

DOI: 10.13382/j.jemi.B2104384

基于 RFRFE 与 ISSA-XGBoost 的变压器故障辨识方法^{*}

王雨虹 王志中

(辽宁工程技术大学电气与控制工程学院 葫芦岛 125105)

摘 要:针对变压器故障诊断精度低的问题,提出了随机森林-递归特征消除(RFRFE)算法与改进麻雀算法(ISSA)优化极端梯度提升树(XGBoost)的变压器故障诊断方法。首先以诊断精度为标准,利用 RFRFE 算法选择重要特征变量,去除冗余特征;然后采用服从均匀分布随机调整策略和莱维飞行策略来对传统麻雀算法(SSA)进行改进,并将 ISSA 与 SSA 和粒子群算法(PSO)进行算法性能测试,证明其分类精度和网络寻优能力均有所提升;最后使用改进的麻雀算法对 XGBoost 相关超参数进行寻优,获取 RFRFE 与 ISSA-XGBoost 相结合的综合故障诊断模型,并与 PSO-XGBoost 和 SSA-XGBoost 故障诊断模型对比诊断效果,结果表明 ISSA-XGBoost 故障诊断率为 91.08%,比 PSO-XGBoost 和 SSA-XGBoost 分别提高了 9.9%、6.93%验证了所提方法能够有效地提高变压器故障诊断性能。

关键词:变压器;故障诊断;RFRFE 算法;麻雀算法;XGBoost

中图分类号: TM42;TN06 **文献标识码:** A **国家标准学科分类代码:** 470.40

Transformer fault identification method based on RFRFE and ISSA-XGBoost

Wang Yuhong Wang Zhizhong

(Faculty of Electrical and Control Engineering, Liaoning Technical University, Huludao 125105, China)

Abstract: Aiming at the problem of low accuracy of transformer fault diagnosis, a random forest-recursive feature elimination (RFRFE) algorithm and an improved sparrow algorithm (ISSA) optimization of the extreme gradient boosting tree (XGBoost) transformer fault diagnosis method are proposed. First, based on the diagnostic accuracy, the RFRFE algorithm is used to select important feature variables to remove redundant features; then the traditional sparrow algorithm (SSA) is improved by the uniform distribution random adjustment strategy and the Levi flight strategy, and the ISSA and SSA and particle swarm optimization (PSO) performs algorithm performance testing, which proves that its classification accuracy and network optimization capabilities have been improved; finally, the improved sparrow algorithm is used to optimize XGBoost related hyperparameters to obtain the synthesis of the combination of RFRFE and ISSA-XGBoost. The fault diagnosis model is compared with the PSO-XGBoost and SSA-XGBoost fault diagnosis models. The results show that the fault diagnosis rate of ISSA-XGBoost is 91.08%, which is 9.9% and 6.93% higher than that of PSO-XGBoost and SSA-XGBoost, respectively. The proposed method can effectively improve the performance of transformer fault diagnosis.

Keywords: transformer; fault diagnosis; RFRFE algorithm; sparrow searched algorithm; extreme gradient boosting (XGBoost)

0 引 言

变压器作为电网中关键枢纽设备之一,在电能转换、分配以及传输等方面起着重要作用^[1]。变压器故障不仅会造成巨大的经济损失,还会危害公共安全,因此及时准

确诊断变压器故障对保障电力系统安全具有重要意义^[2]。

变压器发生故障时,绝缘油中会产生 H₂、CH₄、C₂H₆ 等多种气体,其气体含量与故障类型呈现较强的非线性特征相关性^[3]。因此,基于油中溶解气体分析(dissolved gas analysis, DGA)技术普遍用于变压器故障监测与诊断

之中^[4]。近年来,利用神经网络^[5]、支持向量机(SVM)^[6-7]、极限学习机(ELM)^[8]等人工智能方法结合 DGA 分析的变压器故障智能诊断技术不断发展成熟,相关学者对此展开研究。文献[9]通过 CEABC-WNN 诊断模型对变压器故障进行分析,该模型误判率低,稳定性高;文献[10]通过改进萤火虫寻优支持向量机(IFA-SVM)模型对变压器进行故障诊断,验证了 SVM 具有较好的模式识别特性;文献[11]通过改进型 ELM 用于变压器故障诊断,取得了较好的诊断效果。基于上述故障诊断模型对变压器故障诊断精度确实有所提高,但由于算法本身的不足导致故障诊断方法具有一定的局限性。如神经网络结构还存在系统相对复杂、收敛速度慢、过拟合等^[12]缺点;SVM 本质是一个二分类器,对于变压器故障诊断等多分类问题效果不佳;ELM 容易受初始随机生成参数的影响。

极端梯度提升树(extreme gradient boosting, XGBoost)^[13]是在梯度提升决策树(gradient boosting decision tree, GBDT)的基础上对 boosting 思想的扩展。其基本思想是将多个分类准确率较低的基分类器集成为一个准确可靠的集成分类器,以提升其预测速度与精度^[14]。XGBoost 的超参数配置是建立变压器故障诊断的一个重要过程,人工寻参需使用者具有专业的背景知识和实践经验,这使得调整超参数的过程不易重现。为此,学者们利用智能仿生算法对 XGBoost 超参数进行寻优,文献[15]对故障数据进行无编码比值作为模型输入变量,然后利用遗传算法优化 XGBoost 相关超参数得到故障诊断模型,有效提升了诊断精度。

本文建立改进麻雀算法优化 XGBoost (ISSA-XGBoost)模型进行变压器故障诊断,诊断前首先利用随机森林-递归特征消除(random forest-recursive feature elimination, RFRFE)^[16]算法对故障数据集进行特征选择,去除冗余特征;然后采用经服从均匀分布随机调整策略和莱维飞行策略改进的麻雀算法(sparrow search algorithm, SSA)来优化 XGBoost 的相关超参数;最后将 RFRFE 算法选择的重要特征量作为分类模型输入,减轻 ISSA-XGBoost 模型的计算压力。并与其他变压器故障诊断模型对比分析,验证了本文模型在提高变压器故障诊断精度的优势。

