

北京航空航天大学学报

Journal of Beijing University of Aeronautics and Astronautics

ISSN 1001-5965, CN 11-2625/V

《北京航空航天大学学报》网络首发论文

题目: 基于机器学习的锂离子电池健康状态分类与预测
作者: 高昊天, 陈云霞
DOI: 10.13700/j.bh.1001-5965.2022.0154
收稿日期: 2022-03-17
网络首发日期: 2022-06-01
引用格式: 高昊天, 陈云霞. 基于机器学习的锂离子电池健康状态分类与预测[J/OL]. 北京航空航天大学学报. <https://doi.org/10.13700/j.bh.1001-5965.2022.0154>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

基于机器学习的锂离子电池健康状态分类与预测

高昊天, 陈云霞✉

(北京航空航天大学 可靠性与工程学院, 北京 100191)

*通信作者 E-mail: chenyunxia@buaa.edu.cn

摘要 对锂离子电池进行准确的健康状态预测是电池应用中的一项关键技术。然而, 由于锂离子电池内部复杂的电化学反应体系, 多样的失效机理以及生产差异, 锂离子电池的退化往往呈现出较大的分散性, 这为锂离子电池健康状态的准确预测造成了较大的困难。为此, 本文提出了一种基于机器学习的锂离子电池健康状态分类与预测方法, 首先基于精度约束, 利用双子群优化算法确定训练集数据合适的类别个数及类别范围, 然后基于 Softmax 分类模型根据锂离子电池早期退化数据进行健康状态分类, 使得退化趋势较为接近的电池被分为一类, 最后对每一类电池分别利用 BP(back-propagation)神经网络构建其健康状态预测模型, 从而减小锂离子电池数据的大分散性的影响, 提升锂离子电池的健康状态预测精度。本文方法相比传统方法预测误差降低了 34%以上, 验证了本方法的有效性和优越性。

关键词 锂离子电池; 健康状态预测; 分类; 机器学习; 优化算法

中图分类号 TP391

文献标识码: A

DOI: 10.13700/j.bh.1001-5965.2022.0154

A machine learning based method for Lithium-ion battery state of health classification and prediction

GAO Haotian, CHEN Yunxia✉

(School of Reliability and System Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100191, China)

*Tel.: 010-82338909 E-mail: chenyunxia@buaa.edu.cn

Abstract Accurate state of health (SOH) prediction for lithium-ion batteries is a key technology in battery applications. However, due to the complex electrochemical systems, diverse failure mechanisms, and production differences, the degradation of the lithium-ion batteries often presents large dispersion, which makes it difficult to accurately predicting the SOH of the lithium-ion battery. To solve this problem, this paper proposes a machine learning-based method for classifying and predicting the SOH of lithium-ion batteries. First, based on the accuracy constraints, the double subgroup optimization algorithm is used to determine the appropriate number of categories and category ranges for the training set data. Then, based on the Softmax classification model, lithium-ion batteries are classified according to the early-cycle data, so that the batteries with a similar degradation trend are divided into one class. Lastly, the back propagation neural network is used to build the SOH prediction model for each type of battery to reduce the influence of data dispersion and improve the prediction accuracy. Compared with the traditional method, the prediction error of the proposed method is reduced by more than 34%, which verifies the effectiveness and superiority of this method.

Key words Lithium-ion battery; SOH prediction; classification; machine learning; optimization algorithm

锂离子电池由于具有低生产成本、高能量密度、低自放电率以及较长的寿命等优点^[1-3], 已经被广泛应用于消费电子、电动汽车、储能电站, 乃至卫星船舶等重要领域^[4-5]。然而, 随着充放电循环的不断进行, 锂离子电池的容量、内阻等电化学性能将逐渐退化, 直至最终失效而导致无法满足设备的正常使用需求, 甚至可能造成严重的安全事故。因此, 对锂离子电池进行准确的健康状态 (State of Health, SOH) 预测, 并由此来判断电池是否能满足用户需求以及何时该被更换是一项非常必要的任

收稿日期: 2022-03-17

基金项目: 国家自然科学基金 (基金号 52075019);

Fund: National Natural Science Foundation of China (52075019);

网络首发时间: 2022-06-01 13:21:42 网络首发地址: <https://kns.cnki.net/kcms/detail/11.2625.v.20220530.1830.003.html>

务。然而, 由于锂离子电池复杂的电化学体系、多样的失效机理以及制造差异, 即使同一批次电池, 其退化数据也往往呈现出较大的分散性^[6-7], 这使得传统方法难以精准刻画其退化规律, 为传统方法对锂离子电池进行准确建模及预测带来了一定的挑战。

锂离子电池的 SOH 被定义电池当前循环容量与初始容量之比^[8], 因此, 锂离子电池的 SOH 预测实质上是对电池的未来容量退化情况进行预测。目前, 针对锂离子电池 SOH 预测的研究已经较为深入。**传统的锂离子电池 SOH 预测方法主要分为基于模型的方法和数据驱动的方法这两大类。**

基于模型的方法通过研究锂离子电池退化过程中关键因子的影响, 从而建立起描述电池性能退化的模型^[9]。随着充放电循环的进行, 锂离子电池的性能会逐渐产生退化, 其内部各项物理化学特征如容量、阻抗、电流、电压等也会发生动态的变化, 通过对锂离子电池的动态特征进行建模即可对锂离子电池的退化状况进行预测, 从而实现了对电池健康状态的预测。这类方法主要利用电化学模型^[10], 等效电路模型^[11], 布朗漂移模型^[12]等构建电池退化模型并结合粒子滤波算法^[12-13], 卡尔曼滤波算法^[14-15]等滤波算法对模型的参数进行动态调节从而来刻画电池的退化过程。

