Li-ion battery state of health Prediction based on Long Short-Term Memory Recurrent Neural Network

Jingdong Lin¹, Guansong Yan²* and Chang Wang²

¹College of Automation, Chongqing University, Chongqing, 400044, China
²Control Science and Engineering, Chongqing University, Chongqing, 400044, China
*Corresponding author’s e-mail: 202013021002@cqu.edu.cn

Abstract. The state of health (SOH) prediction of lithium-ion battery is essential for the health management of batteries. At present, the prediction method combined with the extraction of health indicators in charge-discharge process has received extensive attention, however, many studies ignored that the extraction of battery discharge data will be affected by the actual operating conditions, which will affect the effectiveness of health indicators extraction. In this work, a type of recurrent neural network (RNN), which is long short-term memory-RNN (LSTM-RNN), is proposed to prediction the SOH of Li-ion batteries through the data of charging process and capacity. Because the different choice of network parameters will also affect the performance of the model, particle swarm optimization (PSO) algorithm is used to optimize LSTM model. The test results show that this method can effectively predict SOH of battery, and the maximum RMSE is less than 0.01. Compared with the traditional LSTM algorithm, it has higher accuracy.

1. Introduction
As the core component of new energy vehicles, the performance of lithium-ion battery will decline with the increase of charge-discharge cycles in actual use, and lithium-ion battery will become unreliable, which may lead to system failure and even serious safety accidents [1].

Various indicators have been used to evaluate the health state of lithium batteries e.g., the state-of-charge (SOC), state-of-health (SOH), and remaining useful life (RUL) [2]. SOH and RUL are frequently used to reflect the degree of degradation. Capacity is often used to define battery SOH. When the SOH of an electrified vehicle is less than 80%, it needs to be replaced with a new battery. And a prediction method of RUL is proposed by iteratively predicting multi-step SOH. Therefore, accurate prediction of SOH is very important for battery health management.

SOH prediction methods are generally divided into two categories: model-based methods and data-driven methods. The model-based methods use many mathematical models, which needs to know the internal aging mechanism of the battery. By analyzing the physical and chemical processes inside the battery, the dynamic parameters of the battery are described, and the operation mechanism of the battery was obtained, and the degradation model of the battery was established. However, due to the complexity of the battery, it is a highly nonlinear dynamic system, so it is difficult to accurately describe the characteristics of the battery by physicochemical or thermodynamic models [3], and this method is not suitable for SOH prediction in practice.

In recent years, the data-driven based methods have gradually become a research hotspot. Health indicators (HI) are extracted from the data generated by battery charging and discharging process, and a mapping relationship are established between health indicators and SOH of batteries. In this way, the
battery SOH can be predicted. Compared with the establishment of a complex electrochemical reaction model of a battery, this data-driven method of constructing the mapping relationship between HI and SOH is flexible and easy to operate. In past studies, support vector machines (SVM), relevance vector machines (RVM), Kalman filtering (KF), regression analysis and neural networks have been used to predict the battery SOH. Liu et al. [4] designed a method of SOH estimation by extracting the characteristics of the dynamic change of the battery charging and discharging process and modified the SVR kernel function. The results show that the method can effectively estimate the battery health status. Xu et al. [5] proposed a lithium battery SOH estimation method based on improved particle filtering, which can effectively adapt to highly non-linear lithium batteries, but the extracted health indicators are not necessarily optimal.

It has been confirmed that LSTM-NNs are an effective model for processing time-sequential data [7]. According to the increase in the number of cycles, the data generated by battery charging and discharging can be regarded as the quantity related to the time-sequence. Therefore, the SOH prediction problem of the battery can be defined as a time-sequential problem. In a large number of literatures, LSTM has been applied to SOH estimation and RUL prediction of lithium batteries. Ning et al. [8] extracted capacity-related health factors based on the discharge curve of lithium batteries, and the results showed that the predicted maximum MAPE does not exceed 2%.

The above battery SOH prediction methods did not consider that the discharge cycles will be affected by the actual condition when extracting the HI, resulting in the failure to effectively obtain the discharge data. Therefore, this article does not consider the data obtained from the discharge cycles, only extracts the HI that may be related to the battery SOH based on the charging process, and then selects the three most relevant HI in combination with the Grey Relational Analysis (GRA) method. The long short-term memory network is achieved to predict the SOH of batteries. Since the settings of different model parameters have a greater impact on the training results, PSO is used to find the optimal LSTM model parameters.

