

DOI: 10.3901/JME.2023.22.046

基于实车运行数据的锂离子电池健康状态估计^{*}

何洪文^{1,2} 王浩宇^{1,2} 王 勇^{1,2} 李双歧^{1,2}

(1. 北京理工大学电动车辆国家工程研究中心 北京 100081;

2. 北京理工大学机械与车辆学院 北京 100081)

摘要: 准确估计锂离子电池健康状态(State of health, SOH)对电动汽车安全管理具有重要意义, 针对实车数据存在电池状态不完整、工况复杂、数据质量差的问题, 提出面向实车数据的多工况健康因子提取 SOH 联合估计方法。首先, 提出实车运行数据工况重构方法, 将数据划分为行驶片段和充电片段, 降低电池工况复杂性。然后, 分别构建行驶工况和充电工况的 SOH 评价模型用于 SOH 估计。对于行驶工况, 选择内阻作为 SOH 评价指标, 通过等效电路模型辨识内阻参数, 基于 Auto-LightGBM 的电池内阻建模方法估算 SOH; 对于充电工况, 选择容量作为 SOH 评价指标并通过提取恒流充电片段计算电池容量, 再提取容量的影响特征, 建立容量模型并估计电池 SOH。结果表明, 基于内阻和容量的建模方法平均绝对百分比误差均小于 9%。最后, 建立结合充电与放电的 SOH 综合评价模型, 提出融合充放电片段的电池 SOH 联合估计方法, 基于实车运行数据的 SOH 误差在 2% 以内, 并在实验室数据和多辆实车数据上验证方法的可靠性和适应性。

关键词: 锂离子电池; 健康状态估计; 健康因子提取; 实车运行数据; 机器学习

中图分类号: TM912

Lithium-ion Battery State of Health Estimation Based on Real-world Driving Data

HE Hongwen^{1,2} WANG Haoyu^{1,2} WANG Yong^{1,2} LI Shuangqi^{1,2}

(1. National Engineering Laboratory for Electric Vehicles, Beijing Institute of Technology, Beijing 100081;

2. School of Mechanical Engineering, Beijing Institute of Technology, Beijing 100081)

Abstract: Accurately estimating the state of health of lithium-ion batteries is significant for the safety management of electric vehicles. Aiming at the problems of incomplete battery states, complex operating conditions, and poor data quality in real vehicle data, a joint SOH estimation method for extracting health factors in multiple operating conditions for real vehicle data is proposed. Firstly, the method of condition reconstruction of real vehicle operating data is proposed, which divided the real vehicle data into driving segments and charging segments to reduce the complexity of battery operating conditions. Then, the SOH evaluation models of driving conditions and charging conditions are constructed respectively for SOH estimation. For driving conditions, the internal resistance is selected as the SOH evaluation index, and SOH is estimated by the battery internal resistance modeling method based on Auto-LightGBM. For charging conditions, the capacity is selected as the SOH evaluation index and the battery capacity is calculated by extracting the constant-current charging segment. Then the influence characteristics of the capacity are extracted to establish the capacity model and estimate the battery SOH. The results show that the average absolute percentage errors of the modeling methods based on internal resistance and capacity are both less than 9%. Finally, a comprehensive evaluation model of SOH combining charging and discharging is established, and a joint estimation method of battery SOH combining charging and discharging segments is proposed. The SOH error based on real vehicle data is within 2%, and the reliability and adaptability of the proposed method are verified on laboratory data and multiple real vehicle data.

Key words: lithium-ion battery; state of health estimation; extraction of health factors; real-world driving data; machine learning

^{*} 国家自然科学基金资助项目(U1864202)。20221025 收到初稿,
20230306 收到修改稿

0 前言

随着全球新能源汽车保有量迅速增加，动力电池安全技术成为车辆领域研究前沿的核心议题。动力电池健康状态(State of health, SOH)作为电池管理系统(Battery management system, BMS)的一项主要监测数据，其通过计算动力电池当前状态(内阻、容量等)与初始状态的比值^[1]，反映了动力电池的退化情况，是电池状态的重要表征参数。因此，对电池 SOH 准确估计有利于指导动力电池梯级利用，对提高动力电池的安全性和可靠性有重要意义^[2]。

当前，动力电池 SOH 估计方法主要可分为基于模型的方法、差分分析法和数据驱动的方法。基于模型^[3-5]的方法通过建立不同的电池模型并求得与 SOH 相关的模型参数。文献[3]建立等效电路模型，通过参数辨识求得电池内阻或容量，实现 SOH 估计。文献[4]建立电化学模型，通过分析电池内部的实际电化学反应获得锂离子降解及活性物质消耗情况，进而估计电池 SOH。这类基于模型的方法计算量较大，且需要精确建立电池模型。差分分析法主要包括增量容量分析^[6-7]、差分电压分析^[8]、差分热伏安法^[9]和差分机械参数估计^[10]等。差分分析法通过绘制电压、电流、温度、容量、应变等参数之间的差分变化曲线，提取曲线特征分析 SOH 的变化，但这类方法对试验条件要求较高且需要专门的试验设备。数据驱动的方法则根据各类电池数据，利用极限学习器^[11]、支持向量机^[3, 12]、集成学习^[13-14]、深度神经网络^[15-19]、迁移学习^[20]等机器学习方法建立模型估算电池 SOH。近年来，随着新能源汽车大数据平台的建设及计算机算力的提高，基于数据驱动的 SOH 估计方法受到研究者的广泛关注^[21]。

然而，当前基于数据驱动的 SOH 估计方法存在以下几个问题。一方面，数据驱动的 SOH 估计方法主要基于实验室静态充放电环境的电池数据，在真实充、放电环境下，存在着电池工况复杂，部分电池状态无法获取，数据采集精度低，电池充放电循环不完整，计算资源受限等问题^[22]。例如，实车数据中电池容量、内阻等关键状态无法获取，数据常用的采样周期通常为 10 s 或者 30 s、数据精度也较低。另一方面，现有研究主要基于电池容量或内阻作为健康状态评价指标，存在评价指标单一的问题，而考虑多个指标和特征因子的 SOH 估计方法存在计算时间长、复杂度高的问题^[2]。因此，本文提出了基于实车运行数据的多工况健康因子提取及电池

SOH 联合估计方法。首先，提出工况重构方法将大规模实车运行数据划分为行驶工况与充电工况；接着分别提取放电过程内阻和充电过程容量作为 SOH 的表征指标和影响特征；然后，基于不同工况的评价指标，设计 Auto-LightGBM 算法分别对充放电工况的电池容量和内阻建模，进而实现 SOH 估算；最后，提出融合充放电片段健康因子的电池 SOH 联合估计方法，提升了 SOH 估计精度和鲁棒性，并在多辆实车数据和实验室数据上验证了所提出方法的有效性。

1 数据描述及预处理

1.1 数据集信息

本文中使用的数据集为 2021 年全国高校新能源汽车大数据创新创业大赛中提供的数据。数据来源于某地的 10 辆纯电动公交车，采集的时间跨度为一年，保证了采集的数据中包含季节的变化，可以从温度变化的角度分析电池的健康状态。所有公交车的累计行驶里程均超过 10 万 km，确保动力电池已经进入衰退阶段。数据包含了时间、电压、电流、里程、荷电状态(State of charge, SOC)等信息，数据示意图表 1。

表 1 原始数据示意图表

时间	车速 /(km/h)	总电压 /V	总电流 /I	累计里程 /km	SOC	...
2020/7/6 11:36:31	30.6	557	297.3	131 387.8	59	...
2020/7/6 11:36:46	45.4	556.4	27.8	131 388.0	59	...
2020/7/6 11:37:01	33.4	562.3	-25.3	131 388.1	59	...
...