近年来, 数据驱动模型在锂离子电池健康状态预测领域受到了越来越多的关注, 因为相比于基于模型的方法, 数据驱动模型不需要事先建立明确的数学模型来刻画锂离子电池的退化过程, 而是从锂离子电池充放电循环过程中的历史数据出发, 通过采用特定的学习算法, 从锂离子电池的容量、内阻、电压等数据中提取能反应锂离子电池性能退化状态的关键特征, 并训练这些特征与电池的退化过程的回归关系, 从而建立起能刻画和描述电池退化过程的模型, 实现电池退化状态的预测^[16]。这类方法主要利用支持向量机^[17-19], 神经网络^[8,20,21], 贝叶斯预测^[22], BOX-COX 转换^[23]等机器学习方法进行建模。

尽管目前针对锂离子电池 SOH 预测的研究已较为广泛和深入, 但这些传统方法在实际使用时仍然存在着一些不足。对于基于模型的方法而言, 其难以找到一个适用于各类电化体系多样失效模式的准确的锂离子电池退化模型^[24], 并且, 其预测精度主要依赖于滤波算法, 而滤波算法又常常受制于粒子退化问题^[25]; 而对于数据驱动方法而言, 其虽然具有强大的学习能力, 但由于锂离子电池复杂的电化体系、多样的失效机理以及制造差异, 其退化数据常常呈现出较大的分散性, 而大分散性数据会严重干扰数据驱动模型对数据映射规律的学习, 从而极大影响数据驱动模型的预测精度。因此, 为了改善这一问题, 对于大分散性的锂离子电池数据集, 有必要先根据电池的退化趋势进行分类, 将退化规律接近的电池归为一类, 然后对每一类电池单独进行建模预测, 从而实现锂离子电池 SOH 预测精度的提升。其中, 分类模型的训练需要一定量的有标注的训练集数据, 即训练集中各个电池的所属类别需要进行标注, 而这里的类别标注也会影响到分类模型的分类结果, 进而影响到最终预测模型的预测精度。因此, 如何基于最终预测精度的要求进行分类个数确定及类别范围控制也是一个需要考虑的问题。该问题本质上是一个多目标优化问题, 即所划分的各个类的类内误差尽可能小的同时, 类别的个数也尽可能少, 可以通过优化算法来解决。本文**比较了常用的多种优化算法, 最终选定了效果较优的双子群优化算法**^[26]来对该问题进行优化。

基于上述思考, 本文在传统数据驱动方法的基础上, 通过结合优化算法和分类算法来应对大分散性电池退化数据带来的挑战, 提出了一套基于机器学习的锂离子电池 SOH 分类与预测方法, 在实现锂离子电池 SOH 准确分类的同时涵盖实际应用场景中各类退化趋势电池的预测任务并实现预测精度的提升。以某公司实际的磷酸铁锂电池退化数据为例, 对本文方法以及传统数据驱动方法进行了对比验证, 验证了本文方法的有效性和优越性。

1 问题定义

已有训练集电池容量退化数据: $\mathbf{X}_{\text{train}} = [\mathbf{Q}_1, \mathbf{Q}_2, \dots, \mathbf{Q}_n]$, $\mathbf{Q}_i = [Q_{i1}, Q_{i2}, \dots, Q_{im}]^T$, 其中 $\mathbf{Q}_i (1 < i < n)$ 为训练集中第 i 个电池的容量序列数据, $Q_{ij} (1 < j < m)$ 为训练集中第 i 个电池第 j 循环的容量, n 和 m 分别为训练集电池的总个数和电池的总循环数; 已有测试集电池容量数据: $\mathbf{X}_{\text{test}} = [\mathbf{Q}'_1, \mathbf{Q}'_2, \dots, \mathbf{Q}'_n]$,

$\mathbf{Q}_i' = [Q_{i1}', Q_{i2}', \dots, Q_{im}']^T$, 其中 $\mathbf{Q}_i' (1 < i < n')$ 为测试集中第 i 个电池的容量序列数据, $Q_{ij}' (1 < j < m)$ 为测试集中第 i 个电池第 j 循环的容量, n' 为测试集电池的总个数。本文关注的问题便是如何利用训练集数据建立准确的 SOH 预测模型, 从而在只利用测试集电池早期退化数据的情况下, 对其退化后期的容量退化情况进行准确预测。

因为训练集电池容量退化数据的分散性较大, 为了建立更精准的健康状态预测模型, 需要先对电池进行分类, 然后再对每一类电池单独进行建模预测, 而构建分类模型的前提是训练集中各个电池有明确的类别标签。根据以上思路, 该问题可以拆分为以下三个子问题来进行解决:

1. 确定训练集电池合适的类别标签: 进行类别个数确定及类别范围控制, 使得每个类的类内电池较为接近, 保持在一定范围内, 同时, 类的个数应尽可能少。

2. 构建锂离子电池 SOH 分类模型: 利用训练集电池退化早期的容量数据及其类别标签构建锂离子电池 SOH 分类模型, 然后将测试集电池退化早期的容量数据代入分类模型中, 得到其分类结果。

