DOI：10.13234/j.issn.2095-2805.2026.4.122 **中图分类号**：TM615 **文献标志码**：A

基于 IDBO-LightGBM 的光伏阵列故障诊断方法

吴亚钧，王 璐，张金江(中国电源学会会员)
(浙江科技大学自动化与电气工程学院，杭州 310023)

摘要：为提高光伏阵列故障诊断的准确率，提出1种基于改进蜣螂优化IDBO(improved dung beetle optimazation)算法优化轻量级梯度提升机 LightGBM(light gradient boosting machine)的光伏阵列故障诊断方法。通过IDBO对LightGBM中的特定超参数进行寻优，建立基于IDBO-LightGBM的光伏阵列故障诊断模型。为有效区分各种故障，使用从电流-电压曲线中提取的特征点。为验证所提方法的效果，对模型进行仿真分析，将所提方法与LightGBM算法、DBO-LightGBM算法、其他决策树算法和支持向量机算法进行对比，验证了所提方法的稳定性与准确率。

关键词：光伏阵列；故障诊断；改进蜣螂优化算法；轻量级梯度提升机

Fault Diagnosis Method for Photovoltaic Arrays Based on IDBO-LightGBM

WU Yajun, WANG Lu, ZHANG Jinjiang, *Member*, *CPSS*
(School of Automation and Electrical Engineering, Zhejiang University of Science and Technology, Hangzhou 310023, China)

Abstract: To improve the accuracy of fault diagnosis for photovoltaic arrays, a photovoltaic fault diagnosis method based on an improved dung beetle optimization (IDBO) algorithm which is used to optimize a light gradient boosting machine (LightGBM) is proposed. The IDBO algorithm is used to optimize specific hyperparameters in LightGBM, and a fault diagnosis model for photovoltaic arrays based on IDBO-LightGBM is established. To effectively distinguish various faults, the characteristic points extracted from the current-voltage curves are utilized. To validate the effectiveness of the proposed method, simulation analysis is conducted on the model. A comparison is performed among the proposed method, LightGBM algorithm, DBO-LightGBM algorithm, other decision tree algorithms and support vector machine algorithm, and results verify the stability and accuracy of the novel method.

Keywords: Photovoltaic array; fault diagnosis; improved dung beetle optimization (IDBO) algorithm; light gradient boosting machine (LightGBM)

近几年，光伏发电在我国飞速发展，光伏并网发电规模不断扩大[1]。光伏阵列面积大、分布广泛，在恶劣的自然环境下运行容易发生组件老化、局部遮阴、开路和短路等问题[2-3]。这些问题严重影响光伏系统的健康、光伏电站整体的运行及发电效率[4]。因此，对光伏阵列的故障进行准确高效地诊断十分重要。

对于光伏阵列故障诊断的问题，文献[5]利用红外热图像检测光伏阵列出现的热斑效应，但该方法使用成本较高且对红外热图像的精度有较高要求，无法大规模使用；文献[6]通过光伏阵列实测曲线和伏安特性曲线的对比来实现故障诊断，但该方法不能做到实时在线诊断，且需要建立精确的数学模型，实现难度较大。除了上述传统故障诊断方法，

收稿日期：2023-11-02；**修回日期**：2023-12-01，2023-12-18；
录用日期：2023-12-22；**网络首发日期**：2024-04-26
基金项目：浙江省“尖兵”“领雁”研发攻关计划资助项目(2025C01149)
This work is supported by “Pioneer” and “Leading Goose” R&D Program of Zhejiang under the grant 2025C01149

国内外学者也对机器学习方法进行了研究，如神经网络[7-9]、核极限学习机[10]、支持向量机[11]、决策树[12-13]等。这类方法通过大量信息进行故障的分类识别，克服了由环境因素产生的随机性影响。文献[14]通过使用深度神经网络克服数据集中不同故障的红外热成像数据不平衡问题；文献[15]提出了 1 种基于 *I-U* 转换的特征提取方法，通过对比 10 种机器学习分类模型的使用效果，证明其特征提取方法的合理性；文献[16]将随机森林 RF(random forest)集成学习算法用于光伏阵列早期故障的检测和诊断，优化了 RF 模型的泛化性能；文献[17]提出 1 种基于 AdaBoost 集成模型 AEM(AdaBoost ensemble model)的光伏系统故障检测方法，实现了比 RF 集成学习算法更好的诊断效果；文献[18]通过对比 3 种改进决策树算法和 RF 算法对光伏阵列故障的诊断效果，证明 LightGBM 算法在光伏故障诊断方面的有效性。

然而，上述研究大多是针对光伏阵列在某一时刻仅发生 1 种故障的情况，而光伏系统可能会同时出现多种故障，使得研究人员难以建立准确的机理模型和提取合理故障特征[19]；同时，上述方法并未详细考虑参数选取对其算法性能的影响，而参数选择对于算法诊断效果的影响至关重要[20]。

针对上述问题，本文进行的研究主要包括：①对蜣螂优化 DBO(dung beetle optimization)算法进行了 3 个方面的改进，以解决蜣螂优化算法收敛速度慢、容易陷入局部最优等问题；②建立 1 种基于多工况的光伏阵列模型，获取样本数据，并以 *I-U* 特性曲线提取的故障特征作为模型输入，代表不同工况的标签作为模型输出，构建 IDBO 优化 LightGBM 参数的光伏阵列故障诊断方法；③通过引入改进蜣螂优化算法对 LightGBM 进行参数寻优，进而建立 1 种基于 IDBO-LightGBM 的故障诊断模型。

