[文章编号] 1003－4684(2025)04-0037-06

基于辅助健康因子的锂电池剩余寿命预测研究

金浩博[1]，王淑青[1]，袁晓辉[2]

(1 湖北工业大学太阳能高效利用及储能运行控制湖北省重点实验室，湖北 武汉 430068；
2 华中科技大学土木与水利工程学院，湖北 武汉 430074)

[摘 要] 为提高锂离子电池剩余寿命预测准确率，提出了一种基于辅助健康因子的锂电池剩余寿命预测方法，可使用较少的训练数据准确预测锂电池剩余寿命 。从电池容量增量曲线和概率密度曲线中提取描述电池老化的健康因子，用于电池寿命预测模型的辅助训练。对现有用于电池寿命预测的长短期记忆网络模型进行改进，并使用遗传算法对模型的超参数进行优化，从而进一步提高网络模型的预测精度。最后在 3 个公开的数据集中验证方法的有效性。

[关键词] 锂离子电池；剩余寿命预测；长短期记忆网络；健康因子

[中图分类号] TM912 [文献标识码] A

锂电池被广泛应用在储能系统中，例如新能源汽车、储能电站等[1]。但由于锂离子电池的性能会随着时间和使用而恶化，从而导致容量和功率衰减[2]。这一因素极大的压缩了储能系统技术的发展空间[3]。因此，针对锂电池的剩余使用寿命的研究至关重要。

当前，对于锂电池容量衰退预测已有较多文献研究[4]。研究方法主要分为基于模型和基于数据驱动的方法[5-6]。基于模型的方法是研究电池衰退的机理规律，通过建立数学模型对锂电池剩余寿命进行预测。文献[7]提出了一种用于循环负载条件下老化的电化学电池模型。该模型将电池性能与电解质部分摩尔浓度联系起来，实现高精度的计算。文献[8]提出了一种通过识别老化阶段的电池寿命预测方法。基于数据驱动的方法一般是利用神经网络等智能算法直接基于采集到的数据对电池剩余寿命进行预测。文献[9]提出了一种基于改进的麻雀搜索算法优化的长短期记忆网络(long short-term memory，LSTM)的锂离子电池剩余寿命预测方法。文献[10]开发了一个知识约束的机器学习框架，以学习锂离子电池在工作周期内电池性能的随机退化，用于锂离子电池的健康预测。文献[11]从部分放电电压和温度中提取了两个健康指标，即等时间间隔的放电电压差和等时间间隔放电温度差，用于电池寿命预测。文献[12]提出了电池组健康状态和剩余寿命联合预测的预测框架。利用皮尔逊相关系数挖掘充电阶段间接健康指标与容量增量的关联度，并建立长短期记忆神经网络，联合估计电池健康状态和剩余寿命。文献[13]将差分热伏安信号分析方法和递归神经网络数据驱动方法相结合，以实现电池退化跟踪。锂电池在实际工况下参数变化影响因素多，是一个非线性的演化过程，建立电池模型需要高精度的参数辨识方法。而基于数据驱动的方法不需要复杂的机理建模过程，因此该方法在近年来成为了锂电池剩余寿命预测的热门研究方法之一[14]。

基于数据驱动的锂电池剩余寿命预测多使用电池容量作为健康因子训练模型。已有研究对于锂电池的剩余寿命预测将预测起点设置为 50%及以前时，预测效果并不理想。导致该结果的原因主要有两个方面。首先，电池容量数据的获取较为困难，在无法使用过多的循环数据用于模型训练时，会导致模型的预测性能不足。其次，剩余寿命预测模型为小样本模型，其网络结构大多为单层网络用于避免产生过拟合现象。这也导致其学习能力较为有限，无法使用前 50%的容量数据进行较为准确的电池剩余寿命预测。

为了解决上述两个方面问题，本文使用能定量描述电池老化的容量增量(incremental capacity，IC)曲线和概率密度(probability density，PD)曲线作为辅助健康因子，用于扩充可训练数据。其次，使用带 dropout 层的双层 LSTM 网络模型预测电池

