

DOI:10.19651/j.cnki.emt.2108565

基于 PCA 和 GA-BP 神经网络的锂电池容量估算方法^{*}

吴琼 徐锐良 杨晴霞 徐立友

(河南科技大学车辆与交通工程学院 洛阳 471003)

摘要: 针对车用锂离子动力电池容量估算方法精度不高的问题,提出了一种利用遗传算法优化 BP 神经网络的锂离子电池剩余容量估算方法。首先在整理 NASA 锂离子电池数据集后,得到不同健康状态下电池的容量增量曲线峰值。其次将健康因子进行主成分分析对其降维处理,利用遗传算法优化 BP 神经网络的连接权值,对锂离子电池容量进行预测。最后在 NASA 不同型号的电池上应用模型进行了验证。结果表明,所提出的方法可以在不同训练量的情况下准确估算 4 种锂离子电池的容量,其估算的方均根误差小于 2%,且与未使用遗传算法优化的预测结果相比,该方法具有较高的预测精度。

关键词: 锂离子电池;主成分分析;BP 神经网络;遗传算法优化;容量增量曲线;容量估算

中图分类号: TM912 **文献标识码:** A **国家标准学科分类代码:** 460.99

Lithium battery capacity estimation method based on PCA and GA-BP neural network

Wu Qiong Xu RuiLiang Yang Qingxia Xu Liyou

(School of Vehicle and Traffic Engineering, Henan University of Science and Technology, Luoyang 471003, China)

Abstract: Aiming at the problem that the capacity estimation method of lithium-ion battery for vehicle is not high precision, a method of residual capacity estimation of lithium-ion battery based on BP neural network optimized by genetic algorithm is proposed in this paper. First, after collating NASA's lithium-ion battery data set, the peak value of battery capacity increment curve under different health conditions was obtained. Secondly, the health factor was analyzed by principal component analysis to reduce its dimension, and the connection weight of BP neural network was optimized by genetic algorithm to predict the capacity of lithium ion battery. Finally, the model was validated on different NASA batteries. The results show that the proposed method can accurately estimate the capacity of four kinds of lithium ion batteries under different training amounts, and the square mean error of the estimation is less than 2%, and the proposed method has higher prediction accuracy than the prediction results without genetic algorithm optimization.

Keywords: lithium ion battery; principal component analysis; BP neural network; genetic algorithm optimization; capacity increment curve; capacity estimation

0 引言

锂离子电池目前被广泛应用于电动汽车中^[1]。在日常使用中,随着锂电池充放电循环次数的增加,其容量衰减也随之进行。当锂离子电池老化到一定程度后不仅耐用性大幅下降而且存在安全隐患,因此为锂电池的高效利用,并为老化锂电池的及时更换提供关键的参考依据,有效估计当前锂电池的剩余容量具有重大意义^[2]。

当前,数据驱动是锂离子电池容量预测的主要方法之

一^[3]。其中,反向传播(back propagation, BP)神经网络算法具有高度的非线性,自学能力强,被广泛的应用于锂离子电池状态估计。例如通过 BP 神经网络模型对电池荷电状态(state of charge, SOC)进行估算^[4-5]或通过对神经网络模型的梯次利用估计锂电池容量^[6]。然而,单纯的 BP 神经网络模型存在全局搜索能力不足、局部最优等缺点。将遗传算法(genetic algorithm, GA)应用到优化 BP 神经网络中就能弥补这些缺点,遗传算法是一种模拟生物进化过程的人工智能方法,是现阶段最具代表性的一种非

收稿日期:2021-12-09

^{*} 基金项目:国家重点研发计划项目(2016YFD0701002)、河南省科技攻关项目(212102210328)资助

线性拟合方法,可以合理反映具有非线性关系的输入量与输出量间的关联性,遗传算法不仅具有全局最优的特点,而且可以使BP神经网络的连接权值得到优化,能够很好地克服BP神经网络陷入局部最优的缺陷^[7]。因此,本文采用遗传算法对BP神经网络进行优化,从而提高模型的预测精度。

