

基于噪声抑制半监督学习的 锂离子电池 SOH 估计方法

张雨桐¹, 谢军^{1*}, 王春鑫¹, 谢庆¹, 王楠²

- (1. 新能源电力系统国家重点实验室(华北电力大学), 北京 昌平区 102206;
2. 国网综合能源服务集团有限公司, 北京 西城区 100052)

**Lithium-ion Battery SOH Estimation Based on
Noise-suppressed Semi-supervised Learning Method**
ZHANG Yutong¹, XIE Jun^{1*}, WANG Chunxin¹, XIE Qing¹, WANG Nan²

- (1. State Key Laboratory of Alternate Electrical Power System With Renewable Energy Sources
(North China Electric Power University), Changping District, Beijing 102206, China;
2. State Grid Integrated Energy Service Group Co., Ltd., Xicheng District, Beijing 100052, China)

ABSTRACT: Due to the high costs of long-term testing and discrepancies between real-world operating conditions and laboratory environments, labeled data for lithium-ion battery state-of-health (SOH) estimation remains scarce. To address this challenge, this study proposes a semi-supervised SOH estimation method based on self-convolutional neural network (CNN)-transformer with loss regularization (SCT-LR). First, health indicators are extracted through incremental capacity analysis, and a CNN-Transformer hybrid architecture is employed for establishing a mapping model between these indicators and SOH. Second, this paper adopts a self-training semi-supervised learning framework to generate pseudo-labels for unlabeled samples, enabling iterative annotation and utilization of unlabeled data. Finally, to mitigate the impact of label noise in pseudo-labels, this paper introduces a semi-supervised prediction method incorporating a label noise learning module with loss function regularization. The results demonstrate that the proposed method significantly improves SOH estimation accuracy by leveraging unlabeled data. Specifically, it achieves a 12.11% or greater improvement in estimation accuracy compared to supervised learning benchmarks, while the integration of the label noise learning module reduces the average error by 14.41%. This work provides a robust framework for accurate battery SOH prediction in real-world scenarios with limited labeled samples.

KEY WORDS: lithium-ion battery; state of health (SOH); semi-supervised learning; self-training; label noise learning

摘要: 由于长期测试成本高昂及实际工况与实验室条件差异, 导致锂离子电池健康状态(state-of-health, SOH)标签数据匮乏。针对此问题, 该文提出一种基于损失正则化自训练的锂离子电池 SOH 半监督估计方法。首先通过容量增量分析提取健康因子, 并基于卷积神经网络与 Transformer 混合架构构建其与 SOH 的映射模型。其次, 采用自训练半监督学习框架构建无标签样本的伪标签, 通过自我迭代实现无标签样本的标注与利用。最后, 为降低伪标签中标签噪声影响, 提出融合损失函数正则化的标签噪声学习模块的半监督预测方法。结果表明, 所提方法可利用无标签数据显著改善 SOH 估计结果, 尤其是与监督学习基准模型相比, 其估计精度提升 12.11%以上, 标签噪声学习模块的引入使得所提方法的平均误差降低 14.41%。研究结果将有效支撑标签样本匮乏的现场应用场景下的电池 SOH 精确预测工作。

关键词: 锂离子电池; 健康状态; 半监督学习; 自训练; 噪声学习

0 引言

锂离子电池因其能量密度高、循环寿命长、制造成本低等优点被广泛应用于储能电站、电动汽车、航空航天及电子产品等领域^[1-3]。但在长期使用过程中, 电池内部活性锂损失、固体电解质界面(solid electrolyte interphase, SEI)膜生长引发的电池容量衰退严重影响其安全可靠运行, 故需要对电池健康状态(state-of-health, SOH)进行准确评估^[4-6]。

基金项目: 国家重点研发计划项目(2023YFB2408200, 2023YFB2408203)。

National Key R&D Program of China (2023YFB2408200, 2023YFB2408203)。

电池 SOH 与老化机理密切相关, 受多重因素影响且很多参数难以在线测量, 导致其估算难度较高^[7-9]。早期研究工作主要通过电化学模型或等效电路模型来表征电池内部特性及容量退化规律^[10-11]。但是该方法对数据质量、初值选取的要求较高, 且在部署时泛化能力较差, 故逐渐被数据驱动方法取代^[12-14]。数据驱动方法基于机器学习或深度学习模型, 通过挖掘电池数据中的退化规律构建多维输入特征与 SOH 的映射关系。该方法无需深入解析电池老化机理, 但对数据质量及规模具有较强依赖性, 在工程应用中展现出显著优势与发展潜力^[15]。当前, 数据驱动方法主要通过自回归 (auto regressive, AR)、支持向量机 (support vector machine, SVM)、神经网络 (neural network, NN) 等模型预测电池 SOH^[16-18]。对于上述方法而言, 特征输入和模型的选择是影响 SOH 估计误差的关键。李超然等提出一种一维卷积神经网络 (convolutional neural network, CNN) 模型实现锂离子电池的容量估算, 建立电压、电流、温度数据与 SOH 的映射关系, 并于多个公开数据集中验证方法的 SOH 估计效果^[19]。然而电池容量衰减过程具有非单调特性且伴随容量再生现象, 现有方法主要聚焦于容量退化曲线的局部特征提取, 对长程依赖关系的建模能力仍显不足。因此, 李鹏等提出一种基于转置 Transformer 的电池 SOH 融合估计模型, 结果表明自注意力机制可有效提升模型在复杂电池数据集中的表现^[20]。

然而, 上述数据驱动方法均属于监督学习框架下的 SOH 估计, 并未摆脱对海量、高质量标签样本的依赖性, 不可避免地在训练阶段受限于电池容量退化实验所需要的物质成本与时间成本^[21]。由于实际场景中电池长期处于中间荷电状态区间 (20%~80%) 的非全循环充放电状态, 导致无法通过容量测试方法获取与完整充放周期严格对应的 SOH 标签数据^[22-23]。此时, 仅依赖实验室数据对深度学习模型进行训练, 易导致其过度拟合有限的标签样本, 进而影响模型在实际应用场景下的泛化性和鲁棒性^[24]。

为降低模型对标签样本的依赖性, Yarowsky 等提出一种基于自训练 (self-training) 的半监督学习方法并将其应用于自然语言处理任务, 通过生成伪标签样本并将其加入训练集迭代优化模型^[25]。Che 等将该方法迁移至锂电池 SOH 预测领域, 通过伪容

量标签重建策略实现小样本场景下容量衰减轨迹的预测^[26]。类似地, Xiong 等充分利用无标签样本并融入模型训练, 减少对标签样本的数据需求^[27]。此外, 部分研究也采用半监督学习方法挖掘无标签样本特征辅助训练, 通过整合少量标签样本构建输入和输出间映射关系, 并利用无标签样本挖掘数据潜在结构与分布特征以增强模型性能^[28-29]。然而, 上述方法的训练效果高度依赖模型初始化性能, 且易受伪标签噪声干扰, 导致实际部署效果不佳。

为此, 本文提出一种基于损失正则化的自训练 CNN-Transformer (self-CNN-transformer with loss regularization, SCT-LR) 的锂离子电池健康状态半监督估计方法, 通过容量增量分析提取健康因子并将其作为输入特征训练 CNN-Transformer 模型, 并通过自训练 CNN-Transformer (self-CNN-transformer, SCT) 方法为无标签样本生成伪标签以扩充训练集。在伪标签生成过程中, 使用基于标签噪声学习的伪标签置信度优化方法, 利用损失函数正则化技术防止模型过度拟合伪标签噪声, 有效提升模型对伪标签的容忍度以及分布外数据的泛化能力。从而实现标签样本匮乏场景下的锂离子电池 SOH 高精度在线预测。

1 健康因子与特征提取网络

1.1 健康因子提取

锂离子电池在充放电循环的电压平台期内的变化趋势能够较为准确地反映电池容量退化及活性锂损失程度。但电池充放电曲线的平台期内电压变化较为细微, 故需采用容量增量分析 (incremental capacity analysis, ICA) 对其进行处理, 将微小的电压波动值转换为容量增量 (incremental capacity, IC) 曲线内更易识别的波峰。