[收稿日期] 2023－03－05
[基金项目] 国家重点研发计划项目(2021YFC3200405)
[第一作者] 金浩博(1994－)，男，湖北崇阳人，研究方向为电池安全，电网控制与智能制造。

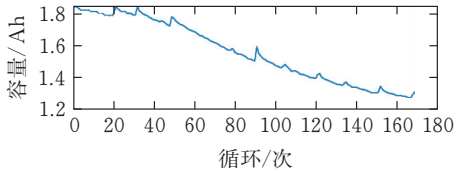
剩余寿命，能够在保证模型精度的同时提高泛化性能，避免过拟合。考虑到网络模型的超参数多、设置复杂等问题，使用遗传算法(genetic algorithm，GA)对网络模型进行超参数优化，减少参数调节繁杂度。最后在 3 个公开的电池数据集上对该网络模型的性能进行验证，表明本文方法能够较为准确地预测锂电池剩余寿命。

1 电池数据集介绍

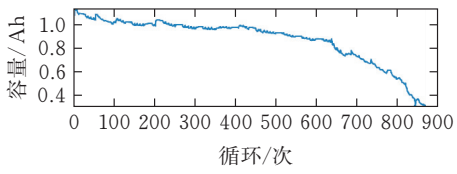
本文所用数据集包括 NASA 随机充放电数据集、马里兰大学 CALCE 数据集、牛津大学电池老化数据集。其中 NASA 电池数据集由美国国家航天局提供，数据中电池充放电均在室温下进行，充电采用恒流恒压充电模式，放电采用恒流放电模式。CALCE 数据集使用的电池阴极材料为 LiCoO2，同样采用恒流恒压充电模式，恒流放电模式。与 NASA 数据集不同的是该数据的测试温度为 1℃。牛津大学电池老化数据集使用 Kokam 钴酸锂软包电池，通过 ARTEMIS 驾驶循环测试得到，每一百次循环测量一次电池容量，因此该数据的有效循环次数较少。本文使用的数据集各项参数如表 1 所示。如图 1 所示，由于电池的型号不同以及测试条件的差异，电池容量衰退的情况也有所差异。

表 1 电池老化数据集参数

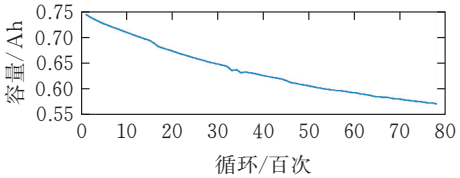
电池编号	额定容量/Ah	恒流充电电流/A	恒流放电电流/A	放电截止电压/V	环境温度/℃
B5	2	1.5	2	2.7	24
Cell1	0.74	1.48	1.48	2.7	40
CS_35	1.1	0.55	0.55	2.7	1



(a)NASA 数据集



(b)CACLE 数据集



(c)牛津大学数据集

图 1 不同数据集电池容量衰退曲线

2 基于辅助健康因子的锂电池剩余寿命预测模型

2.1 长短期记忆网络

长短期记忆网络(long short-term memory，LSTM)是在循环神经网络(Recurrent Neural Network，RNN)的基础上改进而来[14]。一方面 LSTM 解决了 RNN 的梯度消失和梯度爆炸问题，使得该网络能够用在较长的时间序列中，另一方面，解决了 RNN 训练模型需要预先确定最优延迟窗口长度的问题。典型的 LSTM 网络结构如图 2 所示。图中各变量的表达式如下：

$$i_t = \sigma(W_{xi}\,x_t + W_{hi}\,h_{t-1} + W_{ci}\,c_{t-1} + b_i) \tag{1}$$

$$f_t = \sigma(W_{xf}\,x_t + W_{hf}\,h_{t-1} + W_{cf}\,c_{t-1} + b_f) \tag{2}$$

$$c_t = f_t\,c_{t-1} + i_t \tanh\,(W_{xc}\,x_t + W_{hc}\,h_{t-1} + b_c) \tag{3}$$

