Robust prognostics for state of health estimation of lithium-ion batteries based on an improved PSO–SVR model

Taichun Qin a, Shengkui Zeng a,b, Jianbin Guo a,b,*

a School of Reliability and Systems Engineering, Beihang University, Beijing 100191, China
b Science and Technology on Reliability and Environmental Engineering Laboratory, Beijing 100191, China

A R T I C L E   I N   F O

Article history:
Received 25 May 2015
Received in revised form 21 June 2015
Accepted 29 June 2015
Available online xxxx

Keywords:
Lithium-ion battery
Prognostic
State of health
Support vector regression
Particle swarm optimization

A B S T R A C T

State of health (SOH) estimation of lithium-ion batteries is significant for safe and lifetime-optimized operation. In this study, support vector regression (SVR) is employed in battery SOH prognostics, and particle swarm optimization (PSO) is employed in obtaining the SVR kernel parameter. Through a new validation method, the proposed PSO–SVR model in this paper can well grasp the global degradation trend of SOH and is little affected by local regeneration and fluctuations. The case study shows that compared with the eight published methods, the proposed model can obtain more accurate SOH prediction results. Even SOH prediction starts from the cycle near capacity regeneration, the proposed model still can grasp the global degradation trend. Furthermore, the improved PSO–SVR model has great robustness when the training data contain noise and measurement outliers, which makes it possible to get satisfactory prediction performance without pre-processing the data manually.

1. Introduction

With the advantages of high energy density and lightweight, lithium-ion batteries are widely used in various kinds of portable devices, electric cars, spacecraft, etc. [1,2]. It is useful to accurately estimate state of charge (SOC) and state of health (SOH), as they play important roles in safe and reliable usage of lithium-ion batteries [3,4]. However, compared to the relatively mature study on SOC issue, SOH research still stays in the initial stage [5,6].

In general, the material properties and manufacturing assemblies are various from battery to battery [7]. Also, the performance of lithium-ion batteries degrades with many factors, such as environment temperature, and charging and discharging current, depth of discharge [3,8]. Moreover, it is rather difficult to measure parameters of the electrochemical reaction of lithium-ion batteries [4]. Therefore, SOH prognostics of data-driven approach seem more applicable than those of physical-model approach. In recent years, many algorithms based on data-driven approach have been applied to SOH prognostics, such as extended Kalman filter (EKF) [9], particle filtering (PF) [5], fuzzy logic [10], neural networks [11], and regressions [12]. However, with poor robustness, these algorithms are not good at dealing with limited data that contain uncertain information. As prediction errors are accumulating over time, these algorithms cannot be flexibly used in long-term prediction.

The battery capacity can be easily obtained and has a clear physical meaning [13], so capacity becomes the most common used health indicator of batteries. However, the uncertainty of capacity time series, generated by capacity regeneration [14], unexpected fluctuations, measurement error and other factors [7], is a big challenge for battery SOH prognostics. Olivares [15] has detected and isolated the effect of regeneration phenomena within the life-cycle model through a particle-filtering-based prognostics framework. Treating regeneration as a cycle process, Liu [14] has described capacity regeneration and degradation by using combination Gaussian progress function regression. With wavelet analysis, He [16] has separated local regeneration and fluctuation characteristics from global degradation characteristic, which makes it possible to analyze the two characteristics respectively. However, how to obtain accurate global degradation prediction of SOH with raw capacity data is still a topic that is worth studying.

In this paper, based on particle swarm optimization (PSO), a new approach to the optimal selection of support vector regression (SVR) parameters is proposed. This PSO–SVR model can well grasp the global degradation trend, despite the fact that the capacity time series contain regeneration, fluctuations, noise and outliers.

© 2015 Elsevier Ltd. All rights reserved.
2. Methods

2.1. Support vector regression

\( \varepsilon \)-SVR is one of the most common used SVR models [17]. When the dataset \( \{x_i, y_i\} \) presents a linear relationship, the regression function can be given by:

\[
f(x) = \mathbf{w} \phi(x) + b
\]

where \( \mathbf{w} \) and \( b \) denote the undetermined parameters.