如图 1 所示, 随着电池循环次数的增加, IC 曲线主峰呈现出向右下方偏移的趋势, 直观反映容量损失。但 IC 曲线由大量高维离散点组成, 难以直接用于预测模型建模。故通常由 IC 曲线中提取具有代表性的健康因子, 将其作为输入特征用于构建 SOH 预测模型。研究表明^[30-31], IC 曲线主峰峰值 f_{peak} 、峰值电压 v_{peak} 以及主峰峰下面积 S 可有效捕捉电池老化过程中的关键变化, 可作为表征 SOH 的健康因子。对于特定电压窗口 $[V_{\text{smin}}, V_{\text{smax}}]$, (V_i, G_i) 为 IC 曲线任意一点, 则主峰峰值 f_{peak} 与峰值电压 v_{peak} 可表示为:

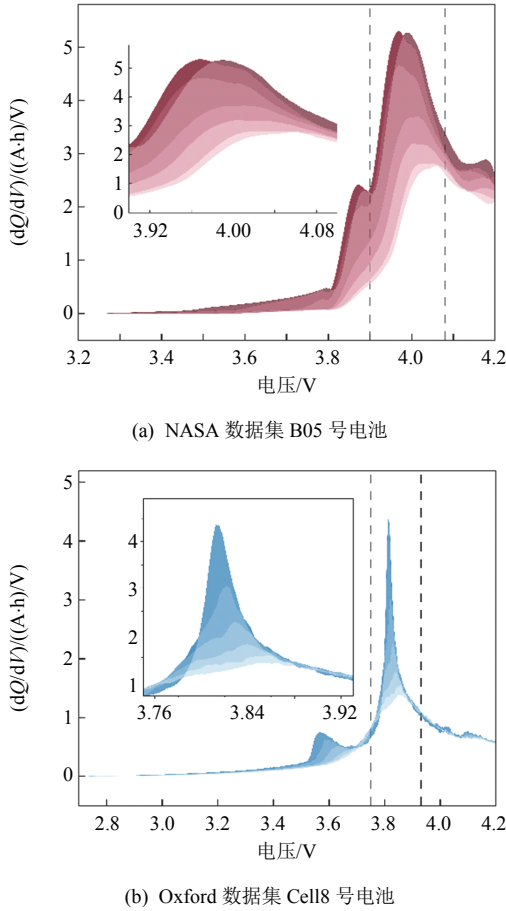


图1 不同循环次数下的容量增量分析曲线

Fig. 1 Incremental capacity analysis curves under different cycle numbers

$$f_{\text{peak}} = \max \{G_i | V_i \in [V_{\text{min}}, V_{\text{max}}]\} \quad (1)$$

$$v_{\text{peak}} = \underset{v \in [V_{\text{min}}, V_{\text{max}}]}{\operatorname{argmax}} f_{\text{peak}} \quad (2)$$

若 (V_j, G_j) 为该特定电压窗口 $[V_{\text{min}}, V_{\text{max}}]$ 第 j 个采样点, 则主峰峰下面积 S 可表示为

$$S = \sum_{j=1}^{m-1} \frac{(V_{j+1} - V_j)(G_{j+1} + G_j)}{2} \quad (3)$$

式中 m 为该特征窗口内采样点总数。

为进一步定量评估所提健康因子与电池 SOH 的相关性, 采用皮尔逊相关系数(pearson correlation coefficient, PCC)进行分析。设电池老化循环总数为 n , 某一健康因子的特征可表示为 $\mathbf{h}=[h_1, h_2, \dots, h_n]$, 与之对应的电池 SOH 值为 $\mathbf{y}=[y_1, y_2, \dots, y_n]$, 则皮尔逊相关系数 P 可表示为

$$P = \frac{\sum_{i=1}^n (h_i - \bar{h})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (h_i - \bar{h})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (4)$$

式中: h_i 为第 i 次循环时该健康因子的观测值; $\bar{h}$ 为该健康因子的样本均值; y_i 为第 i 次循环时对应的电池 SOH 值; $\bar{y}$ 为该电池 SOH 的样本均值。

分析结果如图2所示。以 Oxford 数据集的 Cell8 为例, 其 IC 曲线峰值、峰值电压以及主峰峰下面积与 SOH 的相关系数分别为 0.960 6、-0.956 8 和 0.995 5, 表明, 所选健康因子可以准确跟踪电池 SOH 的变化趋势。

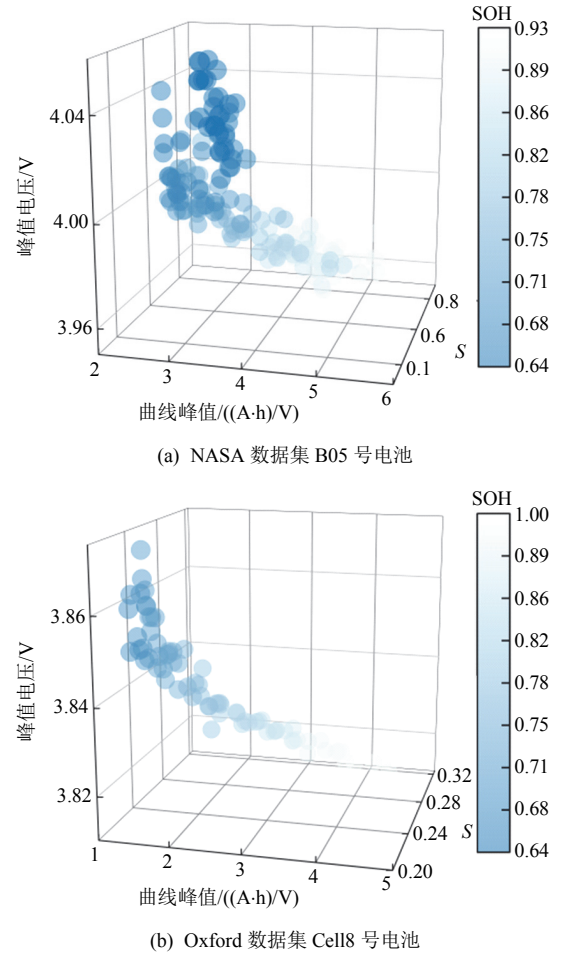


图2 电池健康因子与 SOH 相关性分析

Fig. 2 Correlation analysis between battery health indicators and SOH

1.2 CNN-Transformer 模型

锂离子电池 SOH 估计的核心在于通过分析历史观测数据实现容量轨迹的连续时间预测, 同时建立跨时间尺度的数据关联规律。当前数据驱动方法中, 传统循环神经网络的时序递归架构、卷积神经网络的空间卷积架构, 以及基于 Transformer 的多头注意力堆叠编码结构, 均面临局部突变特征捕获与全局退化趋势建模的协同优化难题。而电池容量退化数据通常呈现非单调衰减趋势并伴随容量再生现象, 对模型的局部动态响应与长期依赖关系的

双重建模能力提出更高的要求。

本文提出的 CNN-Transformer 混合架构通过分层特征融合机制实现退化特征的多尺度建模。设 $\mathbf{X} \in \mathbf{R}^{P \times D}$ 表示健康因子输入张量，其中 P 为时间步长， D 为特征维度。首先，利用一维卷积层从健康因子时序张量中解耦提取局部退化特征，学习容量再生现象衍生的非平稳波动特征。其数学表征为

$$\mathbf{C} = \sigma(\mathbf{X} * \mathbf{W}_c + \mathbf{b}_c) \quad (5)$$

式中： $\mathbf{C}$ 为卷积输出的特征图； $\mathbf{W}_c$ 为卷积核权重矩阵； $\mathbf{b}_c$ 为偏置向量； $\sigma(\cdot)$ 为 RELU 激活函数，用于增强对非平稳波动特征的学习能力。随后，使用 Transformer 模块的多头自注意力编码机制对卷积输出的高维特征嵌入进行全局依赖关系重构，实现退化趋势模式与容量再生扰动特征的协同表征，进而精准预测 SOH 变化趋势。多头自注意力机制操作的数学表征为：

$$\mathbf{Z} = \text{LayerNorm}(\mathbf{C} + \text{MHA}(\mathbf{C})) \quad (6)$$

$$\text{MHA}(\mathbf{C}) = \text{Concat}(\text{head}_1, \dots, \text{head}_h) \mathbf{W}^O \quad (7)$$

