A new method for estimating lithium-ion battery capacity using genetic programming combined model

Hang Yao,Xiang Jia*,Bo Wang, Bo Guo
College of Systems Engineering
National University of Defense Technology
Changsha, China
jiaxiang09@sina.cn

Abstract—**Lithium-ion battery is the main energy source widely used in many fields. Therefore, it is particularly essential for estimating the health of lithium-ion battery accurately, especially in important fields such as aerospace, rail transit and satellite. For lithium-ion battery, the battery capacity is a health index (HI) that best reflects its performance degradation. By estimating the battery capacity, the health status of the lithium-ion battery can be clearly identified. However, there are technical barriers to the direct measurement of battery capacity in engineering, and many characteristics and capacities of lithium-ion batteries have abrupt changes, so that it is difficult to calculate the battery capacity accurately by formula calculation. In this paper, a new method of genetic programming combined model is proposed, which can calculate the capacity of lithium-ion battery by formulating multiple monitored features with a certain precision. Therefore, the functional relationship between multiple features and HI is well measured, which lays a good foundation for the subsequent life prediction of battery.**

Keywords-genetic programming; lithium-ion battery; health index; formula calculation

I. INTRODUCTION

Nowadays, product health management is an important part of improving industrial production efficiency and safety. With the advancement of monitoring technology, using monitoring data to predict product life and achieve health management has become a viable and effective method. Lithium-ion batteries are a fundamental component that is widely used in satellite, aerospace and rail transit[1,2]. In order to meet the needs of operating voltage and power, as well as to extend the life of the battery pack and battery, it is especially important to achieve health management of lithium-ion batteries.

These fields have fully tested the performance status of lithium-ions, and obtained characteristic data of many lithium-ion batteries. Although the battery capacity can directly reflect the health indexs of lithium-ion battery performance degradation, due to the limitations of monitoring technology, the battery capacity cannot be directly measured in actual engineering. In order to solve this problem and realize further management of lithium-ion batteries, many scholars have done a lot of effective work on the state-of-health of lithium-ion battery.

The features of lithium-ion battery such as the temperature, current, voltage or resistance of the lithium-ion battery that are used as HIs to predict the remaining life. For making full use of the performance characteristics of the monitoring, more accurate methods for judging the health status of lithium ions have been applied. Landi and Gross [3] estimated the overall trend of battery status based on the relationship between fuzzy logic processing performance characteristics and battery state. Liu [4] et al. use a combination of nonlinear degradation functions to optimize the autoregressive model to achieve an estimate of battery cycle life. Li [5] et al. used artificial neural networks (ANNs) to predict Remaining Useful Life (RUL) for lithium-ion batteries. Zhang [6] et al. adopt Accelerated Particle Swarm Optimization (APSO) algorithm to optimize the kernel function of Relevance Vector Machine (RVM) to realize the prediction of lithium-ion battery RUL.

Currently, multiple features are mapped to battery capacity by RVM or Support Vector Machine (SVM) methods to achieve residual life prediction. SVM and RVM realize good results in estimating the state-of-health of lithium-ion batteries, and have obtained more in-depth research. Wang et al. [7] established a three-parameter conditional capacity degradation model and used RVM to obtain correlation vectors to represent battery capacity attenuation and cycle life. Simultaneously. Widodo et al. [8] proposed a battery health assessment framework based on discharge voltage sample entropy. Based on discharge voltage sample entropy, Widodo et al. [8] proposed a battery health assessment framework and gave the uncertainty representation. Li et al. [9] established a multi-step prediction model to realize the state-of-health prediction based on average entropy and RVM. Zhang et al. [10] used wavelet to denoise the noise during battery testing, and estimated the robustness of the test by differential evolution-optimized RVM. These methods have good results in the estimation of lithium-ion health state, but these methods can not reflect the functional relationship between the selected features and battery capacity. It is necessary to select the appropriate kernel function in the estimation according to the actual situation.

