

# 基于机器学习流水线的电池健康状态估计方法

Darius Roman<sup>1</sup>, Saurabh Saxena<sup>2, 4</sup>, Valentin Robu<sup>1, 3, 5</sup>, Michael Pecht<sup>2</sup>, and David Flynn<sup>1</sup>

<sup>1</sup>Smart Systems Group, School of Engineering & Physical Sciences, Heriot-Watt University, Edinburgh, EH14 4AS, UK.

<sup>2</sup>Center for Advanced Life Cycle Engineering, University of Maryland, College Park, MD 20742, USA.

<sup>3</sup>Algorithmics Group, TU Delft, Delft, the Netherlands.

<sup>4</sup>Present address: Argonne National Laboratory, Lemont, USA.

<sup>5</sup>Present address: Intelligent and Autonomous Systems Group, CWI, National Research Centre for Mathematics and Computer Science, Amsterdam, the Netherlands.

## 摘要

从便携电子设备到电动汽车, 锂离子电池在各种场景中的应用越发广泛。无论哪种应用场景, 都要求通过设备机载控制器对电池健康状态 (state of health, SOH) 进行实时可靠地估计以保证电池安全运行并最终保证系统的完整性和可靠性。实际应用中通常使用电池容量作为电池性能退化指标, 本文设计并评估了一种基于机器学习流水线的电池容量估计方法, 并在 179 块电池 (分别在不同工况下工作) 的循环数据上进行测试。本文的模型使用了两个参数化方法和两个非参数化方法, 实现了电池 SOH 估计并给出估计结果的置信区间。机器学习流水线模型从充电过程中的部分电压和电流数据中提取了三十个特征并进行自动特征选择和算法标定步骤。当把机器学习流水线模型部署在使用快充策略的电池数据集上时, 算法达到均方根误差 (root mean squared error, RMSE) 仅 0.45% 的预测精度。我们的方法为设计灵活可拓展的数据驱动的电池 SOH 估计模型提供了深刻理解, 同时强调了预测置信区间的重要性。我们提出的机器学习流水线方法结合了实验数据和机器学习建模, 可应用于其他需要实时估计 SOH 的关键部件。

## 1 简介

可充电锂离子电池在许多应用领域中都扮演着重要角色, 包括便携式电子设备、医疗设备、电网的可再生能源消纳环节和电动汽车等。在 2010 年至 2016 年期间, 锂离子电池价格陡降 73%, 在 2017 年达到历史最低点, 为每千瓦时 273 美元 [1], 这为估值高达 650 亿美元的巨大的储能市场打开了大门 [2]。不管应用场景如何, 锂离子电池随着时间而老化, 且在老化过程中, 电池呈现出容量降低、内阻增大的趋势。电池的老化速度受动态运行条件影响, 包括不同的充放电速率、不同的运行电压限制和运行过程中的温度变化。能够不依赖具体机理、使用数据驱动方法实现实时电池退化状态估计对是安全、有效的电池管理系统中非常重要的环节 [3]。电池健康状态可以用来预测电池的预期寿命, 但是通过直接测量电池内部电化学反应参数来进行在线电池健康状态估计的方法目前仍十分有限 [4]。

SOH 是一个用于量化电池的一般状态和与未使用状态相比提供指定性能能力的指标, 这个性能可以用容量或阻抗来衡量。我们选择电池的容量作为其健康指标, 因为容量与电池的能量储存能力有关, 直接影响电池剩余运行时间和寿命。容量衰减估计这一问题已经引起了工业界和学术界的广泛关注 [4–7], 大量的方法已经被提出。现有的实现电池容量退化估计的方法主要包括基于电化学模型的方法 [8–10]、基于等效电路的方法 [11–13] 和数据驱动的方法 [14–19]。电化学模型近似描述电池在操作期间内发生的化学过程, 使用这种方法要求对电池工作机理 (如电极材料和电解质的性质) 的非常详尽的认识, 同时该方法通常利用复杂的偏

微分方程, 使得其对存储空间和计算能力的要求很高。等效电路模型使用具有经验非线性参数的电路组件取代实际电池元件 [9]。与电化学模型相比, 等效电路模型使用较少的输入, 大大减少了需要学习的参数数量。但这种方法的预测性能很大程度上取决于对“等效”假设的准确性。事实上, 假设往往难以拟合实际工作情况, 导致等效电路模型准确性和鲁棒性有限 [8]。此外, 为了在不同的电池荷电状态值下确定等效电路模型参数 (例如欧姆阻抗和并联电阻-电容阻抗), 通常需要进行脉冲放电 [20] 和电化学阻抗谱分析 [10, 21, 22], 然而, 这些测量并不适用于在线应用。