1　光伏阵列故障诊断模型的构建

1.1　光伏模型构建

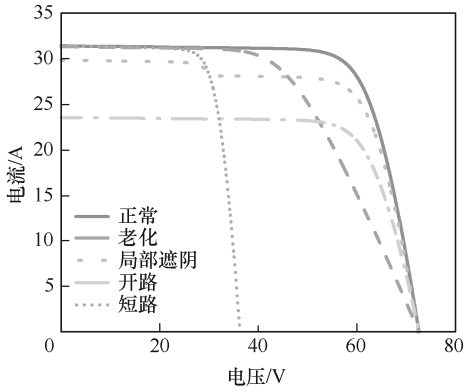
目前，已经提出多种复杂程度和准确性各异的光伏组件建模方法。其中，单二极管模型因高精度和相对简单而被普遍使用[21]。光伏电池单体模型的 *I-U* 方程可表示为

$$I = I_{\mathrm{ph}} - I_{\mathrm{s}}\left[\exp\left(\frac{q(U+R_{\mathrm{s}}I)}{nkT}\right)-1\right]-\frac{U+R_{\mathrm{s}}I}{R_{\mathrm{sh}}} \quad (1)$$

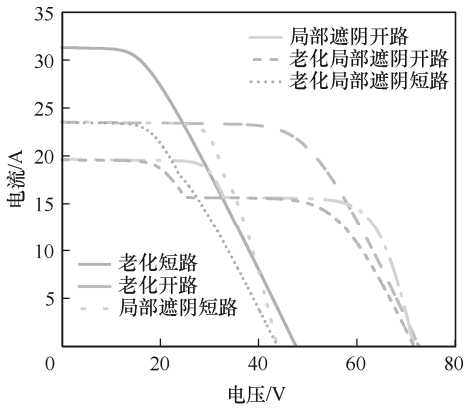
式中：I_{ph} 为光生电流；I_{s} 为二极管的饱和电流；q 为电子电荷；U 和 I 分别为光伏电池的输出电压和电流；R_{s} 和 R_{sh} 分别为模型串联等效电阻和并联等效电阻；n 为二极管的理想因子；k 为玻尔兹曼常数；T 为电池温度。

1.2　光伏阵列故障分析

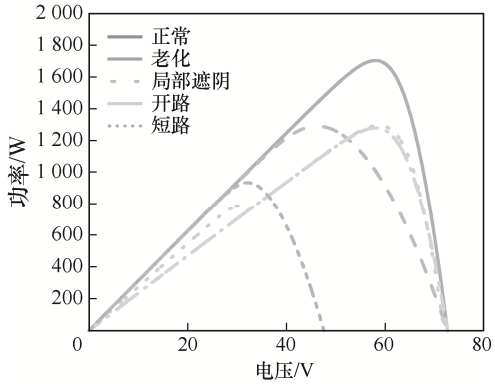
本文采用 MATLAB/Simulink 搭建的光伏仿真模型对任意支路的老化、局部遮阴、开路和短路工况及基于以上工况的复杂工况的输出特性进行模拟，光伏阵列在标准测试条件 STC(standard test condition)(辐照度为 1 000 W/m^2，温度为 25 ℃)下，不同工况对应的 *I-U* 曲线和 *P-U* 曲线如图 1 所示。



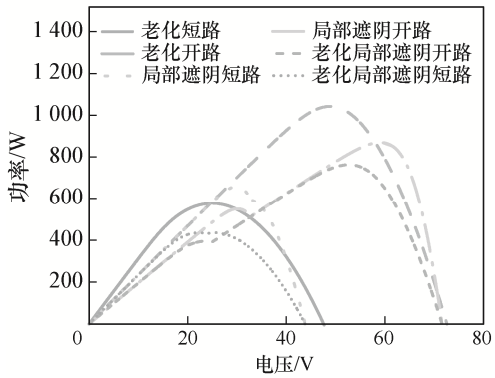
(a) 单故障*I-U*曲线



(b) 多故障*I-U*曲线



(c) 单故障*P-U*曲线



(d) 多故障*P-U*曲线

图 1 不同工况下光伏阵列的 *I-U* 和 *P-U* 曲线

Fig. 1 *I-U* and *P-U* curves for photovoltaic arrays under different working conditions

由图 1 可见，不论出现何种故障情况，光伏阵列的最大功率点电压、最大功率点电流、开路电压、短路电流均会发生明显改变，可以根据所构建的光伏阵列故障仿真模型，对正常、短路、开路、局部遮阴、老化 5 类工况进行仿真。但在实际环境中，光伏阵列往往会发生多重故障同时出现的情况，因此根据实际可能出现的故障类型选取 15 种运行工况作为仿真研究对象，故障诊断模型的输出定义见表 1。

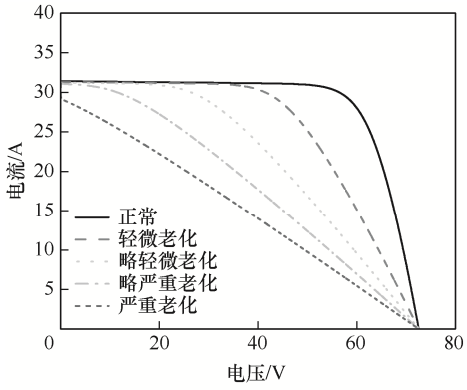
1.3 光伏阵列特征提取

对于光伏阵列中最为常见的 4 类故障情况，本节通过对比不同严重程度故障下光伏阵列 *I-U* 曲线的变化，提取故障诊断的特征量。图 2 为 STC 下，不同严重程度故障下光伏阵列的 *I-U* 曲线。图 2(a)中，最大功率点 MPP(maximum power point)的斜率随着老化严重程度的增加而下降；图 2(b)中，开路光伏组件数量的增加导致光伏阵列普通条件下的短路电流 $I_{sc,ref}$ 与 STC 下的短路电流 $I_{sc,STC}$ 减小；图 2(c)中，局部遮阴可以使 *I-U* 曲线出现拐点和阶跃特征，以及 MPP 的垂直或水平变化；图 2(d)

