Remaining Useful Life Prediction Driven by Multi-source Data for Batteries in Electric Vehicles

Jiahuan Lu¹, Xinggang Li¹, Hao Lei¹, Yonggang Liu², Rui Xiong¹,*

¹National Engineering Laboratory for Electric Vehicles, School of Mechanical Engineering, Beijing Institute of Technology, Beijing, 100081, China
²State Key Laboratory of Mechanical Transmissions, Department of Automotive Engineering, Chongqing University, Chongqing, 400044, China

Abstract

Predicting battery remaining useful life (RUL) is used for early warning of battery aging failure and providing instructions of battery maintenance and recycling. The existing RUL prediction focus too much on decreasing the dependence of aging tests, neglecting the value of test data. In this regard, a battery RUL prediction method driven by multi-source data is proposed for EVs to make full use of the aging test data from other cells. Six lithium-ion batteries were used to verify the effectiveness of the method. The results show that the prediction error is less than only 1 cycle in the case of capacity ‘diving’. In conclusion, the proposed method effectively improves the performance of RUL prediction by using multi-source data, and provides a solution for battery management in the era of big data.

Keywords: remaining useful life, aging, battery, electric vehicle, multi-source data, neural network

1. Introduction

Battery remaining useful life (RUL) is designed for evaluating the remaining useful cycles before the end of life of the batteries in electric vehicles (EVs), which is treated as an important index for evaluating the safety and durability [1,2]. In practical application, accuracy RUL prediction effectively realizes the early warning of battery degradation failure, and scientifically provides instructions for the maintenance and recycling of vehicle batteries. Therefore, the development of battery RUL prediction method is of great significance to ensure the safety and durability of EVs [3].

The research of battery RUL prediction method has always been challenging, because it takes lots of manpower and material resources to complete an aging test for at least six months [4]. More importantly, even if a battery aging test is successfully completed, the battery inconsistencies caused by manufacturing are contained in the test data, making it difficult for researchers to apply the experimental results to other batteries. Therefore, most of the previous battery RUL prediction methods have been devoted to reducing the dependence of aging tests. For example, some researchers predicted RUL by fitting degradation trajectory [5]. To reduce the dependence of off-line tests, they focus on the preprocessing methods of the battery’s own historical data to improve the performance of curve fitting [6,7]. Some researchers use filter observers to predict RUL [8]. They are devoted to the development of filter observers with good convergence and high accuracy to improve the performance of the degradation trajectory model in tracking the battery’s own historical data [9,10]. The battery RUL can be predicted even without the initial references of parameters provided by off-line aging tests. Some researchers treat the RUL prediction as time series prediction, and predict by a time series model without any offline test data [11].

Although the above-mentioned battery RUL methods operate almost without aging tests and achieve good battery RUL prediction, some battery aging test is actually indispensable before practical application. This is because EV manufacturers need to know the aging characteristics of batteries in advance to evaluate the reliability of the system. Consequently, pursuing a non-experimental method for RUL prediction is unnecessary. On the contrary, the common aging mechanism of batteries within the same material is included in the test data, although the aging data of different batteries are inconsistent. Accordingly, the main contribution of this paper is to make full use of the common information from the off-line aging test data of multiple batteries to improve RUL prediction performance.

The rest of the paper is organized as follows: section 2 presents the experimental results prepared for the research; section 3 shows the procedure of the proposed method; section 4 analyses the results of the proposed method; finally, some conclusions are summarized in section 5.
2. Experiment

To support this study, six lithium-manganese-cobalt-oxide 18650-battery cells of the same type, numbered Cell 1 to Cell 6, are employed for aging tests. The specifications of these cells are shown in Table 1.

<table>
<thead>
<tr>
<th>Table 1 Specifications of the experimental batteries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specification item</td>
</tr>
<tr>
<td>Nominal capacity</td>
</tr>
<tr>
<td>Charging cut-off voltage</td>
</tr>
<tr>
<td>Charging standard current</td>
</tr>
<tr>
<td>Charging cut-off current</td>
</tr>
<tr>
<td>Discharge cut-off voltage</td>
</tr>
<tr>
<td>Discharge standard current</td>
</tr>
</tbody>
</table>

The aging tests are carried out at 25°C provided by a DGBELL incubator. The cells are cycled charged and discharged by a LANHE battery test equipment, where each cycle includes a charge profile of constant current constant voltage (CCCV) with 1.2A and a discharge profile of constant current with 1.2A. A rest profile of 1-hour is set between the charge profile and the discharge profile. When the capacity decreases by 20%, the cell is considered to be failure and the test is terminated.