$$o_t = \sigma(W_{xo}\,x_t + W_{ho}\,h_{t-1} + W_{co}\,c_t + b_c) \tag{4}$$

$$h_t = o_t \tanh\,(c_t) \tag{5}$$

式中，i 表示输入门，o 表示输出门，f 表示遗忘门，c 表示细胞状态。W 和 b 表示权重和偏置。σ 和 tanh 分别表示 sigmoid 激活函数和双曲正切激活函数。LSTM 的训练过程大致可分为如下四个步骤：首先按照公式(1－5)前向计算细胞输出，然后通过输出与标签的误差反向传播，继而根据误差调整细胞权重的梯度，最后根据优化方法更新细胞权重。

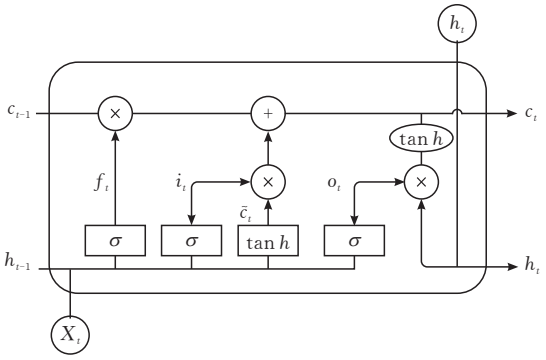


图 2 LSTM 网络结构

由于电池的循环次数有限，当前用于电池数据训练的模型均为小样本，因此使用单层 LSTM 模型能够实现较为准确的预测。但是当模型的训练数据扩充时，单层模型的预测准确性可能无法满足需求，因此本文使用双层的 LSTM 进行建模，更加深层的神经网络模型能够加强模型的拟合能力从而提高预测精度。然而更深的模型结构也会导致过拟合现象的产生。为了解决该问题，选择在模型中加入 dropout 层，此时每个神经元有一定的概率被移除，这样可以使得神经元训练期间保持独立性，同样也就使得特征之间的协同作用被减弱，以避免模型加深后产生的过拟合现象。

2.2　辅助健康因子的提取

在电池剩余寿命预测的相关研究中，大多使用电池容量作为描述电池健康状态的健康因子，当电池的容量下降到额定容量的 70%以下时可以认定电池到达了失效状态，即电池寿命终点（End of Life，EOL）。为了扩充用于电池寿命预测模型的训练数据，选择从 IC 曲线和 PD 曲线中提取相关特征作为模型训练的辅助健康因子。IC 曲线是以电池容量对电压进行微分得到的，而 PD 曲线则是通过数点法计算各区间内电压出现次数。相较基于电压差的特征参数，使用 PD 曲线最高峰值和 IC 曲线最高峰值能够快速定位，特征提取较为方便，无需根据电池型号的差异而调整电压区间。以牛津大学的 Cell1 电池数据为例，如图 3、图 4 所示，电池的 IC 曲线和 PD 曲线的最高峰值都随着电池循环次数的增加不断降低，因此使用其最高峰值作为辅助健康因子能够一定程度描述其电池老化状态。