1 变压器的故障特征选择

1.1 随机森林(RF)和特征重要度分析

RF^[17]采用 bootstrap 重采样技术采样若干个样本集,每个样本集进行节点分裂决策树建模,多棵决策树构建随机森林,通过每棵树预测投票出最终结果。

本文选择变压器故障发生时其内部油和固体绝缘材料分解的常用 7 种主要气体 H_2 、 CH_4 、 C_2H_6 、 C_2H_4 、 C_2H_2 、 CO 、 CO_2 组合形成变压器故障特征数据集^[18]。 n 为训练样本个数,故障特征 X_j 的特征重要性度量 P_j 按照如下步骤进行计算。

1) 初始化 $n = 1$, 训练样本数据构建决策树 T_n 并将袋外数据标记为 K_n^{oob} 。

2) 决策树每个节点处抽取变压器故障数据集的 m_{try} ($m_{try} < 7$) 个特征,以分类能力最佳特征进行节点分裂。每颗决策树不剪枝。

3) 重复步骤 1)~3) 共 $N(7)$ 次,生成含有 $N(7)$ 棵决策树的随机森林。

4) 对于故障特征 X_j , ($j = 1, 2, \dots, 7$), 对 K_n^{oob} 中的特征 X_j 的值进行扰动,扰动后的数据集记为 K_{nj}^{oob} , 使用 T_n 对 K_{nj}^{oob} 进行分类,统计故障样本正确分类的个数,记为 R_{nj}^{oob} 。

5) 根据式(1)计算故障特征 X_j 的特征重要性度量 P_j 。

$$P_j = \frac{1}{N} \sum_{i=1}^n (R_n^{oob} - R_{nj}^{oob}) \quad (1)$$

1.2 RFRFE 算法

RFRFE 算法由特征重要性排序和递归特征消除(recursive feature elimination, RFE)^[19]两部分构成。特征重要性排序:基于初始故障训练集数据使用 RF 计算特征重要性并排序。递归特征消除:删除重要性最小故障特征,剩余故障特征重新测试 RF 分类正确率。逐次迭代,直到所有特征序列搜索完毕^[20]。最后得到特征数目最少、分类精度最高的变压器故障特征集。

2 基于 ISSA-XGBoost 变压器故障辨识模型

2.1 麻雀优化算法

SSA 算法^[21]是一种新型启发式群智能算法。通过模拟麻雀觅食并躲避捕食者的行为等过程实现对模型参数的优化搜索。假设标准 SSA 算法种群数为 N , 在 D 维区域搜索最优解。

发现者位置更新如下:

$$x_{id}^{t+1} = \begin{cases} x_{id}^t \cdot \exp\left(\frac{-i}{\alpha \cdot T}\right), R_2 < ST \\ x_{id}^t + Q \cdot L, R_2 \geq ST \end{cases} \quad (2)$$

式中: x_{id}^{t+1} 表示在 $t+1$ 代第 i 只麻雀的位置; T 为最大迭代次数; α 为均匀随机数, 在 $(0, 1]$; Q 为符合标准正态分布的随机数; L 表示矩阵; R_2 为警戒值; ST 为安全值。

加入者位置更新如下:

$$x_{id}^{t+1} = \begin{cases} Q \cdot \exp\left(\frac{xw_d^t - x_{id}^t}{t^2}\right), & i > \frac{n}{2} \\ xb_d^{t+1} + \frac{1}{D} \sum_{d=1}^D (\text{rand} \{-1, 1\} \cdot |x_{id}^t - xb_d^{t+1}|), & i \leq \frac{n}{2} \end{cases} \quad (3)$$

式中: xw_d^t 表示最劣位置, x_{id}^t 表示最优位置; xb_d^{t+1} 表示种群 $t+1$ 次迭代时麻雀在 d 维最优位置。

侦察者位置更新为:

$$x_{id}^{t+1} = \begin{cases} xb_d^t + \beta(x_{id}^t - xb_d^t), & f_i \neq f_g \\ x_{id}^t + K \left(\frac{x_{id}^t - xw_d^t}{|f_i - f_w| + e} \right), & f_i = f_g \end{cases} \quad (4)$$

式中: β 为步长控制参数; K 为 $[-1, 1]$ 的随机数, 表示麻雀移动方向; f_i 和 f_w 以及 f_g 分布表示当前麻雀的适应度、最劣适应度和最佳适应度。

2.2 改进麻雀算法

传统麻雀优化算法在搜索过程中存在全局搜索能力不足, 局部开发能力较弱, 易陷入局部最优等缺陷。为此从两个方面改进麻雀搜索算法: 1) 服从均匀分布随机调整策略, 受文献[22]启发在发现者中加入非线性控制参数来协调算法全局搜索与局部开发能力; 2) 莱维飞行策略, 对加入者引入莱维飞行策略, 对当前最优解进行扰动, 增强局部开采能力, 避免算法陷入局部最优。

2.2.1 服从均匀分布随机调整策略

引入非线性控制参数对麻雀发现者进行改进协调算法的全局搜索与局部开发能力。

发现者位置更新如下:

$$x_{id}^{t+1} = \begin{cases} c \cdot \left[x_{id}^t \cdot \exp\left(\frac{-i}{\alpha \cdot T}\right) \right], & R_2 < ST \\ x_{id}^t + Q \cdot L, & R_2 \geq ST \end{cases} \quad (5)$$

非线性控制参数 c 公式如下:

$$c(t) = \delta \left(\lg\left(\frac{c_{\max}}{T_{\max}}\right)^{-20t} + \exp\left(\frac{t}{T_{\max}}\right)^{-20(c_{\max} - c_{\min})} \right) \quad (6)$$

式中: $c_{\max}$ 、 $c_{\min}$ 是参数 c 的最大值和最小值; $T_{\max}$ 为最大迭代次数; t 是当前迭代次数; δ 是服从 $[0, T_{\max}]$ 之间均匀分布随机数且 δ 的取值范围介于 0~1。 $c(t)$ 变化曲线如图 1 所示。

由图 1 可知, $c(t)$ 随迭代次数增加呈非线性递减状态, 在算法搜索前期 c 具有较大的值, 发现者有充分时间进行全局探索; 算法搜索后期 c 具有较小的值来实现对某个区域的局部开发。

