimate the SOH in a more realistic environment. However, it has the shortcomings of poor generalization ability, discretization of structure, and low convergence speed. SVM and RVM have strong capabilities in solving small sample, nonlinear and local minimum values. Therefore, they have achieved good results in the SOH estimation and prediction of Li-ion batteries. As a result, these methods have been studied deeply in PHM. Zhang [27] denoised by empirical mode decomposition and then used multi-kernel relevance vector machine to achieve battery capacity prediction. Dong *et al.* [28] estimated the SOH of battery and predicted the remaining service life by supporting a vector regression particle filter. Li *et al.* [29] established a multistep prediction model based on average entropy to predict SOH and RUL.

However, the disadvantages of SVM and RVM in these aspects cannot be ignored: usually, it is a time-consuming task for tuning the soft margin parameter  $C$  through cross-validation methods, sparsity cannot always be achieved, and a high number of support vectors are thus obtained. GPR is a nonparametric model that uses a Gaussian process prior to regression analysis of data. Zhou *et al.* [30] realized the online estimation of Li-ion battery capacity by combining EKF with the GPR model. Simultaneously, the performance of the GPR model depends on the chosen covariance function and its parameters, and the long-term prediction error is large. Although these methods do not require a deep understanding of battery ageing and its associated degradation mechanisms, it is necessary to collect a large amount of data from previous experiments to train the estimation models, which limits the wide use of these methods. Specifically, these methods need to choose the appropriate kernel function and parameters according to the actual situation, and it is not an easy task to achieve the autonomy of the kernel function. Furthermore, the above data-driven method cannot clearly know the potential relationship between the features and SOH.

The remainder of this paper is organized as follows. Section II describes some of the current problems in the SOH estimation of Li-ion batteries. In Section III, the process of obtaining an SOH calculation formula by a genetic programming model is elaborated. Section IV gives experimental verification and comparison of BP neural networks. Finally, this paper is concluded in Section V.

## II. PROBLEM STATEMENT

After an extensive literature review, there are many kinds of battery SOHs that can be reflected in the numerous monitoring parameters of batteries. The SOH of a battery can only be monitored by mechanisms directly, but it is hard to monitor due to the limitation of practical applications. Especially when we find it difficult to know how it works, it is a meaningful issue to choose the parameters that can best reflect the change in battery SOH and construct the corresponding SOH estimation formula such as mechanism.

Genetic programming (GP) model is a very powerful supervised machine learning algorithm and has been successfully applied to classification and regression in many different fields [31], such as network identification [32], fault detection [33] and task scheduling planning [34]. This method has a good ability to approximate nonlinear relationships and is robust to outliers. Although it has good regression performance and has been well used in other areas, such as strategy optimization and feature selection, no similar work has been done to date in estimating the SOH of batteries. Reference [35] used multi-objective GP to synthesize human-understandable HI from sequences of voltages, currents and temperatures streamed via on-vehicle sensors. In paper [36], the GP was used to address the challenge of automatically discovering advanced features, which can well capture fault progression. However, these studies did not give a concise

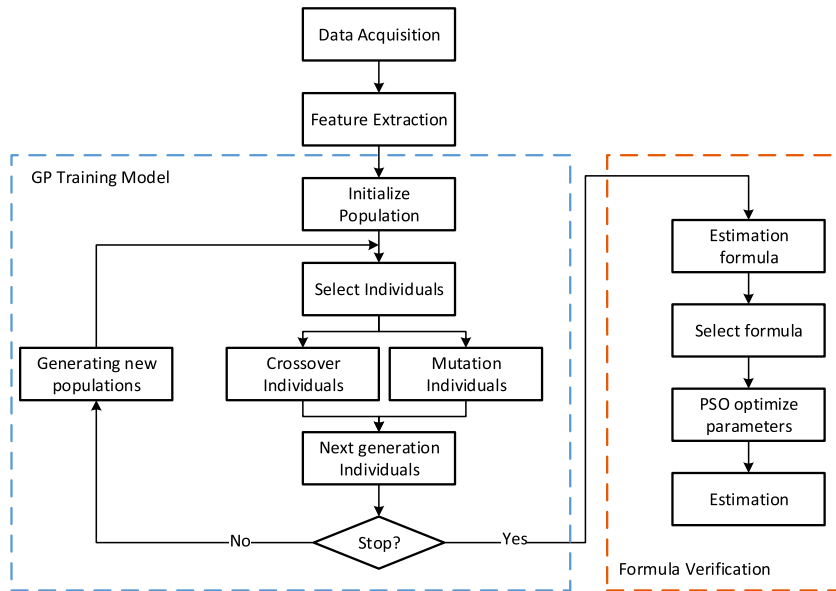


FIGURE 1. The GP model.

model expression to estimate HI and explore the inherent relationships of HI and features.

An accurate multiple regression model is sought for battery SOH estimation, which is constructed through various function changes and different formulas. The purpose is to obtain an accurate and simple battery SOH estimation model. Motivated by this, this paper proposes a method of autonomously determining the regression formula based on the GP regression method to diagnose Li-ion battery SOH. This proposed work aims to fill this gap by using the GP mode for estimating the SOH of Li-ion batteries accurately and exploring the relationship between monitoring parameters, such as current, voltage, temperature and SOH.

This method has several notable features of SOH estimation. First, it has strong randomness in the formula model, which can cover most of the structural space of SOH and features. Next, it has good robustness for the interference of significant noise and input-independent features present in the predictor variables. In addition, compared to other machine learning models, it is only necessary to determine the parameters without having to train the model multiple times.

The SOH is an important index for battery health management. To evaluate the estimation results, the SOH of Li-ion batteries must be defined. Therefore, it is important to understand the definition of SOH clearly, which can be generally defined as [1]:

$$SOH = \frac{C_{actual}}{C_{norm}} \times 100\%, \quad (2)$$

where  $C_{actual}$  represents the maximum practical capacity as measured from the operating battery at the current time.  $C_{actual}$  may fade over time due to the effect of battery ageing.  $C_{norm}$  is the rated capacity from battery manufacturers. We assume that the SOH of Li-ion batteries can be repre-

sented by an unknown function:

$$F_{soh} = f(g_1, g_2, \dots, g_n). \quad (3)$$

where  $F_{soh}$  is the estimated SOH of the Li-ion battery and  $g_i(i = 1, 2, \dots, n)$  are the features extracted from the Li-ion battery monitoring data. Using the genetic programming model, Eq. (3) can be constructed. It has the obvious advantage of avoiding the fixed-function model framework so that the model can independently explore the functional relationship between features and the SOH of Li-ion batteries.