A method of genetic programming combined model is proposed in this paper. According to the multiple

978-1-7281-0860-5/19/$31.00 ©2019 IEEE

characteristics of the monitoring, the model autonomously explores the formula for calculating the battery capacity by the feature, which embodies the relationship between the battery capacity and multiple features. In addition, the model can be simplified while ensuring a certain degree of precision. The remainder of this paper is organized as follows. Section 2 describes some of the current problems in the estimation of lithium-ion battery capacity; In Section 3 the genetically-programming combination model calculate the capacity of lithium-ion battery clearly; Section 4 gives experimental verification and comparison. Finally, the conclusion is found in Section 5.

II. PROBLEM STATEMENT

Intuitively, estimating the health status of a lithium-ion battery by formula, which should be most straightforward. However, it is difficult to measure the battery capacity in engineering, so there are currently a series of methods for estimating the state of lithium-ion batteries, which are difficult to reflect the relationship between the performance characteristics of the input and the state of the output battery.

Fuzzy logic (FL) is used to process measurement data for non-linear and complex systems. In the fuzzy logic of (1), two exponential functions are used to calculate the SOH of a lithium-ion battery by the following equation

$$y_{fit} = a_0 + a_1 e^{-(\frac{x}{\alpha_1})^{\beta_1}} + a_2 e^{-(\frac{x}{\alpha_2})^{\beta_2}} . \quad (1)$$

Where x represents the number of cycles and y is the value of SOH, the rest are the parameters to be determined. First, the health index was calculated using the FL fit curve with an error between 5% and 10%. The error is then reduced by the neural network. Since the model structure of the fuzzy logic has been determined, it is difficult to reflect the relationship between the input features and the SOH. Only the state of the overall trend of the lithium-ion battery can be given, resulting in the accuracy of the model is not high.

RVM can identify patterns of nonlinear systems by analyzing data, which can be used to predict the SOH of lithium-ion battery[11]. However, it does not have a specific formula for expressing the health status of lithium-ions, and does not reflect the relationship between input characteristics and battery capacity; and the choice of kernel function has a great influence on the calculation results of support vector machines, and the calculation process is complicated.

Trying to find a way to avoid the fixed model framework and enable the model to explore on its own to achieve lithium-ion battery estimation. In particular, when the specific model cannot be determined or the kernel function cannot be selected, the capacity of the lithium-ion battery can be estimated directly by monitoring the performance characteristic data. Thus, a formula for estimating the battery capacity by the feature is obtained, which embodies the relationship between battery capacity and multiple features, and simplifies the model while ensuring a certain precision.

III. METHODOLOGY

As shown in Figure 1, a machine learning step of estimating HI to achieve a prediction of the remaining life of the product typically includes data acquisition, feature extraction, feature selection, estimated prediction, and remaining life outcome[12]. The genetic programming combination model mainly described in this section belongs to two steps in the dashed box. Through the model, the feature selection and the target of the battery capacity estimation formula can be achieved.

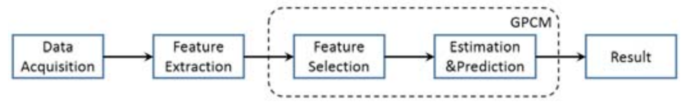

Figure 1. Steps of a general prognostics method

A. Genetic Programming:

The flow chart of the genetic plan is shown in Fig.2, The extracted features and battery capacity are input into the genetic specification model, and the initial calculation formula population is randomly generated, and the individual selection is performed according to the fitness function, and then the intersection and variation are performed. Finally, the evolution is terminated according to the judgment condition, thereby obtaining the optimal formula individual. Similar to genetic algorithms, genetic programming[13-15] has the same important elements as follows:

- Randomly generate an initial genetic population;
- Need 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;
- Termination controllable

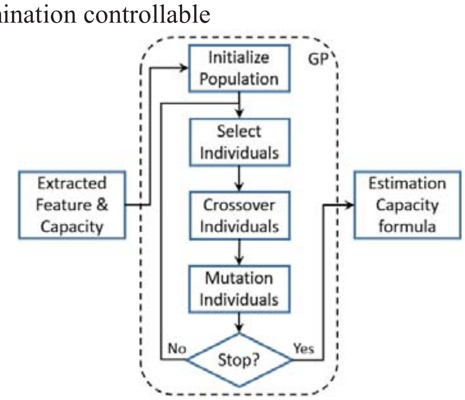

Figure 2. Flowchart of the genetic programminng method

Usually genetic algorithm optimizes a value, and genetic programming differs from genetic algorithm in that the individual of genetic programming is a strategy or function. An individual represents a function or a strategy, so the individual's code includes a series of mathematical operations, such as addition, subtraction, multiplication, division, square, square root, exponential operation, and logarithm operation. .