1.2 数据预处理

如表 2 所示，相对于实验室数据，原始数据中信息过于庞杂，相关参数较多；实车数据的数据精度较差，数据采样周期为 15 s；数据采集条件相比于实验室中存在不足，导致数据中存在较多的缺失和异常值；实车数据中并不包含容量、内阻等信息，因此需要对原始数据进行一定处理以便于后续的分析研究，处理过程包括数据清洗和数据预处理。

表 2 实车行驶数据与实验室数据(以 Mit-Stanford 数据^[23]为例)比较

数据集	采样间隔/s	采样精度	包含信息
实车行驶数据	15	电压/V: 0.1 电流/A: 0.1 温度/℃: 1	电流、电压、 温度、SOC、 累计里程、车 速等
实验室数据 (Mit-Stanford 数据集)	5	电压/V: 0.000 1 电流/A: 0.000 1 温度/℃: 0.000 1	电流、电压、 温度、SOC、 容量、内阻

数据清洗的目的是通过对异常数据的删除与填补,提高数据的质量。异常数据主要包括数据重复、数据值异常及数据缺失。通过比对每帧数据的时间信息可以发现继而删除重复数据。异常值的处理包含两个方面,其一是根据国标中的相关规定与说明删除严重故障或无效数据;其二是删除明显异常的数据,如速度小于 0 的数据。对于缺失的数据,则按照在确保数据质量的前提下尽可能保留数据的原则进行填补,但对于大段缺失电压、电流等关键信息的数据片段则进行删除。清理后的数据还需要进行预处理,删去档位信息等与电池健康状态无关的数据以剔除原始数据中的无用信息。

2 基于行驶数据的动力电池 SOH 估计

本文中估计方法的整体框架如图 1 所示。动力电池健康状态的退化主要体现在容量的衰减与欧姆内阻的增加。在电动汽车的行驶阶段,由于工况复杂导致电流剧烈变化,在本数据使用的采样间隔下无法计算出准确的容量,而利用参数辨识则能够在数据变化较大的情况下获得准确的内阻值,同时由于内阻的增加导致的电池输出功率衰退也是电池健康状态退化的主要表现,因此选择动力电池欧姆内阻值作为基于行驶数据的电池健康状态评价指标。

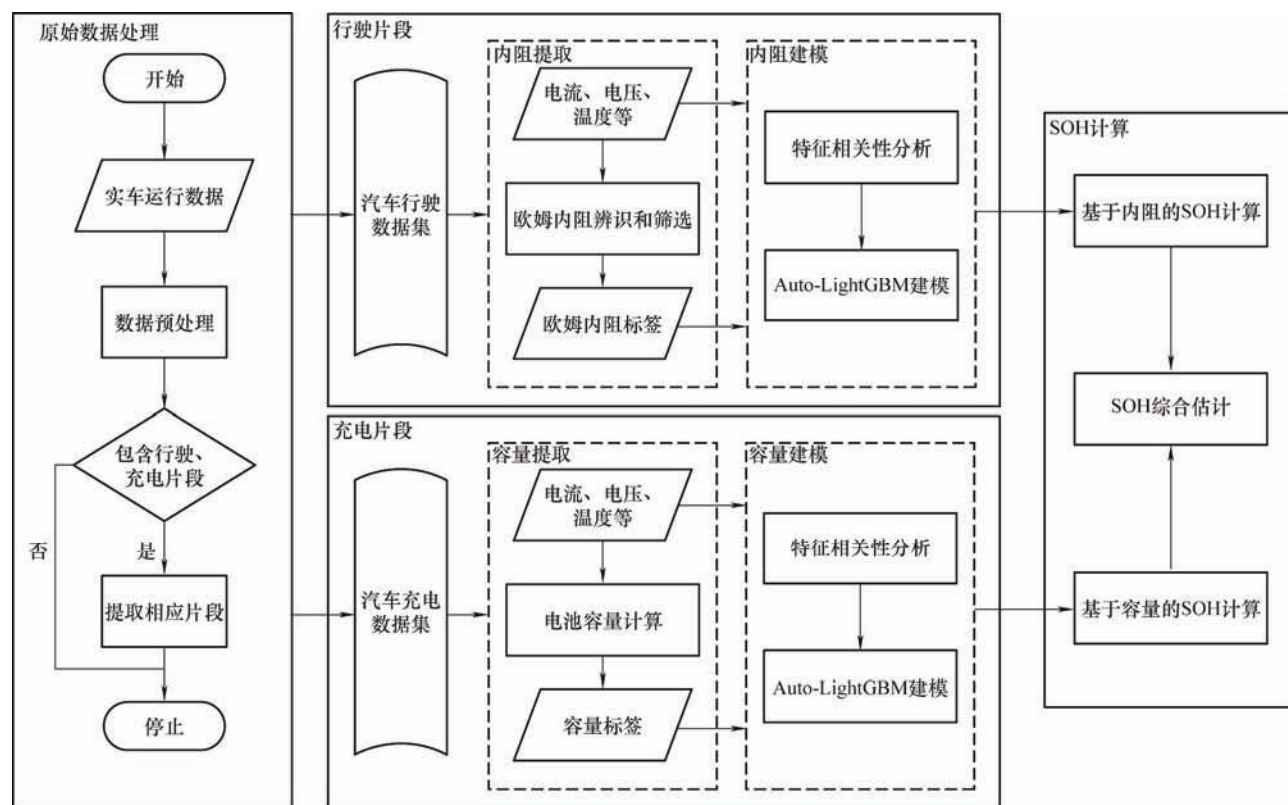


图 1 基于实车数据的电池 SOH 估计方法整体框架

2.1 行驶片段数据提取

原始数据反映了车辆的真实运行状况,包含行驶、充电、长时间停车等多种工况下的数据,因此需要从原始数据中提取行驶片段数据。

由于车辆在实际行驶时存在临时停车情况,且由于制动能量回收系统的存在,行驶片段的数据呈现电流正负混杂、速度有正有零的特征,因此单独使用电流判断无法区别制动能量回收工况与充电工况。而单独使用速度判断则会使得行驶片段的数据提取不完整。在提取行驶片段数据时,可以先提取行驶片段边界,即首先提取所有车速大于零的数据。

再计算每帧数据之间的时间间隔,若两帧数据之间的间隔 t 大于 300 s,则该两帧数据分别为前一行驶片段的结束和后一行驶片段的开始。在获得每一个行驶片段的开始和结束边界后,再从原始数据中利用行驶边界提取出供研究用的完整行驶片段。

考虑提取得到的部分行驶片段时间或行驶里程过短,容易产生较大误差,故删去行驶时间小于 5 min 或累计行驶里程少于 1 km 的行驶数据。同时,在行驶片段内注意到提取的数据中存在数据丢包现象,某些数据帧之间的时间间隔可达 200 s 以上。丢包数据不仅会使得 SOC、累计里程等信息发生突

变, 还会使得某些含有丢包数据的行驶片段数据点过少, 造成较大的误差。但若删除这些数据, 则会进一步扩大两帧数据时间差, 间接上加重了丢包现象, 因此对于这些数据暂时保留, 可以通过后续的参数辨识及异常值检测处理丢包数据。

2.2 动力电池欧姆内阻辨识和异常值筛选

为获得动力电池的准确欧姆内阻值, 需要首先对动力电池建模。由于数据中给出了动力电池组的总电压、总电流数据, 因此在建模时将动力电池包视为整体建模。使用 n -RC 等效电路模型对动力电池建模, 模型结构如图 2 所示。

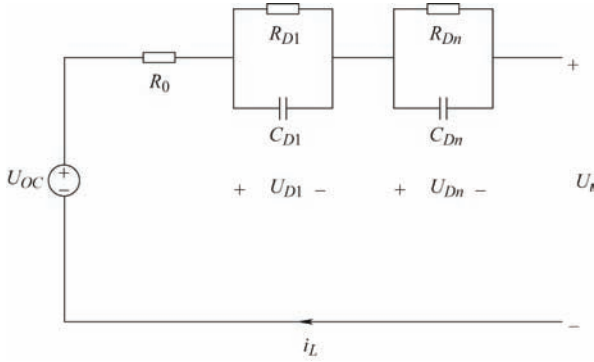


图 2 n -RC 等效电路模型结构图

当前最常用的 n -RC 等效电路模型为 Rint 模型、Thevenin 模型和双极化(Dual polarization, DP)模型, 其分别是 n -RC 模型在 $n = 0, 1, 2$ 时的具体情况。其中, Rint 模型由于没有 RC 网络, 因此不能反映电池的动态特性, 建模精度较差, 而 Thevenin 模型虽然增加了一组 RC 网络以描述电池的极化特性, 但整个模型的动态响应能力依旧有限。DP 模型则采用两组 RC 网络更加准确地描述了电池的极化特性, 其在建模精度与计算复杂度之间保持了较好的平衡^[24], 故此处采用 DP 模型, 其状态方程如式(1)所示

$$\begin{cases} U_t = U_{OC} - i_L R_0 - U_{D1} - U_{D2} \\ \dot{U}_{D1} = -\frac{U_{D1}}{C_{D1} R_{D1}} + \frac{i_L}{C_{D1}} \\ \dot{U}_{D2} = -\frac{U_{D2}}{C_{D2} R_{D2}} + \frac{i_L}{C_{D2}} \end{cases} \quad (1)$$

