

基于IPSO-LightGBM模型的锂离子电池荷电状态预测

任小强, 何 青, 唐晓华

(西南交通大学希望学院信息工程系, 四川 成都 610400)

摘 要: 为有效提升锂离子电池的荷电状态 (State of Charge, SOC) 预测精度, 提出一种基于改进粒子群优化算法 (Improved Particle Swarm Optimization, IPSO) 和轻量级梯度提升机 (Light Gradient Boosting Machine, LightGBM) 的锂离子电池SOC预测模型, LightGBM模型用于构建锂离子电池SOC的预测, IPSO用于优化LightGBM模型的超参数。首先, 对公开数据集进行预处理, 并使用多种策略改进基本粒子群优化算法; 其次, 建立基于IPSO-LightGBM、LightGBM和反向传播 (Back Propagation, BP) 神经网络的锂离子电池SOC预测模型; 最后, 使用马里兰大学提供的电池数据集对三种模型进行实验仿真。结果表明, IPSO-LightGBM模型的预测准确率优于未优化的LightGBM模型和BP神经网络模型。

关键词: 荷电状态; 锂离子电池; 粒子群优化算法; 轻量级梯度提升机

中图分类号: TP181; TM912; U469.72 **文献标志码:** A **文章编号:** 1672-6332 (2024) 04-0049-07

锂离子电池(Lithium-ion Battery, LIB)的荷电状态被定义为剩余可用容量与额定容量的比值^[1], 是估计电动汽车 (Electric Vehicle, EV) 剩余行驶里程的基础。SOC无法直接测量, 一般通过估计获得。不准确的估计结果会引起电池过充、深度放电和高危运行等风险, 从而加剧电池寿命的减少。因此, LIB的SOC准确预测具有学术及工程研究意义。对于LIB的SOC估计, 国内外学者进行了大量的研究, 提出了三种估计方法, 分别为基于模型的方法^[2]、基于试验的方法^[3]和基于数据驱动的方法^[4]。基于模型的方法能较准确地描述电池的动静态特性, 估计过程不需要大量的历史数据, 但构建精确模型所需参数较多, 存在一定难度^[5]。基于试验的方法原理简单、易实现, 实际应用局限性大^[6]。随着机器学习的快速发展与完善, 基于数据驱动的

方法被广泛应用于SOC的估计中^[7]。该方法仅通过训练采集到的数据就可以计算得到准确的SOC, 极大地简化了建模过程, 其泛化性与鲁棒性强。使用数据驱动的方法对SOC 进行估计时, 常借助的工具具有神经网络 (Neural Network, NN)^[8]、深度学习 (Deep Learning, DL)^[9]、支持向量机 (Support Vector Machine, SVM)^[10]和极限学习机 (Extreme Learning Machine, ELM)^[11]等。文献[12]使用BP神经网络对LIB的SOC进行了估计, 并使用麻雀搜索算法 (Sparrow Search Algorithm, SSA) 优化BP神经网络的初始参数, 均方根误差能够控制在2.3%以下。文献[13]利用双向门控循环单元 (Bidirectional Gated Recurrent Unit, BiGRU) 神经网络模型对LIB的SOC进行了估计, 实验结果表明该模型的预测误差控制在0.776%以内。文献[14]利用带有核函数的

[收稿日期] 2024-02-15

[基金项目] 成都市哲学社会科学重点研究基地成都市交通 + 旅游大数据应用技术研究项目 (项目编号: 20231012); 四川省教育信息化与大数据中心项目 (项目编号: DSJZXKT264); 西南交通大学希望学院2024年青年科研项目 (项目编号: 2024053)。

[作者简介] 任小强 (1989-), 男 (汉), 甘肃人, 硕士, 讲师, 主要研究方向: 云计算与数据挖掘。

ELM来提高LIB的SOC估算的准确性和鲁棒性,仿真实验结果表明,ELM在训练精度及泛化能力方面等具有较为明显的优势。轻量级梯度提升机器学习(Light Gradient Boosting Machine, LightGBM)是一种梯度提升树(Gradient Boosting Decision Tree, GBDT)算法,具有训练速度快、预测精度高和内存消耗低等优点,已在各个领域得到了广泛应用^[15-16],但鲜有文献对LightGBM模型的超参数进行优化来提高模型的全局搜索能力^[17]。