相反地, 数据驱动的方法有着诸多好处, 包括 1) 这种方法与具体化学机理无关和 2) 这种方法能够分析更完整的退化过程, 如在小负载工况下的容量退化过程 (这种情形往往会被上述两种方法忽略)。到目前为止, 许多研究已经使用机器学习方法进行电池健康状态估计的分析。增量容量分析和差分电压分析是分析电池老化机理的方法 [23]。几项研究表明 [23–27], 这类方法可以用于离线和在线电池容量性能衰退估计。然而, 这类方法有几个缺点, 为了得到增量容量曲线, 需要对容量-电压曲线进行微分, 这会放大噪音并将其传播到算法中。此外, 这类方法工作的前提是两条曲线覆盖足够的电压范围, 并限制在低充电电流率 (1/5 到 1/25 倍率) 下 [28–30], 以获得高保真曲线。只有当充电时使用低电流值并准确记录容量-电压曲线的关键部分, 才能捕捉到增量容量曲线中的特定峰值点并据此进行在线的电池容量估计。

为了解决上述问题, 一种替代方法是使用原始的电压-时间数据作为算法输入进行训练, 从而消除微分操作 [31, 32]。值得注意的是 Richardson 等人提出了一种方法 [32], 这种方法首先对电压-时间数据进行 Savitsky-Golay 滤波使之变得平滑, 然后将等间距电压值作为高斯过程回归 (GPR) 算法的输入。但是, GPR 的训练过程消耗很多的计算资源 (计算复杂度主要取决于核函数决定) [33], 使得它并不适用于在线电池容量估计场景, 此外, 高计算复杂性使得这一算法很难向大规模电池退化数据集拓展, 最后, GPR, 对输入的电压-时间曲线的部分非常敏感。其他基于贝叶斯的方法, 例如相关向量机 (RVM) [34], 也已用于估计电池容量退化, 但是 RVM 也同样面临因训练速度慢而难以应用的问题, 尤其是当与其他频率主义方法相比时 [35]。Shen [33] 等人提出了加速 GPR 训练过程的方法, 代价是牺牲预测精度。与文献 [32] 使用充电曲线的恒流充电段策略不同, Wang 等人 [36] 使用了恒压充电段作为输入, 通过支持向量回归 (SVR) 来估计容量衰减。SVR 确实比 GPR 速度快, 但这种方法并不能给出预测结果的不确定度。应用上往往需要结合估计结果和结果的不确定度量以提供足够的信息进行决策, SVR 在不确定度估计方面的缺失是难以将它应用在复杂动态系统 (如锂离子电池) 寿命预测的主要原因 [37]。

过去的研究或对 SOH 估计的不确定度评估有限, 或根本没有进行评估 [8–36]。我们提出的模型 (battery health and uncertainty management pipeline, 以下简称 BHUMP), 能够实时电池健康状态估计并给出估计结果的不确定度。BHUMP 对获得的数据流进行分层操作, 首先, 在原始充电曲线分段上进行特征工程研究, 继而算法自动进行离线特征提取、通过对抗样本 (人为添加噪声诱导模型生成错误结果) 进行数据增广并借助四种机器学习算法得到电池健康状态及其不确定度的估计。具体地, 不确定性量化通过校准误差和改进的准确度度量 ( $\alpha$ - $\beta$  准确度区域) 实现。毫无疑问现实世界中电池有不同的设计 [38], 使用不同材料 [39], 为了验证模型的泛化性能, 我们在 179 块电池上对算法进行评估, 这些电池采用不同的结构形式 (圆柱电池、方形电池或软包电池)、不同正极材料 (磷酸铁锂或钴酸锂)、不同充电协议 (恒流充电、恒流-恒压充电或两步快速充电协议) 和不同充放电倍率。

本文优化改进了以往的电池健康状态估计方法, 并针对我们提出的机器学习流水线模型进行了大量测试, 同时定义了用于电池健康状态估计及其不确定性度量问题中使用的机器学习模型中的不确定性的指标。BHUMP 为研究者提供了一套高鲁棒性的电池健康状态估计解决方案, 这套框架对各种不同化学机理的电池在各种工作条件下的状态预测都适用。BHUMP

建立电池模型并从中提取中特征，这些特征能够充分反映电池内部老化过程并估计电池健康状态，这一方法不仅能通过不超过 15 分钟的充放电循环特征实现预测，同时预测精度比以往的方法都要好。我们的算法输出准确的电池健康状态估计及其不确定性，这些输出将被进一步用于估计电池荷电状态和剩余寿命，从而指导电池充放电决策，改善电池性能并最终延长电池寿命。

## 2 机器学习流水线模型

站在机器学习的视角上，电池健康状态估计可以被视为多变量监督学习问题。我们引入一种基于流水线（pipeline）的方法，这种方法首先在电池充电/放电循环数据上实施特征工程特征工程，通过特征工程提取的强相关特征被用于训练机器学习模型（可以是贝叶斯模型或频率模型），最后通过估计预测结果的分布的均值和标准差量化预测结果的不确定度。我们的学习模型包括两个阶段，阶段一为离线流水线构建和训练阶段，阶段二为在线电池健康状态估计阶段。具体地，离线阶段进行特征工程实施、训练数据增广、自动特征选择、回归模型训练和不确定性标定。在线阶段在未知电池容量的条件下使用训练好的机器学习流水线模型进行电池健康状态诊断。辅助材料中归纳了这两个阶段的流程图。