式中各参数的定义如图 2 所示。为便于实际应用, 需要将状态方程离散化, 离散化后的结果为

$$\begin{cases} U_{t,k} = \Phi_{2,k} \theta_{2,k} \\ \Phi = (1 U_{t,k-1} U_{t,k-2} i_{L,k} i_{L,k-1} i_{L,k-2}) \\ \theta = ((1 - \theta_1 - \theta_2) U_{OC,k} \theta_1 \theta_2 \theta_3 \theta_4 \theta_5)^T \end{cases} \quad (2)$$

式中, Φ 为系统数据矩阵, θ 为系统参数矩阵。基于式(2)中的离散化模型, 可利用带遗忘因子的递推最小二乘法(Forgetting factor recursive least squares, FFRLS)辨识相关参数。针对电池系统

$$y_k = \Phi_k \theta_k + e_k \quad (3)$$

式中, y_k 为系统观测值, 此处指测量的端电压, Φ_k 和 θ_k 为数据矩阵和参数矩阵, e_k 为系统噪声, FFRLS 的算法设计流程为

$$\begin{cases} K_k = P_{k-1} \Phi_k^T (\Phi_k P_{k-1} \Phi_k^T + \mu)^{-1} \\ \hat{\theta}_k = \hat{\theta}_{k-1} + K_k (y_k - \Phi_k \hat{\theta}_{k-1}) \\ P_k = \frac{1}{\mu} (I - K_k \Phi_k) P_{k-1} \end{cases} \quad (4)$$

式中, K_k 是增益矩阵, P_k 是协方差误差矩阵, I 为单位矩阵, μ 为遗忘因子, 此处为 0.97。将实车行驶数据中的总电压、总电流按数据矩阵的格式输入, 应用 FFRLS 即可辨识得到参数矩阵, 进而解析出欧姆内阻 $R_{0,k}$ 和电压辨识误差 e_k 的表达式为

$$R_{0,k} = \frac{\hat{\theta}_{2,k}(4) - \hat{\theta}_{2,k}(3) - \hat{\theta}_{2,k}(5)}{1 + \hat{\theta}_{2,k}(1) - \hat{\theta}_{2,k}(2)} \quad (5)$$

$$e_k = y_k - \Phi_k \theta_k \quad (6)$$

对车辆的一个行驶数据片段进行辨识, 得电池电压辨识误差的结果如图 3 所示, 图中还根据原始数据中的总电压值画出了误差限。由图 3 可看出, 电压辨识误差始终小于总电压的 2%, 说明参数辨识结果较好, 合理的辨识精度可用于后续任务中。

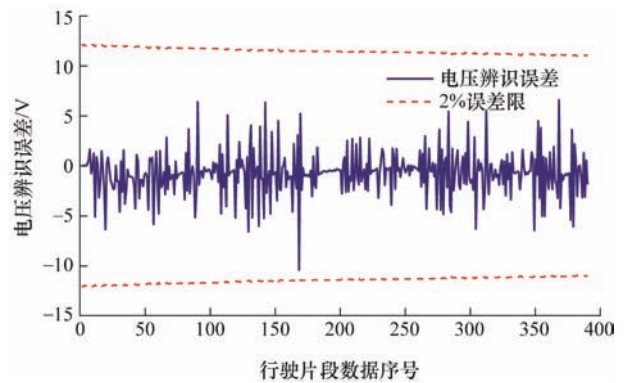


图 3 电压辨识误差结果图

由于实车运行数据存在丢包、噪声的影响,因此导致参数辨识得到的内阻值存在一定的异常值,因此需要对辨识结果进一步处理。此处采用箱型图分析法筛选异常值。箱型图法是基于四分位点内距(Inter-quartile range, IQR)探测异常值,其公式为

$$\begin{cases} \text{IQR} = Q_3 - Q_1 \\ Q_u = Q_3 + 1.5 \times \text{IQR} \\ Q_l = Q_1 - 1.5 \times \text{IQR} \end{cases} \quad (7)$$

式中, Q_1 和 Q_3 分别为上四分位数与下四分位数, Q_l 和 Q_u 分别为筛选下边界和上边界。超出上下边界的数即为异常值。

由于动力电池欧姆内阻值会随着运行时间和运行状态的改变而改变,因此不能对所有内阻辨识结果直接应用箱型图分析,而应该对内阻结果按一定条件分类后再分别筛选异常值。由过往经验可知,温度对欧姆内阻有较大影响,且原始数据中同一温度下均存在若干条电池运行数据,存在足够多分析内阻值的样本,因此按照温度对内阻辨识结果分类并采用箱型图法分析异常值,分析结果如图 4 所示。

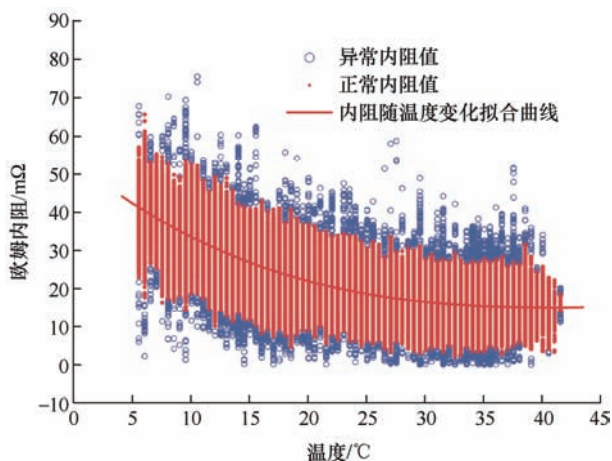


图 4 箱型图筛选结果

根据图 4 可知,经过箱型图分析与筛选,原结果中欧姆内阻辨识为负值及辨识结果过大或过小的异常值均被筛选。对余下的内阻值辨识结果按温度取均值并拟合随温度的变化曲线,可以看出辨识的欧姆内阻随温度增大而减小,符合欧姆内阻与温度之间的特征关系,证明异常值筛选的有效性。

2.3 欧姆内阻影响特征提取

为了减少实车运行情况下测量数据精度、噪声

和丢包等因素对内阻辨识结果的影响,欧姆内阻辨识结果不能直接用于计算 SOH,而应该建立内阻回归模型以分析及预测其变化情况。已有大量研究证明,循环次数、温度、充放电倍率和充放电深度是影响电池衰退的主要因素,但原始数据中并没有给出除温度外的相关数据,需要根据已有的车辆行驶数据提取出对欧姆内阻有影响的参数^[25],排除影响较小或者相互之间存在较大关联的特征,因此首先对原始数据进行特征工程,以分析和提取实车数据中对欧姆内阻变化有重要影响的高价值特征,便于后续利用实车行驶数据对欧姆内阻的建模。

在整车行驶数据中,动力电池的循环次数与车辆累计行驶里程有密切关系。图 5 中分别给出了不同温度下欧姆内阻随累计里程的变化情况及其拟合曲线。在实车数据中,由于累计里程的测量精度限制为 100 m,导致同一行驶片段内不同帧的行驶数据可能有同样的行驶里程,单独依靠行驶里程并不能完全反映电池循环次数,因此同时提取车辆自某一时间起的运行时间,与行驶里程联合反映电池循环次数。在此处,初始运行时间设为第一个行驶片段的起始时间。

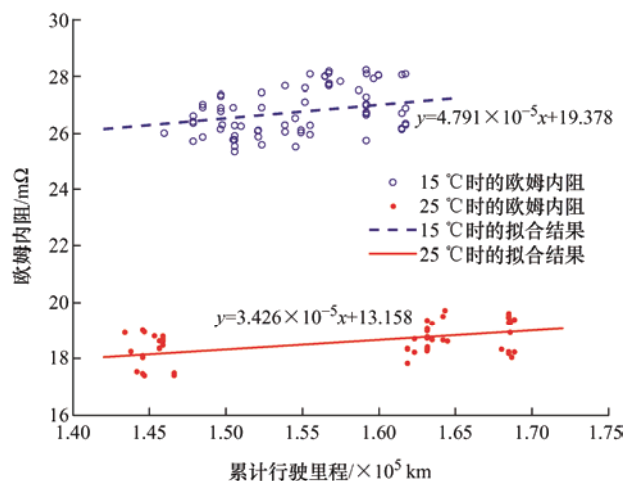


图 5 欧姆内阻随温度和里程的变化图

同理,行驶片段的平均电流能够反映充放电倍率,行驶片段的初始 SOC 与终止 SOC 能反映充放电深度,因此首先选取以上数据作为模型输入。此外,从原始数据中提取总电流、行驶车速等信息。对提取的所有特征和内阻进行 Spearman 相关性分析结果如图 6 所示。

