



数据驱动的锂离子电池全生命周期状态参数评估

柳杰¹, 苗宗成^{2*}, 王清云³

1. 西京学院西安市先进光电子材料与能源转换器件重点实验室, 西安 710123;

2. 西北工业大学光电与智能研究院, 西安 710072;

3. 中国航天科工集团公司第二研究院, 北京 100854

* 联系人, E-mail: miaozongcheng@nwpu.edu.cn

2022-05-13 收稿, 2022-08-15 修回, 2022-10-08 接受, 2022-10-09 网络版发表

国家自然科学基金(52173263)、安徽省杰出青年基金(2108085J11)、陕西省区域创新能力引导计划(2022QFY03-02)和中央高校基本科研业务费-西北工业大学(D5000210825)资助

摘要 机器学习作为人工智能核心技术之一, 通过对当前数据的学习找出复杂问题之间的关系而被广泛应用, 机器学习作为研究锂离子电池状态参数估计的新技术, 相关报道也层出不穷。本文首先回顾了传统模型中的电化学模型和等效电路模型在电池状态估计的研究进展; 然后基于这些模型当前的局限性, 回顾、分析、比较了不同机器学习模型在电池状态参数估计中的应用进展及取得的成绩; 最后指出了当前基于数据驱动的人工智能方法在锂离子电池状态参数评估过程中所面临的问题, 并对未来发展提出了建议。

关键词 锂离子电池, 数据驱动, 机器学习, 参数评估

随着人类对全球变暖、温室气体排放和化石燃料消耗的密切关注, 寻找绿色清洁的替代能源逐渐成为当今面临的首要问题^[1]。据统计我国乘用车数量约3亿辆, 交通运输领域碳排放量占整个经济社会碳排放总量的30%左右^[2]。锂离子电池在新能源汽车的成功应用, 改变了汽车动力的驱动方式, 降低了交通运输领域碳排放总量, 并凭借自身高能量密度和循环稳定性占据了当前电动汽车电池市场。然而到目前为止, 锂离子电池仍然是电动汽车中最昂贵的部件之一^[3], 电池管理系统通过当前车辆驾驶状态和外部环境调整输出功率, 保证锂离子电池安全、稳定、高效的工作并延长电池使用的生命周期^[4]。

电池状态参数的精准估计是电池管理系统(battery management system, BMS)中最重要功能之一, 其中电池的荷电状态(state of charge, SOC)、健康状态(state of health, SOH)和剩余使用寿命(remaining useful life,

RUL)是3个重要参数, 通过SOC来监测电池的剩余用电量、运用SOH来评估电池老化程度、使用RUL诊断电池从当前容量衰退至初始容量80%之前的剩余使用时长。对电池状态参数的精准估计可以使电池在替换或废弃前发挥最大效能, 提高电池使用寿命, 因此在碳中和的背景下, 电池管理系统状态参数评估的研究至关重要。

状态参数评估的报道表明, 研究通常分为传统模型方法和基于数据驱动的机器学习方法^[5]。本文首先简单回顾了传统模型在电池状态参数估计中的应用, 对电化学模型、等效电路模型和融合模型的优缺点进行了对比; 然后概述了机器学习的算法分类、模型评估标准及算法选择时需要考虑的因素。其次, 讨论了基于数据驱动的机器学习方法在状态参数估计中的应用进展, 总结了不同算法构建模型的预测精度和稳健性。最后对未来可能需要研究的方向和所面临的挑战进行

引用格式: 柳杰, 苗宗成, 王清云. 数据驱动的锂离子电池全生命周期状态参数评估. 科学通报, 2023, 68: 644–655

Liu J, Miao Z C, Wang Q Y. Data-driven full life-cycle state parameter assessment of Li-ion batteries (in Chinese). Chin Sci Bull, 2023, 68: 644–655, doi: 10.1360/TB-2022-0550

了展望.

1 传统模型

对电池状态参数的估计模型通常分为电化学模型、等效电路模型和融合模型, 本节介绍了不同模型之间的基本原理与应用进展, 并总结了不同模型之间的优缺点.

1.1 电化学模型

电化学模型通过电池反应机理来描述电池的内部行为. Doyle等人^[6]首次提出并开发了适用于多孔电极被电解液包裹的准二维模型(pseudo two dimensions, P2D). 如图1所示, 其包含了多孔电极理论和浓溶液理论模型, 通过几组耦合偏微分方程(partial differential equation, PDE)作为控制方程描述固相或电解液液相中的电荷守恒. 研究提出一种与充电状态和电池电阻有关的用于预测电池容量衰减的半经验模型, 但该模型仅适用于本次实验中用到的18650型号钴酸锂电池, 在计算过程中要求解几组偏微分方程, 因此需要消耗巨大的算力, 且增加了时间成本^[7]. 保持偏微分方程的其他表达式不变, 只使用一组多项式使固相浓度轮廓更接近抛物线的多孔电极模型(porous electrode mode, PEM), 减少了计算复杂度^[8]. 单粒子模型(single particle

model, SPM)将正负电极视为单个球形的粒子态, 因此在 x 轴上的锂离子固相浓度分布差异和液相电势差异都可以忽略不计, 进而减少了运算量, 提高了计算效率, 这种P2D的简化模型也被应用于电池的状态参数估计^[9,10].

1.2 等效电路模型

与使用繁琐的PDE描述电池的内部电化学机制的电化学模型相比, 等效电路模型(equivalent circuit model, ECM)通过使用电阻器和电容器等电路元件来描述电池动态特性^[11], 常见的电路模型有内阻模型、阻容(resistance capacitance, RC)模型和新一代汽车合作计划(partnership for a new generation of vehicle, PNGV)模型, PNGV模型在1阶RC基础上增加了一个电容, 提高了参数的识别精度. 等效电路模型需要更少的参数, 因此被广泛应用于电池管理系统中^[12~15]. 开路电压(open circuit voltage, OCV)作为等效电路模型中的一个重要元素, 可以通过一组带有经验参数的方程来建立OCV与SOC之间的关系. Zheng等人^[16]在1阶RC下建立了描述电池动态行为的方程, 对比了不同温度下增量型开路电压估计器与低电流开路电压估计器的准确性和稳健性, 结果表明增量型开路电压估计器可以达到更好的预测性能. 等效电路模型有较好的扩展性, 因此常