针对上述问题,本文提出了改进粒子群优化算法IPSO(Improved Particle Swarm Optimization, IPSO)和LightGBM相结合的IPSO-LightGBM模型,对LIB的SOC进行估计。该模型利用IPSO算法对LightGBM模型的超参数进行优化,并通过公开数据集验证了该模型在处理大规模数据集时,也能够对LIB的SOC做出更加准确的预测,具有应用推广价值。

1 荷电状态预测模型

1.1 LightGBM模型

LightGBM模型是由微软研究院在2017年开发的一种实现梯度提升决策树(Gradient Boosting Decision Tree, GBDT)算法的开源机器学习模型^[18],具有更快的训练速度、更低的内存消耗、更高的准确率、高效率的并行训练和快速处理海量数据等优点,被广泛应用于求解各类机器学习任务,其基本原理如式(1)所示, $T(x; \theta_i)$ 为单个决策树, m 表示总迭代轮数(决策树个数), θ_i 为决策树参数。

$$f(x) = \sum_{i=1}^m T(x; \theta_i) \quad (1)$$

LightGBM的分割点选择是基于直方图的算法,如图1所示,先把连续的特征值离散化成 k 个整数,同时构造一个宽度为 k 的直方图,根据直方图的离散值,遍历寻找最优分割点。LightGBM使用基于叶子节点(leaf-wise)的分裂策略,如图2所示,每次分裂选择增益最大的点分裂,而不是每层对每一个点都进行分裂,即按层生长策略(level-wise)。同时LightGBM采用基于梯度的单边采样算法(Gradient-Based One-Side Sampling, GOSS)减少样本数量、互斥特征捆绑算法(Exclusive Feature

Bundling, EFB)减少特征数量、多线程并行计算降低通信和计算开销以及使用高效的内存管理等技术,进一步提高模型的训练精度。此外,LightGBM还提供了C、C++和Python等语言接口,为实时LIB的SOC预测提供了实现方案。

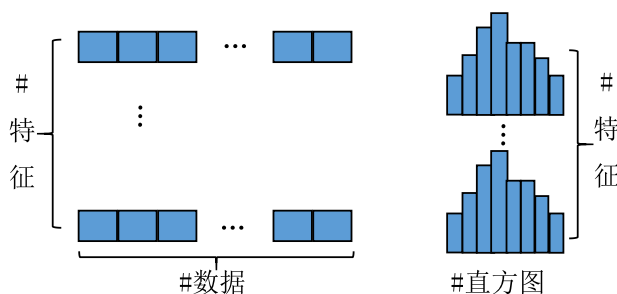


图1 基于直方图寻求最佳分裂点策略

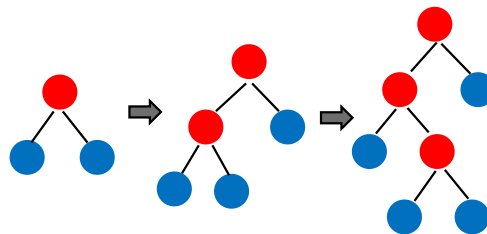


图2 基于叶子节点的分裂策略

1.2 基本粒子群优化算法

PSO算法是一种基于群体协作的启发式算法,每个粒子代表一个可能的解,每个粒子会根据自己的当前个体极值和全局最优解来更新自己的速度和位置,具有运行速度快,参数较少等特点。假设在一个 D 维搜索空间中,有 N 个初始搜索粒子的种群 $G=(G1, G2, \dots, GN)$,第 k 代第 i 个粒子的位置向量为 $X_i^k=(x_{i1}^k, x_{i2}^k, \dots, x_{iD}^k)^T, i=(1, 2, \dots, N)$;相应的飞行速度为 $V_i^k=(v_{i1}^k, v_{i2}^k, \dots, v_{iD}^k)^T$;截止 k 代,第 i 个粒子的个体最佳位置记为 $P_i^k=(p_{i1}^k, p_{i2}^k, \dots, p_{iD}^k)^T$;截止 k 代,整个种群的最佳位置记为 $P_g^k=(p_{g1}^k, p_{g2}^k, \dots, p_{gD}^k)^T$ 。由此,第 $k+1$ 代时,粒子 i 的第 d 维的速度和位置更新公式为:

$$v_{id}^{k+1} = \omega v_{id}^k + c_1 r_1 (p_{id}^k - x_{id}^k) + c_2 r_2 (p_{gd}^k - x_{id}^k) \quad (2)$$

$$x_{id}^{k+1} = x_{id}^k + v_{id}^{k+1} \quad (3)$$

式(2)中, ω 为惯性重系数; $d=1, 2, \dots, D$; $i=1, 2, \dots, N$; x_{id}^{k+1} 是第 t 次迭代时, 第 i 个粒子的位置矢量的第 d 维分量; v_{id}^{k+1} 为相应的速度矢量的第 d 维分量; c_1 和 c_2 表示学习因子; r_1 和 r_2 为(0, 1)之间的随机数。

1.3 多种策略改进粒子群优化算法

1.3.1 非线性递减惯性系数

惯性系数 ω 描述了粒子上一代速度对当前代速度的影响, 它是PSO算法中极其重要的一个参数, 控制其取值大小可有效地调节平衡PSO算法的全局与局部寻优能力。在基本PSO算法中, ω 根据经验选取, 算法迭代过程中保持不变, 易陷入局部最优^[19]。因此, 本文重新构建非线性递减惯性系数, 并引入权重控制因子, 如式(4)所示, 能够确保粒子群在初始状态时以较大的 ω 进行全局开发性搜索, 而在迭代后期又以较小的 ω 进行更为精细化的局部寻优。

$$\omega = \begin{cases} \omega_{\min} - \frac{(\omega_{\max} - \omega_{\min}) \times (f - f_{\min})}{f_{\max} - f_{\min}}, & f < f_{\max} \\ \omega_{\max}, & f \geq f_{\max} \end{cases} \quad (4)$$

式中, f 表示粒子实时的目标函数值; $f_{\max} = \frac{1}{N} \sum_{i=1}^N f(x_i^k)$ 和 $f_{\min} = \min(f(x_i^k))$ 分别表示当前粒子群的平均值和最小目标值。

1.3.2 动态平衡学习因子

学习因子 c_1 、 c_2 是PSO算法中的重要参数, 它的取值大小关系着PSO算法的收敛速度和粒子的多样性, 前者代表个体的自我认知, 后者代表粒子的社会认知, 初期需要关注个体自我认识的能力, 后期则应更注重个体获取社会信息的能力。受正弦余弦算法(Sine Cosin Algorithm, SCA)启发, 引入随机控制因子 λ , 当 $\lambda > 0.5$ 时, 使用SCA算法中的正弦项来代替PSO算法中的学习因子 c_1 与 c_2 , 否则, 使用SCA算法中的余弦项来代替^[20], 改进后的速度更新公式如式(5)所示, 其中, $R_1 = 2 \times (1 - t/T_{\max})$, t 是当前迭代次数, T_{\max} 是设置的最大迭代次数; $r_2 \in [0, 2\pi]$ 。

$$v_{id}^{k+1} = \begin{cases} \omega v_{id}^k + \sin(R_1)r_1(p_{id}^k - x_{id}^k) + \sin(R_1)r_2(p_{gd}^k - x_{id}^k), & \lambda > 0.5 \\ \omega v_{id}^k + \cos(R_1)r_1(p_{id}^k - x_{id}^k) + \cos(R_1)r_2(p_{gd}^k - x_{id}^k), & \lambda \leq 0.5 \end{cases} \quad (5)$$