特征工程分为两种，一种是通过神经网络自动编码器等方法实现自动特征生成/提取 [40, 41]，另一种则是通过领域先验知识进行特征构建 [42, 43]。我们采用基于先验知识的特征建模方法，这种方法选择对目标值贡献最多的特征量。在辅助材料的第一部分，我们还提供了一个旨在揭示电池的部分充电过程是如何量化电池容量退化的潜在机理的假设。辅助材料中的表一总结了我们在电池循环测试中记录的各种测量量。我们的机器学习流水线模型关注充电曲线片段，通过部分充电段数据捕捉电池在充放电循环中的退化信息（图 1 展示了一种典型的片段选取方法）。被选中的充电曲线片段将被用于进行特征工程。

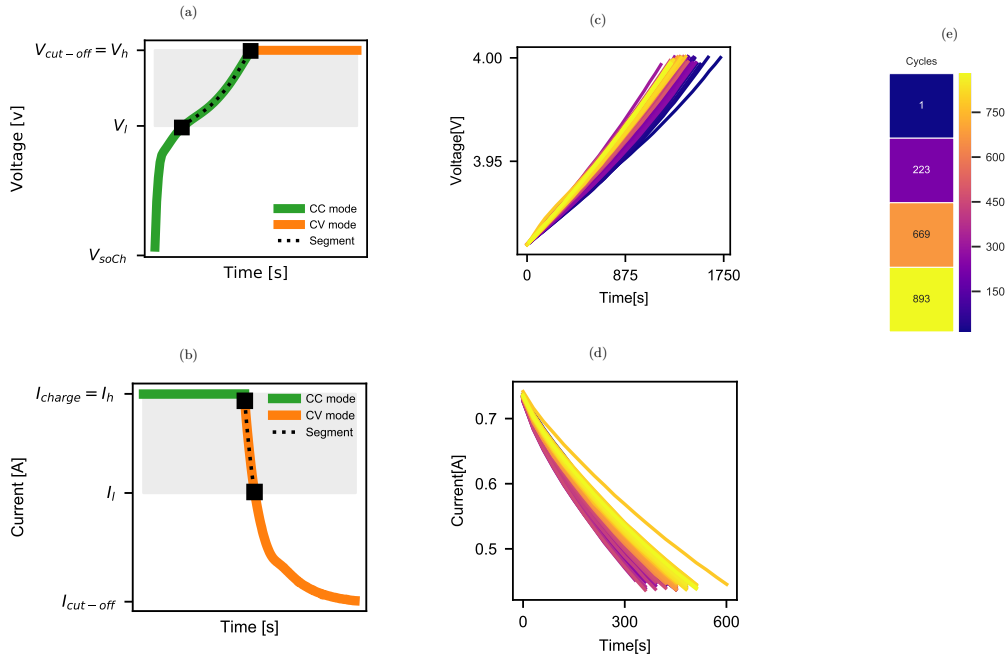


图 1: 软包电池恒定电压-恒定电流协议充放电示意图及其随时间分布曲线 **a** 充电过程中电压曲线, **b** 充电过程电流曲线, **c** 提取的电压曲线段随时间分布示意图, **d** 提取的电流曲线段随时间分布示意图, **e** 随循环圈数增加的老化程度热力图。

我们提出的流水线方法构造了 30 个特征，并使用和文献 [44] 提出的方法类似的基于随

机森林的递归特征消除交叉验证法 (RF-RFE-CV) 选出与待预测结果最相关的特征量 (递归特征消除法的总体性能胜过了其他传统方法 [45, 46], 这是我们选择它的原因)。在训练之前, 我们引入 Goodfellow 等人提出的对抗样本 [47] 进行数据增强, 使用对抗样本的原因是为了改善电池设计或化学成分差异带来的影响。此外, 使用对抗数据进行训练可以使算法对离群样本具有鲁棒性以防止过拟合, 并减少预测分布围绕其均值 (预测值) 的波动。以 Doyle 等人提出的锂电池准二维模型为例 [48], 很多电化学模型生成合成数据的过程也可以被视为一种数据增强过程。这种方法结合了电化学模型和基于数据的模型的潜力, 我们认为未来的研究必须同时整合合成数据。在训练过程中, 我们使用权重衰退法进行正则化以约束模型复杂度, 提高模型泛化能力。

增强的数据样本被作为四种算法的训练输入, 这些算法分别是: 随机森林 (random forest, RF) 和集成深度神经网络 (deep neural network ensemble, dNNe)、贝叶斯岭回归 (Bayesian ridge regression, BRR) 和高斯过程回归 (Gaussian Process Regression, GPR)。与基于贝叶斯主义的 BRR 和 GPR 不同, 频率学派的方法受限于其基本假设和表述形式无法给出预测结果不确定性度量。为了克服这种局限性, 我们考虑了两种改良的基于集成学习的算法: 基于“无穷小刀切法” (infinitesimal jackknife, IJ) 的置信区间估计 [49] 和神经网络集成学习 [50]。为了训练这些模型, 我们使用自适应矩估计 (Adaptive Moment Estimation, Adam) 优化器训练集成神经网络模型, 用随机搜索的方法调整余下所有模型的超参数 [51]。在训练过程中我们发现, 从均匀分布中抽取随机样本对 BRR 和 GPR 的参数训练来说效果最好, 而对 RF 和 dNNe 参数训练, 随机初始化可以得到令人满意的结果。此外, 在超参数调整过程中, 使用了批量交叉验证方法, 每个批次由一个单元代表, 这样的策略既防止了模型的过拟合, 同时模拟了在线部署的情形。工程上应用的机器学习模型需要从预测结果误差和针对预测结果的不确定性量化角度进行严格的性能评估, 为此, 我们先对模型进行重新校准, 然后根据平均绝对百分比误差、均方根误差和不确定性估计指标进行评估。