3. 构建锂离子电池 SOH 预测模型: 对训练集中每一类电池分别进行建模, 得到每个类对应的 SOH 预测模型, 然后根据测试集电池的分类结果, 将其代入到相应的预测模型中, 预测其后续 SOH 退化情况。

2 模型建立

2.1 基于精度约束的分类个数确定及类别范围控制方法

本方法首先需要对训练集数据进行基于精度约束的分类个数确定及类别范围控制, 即通过优化算法确定满足要求的最优的分类个数以及训练集中每个电池所属的类别, 为下一步分类模型的训练打好“标签”。

该问题本质上是一个多目标优化问题, 目标一是使得每个类内各个电池的退化规律都较为接近, 各个电池间的偏差尽可能小, 保持在一定范围内——这是为了使训练出的分类模型所分出的各个类的类内偏差也都较小, 从而使针对每个类进行建模的预测模型能给出准确的预测, 其中, 各个电池间的相对偏差通过(1)来进行计算:

$$D_{ij} = \frac{\sum_{k=1}^m |Q_{ik} - Q_{jk}|}{mQ_0} \quad (1)$$

其中, D_{ij} 代表第 i 号电池与第 j 号电池之间的相对偏差, Q_{ik} 表示第 i 号电池第 k 循环的容量值, Q_0 为额定容量。

目标二是划分的类的个数应尽可能少, 从而使得最终所需的针对每个类的预测模型也尽可能少, 减少模型训练的时间成本。基于上述分析, 该问题的适应度函数可以写成(2)形式:

$$F(\mathbf{x}) = a \times N + b \times \sum_{i=1}^N \sum_{j=2}^{n_i} \sum_{k=1}^{j-1} d_{ijk} \quad (2)$$

其中, a 、 b 为权重系数, N 为分类个数, n_i 为第 i 类中电池的个数, d_{ijk} 是引入惩罚项调整后的电池间的相对偏差, 通过错误!未找到引用源。计算:

$$d_{ijk} = \begin{cases} D_{jk} & D_{jk} < K \\ p \times D_{jk} & D_{jk} \geq K \end{cases} \quad (3)$$

其中, K 为精度约束, p 为惩罚项系数。通过引入惩罚项, 对每所优化出的每个类的类内偏差范围进行控制。

该问题的优化对象是每个电池的类别标签, 是一个整数优化问题。其可行解可以写成(4)的形式:

$$\mathbf{x} = [l_1, l_2, \dots, l_n]^T \quad (4)$$

其中, 可行解 \mathbf{x} 是一个 n 维列向量, l_i 代表第 i 个电池的类别标签, 且 $l_i = 1, 2, \dots, N$ 。该问题可以通

过合适的优化算法来进行优化求解。本文通过对常用的遗传算法^[27]、果蝇优化算法^[28]、教与学优化算法^[29]以及双子群优化算法^[26]等常用的优化算法进行验证对比,最终选择使用双子群优化算法来对该问题进行优化。

双子群优化算法融合了果蝇优化算法与教与学优化算法的优点,该算法将可行解“种群”分为两部分,一部分按照果蝇算法的寻优规则进行迭代寻优,另一部分按照教与学优化算法的规则进行寻优。其中,果蝇优化算法是依据果蝇觅食行为设计出的一种优化算法,属于精英学习策略,即所有个体向当前的最优个体学习。该算法具有很强的局部搜索能力,但容易陷入局部最优解。教与学优化算法模拟班级中学生的学习方式来进行优化,即个体可以向“教师”(即最优个体)学习,也可以“同学”(个体)之间互相学习,还可以以自身为基础进行自学,通过对这三种学习方式进行比较,选择其中最优的学习方式,实现自身位置的更新。该算法因其学习方式的多样性而具有较强的全局搜索能力。而双子群优化算法通过综合这两种算法的优势,获得了更为均衡的局部搜索与全局寻优的能力。该算法的计算过程如下所述:

1.初始化种群规模 $2S$, 最大迭代数 M , 最大搜索半径 R_{\max} , 最小搜索半径 R_{\min} 。初始化 A_1 和 A_2 两个子群,即各随机生成 S 个可行解,其中 A_1 子群将按照果蝇算法的规则进行寻优, A_2 子群按照教与学优化算法的规则进行寻优。

2.计算 A_1 与 A_2 子群中所有个体的适应度,找出其中的最优个体,记录其适应度 $bestFit$ 及其位置 $bestX$ 。

3. A_1 子群寻优:以 $bestX$ 为中心,以 $[R_{\min}, R_{\max}]$ 之间的随机数为半径随机生成 S 个个体。

4. A_2 子群寻优,具体步骤如下:

a) “教”阶段:各个体以(5)的方式向教师学习

$$x_{new}^i = x_{old}^i + l \quad (5)$$

$$l = \text{round}[r_i \times (bestX - T_i \times meanX)] \quad (6)$$

其中, x_{old}^i 与 x_{new}^i 分别为第 i 个个体在学习前后的位置, $\text{round}()$ 为取整函数, $meanX = \sum x / S$, 表示子群中所有个体的平均位置,教学因子 T_i 为 1 或 2 的随机整数,学习因子 r_i 为 0 到 1 之间的随机数。若 x_{new}^i 优于 x_{old}^i , 则用 x_{new}^i 更新 x_{old}^i , 否则不更新。

b) “互相学习”阶段:

对于每一个 A_2 子群中的个体 x_i , 其随机从挑选一个个体 x_j 作为学习对象, x_i 通过比较与 x_j 之间的差异来进行学习,学习公式如错误!未找到引用源。所示:

$$x_{new}^i = \begin{cases} x_{old}^i + r(x_i - x_j) & F(x_i) > F(x_j) \\ x_{old}^i + r(x_j - x_i) & F(x_i) \leq F(x_j) \end{cases} \quad (7)$$

若 x_{new}^i 优于 x_{old}^i , 则用 x_{new}^i 更新 x_{old}^i , 否则不更新。

c) “自学”阶段:

对每一个 A_2 子群中的个体,其以自身为中心,以 $[R_{\min}, R_{\max}]$ 之间的随机数为半径,在该范围内自学。

5.计算 A_1 与 A_2 子群中所有个体的适应度,找出其中的最优个体,记录其适应度 Fit 及其位置 X , 若 $Fit < bestFit$, 则将 $bestFit$ 与 $bestX$ 替换为 Fit 与 X 。

6.判断迭代次数是否小于等于 M , 若是则回到第 3 步, 若否则跳出迭代,完成寻优。

2.2 基于 Softmax 分类模型的锂离子电池 SOH 分类模型

2.1 中的步骤为训练集数据打上了合适的类别标签,之后便可以用于训练锂离子电池的分类模型。当训练集中的类别个数大于 2 时,这就变成了一个多分类问题,而常用的逻辑回归、支持向量机等方法适用于二分类问题,应用在多分类问题上较为复杂且会产生分类误差累积等问题,因此,本文

使用能够便捷处理多分类问题的 Softmax 神经网络来进行分类。

典型的 Softmax 神经网络的结构如图 1 所示, 其中, Softmax 层实际上只是神经网络的一层额外的处理层, 将神经网络的输出转换为一个概率分布, 由此实现分类。Softmax 层的激活函数如(8)所示:

$$\text{Softmax}(y)_i = \frac{e^{y_i}}{\sum_i e^{y_i}} \quad (8)$$

其中, y_i 为原始神经网络的输出。从(8)中可以看出, 原始神经网络的输出被用作置信度来生成新的输出, 而新的输出满足概率分布的所有要求。这个新的输出可以理解为经过神经网络的推导, 一个样本属于各个类别的概率分别是多大。

本文使用图 1 所示的神经网络作为分类模型, 该神经网络的输入为锂离子电池前半循环的容量值——这是为了实现提前一半时间对电池进行准确的分类和预测; 输出为一个 N 维的向量, 代表着该样本属于各个类的概率, 例如, 若输出为 $[1, 0, 0, 0, 0]^T$, 则表示该样本属于第一类的概率为 100%, 属于其他类的概率为 0。利用 2.1 中完成打标的训练集数据对该网络进行训练, 便可得到锂离子电池 SOH 分类模型, 之后将测试集电池前半循环的容量值输入神经网络, 即可完成对测试集电池的分类。

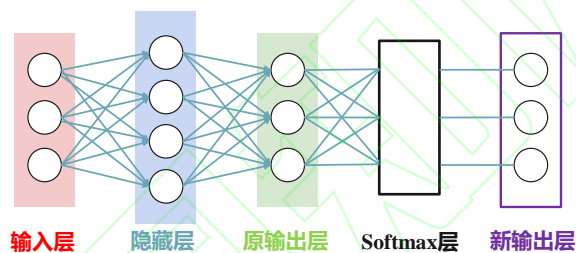


图 1 Softmax 分类模型的网络结构
Fig.1 Network structure of Softmax classification model

2.3 基于神经网络的锂离子电池 SOH 预测模型

神经网络等机器学习模型已被广泛应用于锂离子电池的 SOH 预测。本文选用简单而高效的反向传播神经网络(back-propagation neural network, BPNN)来构建锂离子电池 SOH 预测模型。该神经网络的输入为锂离子电池前半循环的容量值, 输出为后半循环的容量值, 以此学习锂离子电池前期退化情况与后期退化趋势的映射关系, 并实现通过早期退化数据对后续退化情况的预测。

对于训练集中各类退化趋势的锂离子电池分别代入神经网络训练出一个预测模型, 之后根据分类器对测试集电池的分类结果, 将每个测试集电池分别代入到与之对应的预测器中, 从而完成对测试集电池健康状态的预测。

本文方法的整体技术路线如图 2 所示。

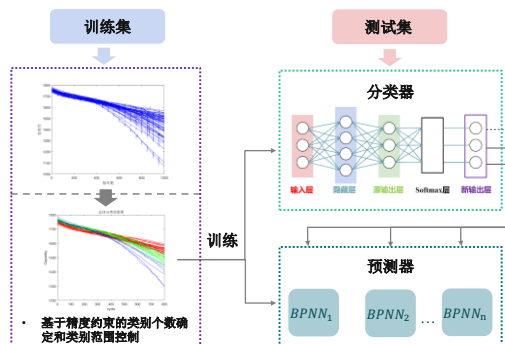


图 2 技术路线图
Fig.2 Implementation route

3 案例应用与分析

3.1 案例应用

本文以某锂离子电池厂商提供的 61 个某体系锂离子电池的实际容量退化数据为例来验证本文方法的有效性和优越性。该批电池在同温度、同充放电条件下测试了 1000 循环，得到了其前 1000 循环的容量退化情况，如图 3 所示。从图中可以看出，即使在相同的测试条件下，电池的退化仍然呈现出很大的分散性，这为电池健康状态的准确预测带来了困难。