1.3.3 重构速度更新机制

为了保持精英学习的轨迹和增加群体的多样性, 使得粒子的搜索区间更广, 能在迭代前期有更大可能搜索到全局最优解, 本文利用个体最优和群体最优的线性组合策略构建群体的均值位置 p_{mix}^k , 它的每一维度都随机选自每个粒子当前 p_g^k 所对应的某一维数据, 且不重复选取同一个粒子的多个维度, p_{mix}^k 的生成过程如图3所示。本文在执行速度更新时加入 p_{mix}^k 改进速度更新方式, 改进后的速度更新方式如式(6)所示, 其中 c_3 , r_3 为(0, 1)之间的随机数。

$$v_{id}^{k+1} = \omega v_{id}^k + c_1 r_1 (p_{id}^k - x_{id}^k) + c_2 r_2 (p_{gd}^k - x_{id}^k) + c_3 r_3 (p_{mix}^k - x_{id}^k) \quad (6)$$

1.4 基于IPSO-LightGBM的锂电池SOC预测

LightGBM算法在很多领域得到了良好的应用, 但该算法具有众多超参数, 不同超参数组合决定了算法的性能, 传统LightGBM模型的超参数通过随机初始化方式所得, 它的精度较低。针对这一问题, 提出将LightGBM模型训练的平均绝对值百分比误差(Mean Absolute Percentage Error, MAPE)作为粒子适应度函数 $f(t)$, 并通过反复迭代得到适应度函数最小的全局最优解, 即得到对应的最优超参数组合, 其次使用最优超参数对应的LightGBM模型进行SOC预测, 总体框架如图4所示。MAPE如式(7)、式(8)所示, 其中, soc_i' 为锂电池样本 i 的SOC预测值, soc_i 为锂电池样本 i 的SOC实际值, h 为预测样本数。

$$f(t) = \frac{1}{1 + MAPE(t)} \quad (7)$$

$$MAPE = \frac{1}{h} \sum_{i=1}^N \left| \frac{soc_i' - soc_i}{soc_i} \right| * 100\% \quad (8)$$