服从均匀分布随机数 δ 让参数 c 在算法搜索前期可能保持较大值, 一定程度上加强全局探索; 后期参数 c 下降缓慢 δ 可能产生较大的值, 让参数 c 实现动态变化,

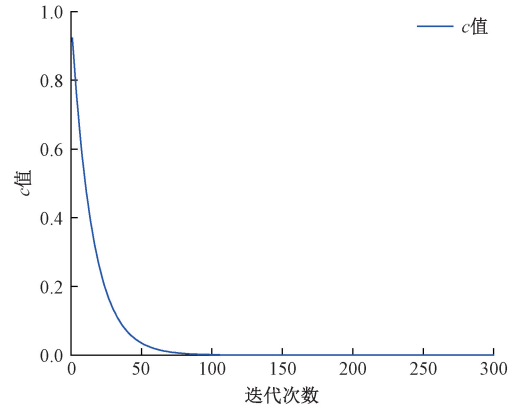


图 1 非线性控制参数 c 变化曲线

Fig. 1 Variation curve of nonlinear control parameter c

从而增强算法局部开发以及在一定程度上加快算法收敛。

2.2.2 莱维飞行策略

莱维飞行^[23]为服从莱维分布的随机搜索方法, 是短距离搜索与偶尔较长距离行走相间的行走方式。加入者中引入莱维飞行策略, 当算法在搜索空间中找到局部最优解时, 对当前最优解进行扰动, 提高了全局寻优能力。

加入者公式更新如下:

$$X_{i,j}^{t+1} = \begin{cases} Q \cdot \exp\left(\frac{xw_d^t - x_{id}^t}{t^2}\right), & i > \frac{n}{2} \\ xb_d^{t+1} + (xb_d^{t+1} - x_{id}^t) \otimes \text{Levy}(\lambda), & \text{其他} \end{cases} \quad (7)$$

式(7)中 $\otimes$ 表示点对点乘法; $\text{Levy}(\lambda)$ 表示服从莱维分布的路径且满足 $1 < \lambda \leq 3$, λ 是幂次数, 带有重尾的概率分布, 较为复杂。

计算 Levy 飞行的搜索路径 $L(\lambda)$ 时, 采用 Mantegna 提出的模拟 Levy 飞行路径的计算公式:

$$s = u / |v|^{1/\beta} \quad (8)$$

式中: s 为 Levy 飞行路径 $L(\lambda)$; $\beta \in (0, 2)$, 本文取 $\beta = 1.5$; 参数 u 、 v 为正态分布随机数, 服从式(9)所示的正态分布; 式(9)所对应正态分布的标准差取值满足式(10)。

$$\begin{cases} u \sim N(0, \sigma_u^2) \\ v \sim N(0, \sigma_v^2) \end{cases} \quad (9)$$

$$\begin{cases} \sigma_u = \left\{ \frac{\Gamma(1+\beta) \sin(\pi\beta/2)}{\Gamma[(1+\beta)/2] 2^{(\beta-1)/2} \beta} \right\} \\ \sigma_v = 1 \end{cases} \quad (10)$$

2.3 算法性能测试

为验证上述改进麻雀优化算法的有效性, 选取 Ackely 和 Schwefel 函数两个测试函数对其进行性能测试, 并与传统 SSA 和粒子群优化算法 (PSO) 进行比较分析。Ackely 函数最中心的位置为最小值, 其四周存在很

多局部极小值,通常检测算法跳出局部最优的能力和全局收敛速度。如图 2 所示,函数表达式如下:

$$f_1(x) = -20 \exp \left(-0.2 \sqrt{\frac{1}{d} \sum_{i=1}^d x_i^2} \right) - \exp \left(\frac{1}{d} \sum_{i=1}^d \cos(2\pi x_i) \right) + 20 \quad (11)$$

其中, $x \in [-32, 32]$, 定义域内全局最小值为 0。

Schwefel 函数较复杂,具有很多局部极小值。如图 3 所示,表达式如下:

$$f_2(x) = - \sum_{i=1}^n -x_i \sin(\sqrt{|x_i|}) \quad (12)$$

其中 $x \in [-500, 500]$, 定义域内全局最小值为 -418.982 9D。

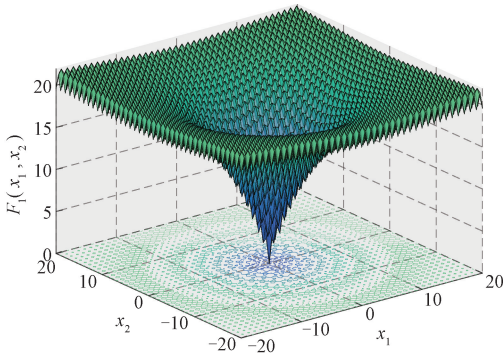


图 2 测试函数 $f_1(x)$

Fig. 2 Test function $f_1(x)$

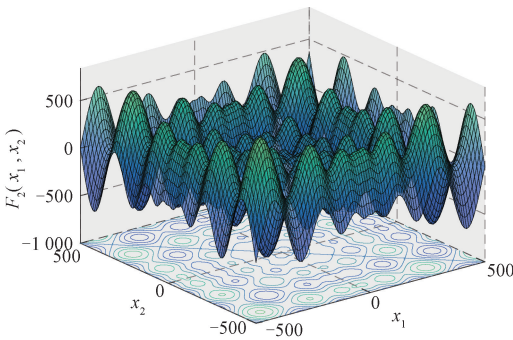


图 3 测试函数 $f_2(x)$

Fig. 3 Test function $f_2(x)$

对测试函数进行寻优测试,最大迭代次数 $T_{\max} = 300$,维数 $Dim=4$,测试结果如图 4 和 5 所示。

在 f_1 函数测试中,3 种寻优算法都能随着迭代次数的增加达到逐渐收敛。PSO 收敛速度慢且适应度值较低,在 300 次的迭代过程中陷入局部最优 60 次;传统 SSA 陷入局部最优 14 次,迭代 91 次时达到收敛状态;而 ISSA 仅在迭代次数为 11 和 13 次时陷入两次局部最优,且迭代 20 次就可达到收敛状态,适应度值为 8.88×10^{-16} 。在 f_2 函数测试中,3 种寻优算法都达到收敛状态且适应度值接近全局最优解,其中,PSO 和 SSA 陷入局部