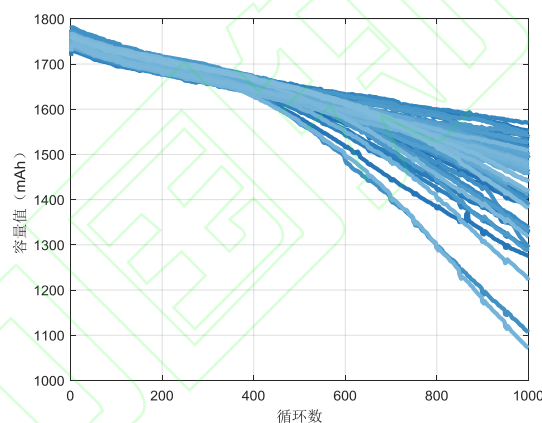


图 3 某批锂离子电池前 1000 循环容量退化曲线
Fig.3 Capacity degradation curve of a batch of lithium-ion batteries in the first 1000 cycles

为了验证本文方法对大分散性数据的有效性，本文以该批电池中序号为偶数号的 30 个电池作为训练集，用于训练分类模型以及预测模型，以序号为奇数号的 31 个电池作为测试集，对测试集电池进行分类并进而将其代入相应类别的预测模型中进行预测，最终得到对测试集电池的健康状态预测结果。本文以前一半循环（即前 500 循环）的容量值作为分类模型与预测模型的输入，即希望能只通过前一半循环的退化信息便对后一半循环的退化情况进行预测。

首先需要进行基于精度约束的类别个数确定及类别范围控制，通过双子群优化算法的优化求解，得出满足精度要求下最少的类别个数以及训练集中各个电池合适的类别标签。经优化后的训练集中各类电池的容量退化情况如图 4 所示。从图中可以看出，根据退化趋势的不同，训练集中的电池被明显分为了四类，每一类内各电池的退化趋势都较为接近而不同类间电池退化趋势的差异较大，即类内差异较小而类间差异较大，这便为训练出一个有效而准确的分类模型提供了数据基础。

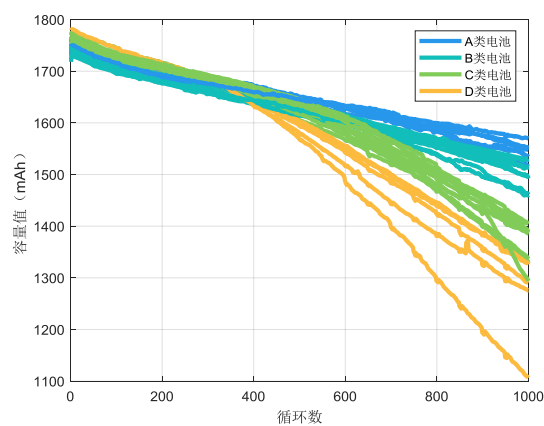


图 4 训练集各类电池容量退化曲线
Fig.4 Capacity curve for each type of battery in the training dataset

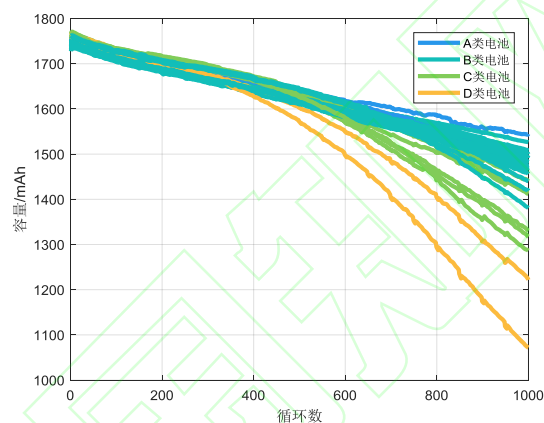


图 5 测试集各类电池容量退化曲线
Fig.5 Capacity curve for each type of battery in the test dataset

将完成打标签处理的训练集代入分类模型中进行训练，完成训练后，将测试集电池前 500 循环的容量值输入到分类模型中，得到测试集电池的分类结果，如图 5 所示。从图中可以大致看出，测试集电池所分出的四类基本处在相应类的训练集电池所在的范围内。为了能更清晰地观察这一结果，本文将每一类电池的训练集与测试集容量曲线单独画在一张图上，如图 6 所示，其中图(a)、图(b)、图(c)、图(d)分别为 A 类、B 类、C 类和 D 类电池中训练集与测试集的情况。从图中可以清晰地看出，四类电池中，测试集电池的容量退化曲线基本都处在训练集电池所涵盖的一个较小的范围内，其分散性远小于原始训练集，从而可以较好地保证对测试集电池健康状态的准确预测。

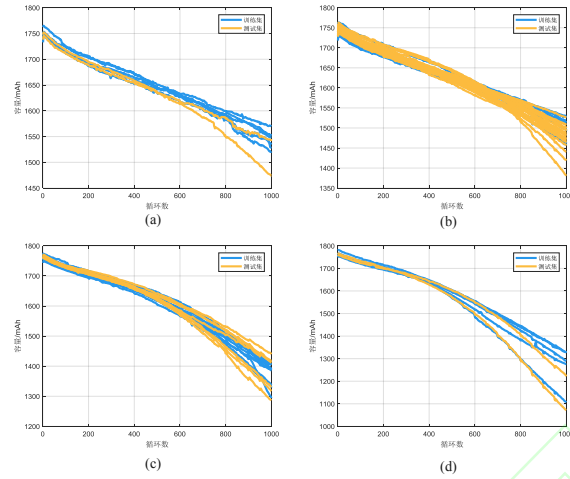


图 6 各类电池训练集与测试集情况
Fig.6 The training dataset and the test dataset for each type of battery