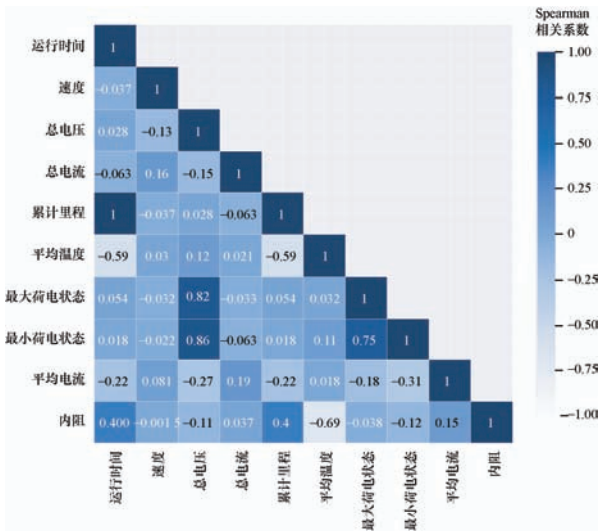


图 6 欧姆内阻相关性分析图

由图中可以看出，运行时间、总电流、累计里程、温度、行驶片段终止 SOC 和片段平均电流与欧姆内阻关系较大，但总电流和终止 SOC 之间存在较大的相关性，故删去总电流数据。运行时间和累计里程虽然完全正相关，但由于上文所述里程精度原因全部保留，最终确定运行时间、累计里程、行驶片段平均电流、行驶片段平均温度和行驶片段终止 SOC 为模型输入。

2.4 欧姆内阻建模及动力电池健康状态估计

前文已经获得了欧姆内阻的辨识值，提取了对内阻有较大影响的特征。接下来，本文采用机器学习方法对内阻建模。

当前针对此类问题常用的机器学习方法包括线性回归(Linear regression, LR)、支持向量机(Support vector machine, SVM)、随机森林(Random forest, RF)、梯度提升决策树(Gradient boost decision tree, GBDT)及深度学习算法中的带全连接层的长短期记忆神经网络(Long short term memory-full connected, LSTM-FC)等。其中，梯度提升决策树模型利用多个弱分类器(决策树)组合成为强学习器，通过决策树迭代训练获得最优模型，是传统机器学习算法中效果较好的方法，包括 XgBoost 和 LightGBM(Light gradient boost machine)等。LightGBM 则在 GBDT 模型的基础上做出相应改变，在不影响训练精确度的条件下极大地加快了 GBDT 模型的训练速度，使其更加适用于工程应用中的大数据处理。表 3 列出了不同的模型对动力电池内阻建模的精度结果和所用时间。建模的精度结果采用方均根误差(Root mean square error, RMSE)、平均绝对百分比误差

(Mean absolute percentage error, MAPE)和模型可决系数 R^2 评估模型在测试集上的准确率，其计算公式为

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \quad (8)$$

$$MAPE = \frac{100\%}{m} \sum_{i=1}^m \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (9)$$

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

式中， y_i 为实际值， \hat{y}_i 为模型预测值， \bar{y} 为实际值的平均值。

表 3 不同算法对内阻的建模结果训练时间

模型名称	方均根误差 RMSE/mΩ	平均绝对百分 分误差 MAPE(%)	可决系数 R^2	训练时间 t/s
LR	5.41	11.08	0.72	0.867
SVM	6.05	13.17	0.69	1.256
RF	4.95	10.54	0.79	4.259
XgBoost	3.92	9.39	0.83	3.217
LightGBM	3.85	9.27	0.86	1.377
LSTM-FC	4.05	9.65	0.82	8.445

由上表可以看出，XgBoost、LightGBM 和 LSTM-FC 三种方法的建模精度明显高于其他方法，而 LightGBM 方法在训练时间方面也存在较大的优势，因此本文选择 LightGBM 方法对内阻建模。

超参数是指对模型训练效果有较大影响的参数，LightGBM 主要的超参数为学习率、树的深度、树的数目和每颗树的叶子数，优化超参数可以提升模型的运行效率和结果。针对 LightGBM 模型，传统的超参数优化方法为网格搜索法，网格搜索法虽然能获得理论上的全局最优超参数，但由于实际工程中往往存在海量的数据，每一组网格参数组合即需要拟合一次模型，因此网格搜索法耗时极长，且离散化的网格对于学习率等非整数的超参数调整效果较差。本文中针对超参数的搜索过程应用贝叶斯优化算法，建立 Auto-LightGBM 模型，可以实现超参数的优化。贝叶斯优化的原理是利用在参数空间中已搜索的点信息确定下一个点，避免多次拟合模型，可以极大地提高搜索速度和最终效果。Auto-LightGBM 算法的运行流程如图 7 所示。

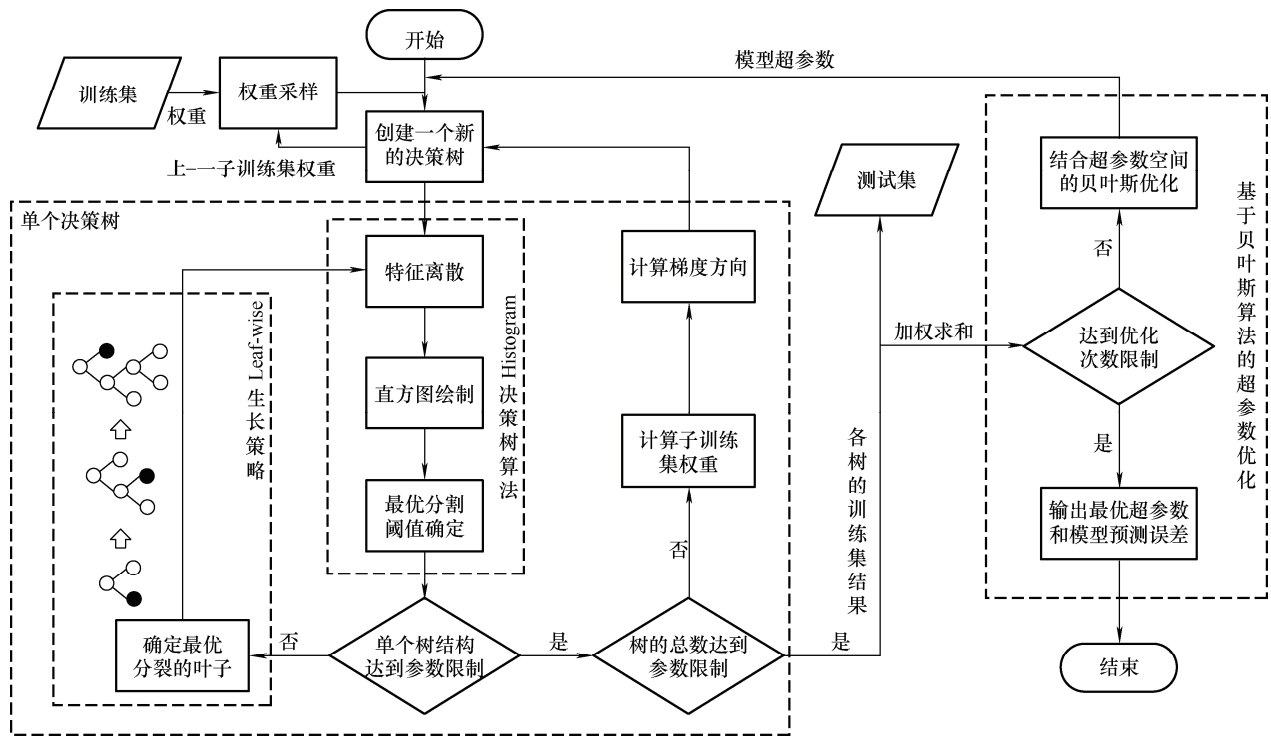


图 7 Auto-LightGBM 算法运行流程

针对 LightGBM 中学习率、树的深度、每棵树的叶子数、树的个数 4 个主要的超参数联合设置超参数空间,利用 Auto-LightGBM 对超参数空间搜索,自动优化的搜索结果如图 8 所示。