### 3 数据集

Group*	I	I	I	I	I	II	III
Dataset	CALCE CS2	CALCE CX2	CALCE PL	NASA 5	NASA11	TRI**	Oxford
Manufacturer	Unknown	Unknown	Unknown	LG Chem	LG Chem	A123 Systems	Kokam
Cathode ***	LiCoO <sub>2</sub>	LiCoO <sub>2</sub>	LiCoO <sub>2</sub>	LiCoO <sub>2</sub>	LiCoO <sub>2</sub>	LiFePO <sub>4</sub>	LiCoO <sub>2</sub> / LiNiMnCoO <sub>2</sub>
Form factor	Prismatic	Prismatic	Pouch	18650 Cylindrical	18650 Cylindrical	18650 Cylindrical	Pouch
# cells	6	6	2	8	25	124	8
Charge	CC-CV	CC-CV	CC-CV	CC-CV	CC-CV	Fast-charge	CC
Discharge	2 regimes	2 regimes	1 regime	2 regimes	7 regimes	1 regime	1 regime

\* Groups based on charge protocol, \*\*Toyota Research Institute, \*\*\*Information from manufacturer, not verified

表 1: 本文使用的数据集概述

我们在一个包含 179 颗电芯的数据集上测试 BHUMP 的性能, 对于数据集的描述参考表 1。这些电芯按照充电协议的不同被划分为三组, 具体地, 采用恒流-恒压 (constant current - constant voltage, CC-CV) 充电协议的电芯被划分进第一组 (47 颗电芯), 采用恒流 (constant current, CC) 充电协议的电芯被划分进第二组 (124 颗电芯), 采用两步快充 (2-step fast-charging) 协议的电芯被划分进第三组 (8 颗电芯)。这种划分很重要, 实施 BHUMP 需要依



据此分组进行特征选择、模型训练和模型评估。对于数据集的更详尽的描述可参考辅助材料中的第四部分。

## 4 算法性能

原文依据充电协议差异将电池划分为三组，三组性能分析内容几乎一致，这里只翻译第一组数据的算法性能评估结果。

### 4.1 第一组数据集上的实验

根据上述机器学习流水线模型的实施步骤，在进行特征工程后选择了 18 个特征作为本组数据的最优特征数量。对于阈值设置，我们将上限电压阈值  $V_h$  设置为 4.2V，下限电压阈值  $V_l$  设置为 3.9V。关于特征选择和训练集/验证集划分的细节请看补充材料。

	MAPE	RMSPE	$C_{score}$	$Sh$	$\alpha$ -accuracy	$\beta$	PEP
BRR	1.52	2.49	84.49	0.021	70.00	0.57	68.92
GPR	1.49	2.24	92.23	0.025	65.00	0.48	71.76
RF	0.72	0.91	100	0.046	92.00	0.29	95.29
dNNe	0.65	0.92	88.01	0.0082	93.00	0.93	97.71

表 2: 38 号电池的预测性能

如图 2 展示了当 dNNe 作为基算法时 BHUMP 模型在一块随机选择的软包电池（编号为 38）上的预测性能，表 2 展示了四种算法在 38 号电池上的性能。该电池在 CC-CV 充电协议下以 0.5C 倍率充电，然后在 4.2V 和 2.7V 之间以 0.5C 倍率进行全深度放电。

比较 dNNe 算法和其余三个算法（RF、BRR 和 GPR）的性能，我们发现前者生成的置信区间相对较小（所有图表均显示相当于 95% 分位数的置信水平，即  $\mu \pm 2\sigma$ ，其中  $\mu$  为预测均值），较小的置信区间表明算法具有更高的主准确度，表现为对应的  $\beta$  值较大。当预测不那么准确时（如当预测一颗电池前几个循环的健康状态时），误差带可以很好反映这种变化。在 38 号电池上，dNNe 算法 dNNe 也实现了最佳平均绝对百分比误差（MAPE）和均方根百分比误差（RMSPE），同时有很高的校准得分。如表 2 所示，38 号电池的预测值的 RMSPE 在 0.65% 到 1.52% 之间，表明四种算法都能实现较高的预测性能。然而，RF 算法校准后预测的分布仍表现出较大波动。

