DOI：10.13234/j.issn.2095-2805.2026.5.248 中图分类号：TM912 文献标志码：A

基于集成模型的锂离子电池 SOH 估计

雷二涛[1]，张浚坤[1]，王瑾瑜[2]，金 莉[1]，马 凯[1]，李 盈[1]
(1.广东电网有限责任公司电力科学研究院，广州 510080；2.北京交通大学电气工程学院，北京 100044)

摘要：准确估计电池健康状态 SOH (state of health)可以有效提升电池系统运行安全性。针对不同工况下 SOH 估计精度不稳定的问题，提出一种基于集成模型的锂离子电池 SOH 估计算法。首先对电池充电阶段开展容量增量 IC(incremental capacity)分析，并提取健康因子作为电池 SOH 的有效特征，接着采用支持向量回归 SVR(support vector regression)模型与反向传播 BP(back propagation)神经网络模型组合构建集成模型，实现对电池 SOH 的在线评估。实验结果表明，所提出的集成模型相较于 2 个子模型最大误差 RMSE(root mean square error)分别下降了 38.1%和 23.7%，均方根误差分别下降了 47.8%和 20.0%，所提方法能更准确估计不同工况下的电池 SOH，削弱了单个模型估计的不稳定性因素。

关键词：锂离子电池；容量增量分析；健康状态估计；集成模型

Estimation of Lithium-ion Battery State of Health Based on Stacking Model

LEI Ertao[1], ZHANG Junkun[1], WANG Jinyu[2], JIN Li[1], MA Kai[1], LI Ying[1]
(1. Electric Power Research Institute of Guangdong Power Grid Co., Ltd., Guangzhou 510080, China; 2. School of Electrical Engineering, Beijing Jiaotong University, Beijing 100044, China)

Abstract: Accurate estimation of the battery state of health (SOH) can effectively improve the safe operation of system. Aiming at the problem of unstable SOH estimation accuracy under different working conditions, a stacking model-based method for lithium-ion battery SOH estimation is proposed. The method firstly carries out incremental capacity (IC) analysis for the charging data, and extracts the health factor as an effective feature, and then constructs an stacking model by stacking the support vector regression (SVR) and back propagation (BP) neural network model to realize the on-line evaluation of the battery SOH. The experimental results show that the maximum error of the stacking model proposed decreases by 38.1% and 23.7%, and the root mean square error (RMSE) decreases by 47% and 20% compared with the two sub-models. The proposed method can more accurately estimate the SOH under various operating conditions, mitigating the instability factor of the single model estimation.

Keywords: Lithium-ion battery; incremental capacity (IC) analysis; estimation of state of health (SOH); stacking model

锂离子电池具有能量密度大、循环寿命长等优势，被大量应用于新能源汽车，并且作为核心零部件决定了新能源汽车的性能与可行驶里程。随着电池使用寿命的逐渐增加，电池内部的电化学反应会导致电池内部的材料发生不可逆的损失，进而影响电池的性能，严重时需要及时更换电池[1]。电池健康状态 SOH (state of health)的准确估计是电池管理系统的重要环节，通过该环节可以对电池的故障进行早期检测，提高电池组运行的安全性与稳定性[2]。因此，准确估计锂离子电池的 SOH 具有重大研究意义[3]。