The test results of the six cells are shown in Figure 1. It is seen that the capacity degradation trajectories of the six cells are different, even though the experimental settings are exactly the same. In the following sections, these very different data will be treated as the multi-source data to improve the RUL prediction.

![Figure 1 Capacity degradation trajectories of the six cells](image)

3. Methodology

3.1 Preprocessing of multi-source data

The results of aging test are originated from multiple cells, which can be treated as multi-source data. For the \(i\)-th of the cells, the result of aging test is expressed as:

\[
\begin{aligned}
&k^i = \{k^i_1, k^i_2, \ldots, k^i_j\}, \quad i = 1, 2, \ldots, N \\
&h^j = \{h^j_1, h^j_2, \ldots, h^j_m\}, \quad j = 1, 2, \ldots, m
\end{aligned}
\] (1)

where the utilization of font bolding represents data set, the superscripts denote the serial number of the cell, the subscripts denote the serial number of sampling points. \(N\) is the scale of multi-source data, \(m\) is the scale of sampling, \(k\) is the cycle number, \(h\) is the corresponding state of health (SOH), which is defined as:

\[
h^j_i = \frac{C^i_j}{C_0} \times 100\%
\] (2)

where the subscript 0 indicates that the battery is in the factory state, \(C\) denotes the maximum capacity.

For convenience, multi-source data are stored in the form of smooth spline functions. For the \(i\)-th of the cells, the smooth spline function \(K(h)\) can be solved by:

\[
\arg \min_{K(h)} \left\{ \rho \sum_{j=1}^{m} (k_j - K'(h_j))^2 + (1 - \rho) \int \left( \frac{d^2 K'}{dh^2} \right)^2 \right\} dh
\] (3)

where \(p\) denotes the smooth coefficient.

After the smoothing step, A set of functions for multi-source data is formed as:

\[
K = \left\{K^1(h), K^2(h), \ldots, K^N(h)\right\}
\] (4)

3.2 Modeling of multi-source data driven

For a given cell to be predicted, the number of cycles it has experienced and the corresponding SOH are available, which is expressed as:

\[
\begin{aligned}
&k^* = \{k^*_1, k^*_2, \ldots, k^*_j\}, \quad j = 1, 2, \ldots, \hat{m} \\
&h^* = \{h^*_1, h^*_2, \ldots, h^*_m\}, \quad j = 1, 2, \ldots, m
\end{aligned}
\] (5)

where the superscript * denotes the cell to be predicted.

To introduce the information from the multi-source data, a classical machine learning method, feedforward neural network (FNN), is employed here to establish a multi-source data driven model.

Firstly, the initialization of FNN is performed. The input includes the historical SOH of the cell to be predicted and the SOH of multi-source data, and the output is the number of cycles of the cell to be predicted. Thus, the dimension of input and output are expressed as:

\[
\begin{aligned}
&D_{\text{input}} = N + 1 \\
&D_{\text{output}} = 1
\end{aligned}
\] (6)

where \(D_{\text{input}}\) and \(D_{\text{output}}\) are the dimension of input and output, respectively. One hidden layer is initialized, and the number of neurons it contains can be determined by:

\[
D_{\text{hidden}} = \left\lfloor \sqrt{D_{\text{input}} + D_{\text{output}}} \right\rfloor + \alpha
\] (7)

where \(D_{\text{hidden}}\) is the number of neurons in the hidden layer, \(\alpha\) is an integer coefficient from 1 to 10.

Secondly, a set of training samples is formulated as:


\[
\begin{align*}
    \mathbf{x} &= \left[ h^*, K^1(h^*), K^2(h^*), \ldots, K^N(h^*) \right] \\
    \mathbf{y} &= \mathbf{k}^* 
\end{align*}
\]

where \( \mathbf{x} \) and \( \mathbf{y} \) denote the training input sample and training output sample, respectively. It is worth noting that training samples need to be standardized before training.