	MAPE	RMSPE	$C_{score}$	$Sh$	$\alpha$ -accuracy	$\beta$	PEP
BRR	4.65	5.54	89.16	0.104	25.76	0.25	36.57
GPR	3.70	4.51	83.62	0.089	32.04	0.29	60.07
RF	2.17	2.70	54.70	0.093	35.94	0.36	65.47
dNNe	3.30	4.26	86.28	0.043	32.14	0.58	63.26

表 3: 第一组数据的平均预测性能

当讨论算法对第一组数据集中所有电池的平均预测性能时（表 3 总结了四种算法在第一组数据集上的平均性能），我们发现 RF 达到了相对较低的校准误差（54.70%），我们认为这归因于我们使用 IJ 方法来估计分布。在实践中，我们往往偏向保守系统，尤其是系统安全性非常重要时。这意味着对应预测值低于真实值（位于  $\alpha$  准确度区域中，如图 3）的样本应该比对应预测值高于真实值的样本多，换言之，适用于工业检测场景的算法的 PEP 指标应接近 100%。同时，容量估计过低会导致算法过于保守。然而，这种行为通过 RMSPE 的增加自然被缓解。参考图 2c 和表 2，dNNe 是保守的，其 PEP 得分最高。

总体上，尽管 RF 算法在 RMSPE 和 MAPE 两个指标上得到了最高的分数（见表 3），但它不能输出良好校准的预测结果，预测的精度也偏低。相比之下，dNNe 模型输出的是一个

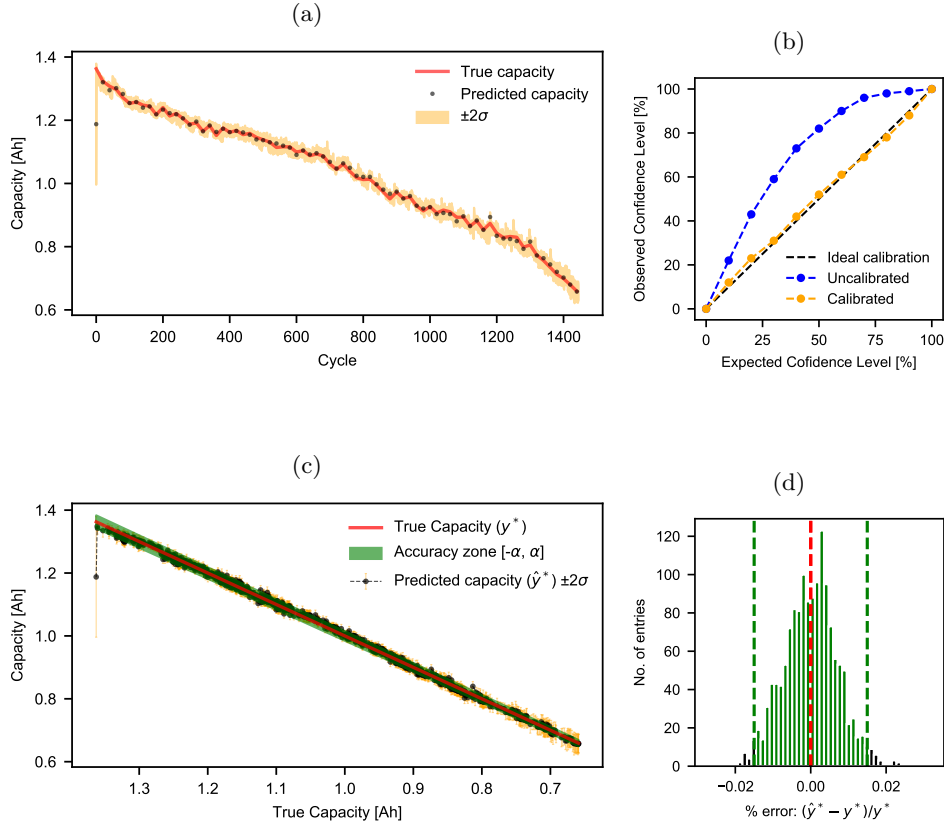


图 2: dNNe 作为基算法在 38 号电池上的预测结果 **a** dNNe 预测的容量-时间曲线, **b** dNNe 算法标定结果, **c** dNNe 预测值和真实值对照, **d** 百分比误差柱状图 (其中  $y^*$  为真实容量,  $\hat{y}^*$  为预测容量)。

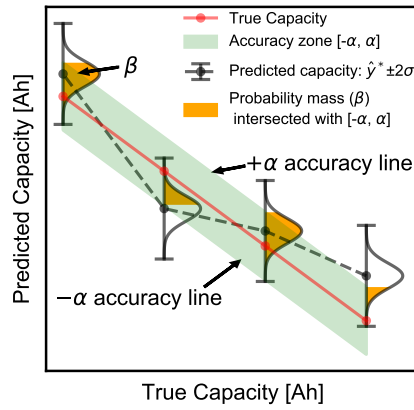


图 3:  $\alpha$ -准确性区域和  $\beta$  概率质量分布示意图

良好校准的预测结果, 平均校准得分仅比理想得分低不到 4%, 其 MAPE 和 RMSPE 指标分别为 1.13% 和 1.56%。

## 5 本文方法在实际场景中适用性的讨论

BHUMP 可以用于电池管理系统，取代传统的等效电路模型进行电池健康状态和电池荷电状态估计。传统方法依赖静态条件（如完全充放电）下的容量信息，而 BHUMP 可以根据充放电片段的测量量进行电池容量退化估计，对不同工况（如随机充放电情形和高倍率充放电情形）适应性更好。BHUMP 可以估计使用两步快充协议充电电池和采用随机放电（ARTEMIS 驾驶协议）情形的电池的容量退化情况，这两个过程对目前电动汽车电池组而言是典型操作。未来的工作可以将模型进一步拓展到采用不同充放电协议的开源电池数据集上 [19]。

