PAPER • OPEN ACCESS

Health State Estimation Method of Lithium Ion Battery Based on NASA Experimental Data Set

View the article online for updates and enhancements.

Health State Estimation Method of Lithium Ion Battery Based on NASA Experimental Data Set

Huaqing Xu, Yanqing Peng [*] **and Lumei Su**

School of electrical engineering and automation, Xiamen University of Technology, Xiamen Fujian 361024, China

*Corresponding author e-mail: pyqxm@163.com

Abstract. Based on the experimental data set of NASA lithium-ion battery, this paper proposes two novel methods for estimating the health status of lithium-ion battery. Firstly, the definition of battery health status is introduced. Secondly, based on the data preprocessing and visualization analysis, four features related to actual capacity degradation are extracted from the data. Thirdly, Two machine learning models, regression tree and random forest, are compared in this work. Both models are used Bootstrap methods for performance evaluation. Finally, The experimental results show that both have high estimation accuracy. The regression tree final model predicts a mean square error of 0.0006, while the random forest final model predicts a mean square error of 0.0002, indicating that the random forest is a better model.

1. Introduction

In recent years, the research on batteries at home and abroad is mainly State of Charge (SOC) and Remaining Useful Life (RUL) [1-10]. The research method is mainly based on the application of battery experimental data using machine learning algorithms for data experiments. In recent years, the types of batteries studied at home and abroad are mainly lead-acid batteries and lithium-ion batteries. Among them, lithium-ion batteries are mostly studied, and the research methods are mainly machine learning algorithm experiments. Domestic research such as: Yang [1] uses support vector machine to predict the SOC of lead-acid batteries; Jiang [2] uses BP neural network and particle swarm optimization autoregressive model to carry out SOC estimation and RUL prediction for lithium-ion batteries. Foreign studies such as: Frisk [7] using random survival forests for RUL prediction of lead-acid batteries; Richardson [10] using Gaussian process regression for RUL prediction of lithium-ion batteries.

Compared with lead-acid batteries, lithium-ion batteries, as a high-performance secondary green battery, have the advantages of high operating voltage, high energy density, low self-discharge rate, no memory effect, and small volume [11], so lithium ions are studied. The battery health assessment method has practical significance.

2. The definition of battery health status

The actual capacity of the battery will gradually decrease with the number of charge and discharge cycles due to aging, etc. Usually, the end of life is determined when the actual battery capacity is lower than the Acceptable Performance Threshold (APT) [2,12]. APT is usually 70% or 80% of rated capacity. State of Health (SOH) is defined as [13] :

$$SOH = \frac{Q_{aged}}{Q_{rate}} \times 100\% \qquad (1)$$

Where Q_{rate} is the rated capacity of the battery when it leaves the factory, and Q_{aged} is the actual capacity.

It can be seen from Eq.(1) that since the rated capacity is constant, the actual capacity of the battery can represent the health of the battery.

3. Data preprocessing and visualization analysis

The experimental data used was from the National Aeronautics and Space Administration (NASA) lithium-ion battery charge and discharge experimental data set [12].

The four battery numbers in the data set are B0005, B0006, B0007, B0018, the model number are all 18650, and the rated capacity is 2Ah. The experiments are all carried out at room temperature of 24℃. The charging experiment was first charged with a constant current of 1.5A until the voltage reached 4.2V, and then in constant voltage mode until the current dropped to 20mA. The discharge process was discharged with a constant current of 2A until the voltage dropped to 2.7V (B0005), 2.5V (B0006), 2.2V (B0007), and 2.5V (B0018), respectively. The charging process and the discharging process in each charging and discharging cycle are started from time 0, and the data of charging and discharging such as voltage, current, temperature and actual capacity are recorded. The experiment stopped after the measured actual capacity was less than the rated capacity of 70% (APT). The actual capacity degradation curve of four batteries is shown in Figure 1.
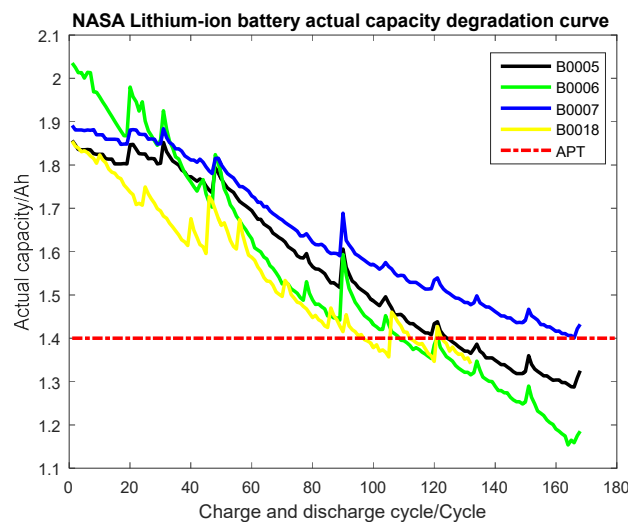