Individual coding can be represented by a tree structure, as shown in the figure 3.

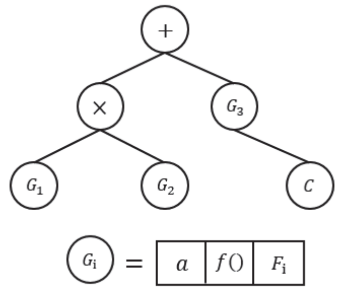

Figure 3. GP individual example

Where node G_i represents the factor, node C is a constant, factor i is equal to the constant a multiplied by the characteristic F_i of a certain mathematical operation $f()$; and the number of factors in the individual is defined as the depth of the tree, then the depth of the individual is 3. Thus the individual represented by the tree structure can be represented by the following formula

$$F = G_1 \times G_2 + G_3 + C. \tag{2}$$

The nodes in the tree structure are all variable, which means the form of the formula can be changed according to the conditions, so that facilitates crossover and mutation operations. After performing the selection operation in the Fig.2 , according to the crossover probability, it is judged whether or not the cross operation is performed. If there is no intersection, each does not change; if a cross operation occurs, the specific figure is as shown in the figure

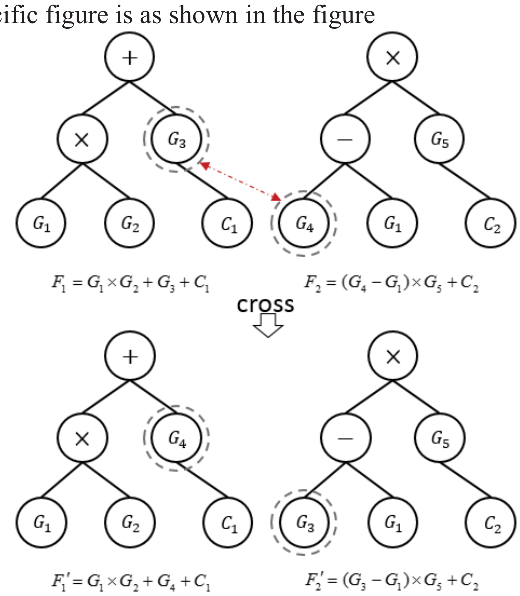

Figure 4. Example of a cross operation

Individual F_1 and individual F_2 are randomly selected, and the same number of nodes are randomly selected from individuals F_1 and F_2 for intersection. In Fig. 4, individuals F_1 and F_2 each select a node G3 and G4, and the two nodes perform cross-interchange, thereby obtaining two new individuals F_1' and F_2'.

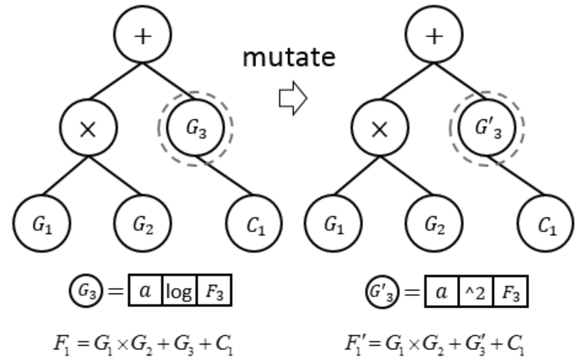

Figure 5. Example of a mutate operation

After the population individual has completed the crossover, according to the flow chart of the genetic plan, the next step is to mutate the individual population, as shown in Figure 2. The individual F_1 is randomly selected, and whether the mutation operation is performed is determined according to the mutation probability. If the mutation is not performed, the individual F_1 is not changed; if the mutation is required, the nodes in the individual are randomly selected, and the whole node or a part is selected to be mutated. In Fig.5 , the logarithm operation variation in the individual F_1 selection node G3 is a square operation, thereby obtaining a new individual F_1'. Among them, individuals who have undergone crossover and mutation can be repeatedly crossed and mutated.