式中： h 为注意力头的数量，每个头均独立计算注意力权重； $\mathbf{W}^O$ 为权重矩阵； $\mathbf{Z}$ 为自注意力机制输出的特征嵌入。

CNN-Transformer 结构图如图 3 所示。

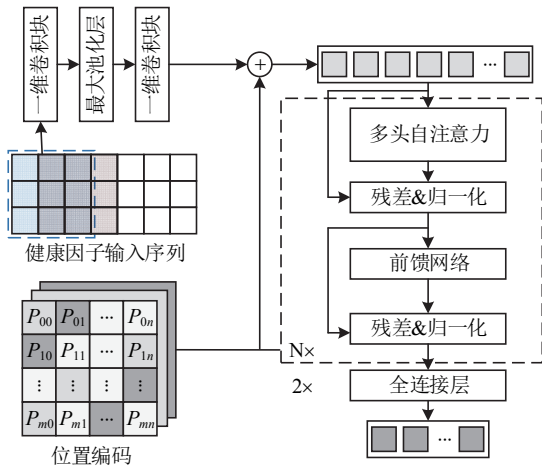


图 3 CNN-Transformer 网络结构

Fig. 3 The proposed CNN-Transformer framework

2 基于 SCT-LR 的电池 SOH 半监督估计方法

2.1 基于半监督学习的 SCT 模型

在锂离子电池健康状态估计研究中，模型不仅需要具备优异的时序特征提取能力，还对标签样本的数量与质量提出更高要求。然而电池 SOH 标签

数据的获取通常依赖耗时的长期退化测试，这些测试过程耗时且成本高昂，导致高质量带标签的退化数据集极其稀缺。

半监督学习通过有效利用标签样本可显著提升模型预测性能，鉴于 self-training 方法与 CNN-Transformer 模型在解决标签样本不足问题上具有互补优势，本文提出一种新的 SCT 模型。其中 self-training 方法的训练流程如图 4 所示。

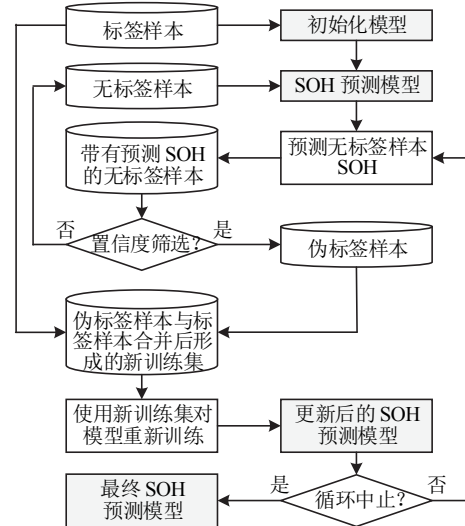


图 4 self-training 方法训练流程图

Fig. 4 Flow chart of self-training

得益于 CNN-Transformer 模型在时间序列数据的局部特征提取和长程依赖建模方面的优势，SCT 模型在训练过程中生成的伪标签具有较高可靠性，从而显著降低误差传播风险。首先基于少量标签样本训练基础时序预测模型得到初始化模型，使其捕获容量退化的基本规律。初始化训练中，标签样本与测试样本的分布差异越小，模型训练效果越优；且保留无标签样本中的分布差异个体可提升模型泛化性能。随后利用该模型预测无标签样本，筛选高置信度结果生成伪标签。通过迭代将高置信度伪标签样本扩充至训练集，增强 SOH 预测模型从无标签样本中挖掘电池老化规律的能力。当无标签样本的预测差异趋于稳定时终止迭代。

在半监督回归框架下，SCT 模型优化的核心挑战在于构建高置信度样本筛选机制。由于现有半监督学习研究多聚焦分类任务，针对回归问题的理论体系与方法验证仍存在明显不足。因此本文以 Zhou 等在 COREG 框架中提出的 k 近邻策略为基础，将其在评估无标签时序数据置信度方面展现出的独特优势迁移至 SCT 模型的训练过程中^[32]。通过构

建基于时序相似度的邻域评估模块,有效提升伪标签筛选精度。设任意一无标签样本及其预测值为 $(x_u, \hat{y}_u)$, 则 x_u 的 k 近邻样本预测值向量为 $\hat{y} = [\hat{y}_1, \hat{y}_2, \dots, \hat{y}_k]$ 。进而可通过计算该无标签样本在其 k 近邻中的预测值的均方误差来获取置信度评价, 其计算公式如下:

$$\Delta = \frac{1}{k} \sum_{m=1}^k (\hat{y}_m - \hat{y}_u)^2 \quad (8)$$

式中: $\hat{y}_m$ 为第 m 个 k 近邻标签值; k 为 k 近邻标签总数。通过选择置信度最高的无标签样本(已赋予伪标签), 并将其重新加入训练集来更新 SCT 模型。

2.2 基于损失正则化的噪声学习模块

尽管在每次扩展训练集时都会根据置信度评估结果选择预测值最稳定的样本进行更新, 但无标签样本的预测伪标签仍然不如真实标签样本的标签更为可靠, 训练过程中产生的标签噪声可能会导致过拟合噪声数据并形成错误的反馈回路, 严重劣化模型性能。因此, 为抑制 self-training 训练过程中由于标签噪声引起的性能损失, 本文提出一种基于损失正则化的标签噪声学习方法, 通过改造模型的损失函数来改善上述问题。该方法在原有半监督损失函数的基础上添加一个 L_2 正则项, 通过该正则项惩罚模型权重的平方限制权重值的大小, 且此过程中模型权重的变化更为平滑, 使得模型对不同特征的输入更为鲁棒。损失函数表达式如下:

$$\ell = \frac{1}{M} \sum_{i=1}^M [(1-f_i) + \alpha_i f_i] \cdot \ell(y_i, \hat{y}_i) + \lambda \|w\|_2^2 \quad (9)$$

式中: M 为标签样本和无标签样本数量的总和; f_i 为指示样本 i 是否为伪标签样本的标志; α_i 用于控制伪标签样本损失权重; $\ell(y_i, \hat{y}_i)$ 为样本 i 的均方误差损失; $\lambda \|w\|_2^2$ 为 L_2 正则项, 其中 λ 用来控制正则化强度。

在模型训练初期, α_i 较小, 此时 L_2 正则项对模型的限制有助于模型更好地学习标签样本特征, 避免过度拟合初始伪标签中的标签噪声。随着训练进行, α_i 逐渐增大, 伪标签数据的影响增大, 此时 L_2 正则化继续平滑地约束模型权重, 从而使模型能够有效利用伪标签数据, 同时避免对错误标签的过度拟合。综上所述, α_i 随训练 epoch 的比例线性增长, 其变化策略的数学表征为

$$\alpha_t = \alpha_s + (\alpha_e - \alpha_s) \frac{t}{T} \quad (10)$$

式中: t 为当前 epoch 索引值; T 为训练总 epoch 数; α_s 、 α_e 分别为初始值和最终值。

2.3 基于 SCT-LR 的电池 SOH 估计框架

本文提取电池充电 IC 曲线的曲线峰值、峰值电压以及主峰峰下面积作为特征参量, 经 1D-CNN 提取容量退化的局部特征后, 利用 Transformer 模型进行退化轨迹的长程依赖建模。在训练过程中, 基于自学习方法建立无标签样本的利用机制, 筛选高置信度伪标签样本并将其逐步扩充至训练集。同时, 利用损失正则化方法对损失函数进行约束, 进一步抑制伪标签生成过程中标签噪声对模型的影响, 建立用于电池 SOH 估计的 SCT-LR 模型。整体算法流程图 5 所示。