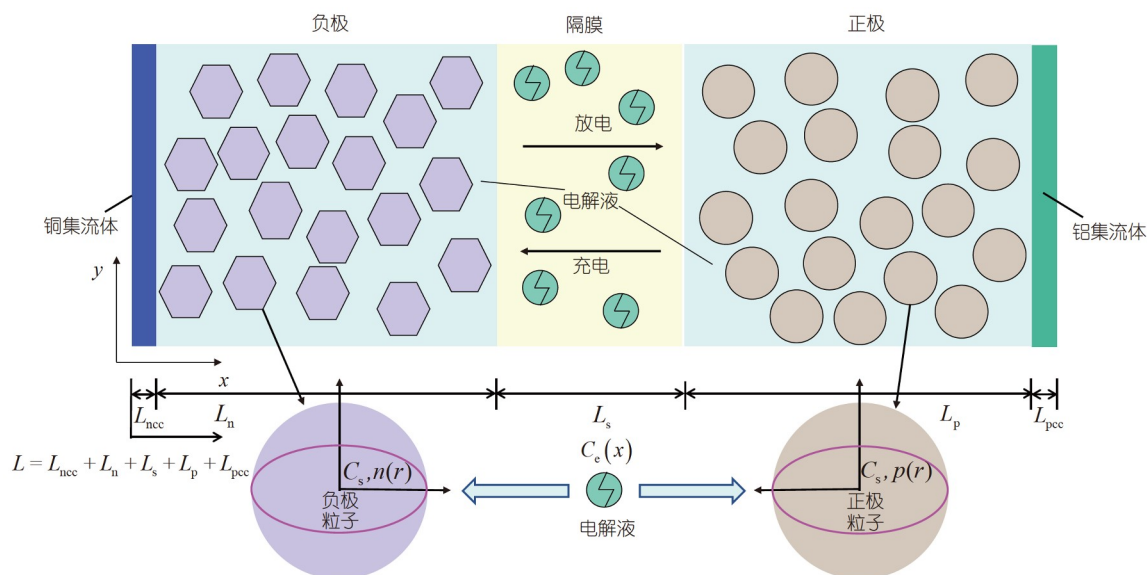


图1 (网络版彩色)电化学模型P2D示意图. L , 沿 x 轴方向电池厚度; L_i , 沿 x 轴方向不同区域电池厚度; C_s , 电极固相浓度; C_e , 电解液液相浓度; $n(r)$, 负极粒子半径; $p(r)$, 正极粒子半径

Figure 1 (Color online) Schematic diagram of electrochemical model P2D. L , battery thickness along x axis; L_i , battery thickness in different areas along the x axis; C_s , electrode solid phase concentration; C_e , electrolyte liquid phase concentration; $n(r)$, negative particle radius; $p(r)$, positive particle radius

结合卡尔曼滤波器完成对状态参数的估计^[17~19]。受噪声波动影响,等效电路模型在参数估计中通常表现出低稳健性。为此,一种自适应的5阶卡尔曼滤波估计器可用于SOC的估计,模型结合了自适应算法和5阶球面径向容积准则的优点,对较大的测量噪声也表现出很好的性能。实验采用2阶RC的双极化电路模型,高斯函数三项式映射出OCV与SOC之间的关系,并在动态应力测试和统一轻型车辆测试程序下,分别进行了无初始测量误差和有初始测量误差测试,验证模型具有较好的估计精度以及纠错能力。所建模型与基于无迹卡尔曼滤波的估计器相比,具有更高的估计精度和稳健性^[20]。

1.3 融合模型

融合模型是在上述提及模型基础上的联合使用,将不同模型的优势结合在一起共同完成电池状态参数的估计^[21],不同模型之间的对比如表1所示。粒子群优化算法(particle swarm optimization, PSO)分别作用在等效电路模型、卡尔曼滤波器完成参数优化和协方差优化,融合后的模型性能优于等效电路模型中的1RC、2RC、3RC和PNGV模型,模型均方根误差(root mean square error, RMSE)均在8.2 mV左右,误差可忽略^[22]。将电化学指标与等效电路模型结合,可以在反映电池内部老化机理的同时减缓模型的计算复杂度。Galatro等人^[23]提出了用于诊断电池健康的工作流程“MSPM-EIS”,将Li等人^[24]应用的改进单粒子模型(modified single particle model, MSPM)作为性能组件,该模块负责完成在高放电倍率下对电池性能的预测,老化组件从拟合的电化学阻抗谱(electrochemical impedance spectroscopy, EIS)曲线中获取指标,量化了固体电解质界面膜生长和析锂现象引起的欧姆损耗和化学降解损耗。

所提出的健康诊断方法能够准确表示锂离子电池的退化行为,且与电化学模型中提到的半经验模型相比具有更好的通用适用性。虽然融合模型将多类数据、算法和模型有效结合,提高了对电池状态参数的估计能力,但引入模型的同时也增加了错误源,且难以制定联合后模型失效的评判标准。

2 机器学习概述

机器学习算法分为有监督学习和无监督学习。有监督学习方法需要完成数据集标注,将变量标记为模型的输入,对应结果标记为输出。根据输出结果分为回归任务和分类任务,回归任务通过拟合数据点,找到回归方程,常见的回归算法有线性回归、随机森林、神经网络等。分类任务的每个样本通常隶属于有限类别之中,其输出结果是离散的,常见的分类算法有逻辑回归、决策树、 k 近邻、支持向量机(support vector machine, SVM)等。与监督学习不同的是,无监督学习是一种缺乏先验信息的学习方式,初始数据无标识,学习最终目标是在数据集中通过模型自身判定将相近的数据块聚集在一起。常见的无监督学习方法包含聚类和降维。

不同机器学习算法思想如下。支持向量机^[25]作为一种二分类模型,最终目的是要在保证分割间距最大化的前提下找到一个超平面完成对样本的划分,对使用核函数的SVM通过提升维度将低维数据映射到高维空间,实现线性可分。 k 近邻^[26]在机器学习中是一种常见的分类和回归方法,新的输入实例通过某种距离度量方法判定给训练样本中距离最近的 k 个实例,即近朱者赤,近墨者黑。 k 近邻与SVM相比更容易处理多分类任务,然而在使用 k 近邻时,不同 k 值在数据集上的表现结果不同,设置过大,致使模型过于简单,输出结果表

表1 模型比较

Table 1 Comparison of models

	传统模型			数据驱动的机器学习模型
	电化学模型	等效电路模型	融合模型	
稳健性	受电池的模型结构和物理参数影响大且当噪声不稳定时性能会变差			模型对噪声和测量偏移等不确定,干扰因素影响小
精度	高	精度不如电化学模型	较高	数据充足下训练出的模型准确性高
优点	考虑内部电化学过程高准确率	模型简单,参数少,原理清晰,计算成本低	增加有用的信息,提高模型预测性能	不需要了解电池内部机理,擅于处理非线性问题
缺点	需要强大的算力,掌握专业的知识,参数多难以建立或识别	模型参数化代价大,准确率没有电化学模型高	难以制定模型的失效准则,引入模型的同时,增加了错误源	模型精度依赖数据集,不可预测的黑箱模型