Figure 1. NASA Lithium-ion battery actual capacity degradation curve

In order to dig out the correlation between the data and the actual capacity degradation, the charging voltage timing curve, charging temperature timing curve, discharge voltage timing curve and discharge temperature timing curve of each charge and discharge cycle of each battery are plotted. The trend of each mode of each battery are similar, the trend of each mode of B0006 battery as shown in Figure 2.
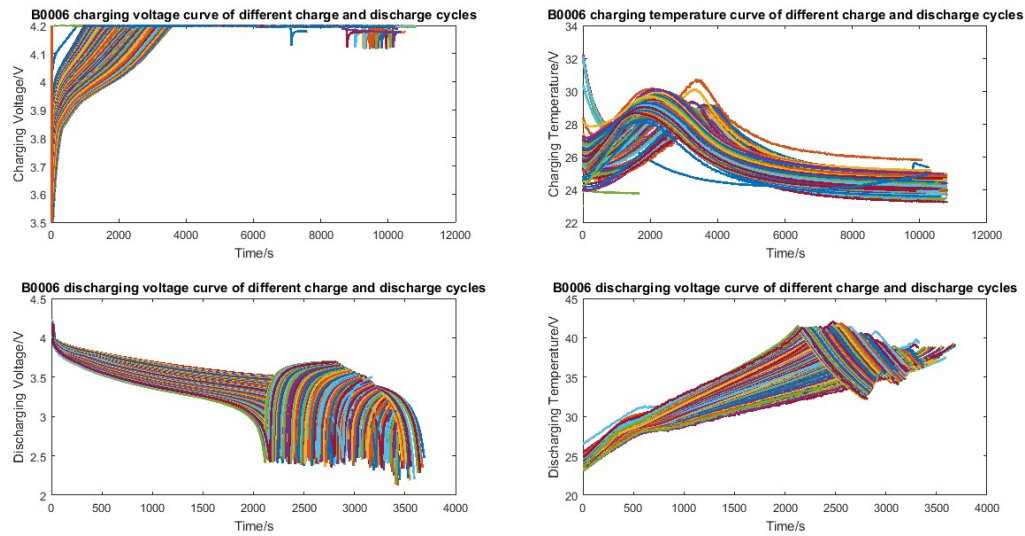
Figure 2. The trend of each mode of B0006 battery

4. Feature extraction

It can be seen from Figure 2. that the time required for the charging voltage of different charging and discharging cycles to reach 4.2V is different. The time required for the charging temperature of different charging and discharging cycles to reach a maximum value is different. And the time required for the discharging voltage of different charging and discharging cycles to 2.5V is different. And the time when the discharge temperature reaches the maximum value in different charge and discharge cycles is different. Therefore, the four types of features of B0006 are extracted as shown in Figure 3.
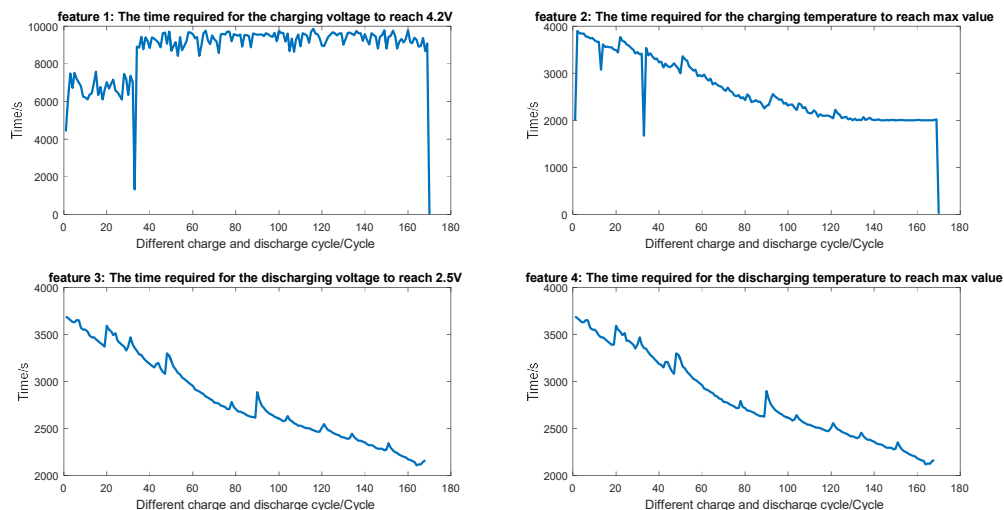

