

引用格式:杨帅,曾文伟,杨凌云,等.基于GOA-SVM的光伏阵列故障诊断方法研究[J].电力科学与技术学报,2024,39(5):172-180.

Citation: YANG Shuai, ZENG Wenwei, YANG Lingyun, et al. Research on fault diagnosis method for photovoltaic array based on GOA-SVM[J]. Journal of Electric Power Science and Technology, 2024, 39(5): 172-180.

基于GOA-SVM的光伏阵列故障诊断方法研究

杨 帅^{1,2}, 曾文伟^{1,2}, 杨凌云³, 黄 瑞^{1,2}, 刘谋海^{1,2}, 易钦逸⁴, 高云鹏⁴

(1. 国网湖南省电力有限公司供电服务中心, 湖南长沙 410004; 2. 智能电气量测与应用技术湖南省重点实验室, 湖南长沙 410004;
3. 国网湖南省电力有限公司, 湖南长沙 410004; 4. 湖南大学电气与信息工程学院, 湖南长沙 410082)

摘 要: 光伏阵列输出功率随机性、波动性强。如果其发生故障, 将严重影响电力系统安全与稳定。针对当前光伏故障诊断的准确率低和收敛速度慢的难题, 提出一种基于蝗虫算法-支持向量机 (grasshopper optimization algorithm-support vector machine, GOA-SVM) 模型的光伏阵列故障诊断方法。首先, 建立光伏阵列等效电路模型, 分析光伏阵列的伏安曲线变化特性; 其次, 考虑环境影响因素和光伏阵列规模非线性变化, 提取反映不同故障特性的特征量, 将数据映射到高维空间进行非线性处理; 最后, 提出蝗虫算法 (grasshopper optimization algorithm, GOA) 优化非线性支持向量机改进方法, 建立 GOA-SVM 光伏阵列故障诊断模型, 并结合实例进行仿真。研究结果表明: 该方法可应用于多种不同规模的光伏阵列模型, 且均能实现对光伏阵列故障的有效诊断, 其对 4×3 光伏阵列规模的数据仿真分类准确率可达 99.808 8%。采用美国国家标准与技术研究院 (National Institute of Standards and Technology, NIST) 公开数据集进行验证, 其故障诊断准确率达到 92.368 2%。与其他方法相比, 该方法的召回率及 F1-Score 均有明显提升。

关 键 词: 光伏阵列; 故障诊断; 多分类; 最优超平面; GOA-SVM 模型

DOI: 10.19781/j.issn.1673-9140.2024.05.018 **中图分类号:** TM615 **文章编号:** 1673-9140(2024)05-0172-09

Research on fault diagnosis method for photovoltaic array based on GOA-SVM

YANG Shuai^{1,2}, ZENG Wenwei^{1,2}, YANG Lingyun³, HUANG Rui^{1,2},
LIU Mouhai^{1,2}, YI Qinyi⁴, GAO Yunpeng⁴

(1. Power Supply Service Center, State Grid Hunan Electric Power Co., Ltd., Changsha 410004, China; 2. Hunan Province Key Laboratory of Intelligent Electrical Measurement and Application Technology, Changsha 410004, China; 3. State Grid Hunan Electric Power Co., Ltd., Changsha 410004, China; 4. College of Electrical and Information Engineering, Hunan University, Changsha 410082, China)

Abstract: The output power of photovoltaic (PV) arrays exhibits strong randomness and volatility. In the event of a fault, it can severely impact the safety and stable operation of the power system. Addressing the challenges of low accuracy and slow convergence in current PV fault diagnosis, this paper proposes a PV array fault diagnosis method based on the grasshopper optimization algorithm-support vector machine (GOA-SVM) model. Firstly, an equivalent circuit model of the PV array is established to analyze the variation characteristics of the PV array's voltage-current curve. Secondly, considering environmental factors and the nonlinear changes in the scale of the PV array, feature quantities reflecting different fault characteristics are extracted, and the data is mapped into a high-dimensional space for nonlinear processing. Finally, an improved method for optimizing the nonlinear support vector machine using GOA is proposed, and a GOA-SVM PV array fault diagnosis model is established, with simulations conducted using practical examples. The research results indicate that this method can be applied to various PV array models of different scales and effectively diagnose faults in PV arrays. For a 4×3 PV array scale, the data simulation classification accuracy can reach 99.8088%. When validated using the publicly available dataset from the national institute of standards and technology

收稿日期: 2023-10-21; 修回日期: 2023-12-11

基金项目: 国家重点研发计划 (2021YFF0602402)

通信作者: 高云鹏 (1978—), 男, 博士, 教授, 主要电力设备故障诊断、智能信息处理等方面的研究; E-mail: gaoyunpeng@hnu.edu.cn

(NIST), the fault diagnosis accuracy achieves 92.3682%. Compared with other methods, this approach demonstrates significant improvements in recall rate and F1-Score.