The optimization object can be expressed as [18]:

\[
\min \frac{1}{2} \| \mathbf{w} \|^2 + C \sum (\xi_i + \xi'_i)
\]

\[
\text{s.t.} \quad \begin{cases} 
    y_i - \mathbf{w} \phi(x_i) - b \leq \varepsilon + \xi_i, \\
    y_i - \mathbf{w} \phi(x_i) + b \leq \varepsilon + \xi'_i, \\
    \xi_i, \xi'_i \geq 0
\end{cases}
\]

where \( \varepsilon \) is the width of \( \varepsilon \)-tube, constant \( C \) (positive) is the penalty degree of the sample whose error is larger than \( \varepsilon \), and slack variables \( \xi_i \) and \( \xi'_i \) are the distances between actual values and the nearer boundary of \( \varepsilon \)-tube.

For nonlinear cases, they can be converted into linear cases through kernel functions. The radial basis function (RBF) kernel is used in this paper:

\[
K(x_i, x_j) = \exp\left(-\frac{1}{2} \frac{\|x_i - x_j\|^2}{\sigma^2}\right)
\]

where \( \sigma \) is the parameter of RBF.

2.2. Particle swarm optimization

PSO was proposed by Kennedy and Eberhart in 1995 [19]. Its basic idea is to obtain the optimal solution (also called particle) by minimizing a fitness function. The velocity and place of every particle will be updated by [20,21]:

\[
v^{k+1}_i = v^k_i + c_1 r_1 (p_{best}^k - x^k_i) + c_2 r_2 (g_{best}^k - x^k_i)
\]

\[
x^{k+1}_i = x^k_i + v^{k+1}_i
\]

where \( v^k_i \) denotes the velocity of particle \( i \) at iteration \( k \), \( x^k_i \) denotes the place of particle \( i \) at iteration \( k \), \( p_{best}^k \) is the optimal solution (individual optimal solution) of particle \( i \) until \( k \)th iteration, \( g_{best}^k \) is global optimal solution of all particles until \( k \)th iteration, \( w \) denotes inertia weight coefficient, \( r_1 \) and \( r_2 \) are random numbers between 0 and 1, and \( c_1 \) and \( c_2 \) are positive acceleration constants.

2.3. Improved PSO–SVR

From Eqs. (2), (3) and (4), the parameters \( \varepsilon \), \( C \) and \( \sigma \) need to be determined when we use \( \varepsilon \)-SVR model for nonlinear regression. Among them, both \( \varepsilon \) (the width of \( \varepsilon \)-tube) and \( C \) (the penalty degree of samples) have clear statistical meanings. They can reflect the commonality of the same type batteries and can be determined based on experience. On the other hand, nuclear parameter \( \sigma \) does not have a definite meaning, and it can reflect the individuality of different batteries. As a result, we expect to obtain an appropriate value of \( \sigma \) for each battery through PSO algorithm based on its earlier capacity time series.

An appropriate criterion is important to evaluate the fitting (or prediction) performance with different values of the nuclear parameter. Among the criteria, mean squared error (MSE) is the most widely used one. However, MSE cannot directly describe the correlation of the fitting (or prediction) curve and the actual SOH curve. Therefore, in this paper, a weighted sum of MSE and coefficient of determination (denoted by \( R^2 \)) is chosen as the criterion to evaluate the fitting (or prediction) effects. The weighted sum of MSE and \( R^2 \) is denoted by \( MC \), where MSE, \( R^2 \) and MC are shown as:

\[
MSE = \frac{1}{N} \sum_{i=1}^{N} (f(x_i) - y_i)^2
\]

\[
R^2 = \frac{\left( \sum_{i=1}^{N} f(x_i)^2 - \left( \sum_{i=1}^{N} f(x_i) \sum_{i=1}^{N} y_i \right) \right)^2}{\left( \sum_{i=1}^{N} y_i^2 - \left( \sum_{i=1}^{N} y_i \right)^2 \right)}
\]

\[
MC = MSE - aR^2
\]

where \( a \) is a MSE and \( R^2 \).

During the training stage, model validation is required to obtain the optimal value of the model parameter. In \( k \)-fold cross-validation, the most commonly used validation, the dataset (training samples) \( D \) is randomly split into \( k \) mutually exclusive subsets \( D_1, D_2, \ldots, D_k \). The model is trained and tested \( k \) times. In each time, it is trained on \( D \) \( \setminus D_i \) and tested on \( D_i \), where \( i \in \{1, 2, \ldots, k\} \). The output of cross-validation is the average of all the \( k \) testing errors.