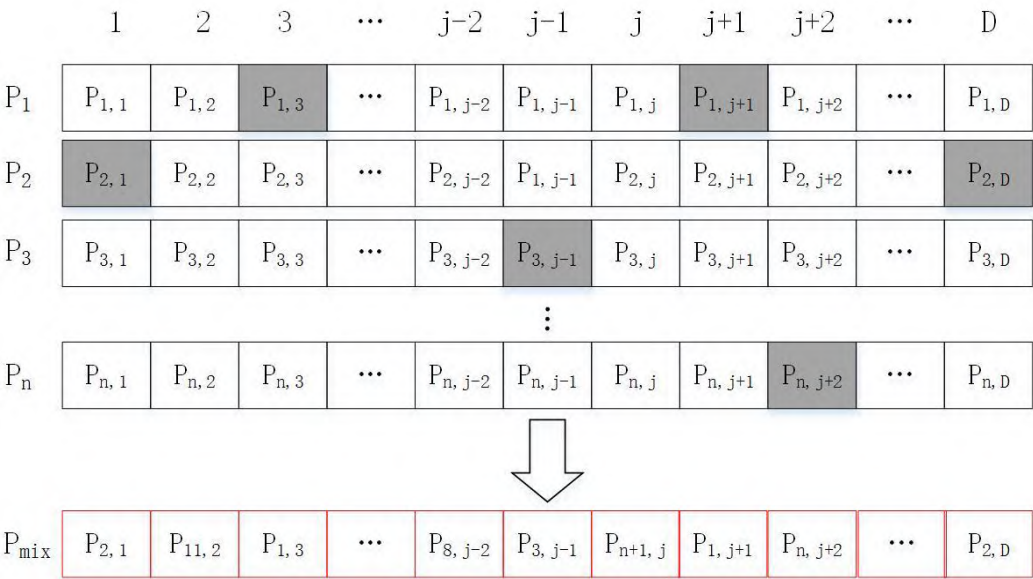


图3 重构粒子生成过程

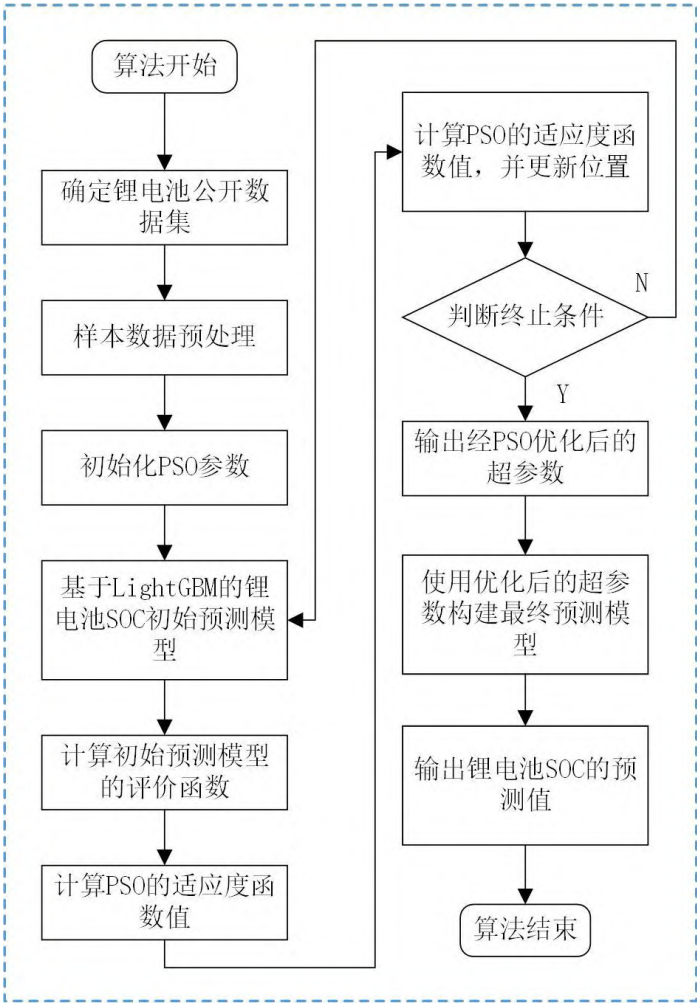


图4 IPSO-LightGBM模型的总体框架

2 仿真实验

2.1 数据集简介与数据预处理

实验采用马里兰大学先进生命周期工程研究中心 (Center for Advanced Life Cycle Engineering, CALCE) 提供的锂电池公开数据集^[21], 从中提取型号为INR18650-20R电池的动态应力测试 (Dynamic Stress Test, DST) 数据集进行模型训练, 该型号电池额定容量为2000mAh。数据预处理时, 首先, 对数据集的缺失值使用均值填充; 然后, 提取特征值和标签序列, 并利用式 (9) 中的标准归一化方法对输入数据进行处理, 将特征数据缩放至 $[0, 1]$ 区间内, 消除数据量纲不统一对训练模型的影响, 提高模型训练的收敛速度。最后, 将预处理后的数据集通过sklearn (全称scikit-learn) 分为训练集、验证集和测试集, 比例设置为6:2:2。

$$SOC_i^* = \frac{SOC_i - SOC_{min}}{SOC_{max} - SOC_{min}} \quad (9)$$

式中, SOC_i 为原始样本数据; soc_i^* 为归一化处理得到的数据; 分别为 SOC_{max} 和 SOC_{min} 分别为总样本数据的最大值和最小值。

2.2 仿真环境及参数设置

本实验基于CPU (Intel Xeon Platinum 8358 @ 2.60GHz)、RAM内存 (16GB)、Linux Lite 5.8操作系统和Python语言作为实验环境, 在sklearn机器学习工具下进行模型搭建及数据处理工作, 相关的参数设置如表1所示。

表1 IPSO-LightGBM模型的参数设置

参数	参数值
学习率	learning_rate=0.005~0.5
决策树的最大深度	max_depth=2~50
最大叶子节点数	num_leaves=10~64
最小权重和	min_child_samples=0.02~1
最小叶子节点样本数	min_child_samples=10~40
弱学习器数量	n_estimators=50~500
学习因子	c3 \in (0, 1)
惯性权值最大值	0.8
惯性权值最小值	0.2
种群数目	s=40
预测样本数	h=10 000
最大迭代次数	Tmax=100
随机选择因子	$\lambda \in (0, 1)$