现为简单的频率统计,划分过小,则导致过拟合.决策树^[27]作为一个树形结构,根节点存放所有训练样本,每个中间节点对应一个判定条件,不同叶子节点代表不同决策结果,从根节点到叶子节点的路径作为一个判定的测试序列.随机森林^[28]是将彼此无关联的一棵棵决策树集成在一起,集成后的随机森林提高了预测的精度,更适合处理高维度的复杂问题,同时在部分数据缺失下也能保持较好的输出,因此具有良好的抗干扰能力. K 均值^[29]作为最简单的方法之一,在聚簇分类任务中被广泛使用,算法思想是随机选取 k 个样本点作为聚簇的质心,计算所有样本点到质心之间的距离,将每个样本点划分至距离质心最近的聚簇中,每次重新计算簇内样本均值改变质心位置,迭代上述步骤直至质心位置不再发生变化. K 均值后的样本点只属于一个聚簇内,且不同聚簇之间不存在重叠嵌套.神经网络^[30]由输入层、隐藏层和输出层组成,输入层中神经元的个数对应输入向量的个数,同时网络为神经元配置权值.除输入层以外,每层网络中的神经元存储上一步长的加权求和,训练过程中网络中的这组权值不断更新,提高模型自身的学习能力.拥有更多层网络结构和节点个数的深度神经网络(deep neural networks, DNN)在训练过程中可以保留更多重要参数,因此具有更强的学习能力.然而深度学习模型^[31]作为一个黑盒子,学习过程不可知,对输出结果缺乏可解释性,且需要大量数据来完成模型的训练.

在处理实际问题时,先将其划分为分类任务、回

归任务或聚簇任务,并参照表2中算法间的比较为模型选择合适算法,同时高精度的模型还与数据集样本的容量和质量呈正相关.机器学习在材料科学领域的不断突破得益于权威数据库,如材料基因项目Material Project^[32~34]和无机晶体结构数据库(the inorganic crystal structure database, ICSD)^[35]的建设.除此之外,数据的获取还可以通过数据挖掘从已发表的文章中收集^[36].准备好的数据集在训练模型前通常需要经过数据清洗完成重复值、缺省值的处理,并根据当前任务需要更改变量类型或标签的重命名.清洗后的数据通过标准化消除特征量纲差异带来的影响,使不同特征的每维数据都能被同等看待,提高估计器的学习能力,其中一类标准化方法计算表达式为

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}. \quad (1)$$

另一种使用标准差完成标准化的计算表达式为

$$x' = \frac{x - \bar{x}}{\delta}. \quad (2)$$

其中, x 为每一维的样本值, $\min(x)$ 与 $\max(x)$ 分别为当前维度下的最小/最大样本值, δ 为样本标准差, $\bar{x}$ 为当前维度下的均值.除了通过数据清洗完成数据预处理外,还应考虑特征维度对模型训练的影响,数据集中特征维度过大不仅会在学习过程中增加训练模型成本,无干扰特征还会影响模型最终的决策能力.机器学习中常用的特征选择方法有不依赖模型的过滤法、以学

表 2 机器学习算法比较

Table 2 Comparison of machine learning algorithms

	算法	精确度	训练时长	线性度	调参数量	优点	缺点
	逻辑回归	高	快	线性	少	计算量小	不容易处理特征之间相关性, 容易欠拟合精度不高
	支持向量机	高	适中	非线性	中	能够处理局部最小值和非线性问题	对核函数的选择敏感, 大规模训练样本速度慢
	线性回归	高	快	线性	少	原理简单易于理解, 算法可解释	难以应对非线性问题
树类结构	梯度提升(XGBoost)					高效可扩展, 正则化防止过拟合	高维空间下效果不好
	决策树(DT)	精准	适中	非线性	中	容易理解和解释	容易过拟合
	随机森林(RF)					对异常值具有稳健性, 精准度好, 容易处理高维数据	低维特征下表现不好
神经网络	深度神经网络(DNN)					准确率高, 学习能力强	黑箱模型训练时间长
	卷积神经网络(CNN)	精准	慢	非线性	多	权值共享减少训练参数	可能会出现梯度消失问题
	循环神经网络(RNN)					很适用于时间序列类型任务	容易出现梯度耗散或梯度爆炸
	K均值	精准	适中	线性	少	收敛快, 大数据集下分类高效	容易收敛至局部最优且容易受到噪声影响

习器性能作为评价特征子集的包裹法以及基于树或森林等算法构建学习器在训练过程中自动完成特征选择的嵌入法, 研究认为包裹法更精准、更有效^[37-39]. 在深度学习中, 如卷积神经网络(convolutional neural networks, CNN), 将输入目标映射至网络神经元后, 由一个指定尺寸大小的卷积核遍历神经元并通过卷积操作提取特征, 连续的卷积操作之间通过池化层保留了更有效特征, 并减少了训练参数数量, 网络经过多层卷积与池化操作提取并保留了最有价值的神经元, 并将输入目标最终映射成一个特征向量传递给全连接层进行识别. 对数据集做完特征工程后, 将其划分为训练集、验证集和测试集, 验证集以训练集的学习结果为模型找到一组最佳的参数, 并由测试集通过评估方式检测模型性能. 分类任务模型采用准确率(accuracy)、精确率(precision, P)和召回率(recall, R)作为性能评价指标. Accuracy计算表达式为

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \quad (3)$$

Precision计算表达式为

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \quad (4)$$

Recall计算表达式为

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \quad (5)$$

F1计算表达式为

$$\text{F1} = \frac{2P \times R}{P + R}. \quad (6)$$

其中, 真正例(true positive, TP): 真实类别为正类, 分类结果也为正类; 假正例(false positive, FP): 真实类别为正类, 分类结果为负类; 假负例(false negative, FN): 真实类别为负类, 分类结果为正类; 真负例(true negative, TN): 真实的类别为负类, 分类结果为负类. 而在回归任务中可以采用平均百分比误差(mean average percentage error, MAPE)、平均绝对误差(mean absolute error, MAE)和RMSE作为性能评价指标. MAPE计算表达式为

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \frac{|y'_i - y_i|}{y_i}. \quad (7)$$

MAE计算表达式为

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y'_i - y_i|. \quad (8)$$

RMSE计算表达式为

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y'_i - y_i)^2}, \quad (9)$$

其中, y'_i 为预测结果, y_i 为实际结果, n 为数据集中样本点个数. 聚簇任务可以采用误差平方和(sum of squared errors, SSE)与轮廓系数(silhouette coefficient, SC)作为评价指标, 低SSE、高SC下对应的聚类效果更好. 其中SSE的计算表达式为

$$\text{SSE} = \sum_{k=1}^K \sum_{x_i \in C_k} |x_i - m_k|^2, \quad (10)$$

$$m_k = \frac{\sum_{x_i \in C_k} x_i}{|C_k|}, \quad (11)$$

其中, C_k 为被划分出的簇, m_k 为 C_k 簇的质心.

SC表达式为

$$\text{SC} = \frac{l_b - l_a}{\max(l_b, l_a)}, \quad (12)$$

其中, l_a 为样本点与聚簇内所有样本点之间的平均距离, l_b 为样本点与相邻最近的聚簇内所有样本点间的平均距离. 机器学习训练模型流程框架如图2所示. 而训练模型过程中, 还需要注意超参数的调整, 即便是在充足训练样本下表现优异的复杂神经网络, 迭代次数的设置及学习速率的选择也会对模型精度带来影响, 训练模型的经验积累最终决定了调节效率.