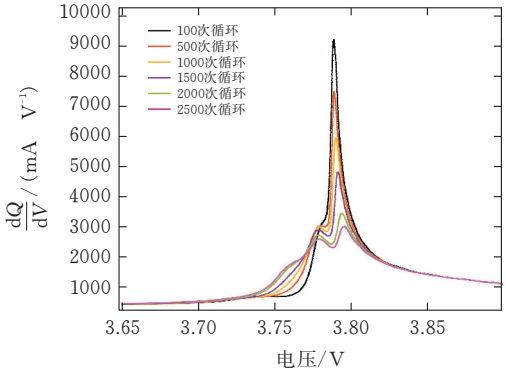


图 3　牛津大学数据集 Cell1 电池 IC 曲线

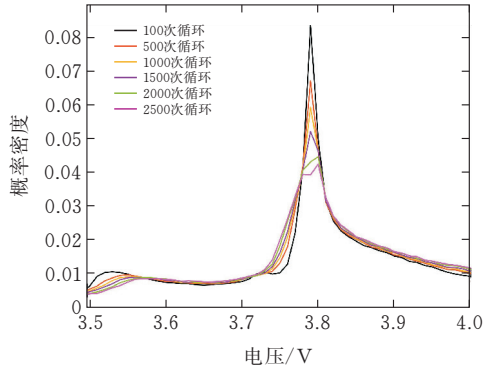


图 4　牛津大学数据集 Cell1 电池 PD 曲线

为了量化电池容量和 IC 峰值以及 PD 峰值之间的关联性，采用皮尔逊相关系数法分析，如：

$$r=\frac{\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\overline{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i-\overline{Y})^2}} \tag{6}$$

式中，X_i 表示序列 X 中的第 i 个元素，$\overline{X}$ 表示序列 X 中所有元素的平均值，Y_i 表示序列 Y 中的第 i 个元素，$\overline{Y}$ 表示序列 Y 中所有元素的平均值。经计算，IC 峰值和 PD 峰值与电池容量之间的相关系数为 0.883 和 0.925，结果表明 IC 峰值和 PD 峰值与电池容量之间具有很强的相关性，能够作为定义电池健康状态的健康因子。

2.3　剩余寿命预测模型建立

本文所提出的方法主要预测流程如图 5 所示。

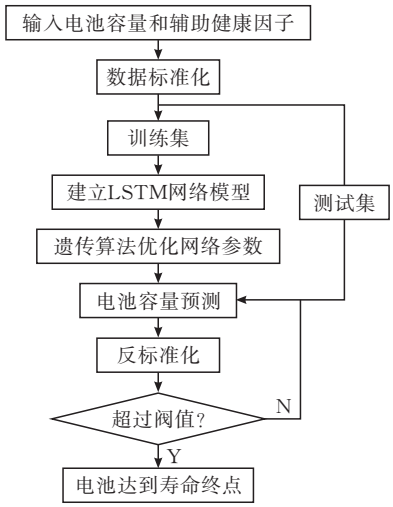


图 5　基于辅助健康因子的锂电池剩余寿命模型预测流程

1）将电池容量数据作为主要的健康因子，用于判断电池健康状态和电池失效标准，将 IC 曲线和 PD 曲线的最大峰值作为辅助健康因子，作为预测模型的训练数据。

2）将数据标准化后分为训练集和测试集，训练集用于网络模型的训练，测试集用于验证模型的性能。然后，使用遗传算法优化电 LSTM 网络模型的超参数，待优化的超参数有第一层网络的神经元个数，第二层网络的神经元个数，dropout 层参数，最大迭代次数。

3）用优化好的 LSTM 模型预测电池容量数据，将预测结果反标准化后得到最终的电池容量，如果容量达到失效阈值，则认为达到了电池寿命的终点，若没有达到失效阈值则认为电池依然能够正常使用。

3　实验分析

3.1　剩余寿命预测模型评价指标说明

为了有效、定量、综合地评价模型预测性能，引用均方根误差（root mean square error，RMSE）、绝对误差（absolute error，AE）和平均绝对误差（mean absolute error，MAE）作为模型的评价指标[15]。其中 RMSE 和 MAE 主要用于评价模型对于容量曲线的预测性能，而 AE 则是用于评价剩余循环寿命的指标。上述指标的计算公式如下：

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N} |C_Y(i) - C(i)|^2} \quad (7)$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |C(i) - C_Y(i)| \quad (8)$$

$$AE = |R - \widehat{R}| \quad (9)$$

$$R^2 = 1 - \frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (10)$$

其中：$C(i)$ 为电池容量实际值；$C_Y(i)$ 为电池容量预测值；N 为循环次数；R 为真实剩余寿命；$\widehat{R}$ 为预测剩余寿命，$\hat{y}_i$ 表示预测值，$\bar{y}$ 表示所有 y 值的平均值，y_i 为所有真实测量值。

