

基于 ICOA 优化 XGBoost 的光伏阵列故障诊断方法

董建业¹ 李红月¹

DONG Jianye LI Hongyue

摘要

基于 XGBoost 模型是识别光伏阵列故障类型的一项重要技术, 由于 XGBoost 参数初始化设定的主观性和随机性, 导致模型在训练和学习时准确度低。针对传统 XGBoost 算法的不足, 文章提出了一种优化长鼻浣熊算法 (ICOA) 优化 XGBoost 初始参数的故障诊断方法。采用 Logistic-Tent 混沌映射、自适应权重因子、Levy 飞行和透镜成像学习策略来优化长鼻浣熊算法 (COA), 降低了算法易陷入局部极值点的可能性。利用 ICOA 算法对 XGBoost 分类算法进行优化, 构建 ICOA-XGBoost 光伏阵列故障诊断模型, 并与其他优化算法模型进行实例对比分析, 验证了改进后的算法在识别光伏阵列故障类别上的有效性和实用性。

关键词

光伏阵列; 故障诊断; 改进长鼻浣熊算法; XGBoost 算法

doi: 10.3969/j.issn.1672-9528.2024.11.012

0 引言

随着能源多元化发展的持续推进, 分布式光伏发电得到广泛发展。但光伏阵列易发生老化、局部遮荫、开路和短路等故障, 因此提高故障诊断准确率, 不仅能提高发电系统的可靠性, 而且能有效降低运维成本^[1]。

目前, 很多机器学习分类算法应用于光伏阵列故障诊断中。吴亚钧等人通过加入 Levy 飞行策略和黄金分割系数改

进蛱蝶算法, 该方法寻优速度快且稳定^[2]。王一鸣等人提出麻雀算法优化支持向量机算法, 有效解决支持向量机算法收敛速度慢的问题^[3]。罗凯元等人引入融合佳点集、分段自适应逃逸能量机制和黄金正弦策略进一步优化哈里斯鹰算法, 显著提高故障诊断准确率^[4]。

综合上述分析, 本文提出一种改进长鼻浣熊算法 (coati optimization algorithm, ICOA) 优化 XGBoost 算法的初始参数, 有效提高模型诊断准确率。首先, 提出优化策略并进行改进; 然后, 将 ICOA 算法与 COA、WOA 和 DBO 算法进

1. 安徽理工大学电气与信息工程学院 安徽淮南 232001

[3] DUAN J Y, GUO L L. Variable-length subsequence clustering in time series[J]. IEEE transactions on knowledge and data engineering. 2022, 34(2): 983-995.

[4] CARON M, BOJANOWSKI P, JOULIN A, et al. Deep clustering for unsupervised learning of visual features[C]// Computer Vision-ECCV 2018. Berlin: Springer, 2018: 139-156.

[5] CHEN F W, PAN S R, JIANG J, et al. DAGCN: dual attention graph convolutional networks[C]//2019 International Joint Conference on Neural Networks. Piscataway, NJ: IEEE, 2019: 1-8.

[6] BONCHI F, GIONIS A, UKKONEN A. Overlapping correlation clustering[J]. Knowledge and information systems, 2013, 35: 1-32.

[7] WANG X S, LI L J, CHENG Y H. An overlapping module

identification method in protein-protein interaction networks[J]. BMC bioinformatics, 2012, 13(S-7):S4.

[8] 陈国艳, 张颖, 梁德群. 基于 BIC 准则的图像分割算法[J]. 辽宁工程技术大学学报 (自然科学版), 2016, 35(11): 1359-1362.

[9] 赖健琼. 自适应 AP 聚类算法研究[J]. 计算机时代, 2022 (4): 38-42.

【作者简介】

张娟 (1981—), 通讯作者 (zhangjuan@mail.taiji.com.cn), 女, 天津人, 本科, 工程师, 研究方向: 数据挖掘、数据中台。

李娜 (1981—), 女, 江西宜丰人, 本科, 助理研究员, 研究方向: 数字经济。

(收稿日期: 2024-08-09)

行对比,证明 ICOA 算法有着更好的收敛性和稳定性;最后将优化算法分类诊断模型与其他诊断模型进行对比,进一步验证所提优化分类诊断模型的优越性。

1 改进长鼻浣熊优化算法 (ICOA)

