

DOI:10.3969/j.issn.2097-0706.2023.10.007

基于改进原子轨道搜索算法优化随机森林分类器的光伏系统故障诊断

PV system fault diagnosis based on random forest classifier optimized by improved atomic orbital search algorithm

杨晓燕¹, 谢满承¹, 郭小璇², 赵岩¹, 陈翀旻¹, 陈子民¹, 廖卓颖^{3*}
YANG Xiaoyan¹, XIE Mancheng¹, GUO Xiaoxuan², ZHAO Yan¹,
CHEN Chongmin¹, CHEN Zimin¹, LIAO Zhuoying^{3*}

(1. 广西电网有限责任公司南宁供电局, 南宁 530001; 2. 广西电网有限责任公司电力科学研究院, 南宁 530023;
3. 中国科学院广州能源研究所, 广州 501640)

(1. Nanning Power Supply Bureau of Guangxi Power Grid, Nanning 530001, China; 2. Electric Power Research Institute, Guangxi Power Grid Company Limited, Nanning 530023, China; 3. Guangzhou Institute of Energy Research, Chinese Academy of Sciences, Guangzhou 501640, China)

摘要: 针对光伏系统故障难以被准确高效地诊断和分类的问题, 提出了一种基于改进原子轨道搜索优化的随机森林(IAOS-RF)算法。此算法在光子的发射和吸收部分引入了自适应权重机制和反向学习机制, 用于更新电子的位置, 能有效加强算法在搜索空间的全面勘探和开发能力。基于一组并网光伏系统故障数据进行算例分析, 对比了 IAOS-RF 算法和几类基准算法的性能差异, 结果显示, IAOS-RF 算法故障分类准确率最高并且可达到 98%, 其诊断结果趋于稳定所需的迭代次数最小, 具有较快的收敛速度。最后针对该算法存在的一些局限性和改进空间, 提出未来需要进一步研究和探讨的问题。

关键词: 光伏发电系统; 故障诊断; 自适应权重; 反向学习机制; 改进原子轨道搜索; 随机森林

中图分类号: TM 615:TK 01*9 **文献标志码:** A **文章编号:** 2097-0706(2023)10-0053-08

Abstract: The random forest classifier optimized by improved atomic orbital search algorithm (IAOS-RF) is applied in PV system fault diagnosis and classification to improve the accuracy and effective. This algorithm introduces adaptive weight mechanism and reverse learning mechanism to photon emission and photon absorption to update the position of electrons, which can effectively enhance the algorithm's comprehensive exploration and development capabilities in the search space. Based on a set of fault data from a grid-connected PV system, the differences in performances between the proposed improved algorithm and fundamental algorithms are compared. The results showed that IAOS-RF has the highest fault classification accuracy among the algorithms, reaching 98%. At the same time, its diagnosis, with a fast convergence rate, requires the least times of iterations to be stable. In the end, in the view of the limitations in the proposed algorithm, the problems need to be improved in the future are discussed.

Keywords: photovoltaic system; fault diagnosis model; adaptive weight; reverse learning mechanism; improved atomic orbital search; random forest

0 引言

随着我国“双碳”目标的提出及碳减排工作的不断深入, 以新能源为主体的新型电力系统快速发展, 光伏发电因可用性高、安装速度快、维护成本

低、能源独立性强, 成为常用的可再生清洁能源之一^[1-2]。国家能源局公布的数据显示, 截至 2022 年年底, 全国累计光伏并网容量约为 3.920 4 亿 kW, 同比增长了 28.12%, 其中集中式光伏电站累计并网容量为 2.344 2 亿 kW, 分布式光伏电站累计并网容量为 1.576 2 亿 kW^[3]。光伏电站系统由组件、逆变器、汇流箱、支架等多个部件组成, 它们通常在恶劣的室外环境中运行, 强风、大雨、冰雹、污染、紫外

基金项目: 广西电网有限责任公司科技项目(GXKJXM20220069)
Science and Technology Project of Guangxi Power Grid Company Limited(GXKJXM20220069)

线、热循环和不当的运输、安装、维护等都会导致光伏系统出现各种故障^[4]。光伏系统部件众多,任何一个环节出现问题,都会影响电站的正常运行^[5],轻则损失发电量(据 Firth 等的统计,光伏阵列故障可导致输出电能减少 18.9%^[6]),重则引起火灾等。光伏系统故障大多发生在直流侧,即模块或阵列本身,也可能发生在交流侧,涉及逆变器或与电网的连接异常^[7]。据统计,组件、逆变器、汇流箱等直流侧设备故障占比高达 90.18%;电缆、箱式变电站、土建、升压站等交流侧设备故障占比达 9.82%^[8]。为了确保光伏阵列可靠运行,近年来许多国内外学者对故障检测和诊断方法进行了广泛研究。