之后, 分别利用四类电池的训练集, 以其前 500 循环的容量值作为输入, 501-1000 循环的容量值作为输出, 代入神经网络进行训练, 得到四个预测模型, 并将测试集电池前 500 循环的容量值代入到相应类的预测模型中, 分别得到测试集中 A、B、C、D 四类电池后续容量退化情况的预测结果, 如图 7 所示。从图中可以看出, 这四类电池的容量预测结果与真实值较为接近, 达到了较好的预测效果, 仅有一小部分电池预测值与真实值偏差较大, 以图 7(d)中靠上方的电池为例来进行说明, 其预测值明显高于真实值是因为训练集中与其接近的电池容量均高于该电池 (如图 6(d)所示), 因此神经网络对于该电池也倾向于给出接近训练集中这些电池的预测结果, 从而导致该电池的预测值大于真实值。为了定量评价预测效果, 本文以测试集中所有电池的平均百分误差(mean percentage error, MPE), 预测结尾点平均百分误差(mean percentage error at the endpoint, MPEE)以及均方根误差(root-mean-squared error, RMSE)作为定量指标来衡量预测误差的大小, 并以平均百分误差为例, 画出了本文方法在该指标下预测误差的分布情况, 结果见表 1 和图 8。从表 1 中可以看出, 测试集电池预测结果的平均百分误差小于 1%, 且从图 8 中可以看出, 绝大部分电池的预测误差都在 1%以内, 说明本文方法能够较为准确地进行锂离子电池健康状态预测, 验证了方法的有效性。

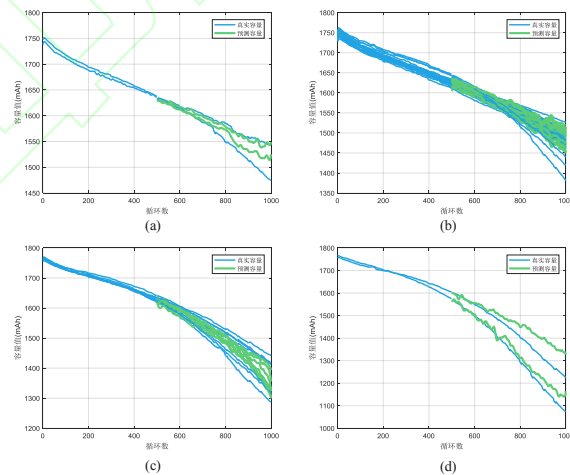


图 7 测试集四类电池容量预测结果
Fig.7 Prediction results for four types of batteries in the test dataset

3.2 对比研究

为了进一步验证本文方法的优越性, 本文方法与未经分类便直接进行预测的传统方法进行了对

比。本文方法按照先分类后预测的思路，本质上是一个分类模型加多个回归模型，而传统方法没有分类的步骤，本质上只是一个单独的回归模型，其直接利用各种机器学习方法来对锂离子电池的退化趋势进行建模回归。为了直观体现因本文方法引入分类而带来的预测精度提升，本节中的传统方法选用与本文方法相同的 BP 神经网络模型，且其网络结构与参数设置与本文方法完全一致。本节中的传统方法直接使用偶数序号的 30 个电池构成的原始训练集数据来对神经网络进行训练，并利用训练好的 BP 神经网络对原始测试集中的 31 个电池的后续容量退化情况进行预测，预测结果如图 9 和表 1 所示。

从图中可以直观地看出，传统方法给出的预测结果波动较大，且有部分电池的预测结果明显偏离了真实趋势，预测效果较差，这说明由大分散性数据直接训练出的神经网络并没有很好地学习到每个样本的规律，而只捕捉到了一个笼统的，大范围内适用的趋势，难以对每个锂离子电池进行准确的健康状态预测；而从定量指标上看，传统方法的平均百分误差在 2% 以上，且从图 8 所示的误差分布情况上看，有许多电池的预测误差在 3% 以上甚至最大误差达到了 6% 以上，预测结果并不理想，而本文方法相比于传统方法在三个误差指标下都降低了 34% 以上，且在 MPE 和 RMSE 这两个指标上降低了接近 60%，预测效果得到了明显的提升，验证了本文方法的优越性。

表 1 本文方法与传统方法预测误差对比
Table 1 Comparison of prediction errors between the proposed method and the traditional method

方法	MPE	MPEE	RMSE
本文方法	0.96%	2.74%	17.45
传统方法	2.33%	4.19%	41.15
误差缩小幅度	58.8%	34.61%	57.59%

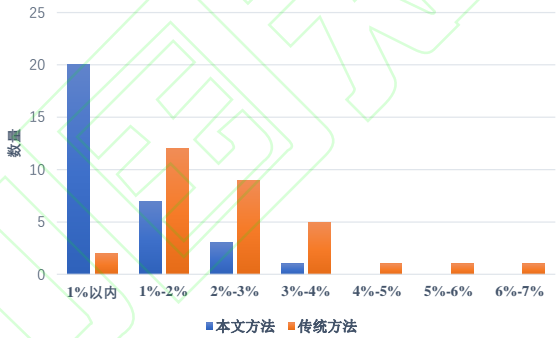


图 8 本文方法与传统方法平均百分误差分布情况
Fig.8 The mean percentage error distribution between the proposed method and the traditional method