另外,对电池进行容量估计时,选取不同的特征参数将影响基于数据驱动方法的估计精度。目前的基于数据驱动的电池状态估算方法大多采用易检测的参数作为模型的输入,如只引入放电电压样本熵作为特征参数,但未分析特征参数间的关联性^[8];以及分别采用放电起始、末端总电压、放电深度、最大回馈电流与电压、电流、温度作为特征参数,不能充分表征电池的退化信息^[9-10]。研究发现^[11],锂离子电池的健康状态与相同时间间隔内电池放电电压差的变化密切相关。因此本文引入锂离子电池容量增量(incremental capacity, IC)分析法,该方法既能辨识电池老化机制,也估算电池健康状态。故本文在使用数据驱动的方法估算电池剩余可用容量时,引入容量增量曲线峰值作为特征参数之一。同时,本文利用主成分分析法(principal component analysis, PCA)将影响动力电池性能衰退的多个特征参数进行融合,既能够全面包含电池的衰退信息,又可以消除特征参数间的冗余信息。基于上述分析,本文提出一种基于主成分分析和遗传优化的BP算法(GA-BP)的锂离子电池容量估算方法,并引入锂离子电池IC曲线峰值作为特征值。该方法不仅在保留电池退化程度相关信息的同时去除冗余,又利用GA算法优化BP神经网络模型的连接权值,进一步提高了BP神经网络模型的估算精度和计算速度。

1 基于主成分分析的健康因子构建

1.1 锂离子电池性能退化状态参数提取

锂离子电池内部化学变化原理非常复杂^[12],能反映其退化状态的物理量难以直接测量,但电池的衰退程度可以通过一些可在线测量获得的状态参数反映。

对于锂离子电池来说,当它以恒流充电或放电时,通过测量锂离子电池的端电压 V ,可以得到 $dQ/dV-V$ 曲线。 $dQ/dV-V$ 曲线通常称为容量增量曲线。 $dQ/dV-V$ 曲线可以反映锂离子电池在不同端电压时单位电压下的充入或放出的电量。锂离子电池的健康状态在该曲线的某些特征上可以得到体现。

以锂离子电池恒流放电过程为例:

$$\frac{dQ}{dV} \approx \frac{\Delta Q}{\Delta V} = \frac{\Delta t \cdot I}{\Delta V} \quad (1)$$

其中, Δt 为时间变化值, I 为恒流放电的电流, ΔV 为 Δt 内的电压变化值。

本文利用美国国家航空航天局(National Aeronautics and Space Administration, NASA)的4组锂离子电池(B5、

B6、B7和B18)充放电循环数据进行后续研究。首先,对B5号电池中各循环中恒流放电阶段的电压曲线进行求导,获取 $dQ/dV-V$ 曲线,第15、60、100、125次循环的IC曲线如图1所示。

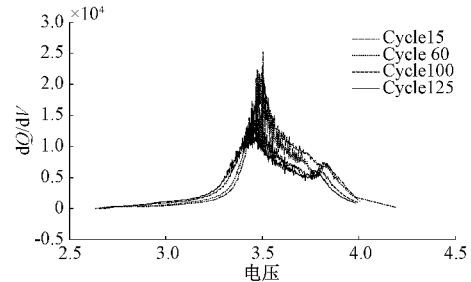


图1 不同循环次数时的IC曲线

在图1中,各循环的IC曲线峰值有较为明显的变化。循环次数越大,即老化程度越高,该峰值越小。在此基础上,可以得到各个循环次数对应的IC曲线峰值和容量值,如图2所示。

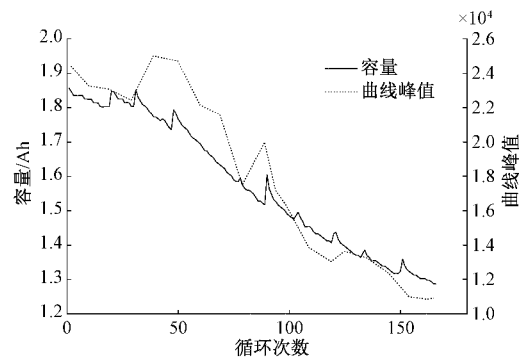


图2 IC曲线峰值与容量随循环次数的变化

从图2可以看出,随着循环次数的增加,锂离子电池容量减小,也就是SOH减小的同时,IC曲线峰值也随之降低,两者具有强烈的相关性。因此,考虑将IC曲线峰值作为估算锂离子电池剩余容量的健康因子。