现有的 SOH 估计方法可分为 3 类：实验估计法、模型估计法和数据驱动法。实验估计法是对电池容量、电池阻抗和电化学阻抗谱等与电池老化相

收稿日期：2024-02-28；修回日期：2024-03-26，2024-04-09；
录用日期：2024-04-17；网络首发日期：2024-06-11

关的特征参数直接进行测量，一般用在研究电池退化机理的理论研究中，具有最高的估计精度，并且对不同类型的电池均具有普适性。但是实验估计法要求也较高，该方法不仅需要具有较高精度的实验测试设备，且可操作性差，只能在实验室条件下进行，在实际运行条件下应用会受到采集精度、采集数据准确度、运行工况等多方面的影响[4]。模型估计法基于电池的老化机理，通过分析与电池老化有关的健康因子来估计电池的 SOH，该方法可以在复杂工况下对电池的内部特性与外部特性进行模拟，其常用模型有电化学模型与等效电路模型[5]。模型估计法的优势在于对不同环境、不同运行工况下的电池均有较好的 SOH 估计方法与较高的估计精度，且可在电池机理层面解释电池的老化过程；其缺点是模型一般较为复杂，计算量较大且数据质量要求高，在线实时估计困难[6]。数据驱动法仅需电池组的可测试数据，如电压、电流、温度等数据，即可从历史序列中挖掘出电池性能的衰退模式，然后使用特定的算法构建模型，最后应用其他电池的输入估计电池的 SOH[7]。数据驱动法的适应能力强，不需要对电池内部复杂的电化学反应过多分析，也不需要准确的电池模型，只需要电池运行的数据即可进行 SOH 的估计，但其估计的精度受数据质量的影响较大，当数据质量较差时估计结果的精度会大大降低。但是相较于实验估计与模型估计法，数据驱动法使用较为简便，是目前电池 SOH 估计中最常见的一种方法[8]。

由上可知，数据驱动法在锂离子电池 SOH 估计领域具有较大的应用场景。本文通过支持向量机回归 SVR(support vector regression)模型与反向传播 BP(back propagation)神经网络模型叠加构建了集成模型，建立了电池充电容量增量 IC(incremental capacity)曲线的特征量到 SOH 的映射关系，集成模型在 2 个子模型的基础上，增加了模型的稳定性。实验结果表明集成模型具有更高的估计精度，对不同工况下的电池具有良好的适应能力。

1　实验设计

本文实验包含 10 节额定容量为 36 A·h、额定电压为 3.6 V 的电池单体，5 种不同温度应力与不同倍率应力的工况。具体工况设置见表 1。

表 1　电池老化测试条件

Tab. 1 Aging test conditions of the experimented battery

电池编号	实验温度/℃	充/放电倍率(C)
B1、B2	25	1.0/1.0
B3、B4	45	1.0/1.0
B5、B6	10	1.0/1.0
B7、B8	25	0.5/1.0
B9、B10	25	1.0/1.5

实验流程包括老化循环实验与定期的性能测试，具体实验步骤如下。

步骤 1 充电选择恒流恒压充电方式，首先以设置的倍率充电至截止电压 4.15 V；然后改为恒压充电，直至充电电流减小至 0.05C 后，停止充电。

步骤 2 放电选择恒流放电方式，以设置的倍率放电至截止电压 2.5 V 后，停止放电。

步骤 3 重复步骤 1 与步骤 2，每间隔 100 次循环实验进行 1 次性能测试。性能测试包括混合脉冲功率特性 HPPC(hybrid pulse power characterization)测试，小电流测试与不同倍率的充放电测试。

步骤 4 在电池衰退至截止条件后，停止循环实验并进行最后一次性能测试，实验结束。

随着电池循环圈数的逐渐增加，电池内部的化学材料不断损失，与此同时，不可逆反应还原的化学产物也在不断累积，导致电池性能不断衰退。本文 10 节电池的容量衰退曲线如图 1 所示。可以看出，不同工况下电池的循环寿命有较大区别，B7 与 B8 电池的循环寿命可达约 900 圈，而 B5 与 B6 电池的循环寿命仅有约 200 圈，电池容量的衰退路径为非线性衰退，且在某些循环下出现了容量恢复现象。上述现象均使得电池的 SOH 估计变得更为复杂，给估计过程带来了更大的不确定性。