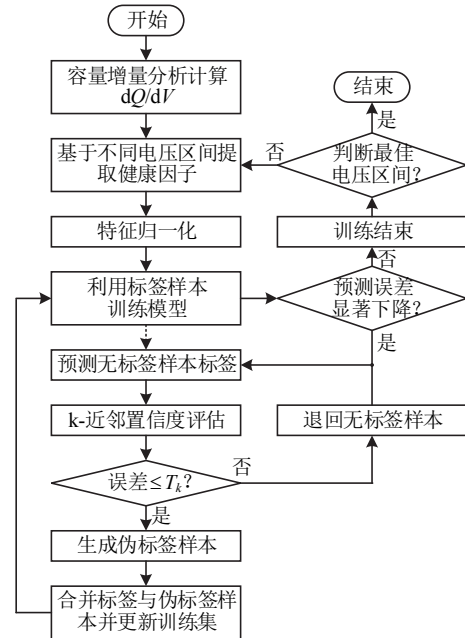


图 5 SCT-LR 模型训练流程图

Fig. 5 Flow chart of SCT-LR model training

3 实验结果与分析

3.1 实验数据及分组设置

本文使用的算力平台主要参数为: i7-14700K CPU、RTX4080-16G GPU, 算法基于 python 语言编写, 使用的深度学习框架为 pytorch。

在研究中, 本文选择主流退化数据中具有代表性的两个样本, 分别是 Oxford 数据集和 NASA 数据集, 表 1 展示两个数据集的细节, 包括电池型号、环境温度、充放电方式以及循环终止条件等信息。

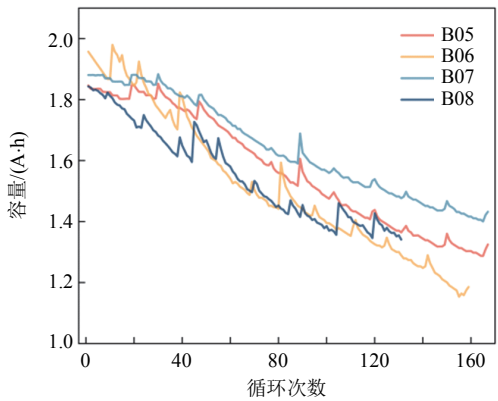
其中, Oxford 数据集包含 8 组标称容量 740 mA·h 的 Kokam 钴酸锂电池编号(Cell1—Cell8), 实验采用每 100 次循环执行 1C 恒流充放电循环以

表 1 电池容量退化数据集参数

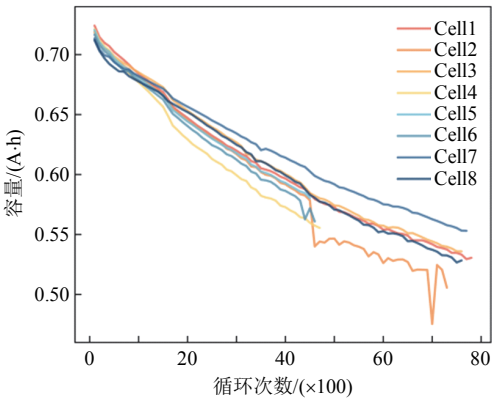
Table 1 Parameters of battery capacity degradation datasets

参数	Oxford	NASA
电池数量	8	4
实验温度/℃	40	24
额定容量/(mA·h)	740	2 000
电压区间/V	2.7~4.2	2.7/2.5/2.2/2.5~4.2
充电模式	恒流(2C)	恒流-恒压(0.75C)
放电模式(1C)	Artemis urban drive cycle	恒流

实现容量标定,当容量衰减至额定容量的 70%时终止实验。NASA 数据集选用 LG Chem 18650 型锂离子电池(标称容量 2.0 A·h),选取 B0005、B0006、B0007 和 B0018 4 组同工况电池(记为 B05—B18)作为分析对象。充电阶段以 1.5 A 恒流充电至 4.2 V 后转为恒压模式直至电流降至 20 mA;放电过程以 2 A 恒流放电至电压阈值 2.7 V 截止。与 Oxford 数据集不同,图 6 中 NASA 电池数据的容量退化曲线可以明显观察到容量再生现象,该特征将在健康状



(a) NASA 数据集



(b) Oxford 数据集

图 6 不同数据集的电池容量退化曲线

Fig. 6 Battery capacity degradation curves across different datasets

态估计时影响模型的预测精度。

为模拟实际应用中标签数据稀缺场景,本研究对 NASA 和 Oxford 数据集分别按照不同比例进行分组与划分。其中标签样本对应实验室环境下完整记录的电池全寿命周期充电电压、电流数据及 SOH 参数,其获取难度更高、耗时更长,但却是数据驱动方法中最为理想的训练数据。相比之下,实际监测数据通常仅包含充电电压与电流时序信息而缺失 SOH 标签,这类数据需通过伪标签标注机制处理后才能用于模型训练。具体的数据集设置如表 2 所示。与实际应用场景不同的是,实验设计中存在标签样本占比高于无标签样本的组别设置,但该分组本质是受限于数据集样本量的固有局限。该设置的目的是在样本总量受限条件下,重点分析标签样本比例与模型预测误差的关联规律。

表 2 电池容量退化数据集设置与分组

Table 2 Configuration and partitioning of battery capacity degradation dataset

数据集	组别	标签样本	无标签样本	测试数据集
NASA 数据集	1	B05	B06, B07	B18
	2	B18	B05, B06	B07
	3	B06	B07, B18	B05
Oxford 数据集	1	Cell1	Cell2, Cell3, Cell4, Cell5, Cell6, Cell7	Cell8
	2	Cell1, Cell3	Cell2, Cell4, Cell5, Cell6, Cell7	Cell8
	3	Cell1, Cell3, Cell4	Cell2, Cell5, Cell6, Cell7	Cell8
	4	Cell1, Cell3, Cell4, Cell5	Cell2, Cell6, Cell7	Cell8
	5	Cell1, Cell3, Cell4, Cell5, Cell6	Cell2, Cell7	Cell8
	6	Cell1, Cell3, Cell4, Cell5, Cell6, Cell7	Cell2	Cell8

3.2 评价指标

采用均方根误差(root mean square error, RMSE)、平均绝对误差(mean absolute error, MAE)

和决定系数 R^2 作为模型性能的评价指标,并将其分别记为 E_{RMSE} 、 E_{MAE} 和 E_R 。RMSE 是残差的标准偏差,其对数据中的异常预测值更为敏感,MAE

表示观测值与预测值之间绝对误差的平均值，主要用来评价模型的平均误差大小。 R^2 用于量化模型预测值对观测值的拟合效果，取值范围为 $[0, 1]$ ，且越接近 1 表示模型拟合效果越好。 E_{RMSE} 、 E_{MAE} 与 E_R 的计算公式分别如下：

$$E_{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}$$

(11)

$$E_{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|$$

(12)

$$E_R = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

(13)

式中： y_i 为实际观测值； $\bar{y}$ 为观测值均值； $\hat{y}_i$ 为模型的预测值； n 为样本数量。

3.3 SCT-LR 模型的 SOH 估计效果

由于 NASA 数据集相较于 Oxford 数据集具有更加明显的容量再生现象，对模型的时序建模能力以及泛化能力要求更高，因此将其作为 SCT-LR 模型的性能评估基准。而 NASA 数据集内 B05、B06、B07 与 B18 电池虽处于相同工况，但因电化学阻抗谱测试频率不同，使得 B18 电池的容量退化轨迹呈现显著差异特征。故实验分组将 B18 电池作为重点验证对象，分析特征分布偏移数据分别作为标签样