### III. METHODOLOGY

#### A. GENETIC PROGRAMMING (GP)

The genetic programming flowchart is shown in Fig. 1. The extracted features and the SOH of Li-ion batteries are input into the genetic programming model. Like the genetic algorithm (GA), the GP is also one kind of machine learning method. Usually, GA optimizes a value, while the difference of GP is that the optimal individual is a strategy or function. First, the initial calculation formula population is randomly generated, and then the individual selection is performed according to the fitness function; further, the intersection and variation are performed. Finally, the evolution is terminated according to the judgement condition while obtaining the optimal formula individually. Similar to GA, GP has the same important elements as follows [36]:

- Randomly generate an initial genetic population;
- Require a training set and fitness function;
- Evaluate the individual’s viability in the population according to the fitness function, and then perform individual screening;
- Individuals in the population perform similar gene manipulations to achieve crossover and mutation;

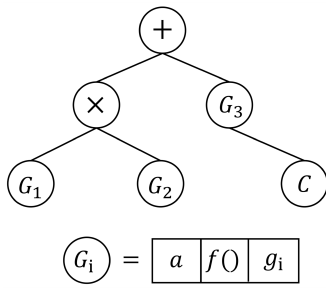


FIGURE 2. GP individual example.

- Termination controllable.

According to the GP model and Eq. (3), a function consists of an independent variables, dependent variables, operators and coefficients. Thus, the extracted features can be linked by operators and coefficients to obtain the calculation formula of SOH for Li-ion batteries. First, a large number of first-generation individuals need to be randomly generated. Each individual in this genetic programming represents a function. In addition, for the independent variables, the individual in the GP model also includes a series of mathematical operations, such as plus, minus, multiplication, division, square, square root, exponential operation and logarithmic operations. An individual coding can be represented by a tree structure, and an illustrative individual is shown in Fig. 2. For this individual, the node  $C$  is a constant, and the node  $G_i$  represents the factor that is a function of the extracted features  $g_i$  in the form of production (logarithm, square, square root, etc.) and constant  $a$ , as

$$G_i = a \cdot f(g_i). \tag{4}$$

Hence, the specific formula represented by the tree structure in Fig. 2 for an illustrative individual is

$$F_{soh} = G_1 \times G_2 + G_3 + C. \tag{5}$$

The depth of the tree structure is defined as the number of factors in an individual. It should be noted that for each individual in the GA model, the tree depth is set to 3, and each node is randomly generated in this paper.

Genetic operations such as crossover and mutation are performed on the node in the tree structure for each individual to determine the model of SOH. After performing the selection operation, it is determined whether the cross operation is performed based on the crossover probability. If there is no intersection, the individual will not change; if it is crossed, using individuals F1 and F2 as examples, the specific operation is shown in Fig. 3. Individuals F1 and F2 are randomly selected, and the number of nodes is randomly selected from individuals F1 and F2 to cross to obtain a new individual. As shown in Fig. 3, nodes  $G_3$  and  $G_4$  in the two individuals are randomly selected and interchanged to obtain two new individuals  $F'1$  and  $F'2$ . When the population of individuals is crossed, the variation in the individual population is simultaneously performed according to the process

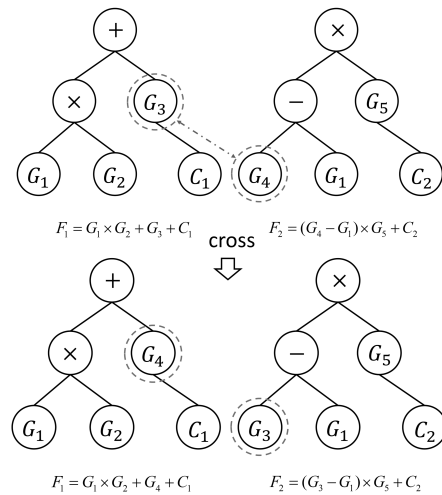


FIGURE 3. An example of a crossing operation.

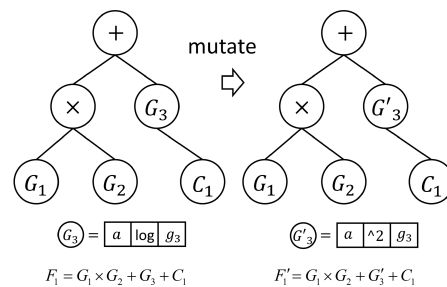


FIGURE 4. An example of a mutating operation.

of genetic programming, as shown in Fig. 1. The individual F1 is randomly selected, and whether the mutation operation is performed is determined or not depends on the mutation probability. If the mutation is not performed, the individual F1 remains the same. In contrast, the nodes in the individual are randomly selected, and the whole node or a part is selected to be mutated. In the example in Fig. 4, for individual F1 selection, the logarithmic operation in node  $G_3$  is mutated to a square operation, resulting in a new individual  $F'1$ . Among them, individuals who have undergone crossover and mutation can be repeatedly crossed and mutated.

### B. FITNESS FUNCTION DESIGN

To minimize the error between the estimations and true values of the SOH, the root mean square error (RMSE) is proposed to measure the individual's fitness as

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (SOH_i - F_i)^2}, \tag{6}$$

where  $SOH_i$  and  $F_i$  represent the true values and estimations of the SOH of the Li-ion battery during the  $i$  cycle, respectively, and  $i = 1, 2, \dots, N$ .

The population individual repeats the selection, crossover, and mutation operations. When the fitness of individual is less

than the threshold or the evolutionary generations reach the maximum iteration numbers, the corresponding individual is the optimal one. The evolution is terminated accordingly.

When the optimal individual is obtained, the model  $F = f_o(g_1, g_2, \dots, g_n)$  can be determined to present the formula between the performance feature parameters and SOH of Li-ion batteries. According to the determined formula, the SOH of Li-ion batteries could be estimated and predicted with newly monitored performance features  $g_{new}$ .

#### IV. EXPERIMENTS AND VERIFICATION

In this section, the source and feature extraction of the performance degradation test data of Li-ion batteries is first introduced to ensure the repeatability and integrity of the algorithm and model. Then, the SOH estimation result of the Li-ion battery obtained by the genetic programming model is given, which is compared with the BP neural network model to verify the applicability and efficiency of the GP model.

##### A. EXPERIMENT DATA AND VERIFICATION DESIGN

In this paper, the test results of commercially available Li-ion 18650-sized rechargeable batteries obtained in the NASA open source database are used for verification. Three different Li-ion battery operating test conditions, including charge, discharge and impedance, can be specified as follows [37]:

**Charging process:** The lithium-ion battery is charged in constant current (CC) mode (current 1.5). After that, the lithium-ion battery continues to charge in constant voltage (CV) mode until the charging current drops to 20 mA, which indicates the end of charging.

**Discharge process:** the lithium-ion battery is discharged in CC mode until the voltage reaches a certain set cut-off voltage.

**Impedance measurement:** The impedance of the lithium-ion battery was measured by EIS frequency scanning in the range of 0.1 Hz ~5 kHz.

The performance of the Li-ion battery degraded due to cyclic charge and discharge, which may be continuous or discontinuous due to impedance measurements during the actual capacity attenuation test. In this paper, two batches of lithium batteries under different experimental conditions are considered:

(1) #5, #6 and #7 as a group are used to train the GP model. From the data, 50% -70% is selected for training commonly used in machine learning, and the rest of the data are used as a preliminary formula validation.

