

An Indirect Prediction Method based on Feature Extraction of Lithium-ion Batteries Remaining Useful Life

Zhiqiang Deng *, Dewen Wang

School of Control and Computer Engineering, North China Electric Power University, Baoding 071003, China

* Corresponding author: Zhiqiang Deng (Email: dengzhiqiang2019@126.com)

Abstract: Predicting the remaining useful life (RUL) of lithium-ion batteries is of paramount significance for the lifespan management of electric vehicles. This study delves into a feature extraction-based approach for indirectly predicting RUL in lithium-ion batteries, circumventing the limitations posed by the challenge of measuring capacity during battery operation. Initially, the health feature (HF) exhibiting strong correlations with capacity is extracted from localized charging voltage intervals, and capacity estimates are obtained through polynomial fitting. Subsequently, a dynamic long short-term memory (D-LSTM) neural network is established, iteratively incorporating prediction outcomes into training data to continuously enhance the model and ultimately accomplish RUL prediction. Experimental results conducted on batteries with varying initial states from the NASA dataset highlight the substantial accuracy and robustness of the proposed D-LSTM method. It outperforms other commonly utilized models, offering the capability for online RUL prediction.

Keywords: Lithium-ion Battery; Remaining Useful Life; Feature Extraction; Long Short-term Memory.

1. Introduction

Due to its characteristics of prolonged cycle life, low self-discharge rates, and exceptional safety performance, the lithium-ion battery has become the primary power source for electric vehicles [1]. However, failures occurring during the degradation of lithium-ion batteries can directly impact the normal operation of devices and even lead to serious safety incidents [2]. In order to address this issue, the prediction of the remaining useful life (RUL) of lithium-ion batteries has become an important task, as it can effectively estimate the degree of degradation of the battery at future moments, thereby providing a crucial guarantee for the safe and reliable operation of energy storage systems.

The RUL of a lithium-ion battery refers to the time or number of cycles over which the battery is expected to continue functioning effectively under specific usage conditions. The critical threshold for battery failure is referred to as end-of-life (EOL), typically defined by 80% of the battery's state of health (SOH), represented using the capacity definition [3]:

$$SOH = Q_n / Q_N \quad (1)$$

In recent years, methods for predicting the RUL of lithium-ion batteries have been broadly categorized into two main groups: model-based approaches and data-driven approaches.

Model-based approaches primarily encompass two types: the equivalent circuit model [4–6] and the electrochemical model [7–9]. The equivalent circuit model is a representation that does not take into account the internal chemical composition of the battery or the corresponding reactions. This model can be constructed based on the battery's electrical characteristics using fundamental electronic components and controllable voltage sources. The equivalent circuit model demonstrates high computational efficiency, rendering it suitable for online battery state prediction. However, achieving a high level of accuracy with this model still

presents certain challenges. The electrochemical model, originating from a spectrum of aging mechanisms, quantitatively delineates the intricate microphysical and chemical processes transpiring within the battery, employing fundamental principles. Notably, this encompasses the P2D model [7] and its simplified derivatives [9]. However, a delicate equilibrium must be ascertained between the domain of applicability and the precision of these models.

The data-driven approach can be perceived as a black-box methodology capable of directly unearthing latent information and evolutionary patterns concerning battery performance alterations from historical battery data. This approach facilitates the prediction of RUL without necessitating an intricate comprehension of precise degradation mechanisms. Noteworthy techniques encompass artificial neural networks (ANN) [10, 11], convolution neural networks (CNN) [12, 13], and long-short-term memory (LSTM) networks [14, 15].

In the realm of data-driven methodologies, aside from the judicious selection of adept models, efficacy is intrinsically tied to the aptness of the chosen health features in encapsulating the battery's degradation trajectory. Given the practical challenges associated with directly measuring attributes like capacity and impedance in real-world scenarios, the extraction of alternative perspectives for health features assumes heightened practical significance.