另外,为了充分反映电池容量的衰减程度,本文选择恒流放电时间、放电功率、放电电压差值、放电平均温度、电池端电压及IC曲线峰值6个具有代表性的电池性能退化参数进行锂离子电池容量预测。

1.2 主成分分析

为防止参数间存在信息重叠、冗余的问题,使后续运算变得复杂,本文利用主成分分析对健康因子去除冗余的融合处理,将累计贡献率超过90%的主成分作为动力电池性能衰退的融合特征因子。

原始指标变量的数据预处理、标准化后记为 x_1, x_2, \dots, x_n ,其综合指标变量即原始数据降维后的主成分记为 $y_1, y_2, \dots, y_m (m \leq n)$,两者之间的关系为:

$$\begin{cases} y_1 = l_{11}x_1 + l_{12}x_2 + \dots + l_{1n}x_n \\ y_2 = l_{21}x_1 + l_{22}x_2 + \dots + l_{2n}x_n \\ \vdots \\ y_i = l_{i1}x_1 + l_{i2}x_2 + \dots + l_{in}x_n \\ \vdots \\ y_m = l_{m1}x_1 + l_{m2}x_2 + \dots + l_{mn}x_n \end{cases} \quad (2)$$

式中:系数 $l_{i2}, l_{i2}, \dots, l_{im} (i=1, 2, \dots, m)$ 构成系数矩阵。

主成分分析算法主要步骤如下:

1) 为消除特征参数间不同的量纲影响,对预处理后的原始数据矩阵 \mathbf{K} 进行正态分布标准化处理,得到标准化矩阵 \mathbf{X} 。

2) 计算相关系数矩阵 \mathbf{R} , 即:

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1p} \\ r_{21} & r_{22} & \dots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \dots & r_{pp} \end{bmatrix} \quad (3)$$

3) 计算特征值与特征向量。

4) 计算贡献率 ω_i 和累计贡献率 ρ 。

$$\omega_i = \lambda_i / \sum_{j=1}^n \lambda_j \quad (4)$$

$$\rho = \sum_{i=1}^m \lambda_i / \sum_{j=1}^n \lambda_j \quad (5)$$

式中: λ_i 和 λ_j 均表示方差特征值。

本文选取累计贡献率超过 90% 的健康因子作为预测模型的输入矩阵。

2 遗传算法优化 BP 神经网络构建

2.1 BP 神经网络模型的建立

神经网络算法具有高度的非线性,强大的自学能力,能够将输入输出之间的内在规律很好地反映^[13],故本文选取 BP 神经网络算法对电池剩余容量进行估计。

BP 神经网络由输入层、隐含层和输出层组成,如图 3 所示。

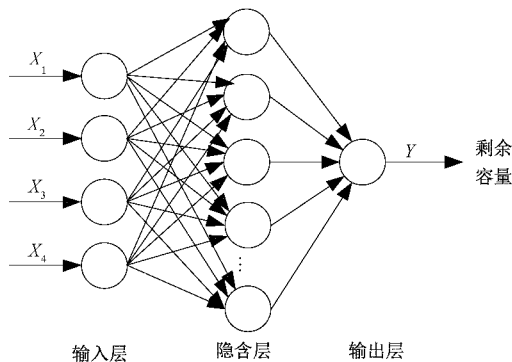


图 3 基于 BP 神经网络的容量预测模型

由经验公式 $p = \sqrt{m+n} + q$, q 取 $[1, 10]$ 之间的常数,经多次实验对比,隐含层数 m 确定为 12 时,仿真效果最优。

2.2 遗传算法优化

锂离子电池容量估计实质上是函数极值寻优问题,利用 BP 神经网络的非线性拟合能力和遗传算法的非线性寻优能力来求函数极值,得到最优的参数解。

遗传算法应用于 BP 神经网络中可以弥补全局搜索能力不足、易陷入局部最优等缺点,它通过选择、交叉以及变异等运算来优化 BP 神经网络的连接权值,从而在不需要具体数学模型的情况下,模拟各个工艺参数的变化规律。遗传算法参数:种群规模为 30,最大进化迭代次数为 60,变异概率为 0.02,交叉概率为默认值。遗传算法优化 BP 神经网络流程如图 4 所示。