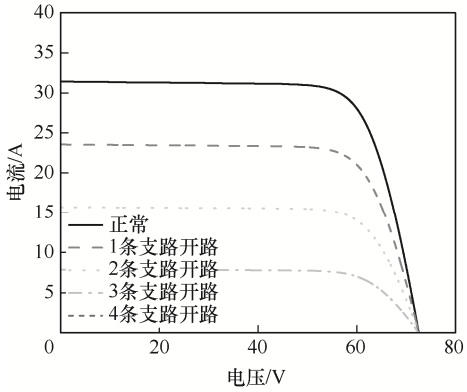
表 1 诊断模型的输出定义

Tab. 1 Output definition of diagnostic model

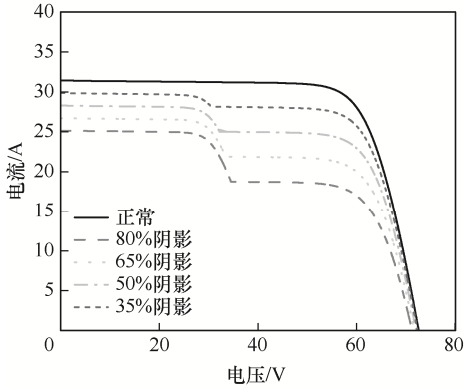
故障类别	工况描述	标签
正常	无故障	0
短路 1	1 个组件短路	1
短路 2	多个组件短路	2
开路 1	1 条支路开路	3
开路 2	多条支路开路	4
老化 1	发生轻微老化	5
老化 2	发生严重老化	6
局部遮阴 1	存在局部阴影	7
局部遮阴 2	全部被阴影覆盖	8
多重故障 1	老化和短路	9
多重故障 2	老化和开路	10
多重故障 3	局部遮阴和短路	11
多重故障 4	局部遮阴和开路	12
多重故障 5	老化、局部遮阴和开路	13
多重故障 6	老化、局部遮阴和短路	14



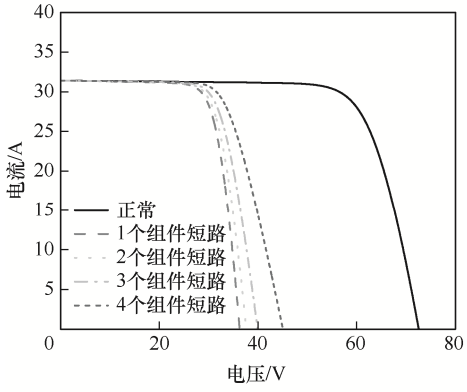
(a) 老化故障



(b) 开路故障



(c) 局部遮阴



(d) 短路故障

图 2　STC 下不同严重程度故障的 *I-U* 曲线

Fig. 2 *I-U* curves for faults of different severities under STC

中，当短路光伏组件数量增加时，STC 下的开路电压 $U_{\mathrm{oc,STC}}$ 将随着普通条件下的开路电压 $U_{\mathrm{oc,ref}}$ 相应减小。

为了有效反映光伏阵列故障信息，可以使用从光伏阵列 *I-U* 曲线中提取的特征点来识别这些故障[19-23]，其故障特征提取的具体公式为

$$\begin{cases} F_1 = \dfrac{U_{\mathrm{oc,ref}}}{yU_{\mathrm{oc,STC}}} \\ F_2 = \dfrac{I_{\mathrm{sc,ref}}}{xI_{\mathrm{sc,STC}}} \end{cases} \tag{2}$$

$$\begin{cases} F_3 = \dfrac{U_{\mathrm{mpp,ref}}}{yU_{\mathrm{oc,STC}}} \\ F_4 = \dfrac{I_{\mathrm{mpp,ref}}}{xI_{\mathrm{sc,STC}}} \end{cases} \tag{3}$$

$$F_5 = \frac{U_{\mathrm{mpp,ref}} I_{\mathrm{mpp,ref}}}{U_{\mathrm{oc,ref}} I_{\mathrm{sc,ref}}} \tag{4}$$

$$\begin{cases} F_6 = \dfrac{I_{\mathrm{mpp,ref}}}{I_{\mathrm{sc,ref}}} \\ F_7 = \dfrac{U_{\mathrm{mpp,ref}}}{U_{\mathrm{oc,ref}}} \\ F_8 = \dfrac{U_{\mathrm{oc,ref}}}{I_{\mathrm{mpp,ref}}} \\ F_9 = \dfrac{U_{\mathrm{mpp,ref}}}{I_{\mathrm{sc,ref}}} \end{cases} \tag{5}$$

$$F_{10} = \frac{I_{\mathrm{mpp,ref}}}{U_{\mathrm{mpp,ref}} - U_{\mathrm{oc,ref}}} \tag{6}$$

式中：y 为每个 PV 串中的 PV 模块的数量；x 为 PV 串的数量；$I_{\mathrm{mpp,ref}}$ 和 $U_{\mathrm{mpp,ref}}$ 分别为普通条件下最大功率点的电流和电压；F_1、F_2 分别为 $U_{\mathrm{oc,ref}}$、$I_{\mathrm{sc,ref}}$ 和 $U_{\mathrm{oc,STC}}$、$I_{\mathrm{sc,STC}}$ 的比例关系；F_3、F_4 分别为 $U_{\mathrm{mpp,ref}}$、$I_{\mathrm{mpp,ref}}$ 和 $U_{\mathrm{oc,STC}}$、$I_{\mathrm{sc,STC}}$ 之间的比例关系；F_5 为填充因子，用于表示组件的故障严重程度；$F_6 \sim F_9$ 分别为 $I_{\mathrm{mpp,ref}}$、$U_{\mathrm{mpp,ref}}$ 与 $I_{\mathrm{sc,ref}}$、$U_{\mathrm{oc,ref}}$ 的比例关系；F_{10} 为伏安特性曲线上的开路点与最大功率的比值。$F_1 \sim F_{10}$ 均可通过 $I_{\mathrm{sc,ref}}$、$U_{\mathrm{oc,ref}}$、$I_{\mathrm{mpp,ref}}$、$U_{\mathrm{mpp,ref}}$ 计算获得。

2　基于 IDBO 优化 LightGBM 的故障诊断方法

2.1　LightGBM 故障诊断模型原理

LightGBM[24]算法是近年来新发布的开源高效分布式梯度提升树算法。相较于其他改进决策树算法，其特点是运行速度快、准确率高[25]。传统梯度提升决策树 GBDT(gradient boosting decision tree)算法很难在精度和效率之间取得较好的平衡，LightGBM 算法可以在加快模型训练速度的同时不对模型精度造成影响[26]。