Key words: photovoltaic array; fault diagnosis; multi-classification; optimal hyperplane; GOA-SVM mode

光伏阵列是光伏发电系统的重要组成部分,由于其长期工作在较为恶劣的环境中,极易引发阵列的开路、短路及逆变器过流、过压、功率管开路等故障,严重影响光伏系统的发电效率^[1-3]。为避免光伏阵列故障导致严重事故,提升电力系统的发电效率,国内外研究机构已展开光伏系统故障诊断技术方面的研究。

近年来,许多学者陆续提出各种先进的光伏故障诊断方法^[4],其大致可分为物理检测法、功率差值法、I-V曲线法、人工智能法等。文献[4]通过测量光照辐射度和环境温度,估计光伏阵列的理论电压、电流以与功率。如果理论功率和实际功率相差超过15%,则判定阵列存在故障。但该法需设置故障阈值,由于光伏阵列长期在户外运行,组件会老化,故其设定的阈值也需定期更新。文献[5]通过无人机拍摄光伏阵列的热成像图,判断热成像图中是否存在显著的亮点来实现光伏阵列的故障诊断,但该方法受天气和复杂环境的影响较大,且红外热成像法需使用较为昂贵的设备,故其故障诊断的精度欠佳,易受外界环境的影响。文献[6-7]通过测量光伏阵列的I-U曲线,从曲线中提取特征,来训练模糊C均值聚类模型,通过训练好的模型来计算待测试样本对各特征隶属度的大小,以实现故障分类,但其数据获取难度较高。文献[8]提出了基于图的半监督学习模型的光伏阵列故障诊断方法,但该模型的输入处理难度较大。人工智能法仅需电气测量数据,且诊断准确率高。其中,支持向量机(supporting vector machine, SVM)模型用于故障诊断,能获得较高准确率。文献[9]采用基于多类SVM的光伏组件故障在线诊断方法,但SVM算法的部分参数的设置对整体诊断精度影响较大。有学者通过遗传算法(genetic algorithm, GA)^[10]、海鸥算法(seagull optimization algorithm, SOA)^[11-12]、人工蜂群算法(artificial bee colony algorithm, ABC)^[13-14]、粒子群算法(particle swarm optimization, PSO)^[15]、灰狼算法(grey wolf optimizer, GWO)^[16]、布谷鸟算法(cuckoo search, CS)^[17]等来优化SVM的相关参数,提高其识别光伏阵列故障类型的准确率,但部分改进算法仍存在收敛速度

慢、容易陷入局部最优等问题^[18-19]。

针对这些问题,本文提出一种基于蝗虫算法—支持向量机(grasshopper optimization algorithm—support vector machine, GOA-SVM)模型的光伏阵列故障诊断方法。先通过建立3种不同规模的光伏阵列等效电路模型,考虑不同程度的短路和开路的光伏阵列故障,分析光伏阵列的伏安曲线变化特性;再考虑环境影响因素和光伏阵列规模非线性变化,构建反映不同故障特性的特征量,并将其映射到高维空间中,将其作为光伏阵列故障诊断方法的输入向量;然后,引入蝗虫优化算法(grasshopper optimization algorithm, GOA),优化非线性SVM,构建基于GOA-SVM的光伏阵列故障诊断模型,解决传统故障诊断方法实现复杂、收敛速度慢的问题;最后,并采用仿真数据进行算例分析,对3种不同规模的光伏阵列模型进行分析,并使用美国国家标准与技术研究院(National Institute of Standards and Technology, NIST)公开数据集进行验证,将所提方法与传统算法进行对比,验证了所提方法的准确性和有效性。

1 光伏阵列故障分析

对光伏阵列中常见的短路、开路这两类故障,本文考虑3×4的光伏阵列,分析其在这两种故障类型下光伏阵列伏安特性曲线的变化情况,并基于伏安特性曲线选取故障诊断的特征量。通过仿真,得到不同工作状态下光伏阵列的I-U曲线和P-U曲线,结果分别如图1、2所示。

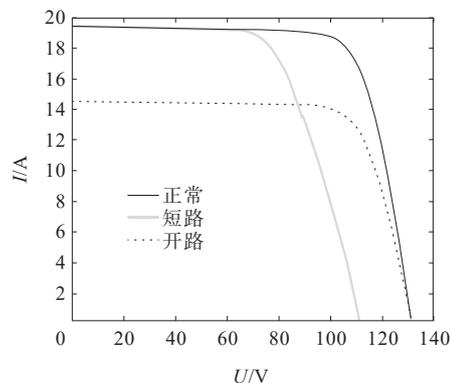


图1 正常状态与不同故障状态下的I-U曲线
Figure 1 I-U curves under normal condition and various fault conditions

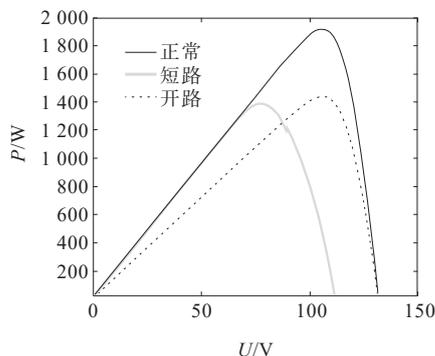


图2 正常状态与不同故障状态下的P-U曲线

Figure 2 P-U curves under normal condition and various fault conditions

1) 短路故障。

从图1、2中可以看出,当光伏阵列发生短路故障时,短路电流 I_{sc} 变化不大,开路电压 U_{oc} 变化很大,且其最大功率点的变化最为显著,因此,选取最大功率点电流 I_m 、电压 U_m 和开路电压 U_{oc} 作为其故障特征。

2) 开路故障。

当光伏阵列发生开路故障时,短路电流 I_{sc} 变化很大,开路电压 U_{oc} 变化不大,且最大功率点的变化显著,因此,选取最大功率点电流 I_m 、电压 U_m 和短路电流 I_{sc} 作为其故障特征。