Hence, in response to the aforementioned research gaps, this paper introduces a feature extraction-based approach for the indirect prediction of RUL. To substantiate the model's effectiveness, experiments were conducted on batteries with various distinct initial states. The results attest to the proposed method's superior precision compared to alternative approaches. The primary contributions of this paper are as follows:

1. This paper introduces a method that extracts the health feature (HF) from specific charging stages, imparting practical significance. Furthermore, this approach

establishes a robust correlation between HI and battery capacity through Pearson correlation coefficients.

2. This paper addresses the challenge of measuring battery capacity directly in real-world scenarios by employing polynomial fitting to establish a mapping between the extracted HF and battery capacity.
3. This paper adopts an incremental learning perspective, progressively predicting forthcoming cycle capacities. By utilizing a dynamic LSTM model that iteratively refines its predictions, the accuracy of RUL estimation is substantially improved.

The remainder of this article is organized as follows: **Section 2** predominantly engages in methodological discourse, encompassing the extraction of HF and delineating the structure and establishment process of D-LSTM. In **Section 3**, the robustness and efficacy of D-LSTM are substantiated through battery validation employing diverse initial states alongside comparative analyses against alternative methodologies. Finally, **Section 4** concludes the paper.

2. Methodology

2.1. Health Feature Extraction

Figure 1 illustrates the charge and discharge voltage profiles of the B5 battery, where darker colors represent more pronounced degradation. It is discernible that, with the progression of degradation, alterations in the curve shapes become apparent, manifesting as shifts in time intervals within equivalent voltage spans. Consequently, it becomes plausible to extract time intervals from the charge and discharge curves, as HF is closely linked to capacity. The computation formula is presented below:

$$HF = T(V_{start}) - T(V_{end}) \quad (2)$$

where V_{start} and V_{end} is the start voltage and end voltage of the voltage interval.

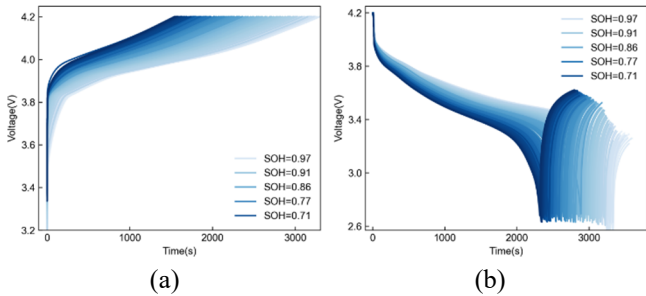


Figure 1. Charge and discharge voltage curves of the B5 battery

However, in the practical utilization of lithium-ion batteries, the discharge process is often non-continuous, and operational conditions exhibit considerable variability. Consequently, acquiring data resembling the stable curves depicted in **Figure 1(a)** presents certain challenges. Moreover, due to the infrequent occurrence of extreme charge and discharge cycles, obtaining complete cycle curves is unrealistic. Thus, to enhance the feasibility of extracting HF from real-world applications, this paper opts to solely consider information from select charging stages.

By employing an iterative approach, the optimal voltage interval of [3.7, 4.1] V was determined. To quantitatively analyze the relationship between the extracted HF and capacity, this paper employs the Pearson coefficient. The

Pearson coefficient serves as a metric for assessing linear correlation and yields values between -1 and 1. Here is the presented calculation formula:

$$Pearson = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (3)$$

where x_i and y_i represent the HF and SOH for the i -th cycle, while \bar{x} and \bar{y} represent the average values of the samples.

The computed results of the Pearson coefficient have been summarized in **Table 1**. Observing the table, it is evident that the coefficient values between all batteries and features surpass 0.9. This observation underscores the presence of an exceedingly strong linear correlation between the extracted HF and capacity across all batteries. Such a relationship indirectly reflects the extent of battery degradation.

Table 1. Correlation coefficient between the capacity and HF.

Battery	Correlation coefficient
B5	0.9974
B6	0.9917
B7	0.9906

Therefore, the approach of polynomial fitting can be employed to estimate capacity values through the mapping of HF. Specifically, a linear function is utilized, as depicted below:

$$C_{est} = a \cdot HF + b \quad (4)$$

where C_{est} is the estimated capacity, while a and b are the constant obtained from the linear regression.