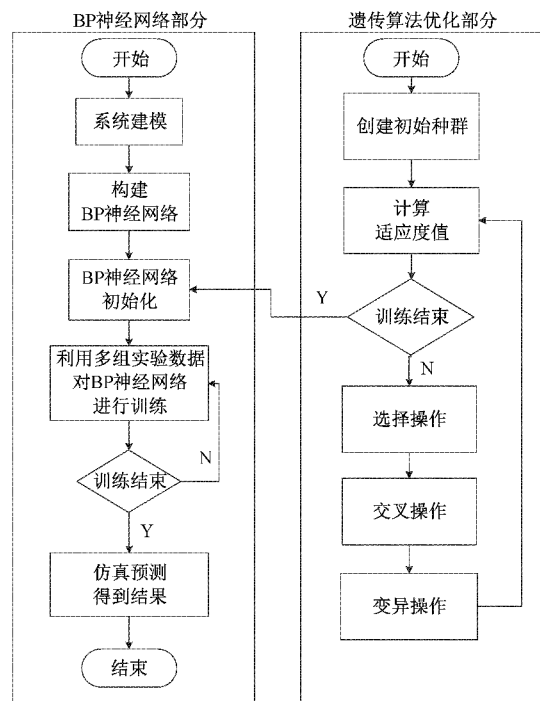


图 4 遗传算法优化 BP 神经网络流程

3 估算结果与分析

3.1 锂离子电池数据分析

试验所用原始数据是来自美国国家航空航天局的 4 组电池 (B5、B6、B7、B18) 充、放电循环数据^[14]。本文以 B5 号电池数据为例,提取放电过程中可以代表电池衰退的 6 个健康因子,即 IC 曲线峰值 H 、恒流放电时间 T 、放电功率 P 、放电电压差值 ΔU 、放电平均温度 T_1 、电池端电压 U 。采用主成分分析算法去除各健康因子之间的冗余部分,用更少的变量表达准确的退化状态。各主成分的特征值、贡献率和累积贡献率如表 1 所示。

表 1 可以看出前 4 个主成分的累计贡献率已经达到 95.381%,所包含的特征参数信息能够较好地反映出原有信息,故利用前 4 个主成分作为动力电池退化的表征参数。

表1 主成分特征值、贡献率、累计贡献率

成分	特征值	贡献率/%	累积贡献率/%
M_1	3.562	53.890	53.890
M_2	1.280	16.285	70.175
M_3	1.012	12.463	82.638
M_4	0.472	12.743	95.381
M_5	0.302	2.316	97.697
M_6	0.216	2.014	99.711