充电过程中的温度变化会进一步将不确定性引入充电段数据的测量，并将其传播到预测算法中。可能的解决方案包括在训练 BHUMP 时将温度作为输入，或将其他现场感知信息如图像（光学图像/数字图像）或 X 射线数据 [52] 等加入输入，从而使得算法能够学习温度、生成的特征和 SOH 指标之间的相关性。这一问题事实上进一步强调了不确定性估计的重要性，单一的健康状态估计结果无法提供下一步决策所需要的完整信息。除了固有的算法偏差，数据集的变化也会影响预测准确度。为了适应数据中的这种变化，BHUMP 实施工程特征得到了 30 个特征量，并利用了无监督的特征选择算法（RF-RFE-CV）选择若干最相关的特征作为 BHUMP 输入。考虑到不同的数据集存在分布差异，我们认为深度学习有可能在未来超过其他算法，因为深度学习的过程几乎不需要用户干预，并且将深度学习模型部署在图形处理单元并结合现代数据存储方案，模型可以进行并行计算并利用与日俱增的算力。此外，当训练集包含的样本数量有限或训练数据与预期应用场景有出入时，可以使用迁移学习提高模型性能。目前针对深度学习开发的计算硬件、计算架构和新的学习算法将加速上述过程，使得相关算法得以在车辆上使用。更具体地说，以 dNNe 为基础的 BHMUP 可以在特定电池上训练，然后结合迁移学习，在不同设计的电池上依据早期衰退信息重新训练模型参数，实现对新电池的高精度预测。进一步的，在数据可用的前提下，BHUMP 可以结合主动学习（Active Learning, AL）被用于采用不同设计、不同化学材料和不同工作温度的电池中。

## 6 结论

目前两种被广泛采用的用于在线电池健康状态估计的方法是基于电池等效电路模型的方法和基于电化学模型的方法。然而，在在线应用中，要兼顾预测准确性和预测结果计算高效性并不容易。我们提出了一种新的基于机器学习流水线的方法，称为 BHUMP。与传统方法相比，我们提出的机器学习流水线有诸多优势，包括其对充电协议和放电倍率的鲁棒性，以及能其在缺少对电池设计、电池电极材料和工作温度的先验知识的条件下进行预测的能力。

本文研究了四种作为 BHUMP 基础的算法，分别是 BRR、GPR、RF 和 dNNe。

研究时关注这些算法的预测误差和量化预测结果不确定性的能力。实验结果表明，模型的精度（这里指的是在测试集上取得的最小误差）取决于样本对应的充电协议。对于采用恒流-恒压充电协议和两步快充协议的电池样本，RF 算法取得了最低的误差，对于采用恒流充电协议的电池样本，BRR 算法取得了最低误差。需要注意的是，在考虑不确定性评估指标时，RF 算法很难校准，并且在预测结果上表现出了很强的乐观倾向。以平均增加 0.43% 的 MAPE 和 0.97% 的 RMSPE 为代价，dNNe 算法的平均校准分数更高，且无论充电协议如何，都能取得第二低的误差。对于采用两步快充协议的电池，dNNe 算法性能最好，实现了 0.45% 的 RMSPE，参照 90% 的置信区间，校准得分为 91.02%。

综上，我们的工作强调了使用部分充电段作为输入，结合多种机器学习模型进行电池退化状态预测的方法。更重要地，我们强调了在实际工程应用中对预测结果进行不确定性量化的重要性。最后，我们证明了我们提出的机器学习流水线方法在计算效率和结果准确性方面的优势。我们认为，这种机器学习管道方法在未来的电池管理系统中将成为标准化应用。

## 7 数据可用性声明

数据集开放获取。具体地, 第一组数据可以在 <https://www.nasa.gov/content/prognostics-center-of-excellence-data-set-repository> 和 <https://web.calce.umd.edu/batteries/data.htm> 下载; 第二组数据可以在 <https://data.matr.io/1/projects/5c48dd2bc625d700019f3204> 下载; 第三组数据可以在 <https://ora.ox.ac.uk/objects/uuid:03ba4b01-cfed-46d3-9b1a-7d4a7bdf6fac> 下载。