2. Preliminaries and Definitions

2.1. Data set description

NASA battery repository is used in this article for experimental verification. Select the data of the four groups of batteries B0005, B0006, B0007 and B0018 under the BatteryAgingARC-FY08Q4 group in the data set. The test environment is shown in Table 1.

| Ambient temperature | Charging mode (CC/ CV) | Discharge mode |
|---------------------|------------------------|----------------|
| B0005               | Room temperature 24    | 2 A constant current | B0006 |
| B0007               | degrees Celsius        | -constant voltage (CV) |       |
| B0018               | charging mode         | discharge         |       |

As shown in Figure 1, after each discharge cycle of the battery, the capacity is calculated by integrating the discharge current over time. With the continuous increase of cycles, the capacity is constantly decreasing due to the nature of the internal structure of the battery. After multiple cycles, its capacity degenerates to 70% of the initial capacity, and the experiment is stopped. The battery SOH based on capacity is defined as

\[ SOH = \frac{c_m}{c_o} \times 100\% \]  

(1)

where \( c_m \) is the maximum available capacity, and \( c_o \) is the rated capacity of the battery.
2.2. Health index extraction

According to the definition of formula (1) in 2.1, SOH is the ratio of the current cycle battery capacity to the rated capacity. This article makes a single-step prediction of SOH based on the historical data of the charge cycles, that is, uses the data of the current cycle and the historical cycles to predict the SOH of the next cycle. The capacity is not extracted during the discharge process, but calculated after the discharge cycle, and the capacity is closely related to SOH. Therefore, capacity is used as an HI for predicting SOH.

As shown in Figure 2, with the increase of charging times, the constant current charging time decreases and the constant voltage charging time increases, which is consistent with the capacity degradation trend. Therefore, the constant current charging time and constant voltage charging time are also regarded as HI reflecting the degradation of battery capacity. By analysing other data of charging process, the average temperature and average voltage are also extracted as HI.

2.3. Correlation analysis

Figure 3 shows the extracted HI change with the number of cycles. It can be seen that the HI extracted in this article are correlated with battery SOH, and the grey relational analysis method is used to quantitatively evaluate the correlation between HI and SOH. The specific calculation results are shown in Table 2.
Table 2 Correlation analysis of HI

| HI                        | SOH   | Constant current charging time | Constant voltage charging time | Average temperature in constant current phase | Average voltage in constant current phase |
|---------------------------|-------|---------------------------------|---------------------------------|-----------------------------------------------|-------------------------------------------|
| Capacity                  | 1.0000| 0.8058                          | 0.6003                          | 0.7252                                        | 0.7766                                     |

According to the GRA results, among the HI extracted based on the charging phase, the battery capacity, constant current charging time and the average voltage of the constant current phase have the highest correlation coefficients with SOH. Therefore, in this article, these three parameters are selected as the ultimately described HI of the battery degradation state.

3. LSTM-based SOH prediction model

3.1. LSTM algorithm structure

Long short-term memory network (LSTM) is an improved recurrent neural network model, which not only has a good learning ability for long-term dependent sequence data that cannot be processed in RNN, but also can overcome the gradient explosion and gradient disappearance in traditional RNN problem. The HI extracted during the cyclic charging and battery SOH are both related to the time sequences. Therefore, the LSTM is adopted for accurate lithium ion battery SOH prediction. Figure 4 shows the LSTM network structure.

Figure 4 LSTM structure

Compared with traditional RNNs, LSTM adds cell state, which can remember long-term dependencies. The forget gate, input gate, and output gate are used by LSTM to control information, and these three gates are converted into values between 0 and 1 by the sigmoid activation function. The calculation formula is shown with

\[
f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)
\]

(2)

\[
i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)
\]

(3)

\[
o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)
\]

(4)

where \(f_t\), \(i_t\) and \(o_t\) are the weight matrices of the forget gate, input gate and output gate. The \(b_f\), \(b_i\) and \(b_o\) are three bias vectors. The \(\sigma\) is the gate activation function, which normally is a sigmoid function, and the \(\tanh\) is the hyperbolic tangent function.