通过时序电流-电压(I - V)测量进行光伏阵列故障诊断是目前较普遍的诊断方法^[9],光伏电站数据采集与监视控制(SCADA)系统通过传感器测量并采集各部件的相关运行参数,这些电压、电流、频率等参数能反映组件的运行状态信息^[10]。对比电压、电流、频率等数据正常运行值和实际值间的差值,经过算法分析得到故障诊断的结果^[9]。在分析计算过程中,常采用群智能算法对核极限学习机^[11-12]、神经网络^[13]、随机森林(Random Forest, RF)^[14]等分类诊断算法进行优化。文献[11]提出了一种基于萤火虫扰动的改进麻雀搜索算法优化传统极限学习机的光伏系统故障诊断方法,能够有效解决传统故障诊断方法收敛速度慢的问题。文献[12]利用 Levy 飞行策略和自适应 t 分布策略对麻雀搜索算法进行改进,再利用改进后的麻雀搜索算法对核极限学习机的核参数和正则化系数进行优化,提高了核极限学习机分类的准确率。文献[13]提出了利用自适应权重粒子群对反向传播(Back Propagation, BP)神经网络优化的光伏阵列故障诊断方法,提高了故障诊断的实时性。文献[14]提出了一种基于级联 RF 的光伏组件在线故障诊断模型,其准确率和收敛时间都优于单一的 RF 模型与 BP 神经网络模型。

神经网络故障诊断方法对样本数量要求高,对相关参数的依赖性强;核极限学习机方法对数据集中存在的离群点过于敏感;而 RF 算法对泛化误差采用无偏估计,具有较强的泛化能力及强抗干扰和抗过拟合特性。因此,本文选择 RF 分类器作为光伏系统故障诊断的基础算法,同时在原子轨道搜索(Atomic Orbital Search, AOS)算法的基础上提出改进的 IAOS 算法,共同构建光伏阵列故障诊断模型。首先,对光伏系统故障进行分类;然后,用 IAOS 算法优化 RF 分类器,形成更加高效的 IAOS-RF 模型;最后,对光伏系统故障进行仿真试验分析。

1 光伏系统故障分类

本文使用实验室可实现的典型并网光伏系统来验证所提算法在最大功率点跟踪(Maximum Power Point Tracking, MPPT)模式下的故障诊断性能。该光伏系统由 $N \times M$ 个光伏组件阵列、反馈传感器、MPPT 控制器、逆变器和升压变压器组成,系统结构如图 1 所示。

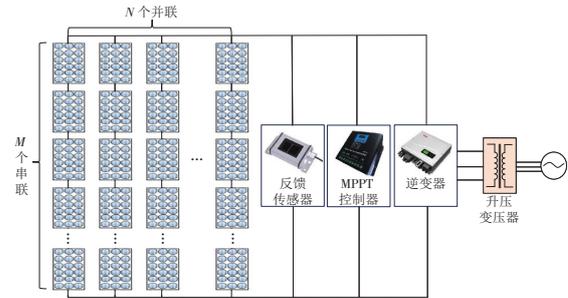


图 1 光伏系统结构

Fig. 1 Structure of a PV system

常见的光伏系统故障只考虑了光伏组件老化、遮阴问题和阵列之间的短路、开路问题^[15-17],很少涉及系统中传感器、控制器、逆变器等元件的故障。本文除了常见的光伏阵列故障,更侧重对不同元件的故障类型进行分类,所述光伏系统不同工作状态包括正常状态、逆变器故障、反馈传感器故障、电网异常故障、局部阴影故障、单支路开路故障、MPPT 控制器故障和升压转换器控制器故障,并对此进行编号。表 1 为以上运行状态及故障成因。

2 算法原理与改进

2.1 AOS 算法基本原理

AOS 算法是在原子搜索算法基础上提出的新型元启发式算法,原子间的相互作用力是该算法数学模型的关键,而电子密度配置和原子吸收或发射能量的基本原理则是 AOS 算法的主要思想^[18]。AOS 算法模拟了电子在高低两种能态之间的转移过程,在该算法中,每个电子代表一个可行解 X ,搜索空间为围绕原子核的电子云,其物理模型可分成不同半径的薄球壳层,也称为电子轨道。每个电子由搜索空间中的可行解 X_i 表示,决策变量 x_j^i 用于定义候选解在搜索空间中的位置,该目标的数学方程为

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_i \\ \vdots \\ X_m \end{bmatrix} = \begin{bmatrix} x_1^1 & x_1^2 & \cdots & x_1^j & \cdots & x_1^d \\ x_2^1 & x_2^2 & \cdots & x_2^j & \cdots & x_2^d \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ x_i^1 & x_i^2 & \cdots & x_i^j & \cdots & x_i^d \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ x_m^1 & x_m^2 & \cdots & x_m^j & \cdots & x_m^d \end{bmatrix},$$

$$i = 1, 2, \dots, m; j = 1, 2, \dots, d, \quad (1)$$

表 1 光伏系统运行状态及故障成因

Table 1 Operation status and fault causes of the PV system

运行状态	故障成因
1 正常状态	无
2 逆变器故障	组件电压过低、输入端子反接、直流开关故障、元件内部电路损坏、接线端子和电缆接头接触不良等
3 反馈传感器故障	接线端子和电缆接头接触不良、元件内部电路损坏、瞬态冲击电流过大等
4 电网异常故障	电网电压和频率过低或过高、电网阻抗过大等
5 局部阴影故障	树叶、积聚的污垢或灰尘以及鸟粪等附着在光伏板表面
6 单支路开路故障	端子附近电力电缆老化、电缆过热、太阳能电池之间连接不牢固、电缆终端松动、接线盒连接器松动以及面板电缆或连接故障
7 MPPT 控制器故障	电池反接、电池损坏、直流负载超出额定电流、机器内部过热、光伏板电压过高或过低、元件内部电路损坏、接线端子和电缆接头接触不良等
8 升压转换器控制器故障	内部绕组铁芯引线故障、外部短路事故频繁、接线端子和电缆接头接触不良等