## 8 代码可用性声明

源代码开放获取, 地址为 <http://doi.org/10.5281/zenodo.4390152>, 作者没有公开数据预处理和特征工程部分的代码。

## 参考文献

- [1] Curry, c. lithium-ion battery costs and market: squeezed margins seek technology improvements & new business models. bloomberg new energy finance <https://data.bloomberglp.com/bnef/sites/14/2017/07/bnef-lithium-ion-battery-costs-and-market.pdf> (5 july 2017).
- [2] Bernhart, w. challenges and opportunities in lithium-ion battery supply. in future lithium-ion batteries 316–334 (royal society of chemistry, 2019).
- [3] G.-.. W. You, S. Park, and D. Oh. Diagnosis of electric vehicle batteries using recurrent neural networks. *IEEE Trans. Indust. Electron.*, 64, 2017.
- [4] A. Barré. A review on lithium-ion battery ageing mechanisms and estimations for automotive applications. *J. Power Sources*, 241, 2013.
- [5] J. Zhang and J. Lee. A review on prognostics and health monitoring of li-ion battery. *J. Power Sources*, 196, 2011.
- [6] A. Farmann, W. Waag, A. Marongiu, and D. U. Sauer. Critical review of on-board capacity estimation techniques for lithium-ion batteries in electric and hybrid electric vehicles. *J. Power Sources*, 281, 2015.
- [7] M. A. Hannan, M. H. Lipu, A. Hussain, and A. Mohamed. A review of lithium-ion battery state of charge estimation and management system in electric vehicle applications: challenges and recommendations. *Renew. Sustain. Energy Rev.*, 78, 2017.
- [8] X. Hu, S. Li, and H. Peng. A comparative study of equivalent circuit models for li-ion batteries. *J. Power Sources*, 198, 2012.
- [9] T. Feng, L. Yang, X. Zhao, H. Zhang, and J. Qiang. Online identification of lithium-ion battery parameters based on an improved equivalent-circuit model and its implementation on battery state-of-power prediction. *J. Power Sources*, 281, 2015.
- [10] D. Andre. Characterization of high-power lithium-ion batteries by electrochemical impedance spectroscopy. ii: Modelling. *J. Power Sources*, 196, 2011.
- [11] Daigle, m. j. & kulkarni, c. s. electrochemistry-based battery modeling for prognostics. in ann. conf. prognostics and health management society 040 (phm, 2013).
- [12] Bole, b., kulkarni, c. s. & daigle, m. adaptation of an electrochemistry-based li-ion battery model to account fordeterioration observed under randomized use. in proc. ann. conf. prognostics and health management society (phm, 2014).
- [13] G. K. Prasad and C. D. Rahn. Model based identification of aging parameters in lithium ion batteries. *J. Power Sources*, 232, 2013.



- [14] K. A. Severson. Data-driven prediction of battery cycle life before capacity degradation. *Nat. Energy*, 4, 2019.
- [15] B. Saha, K. Goebel, S. Poll, and J. Christophersen. Prognostics methods for battery health monitoring using a bayesian framework. *IEEE Trans. Instrum. Measure.*, 58, 2008.
- [16] K. Goebel, B. Saha, A. Saxena, J. R. Celaya, and J. P. Christophersen. Prognostics in battery health management. *IEEE Instrum. Measure. Mag.*, 11, 2008.
- [17] X. Hu, J. Jiang, D. Cao, and B. Egardt. Battery health prognosis for electric vehicles using sample entropy and sparse bayesian predictive modeling. *IEEE Trans. Indust. Electron.*, 63, 2015.
- [18] V. Klass, M. Behm, and G. Lindbergh. A support vector machine-based state-of-health estimation method for lithium-ion batteries under electric vehicle operation. *J. Power Sources*, 270, 2014.
- [19] P. M. Attia. Closed-loop optimization of fast-charging protocols for batteries with machine learning. *Nature*, 578, 2020.
- [20] M. Coleman, W. G. Hurley, and C. K. Lee. An improved battery characterization method using a two-pulse load test. *IEEE Trans. Energy Conv.*, 23, 2008.
- [21] W. Waag, S. Käbitz, and D. U. Sauer. Experimental investigation of the lithium-ion battery impedance characteristic at various conditions and aging states and its influence on the application. *Appl. Energy*, 102, 2013.
- [22] U. Tröltzsch, O. Kanoun, and H.-R. Tränkler. Characterizing aging effects of lithium ion batteries by impedance spectroscopy. *Electrochim. Acta*, 51, 2006.
- [23] C. R. Birkel, M. R. Roberts, E. McTurk, P. G. Bruce, and D. A. Howey. Degradation diagnostics for lithium ion cells. *J. Power Sources*, 341, 2017.
- [24] Y. Li, S. Zhong, Q. Zhong, and K. Shi. Lithium-ion battery state of health monitoring based on ensemble learning. *IEEE Access*, 7, 2019.
- [25] Y. Li. Random forest regression for online capacity estimation of lithium-ion batteries. *Appl. Energy*, 232, 2018.
- [26] Sun, b., ren, p., gong, m., zhou, x. & bian, j. soh estimation for li-ion batteries based on features of ic curves and multi-output gaussian process regression method. *destech trans. environ. energy earth sci.*<https://doi.org/10.12783/dteees/iceee2018/27789> (2018).
- [27] X. Feng. Online state-of-health estimation for li-ion battery using partial charging segment based on support vector machine. *IEEE Trans. Vehic. Technol.*, 68, 2019.
- [28] Y. Li. A quick on-line state of health estimation method for li-ion battery with incremental capacity curves processed by gaussian filter. *J. Power Sources*, 373, 2018.
- [29] M. Dubarry, V. Svoboda, R. Hwu, and B. Y. Liaw. Incremental capacity analysis and close-to-equilibrium ocv measurements to quantify capacity fade in commercial rechargeable lithium batteries. *Electrochem. Solid State Lett.*, 9, 2006.
- [30] C. Weng, Y. Cui, J. Sun, and H. Peng. On-board state of health monitoring of lithium-ion batteries using incremental capacity analysis with support vector regression. *J. Power Sources*, 235, 2013.
- [31] D. Yang, X. Zhang, R. Pan, Y. Wang, and Z. Chen. A novel gaussian process regression model for state-of-health estimation of lithium-ion battery using charging curve. *J. Power Sources*, 384, 2018.
- [32] R. R. Richardson, C. R. Birkel, M. A. Osborne, and D. A. Howey. Gaussian process regression for in situ capacity estimation of lithium-ion batteries. *IEEE Trans. Indust. Inform.*, 15, 2018.
- [33] Shen, y., seeger, m. & ng, a. y. fast gaussian process regression using kd-trees. in *adv. neural information processing systems (nips)* 1225–1232 (2006).
- [34] Saha, b., poll, s., goebel, k. & christophersen, j. an integrated approach to battery health monitoring using bayesian regression and state estimation. in *2007 ieee autotestcon* 646–653 (ieee, 2007).
- [35] D. Ben-Shimon and A. Shmiloivici. Accelerating the relevance vector machine via data partitioning. *Found. Comput. Decision Sci.*, 31, 2006.