从图1、2还可以看出,相较于正常运行状态,当发生短路故障时,光伏阵列的开路电压明显减小;当发生开路故障时,短路电流明显减小。仿真试验过程中收集到的典型样本数据见表1。

表1 光伏阵列仿真样本

Table 1 Simulation sample of PV array

状态类别	行数	列数	U_{mpp}/V	I_{mpp}/A
正常	9	10	385.22	58.98
	4	3	174.77	15.91
	1	1	42.80	5.84
开路	9	10	373.95	58.86
	4	3	171.02	10.63
	1	1	25.53	0.017
短路	9	10	385.22	53.08
	4	3	135.75	15.95
	1	1	15.38	3.10

综上所述,当光伏阵列出现开路、短路故障时,其开路电压、短路电流、最大功率点和最大功率点的电压与电流中至少一种参数会发生明显变化。通过分析3种不同规模的光伏阵列,模型可具有更

好的泛化性。因此,本文选择这些参数作为模型的输入特征。

2 基于GOA-SVM光伏阵列故障诊断

2.1 GOA优化SVM模型

SVM分析数据的监督式学习模型与相关的二分类模型学习算法是识别光伏阵列故障的主要方法。该方法是结构风险最小化的近似实现,SVM能较好地解决小样本、非线性分类以及高维模式识别等问题,具有计算简单、鲁棒性好和泛化性能佳等优点。因此,SVM被广泛应用于回归预测、分类识别、模式识别、函数估计等领域。

在光伏阵列故障识别过程中,假设光伏阵列样本的训练集为 $\{(x_i, y_i), i=1, 2, \dots, n\}$, $x_i \in \mathbb{R}^n$, $y_i \in \{1, 2, 3\}$, x_i 为光伏阵列特征输入向量,包括光伏阵列的行数和列数、光照强度、温度、最大功率点电压和电流等特征分量, y 为类型标签,其分为正常状态、短路故障、开路故障3种状态。设存在分类平面 $d(x) = \omega x + b = 0$, $\omega \in \mathbb{R}^n$ 为权重向量, $b \in \mathbb{R}$ 为偏置,该分类平面能将所有光伏阵列状态进行准确分类,且使分类间隔最大化,则定义分类平面 $d(x) = \omega x + b = 0$ 为最优分类超平面,定义两类样本中距离该分类线最近的向量,且平行于最优分类面的样本为支持向量。二分类支持向量机的原理图如图3所示。在图3中, H 为最优超平面, D_1 和 D_2 为两类不同的样本。在样本空间或者特征空间内构造最优超平面,最大分离这两类样本。为保证模型的泛化能力最大,应使得平面与样本集之间距离最大。

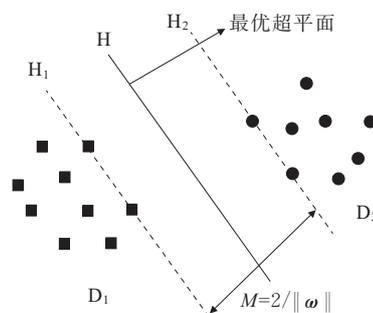


图3 支持向量机原理示意

Figure 3 Schematic diagram of support vector machine

求取分类间隔为最大值的问题等价于求取 $\|\omega\|$ 的最小值问题,即当 $\|\omega\|$ 最小时,该最优超平面(在二维空间中为直线)能正确地将两类样本分开,且两类样本之间的间隙最大,故求解最优分类线的问题可转化为优化问题,即

$$\begin{cases} \min_{\omega, b} \frac{1}{2} \|\omega\|^2 = \min_{\omega, b} \frac{1}{2} \omega^T \omega \\ \text{s.t. } y_i(\omega x_i + b) \geq 1 \end{cases} \quad (1)$$

式中, $i = 1, 2, \dots, n$ 。

考在实际光伏阵列中,温度和光照强度的影响很大,很多光伏阵列数据样本是线性不可分的。因此,引入 Lagrang 乘子 α_i 与惩罚因子 C ,将式(1)转化为对偶优化问题,即

$$\begin{cases} \max_{\alpha} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \alpha_i y_i y_j (x_i \cdot x_j) \\ \text{s.t. } \begin{cases} \sum_{i=1}^n y_i \alpha_i = 0 \\ 0 \leq \alpha_i \leq C, i = 1, 2, \dots, n \end{cases} \end{cases} \quad (2)$$

式中, y_i 为第 i 个训练样本的类别标签,表示样本的分类。

通过非线性变换将(2)转化为某个维特征空间中的线性分类问题,即在高维特征空间中学习线性支持向量机,用核函数 $K(x, x_i)$ 替换式(2)中的内积,求解得到的解即为非线性支持向量机,即

$$f(x) = \text{sgn}[(\omega \cdot x) + b] = \text{sgn}\left[\sum_{i=1}^n y_i \alpha_i K(x \cdot x_i) + b\right] \quad (3)$$

由此可知,SVM 在解决光伏阵列故障诊断问题时极易受核函数与惩罚因子参数的影响,本文用 GOA 对其参数进行寻优。

GOA 是由 Saremi 等^[20]提出的一种新型的全局优化元启发式算法。该算法受成年与幼年蝗虫的大规模移动与觅食的聚集行为启发,具有收敛速度快、搜寻效率高和操作参数少等特点。该算法通过模拟自然界中蝗虫种群行为来解决优化问题,能很好地平衡全局和局部搜索过程,寻优精度高。

