

DOI: 10.19783/j.cnki.pspc.231055

基于 Adaboost-INGO-HKELM 的变压器故障辨识

谢国民, 江海洋

(辽宁工程技术大学电气与控制工程学院, 辽宁 葫芦岛 125105)

摘要: 针对目前变压器故障诊断准确率低的问题, 提出一种多策略集成模型。首先通过等度量映射(isometric mapping, Isomap)对高维非线性不可分的变压器故障数据进行降维处理。其次, 利用混合核极限学习机(hybrid kernel based extreme learning machine, HKELM)进行训练学习, 考虑到 HKELM 模型易受参数影响, 所以利用北方苍鹰优化算法(northern goshawk optimization, NGO)对其参数进行寻优。但由于 NGO 收敛速度较慢, 易陷入局部最优, 引入切比雪夫混沌映射、择优学习、自适应 t 分布联合策略对其进行改进。同时为了提高模型整体的准确率, 通过结合 Adaboost 集成算法, 构建 Adaboost-INGO-HKELM 变压器故障辨识模型。最后, 将提出的 Adaboost-INGO-HKELM 模型与未进行降维处理的 INGO-HKELM 模型、Isomap-INGO-KELM 模型、Adaboost-Isomap-GWO-SVM 等 7 种模型的测试准确率进行对比。提出的 Adaboost-INGO-HKELM 模型的准确率可达 96%, 均高于其他模型, 验证了该模型对变压器故障辨识具有很好的效果。

关键词: 故障诊断; 油浸式变压器; Adaboost 集成算法; 切比雪夫混沌映射; 混合核极限学习机; 等度量映射

Transformer fault identification based on Adaboost-INGO-HKELM

XIE Guomin, JIANG Haiyang

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

Abstract: Aiming at the low accuracy of transformer fault diagnosis, a multi-strategy integrated model is proposed. Firstly, the high dimensional nonlinear indivisible transformer fault data is reduced by isometric mapping (Isomap). The hybrid kernel based extreme learning machine (HKELM) is used for training and learning. Considering that the HKELM model is easily affected by parameters, the northern Goshawk optimization (NGO) algorithm is used to optimize its parameters. However, due to the slow convergence rate of NGO, it is easy to fall into local optimal, and Chebyshev chaotic mapping, optimal learning, and adaptive T-distribution joint strategies are introduced to improve it. At the same time, in order to improve the overall accuracy of the model, the Adaboost-INGO-HKELM transformer fault identification model is constructed by combining the Adaboost integrated algorithm. Finally, the test accuracy of the proposed Adaboost-INGO-HKELM model is compared with that of INGO-HKELM model without dimensionality reduction, Isomap-INGO-KELM model, Adaboost-Isomap-GWO-SVM model. The Adaboost-INGO-HKELM model proposed in this paper can achieve an accuracy of 96%, which is higher than other models, which verifies that the model has a good effect on transformer fault identification.

This work is supported by the National Natural Science Foundation of China (No. 51974151).

Key words: fault diagnosis; oil-immersed transformer; Adaboost integration algorithm; Chebyshev chaotic mapping; hybrid kernel based extreme learning machine; isometric mapping

0 引言

变压器运行状态是影响电力系统安全性的重要

因素^[1-2], 在变压器遭受故障时应及时得到处理, 以避免产生不可估量的经济损失, 甚至发生重大安全事故, 所以对变压器的故障辨识具有重要的意义。

目前我国电力系统中, 应用于油浸式变压器的故障诊断方法主要是溶解气体分析法。利用变压器故障时产生的气体对变压器进行故障诊断已经成

基金项目: 国家自然科学基金项目资助(51974151); 辽宁省教育厅重点实验室基金项目资助(LJZS003)

为了重要的手段^[3-4]。但是这种方法需要具备大量经验,且在气体浓度较高的情况下效果会相对明显,应用有一定的局限性^[5]。目前人工智能技术发展迅速,许多机器学习方法也被应用到变压器的故障辨识中,其中包括支持向量机、人工神经网络、贝叶斯网络、极限学习机等^[6-10]。文献[7]利用支持向量机进行变压器故障诊断,但准确度和泛化能力有待进一步提高。文献[8]利用人工神经网络进行变压器故障诊断,但是神经网络需要大量的数据进行训练才能提高其准确性和可靠性。文献[9]运用贝叶斯网络进行诊断,但是贝叶斯网络算法可能会受先验模型的影响,导致对变压器故障诊断效果不佳。文献[10]提出利用极限学习机进行故障辨识,极限学习机是单隐含层前馈神经网络,因其隐含层无需迭代,学习速度较快、泛化能力较强,故更适用于故障预测。但极限学习机对非线性问题的逼近能力、泛化能力、鲁棒性有限。通过利用核函数的核极限学习机,在获得最小平方优化解后,解决了传统极限学习机在处理非线性问题时的限制^[11]。

传统的机器学习方法学习模型单一,不能有效地保证该模型的分类预测精度。Adaboost 集成算法从数据集中训练不同的弱分类器,将弱分类器按加权组合的方式构成一个强分类器^[12]。Adaboost 通过加权训练样本来降低过拟合风险,同时可以有效处理高维特征空间下的数据样本。核极限学习机(kernel based extreme learning machine, KELM)是在极限学习机(extreme learning machine, ELM)的基础上引入了核函数,通过核函数的映射,可以更好地适应复杂数据分布,比 ELM 具有更好的泛化能力和非线性分类建模能力^[13]。故本文提出利用 Adaboost 集成学习方法来优化混合核极限学习机(HKELM),同时利用改进的北方苍鹰群智能优化算法来优化混合核极限学习机的参数,提升模型的整体性能。

1 变压器故障特征处理

1.1 故障数据的平衡处理

国家电网某分公司的变压器故障数据有正常类别 192 个、中低温过热 85 个、高温过热 63 个、低能放电 105 个和高能放电 50 个。由于高能放电、高温过热故障发生频率较低,易出现样本不平衡问题,此时分类边界会偏向少数类的区域,将会影响模型学习更本质的特征以及模型的鲁棒性^[14-15]。所以采取自适应综合过采样方法(adaptive synthetic sampling, ADASYN)将故障数据扩充为 5 种故障类型数据(均至 192 个)。