3 数据驱动建模的电池状态估计

以数据为驱动的机器学习用于电池生命周期内的状态参数估计, 主要包括荷电状态监测、健康状态诊断和剩余使用寿命的预测. 随着机器学习方法的成熟和近些年在各领域取得的成果及所积累的经验, 用于电池状态估计的手段也从简单的线性回归方法到具有复杂网络结构的神经网络. 表S1总结了机器学习算法创建模型在电池状态参数估计的误差精度和稳健性.

3.1 基于数据驱动建模的电池荷电状态估计

机器学习与EIS相结合对电池的状态评估引起了国内外学者的关注^[40], Babaeiyazdi等人^[41]使用松下NCR18650PF型号锂离子电池的EIS数据, 采用线性回归和高斯过程回归算法完成对SOC的估计. 深度学习算法中的卷积神经网络主要用于图像处理和自然语言处理, 并取得了不错的成绩. 不同的是, 图像处理任务

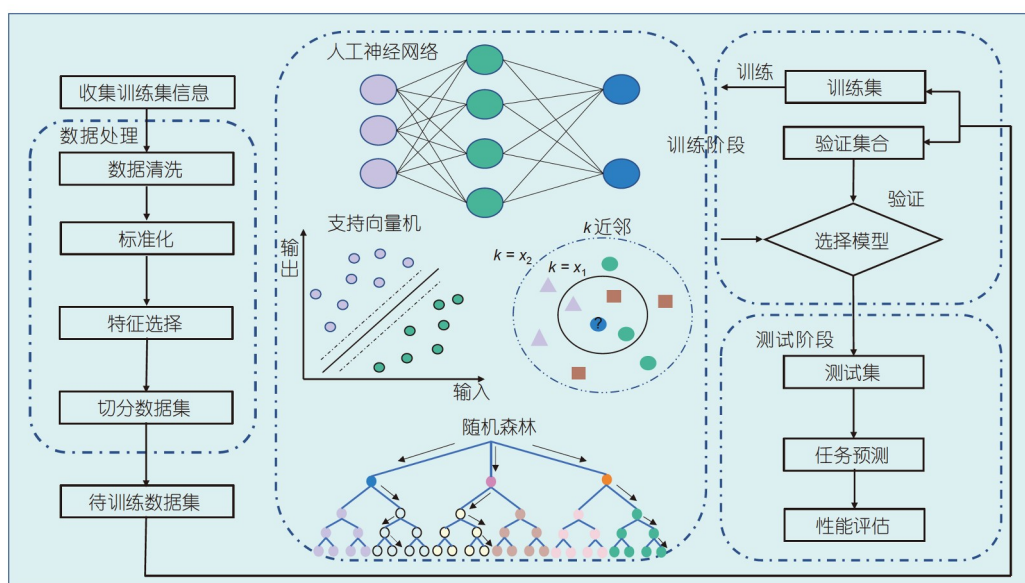


图2 (网络版彩色)机器学习流程框架

Figure 2 (Color online) Machine learning process frameworks

使用的二维卷积, 而电池状态参数估计任务使用的是一维卷积操作。在一维CNN上使用不同尺寸卷积核提取时间序列特征完成SOC的估计^[42]。同时, 具备学习率优化策略的全卷积神经网络算法训练的模型, 在不需要重新训练的前提下完成了不同循环工况驾驶下荷电状态的估计^[43]。长短时神经网络(long short term memory networks, LSTM)于1997年被Hochreiter和Schmidhuber^[44]提出, 单元结构由3个门控单元、2个细胞状态和1个记忆体组成, 内部结构如图3所示。LSTM在电池荷电状态参数估计中, 建立电流、电压和温度与SOC的映射完成电池荷电状态的估计^[45]。滤波器能够减缓LSTM的输出波动, 提高模型电池荷电状态的估计精度, 集成了自适应H-infinity滤波器的长短时神经网络构建的模型MAPE仅为0.57%, 模型具有较高的预测能力^[46]。与单向循环神经网络相比, 双向循环神经网络通过训练过程捕获更多有效信息。在威斯康星大学公布的NCR18650PF电池数据集上, 一种双向堆叠长短时神经网络(stacked bidirectional long short-term memory networks, SBLSTM)实现了SOC估计, 实验结果显示, SBLSTM在不同环境温度下获得了令人满意的估计精度, 平均绝对误差为0.46%。与此同时, 将模型作用在马里兰大学高级生命周期工程中心(Center for Advanced Life Cycle Engineering, CALCE)用INR18650-20R电池采集的负载循环数据, 模型同样实现了低误差精度, 完

成了模型的可扩展性验证^[47]。

3.2 基于数据驱动建模的电池健康状态估计

Shu等人^[48]从电压曲线为最小二乘支持向量机挖掘特征, 并通过遗传算法找到最佳模型参数, 实验在高采样精度下并没有显著提高预测精度, 不同采样时间下表现出的绝对误差均保持在2%以下。基于树的梯度提升模型中, 使用特征相关系数与方差膨胀系数筛选特征, 模型在保证良好估计性能的前提下有效避免了过拟合^[49]。在Li等人^[50]提出的LSTM在线容量估计模型中, 使用亚琛工业大学在三洋UR18650E圆柱型电池采集的负载循环数据, 建立了能够稳健处理噪声的模型且最大平均绝对百分误差仅为2.08%。电池内部不稳定因素对电池老化同样有着重要影响, 如固体电解质界面膜(solid electrolyte interface, SEI)增加了电池阻抗, 低温条件和高倍率充电诱发锂枝晶, 不同老化机制在电池不同组件中发生各自独立的老化反应^[51], 电池内部老化衰减机制如图4所示。深度神经网络不考虑电池内部老化机制, 从美国宇航局(National Aeronautics and Space Administration, NASA)艾姆斯预测卓越中心提供的锂离子电池循环数据集中有效提取了电流、电压和充电容量对应的样本数据, 建立了放电容量为输出的健康状态估计模型^[52], 在大规模数据集下准确反映了复杂数据与健康指标之间的关系。为减缓不同充电策

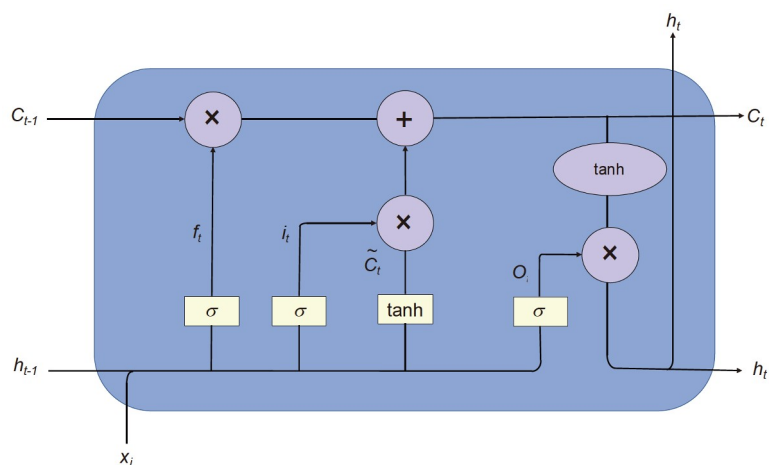


图 3 (网络版彩色)长短时神经网络单元结构. C_{t-1} , 上层细胞单元输出细胞态; C_t , 当前细胞单元输出细胞态; h_{t-1} , 上层细胞单元输出记忆体; h_t , 当前细胞单元输出记忆体; x_t , 当前细胞单元传入信息; f_t , 遗忘门; i_t , 输入门; O_t , 输出门; σ , Sigmoid 激活函数; $\tanh$, 双曲正切激活函数; $\tilde{C}_t$, 候选态