1.1 长鼻浣熊优化算法

长鼻浣熊优化算法 (COA) 是基于长鼻浣熊狩猎行为的优化算法。主要核心是模仿长鼻浣熊的两种自然行为:浣熊对鬣蜥的狩猎行为和浣熊逃离捕食者的行为,该算法具有进化能力强、探索速度快和寻优能力较强等优点^[5]。具体过程如下。

(1) 种群初始化

在寻优空间中随机初始化种群:

$$x_{i,j} = b_j^L + R(b_j^U - b_j^L) \quad (1)$$

式中: $x_{i,j}$ 为个体; b_j^L 为寻优空间的下边界; b_j^U 为寻优空间的上边界; R 为 $[0,1]$ 之间的随机数。

(2) 探索阶段

在此阶段中一半浣熊攻击鬣蜥,其余的浣熊等待鬣蜥改变位置。前一半位置更新表达式为:

$$\begin{cases} x_{i,j}^{p1} = x_{i,j} + R \cdot (G_j - Ix_{i,j}) \\ i = 1, 2, \dots, \left\lfloor \frac{N}{2} \right\rfloor \end{cases} \quad (2)$$

式中: $x_{i,j}^{p1}$ 为第一阶段个体位置; G_j 为鬣蜥的位置; I 为集合 $\{1,2\}$ 中随机选取的一个数; N 为浣熊的数量。

剩余的浣熊位置更新公式为:

$$x_{i,j}^{p1} = \begin{cases} x_{i,j} + R(G_j - Ix_{i,j}), F_{G,j}^g \leq F_{i,j} \\ x_{i,j} + R(x_{i,j} - G_j), F_{G,j}^g > F_{i,j} \\ i = \left\lfloor \frac{N}{2} \right\rfloor + 1, \dots, N \end{cases} \quad (3)$$

式中: G_j^g 为鬣蜥的新位置; $F_{G,j}^g$ 为更新后的目标函数值; $F_{i,j}$ 为更新前的目标函数值。

位置更新后,进行目标函数值对比。如果更新后的个体更优,则更新个体到当前位置;否则保持原状。

(3) 开发阶段

在开发阶段中浣熊会逃避捕食者到其他位置。此时每只浣熊附件随机生成一个位置。

$$\begin{cases} x_{i,j}^{p2} = x_{i,j} + (1-2R)(b_{j,L}^w + R(b_{j,U}^w - b_{j,L}^w)) \\ b_{j,L}^w = \frac{b_j^L}{t}, b_{j,U}^w = \frac{b_j^U}{t}, t = 1, 2, \dots, T \end{cases} \quad (4)$$

式中: $x_{i,j}^{p2}$ 为个体新位置; t 为当前迭代次数; T 为最大迭代次数。

位置更新后,进行目标函数值对比。如果更新后的个体更优,则更新个体到当前位置;否则保持原状。

经过所有阶段位置更新后,则完成一次算法迭代。随后

COA 进行下一次的迭代。将最优位置作为最优解输出。

1.2 长鼻浣熊优化算法的改进

(1) Logistic-Tent 混沌映射初始化种群

在 COA 中种群初始化是随机的,会影响后期全局搜索能力^[6]。引入 Logistic-Tent 混沌映射生成优良的初始化种群,能有效防止在种群初始化不均问题,加快收敛速度,提高寻优的精准度。Logistic-Tent 混沌映射的数学表达式为:

$$x_{n+1} = \begin{cases} \left[rx_n(1-x_n) + \frac{(4-r)}{2}x_n \right] \bmod 1, x_n < 0.5 \\ \left[rx_n(1-x_n) + \frac{(4-r)(1-x_n)}{2} \right] \bmod 1, x_n \geq 0.5 \end{cases} \quad (5)$$

式中: r 为 $[0,4]$ 之间的随机数。

(2) Levy 飞行策略

Levy 飞行指步长的概率分布为重尾分布的随机行走策略^[7],分布函数数学表达式为:

$$\text{Levy}(\beta) = \frac{\mu}{|\nu|^{\frac{1}{\beta}}} \quad (6)$$

式中: $\text{Levy}(\beta)$ 为飞行步长; β 为 $[0,2]$ 之间的随机数; μ 服从 $N(0, \sigma_\mu^2)$ 分布; ν 服从 $N(0, \sigma_\nu^2)$ 分布。

Levy 飞行策略引入后,COA 在开发阶段位置更新公式由式 (4) 变为:

$$x_{i,j}^{p2} = x_{i,j} + (1-2R)(b_{j,L}^w + R(b_{j,U}^w - b_{j,L}^w)) \cdot \text{Levy}(\beta) \quad (7)$$