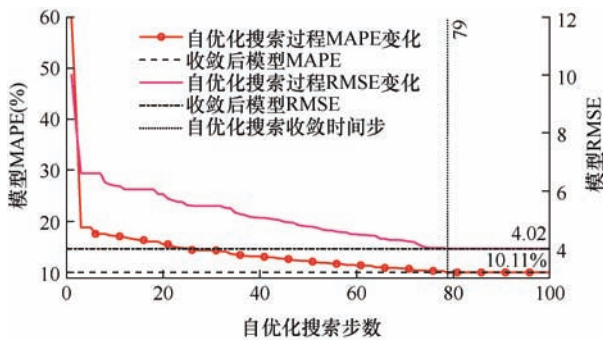


图 8 Auto-LightGBM 的超参数搜索过程

由图中可以看出, Auto-LightGBM 在搜索的第 79 步即收敛,搜索用时 104.469 s。得到最优学习率为 0.392,树的深度为 10,叶子数为 116,树的个数为 152。应用最优超参数训练模型,训练集共包含 2 872 个行驶片段,218 110 帧数据,测试集共包含 507 个行驶片段,38 490 帧数据。利用 RMSE、MAPE 和 R^2 评估模型在测试集上的准确率,参数辨识获得的欧姆内阻值与模型估计结果如图 9 所示。模型估计效果较好, RMSE 为 3.24 mΩ, MAPE 为 8.92%, 整体回归率 R^2 为 0.92。

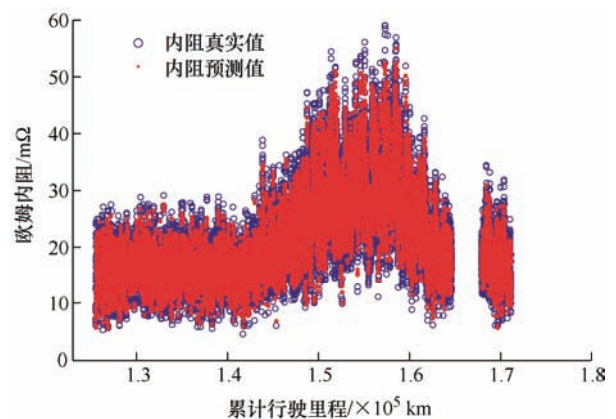


图 9 Auto-LightGBM 预测内阻值与真实值比较图

考虑到实车使用特点,存储多个行驶片段后利用空闲时间统一计算 SOH 较合理,因此统计每个行驶片段参数辨识和模型预测的平均内阻,利用内阻计算 SOH 的公式为

$$SOH_R = \frac{R_{EOL} - R_i}{R_{EOL} - R_{new}} \times 100\% \quad (11)$$

式中, R_i 为当前电池欧姆内阻, R_{new} 为出厂时的电池欧姆内阻,此处用前 20 个片段的平均内阻代替, R_{EOL} 为电池寿命终止时的欧姆内阻,计算时常采用 R_{new} 的两倍^[26]。计算获得其 SOH 随累计行驶里程的变化如图 10 所示,图中的累计行驶里程为该行驶片段结束时的车辆里程。

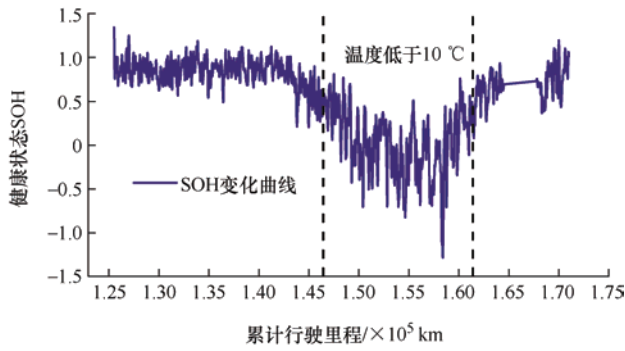


图 10 基于内阻的电池 SOH 变化图

可以看出, 欧姆内阻由于受到温度影响较大, 电池 SOH 在低温时出现不合理值, 这也说明单独依赖内阻判断 SOH 并不合理, 需要结合其他判据综合评价。

3 基于充电数据的动力电池 SOH 估计

在动力电池的一个充电片段内, 充电电流的变化相对平稳, 不适用于参数辨识, 但恒流充电片段有利于计算充电容量, 因此, 本文选择充电过程的电池容量值作为动力电池健康状态评价指标。

相较于行驶数据, 充电数据的特征明显, 提取相对简单。通过车辆充电时电流为负且车速为零即可提取出所有充电数据。提取完所有充电数据后, 同样需要划分充电片段。考虑到实车运行时充电往往间隔较长时间, 因此当两帧数据之间的时间间隔大于 300 s 时即视为不同的充电片段。

3.1 充电容量计算

针对动力电池容量的估计, 已有研究主要采用 ICA、DVT 等差分分析法, 通过计算充电片段电压、温度、容量等数据的差分关系, 提取差分曲线的波峰位置^[27]、波峰幅值^[12]等信息估计动力电池的容量。但这些方法适用于实验室在固定的试验条件下测得的较为精确的电池数据^[28], 实车运行时很难保证运行条件的一致, 且实车数据存在的丢包、错误等现象使得差分分析曲线出现很大波动, 难以准确提取波峰的相关信息, 故在处理实际运行数据时不适合使用差分分析法。通过分析充电片段的电流数据, 发现充电电流在一定区间内重复出现, 这是因为车辆在充电时存在恒流充电片段, 因此可基于恒流充电片段计算其片段容量。

首先提取恒流充电片段, 方法是计算每个充电片段内前后两帧数据总电流的差值, 若总电流差值小于 5 A, 则该两帧数据均为恒流充电。为保证提取的充电片段具有一定长度, 避免恒流充电片段过

短导致计算得到的容量及 SOC 变化过小造成误差, 最终选择连续恒流充电超过 200 s 的恒流充电数据构成最终的充电数据集, 并根据安时计量法采用式 (12) 计算其在恒流充电片段内的充电容量 Q 。提取的恒流充电片段的 SOC 均处在 40%~90% 区间内, 电池均处于稳定工作阶段, 确保了容量计算的可靠性。

$$Q = \sum_0^T I(t) \cdot \Delta t \quad (12)$$

式中, $I(t)$ 为 t 时刻电流, Δt 为两帧数据时间差。为确保后续分析时容量的一致性, 同时提取恒流充电期间电池 SOC 的变化值, 利用式 (13) 计算出电池从容量为零到充满的容量 Q_f 。

$$Q_f = \frac{Q}{\Delta SOC} \quad (13)$$

式中, ΔSOC 为恒流充电片段的 SOC 变化值。由于在单次充电过程中可以认为电池容量不变且原始数据已经给出符合精度要求的电池 SOC, 安时计量法也不涉及温度、循环次数等其他影响容量的因素, 因此通过以上方法计算得到的结果可以作为电池容量的标签信息用于容量的建模。

3.2 充电容量影响特征提取

由于不是每个充电片段都存在连续的恒流充电过程以计算容量, 因此对充电容量的估计同样需要建立多特征预测模型。与第 2 章中欧姆内阻的影响特征提取类似, 同样需要提取对容量有较大影响的特征并建模分析。动力电池的容量同样与温度、循环次数、充放电倍率和深度有关。由于不同的充电片段都相隔较久, 不会出现累计里程因受到精度限制而相同的情况, 因此累计里程已经足以反映循环次数的变化。图 11 展示了在不同温度条件下动力电池的容量随累计里程的变化情况。

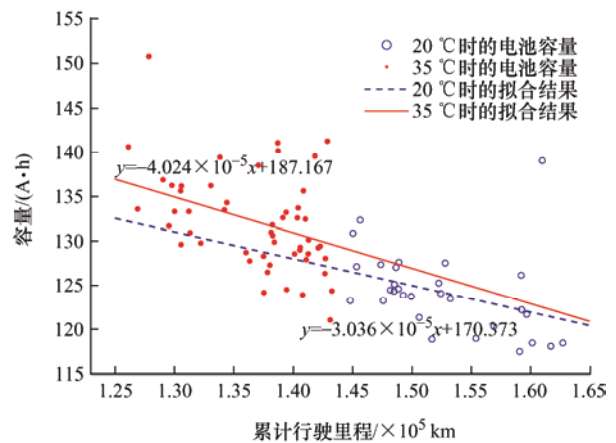


图 11 电池容量随温度和里程的变化图

同样,通过提取充电片段的平均电流反映充放电倍率,提取充电片段的起始与终止 SOC 反映充放电深度。需要注意的是,虽然容量是基于恒流片段计算出来的,但计算平均电流和起始与终止 SOC 时应考虑整个充电片段以更好地反映充电片段的整体特征。此外,将充电片段平均电压纳入可能影响 SOH 的特征中。可得到各个特征的相关性分析结果,如图 12 所示。