(2) #33, #34 and #36 are used to verify the formula results. Twenty percent of the data are selected for parameter optimization, and the formula is verified by realizing an accurate estimation of the SOH of Li-ion batteries.

All experiments were performed at room temperature (24°C). The specific battery situation is shown in Table 1.

##### B. FEATURE EXTRACTION

The charge was performed at a CC of 1.5 A until the voltage reached 4.2 V, and then it continued charging at a CV until the

TABLE 1. Results for training 120 cycles.

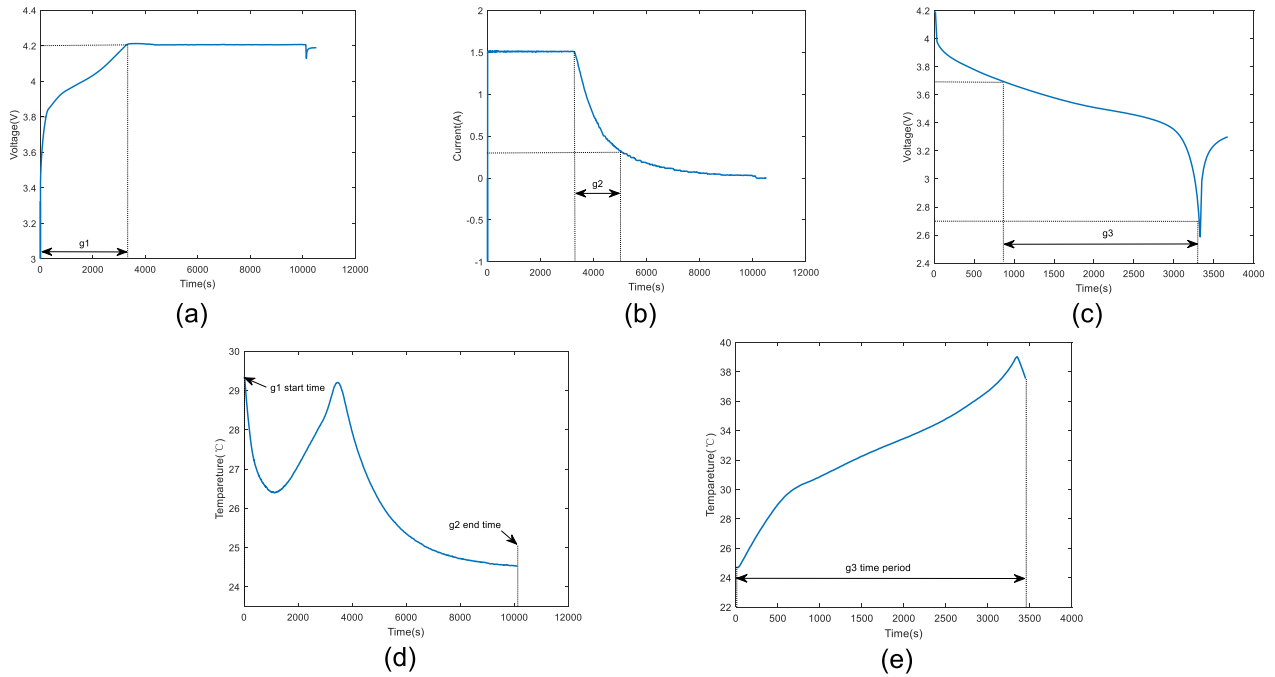
Battery	Discharge current	Cut-off voltage	Number of cycles
#5	2A	2.7V	168
#6	2A	2.5V	168
#7	2A	2.2V	168
#33	4A	2V	197
#34	4A	2.2V	197
#36	2A	2.7V	197

charge current dropped to 20  $\mu$ A. Five performance features can be extracted from the Li-ion battery data, and more details of extracting features can be seen in [15] and [38]:

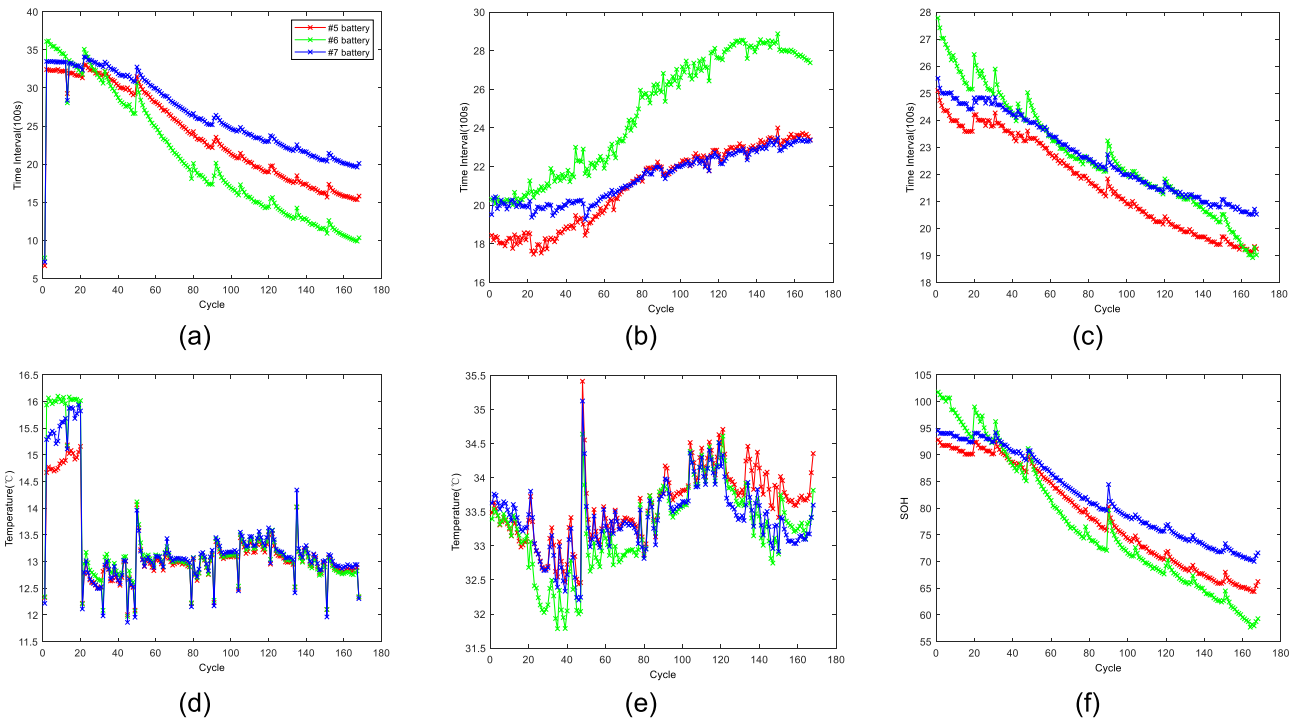
An equal voltage rise charging interval (g1) represents the time it takes for the voltage to rise from a lower value to a higher value during CC charging. The equal electric current drop charging interval (g2) is the time elapsed during a CV charging process in which the current drops from a higher value to a lower value. An equal voltage drop discharge time interval (g3) is extracted from the discharge voltage curve of the battery. The average charge battery temperature (g4) represents the average current temperature between the start time of g1 and the end time of g2. The average battery discharge temperature (g5) is the average battery temperature during the g3 period. In the performance features of Li-ion battery extraction, g1 is set to the time interval of the Li-ion battery CC charging voltage 2.7 V rising to 4.2 V; g2 is set to the Li-ion battery CV discharge current 1.5A falling to 0.3A time interval; g3 is set to a time interval in which the discharge voltage is reduced from 3.7 V to 2.7 V; the performance feature data are regulated to within 0-100, so the value of performance feature g1, g2, and g3 are respectively divided by 100. Finally, the performance features (g1, g2, g3, g4, g5) of the Li-ion battery are obtained.