如图 1 所示,平衡数据前高能放电(黄色)特征数量较少,而平衡数据后的特征如图 2 所示,高能放电特征数量变多,此时各故障类型的特征数量较为均衡。

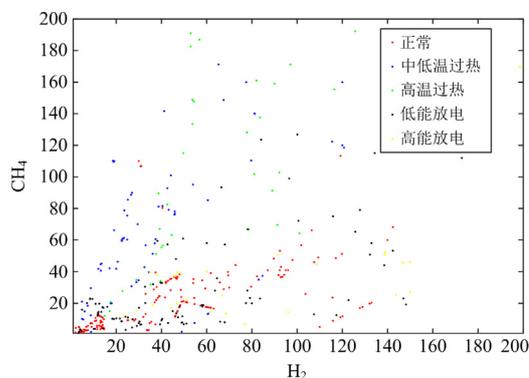


图 1 平衡数据前特征分布图

Fig. 1 Feature distribution plot before balancing data

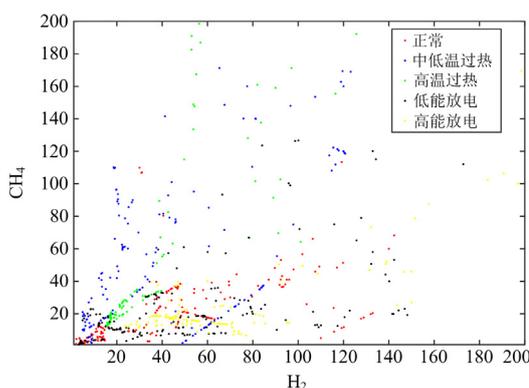


图 2 平衡数据后特征分布图

Fig. 2 Feature distribution plot after balancing data

1.2 故障数据的特征扩充

当变压器出现故障时,变压器内部的绝缘油将会发生热裂解,通过热裂解,绝缘油主要产生 5 种气体^[16]—— H_2 、 CH_4 、 C_2H_2 、 C_2H_4 和 C_2H_6 。

不同故障类型的 5 种气体浓度以及各气体之间的比例也不相同。选取主要产生的 5 种气体浓度以及 15 组各气体间的比值关系作为特征数据,这样可更准确地反映故障类型的特征数据,如表 1 所示,其中, $Z_1 + Z_2 = CH_4 + C_2H_2 + C_2H_4 + C_2H_6$ 。

1.3 故障特征的降维

由于高维数据将会导致计算需求过高,易出现过拟合现象,同时样本点之间的距离会变得稀疏,将可能影响数据的训练与预测结果。故本文采取了等度量映射(isometric mapping, Isomap)算法进行特征提取与降维处理。

表 1 故障特征气体数据

Table 1 Fault characteristic gas data

20维特征				
H ₂	CH ₄	C ₂ H ₂	C ₂ H ₄	C ₂ H ₆
$\frac{CH_4}{H_2}$	$\frac{C_2H_2}{H_2}$	$\frac{C_2H_4}{H_2}$	$\frac{C_2H_6}{H_2}$	$\frac{C_2H_2}{CH_4}$
$\frac{C_2H_4}{CH_4}$	$\frac{C_2H_2}{C_2H_4}$	$\frac{CH_4}{C_2H_6}$	$\frac{C_2H_2}{C_2H_6}$	$\frac{C_2H_4}{C_2H_6}$
$\frac{C_2H_4}{Z_1+Z_2}$	$\frac{CH_4}{Z_1+Z_2}$	$\frac{C_2H_6}{Z_1+Z_2}$	$\frac{C_2H_2}{Z_1+Z_2}$	$\frac{H_2}{Z_1+Z_2}$

等度量映射是一种流形学习，常用来处理非线性数据的降维。其原理是建立测地距离与欧式距离的映射关系来实现对高维数据的降维^[17]。通过考虑数据点之间的最短路径，同时保留各个数据点之间的全局结构信息，使在降维过程中能够更好地保持数据的相对距离关系。

在传统的机器学习方法中，不同数据点之间的距离为欧氏距离^[18]。而在实际应用中，这些数据并不是定义在欧氏空间中，而是极大可能分布在一个流形体中，故无法用来直接度量实际的数据。

在 Isomap 算法中，当计算两点数据的真实距离(即测地距离)时，测地距离才反映流形的真实低维结构。如图 3 所示，蓝色实线为两个样本点之间的测地距离，蓝色虚线代表两个样本点之间的欧式距离。

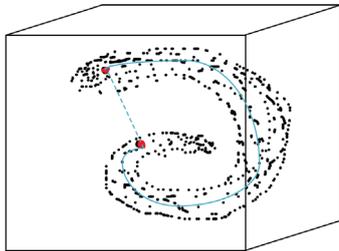


图 3 流形学习数据图

Fig. 3 Popular learning data graph

Isomap 算法步骤如下所述。

步骤 1: 计算数据集中每一对样本点的欧式距离，得到高维欧式距离矩阵 D_1 。

步骤 2: 采用 K 近邻法确定邻接关系，对得到的 k 个近邻点与每个样本点 X_i 之间的距离作为权重，构成带权邻接矩阵 G 。

步骤 3: 利用带权邻接矩阵 G 计算不同数据点之间的最短路径距离矩阵 D_2 。

步骤 4: 对最短路径距离矩阵 D_2 进行多维缩放 (multidimensional scaling, MDS) 处理，得到低维嵌入空间数据集 P 。

Isomap 算法中的残差指在低维空间中，数据点之间的实际距离与这些数据在高维空间中的最短路

径距离间的误差。残差反映了降维过程中算法无法完全保持原始的数据点之间的欧氏距离。采用 Isomap 对变压器故障数据进行降维处理后得到维数与残差关系曲线，如图 5 所示。