Figure 3 (Color online) Cell structure of long short time neural network. C_{t-1} , cellular state output from the upper cellular unit; C_t , cellular state output from the current cellular unit; h_{t-1} , upper state output from the upper cellular unit; h_t , output from the current cellular unit; x_t , memory output from the upper cellular unit; f_t , forget gate; i_t , input gate; O_t , output gate; σ , Sigmoid activation function; $\tanh$, hyperbolic tangent activation function; $\tilde{C}_t$, candidate states

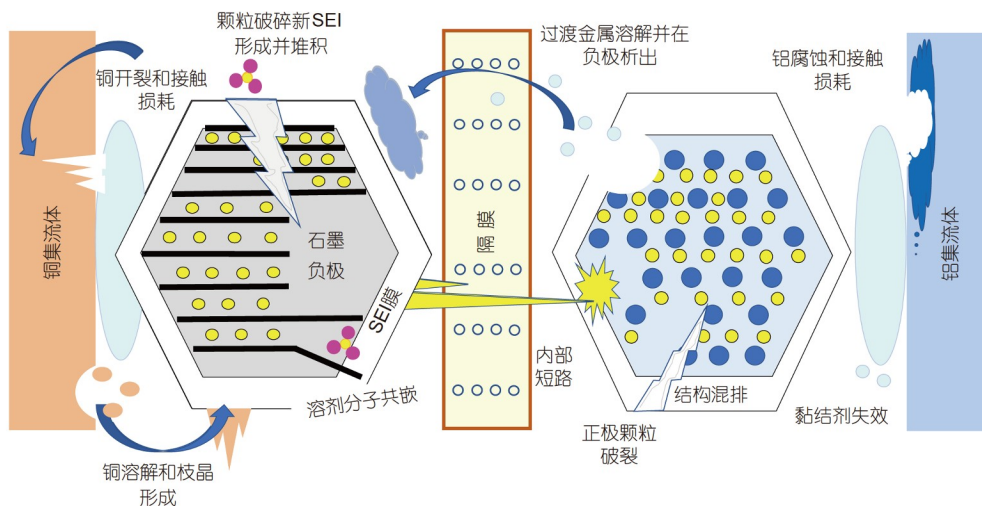


图 4 (网络版彩色)锂离子电池内部老化衰减机制

Figure 4 (Color online) Internal ageing decay mechanisms of lithium-ion batteries

略对电池放电容量和循环次数的影响, 提高模型对不同充放电倍率电池的普适性, Roman课题组^[53]使用机器学习管道技术建立了量化电池健康状态估计模型, 实验首先将179块电池分为3组, 每组遵循不同的充电协议, 然后使用对抗性示例技术扩充数据样本, 接着在基于随机森林的递归特征消除交叉验证方法下筛选了最优特征。最后, 基于随机森林和DNN方法都实现了短时间内的低误差估计, 模型可用于电池在线健康诊断。

3.3 基于数据驱动建模的电池剩余使用寿命估计

目前对电池剩余使用寿命的研究主要集中在数据驱动的机器学习领域, 实验通常从处理样本数据和优化训练模型两个方面减少模型复杂程度, 提高电池剩余使用寿命预测精度。样本数据的处理一般采用降维、特征选择或其他技术。自动编码器(autoencoder, AE)是一种有效的降维和异常检测技术, 在Top-N推荐任务

中取得了优异表现^[54]。鉴于此,一种集成了AE的深度神经网络方法实现了剩余使用寿命的预测,实验选用NASA提供的锂离子电池数据集,并将21维初始设计的特征数据降低至15维,模型对比了基于线性回归、贝叶斯回归和支持向量机方法的训练模型,验证了其有效性^[55]。包裹式特征选择方法以学习器的评价指标作为优化特征子集的评判标准,魏学哲课题组^[56]在基于电池弛豫电压特征变量的容量估计方法中,首先从CALCE电池充放电性能测试数据集中的搁置处提取特征,然后将特征数据输入到XGBoost学习器,接着根据评价指标选择包含了3个特征的特征子集。最后,实验使用随机森林算法实现了1.1%的低误差精度,模型能不依赖电池的历史信息、测试条件和电压区间,准确完成电池容量的估计,为电动车电池管理系统在线预测RUL提供了有力保障。重要性采样(importance sampling, IS)是蒙特卡洛中的一种采样策略,其主要目的是选择合适的分布,增加重要样本的相对频率,最终在一定采样数量上提高估计能力,减少方差^[57]。有研究使用IS选择的样本点提高了模型RUL的估计能力,实验从IFP1865140圆柱电池恒流充电升压曲线处和平台区获取了11个采样点,同时收集相同数量的均匀采样点完成对比实验。实验结果显示,相较于均匀采样训练出来的模型,IS下的前馈神经网络预测精度提高了27.3%^[58],具有更好的性能。经验模态分解(empirical mode decomposition, EMD)与传统信号处理方式中的傅里叶变换相比,能更有效处理非线性、非平稳时间序列信号,其本质过程是将数据序列或信号分解成不同时间尺度下的本征模态函数(intrinsic mode function, IMF)和残差,从而实现信号平稳化^[59]。为此,一种基于EMD的长短时神经网络和高斯过程回归(Gaussian process regression, GPR)的混合模型被用于RUL的预测。在数据预处理阶段中,首先从NASA和CALCE数据集电池数据中提取原始电池容量数据,其次将其分解成两个IMF和残差,最后在容量估计模块中分别由GPR和LSTM算法完成长期信号和再生信号的预测,训练模型获得了RMSE为0.0032的估计精度。另外,实验分别对比了GPR与LSTM方法和基于EMD的LSTM与GPR模型,验证了所建立模型的自适应和有效性^[60]。从优化模型角度可以发现,研究集中在寻找最优超参数及优化机器学习算法,从而实现缩短模型收敛时间,提升模型稳健性和准确性。PSO是一种进化算法,通过迭代找到最优解。对此,在LSTM的基础上,使用PSO找到最优超参数,缩短