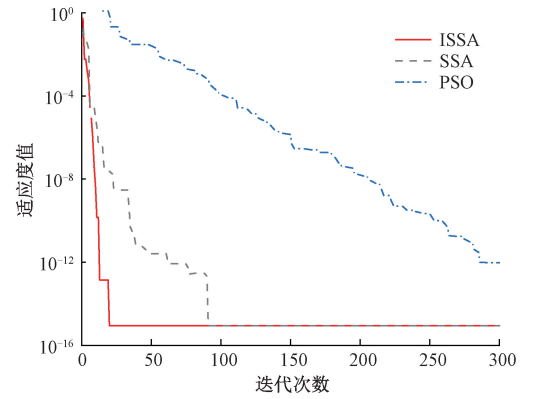


图 4 ISSA、SSA 和 PSO 对 $f_1(x)$ 函数寻优过程

Fig. 4 ISSA, SSA and PSO to optimize the $f_1(x)$ function process

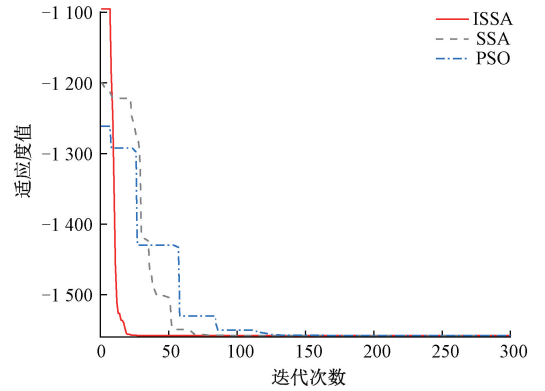


图 5 ISSA、SSA 和 PSO 对 $f_2(x)$ 函数寻优过程

Fig. 5 ISSA, SSA and PSO to optimize the $f_2(x)$ function process

最优的次数分别 16 和 25;ISSA 在第 1、13、27 次迭代时仅有 3 次陷入局部最优,迭代 33 次趋近最优解。通过 f_1 和 f_2 测试函数对 3 种寻优算法进行性能测试,能够说明改进后的麻雀算法在收敛速度和迭代次数以及跳出局部最优能力方面均优于传统 SSA 和 PSO 算法。

2.4 XGBoost 算法

假设变压器故障特征数据集为 Φ , 其中 $\Phi = \{(x_i, y_i) : x_i \in R^m, y_i \in R\}$ 有 n 样本, m 特征。数据集 x_i 对应的标签为 y_i , K 棵回归树 (CART) 集成最终故障预测结果为:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), f_k \in F \quad (13)$$

其中, 每一个函数 f_k 是独立的一棵回归树, F 为 CART 构成的集合; $f_k(x_i)$ 为在第 K 棵 CART 上数据样本集 i 的预测值; $\hat{y}_i$ 为最终的预测结果。XGBoost 的目标函数定义为:

$$Ob_j = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (14)$$

$$\Omega(f_k) = \gamma T + \frac{1}{2} \gamma \|w\|^2 = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (15)$$

式中: y_i 表示样本集真实值, $\sum_{i=1}^n l(y_i, \hat{y}_i)$ 表示损失函数; $\Omega(f_k)$ 为正则项, 预防模型过拟合; T 和 W 分别表示叶子节点的数量及权重; γ 和 λ 为惩罚项, 抑制随着叶子节点增加导致模型更加复杂。设 $\hat{y}_i^{(t)}$ 为在第 t 次迭代中的预测值, 加入新函数 f_t 贪婪优化目标函数, 以提高预测精度。则第 t 次迭代中学习模型的目标函数为:

$$L^{(t)} = \sum_{i=1}^n (l(y_i, \hat{y}_i^{(t-1)}) + f_t(x_i)) + \Omega(f_t) \tag{16}$$

对式(16)中损失函数二阶泰勒公式展开得第 t 次训练简化目标函数为:

$$\tilde{L}^{(t)} \cong \sum_{i=1}^n \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) \tag{17}$$

式中: g_i 为损失函数的一阶导数; h_i 为损失函数的二阶导数。将式(15)代入式(17), 得到最终的目标函数为:

$$\tilde{L}^{(t)} = \sum_{j=1}^T \left[\left(\sum_{i \in I_j} g_i \right) w_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i + \lambda \right) w_j^2 \right] + \gamma T \tag{18}$$

式中: w_j 为二次函数。最优 w_j 为:

$$w_j^* = - \frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \tag{19}$$

代入式(18)中, 得到最优的目标值:

$$L(\widetilde{t}) = - \frac{1}{2} \sum_{j=1}^T \frac{\left(\sum_{i \in I_j} g_i \right)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T \tag{20}$$

式中: $L(\widetilde{t})$ 代表树的结构分数, 越小精度越高。

2.5 ISSA 优化的 XGBoost 模型

为了提高 XGBoost 的分类性能, 利用 ISSA 优化算法对 XGBoost 模型的超参数进行寻优, 以得到更加准确的变压器故障诊断结果。XGBoost 超参数较多, 全部优化对计算机配置要求较高, 运行时间较长。本文最终选取 4 个主要超参数进行优化。

树的数量 $n_estimators$, 该值过大, 占用内存以及训练时间会相应增长; 树的最大深度 max_depth , 能够避免过拟合; 学习率 eta , 控制树的权重; $subsample$ 为控制每棵树, 随机采样比例。XGBoost 相关超参数搜索区间如表 1 所示。

表 1 XGBoost 超参数寻优区间

超参数	寻优范围
$n_estimators$	[10~200]
eta	[0.01~1]
max_depth	[2~10]
$subsample$	[0.1~0.9]