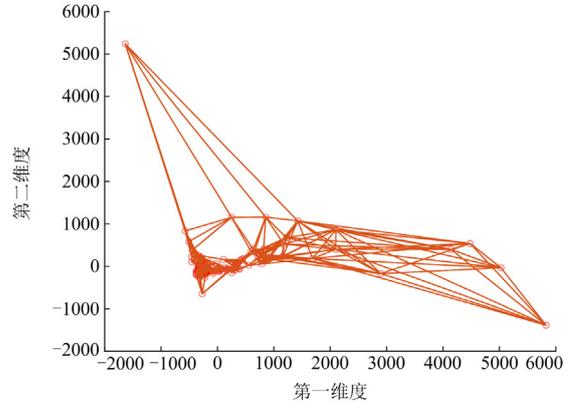


图 4 Isomap 二维邻域图

Fig. 4 Two-dimensional neighborhood map of Isomap

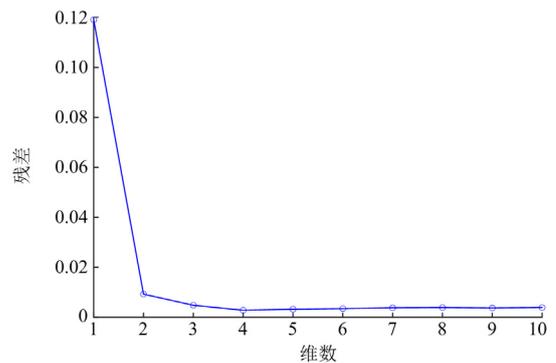


图 5 维数与残差的关系图

Fig. 5 Dimension and residual relationship diagram

根据图 5 所示，由出现的拐点以及曲线稳定情况可知，在维度为 4 时，残差达到最低且趋于稳定，故本文最终将降维的维度确定为 4 维。

2 混合核极限学习机

在对变压器进行故障诊断分类时，大多依靠学习机的分类能力，所以在采用学习机时更注重学习机的分类速度和泛化能力。

极限学习机在训练过程中随机产生隐藏层的输入权重和阈值，极限学习机网络具有泛化能力强、学习速度快等优势^[19]。

核极限学习机是利用核函数的空间映射，提升学习机的非线性分类能力，使模型能够更好地处理复杂数据的分布。通过引入核函数使得算法学习速度加快、泛化能力变强。对于非线性回归或分类具有良好的效果，与传统神经网络相比，计算量和搜

索空间较少^[20]。

混合核极限学习机采用两个核函数, 以增强模型的泛化能力为目的, 其输出模型为

$$\begin{cases} \mathbf{f}(x) = \mathbf{H}\boldsymbol{\beta} \\ \mathbf{H} = \begin{bmatrix} h(x_1) \\ \vdots \\ h(x_N) \end{bmatrix} \end{cases} \quad (1)$$

式中: $\mathbf{f}(x)$ 为输出; \mathbf{H} 为隐含层节点输出; $\boldsymbol{\beta}$ 为隐含层和输出层之间连接权重; $h(\cdot)$ 为隐含层输入。

$$\boldsymbol{\beta} = \mathbf{H}^T \left(\mathbf{H}\mathbf{H}^T + \frac{\mathbf{I}}{c} \right)^{-1} \mathbf{T} \quad (2)$$

式中: c 为正则化参数; \mathbf{I} 为单位矩阵; \mathbf{T} 为训练集目标向量。核极限学习机的矩阵模型为

$$\begin{cases} \boldsymbol{\Omega} = \mathbf{H}\mathbf{H}^T \\ \boldsymbol{\Omega}_{i,j} = h(x_i)h(x_j) = \mathbf{K}(x_i, x_j) \end{cases} \quad (3)$$

核矩阵 $\boldsymbol{\Omega}$ 替代 ELM 中的 $\mathbf{H}\mathbf{H}^T$ 矩阵, $\mathbf{K}(x_i, x_j)$ 为核函数矩阵。学习机模型如图 6 所示。

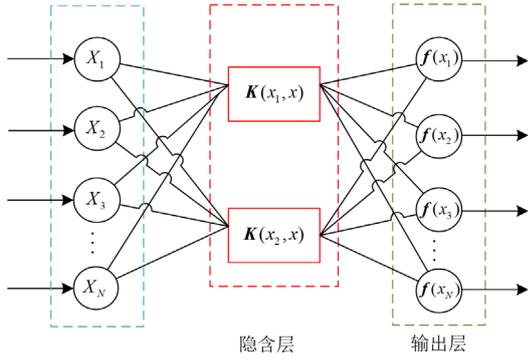


图 6 KELM 模型结构图

Fig. 6 KELM model structure diagram

输出模型为

$$\mathbf{f}(x) = \begin{bmatrix} \mathbf{K}(x_1, x) \\ \vdots \\ \mathbf{K}(x_N, x) \end{bmatrix}^T \left(\boldsymbol{\Omega} + \frac{\mathbf{I}}{c} \right)^{-1} \mathbf{T} \quad (4)$$

由上述可见, 必须定义核参数、核函数加权系数和正则化系数的最优取值集, 以实现泛化能力、性能稳定性和精度之间的协调^[21]。如果仅采用一种核函数, 那么搜索范围会受到限制。多项式 Poly 核函数是全局核函数, 而高斯 RBF 核函数为局部核函数, 本文通过将两种核函数线性相结合, 使学习机在全局和局部范围中都有较好的性能。

Poly 核函数的函数表达式为

$$K_{\text{poly}}(x_i, x_j) = (x_i, x_j + c_1)^d \quad (5)$$

RBF 核函数的函数表达式为

$$K_{\text{RBF}}(x_i, x_j) = \exp\left(-\|x_i - x_j\|^2 / 4\sigma^2\right) \quad (6)$$

式(5)、式(6)中, c_1 、 d 均为多项式 Poly 核函数的核参数; σ 为高斯 RBF 核函数的核参数。混合核函数由 Poly 核函数和 RBF 核函数线性相组合。

$$\begin{cases} K_{\text{H}}(x_i, x_j) = k_1 K_{\text{poly}}(x_i, x_j) + k_2 K_{\text{RBF}}(x_i, x_j) \\ k_1 + k_2 = 1 \end{cases} \quad (7)$$

由两种核函数的线性组合得到混合核极限学习机(HKELM)。同时其参数 c 、 σ 、 c_1 、 d 、 k_1 对混合核极限学习机性能具有很大影响, 与原有的故障诊断模型相比, 参数优化后的故障诊断模型提升了诊断精度^[22], 故需要采用一种群智能优化算法对混合核极限学习机的参数进行联合寻优, 提高模型分类性能。