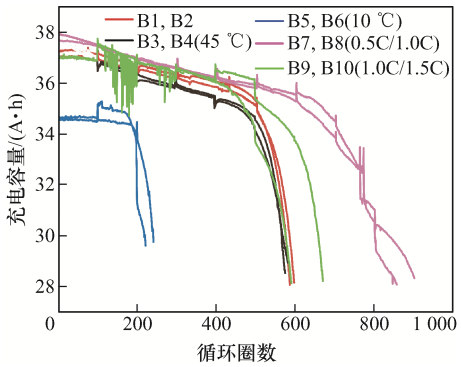


图 1 数据集容量衰退曲线

Fig. 1 Capacity degradation curves of the dataset

2 特征提取与分析

2.1 特征提取

将电池充放电过程中电压-容量(*U*-*Q*)曲线微分即可得到容量增量(IC)曲线。IC 曲线将电压转化为更容易观察的峰值，在锂离子电池全寿命期间的变化状态与电池的 SOH 有着一定联系。该曲线具有较高的敏感性，并且可以从侧面分析电池内部的电化学健康状态。图 2 展示了 B1 电池充电阶段的 *U*-*Q* 曲线与 IC 曲线。IC 曲线的不同峰值与谷值较好地反映出了 *U*-*Q* 曲线的不同阶段，通过分析 IC 曲线中的波峰、波谷以及对应的区间面积等信息估计电池的 SOH。

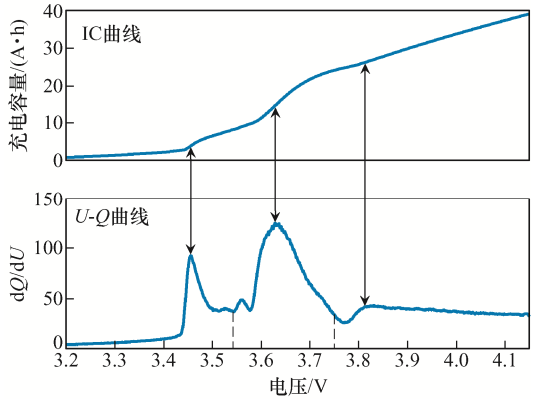


图 2 充电阶段 *U*-*Q* 曲线与 IC 曲线

Fig. 2 *U*-*Q* curve and IC curve during charging stage

图 3 为在性能测试中 IC 曲线随着循环次数增加的变化图，并选择了 S-G 滤波器对曲线噪声进行滤波处理。随着充放电次数的逐渐增加，IC 曲线的峰高与峰位置的变化具有一定的规律性，峰①、峰②、峰③的峰值逐渐降低，同时 3 个峰对应的电压位置均有着一定程度的右偏移。此外，峰③的峰值随着电池老化逐渐变得不明显，在某些电池的老化末期甚至完全消失。由于峰③在老化状态的后期完全无法使用，因此选择峰①和峰②的峰高度与峰位置作为特征，不在峰③上提取特征。

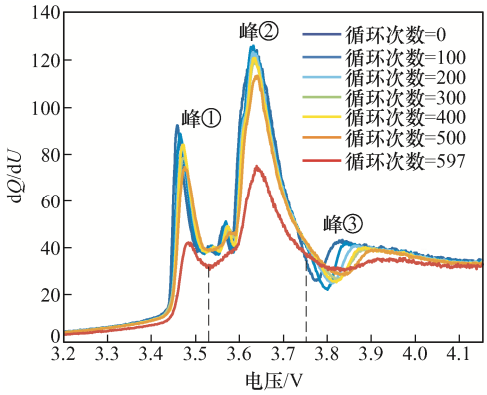


图 3 IC 曲线变化图

Fig. 3 IC curve evolution over the full life cycle

由于电池的衰退信息更多地表现在 IC 曲线的整体而非某些单独的数据点上，因此对整个充电过程进行分段，选择每一段 IC 曲线对应的面积作为模型的输入。选择峰①与峰②2 个峰之间的谷值点作为第一个电压分段点；在第二个电压分段点的选择上，由于峰③的位置在电池的老化末期难以确定，因此选择 3.75 V 为分段点。图 4 展示了部分特征与容量间的相关性。

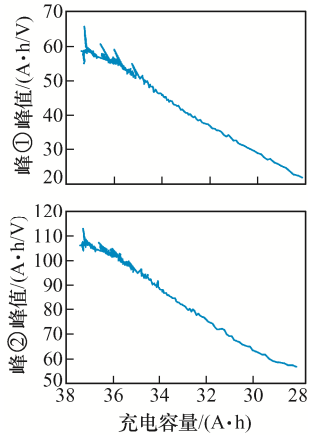


图 4 部分特征与容量变化的关系图

Fig. 4 Selected features vs. capacity curves

2.2 特征分析

本文选择距离相关系数函数作为特征的评价方法，计算公式为

$$\hat{d}_{\mathrm{corr}}(u,v)=\frac{\hat{d}_{\mathrm{cov}}(u,v)}{\sqrt{\hat{d}_{\mathrm{cov}}(u,u)\hat{d}_{\mathrm{cov}}(v,v)}} \tag{1}$$

