On-line Remaining Charging-discharging Cycle Prediction of Lithium-ion Batteries using Cumulative Indicator

Zeyu Luo\(^1\), Hao Chen\(^1\), Xian-Bo Wang\(^{1,2}\) and Zhi-Xin Yang\(^1*\)

\(^1\) State Key Laboratory of Internet of Things for Smart City, and Department of Electromechanical Engineering, University of Macau, Macau SAR, 999078, China
\(^2\) College of Electrical Engineering, Henan University of Technology, Zhengzhou, China, 450001, China

Email: zxyang@um.edu.mo

Abstract. Prediction of the effective number of full charging-discharging cycle is valuable for lithium-ion battery (LIB) replacement and recycling. This paper proposes to construct a cumulative degradation indicator (CDI) to work as a more predictable indicator. The proposed CDI is better than the original degradation indicator (DI) in multiple criteria. In the stage of determining the end-of-life (EoL) threshold, a relevance vector machine (RVM) is introduced to screen a small number of available samples, and to reduce the prediction error. In the experimental verification stage, this paper uses LIB full-life data from NASA to verify the early and long-term prediction performance of RCDC using a small sample. The experimental results show that when the proportion of training data approaches 50%, the prediction error gradually converges to the actual value.

1. Introduction

With the development of automobile industry in decades, million of LiBs installed in vehicles are reaching their end-of-life (EoL), and need to be properly handled to prevent impacts on the environment. The most important thing is how to accurately estimate the remaining useful life (RUL) of LiBs, thereby improving battery utilization and reducing the risk of environmental harmfulness.

To accurately describe the usable cycles of LiBs, this paper uses the remaining charging-discharging cycles (RCDCs) to replace the term RUL. As shown in Eqs. (1)-(3), the indicators state-of-capacity (SoC) and state-of-health (SoH) are commonly selected to measure and evaluate the health status of LiBs \([1-3]\). In addition, the performance of LiBs also varies due to the number of charging-discharging cycles, temperature changes, and the solid electrolyte interface conversion. Therefore, an accurate and robust RUL prediction algorithm is essential to improve performance and optimize the charging-discharging process of LiBs. Generally, 70\%-80\% of SoH is regarded as the threshold range for the battery replacement \([4,5]\).

\[
\text{SoC}(t_i) = \text{SoC}(t_{i-1}) + \frac{1}{Q_f} \int_{t_{i-1}}^{t_i+\Delta t} i_{bat}(t)dt \times 100\% \tag{1}
\]

\[
\text{SoH}(t_i) = \frac{C(t_i)}{C(t_1)} (i = 1, 2, ..., n) \tag{2}
\]

Many studies are based on determination of RCDC, while there are still some problems. If the time series analysis methods are used to predict RCDC, the proportion of training data provided is comparatively small (i.e., usually it is considered the proportion of training data exceeds 50% of the...
overall data can be regarded as an ideal way, it helps to obtain a guaranteed regression result), and thus the early prediction and long term prediction efficiency may be limited. The quantities of training data are relatively small when RCDC is predicted by time series analysis methods. The prediction result however, is considered guaranteed only if more than 50% of overall data is taken as training data. Thus, the efficiency of early and long-term prediction may be limited. Besides, the works in [6,7] revealed that the battery capacity regeneration phenomena (CRP) hinders RCDC prediction directly. Figure 1 shows the challenges met in the prediction.

To deal with the problem of insufficient training data proportion and the challenge brought by the nonlinear CRP, this paper attempts to establish a new cumulative degradation indicator (CDI) to replace the original SoH-based degradation indicator (DI) for RCDC prediction. Then, to establish the "DI-CDI" relationship, the authors uses RVM to obtain a small amount of relevant vectors (RVs), and selects these vectors to conduct curve fitting to determine the EoL threshold of CDI. The proposed CDI is consistent with the original DI in terms of overall monotonousness and synchronization, but the linearity is better.

Figure 1. SoH Curve of B0006 battery.

2. Methodology
The proposed prediction framework is shown in figure 2. It includes online degradation indicator acquisition, discrete wavelet transform (DWT) denoising, cumulative features calculation, RVs selection using RVM, EoL estimation on CDI curve, and RCDC prediction. In this paper, the normalized DI of LIBs is given in equation 3.

\[
DI(t_i) = 1 - SoH(t_i) = 1 - \frac{c(t_i)}{c(t_{1})} \quad (i = 1,2, ..., n)
\]

\[
CDI(t_i) = \frac{\sum_{i=1}^{n} DI(t_i)'(t_i)}{\sqrt{\sum_{i=1}^{n} \text{abs}(DI(t_i)(t_i))}} \quad (i = 1,2, ..., n)
\]

2.1. CDI Calculation
Although the DI of a LIBs can reflect the real-time status of the capacity. It cannot reflect the influence of the past state on the current state, that is, the historical relevance of the time series. And affected by CRP, the DI curve shows non-linear characteristics, which is not conducive to curve fitting, and is not conducive to RCDC prediction. To reveal the historical relevance problem, as shown in equation 4, this paper uses cumulative degradation characteristics to reflect the trend and monotonicity of historical degradation series. As shown in table 1, CDI is far better than the DI curve in terms of monotonicity and linearity.