3 改进北方苍鹰优化算法

3.1 北方苍鹰优化算法

北方苍鹰优化算法(northern goshawk optimization, NGO)是 Mohammad Dehghani 在 2021 年提出的一种模拟北方苍鹰识别和追捕猎物的群智能优化算法^[23-24]。北方苍鹰的两个主要动作被模仿为两个阶段: 1) 猎物识别和攻击; 2) 追击和逃逸操作。

3.1.1 猎物识别阶段(p1)

NGO 算法的搜索机制源于其对猎物的高效搜索和捕获。在第一阶段, 北方苍鹰会随机地选择一个猎物进行攻击, 这种行为过程的数学表达式为

$$P_i = X_k, i = 1, 2, \dots, N, k = 1, 2, \dots, i-1, i+1, \dots, N \quad (8)$$

式中, P_i 为第 i 个北方苍鹰选择的猎物的位置。为提高算法全局搜索能力, k 在 $[1, N]$ 内不为 i 的随机整数, 避免发生只在当前位置搜索的情况。

$$X_{i,j}^{\text{new,p1}} = \begin{cases} x_{x,j} + r(p_{i,j} - Ix_{i,j}) & F_{p_i} < F_i \\ x_{x,j} + r(x_{i,j} - p_{i,j}) & F_{p_i} \geq F_i \end{cases} \quad (9)$$

$$X_i = \begin{cases} X_i^{\text{new,p1}} & F_i^{\text{new,p1}} < F_i \\ X_i & F_i^{\text{new,p1}} \geq F_i \end{cases} \quad (10)$$

式中: F_{p_i} 为其适应度值; X_i 为第 i 个北方苍鹰的当前状态; $X_i^{\text{new,p1}}$ 为其新状态; $X_{i,j}^{\text{new,p1}}$ 是其在 j 维度的新状态; F_i 为第 i 个北方苍鹰的目标函数值; $P_{i,j}$ 为第 i 个北方苍鹰在 j 维度选择的猎物的位置; X_i 为第 i 个北方苍鹰的位置; $F_i^{\text{new,p1}}$ 是与其对应的适应度值; r 为 $[0,1]$ 内的随机数; 为了控制猎物识别的速度和精度, I 取 1 或 2, 当 I 取 1 时, 搜索速度较快, 但精度较低, 当 I 取 2 时, 搜索速度较慢, 但精度较高。

3.1.2 追击和逃逸阶段(p2 开发阶段)

在北方苍鹰追逐到猎物时，猎物会尝试挣脱逃跑，北方苍鹰会继续追击，由于北方苍鹰速度极快，最终会成功捕捉到猎物。此过程模拟的是算法对局部搜索的能力^[25]。假设猎物逃跑的范围在半径为 R 的圆内，则此时的北方苍鹰的行为过程数学表达式为

$$X_{i,j}^{new,p2} = x_{i,j} + R(2r - 1)x_{i,j} \quad (11)$$

$$R = 0.02 \left(1 - \frac{t}{T} \right) \quad (12)$$

$$X_i = \begin{cases} X_i^{new,p2} & F_i^{new,p2} < F_i \\ X_i & F_i^{new,p2} \geq F_i \end{cases} \quad (13)$$

式中： t 为当前迭代次数； T 为最大迭代次数； $X_i^{new,p2}$ 为第二阶段的第 i 只北方苍鹰的最新状态； $F_i^{new,p2}$ 为其对应的适应度值； $X_{i,j}^{new,p2}$ 为捕猎第二阶段的第 i 只北方苍鹰在 j 维度下的新状态。

3.2 改进的北方苍鹰优化算法

3.2.1 Chebyshev 混沌映射策略

在标准的北方苍鹰优化算法中，北方苍鹰的初始位置是随机分布的，因此会出现位置分布不均的情况且易陷入局部最优。因此引入了 Chebyshev 混沌映射策略，Chebyshev 映射是混沌映射的典型代表，其不需要设置系统参数，初始值的设置影响较小，稳定性更强^[26]，其表达式为

$$X_{k+1} = \cos(k \cos^{-1}(X_k)), X_k \in [-1,1] \quad (14)$$

Chebyshev 映射是常见的一种非线性的映射，是利用 Chebyshev 多项式对输入进行变换。此策略将会使北方苍鹰的初始位置更具有随机性，使种群初始化遍历性更加充分，可以有效地避免算法存在搜索盲区。

3.2.2 猎物择优策略

在猎物识别阶段，更新前后的个体进行信息交互，如果更新前的个体适应度很强，则会学习相对正确，如果更新前的个体适应度很弱，盲目进行学习且不能保证其相对正确性。故此阶段先对初始种群进行适应度排序得到集合 N_A ，选出适应度在前 10% 的个体作为集合 N_B ，再进行随机选择学习，保证了更新后的种群质量。

$$N_B = 10\%N_A \quad (15)$$

3.2.3 自适应 t 分布策略

在猎物识别阶段，选出适应度前 10% 的个体进行学习，但这种策略易陷入局部最优，全局开发性能不足。故在后期引入自适应 t 分布策略。自适应 t 分布允许自由度根据数据的不确定性进行估计从而自动调整。本文采用以迭代次数 t 为自适应 t 分布

的自由度。对适应度排序在 10% 后的个体位置进行扰动，使算法具有较强的全局开发能力，具体的表达式为

$$X_k^{t+1} = X_k^t \cdot t \quad (16)$$

式中： X_k^{t+1} 为扰动后的北方苍鹰的位置； X_k^t 为北方苍鹰第 t 次迭代时的位置。

3.2.4 改进后的 NGO 算法测试

传统的 NGO 算法由于在第二阶段盲目地从全部个体中随机挑选学习，导致收敛速度变慢。单峰测试函数 f_6 与多峰测试函数 f_{11} 如表 2 所示，其算法对比结果如图 7、图 8 所示。

表 2 测试函数

Table 2 Test function

函数	维度	变量范围
$f_6(x) = \sum_{i=1}^n [x_i + 0.5]^2$	30	[-30, 30]
$f_{11}(x) = \sum_{i=1}^{30} [x_i^2 - 10\cos(2\pi x_i) + 10]$	30	[-5.12, 5.12]