The outcomes of the linear fitting have been presented in **Figure 2**. It is discernible that estimating capacity indirectly through the extraction of HF proves to be a viable alternative to direct capacity measurements.

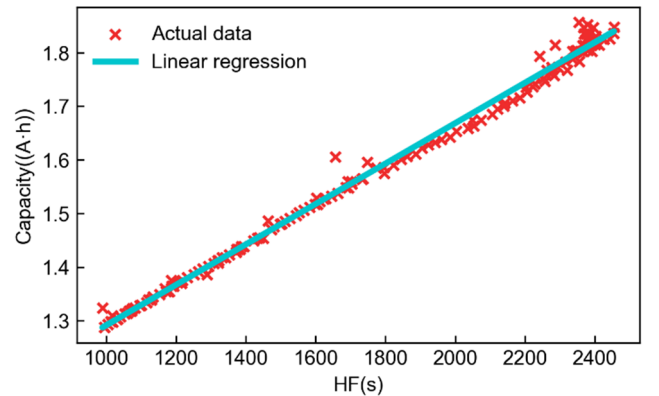


Figure 2. Result of the linear fit

2.2. D-LSTM Model

The D-LSTM model proposed in this paper is built upon the foundation of LSTM, which is an advancement over recurrent neural networks (RNNs). LSTM was introduced by Hochreiter and Schmid Huber with the goal of addressing challenges like vanishing and exploding gradients that arise when training long sequences. LSTM enhances RNNs by incorporating three information control gates: the forget gate, input gate, and output gate. These gate structures empower LSTM to adeptly capture long-term dependencies within

sequences, thereby exhibiting superior performance in processing temporal data. This capability arises from its effective maintenance and updating of memory states. The structure of LSTM is illustrated in **Figure 3**.

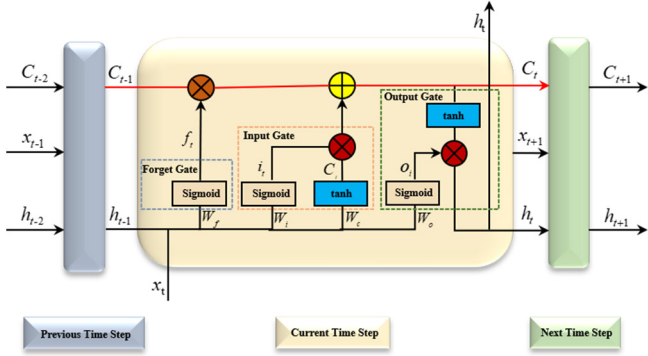


Figure 3. The basic structure of LSTM

The relevant calculation formula is provided below:

$$\begin{cases} f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \\ i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\ \tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \\ C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \\ o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \\ h_t = o_t * \tanh(C_t) \end{cases} \quad (5)$$

where σ is the sigmoid activation function; W_f , W_i and W_o represent the weight matrices for the forget gate, input gate, and output gate respectively; b_f , b_i and b_o are the corresponding bias vectors; \tanh is the hyperbolic tangent activation function.

The workflow for RUL prediction using the D-LSTM model is outlined below:

1. Extract the HF from historical data and employ this data to train the initial D-LSTM model.
2. Establish a linear mapping relationship from health index to capacity and determine the coefficient values of the linear function.
3. Utilize the D-LSTM model to generate health index predictions for the next n cycles.
4. Employ the linear function established in Step 1 to compute capacity predictions for the subsequent n cycles.
5. If the capacity prediction falls below a specified EOL, which corresponds to 80% SOH, output the predicted RUL. If not, incorporate these n -cycle capacity predictions into the training data, fine-tune the D-LSTM model, and then repeat Steps 3-4.

3. Experimental

3.1. Experimental Setup

Before training the D-LSTM model, the HF undergoes standardization, computed as follows:

$$x_{\text{norm}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (6)$$

where x_{norm} represents the standardized data, x , x_{\max}

and x_{\min} represent the original data, the maximum and minimum of the original data, respectively.