Thirdly, Levenberg-Marquardt method is applied to train the multi-source data driven model based on \( \mathbf{x} \) and \( \mathbf{y} \), which can be described as:

\[
q(t + 1) = q(t) - [J^T J + \mu \mathbf{I}]^{-1} J^T e 
\]

where \( t \) is the number of iterations, \( q(t) \) is the connection weight vector or threshold vector between the layers at the \( t \)-th iteration, \( \mathbf{I} \) is the unit matrix, \( J \) is the Jacobian matrix, \( \mu \) is an iteration coefficient that decreases after each successful iteration and increases when the error increases after the trial iteration, \( \mathbf{e} \) is the error vector of the FNN, which is calculated by:

\[
\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} 
\]

where \( \hat{\mathbf{y}} \) is the predicted output of FNN, which is calculated by the propagation equation:

\[
\begin{align*}
    \mathbf{H} &= f \left( \mathbf{X} \cdot \mathbf{W}_1 - \mathbf{b}_1 \right) \\
    \mathbf{Y} &= \mathbf{H} \cdot \mathbf{W}_2 - \mathbf{b}_2 
\end{align*}
\]

3.3 Prediction driven by multi-source data

After training, the FNN can predict the future capacity degradation trajectory by giving a target prediction matrix \( \mathbf{x}^{tar} \):

\[
\mathbf{x}^{tar} = \left[ h^{tar}, K^1(h^{tar}), K^2(h^{tar}), \ldots, K^N(h^{tar}) \right] 
\]

where \( h^{tar} \) is a vector of target SOH. By substituting \( \mathbf{x}^{tar} \) into Equation (11), the cycles that the cell needs to be experienced in the future can be predicted. The RUL of the cell can be solved by:

\[
\mathbf{RUL} = \mathbf{Y}^{tar} - \mathbf{k}^* 
\]

where \( \mathbf{RUL} \) is a vector that consists of RUL values corresponding to different target SOH, \( \mathbf{Y}^{tar} \) is the output of FNN based on \( \mathbf{x}^{tar} \).

![Figure 2 Results of RUL prediction based on MSDD and PF: (a)~(c) start with 95%, 90%, 85% SOH in case 1, (d)~(f) start with 95%, 90%, 85% SOH in case 2, (g)~(i) start with 95%, 90%, 85% SOH in case 3](image-url)
4. Result and discussion

To verify the effectiveness of the proposed method, three cases are set up, as shown in Table 2. A commonly used RUL prediction method, particle filter-based method (PF) [12], is introduced as a benchmark for verification. Notice that the initialization and update of the parameters for PF is based only on the cell’s own historical data. For convenience, the multi-source data driven prediction method is abbreviated as MSDD.

<table>
<thead>
<tr>
<th>Case</th>
<th>Multi-source data</th>
<th>Prediction target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1</td>
<td>Cell 1</td>
<td>Cell 6</td>
</tr>
<tr>
<td>Case 2</td>
<td>Cell 1/2/3</td>
<td>Cell 6</td>
</tr>
<tr>
<td>Case 3</td>
<td>Cell 1/2/3/4/5</td>
<td>Cell 6</td>
</tr>
</tbody>
</table>

The SOH of 95%, 90% and 80% are taken as three starting points for simulation, the results of RUL prediction are shown in Figure 2. From Figure 2 (a) to (c), the MSDD can be driven by only one cell, and the prediction accuracy for three starting points is higher than PF. From Figure 2 (d) to (f), increasing the number of data sources to three do not significantly improve the prediction accuracy, but increase the nonlinear characteristics within the prediction trajectory. This is because different test trajectories in multi-source data contribute to the prediction, which indicates that the MSDD has the potential to perfectly fit the future aging trajectory. From Figure 2 (g) to (i), the nonlinearity of the predicted trajectories is further enhanced when the number of multi-source data is increased to 5. The deterioration of prediction as shown in Figure 2 (g) is due to the concentration of multi-source trajectories in the early stage of aging. Notice that the MSDD successfully predicts the ‘diving’ trajectory as shown in Figure 2 (i), almost coinciding with the actual data, and the RUL error is only 1 cycle. This shows that choosing appropriate multi-source data can effectively solve the problem of capacity ‘diving’ in severe aging stage.

5. Conclusion

This paper proposes a RUL prediction method driven by multi-source data for batteries in EVs, which can effectively make full use of the off-line test data from other cells to improve RUL prediction. Only driven by one cell, the prediction of this method can be more precise than PF. By choosing the appropriate multi-source data, the error of this method is only 1 cycle in the case of capacity ‘diving’.

With the development of cloud computing, this method can directly treat the running data of other EVs equipped with the same type of batteries as the multi-source data, which provides a solution for battery management in the era of big data.

Acknowledgement

This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFB0104100. The systemic experiments of the lithium-ion batteries were performed at the Advanced Energy Storage and Application (AESA) Group, Beijing Institute of Technology.

Reference