LightGBM 算法在 GBDT 算法的基础上提出 2 种优化方法：基于梯度的单边采样 GOSS(gradient-based oneside sampling)和互斥特征捆绑 EFB(exclusive feature bundling)。使用 GOSS 算法可以在不损失学习精度的情况下实现正确的学习，加快学习速度。EFB 可以减少高维数据的特征个数，使损失最小化。

2.2 蜣螂优化算法的改进

蜣螂优化算法[27]是 1 种基于蜣螂滚球、跳舞、繁殖、偷窃、觅食等行为进行建模的新型种群智能优化算法，它能凭借 4 个子种群在给定的种群比例下分别运行 4 种不同的搜索手段，使不同的更新法则在全局搜索和局部搜索之间维持平衡，其种群比例可以按照实际问题予以确定。

传统的蜣螂优化算法的初始解是随机产生的，使得种群多样性不足、寻优效果有限、收敛速度不快。为改善这一问题，使用混沌数来替代原算法中的随机变量，以实现更全面的搜索。本文加入的 Chebyshev 混沌映射方法具有随机性、非周期性等特性，其数学表达式为

$$\chi_{k+1}=\cos[k\cdot\arccos(\chi_k)] \tag{7}$$

式中：χ_k 为第 k 代混沌序列的当前值；k 为当前迭代次数。

在迭代过程中发现滚球蜣螂之间缺乏足够的交流，而黄金正弦算法对于更新种群位置有一定帮助[28]。正弦函数上的点均可以通过黄金正弦算法进行遍历，引入黄金分割系数到位置更新中，旨在局部区域进行充分的探索，既可以让滚球蜣螂的局部开发能力和全局搜索能力达到更好的平衡，也可以让蜣螂的收敛速度和求解精度提高。引入黄金正弦算法后的滚球蜣螂更新位置公式为

$$\begin{aligned}x_i(t+1)=x_i(t)&|\sin(r_1)|-r_2\cdot\sin(r_1)\cdot\\&|c_1X^*-c_2x_i(t)|\end{aligned} \tag{8}$$

式中：$x_i(t)$为第 i 只蜣螂在第 t 次迭代时的当前位置；r_1 和 r_2 为随机数，$r_1\in(0,2\pi]$，$r_2\in(0,\pi]$；X^*为当前迭代周期内的全局最优位置；c_1 和 c_2 为引入黄金分割数后得到的系数，可表示为

$$\begin{cases}c_1=a\tau+b(1-\tau)\\c_2=a(1-\tau)+b\tau\end{cases} \tag{9}$$

式中：a、b 初始值设置为$-\pi$ 和 π，随后 a 和 b 随着目标的变化而改变；τ 为取值是 $(1-\sqrt{5})/2$ 的黄金分割数。

Lévy 飞行是用于模拟动物的随机游走过程的算法，具有扩大种群探索区域和丰富种群多样性的特点。在偷窃蜣螂的位置更新公式中，蜣螂存在陷入局部最优的情况，而使用 Lévy 飞行策略进行位置更新使得偷窃蜣螂搜索范围更广泛，此时偷窃蜣螂的位置更新公式为

$$\begin{aligned}x_i(t-1)=X^{\mathrm{b}}+S\cdot\boldsymbol{g}\cdot(&|x_i(t)-X^*|+\\&|x_i(t)-X^{\mathrm{b}}|)\oplus L(\lambda)\end{aligned} \tag{10}$$

式中：X^{b} 为全局最优位置；S 为步长缩放因子；$\boldsymbol{g}$ 为正态分布随机向量；$\oplus$为点乘；$L(\lambda)$ 为服从 Lévy 飞行分布的步长。

2.3 IDBO-LightGBM 混合模型

由于在默认参数下 LightGBM 模型的故障诊断准确率不高，因此要优化 LightGBM 模型中的指定超参数，以提升其诊断准确率。参数值在给定范围内随机生成，迭代 100 次，参数寻优后的结果见表 2。

表 2 参数寻优结果

Tab. 2 Parameter optimization results

参数	最优值	参数含义
Learning_rate	0.206	学习率，该步长参数用于控制每次迭代中模型参数的更新幅度
n_estimators	200	决策树的数量或迭代次数
max_depth	4	决策树的最大深度限制
min_child_samples	28	1 个叶子节点所需的最小样本数

使用 IDBO 算法对 LightGBM 进行参数寻优，进而建立 IDBO-LightGBM 故障诊断模型，故障诊断具体步骤如下。

步骤 1 输入正常工况和不同故障工况下的光伏阵列样本数据，对数据进行标准化转换，划分训练集与测试集。

步骤 2 初始化 LightGBM 模型的参数和 IDBO 算法的参数，并设置蜣螂种群数量、自变量上下限、最大迭代次数、维度大小等。

步骤 3 根据式(7)对蜣螂种群初始化。

步骤 4 根据各蜣螂种群的位置更新公式计算蜣螂新位置的适应度值。判断当前蜣螂种群的适应度值对比已存储的最佳适应度值是否更佳，若

是，则替换存储的最佳适应度值为当前适应度值，并保存其对应的数值；否则继续存储之前的最佳参数。

步骤 5 判断迭代次数是否达到最大值，若是，则结束迭代，比较所有迭代存储的最优适应度值并输出最优参数；若不是，则返回步骤 3。

步骤 6 以 IDBO 寻得的最优参数值构建最优的 LightGBM 模型进行光伏阵列故障的诊断测试。

3　仿真实验分析

3.1　仿真模型构建

本文通过对表 1 中 15 种标签所对应的不同工况进行仿真，每条标签均有 1 000~1 500 条数据，共计 16 000 条。数据集由 MATLAB/Simulink 构建出的 2×4 的故障仿真模型获得，故障仿真模型如图 3 所示。