During the training of the D-LSTM model, the batch size is set to 20, the number of epochs is set to 200, and the learning rate is set to 0.001. A dropout layer is introduced with a parameter of 0.3 to prevent overfitting. The training process is accelerated using the adaptive moment estimation (Adam) optimization algorithm to aid convergence.

Furthermore, three metrics are adopted to assess the precision and robustness of the proposed method in predicting RUL for lithium-ion batteries. These metrics include RUL predicted error (RUL_{er}), root mean squared error (RMSE), and mean absolute error (MAE).

$$RUL_{\text{er}} = RUL_{\text{predict}} - RUL_{\text{actual}} \quad (7)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \hat{x}_i)^2} \quad (8)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |x_i - \hat{x}_i| \quad (9)$$

where x_i and \hat{x}_i represent the predicted and actual values of capacity for the i -th cycle, respectively.

3.2. Dataset Description

The proposed methodology is validated using the B5, B6, and B7 batteries from the NASA dataset [16]. The operating conditions of the batteries are detailed in **Table 2**.

Table 2. Operating conditions of NASA dataset

Battery	Nominal Capacity (Ah)	Cut-off voltage(V)	Charge current(A)	Discharge current(A)
B5	2.0	2.7	1.5	2
B6	2.0	2.5	1.5	2
B7	2.0	2.2	1.5	2

Figure 3 illustrates the capacity degradation curves for each battery. It's evident that the initial states of the batteries are inconsistent. Additionally, due to factors like capacity regeneration, noticeable fluctuations are present in certain segments of the curves, posing challenges to RUL prediction.

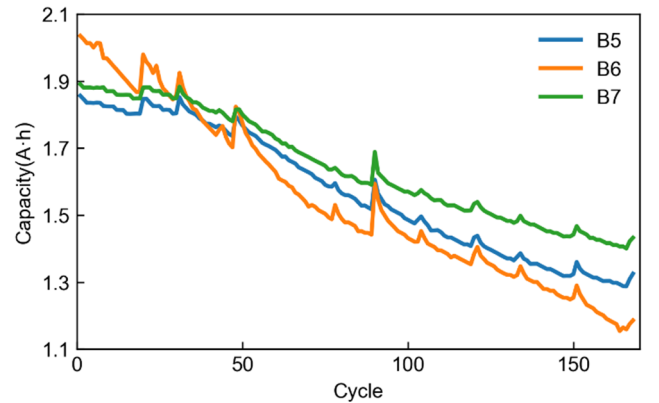


Figure 4. Capacity degradation curves of NASA dataset

3.3. Experimental Results

In this section, the initial 50% of the data is utilized as the training set, and the remaining 50% serves as the testing set. Therefore, the prediction starting point is set at the 84th data point in the dataset. Simultaneously, the EOL for the B5 and B6 batteries is set at 1.38 V, while the EOL for the B7 battery

is set at 1.50 V. To comprehensively validate the superiority of the proposed D-LSTM method, the study also employs other common techniques, including LSTM, CNN, and ANN, as controls. The resulting RUL prediction curves are depicted in Figure 5. It's clearly observable that across all tested batteries, the D-LSTM method outperforms other techniques in terms of accurate RUL prediction, closely aligning with the actual battery degradation curves.