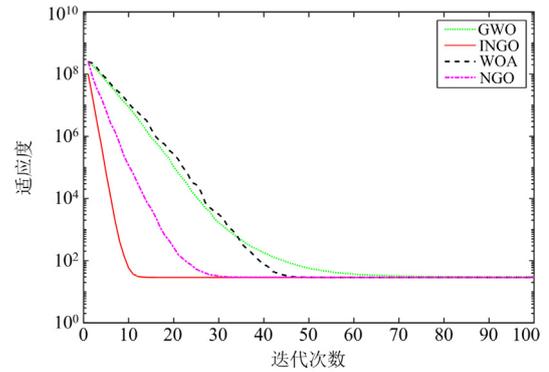


图 7 测试函数 f_6 的结果

Fig. 7 Results of test function f_6

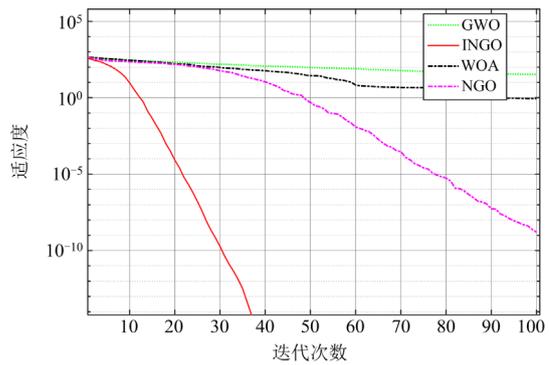


图 8 测试函数 f_{11} 的结果

Fig. 8 Results of test function f_{11}

由图 7 与图 8 可知，无论是单峰测试函数 f_6 还是双峰测试函数 f_{11} ，引入改进策略后的 INGO 算法收敛速度明显快于传统的 NGO 算法。故采用经改进后的 INGO 算法对混合核极限学习机的参数进行

联合寻优。

4 Adaboost算法与集成学习

Adaboost 是一种集成学习算法, 也是代表性的 boosting 策略, 通过最小化错误概率来提高模型的性能^[27], 其原理是通过迭代训练, 由上一次的分类错误率来计算调整下一个弱分类器的权重, 再根据弱分类器的权重动态调整样本的权重, 使下一个弱分类器更关注预测相差较大的样本, 最后若干个弱分类器通过加权组合成一个强分类器, 其具备较强的泛化能力和鲁棒性能, 能够较好地解决过拟合、梯度爆炸等问题^[28]。Adaboost 模型结构如图 9 所示。

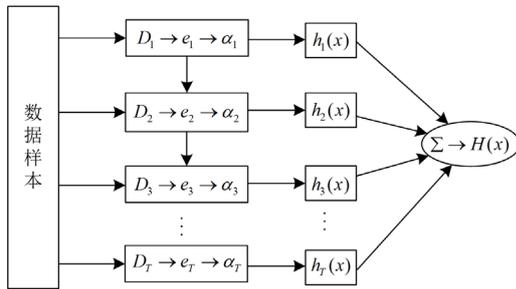


图 9 Adaboost 模型结构图

Fig. 9 Adaboost model structure diagram

1) 初始化样本数据的权重, 使每一个样本数据初始化的权值均相同。

$$w_i = \frac{1}{T} \quad (17)$$

此时样本数据集的初始权重分为

$$D_1(i) = (w_1, w_2, \dots, w_T) = \left(\frac{1}{T}, \dots, \frac{1}{T}\right) \quad (18)$$

2) 使用权重为 $D_t(i)$ 的训练数据集, 通过训练得到第 t 个基本分类器 h_t 。

3) 计算基本分类器 h_t 的分类误差率 e_t 。按照式 (19) 计算第 t 个基本分类器的权重 α_t 。

$$\alpha_t = \frac{1}{2} \ln \left(\frac{1 - e_t}{e_t} \right) \quad (19)$$

4) 更新训练数据集的权重。

$$D_{t+1}(i) = \frac{D_t(i) \exp(-\alpha_t y_i H_t(x_i))}{Z_t} \quad (20)$$

$$Z_t = 2\sqrt{e_t(1 - e_t)}$$

式中, Z_t 是归一化因子。

5) 重复 T 次 2)–4), 由此得到 T 个弱分类器的分类权重 α_t , 构成基本分类器的线性组合。

$$f(x) = \sum_{t=1}^T \alpha_t h_t(x) \quad (21)$$

6) 通过将其加权组合成强分类器。

$$H(x) = \text{sign} \left(\sum_{t=1}^T \alpha_t h_t(x) \right) \quad (22)$$

5 Adaboost-INGO-HKELM模型的故障辨识

Adaboost 通过更注重分类错误样本, 对多个弱分类器加权组成一个强分类器, 既能保证模型的性能, 又具有较强的泛化能力。核函数是核极限学习机的重要因素, 采用混合核函数可以改善模型性能^[29]。其中 c 、 σ 、 c_1 、 d 、 k_1 参数对 HKELM 的性能影响较大, 采用 INGO 算法对各参数进行联合寻优, 获得参数稳定、性能优异的 Adaboost-INGO-HKELM 模型。

构造模型的故障诊断流程如图 10 所示, 具体步骤如下所述。

1) 准备数据集: 采用 Isomap 选取的变压器故障数据, 并进行特征数据降维处理, 将降维后的数据按照 7:3 的比例划分训练集和测试集。

2) 设置 Adaboost 的迭代次数 T , 初始化样本权重 α_t 。

3) 初始化参数: 初始化 INGO 最大迭代次数, 种群数量。设置 HKELM 的正则化系数 c 、Poly 核函数的核系数 c_1 与 d 、RBF 核函数的核系数 σ 、核函数的权重系数 k_1 的寻优范围。