SOH prognostics focus on predicting future degradation trend of lithium-ion batteries by using earlier degradation time series. Therefore, \( k \)-fold cross-validation that treats all training samples equally seems not very reasonable. A new validation method is shown in Fig. 1. In this validation method, the model is tested on all the training samples (dataset \( D \)) rather than on the subset \( D_i \). It is to ensure that the good performance of the model in the prediction stage is due to the good grasping of the degradation trend in the training stage rather than by accident.

Please cite this article as: T. Qin, et al., Robust prognostics for state of health estimation of lithium-ion batteries based on an improved PSO–SVR model, Microelectronics Reliability (2015), http://dx.doi.org/10.1016/j.microrel.2015.06.133
The fitness function is described as Eq. (10) and $F$ is the output:

$$ F = \frac{1}{2K} \left( \sum_{i=1}^{k} M^i C(i) + k M C(k + 1) \right). \quad (10) $$

The fitness function is the weighted average of two parts. In the first part, a small number of training samples are used in each cycle to make the model able for long-term extrapolation. In the second part, the (k-1) parts of the training samples are used to make the model have the ability of backward extrapolation. Moreover, the multiplication factor $k$ in the second part is to balance the effects of the two parts. The smaller the value of $F$ is, the better the performance of the SVR model is.

The steps of the improved PSO–SVR are briefly described as follows:

Step 1: Take first $m$ training data from a total of $M$.
Step 2: Obtain the optimal value of parameter $\sigma$ by PSO. The validation process (Fig. 1) is nested in the PSO algorithm. Specifically, the fitness function (Eq. (10)) is the optimization objective of PSO and the values of parameter $\sigma$ are given by every loop of PSO.
Step 3: Assign the optimal value to $\sigma$, and train the SVR model using first $m$ samples.
Step 4: Use the trained model to predict the future degradation of batteries, and use the last $(M-m)$ data to evaluate the prediction performance.

3. Experiments and analysis

3.1. Raw data of lithium-ion batteries

The lithium-ion battery degradation experiment data involved in this paper are from NASA Ames Prognostics Center of Excellence (PCoE) [22]. The 18650 sized batteries (No. 5, No. 6, No. 7 and No. 45) were run under three different operational conditions (charge, discharge and impedance) at temperature of 24 °C (No. 5, No. 6 and No. 7) or 4 °C (No. 45). Charging current was constant with a level of 1.5A until the battery voltage achieved 4.2 V. Then the batteries were continued in a constant voltage mode until the charge current fell to 20 mA. For batteries No. 5, No. 6 and No. 7, discharge was run at a constant current level of 2A until the battery voltage dropped to 2.7 V, 2.5 V and 2.2 V respectively. For battery No. 45, discharge current level was 1A until the battery voltage dropped to 2 V.

In this paper, the capacity is used to describe battery SOH. The definition is shown as:

$$ \text{SOH} = \frac{C_i}{C_0} \times 100\% \quad (11) $$

where $C_i$ is the capacity at $i$th degradation cycle and $C_0$ is the initial capacity. The SOH values of four batteries are shown in Fig. 2.

3.2. Comparison and analysis of prediction performance

In this paper, batteries Nos. 5–7 are used to illustrate the effectiveness of the improved PSO–SVR model. The capacity degeneration data from cycle 1 to cycle 100 are selected as the training samples to predict the future SOH values, while the data from cycle 101 to cycle 168 are selected for evaluating the prediction performance of the improved PSO–SVR. The training data number and prediction data number are the same as Ref. [14] and Ref. [16], which is helpful for performance comparisons among different prediction methods. Table 1 shows the prediction errors of the improved PSO–SVR and eight published methods (five methods from Ref. [14] and three from Ref. [16]) on batteries Nos. 5–7. Two criteria, mean absolute percentage error (MAPE) and root mean square error (RMSE), are adopted to evaluate the prediction performance of different methods:

$$ \text{MAPE} = \frac{1}{N} \left( \frac{1}{N-1} \sum_{i=1}^{N} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \right) \times 100\% \quad (12) $$

$$ \text{RMSE} = \sqrt{\frac{1}{N} \left( \frac{1}{N-1} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \right)}. \quad (13) $$