式中：u 与 v 为待分析的变量；$\hat{d}_{\mathrm{corr}}(\cdot)$ 为样本距离相关系数估计函数，取值[0, 1]，越接近 1，说明依赖关系越强，为 0 时表示变量相互独立；$\hat{d}_{\mathrm{cov}}(\cdot)$ 为样本距离协方差估计函数。距离相关系数是为了克服 Pearson 相关系数的弱点而构建的，即使 Pearson 相关系数为 0，两变量之间也可能为非线性相关的关系，但如果距离相关系数为 0，则两变量一定相互独立。所有电池特征与充电容量间的距离相关系数见表 2。

表 2　所有电池特征与充电容量间的距离相关系数

Tab. 2 Distance correlation coefficient between all battery features and charge capacity

特征	距离相关系数
峰①高度	0.955 0
峰①位置	0.916 8
峰②高度	0.975 8
峰②位置	0.875 5
第一段 IC 曲线对应面积	0.856 8
第二段 IC 曲线对应面积	0.669 7
第三段 IC 曲线对应面积	0.781 4

所选择的特征与电池的充电容量之间有着较强的相关性，大部分特征在不同工况下与容量间的相关性可达 0.800 0 以上，随着电池容量的逐渐衰减，特征也有着较为规律的变化，可以很好地用来估计电池的 SOH。

3 机器学习算法

3.1 支持向量回归(SVR)

支持向量回归 SVR 是支持向量机 SVM(support vector machine)中一个很重要的分支，回归问题解决的是寻找 1 个最优超平面，可以使得所有的样本点到超平面之间的距离“最短”[9-10]，其基本原理如图 5 所示。

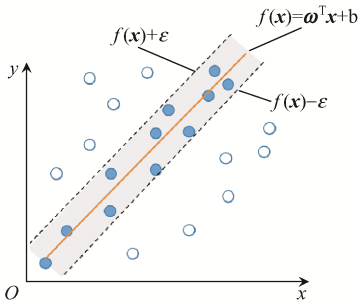


图 5　SVR 基本原理图

Fig. 5 Basic principle diagram of SVR

对于线性硬间隔 SVR，优化目标为

$$\min_{\boldsymbol{w},b}\frac{1}{2}\|\boldsymbol{w}\|^2 \tag{2}$$

式中，$\boldsymbol{w}$ 为权重向量；b 为偏置。其中位于边界内的点满足条件

$$|y_i-(\boldsymbol{w}\boldsymbol{x}_i+b)|\leqslant\varepsilon \tag{3}$$

式中：y_i 为第 i 个样本的真实值；$\boldsymbol{x}_i$ 为第 i 个样本的输入特征向量；ε 为误差容忍度。参数 $\boldsymbol{w}$、b 的确定，不仅需要最大化间隔带，同时也需要最小化损失。

$\boldsymbol{w}$ 和 b 是决定超平面的 2 个参数，经过模型训练可以得到最优的 $\boldsymbol{w}$ 和 b，即可获得最优的超平面。SVR 的代价函数为

$$\sum_{i=1}^{I}l_{\epsilon}[f(\boldsymbol{x}_i),y_i] \tag{4}$$

$$l_{\epsilon}(z)=\begin{cases}0 & |z|\leqslant\varepsilon\\ |z|-\varepsilon & \text{其他}\end{cases} \tag{5}$$

式中：I 为样本总数；$l_{\epsilon}(\cdot)$ 为误差容忍度函数；$f(\boldsymbol{x}_i)$ 为第 i 个样本的预测值；z 为误差项。

从而 SVR 问题可形式化为

$$\min_{\boldsymbol{w},b}\frac{1}{2}\|\boldsymbol{w}\|^2+C\sum_{i=1}^{I}l_{\epsilon}[f(\boldsymbol{x}_i),y_i] \tag{6}$$