式中： m 为搜索空间内电子的数量； d 为可行解的维数。

电子云中电子的初始位置根据以下数学方程随机确定。

$$x_i^j(0) = x_{i,\min}^j + r(x_{i,\max}^j - x_{i,\min}^j),$$

$$i = 1, 2, \dots, m; j = 1, 2, \dots, d, \quad (2)$$

式中： $x_i^j(0)$ 为可行解的初始位置； $x_{i,\min}^j$ 和 $x_{i,\max}^j$ 分别为第 i 个电子第 j 个决策变量的最小和最大界限； r 为 $[0, 1]$ 范围内均匀分布的随机变量。

在数学模型中，每个电子的能量状态即为目标函数值，具有更好目标函数值的可行解代表更低能级的电子，下列向量方程用于包含不同可行解（电子）的目标函数值（能级）。

$$E = [E_1 \ E_2 \ \dots \ E_i \ \dots \ E_m]^T,$$

$$i = 1, 2, \dots, m, \quad (3)$$

式中： E 为目标函数值的向量； E_i 为第 i 个电子的能级。

在基于量子的原子模型中，电子在原子核周围的位置由电子的概率密度函数 (Probability Density Function, PDF) 决定。原子核周围的电子云分为多个虚拟层，PDF 用于确定可行解在这些层中的位置。可行解基于最小化或最大化优化问题按升序或降序进行排序，较好目标函数值的可行解代表较低能级的电子。因此，具有较高 PDF 值的可行解位于内部虚拟电子层，而具有较低 PDF 值的可行解位于外部虚拟电子层，如图 2 所示。

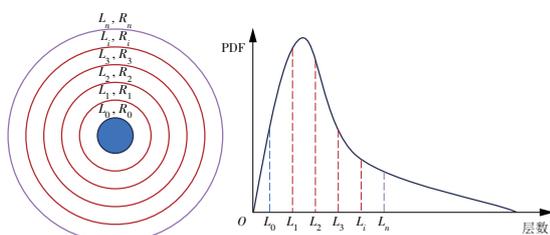


图 2 通过 PDF 确定可行解位置

Fig. 2 Determining the positions of solution candidates by PDF

图 2 所示的原子轨道模型中，随机整数 n 用于表示虚拟层 L 的数量。这些层将整个搜索空间划分为多个部分，以便在数学上表示原子核周围电子的波动行为。这些层的半径表示这些层在核周围的分布方式，其中较小半径的层为核层 (L_0)，较大半径的层为核周围 n 个 (L_1-L_n) 球形层。

虚拟层中可行解位置 X^k 和目标函数值 E^k 的向量方程如下

$$X^k = \begin{bmatrix} X_1^k \\ X_2^k \\ \vdots \\ X_i^k \\ \vdots \\ X_p^k \end{bmatrix} = \begin{bmatrix} x_1^1 & x_1^2 & \dots & x_1^j & \dots & x_1^d \\ x_2^1 & x_2^2 & \dots & x_2^j & \dots & x_2^d \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ x_i^1 & x_i^2 & \dots & x_i^j & \dots & x_i^d \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ x_p^1 & x_p^2 & \dots & x_p^j & \dots & x_p^d \end{bmatrix},$$

$$i = 1, 2, \dots, p; j = 1, 2, \dots, d; k = 1, 2, \dots, n, \quad (4)$$

$$E^k = [E_1^k \ E_2^k \ \dots \ E_i^k \ \dots \ E_p^k]^T,$$

$$i = 1, 2, \dots, p; k = 1, 2, \dots, n, \quad (5)$$

式中： X_i^k 为第 k 个虚拟层中的第 i 个可行解； n 为虚拟创建的最大层数； p 为第 k 个虚拟层的可行解总数； E_i^k 为第 k 个虚拟层中第 i 个可行解的目标函数值。