在 GOA-SVM 的过程中,蝗虫群位置即为 SVM 中的 C 和 g ,其更新计算式为

$$X_i^d = \beta \left[\sum_{j=1, j \neq i}^N \beta \frac{u_d - l_d}{2} s(|x_i^d - x_j^d|) \frac{x_i^d - x_j^d}{d_{ij}} \right] + T_d \quad (4)$$

式中, X_i^d 为多维空间中第 i 个数据点在维度 d 上的位置; d 为 2;蝗虫群位置为 SVM 中惩罚因子 C 和核函数参数 g 的取值; i 和 j 分别为蝗虫个体编号,每个编号代表一个蝗虫在多维空间中的位置; u_d 和 l_d 分别为变量的上界和下界,本文设置其分别为 1 000 和 0.001; T_d 为最优蝗虫的位置,即最优的惩罚因子

和最优核函数参数 g 的组合; d_{ij} 为两只蝗虫间的欧式距离; β 为控制参数,用于平衡算法的全局探索和局部开发;函数 s 为蝗虫个体间的相互作用力。

控制参数 β 的表达式为

$$\beta = \beta_{\max} - t \frac{\beta_{\max} - \beta_{\min}}{T_{\max}} \quad (5)$$

式中, β_{\max} 与 β_{\min} 分别表示递减区间的最大值与最小值; t 表示当前的迭代次数; T_{\max} 表示最大迭代次数。

则蝗虫个体间的相互作用力 $s(r)$ 为

$$s(r) = fe^{-r/l} - e^{-r} \quad (6)$$

式中, r 为蝗虫个体在种群中的影响力因子; f 、 l 分别为吸引强度参数与吸引尺度参数,本文中分别将其设置为 0.5 和 1.5。

本模型选择适应度函数为测试集准确率 f_{fitness} ,其计算式为

$$f_{\text{fitness}} = \frac{N_{\text{test, right}}}{N_{\text{test, total}}} \quad (7)$$

式中, $N_{\text{test, right}}$ 为光伏阵列数据样本测试集正确诊断个数; $N_{\text{test, total}}$ 为光伏阵列数据样本测试集总数。

2.2 基于 GOA-SVM 的光伏阵列故障诊断流程

基于 GOA-SVM 的光伏阵列故障诊断方法的流程如图 4 所示。该算法的具体步骤如下。

- 1) 输入光伏阵列样本数据及其对应的标签数据,并设置训练集和测试集;
- 2) 初始化 GOA 参数,即种群规模 N 、最大迭代次数 T_{\max} 、参数 β_{\max} 和 β_{\min} 、SVM 参数 C 和 g 的取值范围;
- 3) 初始化蝗虫群位置,每只蝗虫的位置都由对应的 (C, g) 组成;
- 4) 利用 SVM 对训练集进行训练,根据式(7)计算出每只蝗虫所对应的适应度值;

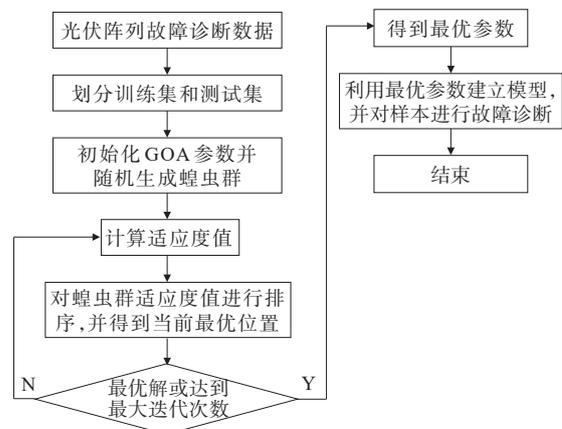


图 4 GOA-SVM 流程

Figure 4 Flowchart of GOA-SVM

5) 根据每只蝗虫的适应度值大小进行排序,得到当前最优蝗虫位置;

6) 根据式(4)~(7)对蝗虫的适应度值和最优位置,不断进行更新计算和重新排序,并检查是否越界;

7) 当种群迭代次数达到最大迭代次数时,输出蝗虫的最优位置所对应的 C 和 g 参数的最优解,否则跳转至步骤6;

8) 利用最优参数建立光伏阵列故障诊断模型,并对样本进行故障诊断。

3 实验验证

3.1 光伏阵列仿真模型

模型输入特征的选择决定能否迅速、精准地对光伏阵列故障类型进行判断。因此,输入特征须直接体现出光伏阵列故障机理特征,易于算法实现。

本文共设置3种不同的光伏阵列工作状态,即正常状态、组件开路故障状态、组件短路故障状态。开路故障通过断开故障光伏阵列与其相邻光伏阵列的导线来模拟;短路故障通过将故障光伏阵列的正负极短接来模拟。本文使用MATLAB数值仿真软件分别搭建 1×1 、 4×3 、 9×10 的串并联光伏阵列。其中, 4×3 的光伏阵列仿真如图5所示, 9×10 的光伏阵列仿真如图6所示。