3.2 遗传算法的优化结果

通过神经网络的非线性拟合能力和遗传算法的非线性能力寻找到最恰当的电池容量与健康因子间的关系,再通过实验予以验证。遗传算法优化神经网络模型的适应度曲线如图5所示。

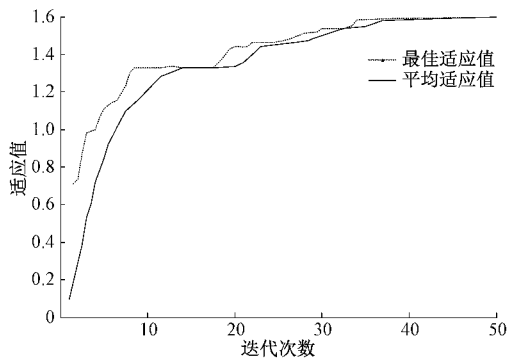


图5 遗传神经网络模型训练的适应度曲线

从图5可以看出,随着迭代次数的增加,适应度曲线逐渐上升,种群的平均适应度也在增加,种群在不断迭代进化的过程中,适应度高的个体竞争胜出,留在种群中,当迭代

35次时,最佳适应度曲线接近一条直线,说明该种群已经达到稳态,种群中适应度高的个体随之达到顶峰,满足遗传算法优化神经网络对权值和阈值的要求。

3.3 容量预测结果及分析

为了验证本文提出的遗传算法优化BP神经网络的锂离子电池容量估算方法的准确性,本节将在NASA锂离子4组电池数据的基础上进行锂离子电池容量估算。在验证时分别对不同数量的训练集以及是否使用遗传算法优化BP神经网络进行电池容量的估计,从而分析不同情况下对电池容量估计的影响。

1)不同训练量对结果的影响

在数据驱动的电池容量估算方法中,60%的数据主要用于训练模型,40%用于测试模型^[15]。为验证所提出模型的准确性和鲁棒性,本文分别选取每个数据集的50%、60%、70%数据作为训练集,剩下的数据作为测试集。数据输入模型后,不同训练量下模型的训练测试结果如图6所示。本文采用平均绝对百分误差 M_{MAPE} 、均方根误差 R_{RMSE} 和平均绝对误差 M_{AE} 来评价试验指标。

$$M_{MAPE} = \frac{1}{N} \sum_{i=1}^N \left| \frac{Y_i - Y'_i}{Y_i} \right| \tag{6}$$

$$R_{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (Y_i - Y'_i)^2} \tag{7}$$

$$M_{AE} = \frac{1}{N} \sum_{i=1}^N |Y_i - Y'_i| \tag{8}$$

式中: Y_i 为第*i*个样本的真实值; Y'_i 为第*i*个样本的预测值; N 为样本数量。

M_{MAPE} 、 R_{RMSE} 和 M_{AE} 的值越小,说明所建立模型得到的预测值与真实值越接近,模型性能越好。模型在不同训练量下的结果指标评价如表2所示。

表2 不同电池在不同训练量下的误差

电池标号及训练量	B5 电池			B6 电池			B7 电池			B18 电池		
	50%	60%	70%	50%	60%	70%	50%	60%	70%	50%	60%	70%
M_{MAPE}	0.034 9	0.015 1	0.015 2	0.033 9	0.035 3	0.022 2	0.021 8	0.005 1	0.018 5	0.030 8	0.003 9	0.030 7
R_{RMSE}	0.051 2	0.023 0	0.023 6	0.059 1	0.054 8	0.037 9	0.036 7	0.008 9	0.001 1	0.044 8	0.006 6	0.046 4
M_{AE}	0.046 9	0.020 9	0.020 2	0.044 1	0.046 0	0.028 7	0.032 4	0.007 4	0.027 3	0.044 0	0.005 4	0.043 9

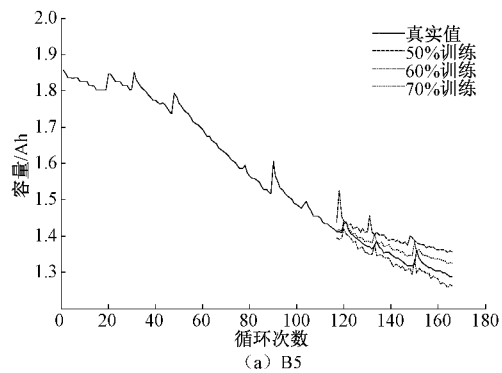
从训练测试结果来看,当使用60%的数据对模型训练时,该模型对每一组电池剩余容量的估算均得到了稳定准确的结果,该结果与真实值非常接近。从表2中的误差值可以看到, M_{MAPE} 最大出现在B6号电池的预测部分,仅为0.0353,最大 R_{RMSE} 为0.0548, M_{AE} 为0.046,说明所建立的模型对B5、B6、B7、B18号电池均能得到较为准确的剩余容量预测结果。