在每个虚拟层中，具有最佳目标函数值的可行解 (E_i^k) 被视为每个虚拟层中具有最低能级的电子，而在所有可行解中具有最佳目标函数值的可行解则是原子中具有最低能级 (E_1) 的电子。从当前虚拟层移除电子所需的结合能与当前电子的位置和目标函数值有关，因此可通过所选择层中所有可行解的位置和目标函数值的平均值来确定该层中可行解的结合状态和结合能，其数学方程为

$$S_B^k = \frac{\sum_{i=1}^p X_i^k}{p},$$

$$i = 1, 2, \dots, p; k = 1, 2, \dots, n, \quad (6)$$

$$E_B^k = \frac{\sum_{i=1}^p E_i^k}{p},$$

$$i = 1, 2, \dots, p; k = 1, 2, \dots, n, \quad (7)$$

式中: S_B^k 和 E_B^k 分别为第 k 层可行解的平均结合态和结合能。

因此,可以通过总搜索空间中所有可行解的位置和目标函数值的平均值得到原子的结合态和结合能。

$$S_B = \frac{\sum_{i=1}^m X_i}{m}, i = 1, 2, \dots, m, \quad (8)$$

$$E_B = \frac{\sum_{i=1}^m E_i}{m}, i = 1, 2, \dots, m, \quad (9)$$

式中: S_B 和 E_B 分别为原子的结合态和结合能; X_i 和 E_i 分别为原子中第 i 个可行解的位置和目标函数值。

电子主要有 2 种位置更新过程,其中光子对电子的作用是主要过程,此时在 $(0, 1)$ 范围内的随机生成数 ϕ 大于光子速率 v_p ($\phi > v_p$);而与粒子或磁场的相互作用等是次要过程,此时 $\phi < v_p$,建立如下数学模型。

(1)当电子能级大于该层结合能($E_i^k > E_B^k$)时,可行解发射光子。

$$X_{i+1}^k = X_i^k + \frac{\alpha_i(\beta_i E_i - \gamma_i S_B)}{k}, \quad i = 1, 2, \dots, p; k = 1, 2, \dots, n, \quad (10)$$

式中: X_i^k 和 X_{i+1}^k 分别为第 k 层第 i 个可行解的当前和未来位置; $\alpha_i, \beta_i, \gamma_i$ 均为包含 $(0, 1)$ 内均匀分布的随机数的向量,用于确定发射光子时释放的能量。

在这个过程中,可行解倾向于发射具有 β 和 γ 能量的光子,以便同时达到原子的结合态(S_B)和原子中具有最低能级(E_1)的最优适应度。

(2)当电子能级小于该层结合能($E_i^k < E_B^k$)时,可行解吸收光子。

$$X_{i+1}^k = X_i^k + \alpha_i(\beta_i E_1^k - \gamma_i S_B^k), \quad i = 1, 2, \dots, p; k = 1, 2, \dots, n, \quad (11)$$

式中: $\alpha_i, \beta_i, \gamma_i$ 均为包含 $(0, 1)$ 内均匀分布的随机数的向量,用于确定吸收光子时所需能量。

在这个过程中,可行解倾向于吸收具有 β 和 γ 能量的光子,以便同时达到第 k 层的结合态(S_B^k)和第 k 层内具有最低能级(E_1^k)的最优解。

(3)可行解与粒子或磁场的相互作用。

$$X_{i+1}^k = X_i^k + r_i, \quad i = 1, 2, \dots, p; k = 1, 2, \dots, n, \quad (12)$$

式中: r_i 为包含 $(0, 1)$ 内均匀分布的随机数的向量,用于确定吸收或发出能量的大小。

2.2 IAOS 算法

在光子的发射、吸收部分的更新策略中引入自适应权重和反向学习机制更新电子位置,可以提高搜索空间的全面勘探和开发能力,更加快速准确地找到最佳值。

2.2.1 自适应权重机制

针对算法末期可能出现停滞现象而易陷入局部最优的问题,引入自适应权重 φ ,增强全局收敛能力,使电子在当前最优位置附近进行随机搜索,自适应权重 φ 的调整公式为

$$\varphi = 1 - \exp\left[-\left(2.5 \times \frac{M - N}{M}\right)^2\right], \quad (13)$$

式中: M 为最大迭代次数; N 为当前迭代次数,此时可行解的位置变化公式更新为

$$X_{i+1}^k = X_i^k + \frac{\alpha_i(\varphi\beta_i E_1 - \gamma_i S_B)}{k}, \quad i = 1, 2, \dots, p; k = 1, 2, \dots, n, \quad (14)$$

$$X_{i+1}^k = X_i^k + \alpha_i(\varphi\beta_i E_1^k - \gamma_i S_B^k), \quad i = 1, 2, \dots, p; k = 1, 2, \dots, n. \quad (15)$$

2.2.2 反向学习机制

利用反向学习策略生成初始电子群。先随机生成初始种群,然后根据初始种群生成其反向种群并从中选择较优的种群作为下一代种群。反向学习策略会选择更靠近的个体作为种群的最初个体,这样每个个体都离最优解更近一步,以便提高种群中所有个体的收敛速度;同时,反向学习策略可通过搜索更多有效区域来提高群体的多样性,增强算法的全局搜索能力。其中针对整体最优个体,对 E_1 进行反向学习。

$$X_{i+1}^* = R(b_l + b_u) - X_{i+1}, \quad (16)$$

式中: X_{i+1}^* 为对整体最优个体 E_1 (原子中最低能量状态的个体)进行反向学习后的个体; b_l 和 b_u 分别控制域的下限和上限; R 为 $(0, 1)$ 内均匀分布的随机数。