本和无标签样本场景下对模型预测误差的影响。为验证所提方法的优越性，本文设置多组对比实验。包括针对 SCT-LR 模型的消融实验对照组，即不包含 1D-CNN 模块的基于损失正则化的自 Transformer(self-transformer with loss regularization, ST-LR)模型以及不包含噪声学习模块的自 Transformer(self-transformer, ST)模型。此外，选取典型时序预测深度学习模型，包括长短期记忆神经网络(long short-term memory, LSTM)和门控循环网络(gated recurrent unit, GRU)作为监督学习对照组，以论证半监督学习框架在数据受限场景下的 SOH 预测性能优势。值得注意的是，监督学习对照组所用训练数据均为表 2 中各数据分组的标签样本，这也与监督学习、半监督学习在相应实际部署条件下能够获取的数据水平相一致。在相同量化环境下，上述模型的推理时间和参数数量如表 3 所示。其 SOH 估计的详细结果如表 4 以及图 7 所示。

表 3 模型推理时间与参数数量

Table 3 Model inference time and parameter count

模型	推理时间/ms	参数数量/($\times 10^6$)
SCT-LR	13.1	1.46
ST-LR	9.3	1.45
ST	9.0	1.45
LSTM	5.5	1.85
GRU	5.8	1.38

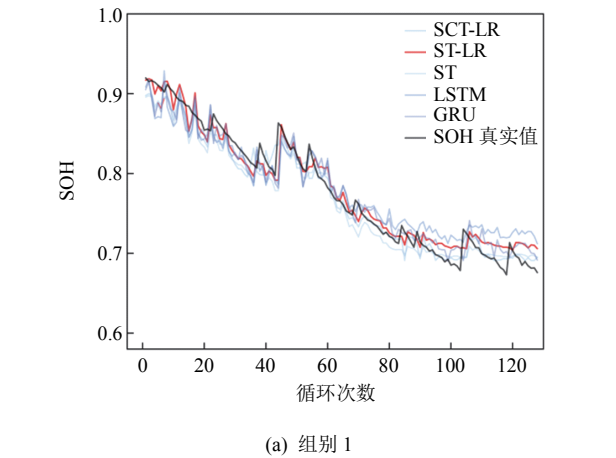
表 4 NASA 数据集的 SOH 估计结果

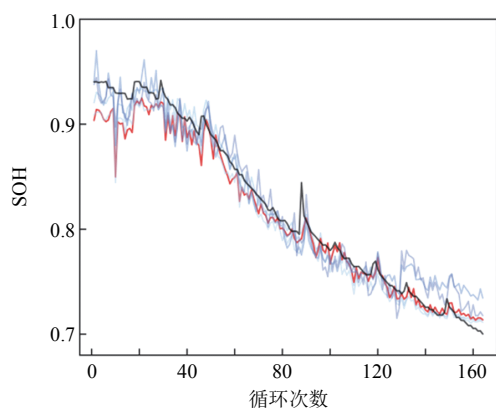
Table 4 SOH estimation results of the NASA dataset

组别	RMSE/%					MAE/%					R^2				
	SCT-LR	ST-LR	ST	LSTM	GRU	SCT-LR	ST-LR	ST	LSTM	GRU	SCT-LR	ST-LR	ST	LSTM	GRU
Grp.1	1.479	1.774	1.947	3.165	2.263	1.220	1.392	1.492	2.363	1.636	0.961 4	0.944 5	0.933 1	0.823 3	0.909 6
Grp.2	1.156	1.639	1.728	2.599	2.323	0.798	1.160	1.309	2.149	1.903	0.978 8	0.957 4	0.952 7	0.892 9	0.914 4
Grp.3	0.916	1.702	2.300	3.006	2.408	1.259	1.259	1.608	2.674	2.125	0.990 5	0.967 3	0.940 4	0.897 9	0.934 5

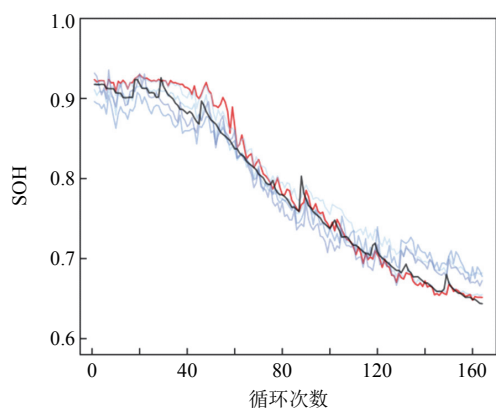
根据表 4 可知，本文所提方法在 NASA 数据集上取得最佳估计精度。其在 3 个数据分组中的平均 RMSE 和 MAE 仅为 1.184%和 1.092%， R^2 值为 0.9769。其估计精度较基于监督学习方法的基准模型提升 12.11%以上。由图 7 可知，SCT-LR 模型的预测值较其他对比方法更为接近实际的电池容量退化轨迹。尤其是在 Grp.3 分组中，无标签样本中包含特征分布偏移数据，使得模型在自训练过程中学习到更鲁棒的特征表示，其 RMSE 和 MAE 较监督学习基准模型中估计误差较小的 GRU 降低 61.96%和 50.75%。综上所述，在数据规模受限场景下，半监督学习方法具备更加精确的 SOH 估计

效果，且特征分布偏移数据作为无标签样本时对模





(b) 组别 2



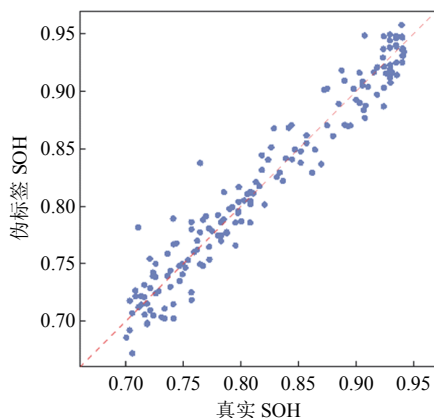
(c) 组别 3

图 7 NASA 数据集电池 SOH 估计结果

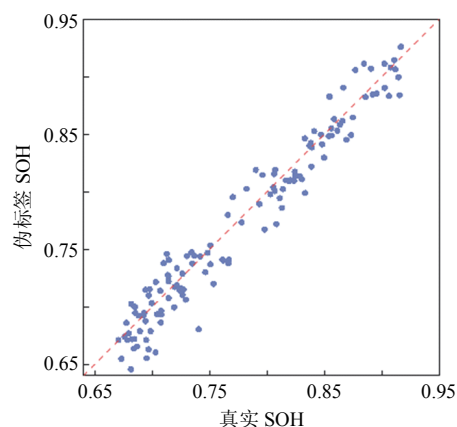
Fig. 7 SOH estimation results for batteries in the NASA dataset

型性能影响最小。

为评估 SCT-LR 模型的伪标签质量,针对 Grp.3 数据组训练过程中无标签数据集生成的伪标签,开展其与真实 SOH 值的差异分析,结果示于图 8。在此过程中,分析使用的 SOH 标签仅用于后期评估,其在模型的训练阶段均未被使用以避免信息泄露问题。在模型收敛后,伪标签 SOH 与真实 SOH 之间的平均 RMSE、MAE 和 R^2 分别为 0.933%、1.281%



(a) B07



(b) B18

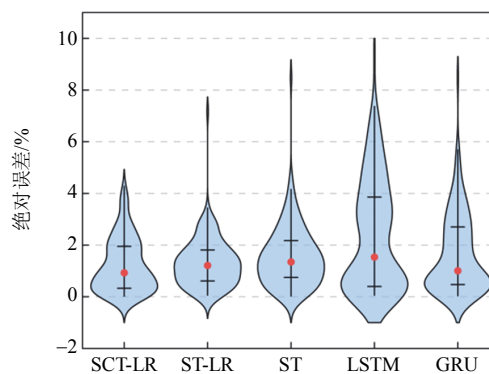
图 8 伪标签 SOH 与真实 SOH 估计结果

Fig. 8 Pseudo-label SOH and true SOH estimation results

和 0.9890。由图 8 亦可知,伪标签 SOH 与真实 SOH 分布在对角线附近,表明大部分伪标签具有较好的准确性,但其分布呈现轻微右偏态,即模型可能倾向于低估电池 SOH。