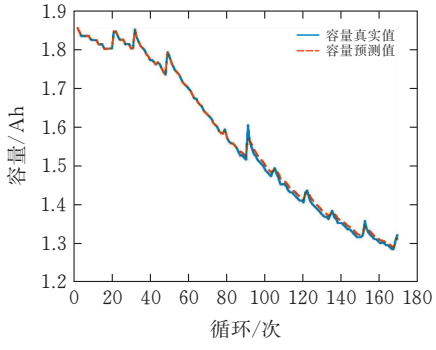
3.2 实验结果分析

为了验证本文方法有效性，设置容量的前 50% 数据以及辅助健康因子的前 50% 数据作为模型的训练集，容量的后 50% 数据作为模型的测试集。设置 LSTM 第一层和第二层的隐藏神经元个数的取值范围为 3～8，设置 dropout 层的参数取值范围为 0.2～0.5，最大迭代次数范围为 50～300。遗传算法的参数设置如表 2 所示。

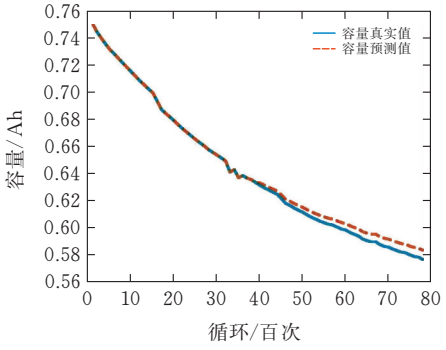
表 2 遗传算法参数设置说明

参数名称	数值	参数名称	数值
迭代次数	300	变异概率	0.1
种群规模	100	交叉概率	0.5

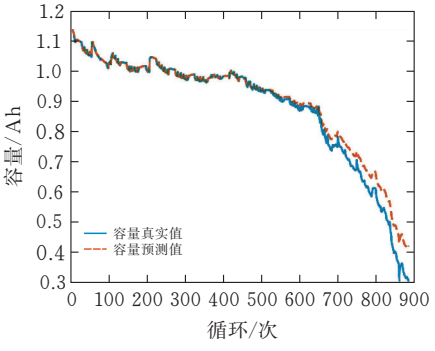
在预测起点为 50% 时各电池的预测结果如图 6 所示，本文提出的 RUL 预测方法模型能够较好地预测电池的容量曲线，虽然在 CALCE 电池数据集中该模型对电池容量的预测性能不理想，但是该电池在容量为 0.77Ah 时已达到了寿命终点，后续电池的容量衰退情况与前期训练的健康因子衰退趋势有所差异，因此对于该数据集的预测性能可以以电池的寿命绝对误差 AE 作为参照指标。而牛津大学数据集中由于电池的衰退较慢，无法在特定循环前达到失效标准，因此对于该数据集设置容量 80% 为电池失效标注。本文模型预测的 RMSE、MAE 以及 AE 指标计算结果见表 3，该模型在电池的寿命绝对误差指标上具有较好的表现。