Figure 3. The four types of features of B0006

In the experimental data of B0006 battery, the number of times of charging is 170, and the number of times of discharge is 168. In order to maintain consistency, it is necessary to remove 2 times of charging data, so 2 data points are removed from feature 1 and feature 2. The 4 minimum values in feature 1 and feature 2 are directly removed. These 4 data points are the measurement abnormal points. In addition, the experimental condition of B0018 battery is exactly the same as that of B0006 battery,

and the model number is also the same. Therefore, the same four types of features are extracted for B0018 battery and the data is combined with B0006. B0018 battery is charged 134 times, discharged 132 times, and the two outliers in the charging data is removed. The total amount of data is 168 + 132 = 300 samples and the actual capacity values corresponding to the specific samples. Although the obvious outliers are removed, there are still slight glitch in the feature sequence due to measurement errors, etc. Therefore, one-time cubical smoothing algorithm with five-point approximation operation is used for each feature sequence to appropriately reduce the measurement error [14].

5. SOH estimation based on machine learning Algorithm

5.1. SOH estimation based on regression tree model

The Classification and Regression Tree (CART) was proposed by Breiman in 1984 [15]. It is a widely used machine learning algorithm that can be used for both classification and regression problems. The core of the classification tree is that the optimal splitting feature is selected based on the Gini index [16], and the regression tree is based on the heuristic algorithm and the least squares method to select the optimal splitting node [17]. The final fitting function is a piecewise constant function, because its core method is the least squares method, so the CART regression tree is also called the least squares regression tree.

The data of 300 samples of 4 features selected in section 4 are used as the input of the model, and the actual capacity values corresponding to the specific samples is the output of the model, and a least squares regression tree model is established. In order to obtain a training set with a more even sample distribution, each of the fourth sample in the B0006 and B0018 batteries is selected as a test set, B0006 is 168/4=42, B0018 is 132/4=33, and a total of 75 test samples. 75/300=25% of the total data set, and the remaining 75% of the samples are used as training sets.

While training the regression tree, the Bootstrap method was used to evaluate the performance [18]. The data outside the bag was used as the verification set, and the experiment was repeated 100 times. The mean square error of the estimated value of each out-of-bag data was calculated, and the mean square error was 0.0291. The standard deviation is 0.0018.

Since the prediction error of the performance evaluation is small, it is determined that the regression tree model is trained by this method, and all the training sets are used to train a certain regression tree model, and then 75 test samples are input, and the prediction mean square error is 0.0006, and the fitting effect is obtained as shown in Figure 4.
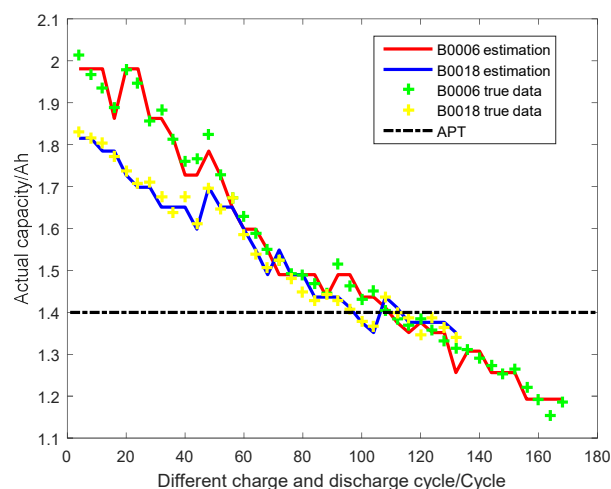

Figure 4. Least squares regression tree fitting effect

It can be seen from the experimental results that the test error is further reduced a lot. The reason is that when the performance evaluation is performed by the Bootstrap method, since the sampling with replacement is applied to the training set, the number of samples actually used for training in each experiment does not exceed the 70% training set. When the number of samples is large enough, theoretically about 36.8% of the samples will become out-of-bag samples, so the training set of the final model is increased by more than 30% compared to the evaluation experiment, and the prediction error will be further reduced. On the other hand, because of the Bootstrap method may change the distribution of real training samples due to its randomness, this may lead to an increase in estimation bias, and the generalization error may also increase [18].