The population individual repeats the selection, crossover, and mutation operations until the termination condition is met. In order to be able to derive the estimation result with the smallest error in the actual result of the battery capacity, the calculation result of the individual formula and the root mean square error of the real capacity can accurately measure the fitness of the individual, and thus the fitness function is

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(C_i - \hat{C}_i)^2}. \tag{3}$$

Where C is the real capacity of lithium-ion battery. $\hat{C}$ represents the estimated value of the battery capacity. N is the amount of data. The termination condition is set in this paper as the optimal individual's fitness meets certain requirements and reaches a certain evolutionary algebra. When the optimal individual is selected, the formula represented by the individual is a formula that can better estimate the battery capacity.

B. Genetic Programming Combination

Due to the actual estimation of the genetic programming formula, sometimes the optimal individual of the population prematurely converges, resulting in the optimal individual satisfying the termination condition of the genetic plan, but the estimation of the capacity of the lithium-ion battery is not very good. By integrating the results of several genetic programming models, the genetic programming combined model can make the estimation results meet the accuracy requirements. The specific model is as follows:

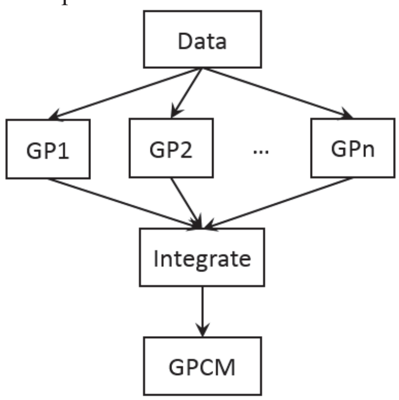

Figure 6. Flowchart of the GPCM

In the GPCM so as to reflect the better individual model is more helpful in combining model, the weight of the model is determined according to the fitness of each GP optimal individual, such as the formula (4),(5) where $RMSE_i$ represents the i-th The fitness of the optimal individual of the GP model, W_i is the weight of the i-th GP model.

$$s_i = 1 - \frac{RMSE_i}{\sum_{i=1}^{n} RMSE_i} \tag{4}$$

$$w_i = \frac{s_i}{\sum_{i=1}^{n} s_i} \tag{5}$$

The genetic programming combination model can be obtained, and the final lithium-ion battery capacity estimation formula can be expressed as:

$$\hat{C} = \sum_{i=1}^{n} w_i C_i \tag{6}$$

C_i represents the battery capacity estimation formula represented by the optimal individual of the i-th model.

IV. APPLICATION AND VALIDATION

To ensure the repeatability and integrity of the algorithm and model, Section 4 first introduces the source and feature extraction of the performance degradation test data for lithium-ion batteries. The effect of the lithium-ion battery capacity estimation formula obtained from a single genetic programming model of different populations is then presented. Finally, the capacity estimation formula of lithium-ion battery obtained by genetic programming model is given and compared with the estimation results of single genetic programming model and RVM model.

The lithium-ion battery performance degradation test data used in algorithm verification and model comparison is from the open source database of NASA. Through the method of extracting the performance degradation characteristics of lithium-ion battery in the literature, F1 is the time interval of the charging voltage of the lithium-ion battery CC 3.7V rising to 4.2V; The time interval of CV discharge current 1.3A falling to 0.3A is F2; F3 is the time interval at which the discharge voltage is 3.7V dropped to 2.7V; F4 represents the average current temperature between the start time of F1 and the end time of F2; and F5 is the average temperature of the battery during the F3 time period.

The five battery characteristics and battery capacity of #05 results in five performance degradation features (F1, F2, F3, F4, F5) of lithium-ion batteries, which are put into genetic programming models of different populations. The number of populations of the model is set to 1000, the crossover probability is 0.7, and the mutation probability is 0.2, the end condition of the GP model is that the fitness of the optimal individual is less than 2%. Figure 7-9 show the estimation results of the optimal individuals obtained from the evolution of 5000 generations of the three sets of genetically normalized models #05.