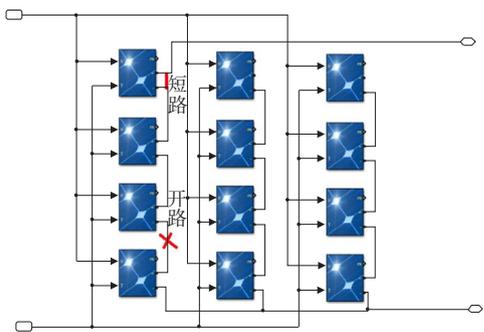


图5 4×3 的光伏阵列仿真

Figure 5 Simulation diagram of a 4×3 photovoltaic array

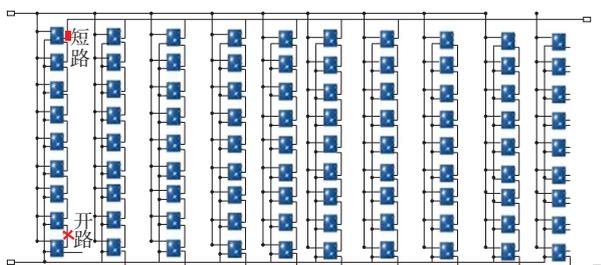


图6 9×10 的光伏阵列仿真

Figure 6 Simulation diagram of a 9×10 photovoltaic array

3.2 仿真实证

对 4×3 光伏阵列规模进行分析,搭建串、并联光伏阵列,得到8 716组数据。其中,包含故障状态的样本数为6 829组,正常状态数据1 887组。在故障状态数据中,开路故障状态数据2 133组,短路故障数据4 696组。将样本数据以7:3的比例随机划分为训练集和测试集。影响SVM模型故障诊断精度的参数分别为惩罚因子 C 和核函数 g ,将未进行优化的SVM模型中惩罚因子 C 和核函数 g 分别设置为1.00和0.59,经诊断后该模型正确诊断样本2 392组,准确率仅为91.472 3%。其中,SVM模型的故障诊断结果和混淆矩阵分别如图7、8所示。

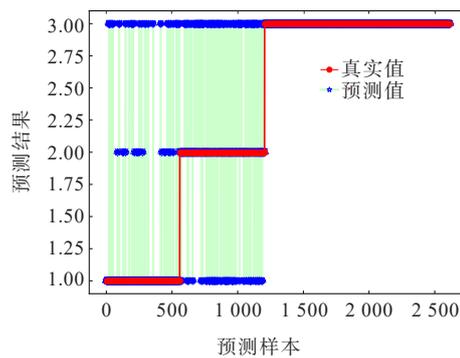


图7 未优化的SVM故障诊断

Figure 7 Unoptimized SVM fault diagnosis diagram

真实分类	I	511	16	32
	II	61	471	114
	III	0	0	1 410
		I	II	III
		预测分类		

图8 未优化的SVM混淆矩阵

Figure 8 Unoptimized SVM confusion matrix diagram

从图7、8可看出,未优化的SVM模型在光伏阵列故障诊断方面的效果欠佳,对数据的分类效果一般,有待进一步优化。使用GOA算法对SVM模型中的惩罚因子 C 和核函数 g 进行参数优化。GOA初始化参数设置为:种群数量为20,最大迭代次数为20,交叉验证折数为5,惩罚因子 C 的优化范围为 $[1 \times 10^{-5}, 1\ 000]$,核函数 g 的优化范围为 $[1 \times 10^{-5}, 1\ 000]$ 。经过迭代计算后,多种优化算法适应度曲线对比图、GOA-SVM模型的故障诊断结果与混淆矩阵分别如图9~11所示。

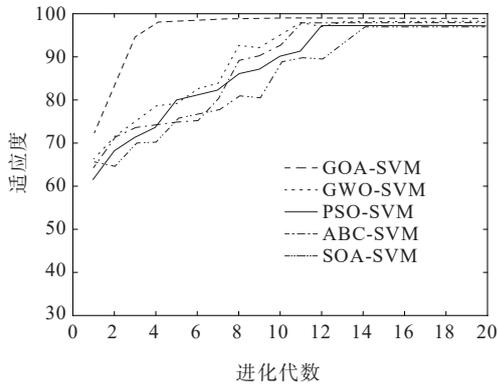


图 9 多种优化算法适应度曲线对比

Figure 9 Comparison of fitness curves of various optimization algorithms

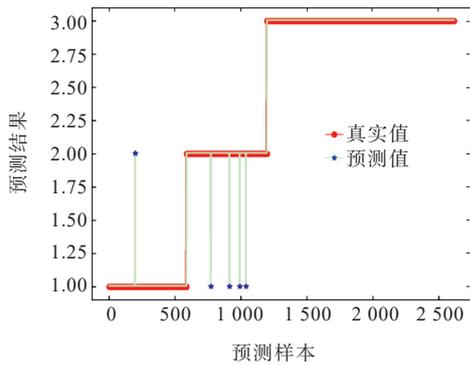


图 10 GOA 优化 SVM 故障诊断

Figure 10 SVM fault diagnosis optimized by GOA

真实分类	I	585	1	0
	II	4	606	0
	III	0	0	1 419
		I	II	III
		预测分类		

图 11 GOA 优化 SVM 混淆矩阵

Figure 11 SVM confusion matrix optimized by GOA

从图 9 可看出,GWO-SVM 算法迭代到第 11 次达到收敛,PSO-SVM 算法和 ABC-SVM 算法迭代到第 12 次达到收敛,SOA-SVM 算法迭代到第 14 次达到收敛,而 GOA 迭代到第 7 次时就开始收敛,收敛速度最快。此时 GOA 最佳适应度为 99.808 8,惩罚因子 C 为 1 000,核函数 g 为 2.498。从图 10、11 可看出,GAO-SVM 模型对测试集的故障诊断准确率为 99.808 8%。