3 实验结果与分析

为验证本工作所提模型对锂电池SOC的预测效果, 将本工作构建的IPSO-LightGBM模型与文献[22]中的LightGBM模型、文献[23]中的BP神经网络模型进行对比, 通过2.1小节中的公开数据集进行训练。其中, 最大迭代次数统一设置为100, LightGBM模型和BP神经网络模型的其它参数设置跟文献[22]和文献[23]相同。仿真测试结果如图5所示, 横坐标为时间, 纵坐标为SOC, 可以看出, 本工作建立的IPSO-LightGBM模型的预测效果优于其它两种模型, 它的预测结果与真实值大致相同, 表现出了更好的预测能力, 验证了IPSO-LightGBM模型的有效性。

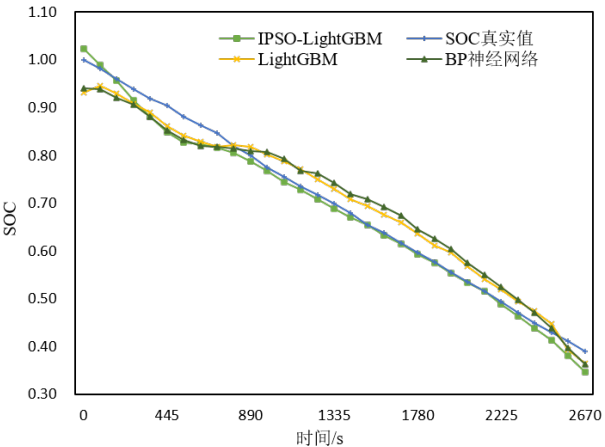


图5 不同模型的SOC估计结果对比图

平均绝对值百分比误差MAPE被用于衡量模型预测效果, 如式 (8) 所示, 其值越接近于0, 代表模型性能和预测精度越好, 不同模型的MAPE迭代曲线如图6所示, 其中横坐标为时间, 纵坐标为平均绝对值百分比误差。从图6可以看出, BP神经网络模型和LightGBM模型的MAPE上下波动范围较大, 最大值分别为12%和10%。IPSO-LightGBM模型可以确保MAPE在一定范围内波动, 平均绝对值百分比误差最大不超过6.5%。

图7表示IPSO-LightGBM模型与LightGBM模型和BP神经网络模型的收敛曲线图, 其中横坐标为迭代次数, 纵坐标为平均绝对百分比误差。从图7可以看出, IPSO-LightGBM模型的迭代次数为13时, 收敛到最优值, 基本达到网络设置精度要求,

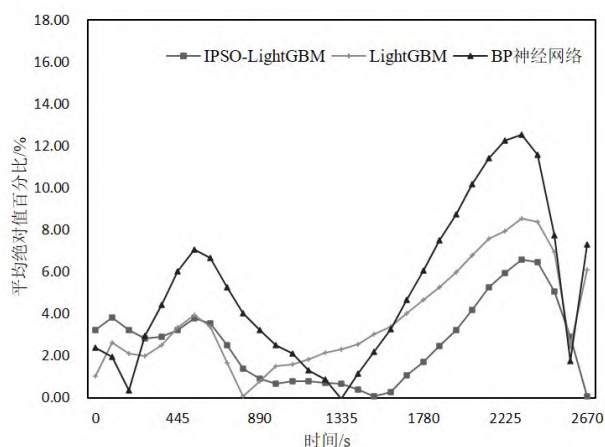


图6 不同模型的平均绝对值百分比误差曲线

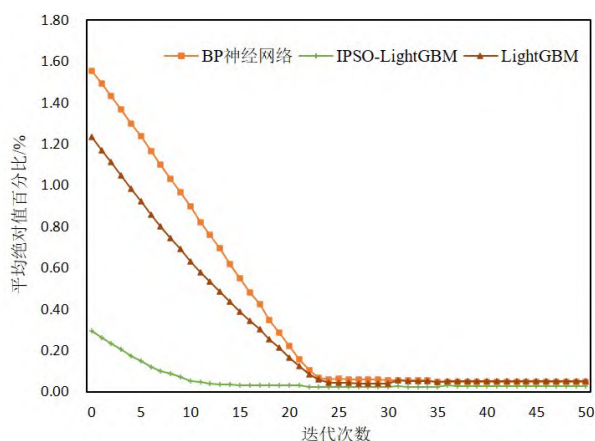


图7 不同模型的仿真误差迭代曲线

而LightGBM模型和BP模型要在19步和24步时，才收敛到最优值。由此可见，IPSO-LightGBM模型表现出更快的收敛速度。

图8为3种预测模型的平均绝对误差（Mean Absolute Error, MAE）、平均绝对值百分比误差（Mean Absolute Percentage Error, MAPE）和均方根误差（Root Mean Square Error, RMSE），从图8可以看出，本工作所建立的IPSO-LightGBM模型与