The forget gate is mainly used to selectively forget the input from the previous node and decide which information should be discarded or retained. The input gate is used to selectively remember the current input, and the output gate is used to output the value of the current memory unit. Through the output of the three gate states, the hidden state and the cell state can be updated to achieve long-term memory and short-term memory functions. The calculation formula is shown with

\[
\hat{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c)
\]

(5)

\[
c_t = f_t \cdot c_{t-1} + i_t \cdot \hat{c}_t
\]

(6)
where $c_i$ represents the state of the cell, which can maintain long-term memory of the input. $h_t$ is the hidden state, which represents the short-term memory of the current input.

3.2. **PSO algorithm framework**

The particle swarm optimization algorithm (PSO) is a random search algorithm based on group cooperation developed by simulating the foraging behavior of a flock of birds. This algorithm is used to find the optimal solution to a specific problem. The implementation process of the PSO algorithm is shown in Figure 5.

![Figure 5 PSO algorithm flow chart](image)

PSO algorithm programming is simple and easy to execute, and can converge to the optimal solution quickly, which is suitable for optimizing parameters. Therefore, the particle swarm algorithm is used to find the optimal LSTM model parameters.

3.3. **An PSO-LSTM-based SOH prediction model**

In this paper, the battery prediction model process based on PSO-LSTM includes data preprocessing, PSO-LSTM model offline training and SOH online prediction.

1) Construct a data set. Normalize the HI proposed in 2.2 from the battery charge and discharge cycle data set. Extract the capacity data and calculate the battery SOH according to (1). The data set is structured as

$$X = \{X_1, X_2, X_3, \cdots, X_m\}$$

$$X_i = \{x_1^i, x_2^i, x_3^i\}$$

$$Y = \{Y_1, Y_2, Y_3, \cdots, Y_m\}$$

where $X_i$ is the vector of the extracted HI, $x_1^i$, $x_2^i$, $x_3^i$ are the capacity, constant current charging time and the average voltage of the constant current phase extracted in the i-th cycle respectively. $Y_i$ is the SOH of the i-th cycle. $m$ is the number of cycles.

2) Model offline training. The time step, the number of hidden layer nodes and the number of training times are determined as the optimization parameters of LSTM algorithm. The parameters optimized by PSO algorithm are input into LSTM model, and other parameters are initialized. The data set obtained in 1) is divided into training set and prediction set. LSTM model is trained by training set.
3) SOH online prediction. After the model is obtained, the charging data of the current cycle is extracted and combined with the past data to construct the input, and input it into the trained model to obtain the battery SOH prediction result of the next cycle.

4. Test and Results
In this work, the B0005, B0006, B0007 and B0018 battery packs are used for model training and verification. In order to further illustrate the predictive performance of the algorithm, the method is compared with the traditional LSTM algorithm, and the absolute percentage error (MAPE) and root mean square error (RMSE) are used to quantitatively evaluate the prediction results of SOH. The evaluation indicators are as follows:

\[
MAPE = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{y_{pred}(i) - y(i)}{y(i)} \right| \times 100\%
\]

\[
RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left[ y_{pred}(i) - y(i) \right]^2}
\]

where \( y_{pred} \) is the predicted value of the SOH, \( y \) is the true value of SOH and \( N \) is the total number of sample predictions.

Battery B0005 contains 168 cycles in total. This article uses only 30% of the data as the training set, that is, 50 cycles as the training set, and the remaining 118 cycles as the test set. The particle swarm optimization results are shown in Table 3. According to this result, the time step of LSTM training is 15, that is, every 15 consecutive samples are used to predict the value of the next sample. Thus, the total number of samples in this data set is 153, the training set has a total of 35, and the test set is 118. The input of this model is that the number of layer nodes is set to 3, the number of output layer nodes is 1, the initial learning rate is 0.001, and Adam is selected as the algorithm optimizer.

| Time step | The number of hidden layer nodes | Training times |
|-----------|----------------------------------|----------------|
| 15        | 203                              | 2162           |

Figure 6 shows the prediction results of the B0005 battery. The figure is the result of de-normalizing the output, in order to better see the prediction effect. The RMSE of SOH predicted by this method is within 0.01, which can achieve high accuracy. It can be seen from Figure 6(b) that the prediction errors of individual points are relatively large. This is because the capacity of the battery will increase after a period of time after the battery is discharged. The current HI extracted in this paper cannot fully reflect this phenomenon. The remaining three sets of batteries are also trained by this method, and the final results are shown in Table 4.