From Table 1, we can see that the improved PSO–SVR method has much better prediction performance than the eight published methods. Among the eight published methods, SE–MGPR has the best prediction results. Both its predictions MAPE and RMSE are the smallest on batteries Nos. 5–7. Therefore, the smallest prediction MAPE and RMSE of the eight published methods on battery No. 5 are 1.38% and 1.20 respectively, while the prediction MAPE and RMSE of the improved PSO–SVR are 0.82% and 0.75 respectively. In the same way, we found that the prediction MAPE and RMSE of the improved PSO–SVR on battery No. 6 are also the smallest. The prediction RMSE of the improved PSO–SVR on battery No. 7 is smaller than that of SE–MGPR, although the prediction MAPE is equal. By comparison, it can be concluded that the improved PSO–SVR could effectively improve the prediction performance of battery SOH. That is mainly because the improved PSO–SVR focuses on

<table>
<thead>
<tr>
<th>Battery No.</th>
<th>Error criteria</th>
<th>Basic GPR$^a$</th>
<th>LGFPR$^a$</th>
<th>QGFPR$^a$</th>
<th>CLGFR$^a$</th>
<th>CQGFPR$^a$</th>
<th>SMK–GPR$^a$</th>
<th>P–MGPR$^a$</th>
<th>SE–MGPR$^a$</th>
<th>Improved PSO–SVR</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>MAPE (%)</td>
<td>12.10</td>
<td>23.0</td>
<td>1.90</td>
<td>1.60</td>
<td>2.10</td>
<td>1.65</td>
<td>1.55</td>
<td>1.38</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>13.03</td>
<td>1.71</td>
<td>1.50</td>
<td>1.36</td>
<td>1.80</td>
<td>1.38</td>
<td>1.30</td>
<td>1.20</td>
<td>0.75</td>
</tr>
<tr>
<td>6</td>
<td>MAPE (%)</td>
<td>27.0</td>
<td>10.30</td>
<td>7.70</td>
<td>10.20</td>
<td>29.0</td>
<td>10.60</td>
<td>2.96</td>
<td>2.93</td>
<td>2.28</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>22.51</td>
<td>6.90</td>
<td>5.12</td>
<td>6.86</td>
<td>20.44</td>
<td>7.08</td>
<td>2.12</td>
<td>2.11</td>
<td>1.66</td>
</tr>
<tr>
<td>7</td>
<td>MAPE (%)</td>
<td>19.20</td>
<td>1.90</td>
<td>5.40</td>
<td>1.70</td>
<td>2.60</td>
<td>1.91</td>
<td>1.09</td>
<td>1.02</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>20.70</td>
<td>5.59</td>
<td>5.52</td>
<td>1.73</td>
<td>2.59</td>
<td>1.88</td>
<td>1.14</td>
<td>1.07</td>
<td>0.57</td>
</tr>
</tbody>
</table>

$^a$ Results of these methods are obtained from reference [14].

$^b$ Results of these methods are obtained from reference [16].

Please cite this article as: T. Qin, et al., Robust prognostics for state of health estimation of lithium-ion batteries based on an improved PSO–SVR model, Microelectronics Reliability (2015), http://dx.doi.org/10.1016/j.microrel.2015.06.133
grasping the global trend of capacity degradation rather than capturing
the local regeneration and fluctuations.
Take battery No. 5 as an example; the SOH prediction results after
cycle 100 are shown in Fig. 3. It can be seen that the prediction values
of the improved PSO-SVR are close to the actual SOH values. Therefore,
after giving a fault threshold, the remaining useful life (RUL) can be
obtained. For different devices, the fault thresholds of batteries may be
different but generally vary from 70% to 80% of the initial SOH [6].
Taking 70% and 75% for example, as shown in Fig. 3, the prediction er-
ers of RUL are 3 cycles and 0 respectively. Moreover, the largest
prediction error is only 6 cycles. The excellent prediction performance
indicates that the improved PSO-SVR can be well applied to RUL
prediction under different fault thresholds.

3.3. The effect of capacity regeneration
Still for battery No. 5, as seen in Fig. 2, the most noticeable regener-
ation is after the 89th cycle, which makes the capacities from the 90th
cycle no lower than the 89th cycle. Therefore, three sets of training data are selected to analyze the influence of regenera-
tion. The numbers of training data are 89 (from cycle 1 to cycle 89),
95 and 100 respectively. Fig. 4 shows the SOH prediction results under
the above three situations, and the results around cycle 160 are enlarged
and given in the upper right of Fig. 4. It can be seen from the three
prediction curves in the enlarged figure that the SOH prediction results
have small differences (only one or two cycles). In addition, the predic-
tion MAPE are 0.89%, 0.88% and 0.82% respectively and the prediction
RMSE are 0.82, 0.77 and 0.75 respectively. The close errors under
three situations demonstrate that regeneration has small influence on
the prediction accuracy of the improved PSO-SVR.