(3) 自适应权重因子

权重因子 ω 对搜索能力有很大影响,当 ω 较大时,全局搜索能力较强。当 ω 较小时,局部搜索能力较强。为权衡全局搜索和局部搜索能力。在迭代初期 ω 值较大,以此来增强初期全局搜索能力^[8]。权重因子 ω 取值表达式为:

$$\omega = \omega_D - (\omega_D - \omega_x) \arctan\left(\frac{\pi t}{2T}\right) \quad (8)$$

式中: $\omega_{\max}$ 为最大惯性权重; $\omega_{\min}$ 为最小惯性权重。本文设置 $\omega_{\max}$ 和 $\omega_{\min}$ 值分别为 0.9、0.4。

自适应权重因子引入后,COA 在探索阶段位置更新公式由式 (3) 变为:

$$x_{i,j}^{p1} = \begin{cases} x_{i,j} + \omega(G_j - Ix_{i,j}), F_{G,j}^g \leq F_{i,j} \\ x_{i,j} + \omega(x_{i,j} - G_j), F_{G,j}^g > F_{i,j} \end{cases} \quad (9)$$

(4) 透镜成像反向学习

针对 COA 算法陷入局部最优值问题,引入透镜成像反向学习策略。通过计算当前位置的反向解,以此扩大寻优范围,增强跳出局部最优值的能力^[9]。具体数学表达式为:

$$\begin{cases} x_{i,j}^N = \frac{b_j^U + b_j^L}{2} + \frac{b_j^U + b_j^L}{2k} - \frac{x_{i,j}}{k} \\ k = \left[1 + \left(\frac{t}{T} \right)^{0.5} \right]^{10} \end{cases} \quad (10)$$

1.3 XGBoost 模型

XGBoost 是一种分布式梯度提升树的模型,该算法是将多种分类器整合为一个大分类器,多种分类器包含线性分类器、分类回归树等。通过对误差的不断分类来解决过拟合、泛化能力差等问题,其主体是不断增加新的数去拟合上次残差计算出的目标函数。相当于生成新的函数拟合模型的残差,以此不断迭代来提高模型的准确率^[10]。具体模型公式为:

$$\hat{y} = \varphi(x_i) = \sum_{k=1}^K f_k(x_i) \quad f_k \in F \quad (11)$$

$$F = \{f(x) = \omega_q(x)\} \quad (q: R^m \rightarrow T, \omega \in R^T) \quad (12)$$

式中: $\hat{y}_i$ 为第 k 轮迭代后树的预测结果, q 为树结构, T 为叶子结点数目, ω 为叶子结点权重值。

在 XGBoost 模型中,目标优化函数公式为:

$$O^{(k)} = \sum_{i=1}^n L(y_i, \hat{y}_i) + \sum_{i=1}^k \Omega(f_i) \quad (13)$$

$$\Omega(f_i) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T \omega_j^2 \quad (14)$$

式中: i 为第 i 个样本; n 为样本的总数; k 为第 k 棵树; $L(y_i, \hat{y}_i)$ 为损失函数,其中: y_i 为样本的真实值, $\hat{y}_i$ 为样本的预测值; $\Omega(f_i)$ 为正则函数,其中 γ 为树上的叶子总数, γ 和 λ 为正则参数,主要用于抑制叶子过度增加导致模型复杂化。