5.2. SOH estimation based on random forest model

The random forest algorithm is the most representative integrated learning algorithm [18], in which the base learner is CART, so the random forest is also applicable to both classification and regression. It not only introduces the training sample interference through the Bootstrap method, but also introduces the attribute interference through the sampling without replacement method, which further improves the generalization performance of the random forest. [18]

The training set and test set of the random forest regression tree model are set up with the selection method in 5.1 subsection. The number of trees is set to 100, and the number of split attributes randomly selected by each tree is 2, and 100 times training experiments are also performed. The out-of-bag samples of the random forest regression model (out-of-bag samples of the last tree of each training random forest model) is estimated, and the mean value of mean square error is 0.0003 with a standard deviation of 0.000055.

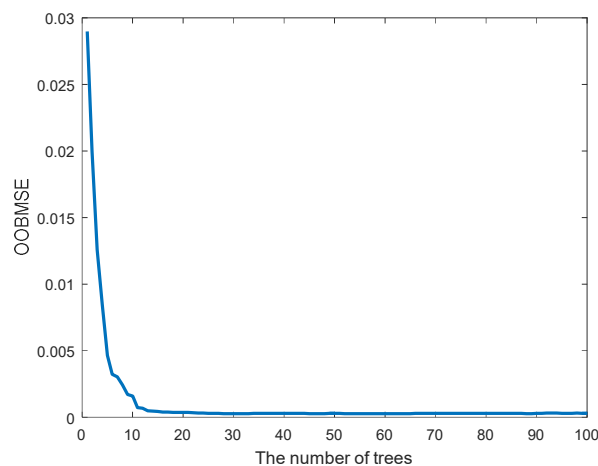The training process of the random forest regression model is shown in Figure 5.

Figure 5. The training process of the random forest regression model

The number of iterations is the number of regression trees trained in random forests. It can be seen that the mean square error converges to a small value before reaching 100 trees. In fact, the random forest is an iterative application Bootstrap method for training. It is proved that the random forest model does not need performance evaluation, and its convergent OOBMSE value can be directly applied as the evaluation of model generalization performance [19-21].

In addition, the random forest can randomly discard a certain feature sequence of the sample outside the bag, and determine the importance degree of a certain feature according to the increase of the prediction error. The judgment results of the four features of the experiment are shown in Figure 6.
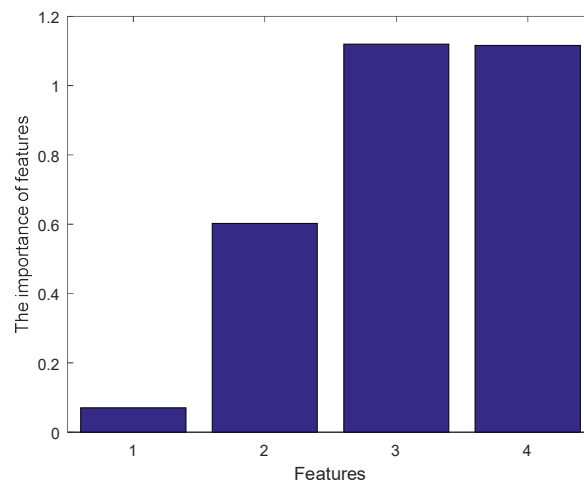
Figure 6. The importance of features

A random forest regression model was used to test 75 test samples with a mean square error of 0.0002, which is 0.0004 smaller than the single regression tree model in subsection 5.1. The fitting effect is shown in Figure 7.
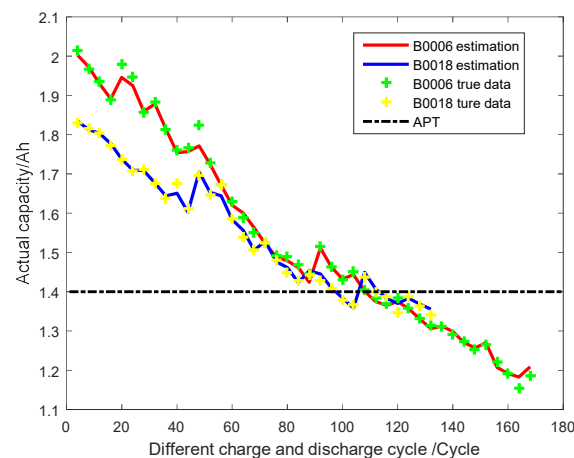

Figure 7. Random forest regression model fitting effect

Compared with the fitting effect of Figure 4., it can be seen that the fitting curve in Figure 7 has no horizontal line segment at all, and the actual capacity value of the test set does not exist for two consecutive identical values, and there are more horizontal lines in Figure 4., so the error is slightly larger.