从光伏阵列故障诊断准确率分析,GOA 对

SVM 模型参数优化使得准确率得到较大提升,相比于未优化的 SVM 模型提高了 8.3365%。该仿真结果表明:在光伏阵列故障诊断时,GOA-SVM 模型比未优化的 SVM 模型更可靠。

为验证所提 GOA-SVM 模型能够准确地诊断光伏阵列的故障,通过对比分析 SVM、KNN 等机器学习模型,以及基于灰狼优化 SVM 模型和基于粒子群算法优化 SVM 模型,采用相同训练集和测试集进行实验,SVM 模型、GWO-SVM 模型、PSO-SVM 模型和 GOA-SVM 等模型的故障诊断结果见表 2 和如图 12 所示。在表 2 中,除之前提及算法外,还采用了 k 邻近算法(k-nearest neighbor, kNN)、梯度提升决策树(gradient boosting decision tree,GBDT)等算法。

表 2 不同算法故障诊断结果

Table 2 Fault diagnosis results of different algorithms

算法	误诊数	准确率/%
分类树	72	97.247 2
KNN	550	78.967 2
决策树	173	93.449 7
随机森林	158	93.944 1
GBDT	65	97.492 5
SVM	257	91.472 7
朴素贝叶斯	195	92.533 8
SOA-SVM	66	97.463 2
ABC-SVM	51	98.043 1
GWO-SVM	48	98.131 6
PSO-SVM	53	97.952 2
GOA-SVM	3	99.808 8

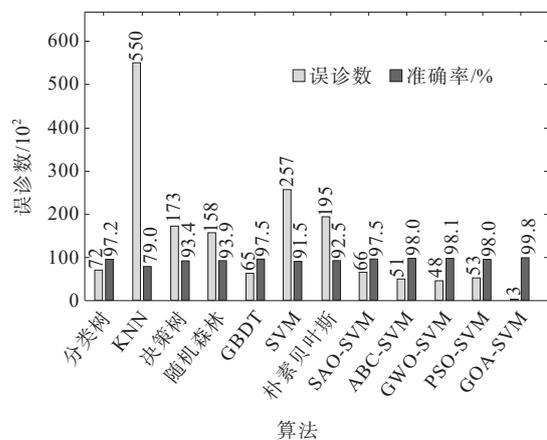


图 12 算法效果比较

Figure 12 Effects comparison of various algorithms

分析表2和图12可知,在分类树、KNN、决策树、随机森林、GBDT、SVM、朴素贝叶斯机器学习模型,GBDT准确率最高,为97.4925%,在SVM其他优化算法中,SOA、ABC、GWO、PSO优化SVM模型准确率分别为97.4632%、98.0431%、98.1316%、97.9522%,而本文提出的GOA-SVM算法准确率可达99.8088%,诊断结果明显优于其他算法。对3种不同光伏阵列规模的仿真数据进行分析,其中1×1光伏阵列有5100组数据,包含正常状态数据1700组,开路状态数据1700组,短路数据1700组;其中10×9光伏阵列有324组数据,包含正常状态数据108组,开路状态数据108组,短路数据108组;将样本数据以7:3的比例随机划分训练集和测试集,结果见表3。

表3 不同光伏阵列规模故障诊断结果

Table 3 Fault diagnosis results of different PV array sizes

阵列规模	测试数据 个数	准确率/ %	精确率/ %	召回率/ %	F1
9×10	97	0.9923	0.9781	0.9640	0.9710
4×3	2615	0.9981	0.9983	0.9972	0.9977
1×1	1530	1.0000	1.0000	1.0000	1.0000

由表3可知,GOA-SVM算法对3种不同光伏阵列规模进行故障诊断时,准确率高、鲁棒性强,其中4×3的光伏阵列准确率为99.81%,F1值为99.77%;1×1的光伏阵列准确率为100%,F1值为100%;9×10的光伏阵列准确率为99.23%,F1值为97.10%。

3.3 实验验证

为了进一步验证本文方法的准确性和有效性,将提出方法针对NIST公开数据集进行分析,共取564522条数据,数据特征包括:光照强度、温度、直流电压、直流电流、直流功率、交流功率,数据标签包括:正常、短路和开路。将样本数据以7:3的比例随机划分训练集和测试集。将本文所提算法应用在该公开数据集上,其适应度曲线、故障诊断结果及混淆矩阵分别如图13~15所示。

从图13可看出,GWO-SVM算法迭代到第12次达到收敛,PSO-SVM算法迭代到第13次达到收敛,而GOA迭代到第10次时开始收敛,收敛速度最快。此时GOA最佳适应度为92.3682,惩罚因子C为1000和核函数g为77.5107。从图14、15中可看出,GOA优化SVM模型对测试集的故障诊断准确率为92.3682%。