##### C. SOH ESTIMATION RESULTS AND FORMULAS

The five performance features of batteries #5, #6, and #7 and the real SOH are shown in Fig. 6. Because the optimal function may be more than one, the function form is endless. To reduce the computational complexity and time consumption, the individual tree depth in this paper is set to 4 to 6. The factor  $G_i$  of the function only considers the square, square root, logarithm and no function operation. The range of the factor  $G_i$  and the constant coefficient of the function is  $(-2, 2)$ . The performance feature data of the Li-ion battery is 168 cycles. The first 120 cycle feature data and the real SOH value of the Li-ion battery are used as the training set, and the remaining 48 cycle feature data are used as the test set. The performance feature data are put into the genetic programming model for training. The number of populations of the model is set to 1000, the crossover probability is 0.75, the mutation probability is 0.05, and the number of iterations is 500. The BP network uses a feedforward neural network with 10 hidden neurons and the Levenberg-Marquardt algorithm.



**FIGURE 5.** A #5 of the battery (a) charge voltage and g1 extraction during the charge cycle; (b) current curve and g2 extraction during the charge cycle; (c) discharge voltage curve and g3 extraction during the discharge cycle; (d) charging temperature curve; (e) discharge temperature curve from the second cycle.

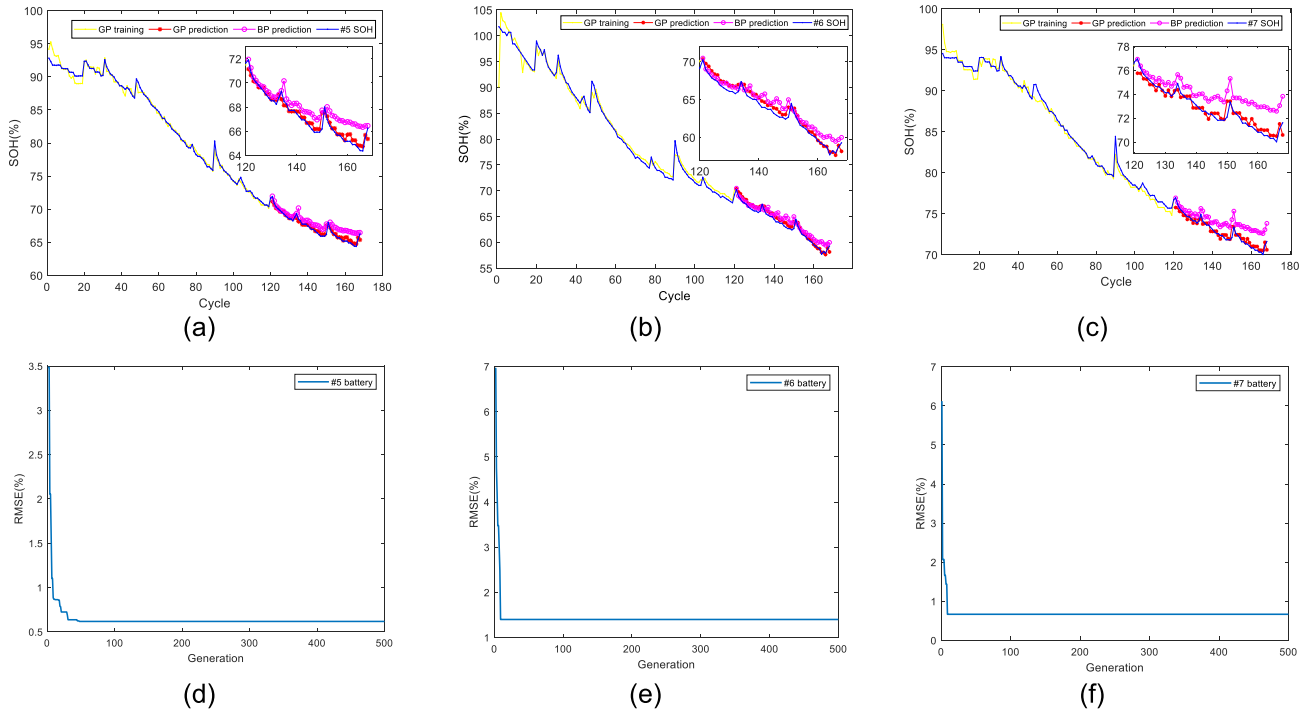


**FIGURE 6.** #5, #6 and #7 battery SOH (f) and 5 performance features (a) g1; (b) g2; (c) g3; (d) g4; (e) g5; trend charts.

The maximum number of trainings is 1000, the training accuracy is 0.001 and the learning rate is 0.01. Figs. 7(a-c) show the three sets of optimal estimation results for batteries #5, #6, and #7 evolved through the genetically regulated model and the BP neural network estimates obtained by the same data

training. Figs. 7(d-f) depict the fitness optimization process for the optimal individual of batteries #5, #6, and #7.

It can be seen from Fig. 7 that after 120 cycles of training, the process of GP model training to find the optimal function converges very quickly, and the optimal individual has been



**FIGURE 7.** The SOH estimation of (a) battery #5 (b) battery #6 (c) battery #7, and the process of fitness optimization of (d) battery #5 (e) battery #6 (f) battery #7.

**TABLE 2.** Results for training 120 cycles.

Battery	$F_{soh}$	GP-RMSE (%)	BP-RMSE(%)
#5	$1.4\sqrt{g_1} + 0.73g_3^{3/2} - 0.44\sqrt{g_4} \log g_1 + 0.47$	0.3827	1.0674
#6	$0.32g_1 \log g_1 + 1.86g_3 + 1.3\sqrt{g_3} + 0.04g_3^2 - 0.48$	0.6628	1.0878
#7	$0.76g_3^{3/2} + 0.73 \log g_2 - 1.71 \log g_4 + 0.41 \log g_5 + 0.92$	0.4271	1.5104

found in fewer than 100 iterations. In addition, the GP model fits well with the SOH of the #5, #6, and #7 Li-ion batteries. Compared to the BP neural network estimation results, the GP model estimates the remaining 48 cycles of SOH more accurately. According to the formulas for obtaining the optimal individual representative of the three Li-ion batteries, calculate the RMSE of the real SOH of Li-ion batteries, as shown in Table 2.