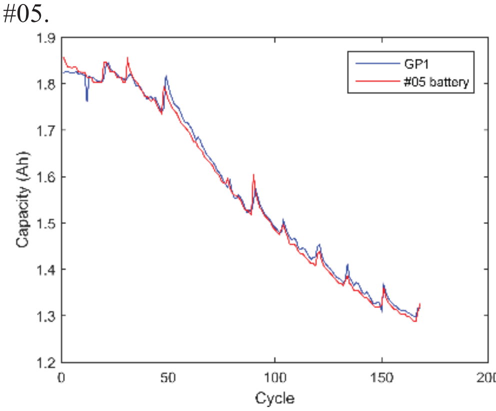

Figure 7. Estimation result of GP1

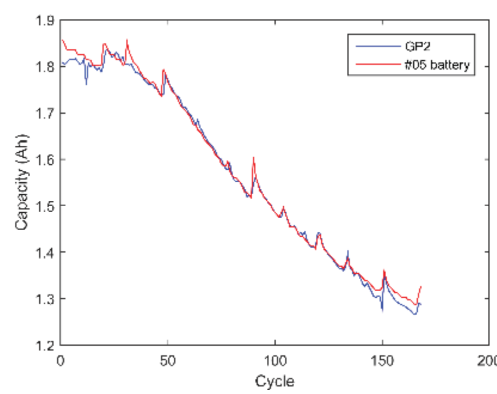

Figure 8. Estimation result of GP2

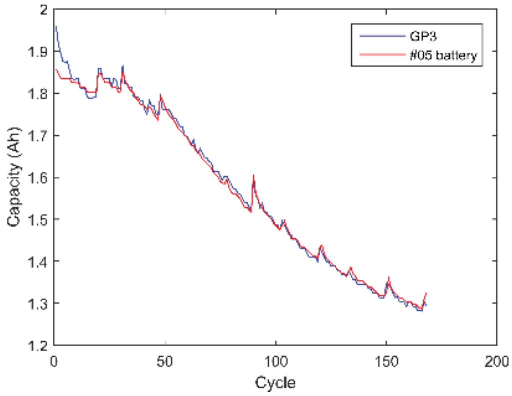

Figure 9. Estimation result of GP3

TABLE I. BATTERY CAPACITY ESTIMATION FORMULA AND ERROR OF DIFFERENT GP MODELS

#5	C_{GPi}	RMSE
GP1	$1.467e^{-4}F_4^2+3.31e^{-4}F_1+0.683$	0.0135
GP2	$1.418\log F_1-0.164\log F_5-0.854\log F_2-0.147$	0.0151
GP3	$1.627e^{-4}F_3\log F_3+0.049\sqrt{F_3}-1.885$	0.0152

It can be seen from the Figs 7-9 that the genetic specification model has a good effect on the estimation of the capacity of the lithium-ion battery, but the estimated deviation in the partial period is slightly larger, and thus the three groups of models are combined to obtain #05. Genetically programming combination model. The following tables □ and □ are the formula and mean square error of the three sets of genetic specification models GP1, GP2 and GP3 of #05, and the weights are respectively obtained by the fitness function.

TABLE II. WEIGHT OF DIFFERENT GP MODEL

Table	GP1	GP2	GP3
w_i	0.3458	0.3277	0.3265

Combining the three genetically programming models of #05, the estimated results of the obtained genetically programming combination model are shown in Fig.10.

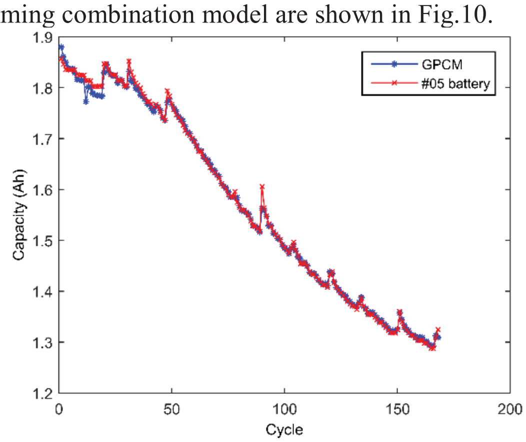

Figure 10. Estimation result of GPCM

The estimated formulas for the #05 lithium-ion battery capacities is

$$\hat{C}_{\#05}=0.3458C_{GP1}+0.3277C_{GP2}+0.3265C_{GP3} \tag{7}$$

The GPCM errors for #05 was 0.0085, which increased by 0.5%, respectively, compared to the optimal single genetically normalized model. By comparing with the actual battery capacity value, the battery capacity estimation obtained by using the genetically programming combination model has achieved good results.