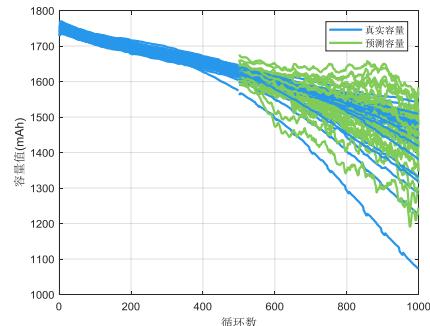


图 9 传统方法对测试集电池的预测结果
Fig.9 Prediction results for the test set batteries using traditional method

3.3 讨论

在机器学习领域,通常认为更多的训练数据会训练出更好的模型,然而对于锂离子电池 SOH 预测模型而言,由于电池退化数据的大分散性,用于建模的数据并非越多越好,这个问题可以通过一个拟合算例来直观地说明。如图 10 所示,对于一个大分散性的数据集,用一条曲线来拟合(图 a 所示)虽然能更好地描述整体,但其拟合误差要明显大于用三条曲线分别拟合(图 b 所示)。神经网络某种程度上也可以被理解为一个回归模型,其训练过程中通过梯度下降法调整各神经元参数来使得模型输出与真实输出之间的误差尽可能小,从而学习到训练集输入与输出之间的映射规律,但当训练集数据分散性过大(即实际上呈现出多种退化规律时),神经网络在参数调整的过程中容易得到多种规律综合的结果,在使全局误差尽可能小的同时会使个体误差因多种规律的干扰而变大——也就是说,更多的训练数据会让模型学到更多的规律,适用于更大范围内的电池预测任务,但模型在变得更普适的同时也会不可避免的牺牲对个体的预测精度,即“泛用”和“精细”不可兼得。图 9 中所呈现的预测结果波动较大且个别电池的预测结果明显偏离正确趋势的现象正是由于神经网络在多种退化规律的“干扰”中学习得较为“混乱”所导致的。而本文方法首先对锂离子电池进行分类并针对每一类电池单独进行建模预测,每一类电池的分散性相比原数据集大大减小,从而使得神经网络可以更好地学习到该类电池的退化规律,并给出更为准确的 SOH 预测。因此,为了提高预测精度,有必要先利用合适的分类算法来对锂离子电池进行分类,然后再对各类电池分别进行建模预测。

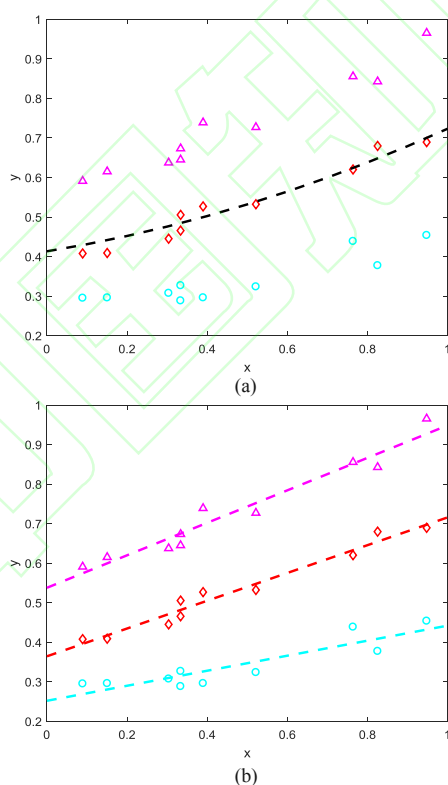


图 10 大分散性数据集拟合算例
Fig.10 An example of fitting a large distributed dataset

4 结论

针对锂离子电池退化数据的大分散性特点,本文提出了一种基于机器学习的锂离子电池健康状态分类与预测方法,主要有以下优点:

- 1) 基于双子群优化算法完成基于精度约束的训练集数据类别个数确定及类别范围控制,确定出

最优的训练集各电池类别标签。

2) 基于 Softmax 分类模型, 利用锂离子电池前一半循环的容量退化数据实现其健康状态的准确分类。

3) 对每一类电池分别构建基于 BP 神经网络的健康状态预测模型, 该模型对所有测试集电池的平均百分误差小于 1%, 且相比传统方法在不同指标上误差降低了 34%—58%, 预测准确度大大提升。

参考文献 (References)