式中，C 为惩罚系数。引入松弛变量 ξ，使 SVR 的要求降低，即允许一些数据点不在区间内，SVR 问题可重写为

$$\min_{\boldsymbol{w},b,\xi_i,\hat{\xi}_i}\frac{1}{2}\|\boldsymbol{w}\|^2+C\sum_{i=1}^{I}[\xi_i,\hat{\xi}_i] \tag{7}$$

3.2 BP 神经网络模型

BP 神经网络是一种拥有 3 层或多层的神经

网络，每一层网络均由若干个神经元构成，且左侧的每一个神经元均与右侧的每一个神经元连接，而每一层内部的上、下神经元之间没有连接[11]。图 6 为只有 1 个隐含层的 BP 神经网络模型。

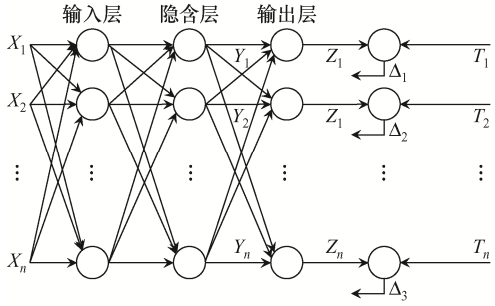


图 6 BP 神经网络模型

Fig. 6 BP neural network model

BP 神经网络的构建就是依据输入信号的前向传播与预测误差的反向传播。假设神经网络结构为图 6 所示的 3 层 BP 神经网络，输入层的节点数目为 n，隐含层的神经元数目为 1，输出层的节点数目为 m。隐含层的输出为

$$H_j = g\left(\sum_{i=1}^{n} w_{ij}x_i + a_j\right) \tag{8}$$

式中：$g(\cdot)$为激励函数；w_{ij} 为输入层与隐含层之间的权重；x_i 为第 i 个样本的输入；a_j 为输入层与隐含层之间的偏置。

输出层的输出为

$$O_k = \sum_{j=1}^{l} H_j w_{jk} + b_k \tag{9}$$

式中：w_{jk} 为隐含层与输出层之间的权重；b_k 为隐含层与输出层之间的偏置。

权重与偏置的更新迭代公式为

$$\begin{cases} w_{ij} = w_{ij} + \eta H_j(1-H_j)x_i\sum_{k=1}^{m} w_{jk}e_k \\ w_{jk} = w_{jk} + \eta H_j e_k \end{cases} \tag{10}$$

$$\begin{cases} a_j = a_j + \eta H_j(1-H_j)\sum_{k=1}^{m} w_{jk}e_k \\ b_k = b_k + \eta e_k \end{cases} \tag{11}$$

式中：η 为学习速率；e_k 为输出层第 k 个神经元的误差项。

3.3 集成模型

SVR 模型可采用不同的核函数向高维空间映射，使用简单，但数据量较大时效果较差，且该模型对初始参数与核函数的选择较为敏感，受缺失数据的影响较大；BP 神经网络模型可以处理高维度数据，适应能力强，但对数据量的需求较大，且其结构较为复杂，调参工作较为烦琐。本文结合 2 个模型的优点，构建新的集成模型。

图 7 为集成模型的框架图，其思路是并行训练各自独立的不同类型的模型，再将每个模型的输出值输入到 1 个训练好的集成计算模型中，将各个模型中的输出结果进行结合。首先构建好电池的健康因子数据；然后将其分别输入到 SVR 模型与 BP 神经网络模型中，得到 2 个 SOH 估计值；最后在集成计算模块中采用数学平均的方法计算两者的平均值，得到最终的 SOH 估计值。