- [36] Z. Wang, S. Zeng, J. Guo, and T. Qin. Remaining capacity estimation of lithium-ion batteries based on the constant voltage charging profile. *PLoS ONE*, 13, 2018.
- [37] Engel, s. j., gilmartin, b. j., bongort, k. & hess, a. prognostics, the real issues involved with predicting life remaining. in 2000 ieee aerospace conf. proc. 00th8484, vol. 6, 457–469 (ieee, 2000).
- [38] Pomerantseva, e., bonaccorso, f., feng, x., cui, y. & gogotsi, y. energy storage: the future enabled by nanomaterials. *science*366, eaan8285 (2019).
- [39] Z. W. Seh, Y. Sun, Q. Zhang, and Y. Cui. Designing high-energy lithium–sulfur batteries. *Chem. Soc. Rev.*, 45, 2016.
- [40] Liu, g., bao, h. & han, b. a stacked autoencoder-based deep neural network for achieving gearbox fault diagnosis. *hindawi math. problems eng.*2018, 5105709 (2018).
- [41] Kanter, j. m. & veeramachaneni, k. deep feature synthesis: towards automating data science endeavors. in 2015 ieee int. conf. data sci. adv. analytics (dsaa) 1–10 (ieee, 2015).
- [42] N. Williard, W. He, M. Osterman, and M. Pecht. Comparative analysis of features for determining state of health in lithium-ion batteries. *Int. J. Prognostics Health Manage.*, 4, 2013.
- [43] Zhang, y. & guo, b. online capacity estimation of lithium-ion batteries based on novel feature extraction and adaptive multi-kernel relevance vector machine. *energies*8, 12439–12457 (2015).
- [44] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik. Gene selection for cancer classification using support vector machines. *Machine Learning*, 46, 2002.
- [45] B. F. Darst, K. C. Malecki, and C. D. Engelman. Using recursive feature elimination in random forest to account for correlated variables in high dimensional data. *BMC Genet.*, 19, 2018.
- [46] B. Gregorutti, B. Michel, and P. Saint-Pierre. Correlation and variable importance in random forests. *Statist. Comput.*, 27, 2017.
- [47] Goodfellow, i. j., shlens, j. & szegedy, c. explaining and harnessing adversarial examples. preprint at <https://arxiv.org/abs/1412.6572> (2014).
- [48] M. Doyle, T. F. Fuller, and J. Newman. Modeling of galvanostatic charge and discharge of the lithium/polymer/insertion cell. *J. Electrochem. Soc.*, 140, 1993.
- [49] S. Wager, T. Hastie, and B. Efron. Confidence intervals for random forests: the jackknife and the infinitesimal jackknife. *J. Machine Learning Res.*, 15, 2014.
- [50] Lakshminarayanan, b., pritzel, a. & blundell, c. simple and scalable predictive uncertainty estimation using deep ensembles. in adv. neural information processing systems (nips) 6402–6413 (curran associates, 2017).
- [51] J. Bergstra and Y. Bengio. Random search for hyper-parameter optimization. *J. Machine Learning Res.*, 13, 2012.
- [52] A. D. Handoko, F. Wei, B. S. Yeo, and Z. W. Seh. Understanding heterogeneous electrocatalytic carbon dioxide reduction through operando techniques. *Nat. Catal.*, 1, 2018.