此外,将 ST-LR 模型与 ST 模型的预测结果与本文所提方法进行对比,以验证 1D-CNN 模块和噪声学习模块的有效性。结果表明,引入 1D-CNN 和噪声学习模块前后,相同模型的表现差异巨大。其中,加入噪声学习模块后,ST-LR 模型于 3 个数据组中的平均 RMSE 和 MAE 分别降低 14.41%和 12.41%。而所提方法在同时加入 1D-CNN 模块和噪声学习模块后,3 个数据组中的平均 RMSE 和 MAE 降低 40.56%和 24.69%, R^2 均值提升 3.69%。

图 9 是各方法在不同数据组中的绝对误差分布图,与 ST-LR 模型和 ST 模型相比,所提方法实现最低的平均误差以及最小的误差分布范围,且在异常值控制方面表现优异。综上所述,所提方法在电池 SOH 估计领域表现优异。通过整合有限标签样本和包含特征分布偏移的无标签样本,其不仅保证模型对典型退化模式的捕捉精度,还进一步通过伪标



(a) 组别 1

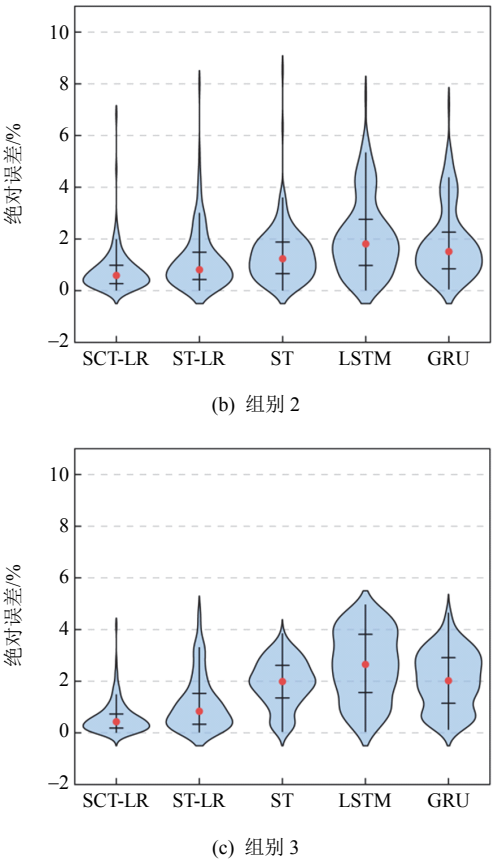


图 9 NASA 数据集电池 SOH 估计绝对误差小提琴图
Fig. 9 SOH estimation absolute error violin plots for NASA dataset

表 5 Oxford 数据集的 SOH 估计结果
Table 5 SOH estimation results of the Oxford dataset

组别	RMSE/%					MAE/%					R ²				
	SCT-LR	ST-LR	ST	LSTM	GRU	SCT-LR	ST-LR	ST	LSTM	GRU	SCT-LR	ST-LR	ST	LSTM	GRU
1	1.712	1.631	1.715	2.013	1.955	1.531	1.341	1.505	1.743	1.702	0.940 0	0.945 6	0.939 8	0.917 1	0.921 8
2	1.176	1.293	1.441	1.620	1.298	0.860	1.062	1.115	1.149	1.020	0.971 7	0.965 8	0.957 5	0.946 3	0.965 5
3	0.354	0.907	1.294	1.386	1.321	0.257	0.801	0.994	1.047	1.078	0.997 4	0.983 2	0.965 7	0.960 7	0.964 3
4	1.025	1.302	1.529	1.562	1.607	0.906	0.997	1.206	1.223	1.362	0.978 5	0.965 3	0.952 2	0.950 1	0.947 2
5	1.365	1.506	1.556	1.829	1.794	1.145	1.236	1.344	1.374	1.504	0.961 9	0.953 8	0.950 5	0.931 6	0.934 2
6	0.724	1.126	1.168	1.101	0.865	0.529	0.872	0.947	0.922	0.640	0.989 3	0.974 1	0.972 1	0.975 2	0.984 7

块对容量退化轨迹中存在的容量再生现象的辨识效果。

然而，所提方法在标签样本数量极度不平衡情况(Grp.1)下表现略差。此时，标签样本匮乏导致模型无法学习到有效特征和模式，进而在早期阶段过度依赖无标签样本的伪标签。但所提方法在 6 个数据组中的估计误差仍保持在较低水平，其平均 RMSE 和 MAE 分别为 1.059%和 0.871 3%。

综上所述，半监督模型的 SOH 估计精度与标签样本比例未呈现严格正相关性，但其在非极端标签匮乏场景下仍保持较高预测性能。而所提方法在

签迭代优化增强模型对差异分布特征的泛化能力。

3.4 标签样本数量的影响

对于半监督学习，标签样本在数据集中所占比例对模型估计性能具有重要影响。首先，根据 3.3 节中对 NASA 数据集的分析结果，选择 Oxford 数据集中容量退化趋势最为异常的 Cell2 作为分析重点，并在各数据组中均将其划分至无标签样本以减少其对模型性能的影响。同时，在数据组 1—6 中逐组增加标签样本比例，以观察各模型在标签样本比例不同情况下的性能表现。其中，监督学习模型依旧仅使用相应数据分组中的标签样本进行训练。结果如表 5 所示。

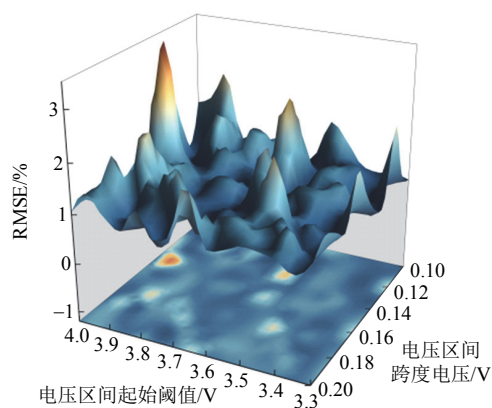
根据表 5 可知，所提方法在数据组 2—6 中均取得较为理想的 SOH 估计精度。尤其在数据组 3 中，所提方法的估计精度显著高于其他基准模型，其 RMSE 与 MAE 相较于监督学习基准模型降低 33.19%和 29.90%以上，R² 值则提升 2.78%以上，表明所提方法在标签样本具有一定数量基础的情况下，可通过组合无标签样本的方式改善 SOH 估计效果。由于 Oxford 数据集的容量退化曲线未呈现显著容量再生现象，因此其估计精度提升幅度低于 NASA 数据集。与之类似，ST-LR 模型在相同工况下也保持较高的估计精度，进一步证明 1D-CNN 模

标签样本匮乏场景下通过有效整合无标签样本数据特征，显著减低 SOH 估计误差，展现出良好的工程应用潜力。

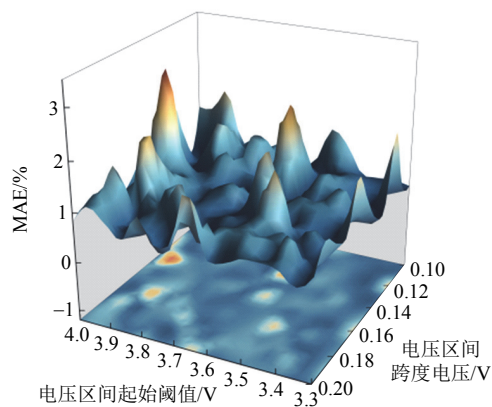
3.5 电压区间对 SOH 估计结果的影响

IC 曲线在不同电压区间内提取得到的健康特征存在差异，且该差异直接影响电池 SOH 估计结果。为进一步分析电压区间起始阈值与跨度范围对 SOH 估计的具体影响，选取 Oxford 数据集的 Grp.3 数据组进行测试，其误差分析和容量增量分析结果如图 10、11 所示，将二者进行对比可发现存在明显关联规律。尽管本节仅针对所提方法在 Grp.3 中