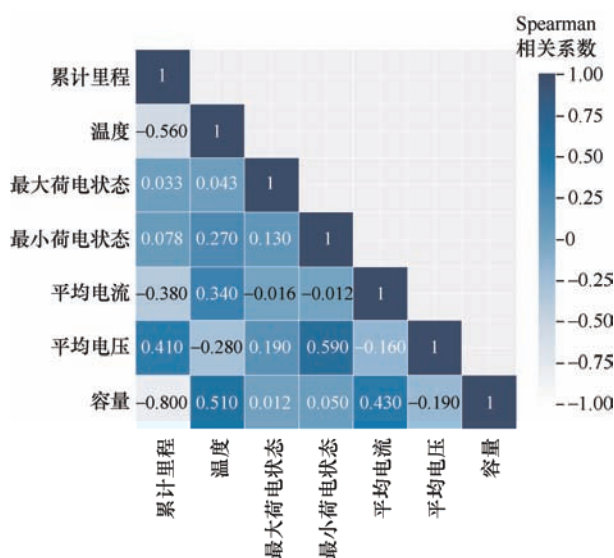


图 12 电池容量相关性分析图

由图 12 可看出,累积里程、充电片段平均电流和平均温度、充电片段的初始 SOC 与终止 SOC 均与容量相关且不存在严重的自相关,因此确定模型输入为累积里程、充电片段平均电流、充电片段平均温度和充电片段的初始 SOC 与终止 SOC。

3.3 充电容量建模及动力电池健康状态估计

与内阻建模类似,同样选择多个机器学习算法建立容量预测模型,其结果如表 4 所示。

表 4 不同算法对容量的建模结果及训练时间

模型名称	方均根误差 RMSE/(A·h)	平均绝对 百分误差 MAPE(%)	可决系数 R^2	训练 时间 t/s
LR	4.82	3.91	0.72	0.077
SVM	5.35	5.15	0.62	0.113
RF	4.12	2.94	0.81	0.364
XgBoost	3.39	2.58	0.90	0.881
LightGBM	3.34	2.52	0.90	0.492
LSTM-FC	3.43	2.72	0.89	4.652

根据表中的结果,LightGBM 方法在训练时间和预测精度方面相较其他方法更有优势,因此对容量的建模同样利用 Auto-LightGBM 算法。将上文确

定的特征作为模型输入, Q_f 作为标签训练模型。通过设置超参数空间,搜索得到最优学习率为 0.282,树的深度为 7,叶子数为 43,树的个数为 186。应用最优超参数训练模型,训练集包含 378 个充电片段,测试集包含 126 个充电片段。训练结果如图 13 所示。模型在测试集上的 RMSE 为 $2.4 \text{ A} \cdot \text{h}$, MAPE 为 2.02%, R^2 为 0.93。

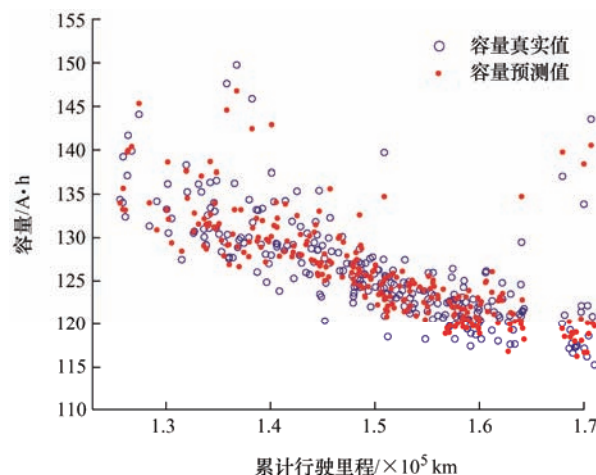


图 13 Auto-LightGBM 预测容量值与真实容量值比较图

根据容量计算 SOH 的公式为

$$\text{SOH}_Q = \frac{Q_i}{Q_m} \times 100\% \quad (14)$$

式中, Q_i 为电池当前容量, Q_m 为电池出厂时的容量。根据模型预测得到的容量值,假设初始状态时电池 SOH 为 100%,根据前 20 个充电片段的实际容量均值计算 $Q_m = 135.97 \text{ A} \cdot \text{h}$,根据最后 20 个充电片段的实际容量求均值计算得到的终止时刻容量 $Q_e = 120.67 \text{ A} \cdot \text{h}$,模型预测的电池 SOH 随行驶里程增量的变化结果见图 14。

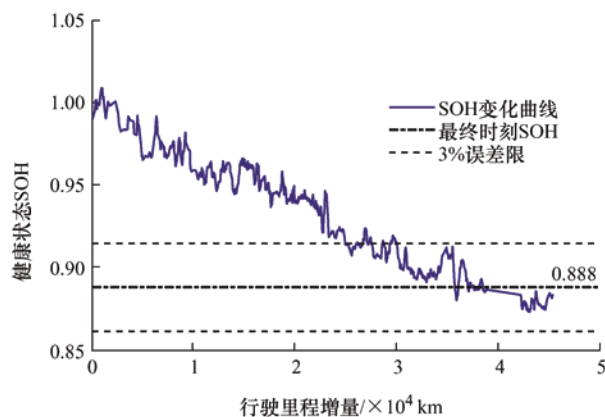


图 14 基于容量计算的电池 SOH 变化图

由图中可看出,在里程增加了 5 万 km 的条件下电池 SOH 衰退了约 12%,且模型预测结果与实际结果误差小于 3%。考虑到实际上车辆此前已经有

12 万 km 的行驶里程,因此 SOH 的衰退属正常范围。由于实车运行过程中动力电池在低温环境下充电时需要先将电池加热到合适的温度,由充电片段数据也可以发现仅存在少量 10 °C 以下的低温充电数据,因此基于容量的 SOH 曲线并没有因为过低的温度而产生不合理的 SOH 估计结果。

4 融合充、放电过程的电池 SOH 估计

4.1 电池 SOH 的联合估计

在上文中,分别针对行驶数据和充电数据提取欧姆内阻和充电容量,进而建模分析了动力电池 SOH 的变化情况。但在实际应用中,由两种方法分别得到的动力电池 SOH 值通常不一致,因此需要融合充放电片段结果对电池 SOH 进行联合估计。

在低温条件下,由上文可知通过欧姆内阻和通过容量获得的电池 SOH 值存在较大差别。这是因为内阻受到温度变化的影响较大。低温时内阻急剧增大且初始内阻是在高温条件下确定的,这使得 SOH 估计结果与常温时出现较大差异,但这并不能反映电池的真实健康状态。而电池容量受温度影响较小,且实际使用中电池不允许过低温度充电,因此在低温时对 SOH 的联合估计应以容量结果为参考。而在高温环境下时,应用单个指标判断电池健康状态存在局限性^[29],因此结合内阻和容量联合判断电池 SOH。

提取所有行驶里程信息相同的充电片段和行驶片段的 SOH 估计结果,先利用式(15)计算得到基于充电片段和行驶片段的 SOH 估计误差,而后再利用式(16)计算得到联合估计的 SOH 结果。

$$r = \frac{|\text{SOH}_c - \text{SOH}_r|}{\text{SOH}_c} \quad (15)$$

$$\text{SOH} = \begin{cases} \text{SOH}_c & r > 0.1 \\ \frac{\text{SOH}_c + \text{SOH}_r}{2} & r \leq 0.1 \end{cases} \quad (16)$$

式中, SOH_c 、 SOH_r 分别为基于容量和内阻计算得到的 SOH, r 为两种方法的 SOH 估计误差。当 r 较大时,说明环境温度较低,由相关性分析知温度对内阻影响更大,因为电池初始内阻由较高温度的 20 个行驶片段的内阻确定,故低温环境下基于内阻的 SOH 的估计结果不合理。且行驶片段存在着较多的 10 °C 以下的低温运行工况,但在充电片段中仅