针对各层最优个体,对 E_1^k 进行反向学习。

$$X_{i+1}^{k*} = R(b_l + b_u) - X_{i+1}^k, \quad (17)$$

式中: X_{i+1}^{k*} 为对当前层最优个体 E_1^k (原子中最低能量状态的个体)进行反向学习后的个体。

通过比较 X_{i+1}^* 与 X_{i+1}^{k*} 的适应度,保留适应度较小的个体作为第 k 层第 i 个电子;更新后变成第 k 层第 $i+1$ 个电子的个体。

2.3 IAOS-RF 算法

利用 IAOS 算法对 RF 的决策树数目和决策树深度进行优化求解,形成更加高效的组合训练模型。选取该模型的期望输出值和实际输出值的均方差为目标函数即适应度值,利用原子轨道中的电子聚集来确定决策树数目和决策树深度,当均方差值最小时,就确定了 RF 的一组理论最优参数。IAOS-RF 模型的具体实现步骤如下。

(1)数据预处理。对采集的原始特征参数进行数据预处理,包括数据归一化和数据打乱等。

(2)算法初始化。初始化 AOS 算法参数和 RF 的最大迭代次数,其中 AOS 算法参数及其常见的取值范围为:轨道数 $n \in [2, 10]$, 原子数 $i \in [2, 50]$, 光子速率 $v_p \in (0, 1)$ 。

(3)计算个体适应度值。每个电子都有一个能量状态,对应各个个体的适应度值。

(4)确定各个层级的结合能和结合态。将各层中的电子位置平均值记为结合态,结合能为各个个体适应度值的平均值。

(5)更新电子位置。引入自适应权重和反向学习机制对光子的发射、吸收部分的更新策略进行改进,更新电子状态。

(6)更新电子最优位置与适应度值。在寻优迭代过程中,对电子的位置向量和适应度进行更新。

(7)确定 RF 分类器最优参数。重复步骤(3)–(6),当达到算法最大迭代次数或满足算法设定精度时,则认为算法趋于收敛,停止迭代,输出理论最优决策树数目和决策树深度给 RF 分类器。

(8)抽取子数据集。采取有放回的抽样方式在原始数据集中进行子数据集的随机抽取。

(9)构建决策树。假设一个子数据集有 X 个属性,决策树的节点需要分裂时,从这些属性中随机抽选出 Y 个属性 ($Y \ll X$);再通过信息增益方式从 Y 个属性中选择一个作为该节点的分裂属性;不断重复这个步骤,直到达到决策深度值为止。

(10)构建 RF。按照步骤(8)–(9)来构建大量若干的决策树直到达到决策树数目值为止,这些子决策树就会构成 RF。

(11)进行故障诊断。将数据集输入不同的子决策树,得到不同的判断结果,判断结果最多的就是 RF 的最佳故障诊断分类结果。

3 算例分析

3.1 数据样本选取

本文数据来自 Azzeddine Bakdi 提供的实验室并网光伏系统故障数据^[19]。在并网光伏系统的 8 种工作状态下分别采集 200 组数据,共 1 600 组,随机抽取 1 200 组数据作为故障诊断的训练样本,其余 400 组数据作为该模型的测试样本。采用的光伏发电阵列工作于最大功率点,关注的电气参数包括光伏阵列电流 I_{pv} 、光伏阵列电压 V_{pv} 、直流电压 V_{dc} 、A 相电流 i_a 、B 相电流 i_b 、C 相电流 i_c 、A 相电压 v_a 、B 相电压 v_b 、C 相电压 v_c ,典型样本数据见表 2。

3.2 试验过程及诊断结果

将 IAOS-RF 算法的最大迭代次数设为 100,用于电子位置测定的光子率设为 0.1,电子数量设为

表 2 光伏阵列典型样本数据

Table 2 Typical sample data for PV arrays

运行状态	I_{pv}	V_{pv}	V_{dc}	i_a	i_b	i_c	v_a	v_b	v_c
1	2.37	90.43	147.95	0.62	0.53	0.16	146.01	120.25	25.59
2	2.23	88.66	147.66	0.66	0.36	0.31	153.80	91.08	63.42
3	2.34	85.33	146.78	0.33	0.07	0.69	125.63	14.94	141.55
4	2.48	90.43	148.24	0.62	0.19	0.40	154.72	63.24	91.31
5	2.05	86.85	145.31	0.34	0.08	0.51	123.49	12.33	143.67
6	2.37	86.91	147.07	0.64	0.46	0.27	154.85	91.81	62.18
7	2.11	92.41	147.36	0.67	0.26	0.44	151.22	49.19	103.51
8	2.27	87.81	149.12	0.74	0.34	0.48	152.94	53.30	102.53

30,虚拟层数设为 5。以上参数对算法性能的影响是相互关联的:较小的最大迭代次数和较高的光子率可加快算法的收敛速度,但也可能降低搜索的精度;较多的电子数量和较大的虚拟层数可以增加全局搜索的能力,但也增加了计算和存储的压力。因此,选择适当的参数组合以平衡算法的速度和品质至关重要。本文参数设定原因如下。

(1)本算法用于故障诊断场景时,目标函数趋于稳定所需的迭代次数基本都少于 100,因此将最大迭代次数设为 100。较小的迭代次数可以在相对较短的时间内得到结果,适用于简单问题或需要较快得到结果的场景,避免算法过度迭代。