Under this training condition, the GP model estimated the RMSE of the SOH of batteries #5, #6 and #7 to be <1%, while the BP neural network estimated the RMSE of SOH to be >1%. The SOH estimation formulas of Li-ion batteries #5, #6 and #7 obtained by the GP model are all different. The number of performance features and the arithmetic function of the factor  $G_i$  in the formula are different, which shows that the formula obtained by the GP model has uncertainty. Similar to the principle of finding the best individual in the GP model, the group passes through the intersection and variation between individuals and then gradually realizes the evolutionary group. The obtained optimal formula is a feasible solution for estimating the SOC of Li-ion batteries.

To further verify the accuracy of the GP model, the optimal formula was obtained by training the 90, 100 and 110 cycles of the #5, #6 and #7 Li-ion batteries, respectively, and then the remaining SOH estimation was performed.

The three Li-ion batteries of #5, #6 and #7 were trained in 90, 100, 110, and 120 cycles, respectively, and then the optimal formula was obtained under different conditions, as shown in Table 3. Compared with the estimation results of the BP neural network, the RMSE of batteries #5 and #7 based on the GP model is 1% lower, while the RMSE of the BP neural network model is approximately 2%. The RMSE of the estimated SOH is much smaller than that of the BP neural network. In the #5 battery estimation results, the GP model is significantly better than the BP neural network model. The RMSE of battery #6 is estimated to be smaller than that of the BP neural network under different training periods. Therefore, we can conclude that the GP model is generally better than the BP neural network model for estimating the SOH of this group of Li-ion batteries. In addition to the estimated SOH results of battery #6 during training for 90 cycles, the RMSE of the SOH and the true value of the Li-ion

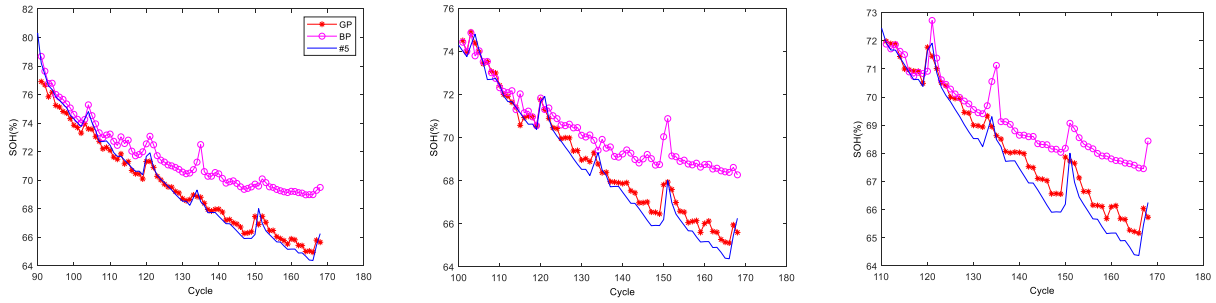


FIGURE 8. #5 Li-ion battery for SOH estimation of 90, 100, 110 cycles of training.

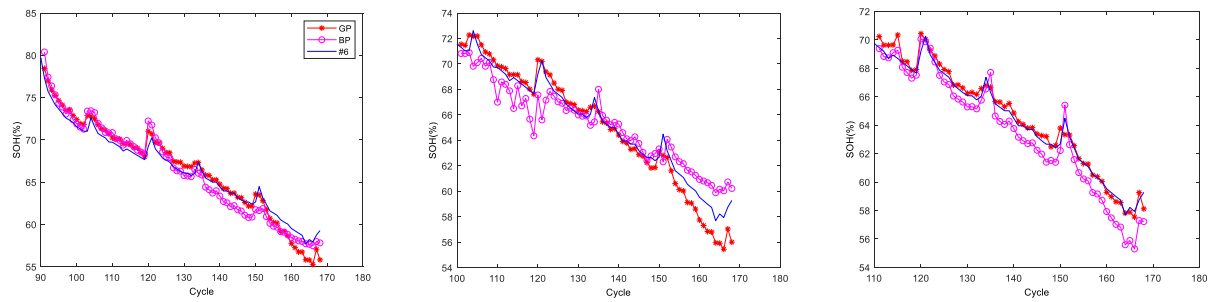


FIGURE 9. #6 Li-ion battery for SOH estimation of 90, 100, 110 cycles of training.

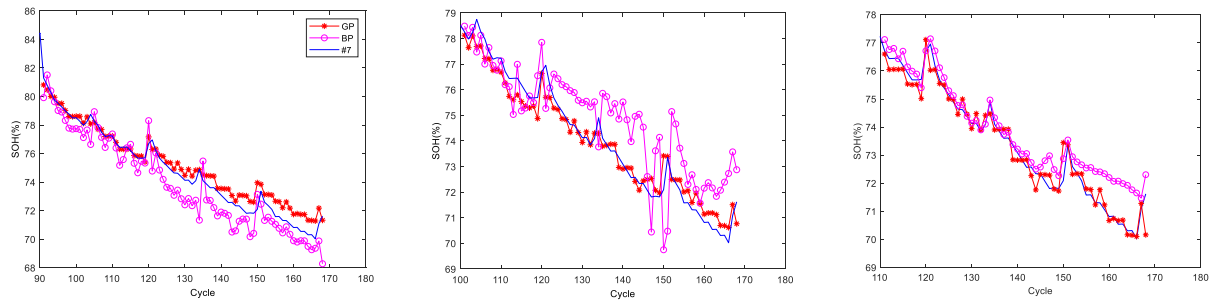


FIGURE 10. #7 Li-ion battery for SOH estimation of 90, 100, 110 cycles of training.

battery were estimated to be below 1% by the GP model. The occasional results with fewer training cycles are more accurate than those trained with more cycles, which needs to be further explained. The reason for this phenomenon is the certain randomness in the cross-mutation process, but the estimation accuracy generally improves as the number of training increases. Based on the above discussion, the GP model can accurately estimate the SOH of Li-ion batteries.