6. Conclusion

This paper first introduces the definition of battery SOH, then data preprocessing section introduces the experimental conditions of NASA lithium-ion battery and data visualization work. By observing the time series curves of charging voltage, charging temperature, discharging voltage and discharging temperature for each charge and discharge cycle, four features are extracted. (The time required for the charging voltage to reach 4.2V in each charge and discharge cycle, the time when the charging temperature reaches the maximum value, the time when the discharging voltage reaches 2.5V, and the time when the discharging temperature reaches the maximum value) are used for SOH estimation. The

least squares regression tree model is compared with the established random forest regression model. The final test set estimation show that the mean square error of a single least squares regression tree is 0.0006, and the mean square error of random forest regression is 0.0002.

References
[1]   Yang Chuankai, et al. "LIBSVM Modeling Method for Life Prediction of Lead-Acid Battery."Distributed Energy(2018).
[2]   Lin Jiang. Research on lithium ion battery SOC estimation and RUL prediction. (2013).
[3]   Chen, Xiongzi, et al. "Probabilistic Residual Life Prediction for Lithium-ion Batteries Based on Bayesian LS-SVR." Acta Aeronautica Et Astronautica Sinica 34.9(2013): 2219-2229.
[4]   Zhang, Yang, X. Jia, and B. Guo. "Bayesian framework for satellite rechargeable lithium battery synthesizing bivariate degradation and lifetime data." Journal of Central South University 25.2(2018): 418-431.
[5]   Li, Lingling, et al. "Remaining Useful Life Prediction for Lithium-Ion Batteries Based on Gaussian Processes Mixture:."Plos One11.9(2016): e0163004.
[6]   Zhang, Lijun, Z. Mu, and C. Sun. "Remaining Useful Life Prediction for Lithium-ion Batteries Based on Exponential Model and Particle Filter."IEEE Access PP.99(2018): 1-1.
[7]   Frisk, E., M. Krysander, and E. Larsson. "Data-driven lead-acid battery prognostics using random survival forests." (2014).
[8]   Zhang, Dong, et al. "Remaining useful life estimation of Lithium-ion batteries based on thermal dynamics."American Control Conference IEEE, 2017.
[9]   Laayouj, Nabil, and H. Jamouli. "Lithium-ion Battery Degradation Assessment and Remaining Useful Life Estimation in Hybrid Electric Vehicle." 2.1(2016):37-44.
[10]   Richardson, Robert R., M. A. Osborne, and D. A. Howey. "Gaussian process regression for forecasting battery state of health." Journal of Power Sources 357(2017): 209-219.
[11]   Guangjun Liu, Zequan Fei, Bin Liang, et al. "Study on the New UPS Application Based on LFP Battery." Mechatronics (2017).
[12]   B. Saha and K. Goebel (2007). "Battery Data Set", NASA Ames Prognostics Data Repository (http://ti.arc.nasa.gov/project/prognostic-data-repository), NASA Ames Research Center, Moffett Field, CA
[13]   Jindong Zhang, Wei Tong, Yening Sun, et al. Summarize of Lithium Battery Status of Health Estimation Method. Journal of Power Supply. (2017), 15(2): 128-134.
[14]   "Analysis of data processing in flood simulation." Shanxi Hydrotechnics (2004).
[15]   BREIMAN L. "Classification and regression tree". Boca Raton: Chapman & Hall/CRC, (1984), 17-23.
[16]   Qiu, Yihui, C. Zhang, and S. Chen. "Research of Patent-value Assessment Indictor System Based on Classification and Regression Tree Algorithm." Journal of Xiamen University (2017).
[17]   Li Hang. "Statistical Learning Method." Beijing: Tsinghua University Press (2012).
[18]   Zhou Zhihua. "Machine Learning." Beijing: Tsinghua University Press, (2016).
[19]   Breiman L. "Bagging predictors" Machine Learning. Machine Learning, (1996), 24.
[20]   Wolpert D H, Macready W G. "An Efficient Method To Estimate Bagging's Generalization Error." Machine Learning, (1999), 35(1): 41-55.
[21]   Tibshirani R. Bias, "Variance and Prediction Error for Classification Rules." (1996).