从图6中可以看到,当采用70%的数据用来训练模型时,模型预测的准确性有所提高,此时B6号电池预测结果的 M_{MAPE} 为0.0222, R_{RMSE} 为0.0379, M_{AE} 为0.0287。当

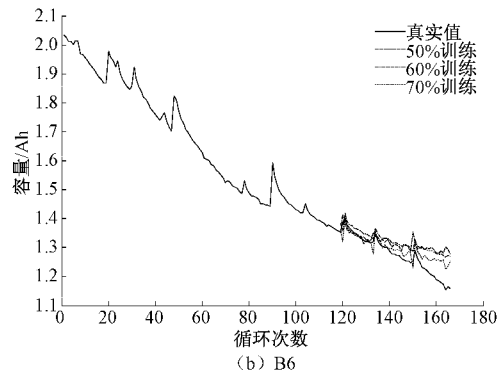
训练量降至50%时,模型的预测精度也只有略微降低,其中B5号电池的估算误差较大, M_{MAPE} 为0.0349, R_{RMSE} 为0.0512, M_{AE} 为0.0469,4种电池的估计精度与采用60%数据作为训练集时没有较大的差别。说明该模型具有良好的鲁棒性和泛化能力,能够在不同训练集数据量的情况下对不同类型电池的剩余可用容量进行准确估计。

2)遗传算法优化BP神经网络对结果的影响

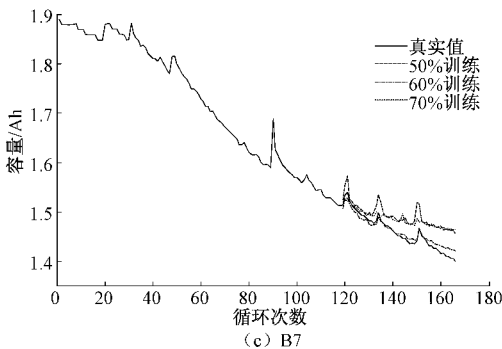
为进一步验证提出的利用遗传算法优化BP神经网络估算模型在精确度上的优势,将采用不同算法优化后的预测值与未采用算法优化后的预测值进行比较,训练测试结