4) 利用 INGO 更新参数, 计算适应度。判断是否寻优完毕, 若完成则得到弱分类器 h_t 。

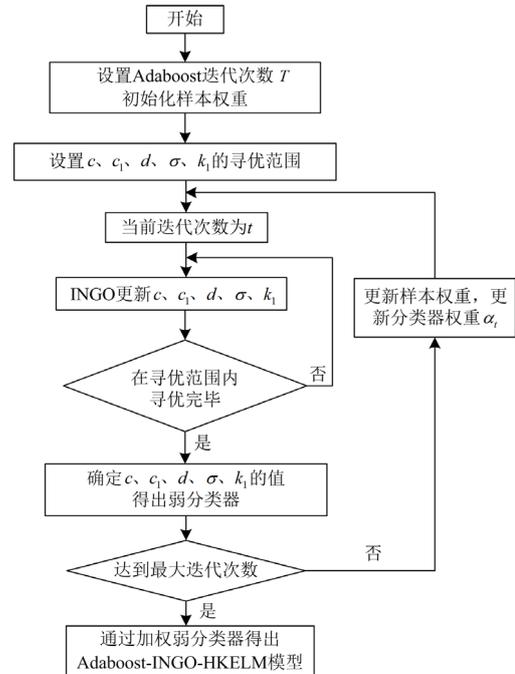


图 10 模型诊断流程图

Fig. 10 Model's diagnostic flow chart

5) 通过计算误差率来更新弱分类器权重 α_i 。

6) 判断是否达到最大迭代次数, 若达到则寻优结束, 通过加权弱分类器输出强分类器。若没达到则回到步骤4)。

6 故障诊断实验分析

从国家电网某分公司变压器故障数据中, 选取 5 种故障类型数据, 经数据平衡处理后的故障类型编号如表 3 所示。

表 3 故障数据

Table 3 Fault data

故障类型	故障编号	训练集样本数	测试集样本数
正常	1	35	15
中低温过热	2	35	15
高温过热	3	35	15
低能放电	4	35	15
高能放电	5	35	15

为使数据平衡, 每种类型按照故障发生频率最小的高能放电数据个数进行测试。正常、中低温过热、高温过热、低能放电、高能放电各 50 个故障数据, 共计 250 个数据, 按照表 3 将数据样本划分训练集和测试集。

1) 未经数据降维的对比实验

在对变压器故障数据诊断时, 学习机的选择尤其重要, 故采用 INGO 算法对 HKELM 的参数寻优, 得到 INGO-HKELM 模型, 对其进行测试的结果如图 11 所示, 准确率为 83%, 正常状态、低能放电及高能放电状态均有偏差, 可能由于高维度的训练数据, 导致测试准确率偏低。

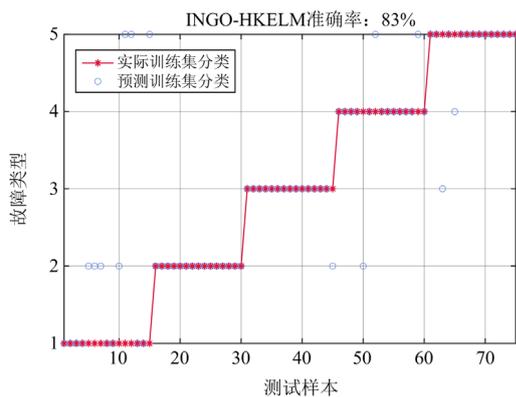


图 11 INGO-HKELM 模型测试结果

Fig. 11 INGO-HKELM model test results

利用灰狼优化算法(grey wolf optimizer, GWO)来优化 HKELM 学习机, 得到的 GWO-HKELM 模型测试结果如图 12 所示, 其准确率仅为 80%。GWO

算法具有收敛速度慢、易陷入局部最优且对初始参数敏感的缺点, 因此测试结果会偏低。同时说明 INGO 结合 HKELM 模型的效果更佳。

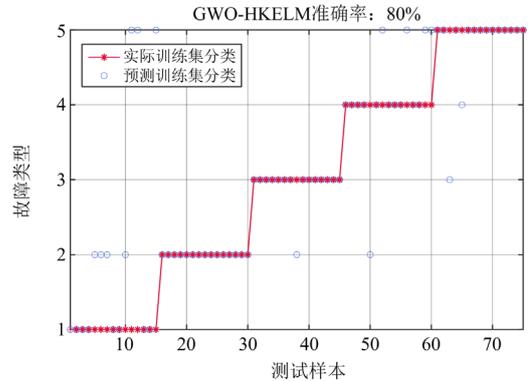


图 12 GWO-HKELM 模型测试结果

Fig. 12 GWO-HKELM model test results

2) 经数据降维后的对比实验

高维数据易发生过拟合、缺乏维度灵活性、计算难度增加等, 故对训练集进行降维处理, 如图 13 所示, 通过对训练集进行 Isomap 降维后, Isomap-INGO-HKELM 模型测试结果为 86.7%, 通过 Isomap 降维有效地提升了模型准确率。

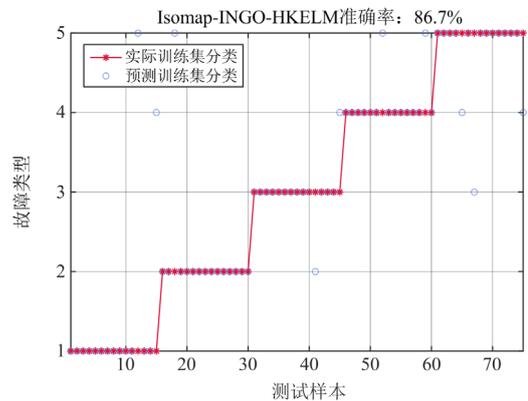


图 13 Isomap-INGO-HKELM 模型测试结果

Fig. 13 Isomap-INGO-HKELM model test results

如图 14 所示, 对 Isomap-INGO-KELM 模型进行测试, 准确率为 84%, 相比 Isomap-INGO-HKELM 模型的准确率有所下降, 说明同时结合混合模型与极限学习机的优点集一体的混合极限学习机, 能更好处理非线性关系数据, 具有较强的泛化能力, 在处理变压器故障数据时可以获得更高的准确率与更好的效果。

如图 15 所示, 对 Isomap-GWO-SVM 模型进行测试, 准确率仅为 82.7%。相比于 Isomap-INGO-KELM 模型与 Isomap-INGO-HKELM 模型的准确率