ISSA 对 XGBoost 相关超参数寻优流程如图 6 所示。

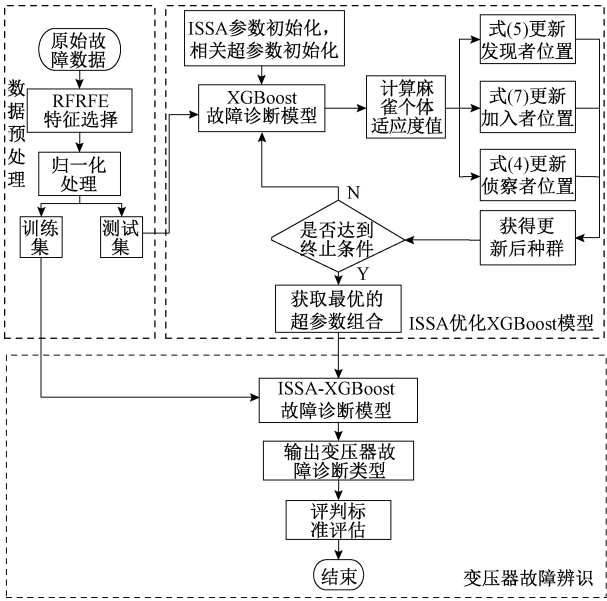


图 6 ISSA 优化 XGBoost 流程

Fig. 6 ISSA optimization XGBoost flow chart

- 1) 设置种群规模 N , 搜索空间维度 D , 最大迭代次数 T 。初始化麻雀种群, 对麻雀位置初始化。
- 2) 将 XGBoost 的相关超参数进行初始化, 包括超参数的寻优范围。
- 3) 计算出当前 XGBoost 超参数下的每一只麻雀的最优和最劣适应度值, 并记录其位置。
- 4) 根据式(5)、(7)、(4)依次随着迭代过程更新发现者、加入者、侦察者位置信息。
- 5) 计算麻雀新位置的适应度值, 若前者更优则保持原有个体不变。
- 6) 更新整个麻雀种群所经历的最优和最劣适应度值, 及其位置。
- 7) 判断是否满足最大迭代次数, 若不满足返回步骤 4) 继续迭代。
- 8) 确定 XGBoost 最优超参数组合, 得到最佳的 XGBoost 故障诊断模型。

3 实验测试及分析

本文数据来源于 IEC TC 10 和国家电网出版的《电网设备检测技术应用典型案例》中的案例数据以及从辽宁某供电公司获得的监测数据, 为保证模型具有良好的普适性和泛化能力, 从中随机选取 505 组有效数据, 在 Anaconda3 集成环境下使用 python3.7 进行多次仿真实验, 验证本文方法在变压器故障诊断中的准确性与可行性。

505 组故障数据随机抽取 101 组作为测试集其余均

为训练集。为消除不同量纲的影响,提高训练速度和分类精度,对故障数据进行归一化处理。辨识变压器故障类型分别是正常、低能放电、高能放电、局部放电、中低温过热、高温过热,编号为 0~5。训练集和测试集样本数据的分布如表 2 所示。

表 2 样本数据的分布

Table 2 Distribution of sample data

故障类型	正常	低能	高能	局部	中低温	高温
训练样本	62	71	73	66	67	65
测试样本	22	12	13	16	16	22

3.1 故障特征选择

RFRFE 算法选择重要特征变量需要确定参数包括随机特征变量个数和决策树数量其中 m_{try} 根据文献[17]选取默认值 $\sqrt{n}$ (n 为训练集特征个数) 决策树数量设置为 1 000。所得故障特征重要度按照由高到低依次排列为 H_2 、 C_2H_6 、 CH_4 、 C_2H_4 、 C_2H_2 、 CO 、 CO_2 。故障特征(个数)与诊断精度的关系如图 7 所示。

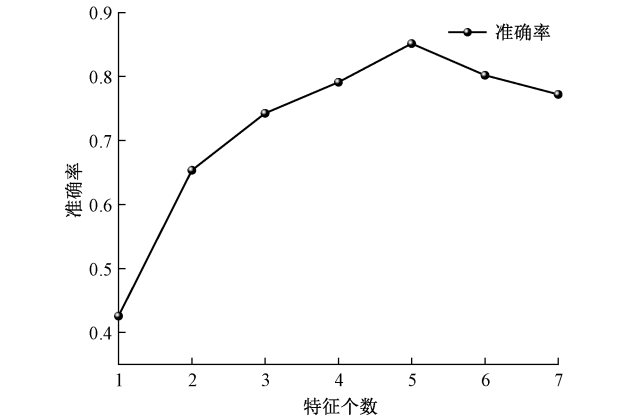


图 7 故障特征与准确率的关系

Fig. 7 The relationship between fault characteristics and accuracy

由图 7 可以看出,特征个数为 7 时,删除重要度最低的特征,剩余特征重新测试分类正确率。随着特征个数减少,分类正确率呈上升趋势。因为删除重要度较低特征能够减小冗余信息对算法性能的影响。特征个数为 5 时,得到最高分类正确率。依次迭代,随着重要度较高故障特征的减少,分类正确率开始逐渐下降。由 RFRFE 算法的选择结果,选择前 5 个特征作为变压器故障诊断的输入。经 RFRFE 算法选择后的部分数据如表 3 所示。

为了说明重要特征变量选择对变压器故障诊断精度的影响。对 RFRFE 故障特征选择前后,诊断精度和运行时间进行对比分析,XGBoost 采用默认参数进行故障诊断,实验结果如表 4 所示。经 RFRFE 特征选择后,降低了故障特征数据的维数,减少了故障诊断模型的输入数据,使得故障诊断的时间和精度都有所提高。

表 3 重要特征样本集部分数据

Table 3 Partial data of important feature sample set

特征气体成分及含量/($\mu\text{L} \cdot \text{L}^{-1}$)					故障类型
H_2	C_2H_6	CH_4	C_2H_4	C_2H_2	
7.5	3.4	5.7	2.6	3.2	正常
31	4.7	6.6	19	67	低能放电
49	25.1	41.6	124	15.7	高能放电
201	9.4	11.2	2.6	0	局部放电
5.5	3.9	9.6	50	1.5	中低温过热
17	11	21	145	4.1	高温过热