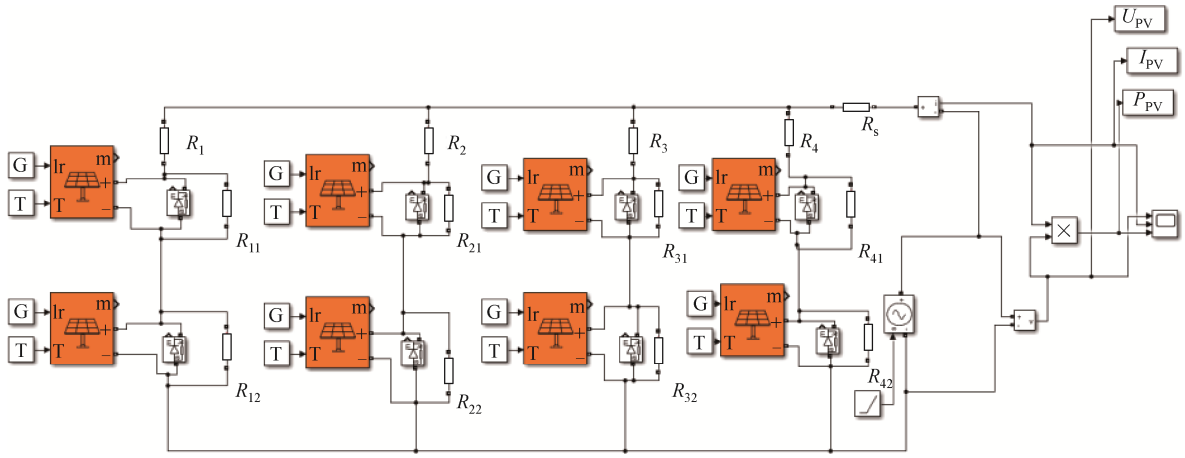


图 3　光伏阵列故障仿真模型

Fig. 3 Simulation model of photovoltaic arrays under fault

设置各光伏阵列模块的环境温度在 10~40 ℃的区间内，辐照度在 200~1 200 W/m^2 的区间内。通过改变电阻 R_s 对光伏阵列的老化情况进行仿真，R_s<1 Ω 为轻微老化，否则属于严重老化；通过调整并联电阻 R_{ij}(i=1, 2；j=1, 2, 3, 4)控制其对应的光伏组件短路情况，以模仿光伏阵列中组件发生短路故障的情况；通过调整串联电阻 R_j(j=1, 2, 3, 4)对光伏阵列开路故障进行仿真；改变每个光伏组件输入的环境温度和辐照度对光伏阵列局部遮阴情况进行仿真。复合故障工况仿真则是对上述 2 种及以上故障同时发生的情况进行仿真。

3.2　样本数据标准化转换

当在非 STC 下发生不同的故障时，I-U 曲线可能会显示出相似的特征。因此，若能将在不同环境条件下的参数准确转换为 STC 状态下的参数，则可有效提高后续识别算法的分类精度。用于转换原始 I-U 曲线的公式为

$$\begin{cases} I_{\mathrm{STC}} = I_{\mathrm{out}} \dfrac{G_{\mathrm{STC}}}{G[1+\beta_i(T-T_{\mathrm{STC}})]} \\ U_{\mathrm{STC}} = \dfrac{U_{\mathrm{out}}}{1+K_{\mathrm{g}}\ln\left(\dfrac{G}{G_{\mathrm{STC}}}\right)+\beta_u(T-T_{\mathrm{STC}})} \end{cases} \tag{11}$$

式中：I_{STC} 和 U_{STC} 分别为 STC 下的输入电流和电压；I_{out} 和 U_{out} 分别为普通条件下的输出电流和电压；G 和 T 分别为普通条件下的辐照度和温度；G_{STC} 和 T_{STC} 分别为 STC 下的辐照度和温度；β_i 为短路电流温度系数；β_u 为开路电压温度系数；K_{g} 为电压的辐照度修正系数。

不同太阳能发电场的光伏阵列具有不同的电压和电流水平。为使所提故障诊断方法具有鲁棒性，有必要对电流和电压归一化，其公式为

$$\begin{cases} I_{\mathrm{n}} = \dfrac{I_{\mathrm{STC}}}{xI_{\mathrm{sc,STC}}} \\ U_{\mathrm{n}} = \dfrac{U_{\mathrm{STC}}}{yU_{\mathrm{oc,STC}}} \end{cases} \tag{12}$$

式中，I_n 和 U_n 分别为归一化后的原始短路电流和开路电压。

3.3 实验过程及诊断结果

3.3.1 算法对比

为了说明所提方法的合理性，使用 IDBO 算法对 2.1 节指定的 LightGBM 超参数进行寻优，将结果与 DBO 算法和 SSA 算法的寻优结果进行对比。其中，IDBO 的迭代次数设为 100，蜣螂种群规模设为 60，各个蜣螂种群的比例为 6:6:7:11。对比结果如图 4 所示，求取适应度值大小的公式为

$$\text{fitness} = 1 - \text{acc} \tag{13}$$

式中：fitness 为适应度值，fitness 越小表示模型诊断越准确；acc 为代入寻优后超参数模型的准确率。

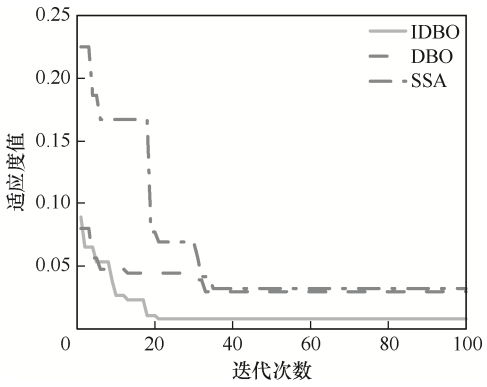


图 4 3 种模型的寻优曲线

Fig. 4 Optimization curves for three models

由图 4 可见，IDBO 算法具有较强的快速收敛能力，在第 21 次迭代时收敛，且其最低适应度值为 0.007 1，最高准确率为 99.29%，很好地满足了准确性要求；DBO 算法和 SSA 算法分别收敛于 0.029 6 和 0.033 2，收敛精度不佳。综上可知，IDBO 算法改进后收敛速度更快，且识别精度更高，证明了本文选用该算法的合理性。