(2)较高的光子率意味着光子更有可能进入较远的位置,扩大搜索空间,以便更好地发现潜在的解,但会降低搜索精度。综合考虑精度和收敛速度要求,选择光子率为 0.1。

(3)较多的电子数量可以提高算法的全局搜索能力,但也增加了计算和存储的成本,选择适当的电子数量可在考虑效率的同时保持算法的搜索品质。为了平衡搜索的速度和精度,将电子数量设为 30。

(4)通过引入虚拟层,算法可以探索更多不同的解并避免陷入局部最优解,较大的虚拟层数将增加计算和存储要求,但可以提高算法的搜索效果。为了增加搜索的灵活性和多样性,将虚拟层数设为 5。

利用 3.1 节采集的数据集,在 IAOS-RF 算法模型中进行迭代训练,将训练结果与 AOS-RF 模型和传统 RF 模型的训练结果进行对比,其中各电气特征参数在诊断过程中的重要性如图 3 所示。其中, I_{pv} , V_{pv} , V_{dc} 这 3 个特征参数占较高比重。由此可见,不同运行状态下,这 3 个参数差异较大,其阈值差能有效帮助识别故障类型。

经过仿真试验得到 IAOS-RF 和 AOS-RF 算法

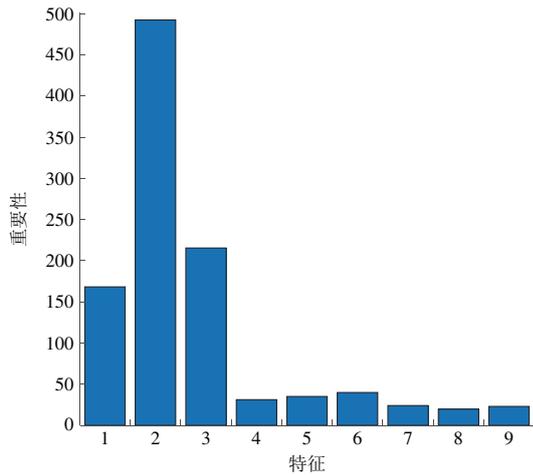


图3 电气特征参数的重要性

Fig. 3 Importance of electrical characteristic parameters

模型的迭代次数和适应度值,结果见表3。由表3可以看出:IAOS-RF算法模型迭代55次左右时目标函数的结果就趋于稳定,而AOS-RF需要迭代70次左右才能趋于稳定,说明改进后的算法可以更好地跳出局部极值点,能以更快的速度收敛到全局最优;同时,IAOS-RF算法模型适应度为0.032,而

AOS-RF算法模型的适应度为0.035,说明改进后的算法寻优精度更好。这是因为引入了自适应权重和反向学习机制对光子发射和吸收部分的更新策略进行改进,用于更新电子状态。自适应权重机制的作用是根据当前搜索状态和解的情况,动态调整光子发射和吸收的权重,使得算法能够更好地平衡探索和利用的能力,更有效地搜索解空间;通过引入反向学习机制,电子的位置将根据光子发射和吸收的结果进行更新,这种更新机制使得电子能够更快地朝着更优解的方向移动,从而加速算法的收敛。

表3 2种算法模型的迭代次数和适应度

Table 3 Iteration times and fitness of the two models

算法模型	迭代次数	适应度
IAOS-RF	55	0.032
AOS-RF	70	0.035

对故障特征参数的样本数据进行训练后,将其余400组测试集数据输入3种模型,得到的测试集的故障诊断分类结果见表4。

表4 3种算法模型诊断结果

Table 4 Diagnostic accuracy of the three models

算法模型	项目	运行状态								平均
		1	2	3	4	5	6	7	8	
RF	误诊样本数	9	9	2	10	3	26	0	1	
	精度/%	76.9	72.7	96.2	77.8	94.3	45.7	100.0	95.8	77.25
AOS-RF	误诊样本数	6	0	0	9	0	0	2	3	
	精度/%	87.2	100.0	100	82.4	100.0	100.0	96.2	94.3	95.00
IAOS-RF	误诊样本数	2	2	0	3	0	0	1	0	
	精度/%	96.0	96.0	100.0	94.0	100.0	100.0	98.0	100.0	98.00

(1)在相同样本数据集下,IAOS-RF算法模型的诊断精度达到了98.00%,略高于AOS-RF的95.00%,二者均远高于RF算法模型的77.25%。这是由于AOS优化了RF分类器的决策树数目和决策树深度,使得分类器中的决策树可以更好地将光伏阵列的故障数据分解为不同维度的细节信息,理论上可以不断逼近实际故障类型的输入输出非线性映射关系。

(2)在测试的400组数据中,IAOS-RF算法模型仅误诊断8组数据,IAOS-RF算法模型的测试集预测结果如图4所示。AOS-RF算法模型误诊断20组数据,传统RF模型则误诊断60组数据。表明改进的IAOS优化RF分类器后,有效增强了故障模型的分类诊断能力,在区域搜索过程中收敛速度和适应度均得到了一定提升,因此可以更准确地对光伏系统故障进行诊断。另外,对于第4种故障,即电网异