模型训练时长,同时引入注意力机制。与不使用PSO的LSTM训练模型相比,误差降低了61.5%^[61]。不同于以往研究中使用CNN实现电池状态参数估计^[62-65]。一种新型混合CNN-LSTM方法被用于RUL的预测,训练模型在NASA电池数据集下实现了低误差精度,实验通过对比LSTM方法验证了CNN-LSTM的有效性^[66]。此外,Zraibi等人^[67]建立了CNN-LSTM-DNN的混合模型,实验通过CNN提取空间特征,LSTM提取时间特征;最后,将LSTM的输出作为DNN的训练样本实现RUL的预测。

4 总结与展望

本文回顾了传统模型与数据驱动的机器学习模型在BMS中SOC、SOH和RUL的研究进展。通过综述传统模型在电池状态参数的估计,发现其所面临问题与挑战,即电化学模型虽然能够模拟出锂离子电池的反应过程且模型精准度较高,但求解偏微分方程的复杂度高,高额计算量限制了电化学模型在实际工程中的应用。等效电路模型虽然计算简单,原理也更容易被理解,但估计精度却不如电化学模型。融合模型集成了不同模型之间的优点,却同时引入了新的错误源。

基于数据驱动的机器学习模型不依赖电池内部反应机理,通过非线性和线性映射来反映电池的动态特性而被广泛应用于电池状态参数估计任务中。虽然机器学习作为一种黑箱模型缺乏物理可解释性且估计结果依赖训练集的质量,但仍不可忽视其在电池状态参数估计领域产生的影响,从数据驱动的锂离子电池状态估计研究现状来看,未来工作应该从以下几个方面开展。

(1) 改进当前高精度模型,跟踪机器学习前沿技术,部署高性能算法完成电池状态参数估计。研究表明长短时神经网络在状态参数估计中有较好的性能,因此对基于长短时神经网络的改进依然需要深入研究。同时要加大对人工智能前沿技术的关注,Transformer在计算机视觉和自然语言处理领域取得了优异的成绩,并在太阳能发电数据集中完成了高精度的发电量预测,但在电池状态参数估计领域却缺少相关文章的报道,未来需要相关领域研究人员从事该方向的研究。

(2) 开发适用于多任务处理的机器学习模型。从机器学习在状态参数估计中使用的数据集及选用特征可以看出,通常会采用相同特征(如电流、电压和温度)下的数据作用在参数估计的模型训练中,但从研究策略上来看,一般采用单模型单状态参数预测的研

究方法. 因此, 结合过往研究取得的成果和积累的经验, 具有高精度的多参数预测模型应该被进一步开发.

(3) 设计电化学与机器学习混合的模型. 电池内部老化等不确定因素会对电池性能产生干扰, 然而当前机器学习在状态估计过程中往往不能反映这种内部机理. 传统模型中的电化学模型从电池自身出发能够反映内部老化机制, 但计算复杂, 泛化能力弱, 并不适用于在线监测. 因此, 未来工作中融合电化学模型后对预

测结果具有解释性的机器学习模型应该引起科研人员的广泛思考.

(4) 整合数据, 规范数据存储标准, 搭建数据库. NASA、斯坦福-丰田研究中心和CALCE相继公开了锂离子电池数据集, 即便如此, 能够为大众科研工作者使用的数据仍然有限. 因此, 构建合理的表间关系、设计科学的表内字段来完成数据库的搭建将有助于完成相关数据集的扩充, 促进状态参数估计领域的发展.

致谢 感谢Tamina Perveen(Quaid-i-Azam University)对我们工作提供的帮助.

参考文献

- Peng J, Zhou Z, Wang J, et al. Residual remaining useful life prediction method for lithium-ion batteries in satellite with incomplete healthy historical data. *IEEE Access*, 2019, 7: 127788–127799
- Wang C F, Cao F, Li M J, et al. Research status and future development of thermal management system for new energy vehicles under the background of carbon neutrality (in Chinese). *Chin Sci Bull*, 2021, 66: 4112–4128 [王丛飞, 曹锋, 李明佳, 等. 碳中和背景下新能源汽车热管理系统研究现状及发展趋势. 科学通报, 2021, 66: 4112–4128]
- Kim T, Song W, Son D Y, et al. Lithium-ion batteries: Outlook on present, future, and hybridized technologies. *J Mater Chem A*, 2019, 7: 2942–2964
- Hu X, Feng F, Liu K, et al. State estimation for advanced battery management: Key challenges and future trends. *Renew Sustain Energy Rev*, 2019, 114: 109334
- Sidhu M S, Ronanki D, Williamson S, et al. State of charge estimation of lithium-ion batteries using hybrid machine learning technique. In: 45th Annual Conference of the IEEE Industrial Electronics Society. 2019. 2732–2737
- Doyle M, Fuller T F, Newman J. Modeling of galvanostatic charge and discharge of the lithium/polymer/insertion cell. *J Electrochem Soc*, 1993, 140: 1526–1533
- Ramadass P, Haran B, White R, et al. Mathematical modeling of the capacity fade of Li-ion cells. *J Power Sources*, 2003, 123: 230–240
- Abada S, Marlair G, Lecocq A, et al. Safety focused modeling of lithium-ion batteries: A review. *J Power Sources*, 2016, 306: 178–192
- Moura S J, Chaturvedi N A, Krstic M, et al. PDE estimation techniques for advanced battery management system-part II: SOH identification. In: 2012 American Control Conference. 2012. 566–571
- Moura S J, Krstic M, Chaturvedi N A, et al. Adaptive PDE observer for battery SOC/SOH estimation. In: 5th Annual Dynamic Systems and Control Division Conference/11th JSME Motion and Vibration Conference. 2012. 101
- Hu X S, Feng F, Liu K L, et al. State estimation for advanced battery management: Key challenges and future trends. *Renew Sust Energ Rev*, 2019, 114: 109334
- Bhangu B S, Bentley P, Stone D A, et al. Nonlinear observers for predicting state-of-charge and state-of-health of lead-acid batteries for hybrid-electric vehicles. *IEEE Trans Veh Technol*, 2005, 54: 783–794
- Xiong R, Sun F C, He H W. Data-driven state-of-charge estimator for electric vehicles battery using robust extended Kalman filter. *Int J Automot Technol*, 2014, 15: 89–96
- Sun F, Xiong R, He H. Estimation of state-of-charge and state-of-power capability of lithium-ion battery considering varying health conditions. *J Power Sources*, 2014, 259: 166–176
- Wang S, Verbrugge M, Wang J S, et al. Multi-parameter battery state estimator based on the adaptive and direct solution of the governing differential equations. *J Power Sources*, 2011, 196: 8735–8741
- Zheng F, Xing Y, Jiang J, et al. Influence of different open circuit voltage tests on state of charge online estimation for lithium-ion batteries. *Appl Energy*, 2016, 183: 513–525
- Sun F, Xiong R, He H. A systematic state-of-charge estimation framework for multi-cell battery pack in electric vehicles using bias correction technique. *Appl Energy*, 2016, 162: 1399–1409
- Rahimi-Eichi H, Ojha U, Baronti F, et al. Battery management system: An overview of its application in the smart grid and electric vehicles. *IEEE Ind Electron Mag*, 2013, 7: 4–16