有少量出现,更加影响了两种方法的 SOH 估计误差,所以此时电池健康状态估计应以容量估计结果为参考。当 r 较小时,温度较高,类比文献[29]中对各个评价指标的权重联合考虑内阻和温度的影响。在 [0.05, 0.15] 区间内以 0.01 为步长,不同 r 的取值下 SOH 联合估计最终值与估计误差如图 15 所示。由图中可看出当 $r \leq 0.1$ 时估计精度相近,但 r 过小时则要求基于内阻的 SOH 估计结果必须非常接近基于容量的 SOH 估计结果,忽略了较多的基于内阻的 SOH 估计信息,降低了联合估计的效果,当 r 在 0.1 到 0.11 的过程中最终 SOH 估计误差出现较大的突变增加。在兼顾融合充电、行驶片段的 SOH 联合估计和 SOH 估计精度的基础上,选择 r 值为 0.1。最终得到的联合估计的 SOH 结果如图 16 所示,SOH 估计结果的误差小于 2%。

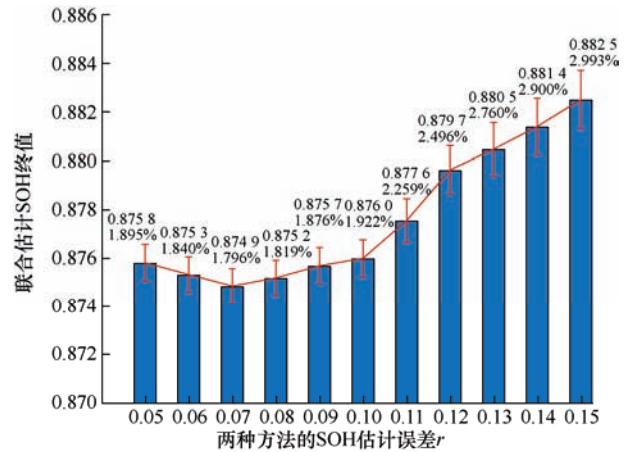


图 15 不同 r 值下的 SOH 联合估计终值及误差

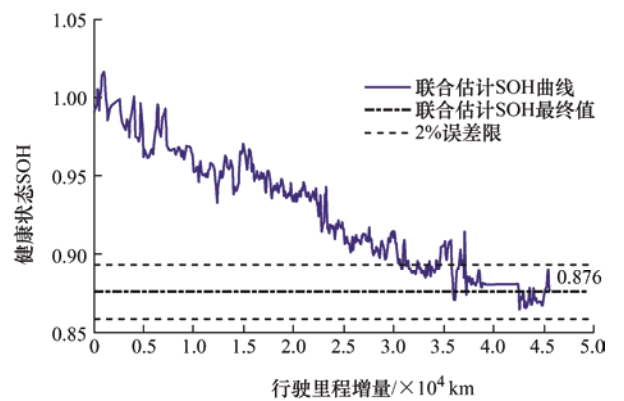


图 16 电池 SOH 联合估计结果

4.2 电池 SOH 联合估计方法的拓展验证

为验证本文中所提出方法的有效性,首先在实验室的电池数据集上进行验证。实验室数据选择 Mit-Stanford 数据集,其对不同的电池采用不同的恒流充电策略,所有电池都以 4C 倍率放电。另外,

数据集中给出了循环次数，因此在利用 Auto-LightGBM 建模时可以替代原方案中的累计行驶里程。对其中一个电池的 SOH 估计结果如图 17 所示。

图 17 中电池 SOH 的真实值根据原始数据包含的每个片段的充电容量计算得出，基于内阻和容量的 SOH 结果则由本文提出的方法得到。由图 17 可以看出，基于内阻和基于容量的 SOH 估计方法取得了较好的估计结果，总体误差较小，证实了基于内阻辨识和恒流片段充电容量的 Auto-LightGBM 建模方法估计电池 SOH 的有效性。

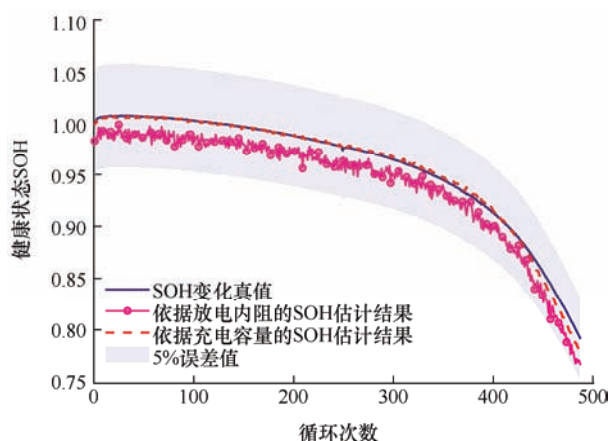


图 17 Mit-Stanford 数据集上的试验结果

接着将文中提出的方法应用于同类型的其他车辆上。图 18 中展示了在其他两辆车辆上的联合 SOH 估计结果，最终的估计误差均小于 2%。图 19 则给出了不同 r 值下两辆车的最终 SOH 联合估计结果和估计误差，图中数据标签为 SOH 百分比估计误差。由结果可看出，不同车辆的 SOH 联合估计误差在 $r = 0.1$ 附近均出现了较大的增加，因此选择 $r = 0.1$ 具有普遍性。

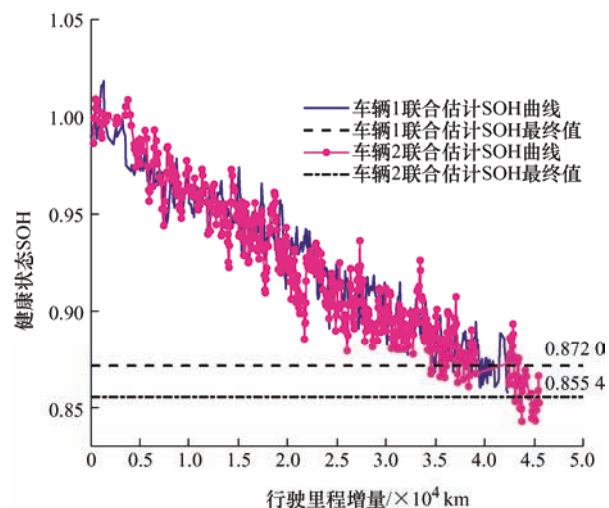


图 18 不同车辆的 SOH 联合估计结果

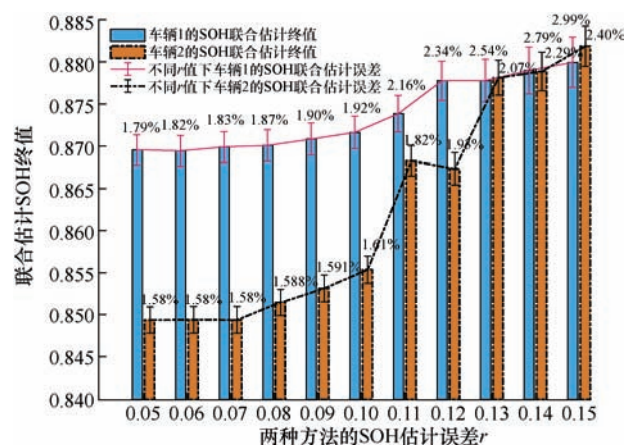


图 19 不同 r 值下的不同车辆 SOH 联合估计结果与误差

5 结论

本文针对实车运行数据的动力电池健康状态估计问题，提出了融合充放电片段健康因子的电池 SOH 联合估计方法。主要结论如下。

(1) 针对实际运行数据工况复杂、数据精度低的问题，提出了一种工况重构方法，将实车运行工况划分，并针对工况变化较大的行驶片段提出参数辨识和筛选方法，降低了复杂工况和低精度数据对健康状态估计的影响。

(2) 针对实车数据中无法获取电池容量等关键特征，构建了行驶工况和充电工况的 SOH 评价模型，提取内阻和容量作为健康状态评价因子，实现了对健康状态关键影响特征的提取和相关性分析。设计了超参数自优化的 LightGBM 模型对健康状态评价因子建模，结果表明，不同评价因子的模型平均绝对百分比误差均小于 9%。对比 LR、RF、SVM、XgBoost、LSTM-FC 等多种机器学习算法，Auto-LightGBM 具有良好的预测精度和实时性。

(3) 针对放电过程中内阻受温度影响显著而充电过程不能反映电池复杂的运行工况的问题，提出了融合充、放电过程的 SOH 联合估计方法，建立了结合充电与放电的 SOH 综合评价模型，试验结果表明联合估计误差小于 2%，相比单一充、放电过程的 SOH 估计精度更高。并在实验室数据和多辆实车数据上验证了本文方法的可靠性和适应性。

参考文献

- [1] HU Xiaosong, FENG Fei, LIU Kailong, et al. State estimation for advanced battery management: Key challenges and future trends[J]. Renewable and Sustainable Energy Reviews, 2019, 114: 109334.