常故障,IAOS-RF算法和AOS-RF算法的误诊率均为最高。究其原因,可能是样本不平衡以及算法对特定故障类型的鲁棒性不足等。

针对这些问题,提出以下能增强算法对各种故

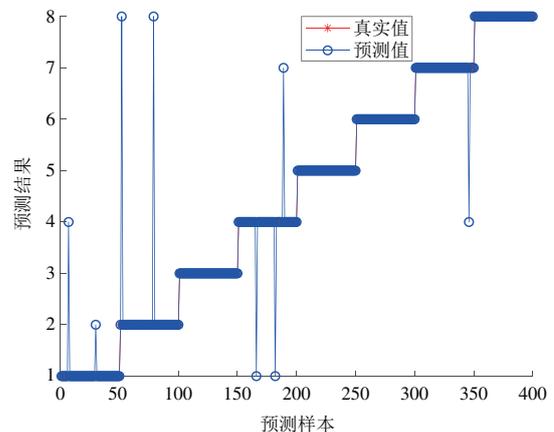


图4 IAOS-RF算法模型测试集预测结果

Fig. 4 Predicted results of the testing set by IAOS-RF algorithm

障类型适应能力的方法。

(1)数据清洗和预处理。引入更多的数据清洗和预处理步骤,减少噪声,处理缺失数据等,以提高算法对数据质量的鲁棒性并降低误诊断的风险。

(2)特征选取和特征工程。使用领域专家等知识,结合模型进行自动筛选和特征提取,进一步改善特征表示能力,优化特征选择的过程。

4 结论

针对光伏系统故障难以被准确高效地诊断和分类问题,利用 IAOS-RF 算法模型对光伏系统进行故障诊断,得到如下结论。

(1)IAOS 具有更强的寻优能力和更快的收敛速度,可有效对光伏系统故障进行诊断,不受模型其他参数影响。

(2)相较于传统 RF 和 AOS-RF 算法模型, IAOS-RF 算法模型可以更加迅速准确地识别出光伏阵列的各类故障,在故障诊断效率和精度上都有明显的优势。

(3)利用 IAOS-RF 算法建立基于数据驱动的诊断模型,在样本有限的情况下具有良好的泛化能力,能够有效提高光伏发电阵列故障分类的准确性,仿真证明该方法的故障诊断准确率可达 98%。

IAOS-RF 算法在光伏系统故障诊断方面取得了显著的成果,但仍存在一些局限性和改进空间,需要进一步研究和探讨。

(1)在现实应用中,光伏系统故障数据可能受到多种因素的影响,如噪声、缺失数据等。因此,如何处理和优化这些数据质量问题以及如何更准确地选择和提取故障特征,仍然是一个值得研究的方向。

(2)尽管 IAOS-RF 算法能够有效地处理光伏系统的故障诊断问题,但结合其他机器学习方法可能进一步提升其性能。例如,可以考虑结合深度学习方法,如卷积神经网络或循环神经网络,以便更好地解决特征提取和模式识别问题。

(3)随着光伏系统的不断发展,其结构和运行情况可能变得更加复杂。如何进一步提升 IAOS-RF 算法的泛化能力,以适应各种系统结构和工况变化,是未来研究的一个关键方向。

参考文献:

[1]ZHANG Y, ETEMAD A H. Fault detection for photovoltaic systems based on multiresolution signal decomposition and fuzzy inference systems [J]. IEEE Transactions on Smart

Grid, 2017, 8 (3): 1274-1283.

[2]王志峰,何雅玲,康重庆,等.明确太阳能热发电战略定位促进技术发展[J].华电技术,2021,43(11):1-4.

WANG Zhifeng, HE Yaling, KANG Chongqing, et al. Strategic positioning of solar thermal power generation to promote technological progress [J]. Huadian Technology, 2021, 43(11):1-4.

[3]国家能源局.2022年光伏发电建设运行情况[EB/OL].(2023-02-17)[2023-06-01].http://www.nea.gov.cn/2023-02/17/c_1310698128.htm.

[4]丁坤,刘增泉,张经纬,等.基于图像奇异值分解的局部遮挡光伏阵列输出特性建模研究[J].综合智慧能源,2023,45(2):53-60.

DING Kun, LIU Zengquan, ZHANG Jingwei, et al. Modeling for output characteristics of the partially occluded photovoltaic array based on image singular value decomposition [J]. Integrated Intelligent Energy, 2023, 45 (2): 53-60.

[5]ALAM M K, KHAN F, JOHNSON J, et al. A comprehensive review of catastrophic faults in PV arrays: Types, detection, and mitigation techniques [J]. IEEE Journal of Photovoltaics, 2015, 5(3): 982-997.

[6]FIRTH S K, LOMAS K J, REES S J. A simple model of PV system performance and its use in fault detection [J]. Solar Energy, 2010, 84(4): 624-635.

[7]MADETI S R, SINGH S. A comprehensive study on different types of faults and detection techniques for solar photovoltaic system [J]. Solar Energy, 2017, 158: 161-185.