表 4 RFRFE 特征选择前后对比

Table 4 Comparison before and after RFRFE feature selection

	原始特征	重要特征
运行时间/s	10.921	5.317
诊断精度/%	74.25	86.13

3.2 不同故障诊断模型诊断结果分析

RFRFE 算法选择后的重要特征作为诊断模型输入,分别选取 K 近邻(KNN)、SVM、ELM、RF 四种主流监督学习模型在默认参数下进行训练,并将测试结果与传统 XGBoost 模型进行对比,结果如表 5 所示。由表 5 可知,XGBoost 模型在变压器故障诊断中对正常和高能放电较为敏感,对高温过热和低能放电诊断效果略差,但综合故障诊断精度在 5 种模型中最高。

表 5 不同模型诊断精度

Table 5 Diagnosis accuracy of different models

故障类型	诊断正确率				
	KNN	SVM	ELM	RF	XGBoost
正常	0.836 3	0.909 0	0.909 0	0.954 5	1
低能放电	0.583 3	0.666 7	0.833 3	0.916 6	0.750 0
高能放电	0.769 2	0.692 3	0.769 2	0.846 1	0.923 0
局部放电	0.937 5	0.687 5	0.687 5	0.812 5	0.875 0
中低温过热	0.875 0	0.75	0.625 0	0.75	0.812 5
高温过热	0.545 5	0.590 9	0.636 4	0.681 8	0.727 3
总计	0.762 3	0.722 7	0.742 6	0.821 7	0.851 4

为了对 XGBoost 的计算稳定性进行测试,5 种方法均运行 30 次,每次得到的结果进行统计分析,对比各方法故障诊断准确率,如表 6 所示。

表 6 不同模型重复训练结果

Table 6 Repeated training results of different models

基准模型	最高准确率	最低正确率	平均正确率
KNN	0.782 1	0.712 8	0.763 2
SVM	0.752 4	0.693 0	0.732 6
ELM	0.792 0	0.702 9	0.740 5
RF	0.851 4	0.742 5	0.797 9
XGBoost	0.871 2	0.801 9	0.842 4

由表 6 可见,基于 XGBoost 的变压器故障诊断模型的 30 次平均故障诊断正确率为 0.842 4。与其他 4 种主

流监督学习模型相比,基于 XGBoost 的变压器故障诊断模型具有最高的平均正确率。

3.3 不同 XGBoost 模型诊断结果分析

由于 XGBoost 模型的超参数对其训练学习效果有着显著的影响,因此采用 ISSA 对设定的 4 个超参数同时优化,并且与 SSA 和 PSO 的优化效果进行对比分析。ISSA、SSA 和 PSO 中相关参数设置为种群数 40,最大迭代次数为 80,空间维数为 4。适应度值以变压器故障诊断精度为标准,适应度变化曲线如图 8 所示。

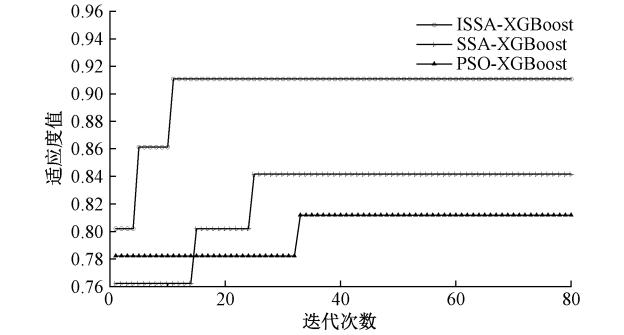


图 8 适应度变化曲线
Fig. 8 Fitness change curve

从图 8 可以看出,3 种优化算法中标准 PSO 收敛速度较慢且适应度值最低。标准 SSA 算法由于全局搜索能力较弱局部开发能力不强,所以在迭代 25 次后才达到全局最优解。而引入非线性控制参数 c 和莱维飞行的 ISSA 在迭代 11 次时就达到收敛状态且适应度值最高,说明改进的 ISSA 与其他两种优化算法相比,具有更强的全局寻优能力和收敛速度。

3.4 诊断结果及对比分析

101 组测试集样本数据用于测试训练后 ISSA-XGBoost 故障诊断模型的诊断精度,分类结果与 SSA 和 PSO 优化 XGBoost 诊断结果进行对比分析。各方法诊断结果如表 7 所示,分类准确率如图 9~11 所示。

表 7 变压器故障诊断正确率结果

Table 7 Transformer fault diagnosis results (%)			
故障类型	PSO	SSA	ISSA
正常	90.90	95.45	100
低能放电	91.66	75	83.33
高能放电	84.61	69.23	92.30
局部放电	68.75	81.25	93.75
中低温过热	75	93.75	87.50
高温过热	77.27	81.81	86.36

由图 9 可以得出,利用 PSO-XGBoost 模型进行变压器故障诊断时,2 个正常诊断为高温过热和中低温过热;1 个低能放电诊断为局部放电;2 个高能放电诊断为正常和高温过热;5 个局部放电诊断为中低温过热、高能放电和低能放电;4 个中低温过热诊断为高能放电和高温过热;5 个高温过热诊断为正常、局部放电、中低温过热以