有所下降, 说明极限学习机在对变压器故障辨识应用上更优于支持向量机。

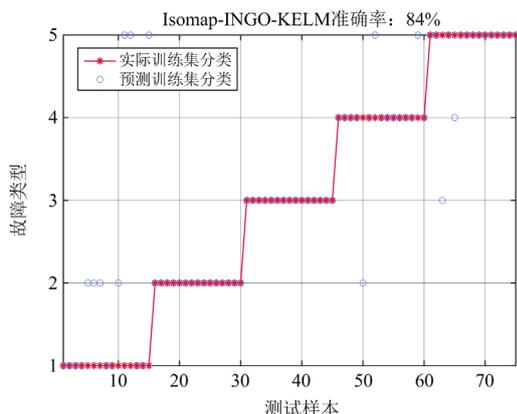


图 14 Isomap-INGO-KELM 模型测试结果
Fig. 14 Isomap-INGO-KELM model test results

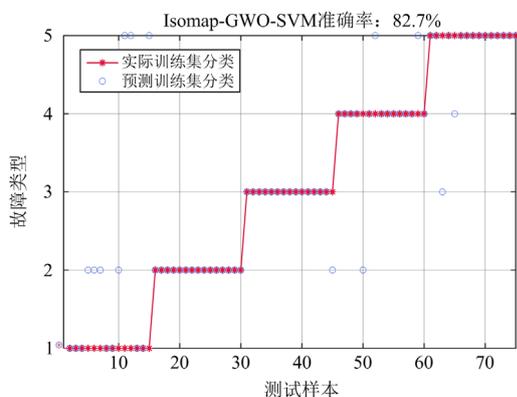


图 15 Isomap-GWO-SVM 模型测试结果
Fig. 15 Isomap-GWO-SVM model test results

3) 引入 Adaboost 集成算法进行对比实验

通过对 Adaboost-Isomap-GWO-SVM 模型和 Adaboost-Isomap-INGO-KELM 模型进行测试, 结果如图 16、图 17 所示, 准确率分别为 88%、93.3%。

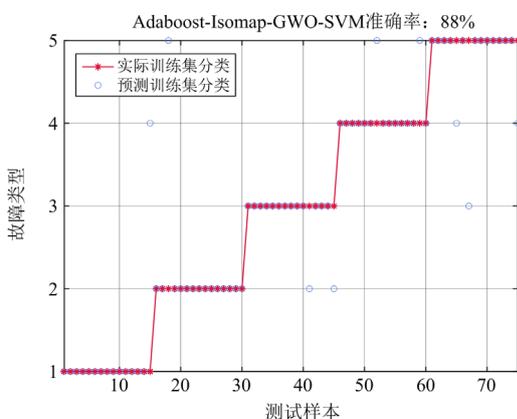


图 16 Adaboost-Isomap-GWO-SVM 模型测试结果
Fig. 16 Adaboost-Isomap-GWO-SVM model test results

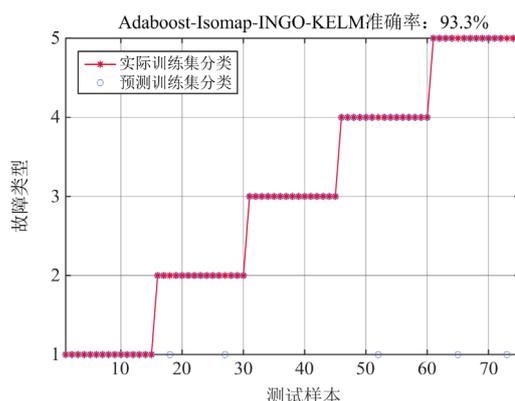


图 17 Adaboost-Isomap-INGO-KELM 模型测试结果
Fig. 17 Adaboost-Isomap-INGO-KELM model test results

通过 Adaboost 集成算法的结合, 弱分类器加权组成一个强分类器, 使 Isomap-GWO-SVM 模型和 Isomap-INGO-KELM 模型的准确率分别提高了 5.3%和 7.3%。

4) 引入强集成双核函数模型试验

基于 INGO-HKELM 模型结合 Adaboost 集成算法, 构建 Adaboost-Isomap-INGO-HKELM 模型并进行测试, 如图 18 所示, 准确率达到 96%, 仅有 3 个训练样本预测失误, 说明 Adaboost-INGO-HKELM 模型在变压器故障诊断中可以获得较高的准确率。

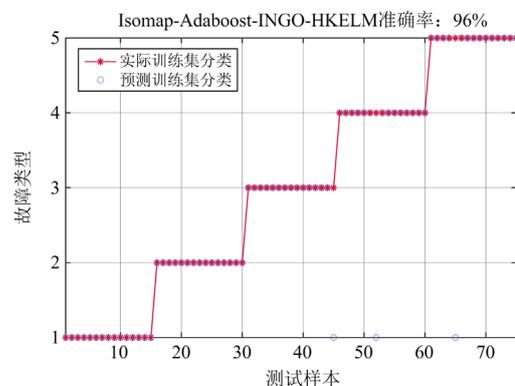


图 18 Adaboost-INGO-HKELM 模型测试结果
Fig. 18 Adaboost-INGO-HKELM model test results

表 4 不同的模型准确率对比

Table 4 Comparison of accuracy of different models

模型名称	准确率/%
INGO-HKELM	83.0
GWO-HKELM	80.0
Isomap-INGO-HKELM	86.7
Isomap-INGO-KELM	84.0
Isomap-GWO-SVM	82.7
Adaboost-Isomap-GWO-SVM	88.0
Adaboost-Isomap-INGO-KELM	93.3
Adaboost-Isomap-INGO-HKELM	96.0

7 实际应用测试

为验证 Adaboost-INGO-HKELM 模型对变压器故障诊断的实际应用效果, 采取国家电网某供电公司东郊 66 kV 变电站于 2023 年 8 月 31 日离线变压器绝缘油样品记录报告单的数据, 如图 19 所示。

将两组故障类型为正常的的数据替换为测试组数据集的第 30、40 个数据后, 测试结果如图 20 所示, 第 30、40 个数据实际故障类别为正常, 从图中可以看出预测分类与实际分类相同, 均为正常状态, 与供电公司样品记录报告结果一致。

一级单位	变电检修工区	
二级单位	东郊66kV变电站	东郊66kV变电站
设备名称	油桶	本体
样品编号/文件名	自动_东郊66kV变电站_油桶_2023-08-31_12584E_3_4_4	自动_东郊66kV变电站_本体_2023-08-31_12585D_4_1_4
分析日期	2023-08-31	2023-08-31
备注		
相别		
CH ₄ (uL/L)	0.76	1.61
C ₂ H ₄ (uL/L)	0.01	0.10
C ₂ H ₆ (uL/L)	0.03	0.23
C ₂ H ₂ (uL/L)	0.01	0.02
H ₂ (uL/L)	0.32	12.41
C ₀ (uL/L)	1.13	48.33
C ₀ 2 (uL/L)	292.76	236.71
Voc (uL/L)	0.81	1.96
注意值	C ₂ H ₂ <=0uL/L, H ₂ <=150uL/L, Voc<=150uL/L	C ₂ H ₂ <=5uL/L, H ₂ <=150uL/L, Voc<=150uL/L
分析意见:	未发现异常!	未发现异常!
报告时间:	2023-08-31	