[8]张永明,邱涛,任俊杰.一种光伏电站智能监控分析平台及方法与流程:CN111010084B[P].2021-09-24.

[9]乔苏朋,杨艳,陈世群,等.光伏阵列故障检测方法综述 [J].电气技术,2021,22(7):1-6.

QIAO Supeng, YANG Yan, CHEN Shiqun, et al. Review on photovoltaic array diagnosis methods [J]. Electrical Engineering, 2021, 22(7): 1-6.

[10]熊昌全,温贤茂,张宇宁,等.光伏组件发电故障诊断方法研究[J].科技创新与应用,2023,13(9):77-81,85.

XIONG Changquan, WEN Xianmao, ZHANG Yuning, et al. Research on fault diagnosis methods for photovoltaic module power generation [J]. Technology Innovation and Application, 2023, 13(9): 77-81, 85.

[11]赵靖英,吴晶晶,张雪辉,等.基于萤火虫扰动麻雀搜索算法-极限学习机的光伏阵列故障诊断方法研究[J].电网技术,2023,47(4):1612-1625.

ZHAO Jingying, WU Jingjing, ZHANG Xuehui, et al. Fault diagnosis of photovoltaic arrays based on sparrow search algorithm with firefly perturbation-extreme learning machine [J]. Power System Technology, 2023, 47 (4) : 1612-1625.

[12]武文栋,施保华,郑传良,等.基于KPCA-ISSA-KELM的

- 光伏阵列故障诊断方法[J]. 智慧电力, 2022, 50(11): 69-76.
- WU Wendong, SHI Baohua, ZHENG Chuanliang, et al. Fault diagnosis method of PV array based on KPCA-ISSA-KELM[J]. Smart Power, 2022, 50(11): 69-76.
- [13] 何根新, 周俊杰, 周家兴, 等. 基于自适应权重粒子群优化BP神经网络的光伏阵列故障诊断方法研究[J]. 电子设计工程, 2021, 29(9): 75-79.
- HE Genxin, ZHOU Junjie, ZHOU Jiaying, et al. Research on fault diagnosis method of photovoltaic array based on BP neural network with adaptive weight particle swarm optimization[J]. Electronic Design Engineering, 2021, 29(9): 75-79.
- [14] 叶进, 卢泉, 王钰淞, 等. 基于级联随机森林的光伏故障诊断模型研究[J]. 太阳能学报, 2021, 42(3): 358-362.
- YE Jin, LU Quan, WANG Yusong, et al. Research on PV fault diagnosis model based on cascade random forest[J]. Acta Energetica Solaris Sinica, 2021, 42(3): 358-362.
- [15] 吴文涛, 陈志聪, 吴丽君, 等. 利用改进初始化的残差-密集连接网络的光伏阵列故障诊断[J]. 福州大学学报(自然科学版), 2022, 50(2): 192-197.
- WU Wentao, CHEN Zhicong, WU Lijun, et al. A fault diagnosis of photovoltaic array based on improved initialization residual-dense connection network [J]. Journal of Fuzhou University (Natural Science Edition), 2022, 50(2): 192-197.
- [16] 余玲珍, 杨靖, 龙道银, 等. 典型故障条件下光伏阵列建模与仿真[J]. 电网与清洁能源, 2020, 36(10): 103-111, 118.
- YU Lingzhen, YANG Jing, LONG Daoyin, et al. Modeling and simulation of photovoltaic arrays under typical fault conditions[J]. Power System and Clean Energy, 2020, 36(10): 103-111, 118.
- [17] 段震清, 孙建民, 梁凌, 等. 基于XGBoost算法的光伏阵列故障诊断方法研究[J]. 太阳能, 2023, 345(1): 62-68.
- DUAN Zhenqing, SUN Jianmin, LIANG Ling, et al. Research on fault diagnosis method of photovoltaic system based on XGBoost algorithm[J]. Solar Energy, 2023, 345(1): 62-68.
- [18] MAHDI A. Atomic orbital search: A novel metaheuristic algorithm[J]. Applied Mathematical Modelling, 2021, 93: 657-683.
- [19] BAKDI A, GUICHI A, MEKHILEF S, et al. GPVS-faults: Experimental data for fault scenarios in grid-connected PV systems under MPPT and IPPT modes[EB/OL]. [2023-06-01]. <https://doi.org/10.17632/n76t439f65.1>.
- (本文责编: 刘芳)
-
- 收稿日期: 2023-06-16; 修回日期: 2023-07-19
 上网日期: 2023-08-18; 附录网址: www.ienergy.cn
- 作者简介:
 杨晓燕(1985), 女, 高级工程师, 从事营销管理和数字化研究, yangxy0041@nng.gx.csg.cn;
 谢满承(1986), 男, 工程师, 从事营销管理和计量线损研究, xiemc7738@nng.gx.csg.cn;
 郭小璇(1986), 女, 高级工程师, 博士, 从事综合能源与需求侧管理技术研究, guo_xiaoxuan@163.com;
 廖卓颖*(1998), 女, 在读博士研究生, 从事分布式发电及微电网技术研究, liaozy@ms.giec.ac.cn.
 *为通信作者。