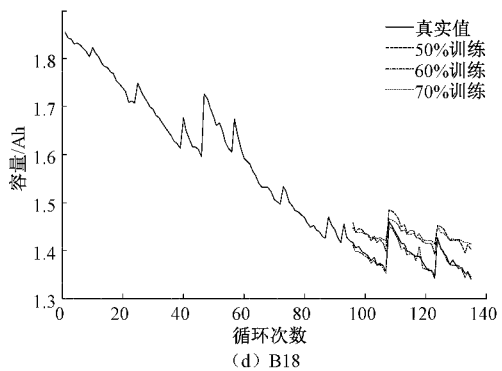
(a) B5



(b) B6

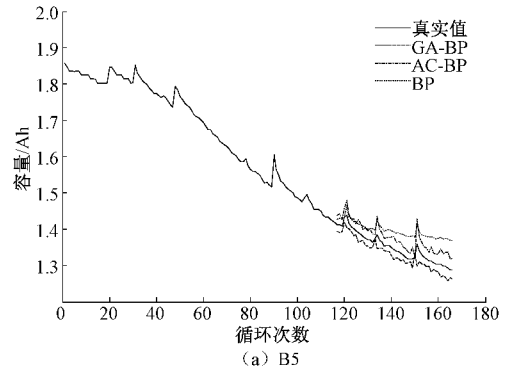


(c) B7

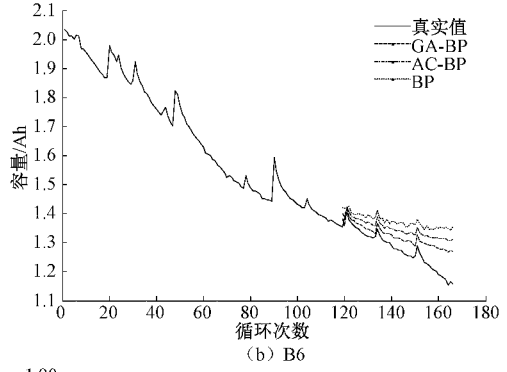


(d) B18

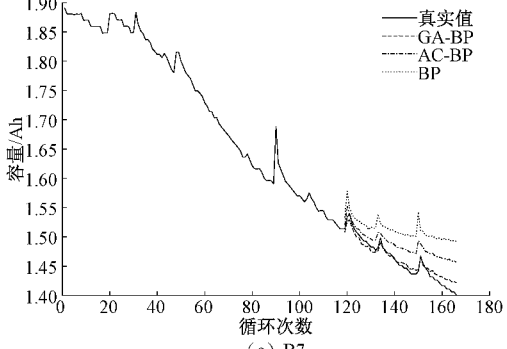
图 6 B5、B6、B7、B18 号电池在不同训练量下的训练测试结果



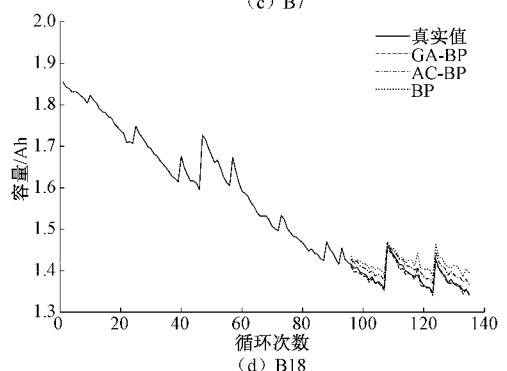
(a) B5



(b) B6



(c) B7



(d) B18

图 7 B5、B6、B7、B18 号电池采用不同算法优化的训练测试结果

果如图 7 所示,训练测试结果指标评价如表 3 所示。

蚁群 (ant colony, AC) 算法是通过模拟自然界蚂蚁的寻径方式得到的一种启发式随机搜索算法,蚁群算法具有全局搜索能力,分布式计算和鲁棒性强等特点,有利于加快 BP 传播神经网络的收敛速度,避免易陷入局部极值的

问题,提高建模精度,因此 AC-BP 算法常被用来建立预测模型。从图 7 可以看到,当未采用遗传算法优化时,模型的预测结果精度较差,最大的误差发生在 B6 号电池预测部分, M_{MAPE} 为 0.0706, R_{RMSE} 为 0.1062, M_{AE} 为 0.0964。当采用蚁群算法优化时,模型的预测结果精度略得到提

表3 不同电池在采用不同算法优化时的误差

电池标号及 优化方法	B5 电池			B6 电池			B7 电池			B18 电池		
	BP	AC-BP	GA-BP	BP	AC-BP	GA-BP	BP	AC-BP	GA-BP	BP	AC-BP	GA-BP
M_{MAPE}	0.033 5	0.021 7	0.015 8	0.070 6	0.049 1	0.035 3	0.034 5	0.017 2	0.005 1	0.021 5	0.0016 8	0.003 9
R_{RMSE}	0.050 9	0.038 2	0.023 0	0.106 2	0.080 4	0.054 8	0.056 7	0.026 0	0.008 9	0.010 3	0.008 1	0.006 6
M_{AE}	0.215 8	0.179 1	0.144 5	0.096 4	0.065 9	0.046 0	0.052 0	0.029 6	0.007 4	0.030 4	0.019 7	0.005 4

高,最大的误差发生在 B6 号电池预测部分, M_{MAPE} 为 0.049 1, R_{RMSE} 为 0.080 4, M_{AE} 为 0.065 9。当采用遗传算法优化后,由测试训练结果可以看到,该模型对 4 组电池的剩余可用容量同样得到了稳定准确的估算结果,该结果与真实值数据非常接近,且优于蚁群算法优化后的结果。