- [1] Dunn B, Kamath H, Tarascon J M. Electrical Energy Storage for the Grid: A Battery of Choices[J]. Science, 2011, 334(6058):928-935.
- [2] Nykvist B, Nilsson M. Rapidly falling costs of battery packs for electric vehicles[J]. Nature Climate Change, 2015, 5(4):329-332.
- [3] Raccuglia P, Elbert K C, Adler P D F, et al. Machine-learning-assisted materials discovery using failed experiments[J]. Nature, 2016, 533(7601):73.
- [4] Gao L, Liu S, Dougal R A. Dynamic lithium-ion Battery model for system simulation[J]. IEEE Transactions on Components and Packaging Technologies, 2002, 25(3):495-505.
- [5] Nishi Y. Lithium ion secondary batteries: past 10 years and the future[J]. Journal of Power Sources, 2001, 100(1/2):101-106.
- [6] Burns J C, Kassam A, Sinha N N, et al. Predicting and Extending the Lifetime of Li-Ion Batteries[J]. Journal of The Electrochemical Society, 2013, 160(9):A1451-A1456.
- [7] Baumhöfer T, Brühl M, Rothgang S, et al. Production caused variation in capacity aging trend and correlation to initial cell performance[J]. Journal of Power Sources, 2014, 247:332-338.
- [8] Dai H, Zhao G, Lin M, et al. A novel estimation method for the state of health of lithium-ion battery using prior knowledge-based neural network and markov chain[J]. IEEE Transactions on Industrial Electronics, 2018, 66(10):7706-7716.
- [9] Barré B, Deguilhem S, Grolleau, et al. A review on lithium-ion battery ageing mechanisms and estimations for automotive applications[J]. Journal of Power Sources, 2013, 241:680-689.
- [10] Lyu C, Lai Q, Ge T, et al. A lead-acid battery's remaining useful life prediction by using electrochemical model in the particle filtering framework[J]. Energy, 2017, 120:975-984.
- [11] Pastor-Fernández C, Chouchelamane GH, Widanage W, et al. A comparison between electrochemical impedance spectroscopy and incremental capacity-differential voltage as Li-ion diagnostic techniques to identify and quantify the effects of degradation modes within battery management systems[J]. Journal of Power Sources, 2017, 360:301-318.
- [12] Dong G, Chen Z, Wei J, et al. Battery health prognosis using brownian motion modeling and particle filtering[J]. IEEE Transactions on Industrial Electronics, 2018, 65(11): 8646-8655.
- [13] Hu C, Jain G, Tamirisa P, et al. Method for estimating capacity and predicting remaining useful life of lithium-ion battery[J]. Applied Energy, 2014, 126:182-189.
- [14] Batzel T D, Swanson D C. Prognostic Health Management of Aircraft Power Generators[J]. IEEE Transactions on Aerospace and Electronic Systems, 2009, 45(2): 473-482.
- [15] Sepasi S, Ghorbani R, Liaw B Y. Inline state of health estimation of lithium-ion batteries using state of charge calculation[J]. Journal of Power Sources, 2015, 299: 246-254.
- [16] Xu F, Yang F, Fei Z, et al. Life prediction of lithium-ion batteries based on stacked denoising autoencoders[J]. Reliability Engineering and System Safety, 2021, 208:107396.
- [17] Wei J, Dong G, Chen Z. Remaining useful life prediction and state of health diagnosis for lithium-ion batteries using particle filter and support vector regression[J]. IEEE Transactions on Industrial Electronics, 2018, 65(7):5634-5643.
- [18] Hu J, Hu J, Lin H, et al. State-of-charge estimation for battery management system using optimized support vector machine for regression[J]. Journal of Power Sources, 2014, 269:682-693.
- [19] Wang D, Miao Q, Pecht M. Prognostics of lithium-ion batteries based on relevance vectors and a conditional three-parameter capacity degradation model[J]. Journal of Power Sources, 2013, 239:253-264.
- [20] You G, Park S, Oh D. "Diagnosis of electric vehicle batteries using recurrent neural networks[J]. IEEE Transactions on Industrial Electronics, 2017, 64(6):4885-4893.
- [21] Zhang Y, Xiong R, He H, et al. Long Short-Term Memory Recurrent Neural Network for Remaining Useful Life Prediction of Lithium-Ion Batteries[J]. IEEE Transaction on Vehicular Technology, 2018, 67(7):5695-5705.
- [22] Hu X, Jiang J, Cao D, et al. Battery health prognosis for electric vehicles using sample entropy and sparse bayesian predictive modeling[J]. IEEE Transactions on Industrial Electronics, 2015, 63(4):2645-2656.
- [23] Zhang Y, Xiong R, He H, et al. Lithium-ion battery remaining useful life prediction with box-cox transformation and monte carlo simulation[J]. IEEE Transactions on Industrial Electronics, 2018, 66, (2):1585-1597.
- [24] Waag W, Fleischer C, Sauer D U. Critical review of the methods for monitoring of lithium-ion batteries in electric and hybrid vehicles[J]. Journal of Power Sources, 2014, 258:321-339.
- [25] Liu Z, Sun G, Bu S, et al. Particle learning framework for estimating the remaining useful life of lithium-ion batteries[J]. IEEE Transaction on Instrumentation and Measurement, 2017, 66(2):280-293.
- [26] 李栋. 智慧水务应用架构设计与关键技术研究[D]. 西安: 西北工业大学, 2019: 106-111.
- [27] LI D. Smart water application architecture design and key technology research[D]. Xian: Northwestern Polytechnical University, 2019:106-111(in Chinese).
- [27] Morris G M, Goodsell D S, Halliday R S, et al. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function[J]. Journal of Computational Chemistry, 2015, 19(14):1639-1662.
- [28] Pan W T. A new Fruit Fly Optimization Algorithm: Taking the financial distress model as an example[J]. Knowledge-Based Systems, 2012, 26(2):69-74.
- [29] Rao R V, Savsani V J, Vakharia D P. Teaching-learning-based optimization: A novel method for constrained mechanical design optimization problems[J]. Computer-Aided Design, 2011, 43(3):303-315.