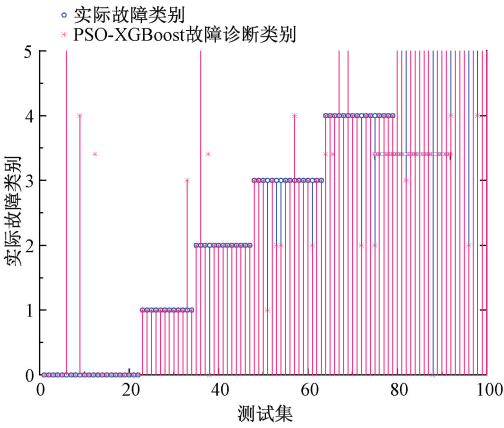


图 9 PSO-XGBoost 故障诊断结果
Fig. 9 PSO-XGBoost fault diagnosis result

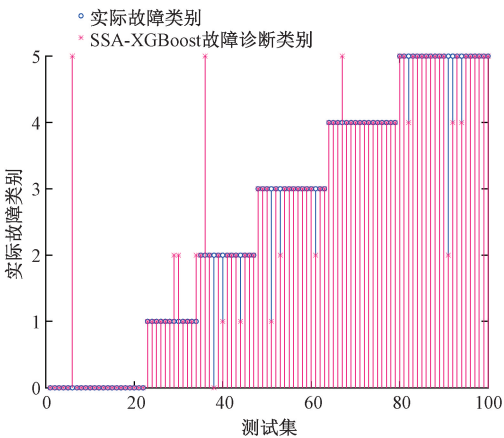


图 10 SSA-XGBoost 故障诊断结果
Fig. 10 SSA-XGBoost fault diagnosis result

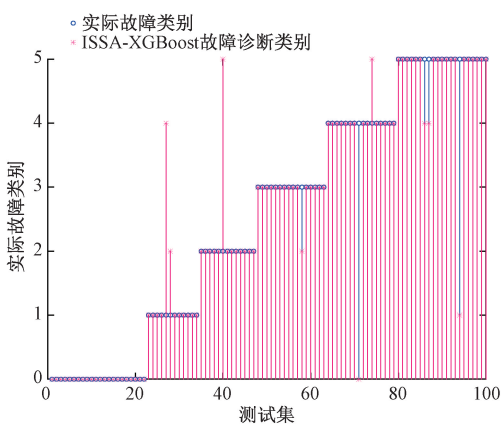


图 11 ISSA-XGBoost 故障诊断结果
Fig. 11 ISSA-XGBoost fault diagnosis result

及高能放电。综合故障诊断率为 81.18%。

由图 10 可以得出,利用 SSA-XGBoost 模型进行变压器故障诊断时,1 个正常诊断为高温过热;3 个低能放电诊断为高能放电;4 个高能放电诊断为正常、高温过热和低能放电;3 个局部放电诊断为高能放电和低能放电;1 个中低温过热诊断为高温过热;4 个高温过热诊断为中

低温过热和高能放电。综合故障诊断率为 84.15%。

由图 11 可以得出,利用 ISSA-XGBoost 模型进行变压器故障诊断时,2 个低能放电诊断为高能放电和中低温过热;1 个高能放电诊断为高温过热;1 个局部放电诊断为高能放电;2 个中低温过热诊断为高温过热和正常;3 个高温过热诊断为中低温过热和低能放电。综合故障诊断率为 91.08%。

结果表明,变压器故障样本集经过 RFRFE 算法特征选择后,ISSA-XGBoost 的变压器故障诊断模型,相比 SSA-XGBoost 和 PSO-XGBoost 故障诊断模型,具有更高的故障识别准确率。

4 结 论

本文针对变压器故障诊断精度较低的问题,充分利用变压器油中气体成分的特征信息进行有监督学习,建立了基于 RFRFE-ISSA-XGBoost 变压器故障诊断模型。RFRFE 算法对变压器故障样本数据进行重要特征选择,能够去除冗余特征,减少变量之间的相关性,缩小训练时间以及提高诊断精度。采用服从均匀分布随机调整策略和莱维飞行策略改进麻雀算法协调了全局搜索与局部开发的平衡并且改善了寻优过程中易陷入局部最优的缺陷。采用 ISSA 优化 XGBoost 的相关超参数,能够有效提高 XGBoost 变压器故障诊断模型的泛化能力和精度。同其他主流监督学习模型以及 PSO-XGBoost 和 SSA-XGBoost 故障诊断模型相比,结果表明本文所提方法能更准确的对变压器故障信息做出快速、精确、可靠诊断,具有一定理论研究和工程实际意义。

参考文献

- [1] 廖伟涵,郭创新,金宇,等. 基于四阶段预处理与 GBDT 的油浸式变压器故障诊断方法[J]. 电网技术, 2019, 43(6): 2195-2203.
LIAO W H, GUO CH X, JIN Y, et al. Fault diagnosis method of oil-immersed transformer based on four-stage preprocessing and GBDT[J]. Power System Technology, 2019, 43(6): 2195-2203.
- [2] 张朝龙,何怡刚,杜博伦,等. 基于深度学习的电力变压器智能故障诊断方法[J]. 电子测量与仪器学报, 2020, 34(1): 81-89.
ZHANG CH L, HE Y G, DU B L, et al. Intelligent fault diagnosis method of power transformer based on deep learning[J]. Journal of Electronic Measurement and Instrumentation, 2020, 34(1): 81-89.
- [3] 刘可真,苟家其,骆钊,等. 基于粒子群优化-长短期记忆网络模型的变压器油中溶解气体浓度预测方法[J]. 电网技术, 2020, 44(7): 2778-2785.
LIU K ZH, GOU J X, LUO ZH, et al. Prediction

method of dissolved gas concentration in transformer oil based on particle swarm optimization-long short-term memory network model[J]. Power System Technology, 2020, 44(7): 2778-2785.

- [4] DAI J, SONG H, SHENG G, et al. Dissolved gas analysis of insulating oil for power transformer fault diagnosis with deep belief network[J]. IEEE Transactions on Dielectrics and Electrical Insulation, 2017, 24(5): 2828-2835.
- [5] 邓聪颖,叶波,苗建国,等. 基于 K-means++ 聚类与概率神经网络的数控机床变位姿动态特性模糊评估[J]. 仪器仪表学报, 2020, 41(12): 227-235.
DENG C Y, YE B, MIAO J G, et al. Fuzzy evaluation of dynamic characteristics of CNC machine tools based on K-means++ clustering and probabilistic neural network[J]. Chinese Journal of Scientific Instrument, 2020, 41(12): 227-235.
- [6] 杨晓敏. 改进灰狼算法优化支持向量机的网络流量预测[J]. 电子测量与仪器学报, 2021, 35(3): 211-217.
YANG X M. Improved gray wolf algorithm to optimize support vector machine for network traffic prediction[J]. Journal of Electronic Measurement and Instrumentation, 2021, 35(3): 211-217.
- [7] 李黄曼,张勇,张瑶. 基于 ISSA 优化 SVM 的变压器故障诊断研究[J]. 电子测量与仪器学报, 2021, 35(3): 123-129.
LI H M, ZHANG Y, ZHANG Y. Research on transformer fault diagnosis based on ISSA optimized SVM[J]. Journal of Electronic Measurement and Instrumentation, 2021, 35(3): 123-129.
- [8] 张淑清,苑世钺,姚玉永,等. 基于 ICEEMD 及 AWOA 优化 ELM 的机械故障诊断方法[J]. 仪器仪表学报, 2019, 40(11): 172-180.
ZHANG SH Q, YUAN SH Y, YAO Y Y, et al. A mechanical fault diagnosis method based on ICEEMD and AWOA to optimize ELM[J]. Chinese Journal of Scientific Instrument, 2019, 40(11): 172-180.
- [9] 贾亦敏,史丽萍,严鑫. 基于精英混沌蜂群算法优化小波神经网络的变压器故障诊断[J]. 高压电器, 2020, 56(8): 230-236.
JIA Y M, SHI L P, YAN X. Transformer fault diagnosis using wavelet neural network based on elite chaos artificial bee colony algorithm[J]. High Voltage Apparatus, 2020, 56(8): 230-236.
- [10] 李俊,冯俊杰,武文吉,等. 基于改进萤火虫算法和多元支持向量机的变压器故障诊断[J/OL]. 电测与表: 1-6 [2021-03-27]. <http://kns.cnki.net/kcms/detail/23.1202.TH.20201225.1450.002.html>.
LI J, FENG J J, WU W J, et al. Fault diagnosis of power transformer based on improved firefly algorithm and