1.4 ICOA-XGBoost 模型建立

基于上述的研究,提出一种 ICOA 优化 XGBoost 参数的光伏阵列故障模型,具体流程如图 1 所示。

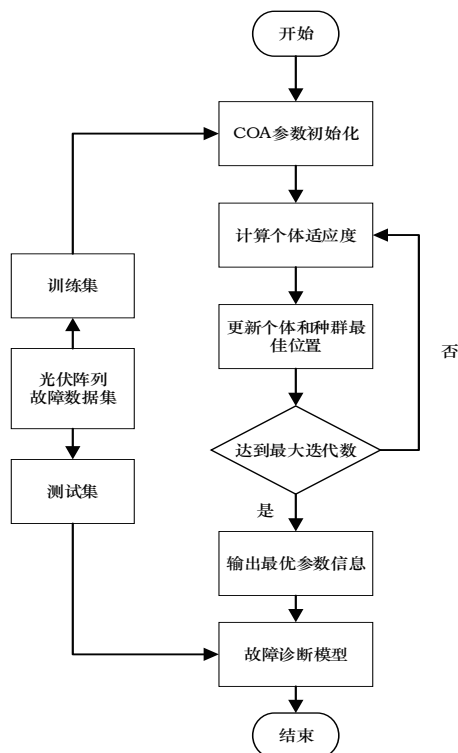


图 1 基于 ICOA-XGBoost 故障诊断流程图

2 ICOA 算法性能测试

为了验证 ICOA 算法性能,将 COA 算法、DBO 算法和 WOA 算法作为对比试验,设置最大迭代次数为 200,种群规模为 30,维数为 30,使用表 1 中的 4 个基准函数,其中 F1 和 F2 为单峰函数, F3 和 F4 为多峰函数。使用测试指标为 mean 和 std。表 2 为 30 次测试后的平均测试结果^[11]。实验结果如表 2 所示。

表 1 基准测试函数

编号	函数	搜索区间	理论极值
F1	$f_1(x) = \sum_{i=1}^n x_i^2$	[-100,100]	0
F2	$f_2(x) = \sum_{i=1}^n x_i + \prod_{i=1}^n x_i $	[-10,10]	0
F3	$f_3(x) = \sum_{i=1}^n [x_i^2 - 10 \cos(2\pi x_i) + 10]$	[-5.12,5.12]	0
F4	$f_{i1} = 1 + \frac{1}{4000} \sum_{i=1}^n x_i^2 - \prod_{i=1}^n \cos\left(\frac{x_i}{\sqrt{i}}\right)$	[-600,600]	0

表 2 测试函数结果对比

函数	指标	ICOA	COA	WOA	DBO
F1	mean	0	0	$1.131e^{-3}$	$3.630e^{-3}$
	std	0	0	$1.712e^{-3}$	$1.131e^{-3}$
F2	mean	0	$7.158e^{-72}$	$1.874e^{-20}$	$4.237e^{-19}$
	std	0	$3.795e^{-71}$	$5.474e^{-20}$	$1.651e^{-18}$
F3	mean	0	0	8.379	$4.520e^{-1}$
	std	0	0	17.579	1.418
F4	mean	0	0	$1.172e^{-1}$	$1.410e^{-2}$
	std	0	0	$2.763e^{-1}$	$3.750e^{-2}$

由表 2 可知, ICOA 总能找到理论极值, 且 std 为 0 说明 ICOA 寻优能力强且稳定, 而 COA 算法也能得到理论极值点, 但寻优需迭代次数比 ICOA 多。表明了该算法所采用的优化策略, 既能加快算法的收敛速度, 也加强了寻优能力。

3 ICOA-XGBoost 光伏故障诊断仿真实验

3.1 实验设计

在实际故障中, 光伏故障一般为短路、开路、老化、阴影遮挡 4 种故障。本文通过 Simulink 仿真搭建光伏阵列仿真模型, 以此获取不同状态下的数据。该阵列主要由 4×4 的光伏组件构成串并联结构^[12]。

其中, 短路故障仿真采用连接线将组件短路连接; 开路故障仿真采用组件直接断开的方式; 阴影遮挡故障仿真采用辐照度 $G = 200 \text{ W/m}^2$; 老化故障仿真采用串联电阻方式, 主线路串联 5Ω 为轻微老化, 串联 10Ω 为严重老化。阵列运行条件在标准情况下 ($T = 25 \text{ }^\circ\text{C}$, $G = 1000 \text{ W/m}^2$)。本文选取 U_{oc} 、 I_{sc} 、 U_{mp} 、 I_{mp} 作为故障特征参数。在 Simulink 仿真

实验中，光照辐照度取值范围为 $G \in [300,1000] \text{W/m}^2$ ，步长为 20W/m^2 ；组件环境温度为 $T \in [15,45]^\circ\text{C}$ ，步长为 2°C 。以此收集到 5250 组数据。按 8:2 比例将数据分为训练集和测试集。最后将得到的数据输入到分类模型中进行故障分类。

3.2 仿真结果分析

为了比较 ICOA-XGBoost 的性能，保持基本参数不变。（最大迭代数为 30，种群规模为 20）。将数据输入到 COA-XGBoost、WOA-XGBoost、DBO-XGBoost 模型中进行故障诊断，实验结果适应度寻优图如图 2 所示，准确率如表 3 所示。