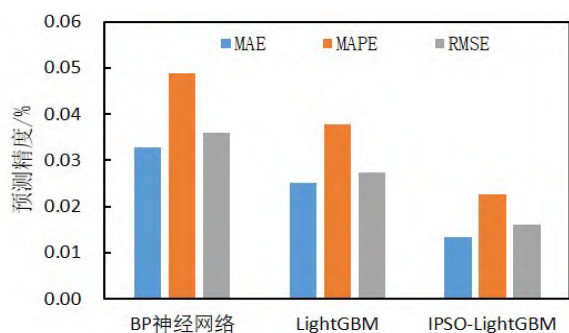


图8 不同模型的预测精度对比

LightGBM模型和BP神经网络模型相比，平均绝对误差、平均绝对值百分比误差和均方根误差最小，分别为0.012 %、0.023 %、0.019 %，验证了本工作所提的IPSO-LightGBM模型有更好的预测精度，对于锂电池SOC有着较好的估计效果。

4 结语

本文建立了一种基于IPSO-LightGBM的锂电池SOC预测模型，首先，采用粒子群优化算法寻找LightGBM模型的最优超参数，其次，使用最优超参数对应的LightGBM模型对锂电池SOC进行预测。在实验阶段，以马里兰大学先进生命周期工程研究中心的锂电池数据集为样本集，采用平均绝对百分比误差作为IPSO算法的适应度函数，对IPSO-LightGBM、LightGBM和BP神经网络模型进行仿真验证，对比了各个模型的预测结果。结果表明，IPSO-LightGBM模型的预测效果优于LightGBM和BP神经网络模型，验证了IPSO算法对LightGBM模型超参数优化的有效性。本文未考虑温度对锂电池SOC的影响，后续的研究将综合考虑更多的影响因素，同时结合更大规模的数据对锂电池SOC进行预测，进一步提高储能锂电池SOC估算的精度。

参考文献：

- [1] 万广伟,张强.锂离子电池SOC评估方法研究进展[J].电源技术,2023,47(9):1122-1125.
- [2] 吴逸洲,刘艳,祝现染,等.基于多模型融合的锂离子电池SOC自适应估计[J].电源技术,2023,47(9):1158-1163.
- [3] 董浩,毛玲,屈克庆,等.基于温度和SOC的锂离子电池特征提取及SOH估计[J].浙江大学学报(工学版),2023,57(7):1470-1478.
- [4] 于秋月,刘江岩,何林,等.基于机器学习的锂离子电池荷电状态多步预测[J].汽车工程学报,2023,13(4):586-596.
- [5] 尚彦赞,宋红为,杨照光,等.基于二阶RC模型的锂电池充放电特性分析[J].高压电器,2023,59(7):87-94.
- [6] 姚昌兴,李昕,邢丽坤.基于多时间尺度锂电池在线参数辨识及SOC和SOH估计[J].重庆工商大学学报(自然科学版),2023,40(5):48-54.
- [7] 张志冬,李云伍,李杨柳,等.几种机器学习算法的锂电池SOC估计研究[J].重庆理工大学学报(自然科学),2023,37(9):40-48.