- 19 Yu Q Q, Xiong R, Lin C. Online estimation of state-of-charge based on the H infinity and unscented Kalman filters for lithium ion batteries. *Energy Procedia*, 2017, 105: 2791–2796
- 20 Linghu J, Kang L, Liu M, et al. Estimation for state-of-charge of lithium-ion battery based on an adaptive high-degree cubature Kalman filter. *Energy*, 2019, 189: 116204
- 21 Liao L, Kottig F. Review of hybrid prognostics approaches for remaining useful life prediction of engineered systems, and an application to battery life prediction. *IEEE Trans Rel*, 2014, 63: 191–207
- 22 Bian X, Wei Z, He J, et al. A two-step parameter optimization method for low-order model-based state-of-charge estimation. *IEEE Trans Transp Electr*, 2021, 7: 399–409
- 23 Galatro D, Da Silva C, Romero D A, et al. Battery health diagnosis approach integrating physics-based modeling with electrochemical impedance spectroscopy. *Energy Technol*, 2022, 10: 2100942
- 24 Li J, Lotfi N, Landers R G, et al. A single particle model for lithium-ion batteries with electrolyte and stress-enhanced diffusion physics. *J Electrochem Soc*, 2017, 164: A874–A883
- 25 Li X, Wang H M, Gu B, et al. Data sparseness in linear SVM. In: 1st International Workshop on Social Influence Analysis/24th International Joint Conference on Artificial Intelligence. 2015. 3628–3634
- 26 Zhang S, Li X, Zong M, et al. Efficient kNN classification with different numbers of nearest neighbors. *IEEE Trans Neural Netw Learn Syst*, 2018, 29: 1774–1785
- 27 Kotsiantis S B. Decision trees: A recent overview. *Artif Intell Rev*, 2013, 39: 261–283
- 28 Biau G. Analysis of a random forests model. *J Mach Learn Res*, 2010, 13: 1063–1095
- 29 Celebi M E, Kingravi H A, Vela P A. A comparative study of efficient initialization methods for the *k*-means clustering algorithm. *Expert Syst Appl*, 2013, 40: 200–210
- 30 Rumelhart D E, Hinton G E, Williams R J. Learning internal representation by back-propagation errors. *Neural Netw*, 1986, 323: 533–536
- 31 Schmidhuber J. Deep learning in neural networks: An overview. *Neural Netw*, 2015, 61: 85–117
- 32 Shandiz M A, Gauvin R. Application of machine learning methods for the prediction of crystal system of cathode materials in lithium-ion batteries. *Comput Mater Sci*, 2016, 117: 270–278
- 33 Joshi R P, Eickholt J, Li L, et al. Machine learning the voltage of electrode materials in metal-ion batteries. *ACS Appl Mater Interfaces*, 2019, 11: 18494–18503
- 34 Liu B, Yang J, Yang H L, et al. Rationalizing the interphase stability of Li doped-Li₇La₃Zr₂O₁₂ via automated reaction screening and machine learning. *J Mater Chem A*, 2019, 7: 19961–19969
- 35 Sarkar T, Sharma A, Das A K, et al. A neural network based approach to predict high voltage Li-ion battery cathode materials. In: 2nd International Conference on Devices, Circuits and Systems, 2014
- 36 Wang G, Fearn T, Wang T, et al. Machine-learning approach for predicting the discharging capacities of doped lithium nickel-cobalt-manganese cathode materials in Li-ion batteries. *ACS Cent Sci*, 2021, 7: 1551–1560
- 37 Jain A, Zongker D. Feature selection: Evaluation, application, and small sample performance. *IEEE Trans Pattern Anal Machine Intell*, 1997, 19: 153–158
- 38 Shah S M S, Shah F A, Hussain S A, et al. Support vector machines-based heart disease diagnosis using feature subset, wrapping selection and extraction methods. *Comput Electr Eng*, 2020, 84: 106628
- 39 Fu G H, Wu Y J, Zong M J, et al. Hellinger distance-based stable sparse feature selection for high-dimensional class-imbalanced data. *BMC Bioinf*, 2020, 21: 121
- 40 Zhang Y, Tang Q, Zhang Y, et al. Identifying degradation patterns of lithium ion batteries from impedance spectroscopy using machine learning. *Nat Commun*, 2020, 11: 1706
- 41 Babaeiyazdi I, Rezaei-Zare A, Shokrzadeh S. State of charge prediction of EV Li-ion batteries using EIS: A machine learning approach. *Energy*, 2021, 223: 120116
- 42 Bhattacharjee A, Verma A, Mishra S, et al. Estimating state of charge for xEV batteries using 1D convolutional neural networks and transfer learning. *IEEE Trans Veh Technol*, 2021, 70: 3123–3135
- 43 Hannan M A, How D N T, Lipu M S H, et al. SOC estimation of Li-ion batteries with learning rate-optimized deep fully convolutional network. *IEEE Trans Power Electron*, 2020, 36: 7349–7353
- 44 Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput*, 1997, 9: 1735–1780
- 45 Chemali E, Kollmeyer P J, Preindl M, et al. Long short-term memory networks for accurate state-of-charge estimation of Li-ion batteries. *IEEE Trans Ind Electron*, 2018, 65: 6730–6739
- 46 Chen Z, Zhao H, Shu X, et al. Synthetic state of charge estimation for lithium-ion batteries based on long short-term memory network modeling and adaptive H-infinity filter. *Energy*, 2021, 228: 120630