从表 3 可以看到,在采用遗传算法优化后,最大的误差值出现在 B6 号电池预测部分, M_{MAPE} 仅为 0.035 3, R_{RMSE} 为 0.054 8, M_{AE} 为 0.046 0,与未采用遗传算法优化及采用蚁群算法优化的结果相比误差值大幅降低。此外,另外 3 种电池的估算误差也与未采用遗传算法优化及采用蚁群算法优化的结果有所降低。说明所提出的采用遗传算法优化所建立的容量估算模型能在不同类型电池的剩余可用容量估算中得到精确的结果。

4 结 论

本文在提出将容量增量曲线峰值作为特征值的基础上建立 GA-BP 神经网络模型,对锂离子电池剩余可用容量进行估计研究。具体而言,本文首先采用主成分分析法对表征电池衰退程度的 6 个健康因子进行降维,然后通过遗传算法优化 BP 神经网络的连接权值,建立基于遗传算法优化的 BP 神经网络模型,进而对锂离子电池容量进行预测。在 NASA 锂离子电池数据集上的实验结果表明,本文提出的估算模型可以在不同训练量的情况下有效估算锂离子电池的剩余可用容量,并与未使用遗传算法优化及采用蚁群算法优化的估算结果相比,在估算精度上有了大幅度提高。该方法能够高效、准确地估算锂电池老化后的剩余可用容量。

参考文献

- [1] 魏业文,李应智,曹斌,等.含 Buck 电路的锂电池低功耗电量均衡技术研究[J].电工技术学报,2018,33(11):2575-2583.
- [2] WANG X, WEI X, DAI H. Estimation of state of health of lithium-ion batteries based on charge transfer resistance considering different temperature and state of charge[J]. Journal of Energy Storage, 2019, DOI: 10.1016/j.est.2018.11.020.
- [3] 刘大同,宋宇晨,武巍,等.锂离子电池组健康状态估计综述[J].仪器仪表学报,2020,41(11):1-18.

- [4] 张传伟,李林阳,赵东刚.基于 BP 神经网络法估算动力电池 SOC[J].电源技术,2017,41(9):1356-1357,1368.
- [5] 杨学平,王正江,蒋超宇,等.基于 BP 神经网络法研究锂电池荷电状态[J].材料导报,2019,33(S2):53-55.
- [6] 王玉坤,姜久春,张彩萍,等.基于神经网络模型的梯次利用锂电池容量估计[J].电源技术,2014,38(4):632-635.
- [7] 张彦粉,魏华,葛纪者,等.基于遗传算法优化 BP 神经网络的可食用油墨粘度的预测[J].包装工程,2021,42(19):49-54.
- [8] WIDODO A, SHIM M, CAESARENDRA W, et al. Intelligent prognostics for battery health monitoring based on sample entropy [J]. Expert Systems with Applications, 2011, 38(9): 11763-11769.
- [9] 梁丹阳,程相,郝建国,等.基于多特征融合的动力电池 RUL 预测[J].中国测试,2021,47(12):149-156.
- [10] 张岸,杨春德.基于 GAN-CNN-LSTM 的锂电池 SOH 估计[J].电源技术,2021,45(7):902-906.
- [11] 陈继永,吴兆宏,李金喜.基于容量增量法的防爆锂电池老化指标分析[J].工矿自动化,2019,45(12):29-34.
- [12] 郝世宇,殷会飞,杨茹,等.基于 AEKF 的锂离子动力电池荷电状态估计[J].国外电子测量技术,2021,40(9):49-53.
- [13] 张少宇,伍春晖,熊文渊.采用门控循环神经网络估计锂离子电池健康状态[J].红外与激光工程,2021,50(2):236-243.
- [14] YANG Z, BO G. Online capacity estimation of lithium-ion batteries based on novel feature extraction and adaptive multi-kernel relevance vector machine [J]. Energies, 2015, 8(11): 12439-12457.
- [15] 张雪霞,高雨璇,陈维荣.基于数据驱动的质子交换膜燃料电池寿命预测[J].西南交通大学学报,2020,55(2):417-427.

作者简介

吴琼,硕士研究生,主要研究方向为动力电池状态估计。

徐锐良(通信作者),副教授,硕士研究生导师,主要研究方向为车辆现代设计方法。
E-mail: lyxurl@163.com