- multi-classification support vector machine [J/OL]. Electrical Measurement & Instrumentation; 1-6 [2021-03-27]. <http://kns.cnki.net/kcms/detail/23.1202.TH.20201225.1450.002.html>.
- [11] 王春明, 朱永利. 基于深度降噪极限学习机的变压器故障诊断[J]. 电测与仪表, 2019, 56(15): 143-147. WANG CH M, ZHU Y L. Transformer fault diagnosis based on deep de-noising extreme learning machine[J]. Electrical Measurement & Instrumentation, 2019, 56(15): 143-147.
- [12] 聂晓音, 谢刚, 李洋, 等. 基于栈式相关性稀疏自编码的电力通信网故障诊断[J]. 电力系统保护与控制, 2019, 47(19): 158-163. NIE X Y, XIE G, LI Y, et al. Fault diagnosis sparse autoencoder[J]. Power System Protection and Control, 2019, 47(19): 158-163.
- [13] CHEN T, GUESTRIN C. XGBoost: A scalable tree boosting system [C]. 22nd ACM SIGKDD International Conference, 2016.
- [14] 滕伟, 黄乙珂, 吴仕明, 等. 基于XGBoost与LSTM的风力发电机绕组温度预测[J/OL]. 中国电力: 18 [20210405]. <http://kns.cnki.net/kcms/detail/11.3265.tm.20210324.1514.004.html>. TENG W, HUANG Y K, WU SH M, et al. Wind turbine generator winding temperature prediction based on XGBoost and LSTM [J/OL]. Electric Power: 18 [20210405]. <http://kns.cnki.net/kcms/detail/11.3265.tm.20210324.1514.004.html>.
- [15] 张又文, 冯斌, 陈页, 等. 基于遗传算法优化XGBoost的油浸式变压器故障诊断方法[J]. 电力自动化设备, 2021, 41(2): 200-206. ZHANG Y W, FENG B, CHEN Y, et al. Fault diagnosis method for oil-immersed transformer based on XGBoost optimized by genetic algorithm[J]. Electric Power Automation Equipment, 2021, 41(2): 200-206.
- [16] GRANITTO P M, FURLANELLO C, BIASIOLI F, et al. Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products[J]. Chemometrics & Intelligent Laboratory Systems, 2006, 83(2): 83-90.
- [17] BREIMAN L. Random forests[J]. Machine Learning, 2001, 45(1): 5-32.
- [18] BARBOSA F R, ALMEIDA O M, BRAGA A P S, et al. Application of an artificial neural network in the use of physicochemical properties as a low cost proxy of power transformers DGA data [J]. IEEE Transactions on Dielectrics & Electrical Insulation, 2012, 19(1): 239-246.
- [19] HU J, MARCULESCU R. Energy-and performance-aware mapping for regular NoC architectures [J]. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2005, 24(4): 551-562.
- [20] 商强, 林赐云, 杨兆升, 等. 基于变量选择和核极限学习机的交通事件检测[J]. 浙江大学学报(工学版), 2017, 51(7): 1339-1346, 1445. SHANG Q, LIN C Y, YANG ZH SH, et al. kernel extreme learning machine [J]. Journal of Zhejiang University (Engineering Science), 2017, 51(7): 1339-1346, 1445.
- [21] XUE J, SHEN B. A novel swarm intelligence optimization approach: Sparrow search algorithm [J]. Systems Science & Control Engineering An Open Access Journal, 2020, 8(1): 22-34.
- [22] 何庆, 林杰, 徐航. 混合柯西变异和均匀分布的蝗虫优化算法 [J/OL]. 控制与决策: 1-10 [2021-05-08]. <https://doi.org/10.13195/j.kzyjc.2019.1609>. HE Q, LIN J, XU H. Locust optimization algorithm based on hybrid Cauchy mutation and uniform distribution [J/OL]. Control and decision: 1-10 [2021-05-08]. <https://doi.org/10.13195/j.kzyjc.2019.1609>.
- [23] VISWANATHAN G M, AFANASYEV V, BULDYREV S V, et al. Levy flights search patterns of biological organisms[J]. Physica A, 2001, 295(1): 85-88.

作者简介



E-mail: yuhong0804001@126.com

Wang Yuhong received her B. Sc. degree from Liaoning Technical University in 2002, M. Sc. degree from Liaoning Technical University in 2008, Ph. D. degree from Liaoning Technical University in 2020. Now she is an associate professor at Liaoning Technical University. Her main research interests include coal mine safety detection, electrical information detection technology and fault diagnosis.



Wang Zhizhong (Corresponding author) graduated from Shandong Agricultural Engineering College in 2015. Now he is a M. Sc. candidate at Liaoning Technical University. His main research interests include electrical information detection technology and fault diagnosis.

王志中 (通信作者), 2015 年于山东农业工程学院毕业, 现为辽宁工程技术大学硕士研究生, 主要研究方向为电气信息检测技术与故障诊断。

E-mail: 766659596@qq.com