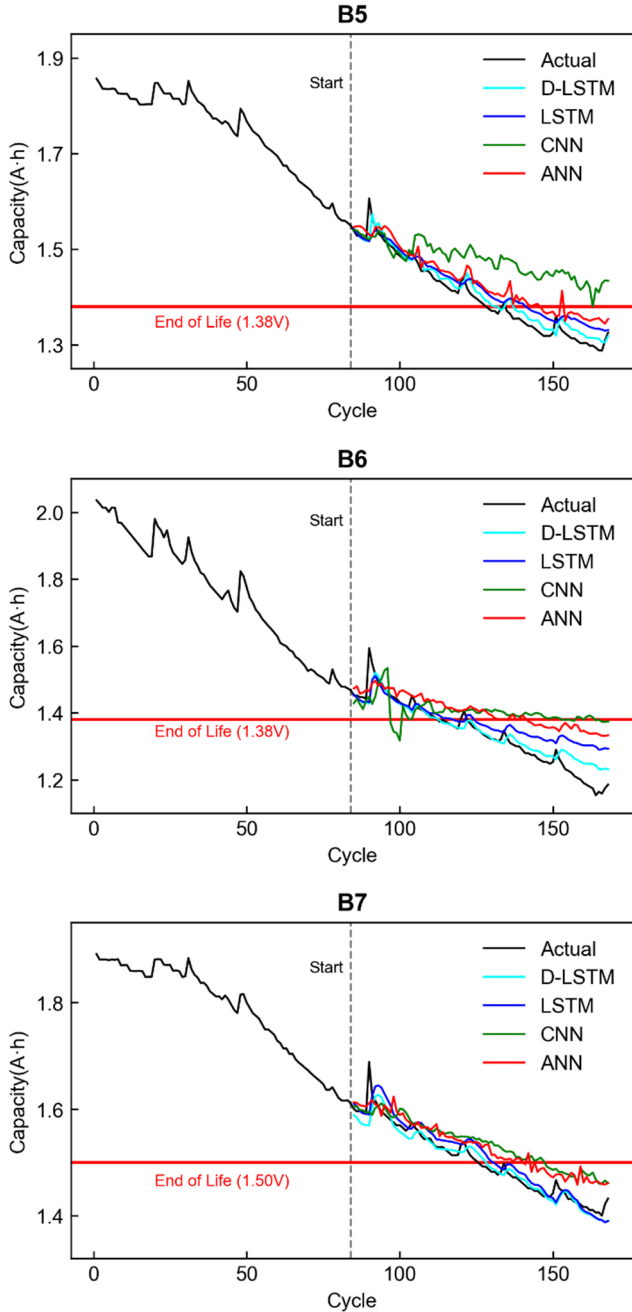


Figure 5. RUL prediction curves for different methods

Table 3 presents the calculated values for MAE, RMSE, and RUL errors, showcasing that the D-LSTM method exhibits the smallest errors across all batteries. Notably, when predicting the EOF for the B6 battery, the D-LSTM method accurately forecasts its RUL with a RUL error of 0. Additionally, on the B7 battery, D-LSTM achieves a MAE of 0.0123 and an RMSE of 0.0191, both of which are the lowest values among all prediction outcomes. Furthermore, the performance of the LSTM method also surpasses that of CNN and ANN, highlighting LSTM's prowess in temporal

prediction. These results lead to the conclusion that the proposed D-LSTM method boasts exceptional dynamic predictive capabilities, achieving the utmost precision in RUL forecasting.

Table 3. Prediction performance of different methods

B5	D-LSTM	LSTM	CNN	ANN
MAE	0.0150	0.0247	0.0783	0.0346
RMSE	0.0187	0.0288	0.0895	0.0390
RUL _{er}	3	13	34	17
B6	D-LSTM	LSTM	CNN	ANN
MAE	0.0226	0.0435	0.0886	0.0727
RMSE	0.0354	0.0601	0.1081	0.0873
RUL _{er}	0	5	-16	22
B7	D-LSTM	LSTM	CNN	ANN
MAE	0.0123	0.0168	0.0361	0.0305
RMSE	0.0191	0.0218	0.0412	0.0348
RUL _{er}	3	5	22	15

4. Conclusion

This paper proposes an effective solution to the indirect prediction challenge of lithium-ion battery RUL, mitigating the limitations associated with impractical capacity measurement in real-world applications, a matter of significant importance. This approach leverages HF extracted from partial charging voltage profiles to effectively substitute for capacity, accurately depicting the battery's degradation trends. Simultaneously, the introduced D-LSTM model dynamically adapts its parameters, addressing the accuracy decline issue in later stages of LSTM predictions and thus ensuring forecast stability. The experimental results on the NASA dataset convincingly demonstrate the precision and robustness of this method, with RUL errors within three cycles distinctly outperforming those of the compared LSTM, CNN, and ANN models.