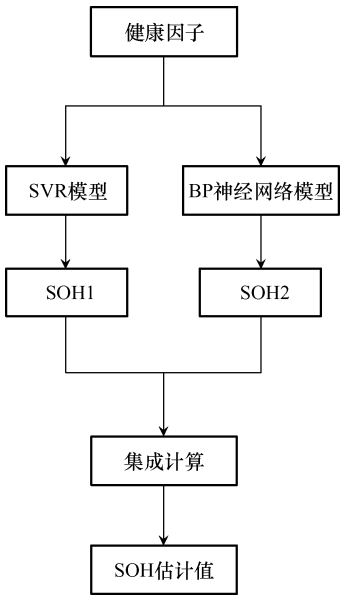


图 7 集成模型框架图

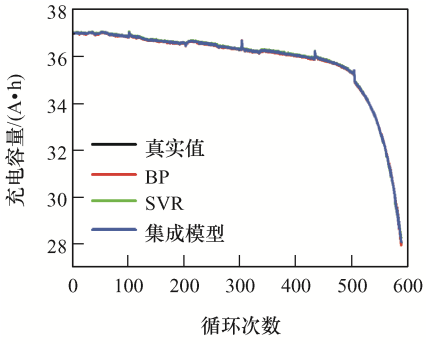
Fig. 7 Stacking model framework diagram

4 SOH 估计结果

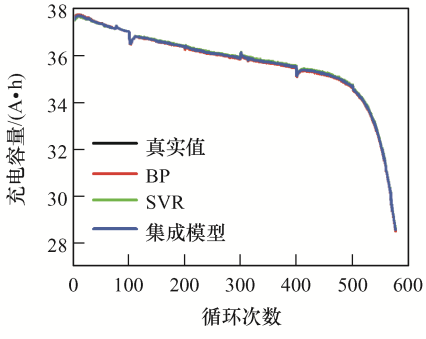
10 节电池的衰退路径都呈现出从近似线性衰退开始，逐渐过渡到加速衰退的变化过程，但是不同的实验工况导致电池的衰退速率也有差

异，本文基于容量增量法提取的 7 个特征，并结合上述的 3 种机器学习算法，构建了 3 种 SOH 估计模型。

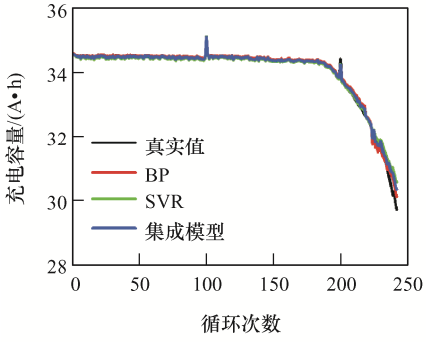
在数据集的划分上，选择 4 节电池作为训练集，分别为 B1、B3、B7、B9，其余 6 节电池作为验证集。SVR 模型中惩罚因子为 6.0，径向基函数参数为 0.1。BP 神经网络模型中迭代次数设置为 1 000 次，预测误差阈值为 10^{-6}，学习率为 0.01。所有电池在所有模型下的估计结果如图 8 所示。采用最大绝对百分比误差、平均绝对百分比误差和均方根百分比误差计算模型在验证集上的误差，具体计算方法见表 3，其中 $\hat{y}_i$ 为第 i 个样本的估计值。