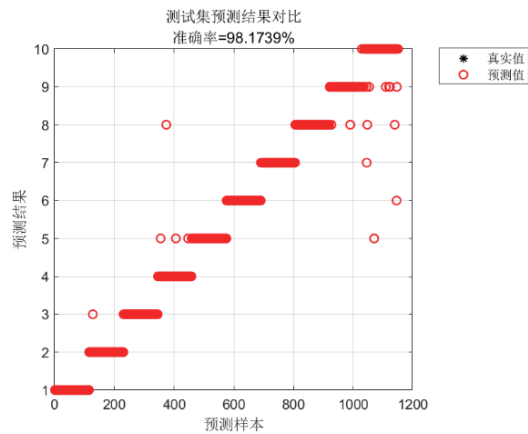


图 2 ICOA-XGBoost 故障诊断模型结果

表 3 不同模型的诊断准确率和系数

分类算法	准确率 /%
ICOA-XGBoost	98.173 9
COA-XGBoost	96.256 4
DBO-XGBoost	91.826 1
WOA-XGBoost	90.784 5

由图 2 可知，ICOA-XGBoost 算法准确率达到 98.173 9%，ICOA 优化 XGBoost 算法效果好，相比 COA-XGBoost 算法准确率提高 1.917 5%。相比算法提高 6.347 8%~7.389 4%。实验结果表明改进策略能有效提高故障识别率，解决 XGBoost 算法参数选择问题。

4 结论

提出的 ICOA 算法与 XGBoost 结合应用于光伏阵列故障诊断中，实验对 4 种类型故障进行仿真模拟，引入 Logistic-Tent 混沌映射初始化种群、自适应权重因子、Levy 飞行策略和透镜成像学习策略，对基础 COA 算法进行改进，从而得到 ICOA 算法。通过与传统 COA 算法、WOA 算法、DBO 算进行比较，实验表明该改进算法有更好的稳定性和寻优能力。通过对 4 种模型进行故障仿真分析，实验表明 ICOA-XGBoost 算法相比其他模型具有更高的准确率。

参考文献：

[1] 颜奕,黄宇,陈鸣,等.基于电容阵列 I-V 测试仪的光伏 STC 曲线拟合研究[J].电子测量与仪器学报,2017,31(2):272-279.

[2] 吴亚钧,王璐,张金江.基于 IDBO-LightGBM 的光伏阵列故障诊断方法[J/OL].电源学报,1-15[2024-05-14].http://kns.cnki.net/kcms/detail/12.1420.TM.20240426.1116.048.html.

[3] 王一鸣,许颇,张凌翔,等.基于 SSA-SVM 的光伏阵列故障诊断模型[J].电子设计工程,2024,32(13):46-49.

[4] 罗凯元,刘光宇,毛嘉铭.基于改进 BiLSTM 模型的光伏阵列故障诊断[J/OL].电源学报,1-13[2024-05-11].http://kns.cnki.net/kcms/detail/12.1420.tm.20240517.1702.001.html.

[5] 刘建锋,李志远,周亚茹.基于漏磁场和 ICOA-ResNet 的变压器绕组早期故障诊断[J].电力系统保护与控制,2024,52(9):99-110.

[6] 秦秋霞,梁仲月,徐毅.基于 Logistic-Tent 混沌映射和位平面的图像加密算法[J].大连民族大学学报,2022,24(3):245-252.

[7] 唐浩漾,王亦凡,秦波,等.基于改进 SSA 优化 SVM 的滚动轴承故障诊断方法[J].计算机技术与发展,2024,34(5):175-182.

[8] 黄鹤,高永博,茹锋,等.基于自适应黏菌算法优化的无人机三维路径规划[J].上海交通大学学报,2023,57(10):1282-1291.

[9] 史洪岩,蔡志豪.基于自适应权重与透镜成像学习的麻雀算法[J].电子设计工程,2024,32(5):13-18+24.

[10] 谭海旺,杨启亮,邢建春,等.基于 XGBoost-LSTM 组合模型的光伏发电功率预测[J].太阳能学报,2022,43(8):75-81.

[11] 周娇.基于图像分割的群智能算法改进研究[D].贵阳:贵州大学,2021.

[12] 高天.基于增强核极限学习机的光伏阵列状态检测研究[D].淮南:安徽理工大学,2024.

【作者简介】

董建业(2000—),男,江苏淮安人,硕士研究生,研究方向:光伏阵列故障检测。

李红月(1978—),女,安徽萧县人,硕士,副教授,研究方向:电力电子与电力传动。

(收稿日期:2024-08-02)