![Figure 6](image_url)

Figure 6 Prediction results and corresponding errors of B0005 battery

(a) Prediction result
(b) Corresponding relative error
Table 4  B0005, B0006, B0018 batteries training result

|        | B0006 | B0007 | B0018 |
|--------|-------|-------|-------|
| RMSE   | 0.0095| 0.0092| 0.0088|
| MAPE   | 0.0007| 0.0006| 0.0005|

5. Conclusion and Future Works
This paper implements a lithium-ion battery health state prediction model based on charging data, which can effectively perform a single-step prediction of lithium-ion battery SOH. According to the prediction results, the battery health status can be learned in time, so as to realize related battery health management. This paper uses lithium battery constant current charging data, capacity and average voltage during constant current phase as HI, and based on a data-driven method, using the PSO-LSTM model for learning and training. The simulation experiment results show that the extracted health factors can effectively reflect the battery degradation state. And this method can effectively improve the accuracy of SOH prediction, which the predicted maximum RMSE is less than 0.01. However, the research still has some shortcomings. This article is based on the complete battery cycle charging data to predict the battery health status, but the actual use process may cause incomplete charging data collection for some reasons, which will affect the extraction of health factors. Therefore, how to make predictions based on incomplete battery charging data requires further research.

References
[1] Zhang, Y.Z., Xiong, R., He, H.W., Pecht, M.G., (2018) Long Short-Term Memory Recurrent Neural Network for Remaining Useful Life Prediction of Lithium-Ion Batteries. IEEE Transactions on Vehicular Technology, 67(7): 5695-5705.
[2] Li, P.H., Zhang, Z.J., Xiong, Q.Y., Ding, B.C., Hou, J., Luo, D.C., Rong, Y.J., Li, S.Y., (2020) State-of-health estimation and remaining useful life prediction for the lithium-ion battery based on a variant long short term memory neural network. Journal of Power Sources, 459.
[3] Jiang, Y.H., Yu, Y.F., Huang, J.Q., Cai, W.W., Marco, J., (2021) Li-ion battery temperature estimation based on recurrent neural networks. Science China Technological Sciences, 64.
[4] Zheng, Y.L., Liu, J.J., Zhao, H.W., Wang, C.Y., (2020) A New Lithium-Ion Battery SOH Estimation Method Based on an Indirect Enhanced Health Indicator and Support Vector Regression in PHMs, Energies, 13(4).
[5] Xu, C., Li, L.W., Yang, Y.X., Wang, K., (2020) Lithium-ion battery SOH estimation based on improved particle filter. Energy Storage Science and Technology, 9(06): 1954-1960.
[6] Yayan, U., Arslan A.T., Yucel, H., (2021) A Novel Method for SoH Prediction of Batteries Based on Stacked LSTM with Quick Charge Data. Applied Artificial Intelligence, 35(6).
[7] Greff, K., Srovastava, R.K., Koutnik, J., Bas, S.R., Schmidhuber, J., (2017) LSTM: A Search Space Odyssey. IEEE transactions on neural networks and learning systems, 28(10).
[8] Ning, Q.J., Shi, M.Z., Shi, Y.S., Ding, E.S., Hong, Y.T., Ou, Y., (2021) Lithium-ion battery life prediction method based on optimal charge and discharge curve. Journal of Shanxi University of Science and Technology, 39(02): 153-160.
[9] Saha, B., Goebel, K., (2007) Battery data set. NASA AMES prognostics data repository.
[10] Fotouhi, Abbas, A., Daniel, J.P., Karsten, L., Stefano, W., (2016) A review on electric vehicle battery modelling: From Lithium-ion toward Lithium-Sulphur. Renewable and sustainable energy reviews, 56: 1008-1021.