的 SOH 估计结果进行分析,但其流程与结果均可推广至其他数据集中。



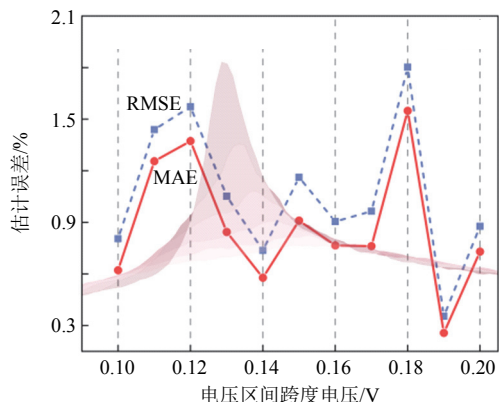
(a) 起始阈值与跨度电压对 RMSE 的影响



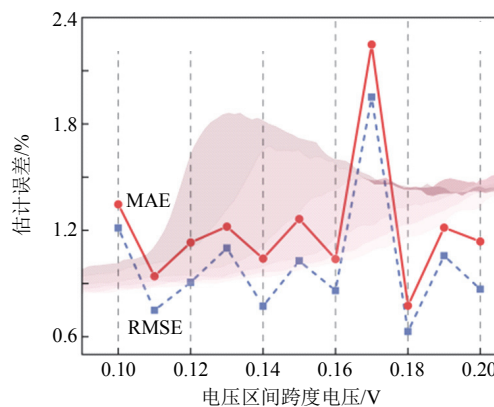
(b) 起始阈值与跨度电压对 MAE 的影响

图 10 电压区间对电池 SOH 估计结果的影响
Fig. 10 Impact of voltage range on battery SOH estimation

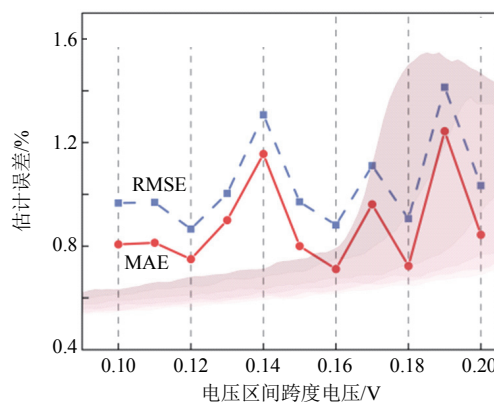
对 Grp.3 数据组的样本,在 3.3~4.2 V 电压区间内采用 1.2 节所述方法进行健康因子提取,其中起始电压设定为 3.3~4.0 V。并在固定跨度范围研究基础上增设[0.1 V,0.2 V]间隔参数,系统分析电压跨度范围对电池 SOH 估计结果的影响。所提方法在 Grp.3 数据组中采用不同电压区间提取健康特征时



(a) 电压区间起始阈值: 3.75 V



(b) 电压区间起始阈值: 3.50 V



(c) 电压区间起始阈值: 3.40 V

图 11 健康因子特征形态对电池 SOH 估计结果的影响
Fig. 11 Impact of health indicator feature morphology on battery SOH estimation

的 SOH 估计误差分布如图 10 所示,其中黄色区域表征全局 RMSE、MAE 较高,表明估计性能较弱;蓝色区域则对应较低的误差指标,反映更加优异的估计效果。显然,蓝色区域与图 1(b)中的 Cell8 容量增量曲线主峰区域高度重合。同时,可明显观察到, [3.55 V,3.65 V]伪峰区域的估计误差虽出现下降趋势,但其降幅显著低于主峰区域。这表明不同特征区域对 SOH 估计精度的贡献度存在显著差异,验证特征选择策略的必要性。SOH 预测精度与电压区间参数间并非为简单的线性关联关系,需结合 IC 曲线进行具体分析。如图 11 所示,通过对图 10 进行切片分析可知,起始电压分别为 3.75、3.50 和 3.40 V 时,不同跨度范围对 SOH 估计误差呈现显著差异性影响。在图 11(a)中,随着电压跨度范围增大,区间内主峰特征逐渐丰富。当选取[3.75 V,3.89 V]电压区间时, RMSE 与 MAE 分别为 0.737%和 0.578%,此时所选区间已完全覆盖主峰特征区域。但当电压跨度继续增至 0.18 V 时, RMSE 和 MAE 分别降至 0.354%和 0.257%,此时误差指标方达到

全局最低值,该现象与固定跨度范围的研究结论存在差异。即当健康特征选择区间位于 IC 曲线主峰区域时,最佳采样区间需在完整覆盖主峰特征的基础上适当延拓,此时 SOH 估计性能方达最优。该结论同样可在图 11(b)的伪峰区间内得到证明,其最低估计误差出现在[3.50 V, 3.68 V]区间。另一方面,如图 11(c)所示,当电压区间未完全覆盖波峰特征时,健康因子与 SOH 相关性减弱导致估计误差显著增加。这表明起始电压宜选取于主峰临近区域以获取更精确稳定的 SOH 估计结果。综上,通过 IC 分析与电压区间分析选取最佳采样窗口,可为工程实践中提升 SOH 估计精度提供关键指导。

4 结论

针对标签样本匮乏场景下电池 SOH 估计精度低的问题,本文提出一种基于半监督学习的 SCT-LR 方法,以充分利用现场无标签样本改善 SOH 估计效果。主要结论如下:

1) 所提方法于 NASA 数据集进行 SOH 预测时取得最佳估计精度,其平均 RMSE 和 MAE 仅为 1.184%和 1.092%, R^2 值为 0.9769,综合性能优于 LSTM 等监督学习模型,具备工程应用价值。基于标签噪声学习的伪标签置信度优化方法利用损失函数正则化技术防止模型过度拟合伪标签噪声。模型加入噪声学习模块后,其在 3 个数据组中的平均 RMSE 和 MAE 分别降低 14.41%和 12.41%。

2) 利用 Oxford 数据集分析数据集内标签样本比例对模型性能的影响,结果表明,所提方法在 6 个数据组中的估计误差均保持在较低水平,其平均 RMSE 和 MAE 分别为 1.059%和 0.8713%。且实际应用情况下模型性能与标签样本数量并未呈现严格正相关性,一定数量无标签样本所包含的电池退化模式与特征可显著改善 SOH 估计精度。

3) 健康因子提取的电压区间对 SOH 估计结果具有重要影响,所提方法以 Oxford 数据集为例,验证最佳采样区间需在完整覆盖主峰特征的基础上适当延拓,方能取得最佳 SOH 估计精度,为工程实践中确定采样窗口提供关键指导。

参考文献

[1] 樊亚翔,肖飞,许杰,等. 基于充电电压片段和核岭回归的锂离子电池 SOH 估计[J]. 中国电机工程学报, 2021, 41(16): 5661-5669.

FAN Yaxiang, XIAO Fei, XU Jie, et al. State of health estimation of lithium-ion batteries based on the partial charging voltage segment and kernel ridge regression[J]. Proceedings of the CSEE, 2021, 41(16): 5661-5669(in Chinese).

[2] ZHANG Kaixuan, CHEN Cheng, ER Lixin, et al. Robust state-of-charge estimation for LiFePO₄ Lithium-ion batteries with pronounced voltage plateau regions[J]. Applied Energy, 2025, 401: 126755.

[3] DEMIRCI O, TASKIN S, SCHALTZ E, et al. Review of battery state estimation methods for electric vehicles-Part II: SOH estimation[J]. Journal of Energy Storage, 2024, 96: 112703.

[4] PENG Simin, ZHU Junchao, WU Tiezhou, et al. SOH early prediction of lithium-ion batteries based on voltage interval selection and features fusion[J]. Energy, 2024, 308: 132993.

[5] 刘芳,刘新慧,苏卫星,等. 基于全新电热耦合模型的锂电池关键状态在线联合估计方法[J]. 中国电机工程学报, 2024, 44(S1): 202-214.

LIU Fang, LIU Xinhui, SU Weixing, et al. Online joint estimation method for key states of lithium battery based on a new electro-thermal coupling model[J]. Proceedings of the CSEE, 2024, 44(S1): 202-214(in Chinese).