- [8] 于仲安, 邵昊晖, 陈可怡. 基于IGWO-BP神经网络的锂离子电池SOC估计[J]. 电源技术, 2023, 47(9): 1153-1157.
- [9] 朱元富, 贺文武, 李建兴, 等. 基于Bi-LSTM/Bi-GRU循环神经网络的锂电池SOC估计[J]. 储能科学与技术, 2021, 10(3): 1163-1176.
- [10] 徐东辉. 基于混沌时间序列LS-SVM的车用锂离子电池SOC预测研究[J]. 车用发动机, 2019(2): 67-71.
- [11] 寇发荣, 罗希, 门浩, 等. 基于特征优选与改进极限学习机的锂电池SOC估计[J]. 储能科学与技术, 2023, 12(4): 1234-1243.
- [12] 张淞, 周永军, 蒋淑霞, 等. 基于SSA-BP改进EKF算法的锂电池SOC估算[J]. 电源技术, 2023, 47(8): 1050-1054.
- [13] 陈梦宇, 张杰. 基于PSO-SG-BiGRU模型估算锂电池SOC[J]. 武汉理工大学学报, 2023, 45(7): 117-123, 131.
- [14] 刘世林, 李德俊, 姚伟, 等. 基于核极限学习机与容积卡尔曼滤波融合的锂电池荷电状态估计[J]. 湖南大学学报(自然科学版), 2023, 50(10): 51-59.
- [15] 范桂英, 汤军, 高贤君, 等. 基于LightGBM的南阳市西部地区山洪灾害风险评价[J]. 中国农村水利水电, 2023(8): 135-141, 156.
- [16] 柯日宏, 吴升, 柯玮文. 一种识别共享单车潮汐点的时空模型和基于KNN-LightGBM的租还需求预测方法[J]. 地球信息科学学报, 2023, 25(4): 741-753.
- [17] 张博, 向旭, 贾俊龙, 等. 基于LightGBM的天然气管道周围滑坡灾害预测方法[J]. 吉林大学学报(理学版), 2023, 61(2): 338-346.
- [18] 谢军飞, 张海清, 李代伟, 等. 基于Lightgbm和XGBoost的优化深度森林算法[J]. 南京大学学报(自然科学), 2023, 59(5): 833-840.
- [19] 冯茜, 李擎, 全威, 等. 多目标粒子群优化算法研究综述[J]. 工程科学学报, 2021, 43(6): 745-753.
- [20] 雍龙泉, 黎延海, 贾伟. 正弦余弦算法的研究及应用综述[J]. 计算机工程与应用, 2020, 56(14): 26-34.
- [21] 谢旭, 高晗, 毕贵红, 等. 基于AdaBoost-Elman算法的锂离子电池荷电状态估计方法[J]. 电力科学与工程, 2021, 37(2): 48-56.
- [22] 肖迁, 穆云飞, 焦志鹏, 等. 基于改进LightGBM的电动汽车电池剩余使用寿命在线预测[J]. 电工技术学报, 2022, 37(17): 4517-4527.
- [23] 徐元中, 曹翰林, 吴铁洲. 基于SA-BP神经网络算法的电池SOH预测[J]. 电源技术, 2020, 44(3): 341-345.

Prediction of state of charge for lithium-ion batteries based on IPSO-LightGBM model

REN Xiaoqiang, HE Qing, TANG Xiaohua

(Information Engineering Department of Hope College, Southwest Jiaotong University, Chengdu, Sichuan, China 610400)

Abstract: To effectively improve the accuracy of State of Charge (SOC) prediction for lithium-ion batteries, a lithium-ion battery SOC prediction model based on Improved Particle Swarm Optimization (IPSO) algorithm and Light Gradient Boosting Machine (LightGBM) is proposed. The LightGBM model is used to construct SOC prediction for lithium-ion batteries, and IPSO is used to optimize the hyperparameters of the LightGBM model. Firstly, the public dataset is preprocessed, and multiple strategies are used to enhance the basic particle swarm optimization algorithm; secondly, lithium-ion battery SOC prediction models based on IPSO-LightGBM, LightGBM, and Back Propagation (BP) neural networks are established; finally, experimental simulations on the three models are conducted using the battery dataset provided by the University of Maryland. The results indicate that the prediction accuracy of the IPSO-LightGBM model surpasses that of the unoptimized LightGBM model and the BP neural network model.

Key words: state of charge, lithium-ion battery, particle swarm optimization algorithm, lightweight gradient boosting machine

【责任编辑: 杨立衡】