3.4. Robustness analysis
The training data usually contain some unexpected errors, and their
values are generally random [7]. In order to analyze the anti-noise per-
formance of the improved PSO-SVR, a Gaussian noise is added to the
raw capacity samples of battery No. 5. Fig. 5 shows the training data
with noise and SOH prediction results. Through calculating the prediction
errors from cycle 101 to cycle 168, the prediction MAPE and RMS
on battery No. 5 are 0.97% and 0.81 respectively. They are close to
0.82% and 0.75, which are the prediction errors with raw data and can
be read from the last row of Table 1. This means that the improved
PSO-SVR can achieve relatively accurate SOH prediction results under
some noise conditions. Battery No. 45 is selected to analyze the robustness of the improved PSO-SVR because its raw capacity data contain outliers. As shown in
Fig. 6, the two outliers are at cycle 20 and cycle 66. SOH prediction
results using 30 raw SOH data (from cycle 1 to cycle 30) as the training samples are shown in Fig. 6(A). By contrast, the prediction results after eliminating the outlier at cycle 20 are shown in Fig. 6(B). The outlier is replaced by the mean value of its two adjacent samples, which is the most common method of eliminating outliers [6]. From Fig. 6(A) and (B), it is clear that the two sets of prediction results after cycle 30 are almost the same. In other words, the improved PSO–SVR can get high accuracy prediction results without eliminating the outliers manually.

Moreover, the effect of the amount of outliers is thoroughly investigated. Still take the first 30 SOH data of battery No. 45 as the training samples, but the difference is that more outliers are added into the raw SOH data. The corresponding cycle numbers of the randomly produced outliers are shown in Table 2 and the values of SOH at these cycles are replaced by zero. The outliers are added into the SOH time series one by one to analyze the robustness of the improved model. With the increase in the amount of the outliers, the changes of the prediction MAPE and RMSE are shown in Fig. 7. The prediction errors remain stable when there are no more than six outliers (20% of the training samples). After that, the prediction errors gradually increase to 3.2% (MAPE) and 2.2 (RMSE) when the amount of the outliers increases to fourteen. Then, the prediction MAPE jumps to 48% and the prediction RMSE soars to 29 when there are fifteen outliers (50% of the training samples). After that, the prediction errors reach to exceed 20%. Therefore, the prediction errors of the improved model are stable and small in most cases. It indicated that the improved model has excellent robustness and can be well applied to SOH prognostics for the training samples contain some outliers.

4. Conclusion

To obtain more accurate long-term prediction results of SOH with limited raw capacity data, we have proposed an improved PSO–SVR model. This model was aimed at grasping the global degradation trend without focusing on local fluctuations. The criterion composed of MSE and R² was used to evaluate the performance from two different angles in the parameter optimization process. A fitness function for PSO and a new validation method were proposed to make the model have the ability to grasp the global degradation trend and realize long-term backward extrapolation.

Case studies were carried out to evaluate the performance of the proposed model. The results showed that: 1) using the previous capacity data of the current battery, this model can achieve more accurate SOH prognostics; 2) this model can get satisfactory results of RUL estimation under different fault thresholds; 3) even starting from the cycle near capacity regeneration, this model can also have good prediction performance; and 4) this model has great robustness and can get accurate prediction results when the data contain noise and outliers.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (NSFC: 61304218) and Beijing Higher Education Young Elite Teacher Project (YETP1123).

References


Table 2

<table>
<thead>
<tr>
<th>Outlier</th>
<th>Cycle number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>17</td>
</tr>
<tr>
<td>4</td>
<td>21</td>
</tr>
<tr>
<td>5</td>
<td>24</td>
</tr>
<tr>
<td>6</td>
<td>13</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>30</td>
</tr>
<tr>
<td>9</td>
<td>22</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>Cycle number</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>29</td>
</tr>
<tr>
<td>Cycle number</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>30</td>
</tr>
<tr>
<td>Cycle number</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>16</td>
</tr>
</tbody>
</table>

Fig. 7. Errors with different amounts of outliers.