3.3.2 不同工况诊断结果分析

为了说明所提方法的有效性，分别使用 IDBO-LightGBM 故障诊断模型和改进前的 DBO-LightGBM 对 3.1 节仿真得到的数据集标准化转换后进行训练与测试。

LightGBM 模型在默认参数下虽然拥有分类能力，但准确率不高，对超参数的选取直接影响 LightGBM 的故障分类精度。使用 LightGBM 分类模型的默认参数，所得未经优化的 LightGBM 模型诊断结果如图 5 所示，其运行工况标签的含义见表 1。

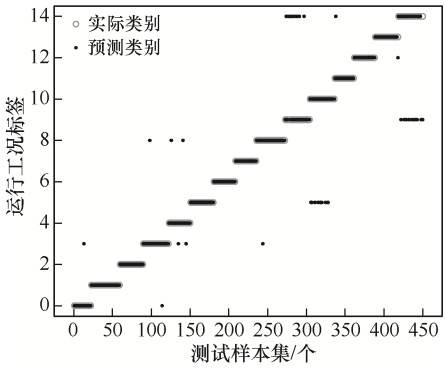


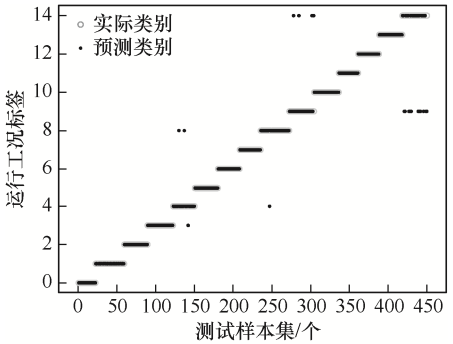
图 5 LightGBM 模型的故障诊断结果

Fig. 5 Fault diagnosis results of LightGBM model

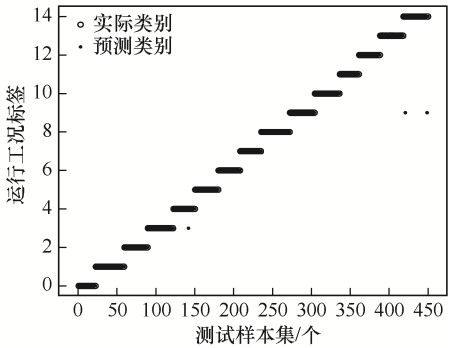
由图 5 可见，LightGBM 模型在默认参数输入的情况下，对测试样本中的标签 0、3、4、8、9、10、11、13、14 均造成了错误的诊断，且诊断结果表明该模型所得到的诊断正确率仅为 90.44%。可以看出，LightGBM 模型在默认参数输入的情况下诊断能力不够理想，有待加强。

将 IDBO 算法和 DBO 算法对 2.3 节指定的超参数寻优后的值输入 LightGBM 模型后，其故障诊断模型的可视化结果如图 6 所示，其运行工况标签的含义见表 1。

由图 6 可见，将寻优后的超参数值输入 LightGBM 模型后，LightGBM 模型对于光伏阵列的故障诊断效果明显提升。但 DBO-LightGBM 模型的诊断对于光伏阵列运行中特定的几种工况的准确率仍然不超过 95%，说明该模型的诊断效果不佳。优化后的 IDBO-LightGBM 模型有 3 组测试样本发生了错误诊断，分别是将 2 条标签 14(老化、局部遮阴和短路)数据诊断为标签 9(老化和短路)，以及 1 条标签 4(多条支路开路)数据诊断为标签 3(1 条支路开路)，模型诊断的准确率达到了 99.29%，说明了该模型在光伏阵列故障诊断方面的有效性。



(a) DBO-LightGBM模型



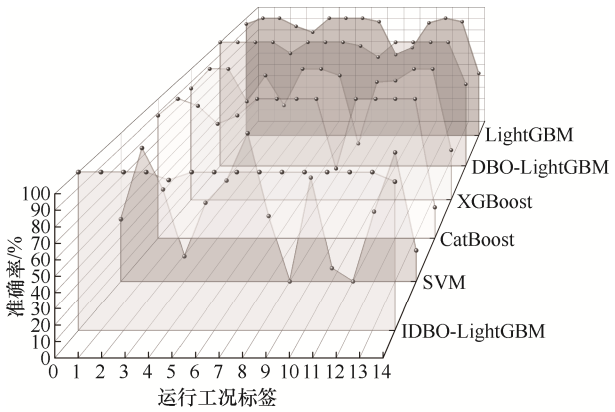
(b) IDBO-LightGBM模型

图 6　IDBO-LightGBM 模型和 DBO-LightGBM 模型的故障诊断结果

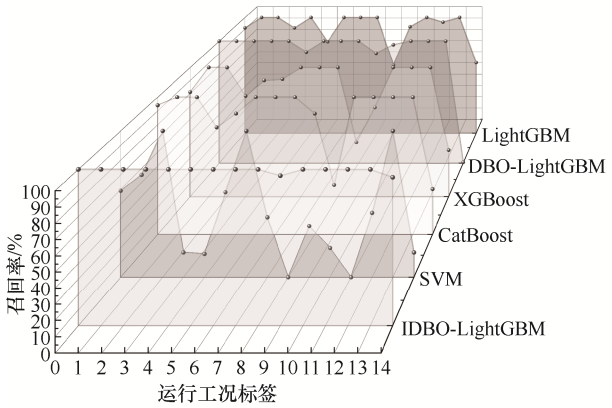
Fig. 6 Fault diagnosis results of IDBO-LightGBM and DBO-LightGBM models

3.3.3　算法对比结果分析

为了说明所提方法的优越性，在使用同样的训练样本和测试样本的前提下，将本文算法对于不同运行工况的诊断准确率与默认参数输入的 LightGBM、使用 DBO 进行参数寻优的 LightGBM、其他改进决策树算法(XGBoost、CatBoost)及支持向量机 SVM(support vector machine)进行对比，如图 7 所示，其运行工况的标签含义见表 1。