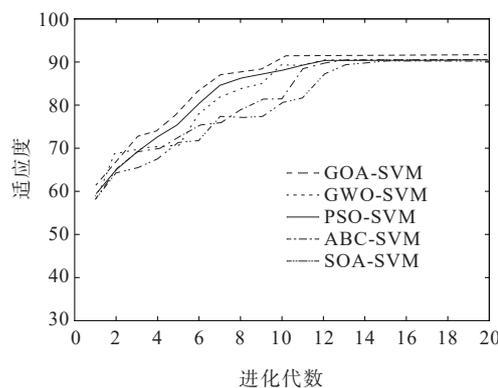


图13 多种优化算法适应度曲线对比

Figure 13 Comparison of fitness curves of various optimization algorithms

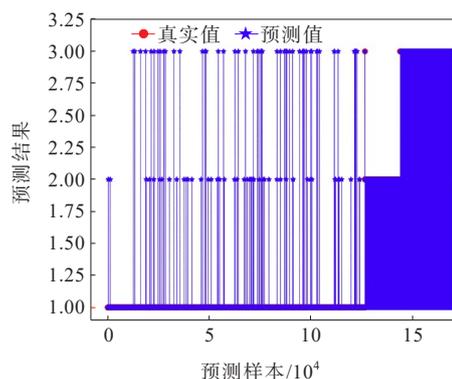


图14 GOA优化SVM故障诊断

Figure 14 SVM fault diagnosis optimized by GOA

真实分类	I	123 474	1 254	1 654
	II	3 678	11 544	2 120
	III	2 537	1 682	21 413
		I	II	III

图15 GOA优化SVM混淆矩阵

Figure 15 SVM confusion matrix optimized by GOA

对比分析SVM、KNN等机器学习模型,基于GWO优化的SVM模型和基于PSO优化的SVM模型,采用相同训练集和测试集进行实验,SVM模型、GWO-SVM模型、PSO-SVM模型和本文GOA-SVM等模型的故障诊断结果如表4和图16所示。

由表4和图16可知,在分类树、KNN、决策树、随机森林、GBDT、SVM、朴素贝叶斯机器学习模型中,分类树准确率最高,为89.3744%。在SVM的

优化算中,SOA、ABC、GWO、PSO 优化 SVM 模型准确率分别为 90.113 2%、90.469 2%、90.895 1%、90.265 8%,而本文的 GOA-SVM 准确率能达到 92.368 2%,由此可知,本文提出算法应用于实际工况时更优于现有其他算法的。

表 4 不同算法故障诊断结果

Table 4 Fault diagnosis results of different algorithms

算法	误诊数	准确率/%
分类树	17 995	89.374 4
KNN	33 048	80.486 3
决策树	28 680	83.065 2
随机森林	25 427	84.986 1
GBDT	21 197	87.483 5
SVM	21 316	87.413 7
朴素贝叶斯	27 861	83.548 6
SOA-SVM	16 744	90.113 2
ABC-SVM	16 140	90.469 2
GWO-SVM	15 419	90.895 1
PSO-SVM	16 485	90.265 8
GOA-SVM	12 924	92.368 2

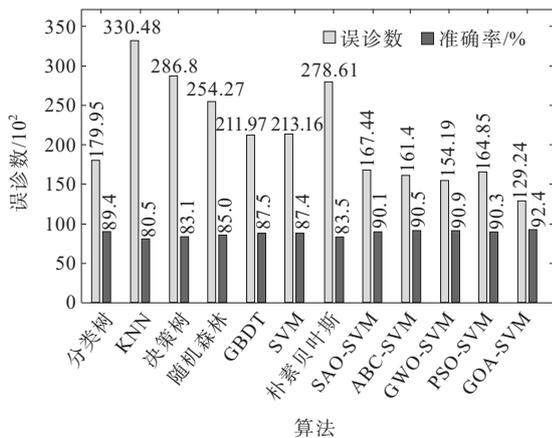


图 16 公开数据算法效果比较

Figure 16 Comparison of algorithm performance on public data

4 结语

本文提出了一种基于 GOA-SVM 算法的光伏阵列故障诊断方法,该方法可有效光伏阵列辨识短路、断路故障状态。仿真与实验结果均表明:对正常工作、短路和断路 3 种状态,本文提出的 GOA-SVM 算法的仿真数据故障诊断准确率达到 99.808 8%;使用 NIST 公开数据集进行验证,其故障诊断准确率达到 92.368 2%,且只需用较少迭代次数即可达

到全局最优解,收敛速度优势明显。本文提出方法针对不同规模的光伏阵列准确率高、鲁棒性强,在同等样本规模条件下该算法具有较高的准确率及更快的收敛性,可有效实现对光伏阵列的故障诊断与判别分析。

参考文献:

- [1] 张长庚.考虑含光伏接入的配电网故障特性研究[J].电网与清洁能源,2022,38(12):138-146.
ZHANG Changgeng.Research on fault characteristics of the distribution network considering access of photovoltaic[J].Power System and Clean Energy,2022,38(12):138-146.
- [2] 陆江,袁廷璧,彭咏龙,等.基于统计推断的异常光伏组件识别方法[J].智慧电力,2023,51(11):38-44.
LU Jiang, YUAN Tingbi, PENG Yonglong, et al. Anomalous photovoltaic module identification method based on statistical inference[J]. Smart Power, 2023, 51(11):38-44.
- [3] 李景歌,荣娜,陈庆超.基于生成对抗网络的分布式光伏窃电数据增强方法[J].电力科学与技术学报,2022,37(5):181-190.
LI Jingge, RONG Na, CHEN Qingchao. A data augmentation method for distributed photovoltaic electricity theft using generative adversarial network[J]. Journal of Electric Power Science and Technology,2022, 37(5):181-190.
- [4] 艾上美,周剑峰,张必朝,等.基于改进 SSD 算法的光伏组件缺陷检测研究[J].智慧电力,2023,51(12):53-58.
AI Shangmei, ZHOU Jianfeng, ZHANG Bichao, et al. Defect detection of photovoltaic modules based on improved SSD algorithm[J]. Smart Power, 2023, 51(12): 53-58.
- [5] 陈宇航,闫腾飞,谢添,等.基于功率损失和 $U-I$ 特性综合考虑的光伏组件故障诊断方法[J].电机与控制应用, 2016,43(11):92-97.
CHEN Yuhang, YAN Tengfei, XIE Tian, et al. A novel fault diagnosis method of photo voltaic module based on power loss and U-I characteristics[J].Electric Machines & Control Application,2016,43(11):92-97.
- [6] 毕锐,丁明,徐志成,等.基于模糊 C 均值聚类的光伏阵列故障诊断方法[J].太阳能学报,2016,37(3):730-736.
BI Rui, DING Ming, XU Zhicheng, et al. PV array fault diagnosis based on FCM[J].Acta Energiae Solaris Sinica, 2016,37(3):730-736.
- [7] 何延一,李胜男,陈亦平,等.基于最优光伏阵列重构的电网调频策略研究[J].电力系统保护与控制,2022,50(1):124-132.
HE Tingyi, LI Shengnan, CHEN Yiping, et al. Optimal PV array reconfiguration-based power grid frequency