(a)NASA 数据集



(b)牛津大学数据集



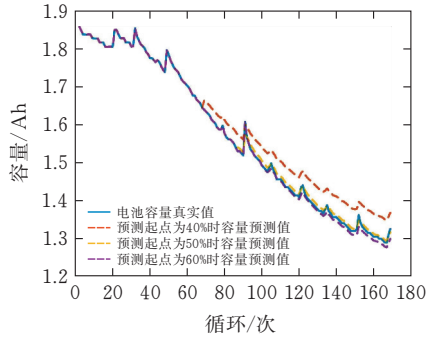
(c)CALCE 数据集

图 6 模型在预测起点为 50% 时的预测性能

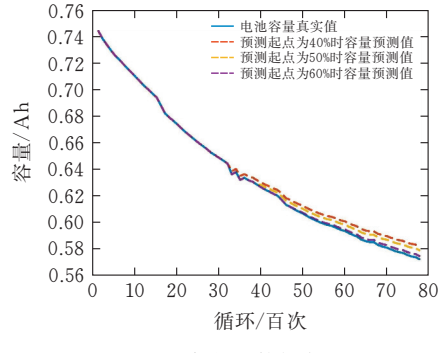
表 3 预测起点为 50% 时模型预测性能

电池编号	RMSE	MAE	R^2	AE
B5	0.0093	0.0084	0.9948	2
Cell1	0.0048	0.0046	0.9953	3
CS_35	0.0331	0.0287	0.9586	1

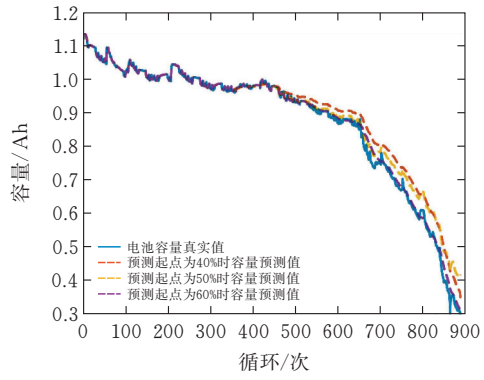
为了充分验证模型的有效性，在 40% 和 60% 的预测起点下进行预测，预测结果如图 7 所示。



(a)NASA 数据集



(b)牛津大学数据集



(c)CALCE 数据集

图 7　模型在不同预测起点下的性能

本文所提模型在 NASA 电池数据集中设置预测起点为 50%以上时，预测效果较好，但在预测起点为 40%时，模型的性能较差，这是因为 NASA 电池数据在测试一段时间后会有静置操作，但是静置过程并无明显规律，因此在训练数据不足时该模型的预测效果不理想。而在 CALCE 数据集和牛津大学数据集中，本文所提方法的预测效果较准确，不论在数据长度较短的牛津大学数据集还是数据长度较大的 CALCE 数据集中都能有较好的预测性能。在不同预测起点的详细性能指标见表 4。

表 4　预测起点为 40%和 60%时模型预测性能

电池编号	预测起点	RMSE	MAE	R^2	AE
B5	60%	0.0485	0.0473	0.9988	1
	40%	0.1046	0.0089	0.8304	11
Cell1	60%	0.0067	0.0063	0.9996	1
	40%	0.0015	0.0014	0.9891	5
CS_35	60%	0.0499	0.0403	0.9948	0
	40%	0.0176	0.0127	0.9422	4

由表 4 数据可知，本文所提方法不仅在电池容量曲线的预测性能上较为优秀，而且在剩余寿命的绝对误差上具有良好的预测性能。

4　结论

本文针对锂电池健康寿命问题提出的基于辅助健康因子的锂电池剩余寿命预测模型方法，能有效预测电池容量衰退情况，实现对电池剩余寿命的间接预测。该方法在较大数据集和较小数据集中都有良好的表现。辅助的健康因子能够有效扩充电池训练模型，使得该模型能够在训练集为 50%时表现出良好的预测性能。随着电池数据集加入训练模型的数量比例增加，预测效果也有所提升。遗传算法优化的 LSTM 网络能够使模型的超参数得到优化，使建模过程得到简化，提升模型的预测性能。

[参 考 文 献]

[1] GE M F, LIU Y, JIANG X, et al. A review on state of health estimations and remaining useful life prognostics of lithium-ion batteries[J]. Measurement, 2021, 174(01): 109057.

[2] LI X, HU Y, XUE B, et al. State-of-health estimation for the lithium-ion battery based on gradient boosting decision tree with autonomous selection of excellent features[J]. International Journal of Energy Research, 2022, 46(02):1756-1765.

[3] CHEN Y, KANG Y, ZHAO Y, et al. A review of lithium-ion battery safety concerns: The issues, strategies, and testing standards[J]. Journal of Energy Chemistry, 2021, 59(01):83-99.