(a) 不同算法的诊断准确率对比



(b) 不同算法的诊断召回率对比

图 7　不同算法的诊断结果对比

Fig. 7 Comparison of diagnostic results among different algorithms

由图 7(a)和(b)可见，以上算法对于标签 7 的召回率均为 100%，对于标签 2、6、13 的诊断准确率均为 100%，这说明在特定情况下，以上算法均有出色表现，没有出现错误的诊断。其中，SVM 算法在其他运行工况的诊断表现相比其他决策树算法较差。

在对标签 9 和标签 14 的诊断结果上，算法普遍效果不佳，但 IDBO-LightGBM 对标签 9 的诊断准确率达到 96.12%，相比 DBO-LightGBM 的 91.50%和 LightGBM 的 59.37%分别提高了 4.62%和 36.75%。对标签 14 的诊断准确率仅有 IDBO-LightGBM 达到 95%以上，对比其他算法有较大提升。

除上述标签外，各改进决策树算法对其他标签均有良好的诊断效果，而参数寻优后的 IDBO-LightGBM 和 DBO-LightGBM 对比默认参数的 LightGBM 的诊断效果均有所提高。

因此，根据图 7 可以认为 IDBO-LightGBM 在多数工况下均具有明显的优越性，因为其在这些情况下具有较高的诊断能力，而其他算法则在某些情况下可能会出现较低的准确率。

4　结论

本文提出了 1 种基于 IDBO-LightGBM 的光伏阵列故障诊断方法，能够诊断光伏阵列不同类型的

故障状况。研究分析了多种类型的光伏阵列故障，构建了融合多种故障特征量的输入向量。建立了基于 IDBO-LightGBM 算法的故障诊断模型，搭建了光伏阵列仿真模型，从而验证了算法的有效性。通过仿真分析和验证，得出以下结论。

(1)本文对 DBO 算法进行了 3 个方面的改进：其一，利用 Chebyshev 混沌映射初始化蜣螂种群的位置；其二，引入黄金正弦策略，使滚球蜣螂能够对局部区域进行充分搜索；其三，引入 Lévy 飞行，使偷窃蜣螂个体去到更广的搜索空间。

(2)本文提出 1 种基于多工况的光伏阵列模型，并以 *I-U* 特性曲线提取的故障特征作为模型输入，代表不同工况的标签作为模型输出，构建 IDBO 优化 LightGBM 参数的光伏故障诊断方法。

(3)通过引入改进蜣螂优化算法对 LightGBM 进行参数寻优，进而建立 1 种基于 IDBO-LightGBM 的故障诊断模型。通过与 SVM 算法、默认参数输入的 LightGBM 算法、使用 DBO 进行参数寻优的 LightGBM 算法及其他改进决策树算法的对比，证明了本文方法的合理性、有效性和优越性。实验结果表明，基于 IDBO-LightGBM 的故障诊断模型可以更准确地识别出正常、老化、局部遮阴、短路和开路等故障，以及以上工况同时发生的复合故障。

参考文献：

[1] 陈国平, 董昱, 梁志峰. 能源转型中的中国特色新能源高质量发展分析与思考[J]. 中国电机工程学报, 2020, 40(17): 5493-5505.
Chen Guoping, Dong Yu, Liang Zhifeng. Analysis and reflection on high-quality development of new energy with Chinese characteristics in energy transition [J]. Proceedings of the CSEE, 2020, 40(17): 5493-5505. (in Chinese)

[2] Osmani K, Haddad A, Lemenand T, et al. A critical review of PV systems' faults with the relevant detection methods [J]. Energy Nexus, 2023, 12: 100257.

[3] 马铭遥, 王海松, 马文婷, 等. 基于 S-V 特性分析的晶硅光伏组件阴影遮挡故障诊断[J]. 太阳能学报, 2022, 43(9): 64-72.
Ma Mingyao, Wang Haisong, Ma Wenting, et al. Partial shadow fault diagnosis of crystalline silicon photovoltaic module based on S-V characteristic analysis [J]. Acta Energiae Solaris Sinica, 2022, 43(9): 64-72. (in Chinese)

[4] 王永光, 戈庆长. 分布式光伏发电系统多峰值 MPPT 优化研究[J]. 电力科学与技术学报, 2020, 35(2): 3-11.
Wang Yongguang, Ge Qingchang. Research on multi-peak MPPT optimization of photovoltaic power generation system [J]. Journal of Electric Power Science and Technology, 2020, 35(2): 3-11. (in Chinese)

[5] 蒋琳, 苏建徽, 李欣, 等. 基于可见光和红外热图像融合的光伏阵列热斑检测方法[J]. 太阳能学报, 2022, 43(1): 393-397.
Jiang Lin, Su Jianhui, Li Xin, et al. Hot spot detection of photovoltaic array based on fusion of visible and infrared thermal images [J]. Acta Energiae Solaris Sinica, 2022, 43(1): 393-397. (in Chinese)

[6] Pei Tingting, Li Li, Zhang Jiangfeng, et al. Module block fault locating strategy for large-scale photovoltaic arrays [J]. Energy Conversion and Management, 2020, 214: 112898.

[7] Mustafa Z, Awad A S A, Azzouz M, et al. Fault identification for photovoltaic systems using a multi-output deep learning approach [J]. Expert Systems with Applications, 2023, 211: 118551.

[8] Liu Yongjie, Ding Kun, Zhang Jingwei, et al. Fault diagnosis approach for photovoltaic array based on the stacked auto-encoder and clustering with *I-V* curves [J]. Energy Conversion and Management, 2021, 245: 114603.

[9] Lu Xiaoyang, Lin Peijie, Cheng Shuying, et al. Fault diagnosis for photovoltaic array based on convolutional neural network and electrical time series graph [J]. Energy Conversion and Management, 2019, 196: 950-965.