- [2] 王震坡, 王秋诗, 刘鹏, 等. 大数据驱动的动力电池健康状态估计方法综述[J]. 机械工程学报, 2023, 59(2): 151-168.
- WANG Zhenpo, WANG Qiushi, LIU Peng, et al. Review on techniques for power battery state of health estimation driven by big data methods[J]. Journal of Mechanical Engineering, 2023, 59(2): 151-168.
- [3] TAN Xiaojun, ZHAN Di, LÜ Pengxiang, et al. Online state-of-health estimation of lithium-ion battery based on dynamic parameter identification at multi timescale and support vector regression[J]. Journal of Power Sources, 2021, 484: 229233.
- [4] SADABADI K, JIN Xin, RIZZONI G. Prediction of remaining useful life for a composite electrode lithium ion battery cell using an electrochemical model to estimate the state of health[J]. Journal of Power Sources, 2021, 481: 228861.
- [5] WENG Caihao, SUN Jing, PENG Huei. A unified open-circuit-voltage model of lithium-ion batteries for state-of-charge estimation and state-of-health monitoring[J]. Journal of Power Sources, 2014, 258: 228-237.
- [6] TIAN Jinpeng, XIONG Rui, YU Quanqing. Fractional-order model-based incremental capacity analysis for degradation state recognition of lithium-ion batteries[J]. IEEE Transactions on Industrial Electronics, 2019, 66(2): 1576-1584.
- [7] ZHANG Shuzhi, ZHAI Baoyu, GUO Xu, et al. Synchronous estimation of state of health and remaining useful lifetime for lithium-ion battery using the incremental capacity and artificial neural networks[J]. Journal of Energy Storage, 2019, 26: 100951.
- [8] SHIBAGAKI T, MERLA Y, OFFER G. Tracking degradation in lithium iron phosphate batteries using differential thermal voltammetry[J]. Journal of Power Sources, 2018, 374: 188-195.
- [9] 林名强, 吴登高, 郑耿峰, 等. 基于表面温度和增量容量的锂电池健康状态估计[J]. 汽车工程, 2021, 43(9): 1285-1290.
- LIN Mingqiang, WU Denggao, ZHENG Gengfeng, et al. Estimation method of state of health of lithium battery based on surface temperature and incremental capacity[J]. Automotive Engineering, 2021, 43(9): 1285-1290.
- [10] CANNARELLA J, ARNOLD C. State of health and charge measurements in lithium-ion batteries using mechanical stress[J]. Journal of Power Sources, 2014, 269: 7-14.
- [11] PAN Haihong, LÜ Zhiqiang, WANG Huimin, et al. Novel battery state-of-health online estimation method using multiple health indicators and an extreme learning machine[J]. Energy, 2018, 160: 466-477.
- [12] WENG Caihao, CUI Yujia, SUN Jing, et al. On-board state of health monitoring of lithium-ion batteries using incremental capacity analysis with support vector regression[J]. Journal of Power Sources, 2013, 235: 36-44.
- [13] GOU Bin, XU Yan, FENG Xue. An ensemble learning-based data-driven method for online state-of-health estimation of lithium-ion batteries[J]. IEEE Transactions on Transportation Electrification, 2021, 7(2): 422-436.
- [14] MENG Jinhao, CAI Lei, STROE D, et al. An optimized ensemble learning framework for lithium-ion battery state of health estimation in energy storage system[J]. Energy, 2020, 206: 118140.
- [15] 贾俊, 胡晓松, 邓忠伟, 等. 数据驱动的锂离子电池健康状态综合评分及异常电池筛选[J]. 机械工程学报, 2021, 57(14): 141-149, 159.
- JIA Jun, HU Xiaosong, DENG Zhongwei, et al. Data-driven comprehensive evaluation of lithium-ion battery state of health and abnormal battery screening[J]. Journal of Mechanical Engineering, 2021, 57(14): 141-149, 159.
- [16] WANG Cunsong, LU Ningyun, WANG Senlin, et al. Dynamic long short-term memory neural-network- based indirect remaining-useful-life prognosis for satellite lithium-ion battery[J]. Applied Sciences, 2018, 8(11): 2078.
- [17] LI Yihuan, LI Kang, LIU Xuan, et al. Lithium-ion battery capacity estimation—A pruned convolutional neural network approach assisted with transfer learning[J]. Applied Energy, 2021, 285: 116410.
- [18] 周子游, 刘永刚, 杨阳, 等. 考虑混杂充电数据的锂离子电池容量估计[J]. 机械工程学报, 2021, 57(14): 1-9.
- ZHOU Ziyu, LIU Yonggang, YANG Yang, et al. Capacity estimation of lithium ion battery considering hybrid charging data[J]. Journal of Mechanical Engineering, 2021, 57(14): 1-9.
- [19] SHEN Sheng, SADOUGHI M, CHEN Xiangyi, et al. A deep learning method for online capacity estimation of lithium-ion batteries[J]. Journal of Energy Storage, 2019, 25: 1000817.
- [20] CHE Yunhong, DENG Zhongwei, LIN Xianke, et al.

- Predictive battery health management with transfer learning and online model correction[J]. IEEE Transactions on Vehicular Technology, 2021, 70(2): 1269-1277.
- [21] HE Hongwen, SUN Fengchun, WANG Zhenpo, et al. China's battery electric vehicles lead the world: Achievements in technology system architecture and technological breakthroughs[J]. Green Energy and Intelligent Transportation, 2022, 1(1): 2773-1537.
- [22] VICHARD L, RAVEY A, VENET P, et al. A method to estimate battery SOH indicators based on vehicle operating data only[J]. Energy, 2021, 225: 120235.
- [23] ATTIA P, GROVER A, JIN N, et al. Closed-loop optimization of fast-charging protocols for batteries with machine learning[J]. Nature, 2020, 578(7795): 397-402.
- [24] HE Hongwen, XIONG Rui, GUO Hongqiang. Online estimation of model parameters and state-of-charge of LiFePO₄ batteries in electric vehicles[J]. Applied Energy, 2012, 89(1): 413-420.
- [25] LI Yuanyuan, STROE D, CHENG Yuhua, et al. On the feature selection for battery state of health estimation based on charging-discharging profiles[J]. Journal of Energy Storage, 2021, 33: 102122.
- [26] GE Mingfeng, LIU Yiben, JIANG Xingxing, et al. A review on state of health estimations and remaining useful life prognostics of lithium-ion batteries[J]. Measurement, 2021, 174: 109057.
- [27] LI Yi, MONEM M, GOPALAKRISHNAN R, et al. A quick on-line state of health estimation method for Li-ion battery with incremental capacity curves processed by Gaussian filter[J]. Journal of Power Sources, 2018, 373: 40-53.
- [28] WANG Zhenpo, MA Jun, ZHANG Lei. State-of-health estimation for lithium-ion batteries based on the multi-island genetic algorithm and the gaussian process regression[J]. IEEE Access, 2017, 5: 21286-21295.
- [29] 宋春宝. 数据驱动的车用锂离子动力电池健康状态表征与评价方法研究[D]. 北京: 北京理工大学, 2021.
- SONG Chunbao. Research on data driven health status characterization and evaluation method of vehicle lithium-ion power battery[D]. Beijing: Beijing Institute of Technology, 2021.
-
- 作者简介: 何洪文, 男, 1975 年出生, 博士, 教授, 博士研究生导师。主要研究方向为纯电驱车辆动力传动及其控制、燃料电池汽车系统集成及综合控制、车辆网联化智能控制理论与方法。
E-mail: hwhebit@bit.edu.cn
- 王浩宇, 男, 1999 年出生。主要研究方向为电动汽车动力电池大数据与状态估计。
E-mail: 3120210355@bit.edu.cn
- 王勇(通信作者), 男, 1995 年出生, 博士研究生。主要研究方向为智能网联新能源汽车、数据驱动的复杂系统建模与控制。
E-mail: 17862709675@163.com
- 李双歧, 男, 1996 年出生, 博士研究生。主要研究方向为电动汽车动力电池大数据与状态估计。
E-mail: sqli9966@gmail.com