- 47 Bian C, He H, Yang S. Stacked bidirectional long short-term memory networks for state-of-charge estimation of lithium-ion batteries. *Energy*, 2020, 191: 116538
- 48 Shu X, Li G, Shen J, et al. A uniform estimation framework for state of health of lithium-ion batteries considering feature extraction and parameters optimization. *Energy*, 2020, 204: 117957
- 49 Zhang Z, Li L, Li X, et al. State-of-health estimation for the lithium-ion battery based on gradient boosting decision tree with autonomous selection of excellent features. *Int J Energy Res*, 2022, 46: 1756–1765
- 50 Li W, Sengupta N, Dechent P, et al. Online capacity estimation of lithium-ion batteries with deep long short-term memory networks. *J Power Sources*, 2021, 482: 228863
- 51 Vetter J, Novák P, Wagner M R, et al. Ageing mechanisms in lithium-ion batteries. *J Power Sources*, 2005, 147: 269–281
- 52 Shen S, Sadoughi M, Chen X, et al. A deep learning method for online capacity estimation of lithium-ion batteries. *J Energy Storage*, 2019, 25: 100817
- 53 Roman D, Saxena S, Robu V, et al. Machine learning pipeline for battery state-of-health estimation. *Nat Mach Intell*, 2021, 3: 447–456
- 54 Wu Y, Dubois C, Zheng A, et al. Collaborative denoising auto-encoders for top-N recommender systems. In: 9th Annual ACM International Conference on Web Search and Data Mining, 2016. 153–162
- 55 Ren L, Zhao L, Hong S, et al. Remaining useful life prediction for lithium-ion battery: A deep learning approach. *IEEE Access*, 2018, 6: 50587–50598
- 56 Zhu J, Wang Y, Huang Y, et al. Data-driven capacity estimation of commercial lithium-ion batteries from voltage relaxation. *Nat Commun*, 2022, 13: 2261
- 57 Sadowsky J S. A new method for Viterbi decoder simulation using importance sampling. *IEEE Trans Commun*, 1990, 38: 1341–1351
- 58 Wu J, Zhang C, Chen Z. An online method for lithium-ion battery remaining useful life estimation using importance sampling and neural networks. *Appl Energy*, 2016, 173: 134–140
- 59 Santhosh M, Venkaiah C, Vinod Kumar D M. Ensemble empirical mode decomposition based adaptive wavelet neural network method for wind speed prediction. *Energy Convers Manage*, 2018, 168: 482–493
- 60 Liu K, Shang Y, Ouyang Q, et al. A data-driven approach with uncertainty quantification for predicting future capacities and remaining useful life of lithium-ion battery. *IEEE Trans Ind Electron*, 2021, 68: 3170–3180
- 61 Qu J, Liu F, Ma Y, et al. A neural-network-based method for RUL prediction and SOH monitoring of lithium-ion battery. *IEEE Access*, 2019, 7: 87178–87191
- 62 Qian C, Xu B, Chang L, et al. Convolutional neural network based capacity estimation using random segments of the charging curves for lithium-ion batteries. *Energy*, 2021, 227: 120333
- 63 Yang Y. A machine-learning prediction method of lithium-ion battery life based on charge process for different applications. *Appl Energy*, 2021, 292: 116897
- 64 Saxena S, Ward L, Kubal J, et al. A convolutional neural network model for battery capacity fade curve prediction using early life data. *J Power Sources*, 2022, 542: 231736
- 65 Saxena S, Ward L, Kubal J, et al. A convolutional neural network model for battery capacity fade curve prediction using early life data. *J Power Sources*, 2022, 542: 231736
- 66 Ren L, Dong J, Wang X, et al. A data-driven Auto-CNN-LSTM prediction model for lithium-ion battery remaining useful life. *IEEE Trans Ind Inf*, 2021, 17: 3478–3487
- 67 Zraibi B, Okar C, Chaoui H, et al. Remaining useful life assessment for lithium-ion batteries using CNN-LSTM-DNN hybrid method. *IEEE Trans Veh Technol*, 2021, 70: 4252–4261
- 68 Chemali E, Kollmeyer P J, Preindl M, et al. State-of-charge estimation of Li-ion batteries using deep neural networks: A machine learning approach. *J Power Sources*, 2018, 400: 242–255
- 69 Anton J C A, Nieto P J G, Viejo C B, et al. Support vector machines used to estimate the battery state of charge. *IEEE Trans Power Electron*, 2013, 28: 5919–5926
- 70 Xue Z, Zhang Y, Cheng C, et al. Remaining useful life prediction of lithium-ion batteries with adaptive unscented Kalman filter and optimized support vector regression. *Neurocomputing*, 2020, 376: 95–102
- 71 Patil M A, Tagade P, Hariharan K S, et al. A novel multistage support vector machine based approach for Li ion battery remaining useful life estimation. *Appl Energy*, 2015, 159: 285–297

Summary for “数据驱动的锂离子电池全生命周期状态参数评估”

Data-driven full life-cycle state parameter assessment of Li-ion batteries

Jie Liu¹, Zongcheng Miao^{2*} & Qingyun Wang³

¹ Xi'an Key Laboratory of Advanced Photo-electronics Materials and Energy Conversion Device, Xijing University, Xi'an 710123, China;

² School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an 710072, China;

³ The Second Institute of China Aerospace Science and Technology Corporation, Beijing 100854, China

* Corresponding author, E-mail: miaozongcheng@nwpu.edu.cn

The battery management system (BMS) is a vital link between the power battery, the onboard system, and the engine. Adjusting the output power depending on the current driving situation of the vehicle and the surrounding environment ensures the lithium-ion battery's safe, stable, and effective operation. It is essential to optimize the charge and discharge rate of the cell and prolong its lifetime. The battery management system has three key estimation parameters—State of charge (SOC), state of health (SOH), and remaining useful life (RUL)—That have key reference values for the battery's current use. As the above state characteristics cannot be computed directly from the battery, a broad and highly predictive approach must be developed to estimate the condition parameters. This review analyzes the development of the application of battery state parameter estimation, explains the model concepts, compares the differences between models, and explores illustrative typical applications. Patterns are categorized into traditional models and data-driven machine learning models in this review.

The traditional model's electrochemical model provides a highly accurate description of the internal behavior of the battery through the battery reaction mechanism and explains the charge transfer process between electrodes from a chemical theory perspective. However, the electrochemical model is not suited for online state parameter estimation due to its maximum complexity and high computational stress among the standard models. The equivalent circuit model, which employs circuit components like resistors and capacitors to describe the dynamic features of the battery, is based on a simpler idea than the complicated partial differential equations that are used to describe the battery's fundamental electrochemical operations. However, because of the low state parameter estimation precision, it is not compatible with emerging battery management systems with high state parameter estimation accuracy. To enhance the performance of the model, the fusion model integrates the features of different models. Though, the application of fusion models for battery management systems needs more research due to the inclusion of models with additional sources of error and the challenge of creating failure criteria for the fused models. The commonly used machine learning algorithms, such as a k -nearest neighbor, support vector machine, random forest, and artificial neural network models, are briefly introduced in this review. It also provides a systematic overview of the model training process from data pre-processing to model evaluation. This review also offers a comprehensive overview of the whole model training process, from data pre-processing through algorithm selection to model performance assessment, and it lists the variables to be considered when choosing an algorithm. Finally, this review discusses the development of data-driven machine learning models for state parameter estimation and examines the estimated performance of models created by the techniques employed in representative papers. The research suggests that long short-term memory network-based models have better estimation performance for estimating battery condition parameters. The data-driven machine learning model is not ideal and still requires improvement since the model that the machine learning algorithm learned is a black box model lacks interpretation of the results and places strict constraints on the source and quality of data.

On this basis, it is proposed that existing or more sophisticated machine learning models are enhanced and their estimation accuracy and stability are evaluated. In order to achieve joint estimation models with high accuracy and scalability, machine learning models for multi-tasking also need to be developed. Furthermore, to overcome the unpredictability of machine learning for the internal aging of batteries, it is also advised to create fusion models using a combination of electrochemical models and computer algorithms. To increase the amount of data available to researchers and aid in the advancement of artificial intelligence in the field of state parameter estimation, it is also advised to integrate experimental data, standardize data storage standards, and create databases.

lithium-ion battery, data-driven, machine learning, parameter evaluation

doi: [10.1360/TB-2022-0550](https://doi.org/10.1360/TB-2022-0550)