[10] 任晓琳, 杨奕, 高龙, 等. 基于 IBA-KELM 模型的 TCT 光伏阵列故障诊断研究[J]. 电源学报, 2023, 21(5): 67-74.
Ren Xiaolin, Yang Yi, Gao Long, et al. Research on fault diagnosis of TCT photovoltaic array based on IBA-KELM model [J]. Journal of Power Supply, 2023, 21(5): 67-74. (in Chinese)

[11] 付华, 张辉. 基于故障电流特征的多端柔性直流电网故障识别[J]. 电源学报, 2025, 23(5): 174-184.
Fu Hua, Zhang Hui. Fault identification of multi-terminal VSC-HVDC grid based on fault current characteristics [J]. Journal of Power Supply, 2025, 23(5): 174-184. (in Chinese)

[12] 顾崇寅, 徐潇源, 王梦圆, 等. 基于 CatBoost 算法的光伏阵列故障诊断方法[J]. 电力系统自动化, 2023, 47(2): 105-114.
Gu Chongyin, Xu Xiaoyuan, Wang Mengyuan, et al. CatBoost algorithm based fault diagnosis method for photovoltaic arrays [J]. Automation of Electric Power Systems, 2023, 47(2): 105-114. (in Chinese)

[13] 刘行行, 帕孜来·马合木提, 程志江, 等. 基于 XGBoost 的光伏阵列故障诊断方法研究[J]. 电子测量技术, 2023, 46(12): 8-14.
Liu Hanghang, Pazilai·Mahemuti, Cheng Zhijiang, et al. Research on fault diagnosis method of photovoltaic array based on extreme gradient boosting [J]. Electronic Measurement Technology, 2023, 46(12): 8-14. (in Chinese)

[14] Kellil N, Aissat A, Mellit A. Fault diagnosis of photovoltaic modules using deep neural networks and infrared images under Algerian climatic conditions [J]. Energy, 2023, 263: 125902.

[15] Ding Kun, Chen Xiang, Jiang Meng, et al. Feature extraction and fault diagnosis of photovoltaic array based on current-voltage conversion [J]. Applied Energy, 2024, 353: 122135.

[16] Chen Zhicong, Han Fuchang, Wu Lijun, et al. Random forest based intelligent fault diagnosis for PV arrays using array voltage and string currents [J]. Energy Conversion and Management, 2018, 178: 250-264.

[17] Lodhi E, Wang Feiyue, Xiong Gang, et al. An AdaBoost ensemble model for fault detection and classification in photovoltaic arrays [J]. IEEE Journal of Radio Frequency Identification, 2022, 6: 794-800.

[18] Adhya D, Chatterjee S, Chakraborty A K. Performance assessment of selective machine learning techniques for improved PV array fault diagnosis [J]. Sustainable Energy, Grids and Networks, 2022, 29: 100582.

[19] He Zengxiang, Chu Pengpeng, Li Chenxi, et al. Compound fault diagnosis for photovoltaic arrays based on multi-label learning considering multiple faults coupling [J]. Energy Conversion and Management, 2023, 279: 116742.

[20] 郭伟欣, 张志东. 基于 Jaya 算法的遮阴条件下光伏 MPPT 控制研究[J]. 电源学报, 2025, 23(5): 105-111.
Guo Weixin, Zhang Zhidong. Research on photovoltaic MPPT control under shading conditions based on Jaya algorithm [J]. Journal of Power Supply, 2025, 23(5): 105-111. (in Chinese)

[21] 吴春华, 易苑, 李智华, 等. 考虑参数权重与分层映射的光伏组件健康程度检测[J]. 电工技术学报, 2024, 39(15): 4856-4867.
Wu Chunhua, Yi Yuan, Li Zhihua, et al. Health detection of photovoltaic modules considering parameter weights and hierarchical mapping [J]. Transactions of China Electrotechnical Society, 2024, 39(15): 4856-4867. (in Chinese)

[22] 刘笑, 杨建, 李力, 等. 基于机器学习的带被动阻尼直流微电网系统的稳定性检测[J]. 电工技术学报, 2024, 39(8): 2281-2293, 2324.
Liu Xiao, Yang Jian, Li Li, et al. Stability detection of DC microgrid systems with passive damping based on machine learning [J]. Transactions of China Electrotechnical Society, 2024, 39(8): 2281-2293, 2324. (in Chinese)

[23] Tao Peng, Shen Hongtao, Zhang Yangrui, et al. Status forecast and fault classification of smart meters using LightGBM algorithm improved by random forest [J]. Wireless Communications and Mobile Computing, 2022, 2022(1): 3846637.

[24] Shongwe S, Hanif M. Comparative analysis of different single-diode PV modeling methods [J]. IEEE Journal of Photovoltaics, 2015, 5(3): 938-946.

[25] Ke G, Meng Q, Finley T, et al. LightGBM: A highly efficient gradient boosting decision tree [C]// Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, California, USA: Curran Associates Inc., 2017: 3149-3157.

[26] 胡澜也, 蒋文博, 李艳婷. 基于 LightGBM 的风力发电机故障诊断[J]. 太阳能学报, 2021, 42(11): 255-259.

Hu Lanye, Jiang Wenbo, Li Yanting. Fault diagnosis for wind turbine based on lightgbm [J]. Acta Energiae Solaris Sinica, 2021, 42(11): 255-259. (in Chinese)

[27] Xue Jiankai, Shen Bo. Dung beetle optimizer: A new meta-heuristic algorithm for global optimization [J]. The Journal of Supercomputing, 2023, 79(7): 7305-7336.

[28] 刘成汉，何庆. 融合多策略的黄金正弦黑猩猩优化算法[J]. 自动化学报, 2023, 49(11): 2360-2373.
Liu Chenghan, He Qing. Golden sine chimp optimization algorithm integrating multiple strategies [J]. Acta Automatica Sinica, 2023, 49(11): 2360-2373. (in Chinese)



吴亚钧

作者简介：

吴亚钧(2000—)，男，硕士研究生。研究方向：光伏发电故障诊断技术。E-mail：222207855020@zust.edu.cn。

王璐(1999—)，男，硕士研究生。研究方向：微电网能源调度。E-mail：222107855029@zust.edu.cn。

张金江(1971—)，男，中国电源学会会员，通信作者，博士，副教授。研究方向：智能电网风险评估与运行控制。E-mail：jjzhang@zust.edu.cn。

(责任编辑：耿越)



公益广告来源：阜平县融媒体中心（电视台）