The optimal formulas obtained for the #5, #6 and #7 Li-ion batteries at different cycles were analysed. Feature selection is automatically realized in the iterative optimization process of the GP model, and the performance features contained in the final optimal formula have a certain correlation with the SOH of Li-ion batteries. Numerically, features  $g_1$ ,  $g_2$  and  $g_3$  appear more frequently in the formula than performance features  $g_4$  and  $g_5$ . Thus, the SOH contribution for the formula is greater than the performance features  $g_4$  and  $g_5$ , so the features  $g_1$ ,  $g_2$  and  $g_3$  are more able to reflect the change in

the SOH of Li-ion batteries. In particular, the optimal formula obtained in the #7 battery training 110 cycle and RMSE is very small.

$$F_{soh} = -0.02g_3^{3/2} + 0.2g_3 - 1.47 \quad (7)$$

By analysing the optimal formula, the GP model can help analyse the influence of performance features on the SOH of Li-ion batteries and help to summarize the functional relationship between the SOH and performance features of Li-ion batteries. It is proven that the performance feature  $g_3$  has a certain functional relationship with the SOH of Li-ion batteries, and a function form of the performance feature  $g_3$  with the SOH of Li-ion batteries can be obtained. On the basis of this, SOH can be represented by this model:

$$F_{soh} = a_0g_3^{3/2} + a_1g_3 + c \quad (8)$$

The selected # 33, # 34 and # 36 batteries were used to verify the obtained formula 8. The first 20% of the data (40



TABLE 3. Results for training 90, 100, 110 and 120 cycles.

Battery	Train cycle	$F_{soh}$	GP-RMSE(%)	BP-RMSE(%)
#5	90	$1.34\sqrt{g_1} - 0.74g_2 + 3.67g_3 + 1.6\sqrt{g_3} - 1.76$	0.4773	2.4758
	100	$2.41\sqrt{g_1} - 0.2\log g_1 + 0.51g_2 + 0.12g_3^2 - 0.33$	0.5261	2.1201
	110	$1.5\sqrt{g_1} + 1.86g_3 + 0.07g_3^2 - 0.99\log g_3 - 1.71$	0.5715	1.7247
	120	$1.4\sqrt{g_1} + 0.73g_3^{3/2} - 0.44\sqrt{g_4} \log g_1 + 0.47$	0.3827	1.0674
#6	90	$1.45\sqrt{g_1} + 1.13\sqrt{g_2} + 0.12g_3^2 + 0.12\log g_3 - 0.51$	1.1216	1.2086
	100	$2.63\sqrt{g_1} + 0.15g_3 + 1.47\sqrt{g_3} + 0.1g_3^2 + 0.86\log g_4 - 0.28$	0.9949	1.3759
	110	$4.46g_3 - 0.98g_2 - 1.13$	0.5053	1.0844
	120	$0.32g_1 \log g_1 + 1.86g_3 + 1.3\sqrt{g_3} + 0.04g_3^2 - 0.48$	0.6628	1.0878
#7	90	$0.73\sqrt{g_1} - 0.32\log g_1 + 1.96g_3 + 0.07g_3^2 + 0.23\sqrt{g_4} - 0.05$	0.6640	1.1834
	100	$1.84\sqrt{g_1} + 0.41g_2 + 1.62\sqrt{g_3} + 1.11g_3^2 - 1.65$	0.4885	1.4888
	110	$-0.02g_3^{3/2} + 0.2g_3 - 1.47$	0.3982	0.6939
	120	$0.76g_3^{3/2} + 0.73\log g_2 - 1.71\log g_4 + 0.41\log g_5 + 0.92$	0.4271	1.5104

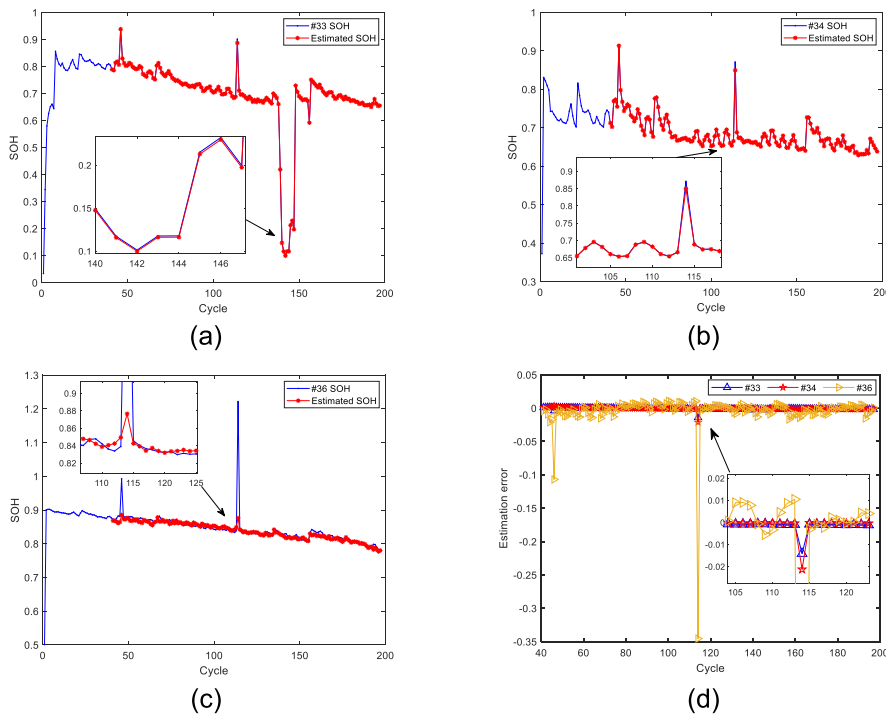


FIGURE 11. Li-ion battery for SOH estimation of 40 cycles of training: (a) #33, (b) #34, (c) #36 and (d) estimation error.

cycles) were used to train the particle swarm algorithm to optimize the parameters  $a_0, a_1, c$ .

It is obvious that the SOH estimation effect for batteries #33 and #34 is very good according to Fig. 11 and Table 4. Both the RMSE and MAE of the estimated results are less than 0.2%. The RMSE is close to 3%, and the MAE is smaller for the estimation of #36. According to Table 4, the estimation errors of # 33 and # 34 are small, and the estimation error of # 36 is relatively large. However, there is a good estimation

result for the overall change trend of the SOH of battery #36, which can be confirmed at MAE = 0.81%. There are two large estimation errors of battery #36, which leads to a significant increase in the estimated RMSE. It can be seen from Fig. 11(c) that the actual SOH value is greater than 1, so it is reasonable to suspect that the monitoring data at these two places are abnormal. It can be concluded that Eq. (8) can achieve accurate estimates of # 33, # 34 and # 36 battery SOH.

TABLE 4. Results for the #33\_#34\_#36 battery.

Battery	$a_0$ $a_1$ $c$	RMSE(%)	MAE(%)
#33	[0.0004,0.0545,0.0045]	0.17	0.11
#34	[0.0009,0.0517,0.0175]	0.19	0.08
#36	[0.0033,0.0154, 0.0160]	2.97	0.81

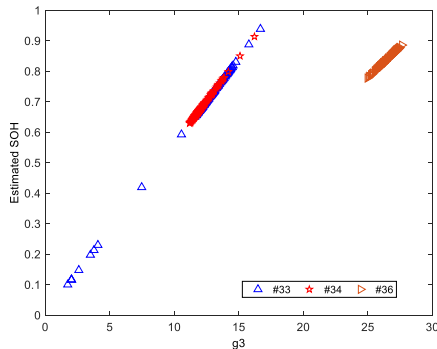


FIGURE 12. The result for g3 estimate the SOH of #33, #34 and #36.