To evaluate the monotonicity and trend of degradation characteristics, this paper introduces two indicators: monotonicity and trending index [8]. The expressions of the two indicators are shown in equation 5-6. The value of \(M\) ranges from 0 to 1. \(M = 1\) indicates the highest monotonicity, and
$M = 0$ shows there is no monotonicity. $T$ represents the correlation between CDI and the charging-discharging cycles $t_i$, and its value range is between -1 and 1. $T = 1$ indicates the best correlation.

$$\text{Monotonicity: } M = \frac{N_p - N_n}{n - 1}$$

$$\text{Trend: } T = \frac{n \sum_{i=1}^{n} (x(t_i) - \bar{x}) (\Sigma_{i=1}^{n} x(t_i) - \sum_{i=1}^{n} x(t_i))}{\sqrt{n \sum_{i=1}^{n} (x(t_i) - \bar{x})^2 (\Sigma_{i=1}^{n} x(t_i) - \sum_{i=1}^{n} x(t_i))^2}}$$

2.2. RVM Regression for RV Selection and Curve Fitting

RVM regression is a machine learning algorithm that combines Bayesian theory with SVM. Compared with SVM, the kernel function of RVM has no special restrictions, and its sparsity is better [9]. Fewer hyperparameters reduce the calculation amount of the kernel function. Given a data series $\{D_{l_i}, CDI_i\}_{i=1}^{n}$, RVM regression describes the links between target and input variables using a linear model of the following expression.

$$z_i = y(D_{l_i}) + \epsilon_i$$

where $y(D_{l_i})$ is the linear combination of kernel function, $\epsilon_i = \sigma^2$ meets the normal distribution. $y(D_{l_i})$ can be expressed as:

$$y(D_{l_i}) = \omega^T K(D_{l_i}, D_{l_i})$$

where $\omega = (\omega_1, \omega_2, ..., \omega_n)$ is the weight and $K(\cdot)$ is a predetermined kernel function. More details about the choice of kernel functions can be found in [9].

| Table 1. CDI Fitness Evaluation. |
|---------------------------------|
|                                |
| **B0005** | **B0006** | **B0007** | **B0018** |
|----------|----------|----------|----------|
| DI       | CDI      | DI       | CDI      |
| 0.5689   | 0.9940   | 0.6766   | 0.9940   |
| 0.7089   | 0.9940   | 0.4371   | 0.9940   |
| 0.6794   | 0.9940   | 0.6794   | 0.9924   |

**Monotonicity:** 0.5689 0.9940 0.6766 0.9940 0.4371 0.9940 0.6794 0.9924

**Trend:** 0.9877 0.9962 0.9817 0.9993 0.9881 0.9972 0.9696 0.9995
When the weight corresponding to these hyperparameters is zero, the sparsity of RVM is met. The inputs corresponding to the remaining nonzero weights are called RVs. As shown in figure 3(a), the red circles are the selected RVs, which are then used for polynomial fitting (as shown in figure 3(b)).

According to equation 9, this paper adopts a second-order polynomial fitting model $g(t_i)$ to obtain the fitting data of CDI in the time ranging from $t_i = t_{1}$ to $t_i = t_{EoL}$. The EoL is deemed to have been reached when the first time CDI satisfies the threshold condition $CDI(t_{EoL})^* \leq CDI(t_{EoL})$. Then, RCDC can be given in equation 10, and the RCDC error shown in equation 11.

$$CDI(t_i \rightarrow t_{EoL})^* = g(t_i) = a_1 \cdot t_i^2 + a_2 \cdot t_i + a_3$$  \hspace{1cm} (9)

$$RCDC = t_{EoL} - t_i$$  \hspace{1cm} (10)

$$Er = t_{EoL} - t_{EoL}$$  \hspace{1cm} (11)

where $a_i=1,2,3$ denote the weights of model $g(t_i)$, $t_{EoL}$ is the predicted endpoint, while $t_{EoL}$ means the real endpoint.

3. Experimental Verification

The battery capacity degradation data used in this study comes from NASA PCoE. There are four sets of batteries: B0005, B0006, B0007, and B0018, all of which are commercial 18650 LIBs.

According to Eq. (4), DI and CDI curves of remaining three sets of LIBs are shown in figure 4. As shown in table 1, the monotonicity and trend indexes of CDIs show better than DIs of three groups of LIBs. In order to facilitate the comparison of curve shapes, this paper multiplies the amplitude of DI by ten times. Obviously affected by the CPR effect, there are many depressions on the DI curves, and these local declining trends may affect the RCDC performance. However, it is not difficult to see that the CDI curves of the three groups of LIBs all increase monotonically, and there is no local fluctuation that occurs on original DI curves. As shown in figure 5, the gray curve reflects the "DI-CDI"
relationship, which is nonlinear. In this paper, the RVM regression method is used to determine the potential RVs. Based on these RVs (e.g., as shown in figure 3(a)), a simple first-order polynomial fitting function can be established, this function is helpful to determine the CDI threshold of the EoL location.