However, it's important to note that the utilized NASA dataset encompasses solely one type of battery (the NCA battery). Future research could incorporate a broader range of battery types to explore the method's applicability across various battery categories, thereby expanding its scope.

References

- [1] H. Tian, P. Qin, K. Li, et al, A review of the state of health for lithium-ion batteries: Research status and suggestions, Journal of Cleaner Production. 261 (2020) 120813.
- [2] Y. Song, D. Liu, Y. Hou, et al, Satellite lithium-ion battery remaining useful life estimation with an iterative updated RVM fused with the KF algorithm, Chinese Journal of Aeronautics. 31 (2018) 31–40.
- [3] M.-F. Ng, J. Zhao, Q. Yan, et al, Z.W. Seh, Predicting the state of charge and health of batteries using data-driven machine learning, Nat Mach Intell. 2 (2020) 161–170.
- [4] Y. Ma, Y. Chen, X. Zhou, et al, Remaining Useful Life Prediction of Lithium-Ion Battery Based on Gauss–Hermite Particle Filter, IEEE Trans. Contr. Syst. Technol. 27 (2019) 1788–1795.
- [5] J. Yang, W. Fang, J. Chen, et al, A lithium-ion battery remaining useful life prediction method based on unscented particle filter and optimal combination strategy, Journal of Energy Storage. 55 (2022) 105648.

- [6] G. Dong, Z. Chen, J. Wei, et al, Battery Health Prognosis Using Brownian Motion Modeling and Particle Filtering, *IEEE Trans. Ind. Electron.* 65 (2018) 8646–8655.
- [7] M. Doyle, T.F. Fuller and J. Newman, Modeling of Galvanostatic Charge and Discharge of the Lithium/ Polymer/ Insertion Cell, *J. Electrochem. Soc.* 140 (1993) 1526–1533.
- [8] V. Laue, F. Röder and U. Krewer, Practical identifiability of electrochemical P2D models for lithium-ion batteries, *J Appl Electrochem.* 51 (2021) 1253–1265.
- [9] H. Pang, L. Mou, L. Guo, et al, Parameter identification and systematic validation of an enhanced single-particle model with aging degradation physics for Li-ion batteries, *Electrochimica Acta.* 307 (2019) 474–487.
- [10] J. Wu, C. Zhang and Z. Chen, An online method for lithium-ion battery remaining useful life estimation using importance sampling and neural networks, *Applied Energy.* 173 (2016) 134–140.
- [11] W. Qin, H. Lv, C. Liu, et al, P. Jahanshahi, Remaining useful life prediction for lithium-ion batteries using particle filter and artificial neural network, *IMDS.* 120 (2019) 312–328.
- [12] L. Ren, J. Dong, X. Wang, et al, A Data-Driven Auto-CNN-LSTM Prediction Model for Lithium-Ion Battery Remaining Useful Life, *IEEE Trans. Ind. Inf.* 17 (2021) 3478–3487.
- [13] Y. Toughzaoui, S.B. Toosi, H. Chaoui, et al, State of health estimation and remaining useful life assessment of lithium-ion batteries: A comparative study, *Journal of Energy Storage.* 51 (2022) 104520.
- [14] Y. Zhang, R. Xiong, H. He, et al, Long Short-Term Memory Recurrent Neural Network for Remaining Useful Life Prediction of Lithium-Ion Batteries, *IEEE Trans. Veh. Technol.* 67 (2018) 5695–5705.
- [15] J. Qu, F. Liu, Y. Ma, et al, A Neural-Network-Based Method for RUL Prediction and SOH Monitoring of Lithium-Ion Battery, *IEEE Access.* 7 (2019) 87178–87191.
- [16] B. Saha and K. Goebel, Modeling Li-ion Battery Capacity Depletion in a Particle Filtering Framework, (2009).