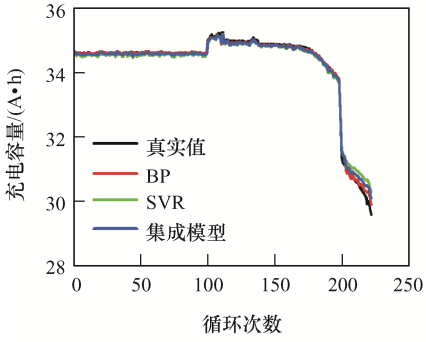
(a) B2电池



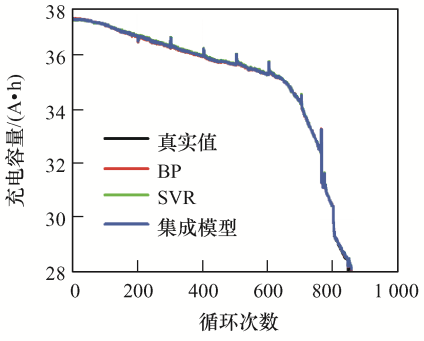
(b) B4电池



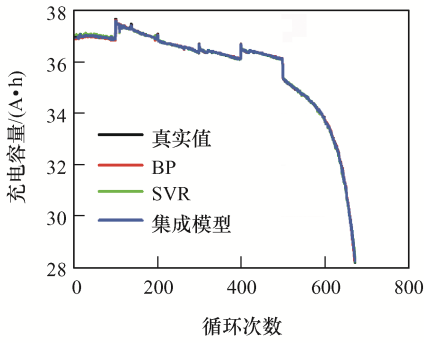
(c) B5电池



(d) B6电池



(e) B8电池



(f) B10电池

图 8　验证集电池估计结果

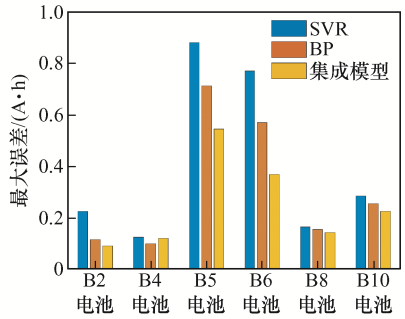
Fig. 8 Estimation results of the validation set

所有模型的误差对比如图 9 所示，误差统计结果见表 4。由于电池工况的不同，单个模型输出的估计结果可能会存在不稳定的现象，如 B5 电池和

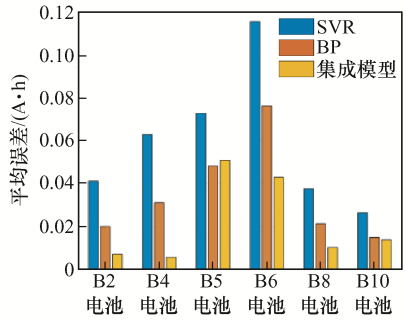
表 3　误差评价体系

Tab. 3 Error evaluation system

误差	表达式
最大绝对百分比误差	$\max\limits_{1\leqslant i\leqslant n}\lvert(y_i-\hat{y}_i)/y_i\rvert$
平均绝对百分比误差	$\dfrac{1}{N}\sum\limits_{i=1}^{N}\lvert(y_i-\hat{y}_i)/y_i\rvert$
均方根百分比误差	$\sqrt{\dfrac{1}{N}\sum\limits_{i=1}^{N}[(y_i-\hat{y}_i)/y_i]^2}$



(a) 最大误差



(b) 平均误差

图 9 所有模型误差对比图

Fig. 9 Error comparison of all models

表 4 单一模型与所提模型误差统计结果

Tab. 4 Error statistics results of single models and the proposed model

估计模型	最大绝对百分比误差/%	平均绝对百分比误差/%	均方根百分比误差/%
支持向量机	2.91	0.14	0.23
BP 神经网络	2.36	0.08	0.15
集成模型	1.80	0.04	0.12

B6 电池工况与其他电池不同， SVR 与 BP 模型对其的估计精度远低于其他电池，而集成模型估计结果的最大误差相较于其他 2 个模型，有着小幅度下降。由表 4 可知，集成模型结合各个算法的优势，在单一模型的基础上进行了优化，使得模型在验证集上的估计结果相较于 2 个子模型的最大绝对百分比误差分别下降了 38.1%和 23.7%，平均绝对百分比误差分别下降了 71.4%和 50.0%，均方根百分比误差分别下降了 47.8%和 20.0%。实验结果表明，该模型相较于其他的估计模型，估计精度更高，尤其对未预见的工况，所提算法的估计效果具有较高的优势。

5 结语

本文提出了一种基于集成模型的 SOH 估计方法，通过分析 IC 曲线构建了与容量衰退曲线相关性较高的 7 个健康因子，用于 SOH 估计模型的输入。SOH 估计模型选择支持向量机回归模型与 BP 神经网络模型作为集成模型的子模型。通过 10 节三元锂离子电池在 5 种不同工况下全寿命周期内的数据进行实验验证，结果表明构建的集成模型相较于 2 个子模型，最大绝对百分比误差分别下降了 38.1%和 23.7%，平均绝对百分比误差分别下降了 71.4%和 50.0%，均方根百分比误差分别下降了 47.8%和 20.0%。本文提出的方法可对不同工况、不同寿命的电池开展 SOH 评估，且具有较高的精度。相较于单一模型，所开发的方法在未预见工况中 SOH 估计效果极佳，对不同工况下电池的安全管理、延长使用寿命均有指导意义。

参考文献：

[1] Sun Fengchun. Green energy and intelligent transportation—promoting green and intelligent mobility [J]. Green Energy and Intelligent Transportation, 2022, 1(1): 100017.