- regulation strategy[J]. *Power System Protection and Control*,2022,50(1):124-132.
- [8] ZHAO Y, BALL R, MOSESIAN J, et al. Graph-based semi-supervised learning for fault detection and classification in solar photovoltaic arrays[J]. *IEEE Transactions on Power Electronics*, 2015, 30(5): 2848-2858.
- [9] WANG L X, LIU J, GUO X G, et al. Online fault diagnosis of photovoltaic modules based on multi-class support vector machine[C]//2017 Chinese Automation Congress (CAC). Jinan, China. IEEE, 2017:4569-4574.
- [10] 郭浩然,李泽滔. 遗传算法优化支持向量机的光伏阵列故障诊断研究[J]. *智能计算机与应用*,2019,9(5):58-62.
GUO Haoran, LI Zetao. Research on fault diagnosis of photovoltaic array based on Support Vector Machine optimized by Genetic Algorithm[J]. *Intelligent Computer and Applications*,2019,9(5):58-62.
- [11] 孙培胜,陈堂贤,程陈,等. 基于 SOA-SVM 模型的光伏阵列故障诊断研究[J/OL]. *电源学报*:1-13[2023-10-14]. <http://kns.cnki.net/kcms/detail/12.1420.TM.20220823.1109.002.html>.
SUN Peisheng, CHEN Tangxian, CHENG Chen et al. Research on fault diagnosis of photovoltaic array based on soa-svm model [J/OL]. *Journal of Power Sources*:1-13 [2023-10-14]. <http://kns.cnki.net/kcms/detail/12.1420.TM.20220823.1109.002.html>.
- [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.
- [13] 刘开石,李田泽,刘东,等. 基于 ABC-SVM 算法的光伏阵列故障诊断[J]. *电源技术*,2021,45(9):1171-1174.
LIU Kaishi, LI Tianze, LIU Dong, et al. Fault diagnosis of PV array based on ABC-SVM algorithm[J]. *Chinese Journal of Power Sources*,2021,45(9):1171-1174.
- [14] 董广凯,黄嵩,董佳林,等. 定功率跟踪控制的光伏并网低电压穿越策略[J]. *高压电器*,2023,59(3):93-98+107.
DONG Guangkai, HUANG Song, DONG Jialin, et al. Low voltage ride-through strategy of grid-connected photovoltaic generation with constant power tracking control[J]. *High Voltage Apparatus*, 2023, 59(3): 93-98+107.
- [15] 林培杰,陈志聪,吴丽君,等. 一种 PSO-SVM 的光伏阵列故障检测与分类[J]. *福州大学学报(自然科学版)*,2017,45(5):652-658.
LIN Peijie, CHEN Zhicong, WU Lijun, et al. Fault detection and classification for photovoltaic arrays based on PSO-SVM[J]. *Journal of Fuzhou University (Natural Science Edition)*,2017,45(5):652-658.
- [16] 程陈,陈堂贤,孙培胜,等. 基于灰狼算法的光伏组件故障诊断模型优化[J]. *电源技术*,2022,46(6):684-687.
CHENG Chen, CHEN Tangxian, SUN Peisheng, et al. Optimization of fault diagnosis model for photovoltaic module based on gray wolf algorithm[J]. *Chinese Journal of Power Sources*,2022,46(6):684-687.
- [17] 张杰,易辉,张霞,等. 基于布谷鸟算法的光伏组件故障诊断模型优化[J]. *电源技术*,2020,44(1):76-79.
ZHANG Jie, YI Hui, ZHANG Xia, et al. Optimization of fault diagnosis model for photovoltaic module based on cuckoo algorithm[J]. *Chinese Journal of Power Sources*, 2020,44(1):76-79.
- [18] 蔡正梓,程海兴,陈茜,等. 基于 PSO-SVM 的变电站视频监控火灾识别算法[J]. *自动化与仪表*,2022,37(7):58-62+67.
CAI Zhengzi, CHENG Haixing, CHEN Qian, et al. Substation video surveillance fire recognition algorithm based on PSO-SVM[J]. *Automation & Instrumentation*, 2022,37(7):58-62+67.
- [19] WANG J J, GAO D D, ZHU S K, et al. Fault diagnosis method of photovoltaic array based on support vector machine[J]. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*,2023,45(2):5380-5395.
- [20] SAREMI S, MIRJALILI S, LEWIS A. Grasshopper optimisation algorithm: theory and application[J]. *Advances in Engineering Software*,2017,105:30-47.