[4] WEI Y, WU D. Prediction of state of health and remaining useful life of lithium-ion battery using graph convolutional network with dual attention mechanisms [J]. Reliability Engineering and System Safety, 2023, 230(01): 108947.

[5] ZHOU W, PAN Z B, WANG J, et al. Review on acoustic emission source location, damage recognition and lifetime prediction of fiber-reinforced composites [J]. Journal of Materials Science, 2023, 58(02):583-607.

[6] WEI M, YE M, WANG Q, et al. State-of-health estimation and remaining useful life prediction of lithium-ion batteries based on extreme learning machine[J]. Journal of Physics: Conference Series, 2021, 1983 (01):012058.

[7] ASHWIN T R, CHUNG Y M, WANG J. Capacity fade modelling of lithium-ion battery under cyclic loading conditions[J]. Journal of Power Sources, 2016, 328(OCT.1):586-598.

[8] ZHANG Y Z, XIONG R, HE H W, et al. Aging characteristics-based health diagnosis and remaining useful life prognostics for lithium-ion batteries[J]. eTransportation, 2019, 1(01):100004.

[9] LIU Y, SUN J, SHANG Y, et al. A novel remaining useful life prediction method for lithium-ion battery based on long short-term memory network optimized by improved sparrow search algorithm[J]. Journal of Energy Storage, 2023, 61(01): 106645.

[10] BAI G, SU Y, RAHMAN M M, et al. Prognostics of Lithium-Ion batteries using knowledge-constrained machine learning and Kalman filtering[J]. Reliability Engineering & System Safety, 2023, 231(01): 108944.

[11] YANG Y, WEN J, LIANG J, et al. Remaining useful life prediction for lithium-ion batteries based on the partial voltage and temperature[J]. Sustainability, 2023, 15(02): 1602.

[12] KWON S, HAN D, PARK J, et al. Joint state-of-health and remaining-useful-life prediction based on multi-level long short-term memory model prognostic framework considering cell voltage inconsistency reflected health indicators[J]. Journal of Energy Storage, 2022, 55(01): 105731.

[13] MA B, YANG S, ZHANG L, et al. Remaining useful life and state of health prediction for lithium batteries based on differential thermal voltammetry and a deep-learning model[J]. Journal of Power Sources, 2022, 548(01): 232030.

[14] 唐万哲. 基于数据驱动方法的锂离子电池剩余寿命预测研究[D].武汉:武汉理工大学,2021.

[15] 唐伦,蒲昊,汪智平, 等.基于注意力机制 ConvLSTM 的 UAV 节能预部署策略[J].电子与信息学报,2022, 44(03):960-968.

Research on Remaining Life Prediction of Lithium Ion Batteries Based on Auxiliary Health Factor

JIN Haobo[1], WANG Shuqing[1], YUAN Xiaohui[2]

(1 *Hubei Key Laboratory for High-efficiency Utilization of Solar Energy and Operation Control of Energy Storage System, Hubei Univ. of Tech., Wuhan* 430068, *China*;
2 *School of Civil and Hydraulic Engin., Huazhong Univ. of Sci. and Tech., Wuhan* 430074, *China*)

Abstract: To improve the accuracy of the remaining life prediction of lithium ion batteries, an auxiliary health factor based remaining life prediction method for lithium batteries is proposed, which can accurately predict the remaining life of lithium batteries using less training data. Health factors describing battery aging are extracted from the incremental battery capacity curve and probability density curve and used for the auxiliary training of the battery life prediction model. The existing Long Short Term Memory (LSTM) model for battery life prediction is improved and the hyperparameters of the model are optimized using Genetic Algorithm (GA) to improve the prediction accuracy of the model. Finally, the effectiveness of the proposed method is verified in three publicly available datasets.

Keywords: lithium ion battery; remaining life prediction; long and short-term memory networks; health factors

[责任编校: 闫 品]