![Figure 5](image1.png)

**Figure 5.** DI-CDI curve fitting. The actual EoL threshold indicated by blue circle, and the predicted result by red circle. (a). B0006. (b). B0007. (c). B0018.

The overall performance of the RCDC prediction to all groups of LIB data is shown in figures 6-7(a), the prediction errors are shown in figures 6-7(b). When \( t_i > 60 \), the prediction error is relatively small and will gradually be consistent with the actual observed value. It means that the available observation data account for 35.7\% of the overall data dimension.

As shown in figures 6-7, when the prediction point \( t_i \) is between 60 and 80 (accounting for 35\% to 47\% of the total data length), the prediction error of RCDC gradually decreases. When \( t_i = 26 \) for the B0006 battery (see figure 7), the prediction error has gradually converged to an acceptable range. It shows that the method proposed in this paper can indeed realize the early and long-term prediction of RCDC.

![Figure 6](image2.png)

**Figure 6.** RCDC Prediction performance of B0005. (a). RCDC prediction. (b). prediction error.

![Figure 7](image3.png)

**Figure 7.** RCDC Prediction performance of B0006. (a). RCDC prediction. (b). prediction error.

4. Conclusion
This paper adopts the idea of online prediction, applies the original DI degradation data of lithium-ion batteries to obtain a new degradation index CDI employing cumulative summation. It is significantly better than DI in monotonicity, trend, and linearity. Based on the fitting curve of DI-CDI, this paper introduces RVM to select the available DI variables for RV selection. The chosen RVs are further used for polynomial fitting to determine the equivalent EoL threshold on the CDI curve, which provides a prediction basis for the RCDC estimation on the CDI curve. The experimental results show
that when the proportion of training data approaches 50%, the prediction error gradually converges to the actual value. The proposed method can also be transplanted to commonly used portable terminal devices.

Acknowledgments
This work was funded in part by the Science and Technology Development Fund, Macau SAR (Grant no. 0018/2019/AKP, 0008/2019/AGJ, and SKL-IOTSC-2021-2023), in part by the Ministry of Science and Technology of China (Grant no. 2019YFB1600700), in part by the Guangdong Basic and Applied Basic Research Foundation (Grant no. 2020B1515130001), in part by the Science Foundation of Henan University of Technology (Grant no. 2019BS004, 2016QNJH02), in part by the Cultivation Programme for Young Backbone Teachers in Henan University of Technology (Grant no. 21420171), in part by the Natural Science Project of Henan Province (Grant no. 202300410117, 20A470003, and 202102210301).

References
[1] Liang P, Deng C, Wu J, Li G, Yang Z & Wang Y 2019 Intelligent fault diagnosis via semisupervised generative adversarial nets and wavelet transform *IEEE Transactions on Instrumentation and Measurement* **69**(7): 4659-4671.
[2] Liang P, Deng C, Wu J, Yang Z, Zhu J & Zhang Z 2020 Single and simultaneous fault diagnosis of gearbox via a semi-supervised and high-accuracy adversarial learning framework *Knowledge-Based Systems* **198**: 105895.
[3] Wang X B, Yang Z X, Wong P K & Deng C 2019 Novel paralleled extreme learning machine networks for fault diagnosis of wind turbine drivetrain *Memetic Computing* **11**(2): 127-142.
[4] Liang P, Deng C, Wu J, Li G, Yang Z & Wang Y 2019 Intelligent fault diagnosis via semisupervised generative adversarial nets and wavelet transform *IEEE Transactions on Instrumentation and Measurement* **69**(7): 4659-4671.
[5] Wang X B, Yang Z X & Yan X A 2017 Novel particle swarm optimization-based variational mode decomposition method for the fault diagnosis of complex rotating machinery *IEEE/ASME Transactions on Mechatronics* **23**(1): 68-79.
[6] Zhong J H, Wong P K & Yang Z X 2016 Simultaneous-fault diagnosis of gearboxes using probabilistic committee machine *Sensors* **16**(2): 185.
[7] Zhong J H, Liang J, Yang Z X, Wong P K & Wang X B 2016 An effective fault feature extraction method for gas turbine generator system diagnosis *Shock and Vibration*.
[8] Yang Z, Wong P K, Vong C M, Zhong J & Liang J 2013 Simultaneous-fault diagnosis of gas turbine generator systems using a pairwise-coupled probabilistic classifier *Mathematical Problems in Engineering*.
[9] Yang Z, Hoi W I & Zhong J 2011 Gearbox fault diagnosis based on artificial neural network and genetic algorithms *Proceedings 2011 International Conference on System Science and Engineering*: 37-42.