V. CONCLUSION

A genetic programming combination model was proposed, and the model autonomously explored the functional relationship between the capacity of the lithium-ion battery and the selected characteristics, thus giving a formula for estimating the capacity of the lithium-ion battery. Thus, a formula for estimating the battery capacity by the feature is obtained, which embodies the relationship between battery capacity and multiple features, and achieves accurate estimation of lithium-ion battery capacity.

ACKNOWLEDGMENT

The authors gratefully acknowledge the financial supports for this research from the National Natural Science Foundation of China (with granted number 71801219 and 61573370).

REFERENCES

[1] L. Guo et al, "A recurrent neural network based health indicator for remaining useful life prediction of bearings," Neurocomputing, vol. 240, pp. 98-109, 2017.

[2] R. Khelif et al, "Direct Remaining Useful Life Estimation Based on Support Vector Regression," IEEE Transactions on Industrial Electronics, vol. 64, (3), pp. 2276-2285, 2017.

[3] M. Landi and G. Gross, "Measurement Techniques for Online Battery State of Health Estimation in Vehicle-to-Grid Applications," IEEE Transactions on Instrumentation and Measurement, vol. 63, (5), pp. 1224-1234, 2014.

[4] D. Liu et al, "Lithium-ion battery remaining useful life estimation based on fusion nonlinear degradation AR model and RPF algorithm," Neural Computing and Applications, vol. 25, (3), pp. 557-572, 2014.

[5] X. Li et al, "Remaining useful life prediction for lithium-ion batteries based on a hybrid model combining the long short-term memory and Elman neural networks," Journal of Energy Storage, vol. 21, pp. 510-518, 2019.

[6] 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, (11), pp. 12439-12457, 2015.

[7] D. Wang, Q. Miao and M. Pecht, "Prognostics of lithium-ion batteries based on relevance vectors and a conditional three-parameter capacity degradation model," Journal of Power Sources, vol. 239, pp. 253-264, 2013.

[8] A. Widodo et al, "Intelligent prognostics for battery health monitoring based on sample entropy," Expert Systems with Applications, vol. 38, (9), pp. 11763-11769, 2011.

[9] H. Li et al, "Intelligent Prognostics for Battery Health Monitoring Using the Mean Entropy and Relevance Vector Machine," Ieee Transactions on Systems Man Cybernetics-Systems, vol. 44, (7), pp. 851-862, 2014.

[10] C. Zhang et al, "Prognostics of Lithium-Ion Batteries Based on Wavelet Denoising and DE-RVM," Computational Intelligence and Neuroscience, vol. 2015, pp. 918305-8, 2015.

[11] Y. SONG et al, "Satellite lithium-ion battery remaining useful life estimation with an iterative updated RVM fused with the KF algorithm," Chinese Journal of Aeronautics, vol. 31, (1), pp. 31-40, 2018.

[12] Yaguo Lei et al, " Machinery health prognostics: A systematic review from data acquisition to RUL prediction," Mechanical Systems and Signal Processing, vol. 104, pp. 799-834, 2018.

[13] L. Liao, "Discovering Prognostic Features Using Genetic Programming in Remaining Useful Life Prediction," IEEE Transactions on Industrial Electronics, vol. 61, (5), pp. 2464-2472, 2014.

[14] J. Coble and J. W. Hines, ‘‘Identifying optimal prognostic parameters from data: A genetic algorithms approach,’’ in *Proc. Annu. Conf. Prognost. Health Manage. Soc.*, 2009, pp. 1–11

[15] R. Sun, F. Tsung, and L. Qu, ‘‘Combining bootstrap and genetic programming for feature discovery in diesel engine diagnosis,’’ *Int. J. Ind.Eng.*, vol. 11, no. 3, pp. 273–281, 2004