[6] 王睿茜,周星,王羽,等. 基于微分曲线的锂离子电池老化诊断研究综述[J]. 中国电机工程学报, 2024, 44(22): 8920-8935.

WANG Ruixi, ZHOU Xing, WANG Yu, et al. A review of degradation diagnosis of lithium-ion batteries based on differential curves[J]. Proceedings of the CSEE, 2024, 44(22): 8920-8935(in Chinese).

[7] FU Juncheng, WU Chunling, WANG Jiwei, et al. Lithium-ion battery SOH prediction based on VMD-PE and improved DBO optimized temporal convolutional network model[J]. Journal of Energy Storage, 2024, 87: 111392.

[8] YI Zhenxiao, WANG Shi, LI Zhaoting, et al. A novel approach for state of health estimation and remaining useful life prediction of supercapacitors using an improved honey badger algorithm assisted hybrid neural network[J]. Protection and Control of Modern Power Systems, 2024, 9(6): 1-18.

[9] 颜宁,李吉洋,李相俊,等. 考虑电热耦合特性的电池模组多状态协同估计方法研究[J]. 中国电机工程学报, 2025, 45(16): 6340-6353.

- YAN Ning, LI Jixiang, LI Xiangjun, et al. Research on multi-state cooperative estimation method of battery modules considering electro-thermal coupling characteristics[J]. Proceedings of the CSEE, 2025, 45(16): 6340-6353(in Chinese).
- [10] YUAN Tianqing, ZHU Minhong, BAI Jing, et al. A GPR-EDM-UPF framework with false data detection and correction for generalized estimation of SOH in batteries under cyber attacks[J]. Journal of Power Sources, 2024, 623: 235449.
- [11] LI Chaofan, YANG Lin, LI Qiang, et al. SOH estimation method for lithium-ion batteries based on an improved equivalent circuit model via electrochemical impedance spectroscopy[J]. Journal of Energy Storage, 2024, 86: 111167.
- [12] FAHMY H M, HASANIEN H M, ALSALEH I, et al. State of health estimation of lithium-ion battery using dual adaptive unscented Kalman filter and Coulomb counting approach[J]. Journal of Energy Storage, 2024, 88: 111557.
- [13] MU Xianmin, ZHANG Jiahao, LI Guanlin, et al. Estimating SOC and SOH of energy storage battery pack based on voltage inconsistency using reference-difference model and dual extended Kalman filter[J]. Journal of Energy Storage, 2024, 81: 110221.
- [14] PANG Hui, CHEN Kaiqiang, GENG Yuanfei, et al. Accurate capacity and remaining useful life prediction of lithium-ion batteries based on improved particle swarm optimization and particle filter[J]. Energy, 2024, 293: 130555.
- [15] 王琛, 闵永军. 基于容量增量曲线与 GWO-GPR 的锂离子电池 SOH 估计[J]. 储能科学与技术, 2023, 12(11): 3508-3518.
- WANG Chen, MIN Yongjun. SOH estimation of lithium-ion batteries based on capacity increment curve and GWO-GPR[J]. Energy Storage Science and Technology, 2023, 12(11): 3508-3518(in Chinese).
- [16] KIM S, LEE P Y, LEE M, et al. Improved State-of-health prediction based on auto-regressive integrated moving average with exogenous variables model in overcoming battery degradation-dependent internal parameter variation[J]. Journal of Energy Storage, 2022, 46: 103888.
- [17] ZHANG Ji' ang, WANG Ping, GONG Qingrui, et al. SOH estimation of lithium-ion batteries based on least squares support vector machine error compensation model[J]. Journal of Power Electronics, 2021, 21(11): 1712-1723.
- [18] WEN Jianping, CHEN Xing, LI Xianghe, et al. SOH prediction of lithium battery based on IC curve feature and BP neural network[J]. Energy, 2022, 261: 125234.
- [19] 李超然, 肖飞, 樊亚翔, 等. 基于卷积神经网络的锂离子电池 SOH 估算[J]. 电工技术学报, 2020, 35(19): 4106-4119.
- LI Chaoran, XIAO Fei, FAN Yaxiang, et al. An approach to lithium-ion battery SOH estimation based on convolutional neural network[J]. Transactions of China Electrotechnical Society, 2020, 35(19): 4106-4119(in Chinese).
- [20] 李鹏, 葛儒哲, 董存, 等. 基于转置 Transformer 模型的电化学储能自适应 SOH 估计方法[J]. 高电压技术, 2025, 51(6): 2945-2954.
- LI Peng, GE Ruzhe, DONG Cun, et al. Adaptive SOH estimation method based on the transpose transformer model for electrochemical energy storage[J]. High Voltage Engineering, 2025, 51(6): 2945-2954(in Chinese).
- [21] YOU G W, PARK S, OH D. Real-time state-of-health estimation for electric vehicle batteries: A data-driven approach[J]. Applied Energy, 2016, 176: 92-103.
- [22] SHAO Kui, ZHAI Chao, ZHANG Chaolong, et al. An FE-S-BiLSTM and heatmap-based state-of-health estimation method for lithium-ion batteries[J]. IEEE Sensors Journal, 2025, 25(10): 17727-17738.
- [23] ZHANG Chaolong, LUO Laijin, YANG Zhong, et al. Flexible method for estimating the state of health of lithium-ion batteries using partial charging segments[J]. Energy, 2024, 295: 131009.
- [24] YE Min, WANG Qiao, YAN Lesen, et al. Enhanced robust capacity estimation of lithium-ion batteries with unlabeled dataset and semi-supervised machine learning[J]. Expert Systems with Applications, 2024, 238: 121892.
- [25] YAROWSKY D. Unsupervised word sense disambiguation rivaling supervised methods[C]// Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics. Cambridge: Association for Computational Linguistics, 1995: 189-196.
- [26] CHE Yunhong, STROE D I, HU Xiaosong, et al. Semi-supervised self-learning-based lifetime prediction for batteries[J]. IEEE Transactions on Industrial Informatics, 2023, 19(5): 6471-6481.

- [27] XIONG Rui, TIAN Jinpeng, SHEN Weixiang, et al. Semi-supervised estimation of capacity degradation for lithium ion batteries with electrochemical impedance spectroscopy[J]. Journal of Energy Chemistry, 2023, 76: 404-413.
- [28] LI Yuanyuan, SHENG Hanmin, CHENG Yuhua, et al. State-of-health estimation of lithium-ion batteries based on semi-supervised transfer component analysis[J]. Applied Energy, 2020, 277: 115504.
- [29] LIN Chuanping, XU Jun, MEI Xuesong. Improving state-of-health estimation for lithium-ion batteries via unlabeled charging data[J]. Energy Storage Materials, 2023, 54: 85-97.
- [30] LEE G, KWON D, LEE C. A convolutional neural network model for SOH estimation of Li-ion batteries with physical interpretability[J]. Mechanical Systems and Signal Processing, 2023, 188: 110004.
- [31] 方斯顿, 刘龙真, 孔赖强, 等. 基于双向长短期记忆网络含间接健康指标的锂电池 SOH 估计[J]. 电力系统自动化, 2024, 48(4): 160-168.
- FANG Sidun, LIU Longzhen, KONG Laiqiang, et al. State-of-health estimation for lithium-ion batteries incorporating indirect health indicators based on Bi-directional long short-term memory networks[J]. Automation of Electric Power Systems, 2024, 48(4): 160-168(in Chinese).
- [32] ZHOU Zhihua, LI Ming. Semisupervised regression with cotraining-style algorithms[J]. IEEE Transactions on Knowledge and Data Engineering, 2007, 19(11): 1479-1493.



张雨桐

在线出版日期: 2026-02-06。

收稿日期: 2025-05-29。

作者简介:

张雨桐(1997), 男, 博士研究生, 研究方向为储能电池系统健康管理与故障诊断, zyt_ncepu@163.com;

*通信作者: 谢军(1988), 男, 博士, 副教授, 硕士生导师, 研究方向为电气设备智能诊断等, junxie@ncepu.edu.cn。

(编辑 胡琳琳, 张宁)