[2] 成燕, 郑林弘, 刘江. 基于模态分解和时间序列的锂电池健康状态估计[J]. 电源学报, 2025, 23(5): 241-250.
Cheng Yan, Zheng Linhong, Liu Jiang. Estimation of state-of-health of lithium-ion batteries based on modal decomposition and time series [J]. Journal of Power Supply, 2025, 23(5): 241-250. (in Chinese)

[3] Pham T, Truong L, Bui H, et al. Towards channel-wise bidirectional representation learning with fixed-point positional encoding for SOH estimation of lithium-ion battery [J]. Electronics, 2023, 12(1): 98.

[4] 赵靖英, 黄麟然, 姚帅亮. 考虑容量再生的锂离子电池健康状态估计方法[J]. 电源学报, 2026, 24(1): 242-251.
Zhao Jingying, Huang Linran, Yao Shuailiang. State-of-health estimation method for lithium-ion battery conside-

ring capacity regeneration [J]. Journal of Power Supply, 2026, 24(1): 242-251. (in Chinese)

[5] Fang Linlin, Li Junqiu, Peng Bo. Online estimation and error analysis of both SOC and SOH of lithium-ion battery based on DEKF method [J]. Energy Procedia, 2019, 158: 3008-3013.

[6] Wesskamp P, Reitemeyer S, Melbert J. Online capacity estimation for automotive lithium-ion cells incorporating temperature-variation and cell-aging [J]. SAE Technical Paper Series, 2017, 1: 2017-1-1191.

[7] Lamorgese A, Mauri R, Tellini B. Electrochemical-thermal P2D aging model of a $LiCoO_2$/graphite cell: capacity fade simulations [J]. Journal of Energy Storage, 2018, 20: 289-297.

[8] Li Alan g, Mayilvahanan K, West A C, et al. Discrete-time modeling of Li-ion batteries with electrochemical overpotentials including diffusion [J]. Journal of Power Sources, 2021, 500: 229991.

[9] 余俊泽, 夏显威, 雷春俊, 等. 基于贝叶斯优化的支持向量回归模型对电能表在线率的预测[J]. 广东电力, 2023, 36(9): 72-79.
Yu Junze, Xia Xianwei, Lei Chunjun, et al. Prediction of electricity meter online rate using support vector regression model based on Bayesian optimization [J]. Guangdong Electric Power, 2023, 36(9): 72-79. (in Chinese)

[10] Lee H, Kim D, Gu J H. Prediction of food factory energy consumption using MLP and SVR algorithms [J]. Energies, 2023, 16(3): 1550.

[11] Yu Quanqing, Liu Yukun, Long Shengwen, et al. A branch current estimation and correction method for a parallel connected battery system based on dual BP neural networks [J]. Green Energy and Intelligent Transportation, 2022, 1(2): 100029.



雷二涛

作者简介：

雷二涛(1987—)，男，博士研究生，高级工程师。研究方向：电能质量，新能源及储能。E-mail：leiertao2008@163.com。

张浚坤(1992—)，男，通信作者，博士研究生，工程师。研究方向：新能源，储能及特种电源。E-mail：junkunzhang@126.com。

王瑾瑜(1996—)，女，博士。研究方向：锂离子电池寿命预测。E-mail: 18117023@bjtu.edu.cn。

金莉(1989—)，女，博士研究生，高级工程师。研究方向：电力电子在电力系统中的应用，新能源及储能技术。E-mail：dkyjinli@163.com。

马凯(1985—)，男，硕士研究生，正高级工程师。研究方向：柔性直流，电化学储能技术。E-mail：makai20023711@163.com。

李盈(1994—)，女，硕士研究生，工程师。研究方向：电力电子与储能技术。E-mail：liying2014ncepu@163.com。

(责任编辑：郭凯凯)