TABLE 5. Results for two methods.

Battery #5	RMSE(%)	MR(%)	$R^2$
GP-PSO	0.54	0.67	0.9918
MK-RVM	0.72	0.45	0.9817

It is apparent that Table 4 that the coefficient of the  $3/2$  power term is much smaller than the power of the first power term, and there is a linear relationship with SOH. This can also be confirmed by Fig. 12.

In order to further verify the proposed method, battery #5 was selected to train the same 84 training cycles according to the test method of Multi-Kernel RVM estimation SOH in [38]. It is obvious that the accuracy is not significantly different and both methods can accurately estimate the SOH of battery, while the RMSE and  $R^2$  of the GP model are better than MK-RVM from Table 4. Because the RMSE of the GP model is smaller and  $R^2$  is closer to 1, the SOH estimation result is more stable and fits better through the GP model. Furthermore, the proposed method is easier to visualize by expression.

V. CONCLUSION

In this paper, a GP model is provided for estimating the SOH of Li-ion batteries based on the characteristic data. Data were collected from the data repository of the NASA Ames Prognostics Center of Excellence (PCoE). This paper selects two batches of Li-ion battery data by extracting the features of Li-ion batteries during each charge and discharge cycle. The feature data and the SOH of the battery are input to the GP model for training, and then a formula is obtained for estimating the SOH of the Li-ion battery. According to the principle of the GP model iteratively searching for the

optimal formula, the obtained SOH estimation formula of Li-ion batteries is not unique and can greatly explore the optimal formula. When the training cycles or batteries are different, the estimation formula will be different. To verify that the SOH of the Li-ion battery can be well estimated by the obtained formula, the SOHs of batteries #5, #6 and #7 are estimated by selecting three different training cycles and compared with the BP neural network estimation results. Then, using the batteries of # 33, # 34 and # 36 to verify the results of the estimation formula, the relationship between features and SOH is explored. It can be concluded that the GP model can trace the change in SOH well through the extracted features, and the proposed method has good robustness.

The SOH estimation model is derived from training data such as traditional machine learning. The difference is that the method focuses on obtaining the function of estimating SOH, and the relationship between the performance features and the SOH of Li-ion batteries is constructed. Although the traditional machine learning model could accurately estimate the Li-ion battery SOH, it is difficult to conclude the possible relationship between extracted features and the SOH of batteries. According to the optimal formula obtained by the GP model, we can summarize the influence of features on the SOH of Li-ion batteries and even obtain the expression between the features and SOH.

This method can also be applied in electronics fields. In particular, performance degradation studies are usable when product performance degradation is not subject to typical degradation processes, and products have certain features and health indexes. When it is desired to obtain a functional relationship between HI and features, the model formula can be optimized autonomously according to this method to realize the estimation of HI.

REFERENCES

- [1] M. Bercibar, I. Gandiaga, I. Villarreal, N. Omar, J. Van Mierlo, and P. Van den Bossche, "Critical review of state of health estimation methods of li-ion batteries for real applications," *Renew. Sustain. Energy Rev.*, vol. 56, pp. 572–587, Apr. 2016.
- [2] C. Zhang, Y. He, L. Yuan, S. Xiang, and J. Wang, "Prognostics of lithium-ion batteries based on wavelet denoising and DE-RVM," *Comput. Intell. Neurosci.*, vol. 2015, May 2015, Art. no. 918305.
- [3] S. Dey, S. Mohon, P. Pisu, and B. Ayalew, "Sensor fault detection, isolation, and estimation in lithium-ion batteries," *IEEE Trans. Control Syst. Technol.*, vol. 24, no. 6, pp. 2141–2149, Nov. 2016.
- [4] Z. Ma, J. Jiang, W. Shi, W. Zhang, and C. C. Mi, "Investigation of path dependence in commercial lithium-ion cells for pure electric bus applications: Aging mechanism identification," *J. Power Sources*, vol. 274, pp. 29–40, Jan. 2015.
- [5] R. Xiong, L. Li, and J. Tian, "Towards a smarter battery management system: A critical review on battery state of health monitoring methods," *J. Power Sources*, vol. 405, pp. 18–29, Nov. 2018.

- [6] K.-T. Lee, M.-J. Dai, and C.-C. Chuang, "Temperature-compensated model for lithium-ion polymer batteries with extended Kalman filter state-of-charge estimation for an implantable charger," *IEEE Trans. Ind. Electron.*, vol. 65, no. 1, pp. 589–596, Jan. 2018.
- [7] S. Yuan, H. Wu, and C. Yin, "State of charge estimation using the extended Kalman filter for battery management systems based on the ARX battery model," *Energies*, vol. 6, no. 1, pp. 444–470, 2013.
- [8] S. Li, S. Pischinger, C. He, L. Liang, and M. Stapelbroek, "A comparative study of model-based capacity estimation algorithms in dual estimation frameworks for lithium-ion batteries under an accelerated aging test," *Appl. Energy*, vol. 212, pp. 1522–1536, Feb. 2018.
- [9] S. Sepasi, R. Ghorbani, and B. Y. Liaw, "Inline state of health estimation of lithium-ion batteries using state of charge calculation," *J. Power Sources*, vol. 299, pp. 246–254, Dec. 2015.
- [10] R. Xiong, F. Sun, Z. Chen, and H. He, "A data-driven multi-scale extended Kalman filtering based parameter and state estimation approach of lithium-ion polymer battery in electric vehicles," *Appl. Energy*, vol. 113, pp. 463–476, Jan. 2014.
- [11] X. Xu, C. Yu, S. Tang, X. Sun, X. Si, and L. Wu, "Remaining useful life prediction of lithium-ion batteries based on Wiener processes with considering the relaxation effect," *Energies*, vol. 12, no. 9, p. 1685, 2019.
- [12] W. Xian, B. Long, M. Li, and H. Wang, "Prognostics of lithium-ion batteries based on the verhulst model, particle swarm optimization and particle filter," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 1, pp. 2–17, Jan. 2014.
- [13] Q. Miao, L. Xie, H. Cui, W. Liang, and M. Pecht, "Remaining useful life prediction of lithium-ion battery with unscented particle filter technique," *Microelectron. Rel.*, vol. 53, no. 6, pp. 805–810, Jun. 2013.
- [14] J. Wei, G. Dong, and Z. Chen, "Remaining useful life prediction and state of health diagnosis for lithium-ion batteries using particle filter and support vector regression," *IEEE Trans. Ind. Electron.*, vol. 65, no. 7, pp. 5634–5643, Jul. 2018.
- [15] D. Liu, X. Yin, Y. Song, W. Liu, and Y. Peng, "An on-line state of health estimation of lithium-ion battery using unscented particle filter," *IEEE Access*, vol. 6, pp. 40990–41001, 2018.
- [16] X. Tang, C. Zou, K. Yao, G. Chen, B. Liu, Z. He, and F. Gao, "A fast estimation algorithm for lithium-ion battery state of health," *J. Power Sources*, vol. 396, pp. 453–458, Aug. 2018.
- [17] C. Weng, Y. Cui, J. Sun, and H. Peng, "On-board state of health monitoring of lithium-ion batteries using incremental capacity analysis with support vector regression," *J. Power Sources*, vol. 235, pp. 36–44, Aug. 2013.
- [18] Y. Li, M. Abdel-Monem, R. Gopalakrishnan, M. Berecibar, E. Nanini-Maury, N. Omar, P. van den Bossche, and J. Van Mierlo, "A quick on-line state of health estimation method for li-ion battery with incremental capacity curves processed by Gaussian filter," *J. Power Sources*, vol. 373, pp. 40–53, Jan. 2018.
- [19] L. Wang, X. Zhao, L. Liu, and C. Pan, "State of health estimation of battery modules via differential voltage analysis with local data symmetry method," *Electrochimica Acta*, vol. 256, pp. 81–89, Dec. 2017.
- [20] M. S. H. Lipu, M. A. Hannan, A. Hussain, M. M. Hoque, P. J. Ker, M. H. M. Saad, and A. Ayob, "A review of state of health and remaining useful life estimation methods for lithium-ion battery in electric vehicles: Challenges and recommendations," *J. Cleaner Prod.*, vol. 205, pp. 115–133, Dec. 2018.
- [21] G.-W. You, S. Park, and D. Oh, "Real-time state-of-health estimation for electric vehicle batteries: A data-driven approach," *Appl. Energy*, vol. 176, pp. 92–103, Aug. 2016.
- [22] D. Yang, Y. Wang, R. Pan, R. Chen, and Z. Chen, "State-of-health estimation for the lithium-ion battery based on support vector regression," *Appl. Energy*, vol. 227, pp. 273–283, Oct. 2018.
- [23] R. R. Richardson, C. R. Birkl, M. A. Osborne, and D. A. Howey, "Gaussian process regression for *in situ* capacity estimation of lithium-ion batteries," *IEEE Trans. Ind. Informat.*, vol. 15, no. 1, pp. 127–138, Jan. 2019.
- [24] D. Yang, X. Zhang, R. Pan, Y. Wang, and Z. Chen, "A novel Gaussian process regression model for state-of-health estimation of lithium-ion battery using charging curve," *J. Power Sources*, vol. 384, pp. 387–395, Apr. 2018.
- [25] A. A. Hussein, "Capacity fade estimation in electric vehicle li-ion batteries using artificial neural networks," *IEEE Trans. Ind. Appl.*, vol. 51, no. 3, pp. 2321–2330, May 2015.
- [26] G.-W. You, S. Park, and D. Oh, "Diagnosis of electric vehicle batteries using recurrent neural networks," *IEEE Trans. Ind. Electron.*, vol. 64, no. 6, pp. 4885–4893, Jun. 2017.
- [27] C. Zhang, Y. He, L. Yuan, and S. Xiang, "Capacity prognostics of lithium-ion batteries using EMD denoising and multiple kernel RVM," *IEEE Access*, vol. 5, pp. 12061–12070, 2017.
- [28] H. Dong, X. Jin, Y. Lou, and C. Wang, "Lithium-ion battery state of health monitoring and remaining useful life prediction based on support vector regression-particle filter," *J. Power Sources*, vol. 271, pp. 114–123, Dec. 2014.
- [29] H. Li, D. Pan, and C. L. P. Chen, "Intelligent prognostics for battery health monitoring using the mean entropy and relevance vector machine," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 44, no. 7, pp. 851–862, Jul. 2014.
- [30] D. Zhou, H. Yin, W. Xie, P. Fu, and W. Lu, "Research on online capacity estimation of power battery based on EKF-GPR model," *J. Chem.*, vol. 2019, pp. 1–9, Jan. 2019.
- [31] M. T. Ahvanooy, L. Qianmu, W. Ming, and W. Shuo, "A survey of genetic programming and its applications," *KSI Trans. Internet Inf. Syst.*, vol. 13, pp. 1765–1794, Apr. 2019.
- [32] B. Ma, X. Jiao, F. Meng, F. Xu, Y. Geng, R. Gao, W. Wang, and Y. Sun, "Identification of gene regulatory networks by integrating genetic programming with particle filtering," *IEEE Access*, vol. 7, pp. 113760–113770, 2019.
- [33] Z. Qu, Y.-X. Chen, L. Liu, Y. Xie, and Q. Zhou, "The algorithm of concrete surface crack detection based on the genetic programming and percolation model," *IEEE Access*, vol. 7, pp. 57592–57603, 2019.
- [34] J. MacLachlan, Y. Mei, J. Branke, and M. Zhang, "Genetic programming hyper-heuristics with vehicle collaboration for uncertain capacitated arc routing problems," *Tech. Rep.*, 2019.
- [35] Y. Echevarría, C. Blanco, and L. Sánchez, "Learning human-understandable models for the health assessment of li-ion batteries via multi-objective genetic programming," *Eng. Appl. Artif. Intell.*, vol. 86, pp. 1–10, Nov. 2019.
- [36] L. Liao, "Discovering prognostic features using genetic programming in remaining useful life prediction," *IEEE Trans. Ind. Electron.*, vol. 61, no. 5, pp. 2464–2472, May 2014.
- [37] P. Khumprom and N. Yodo, "A data-driven predictive prognostic model for lithium-ion batteries based on a deep learning algorithm," *Energies*, vol. 12, no. 4, p. 660, 2019.
- [38] Y. Zhang and B. Guo, "Online capacity estimation of lithium-ion batteries based on novel feature extraction and adaptive multi-kernel relevance vector machine," *Energies*, vol. 8, no. 11, pp. 12439–12457, 2015.



**HANG YAO** received the B.S. degree in system engineering from the National University of Defense Technology (NUDT), China, where he is currently pursuing the master's degree in management science and engineering with the College of Systems Engineering. His research interests include reliability and remaining useful life prediction.



**XIANG JIA** received the Ph.D. degree in management science and engineering from the National University of Defense Technology (NUDT), China. He is currently a Lecturer of management science and engineering with the College of Systems Engineering, NUDT. His research interest includes system reliability.



**QIAN ZHAO** received the master's degree in management science and engineering from the National University of Defense Technology (NUDT), China. He is currently a Lecturer with the College of Information and Communication, NUDT. His research interests include reliability and remaining useful life prediction.



**BO GUO** received the B.S. degree in mathematics from the Huazhong Institute of Technology, the M.S. degree in system engineering from the National University of Defense Technology (NUDT), China, and the Ph.D. degree in engineering management from the Tokyo University of Science. He is currently a Professor with the College of Systems Engineering, NUDT. His research interests mainly include system reliability and project management.

...



**ZHI-JUN CHENG** received the Ph.D. degree in management science and engineering from the National University of Defense Technology (NUDT), China. She is currently an Associate Professor with the College of Systems Engineering, NUDT. Her research interest mainly includes prognostics health management.