图 19 某供电公司样品记录报告单

Fig. 19 A sample record report sheet of a power supply company

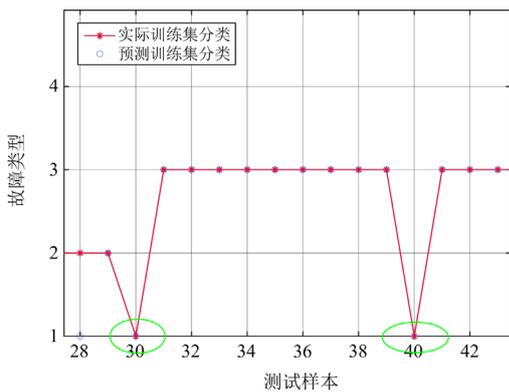


图 20 实际应用模型测试结果

Fig. 20 Application model test results in practice

8 结论

针对电力系统变压器故障数据较为复杂且故障诊断精度低的问题, 首先对原始数据不平衡问题进行 ADASYN 处理, 再进行数据特征扩充, 并利用

Isomap 对训练数据进行降维。其次对核极限学习机的核函数采用双核函数线性结合构成混合核极限学习机, 再利用多种策略改进的 INGO 算法来寻优 HKELM 模型参数, 以提升 HKELM 模型的分类精度和其泛化性能, 经测试, 准确率达到 86.7%。

利用 Adaboost 集成算法的优势, 根据上一个弱分类器的误差来调整样本数据的权重分布, 而下一个弱分类器在训练学习时更加关注误差较大的训练样本。在训练的过程中逐步提升弱分类器 HKELM 的准确性, 并通过训练样本进行权重调整, 从而调整弱分类器权重, 加权组合成强分类器, 构建 Adaboost-INGO-HKELM 模型, 最终测试准确率提升至 96%。

所提模型与 Adaboost-Isomap-GWO-SVM 模型相比, 其准确率提高了 8%, 与 Adaboost-Isomap-INGO-KELM 模型相比, 准确率提高了 2.7%。说明改进的 INGO 算法与改进的对 KELM 以及同时结合 Adaboost 集成算法均对变压器故障辨识体现出良好的效果。最后用近期某供电站离线色谱分析测试数据结果进行测试, 验证了本文模型在变压器故障辨识应用上有着良好性能。

参考文献

- [1] 段中兴, 周孟, 宋婕菲, 等. 基于 ISFLA-SVM 的变压器故障诊断[J]. 高压电器, 2023, 59(1): 126-133. DUAN Zhongxing, ZHOU Meng, SONG Jiefei, et al. Fault diagnosis of transformer based on ISFLA-SVM[J]. High Voltage Apparatus, 2023, 59(1): 126-133.
- [2] 龚泽威一, 饶桐, 王钢, 等. 基于改进粒子群优化 XGBoost 的变压器故障诊断方法[J]. 高压电器, 2023, 59(8): 61-69. GONG Zewei-yi, RAO Tong, WANG Gang, et al. Fault diagnosis method of transformer based on improved particle swarm optimization XGBoost[J]. High Voltage Apparatus, 2023, 59(8): 61-69.
- [3] 肖云波, 范菁, 张宜, 等. 基于改进粒子群算法与油中溶解气体的变压器故障诊断的研究[J]. 电子测量技术, 2021, 44(18): 122-128. XIAO Yunbo, FAN Jing, ZHANG Yi, et al. Research on transformer fault diagnosis based on improved particle swarm algorithm and dissolved gas in oil[J]. Electronic Measurement Technology, 2021, 44(18): 122-128.
- [4] 李强, 张睿智, 范晓丹, 等. 基于特征气体和比值准则的 RAPO 植物油变压器故障诊断方法[J]. 高压电器, 2023, 59(9): 294-303. LI Qiang, ZHANG Ruizhi, FAN Xiaodan, et al. Fault diagnosis method of RAPO vegetable oil transformer

- based on characteristic gas and ratio criterion[J]. *High Voltage Apparatus*, 2023, 59(9): 294-303.
- [5] 陈欢, 彭辉, 舒乃秋, 等. 基于蝙蝠算法优化最小二乘双支持向量机变压器故障诊断[J]. *高电压技术*, 2018, 44(11): 3664-3671.
CHEN Huan, PENG Hui, SHU Naiqiu, et al. Fault diagnosis of transformer based on LS-TSVM optimized by bat algorithm[J]. *High Voltage Engineering*, 2018, 44(11): 3664-3671.
- [6] 刘迪迪, 王洋, 刘辉乾, 等. 基于 ADASYN 平衡化数据集的 POA-SVM 变压器故障诊断[J]. *电网与清洁能源*, 2023, 39(8): 36-44.
LIU Didi, WANG Yang, LIU Huiqian, et al. POA-SVM transformer fault diagnosis based on ADASYN Balanced data set[J]. *Power System and Clean Energy*, 2023, 39(8): 36-44.
- [7] 周晓华, 冯雨辰, 陈磊, 等. 改进秃鹰搜索算法优化 SVM 的变压器故障诊断研究[J]. *电力系统保护与控制*, 2023, 51(8): 118-126.
ZHOU Xiaohua, FENG Yuchen, CHEN Lei, et al. Transformer fault diagnosis based on SVM optimized by bald eagle search algorithm[J]. *Power System Protection and Control*, 2023, 51(8): 118-126.
- [8] 余长厅, 黎大健, 汲胜昌, 等. 基于振动噪声及 BP 神经网络的变压器故障诊断方法研究[J]. *高压电器*, 2020, 56(6): 256-261.
YU Changting, LI Dajian, JI Shengchang, et al. Research on transformer fault diagnosis method based on vibration noise and BP neural network[J]. *High Voltage Apparatus*, 2020, 56(6): 256-261.
- [9] 何宁辉, 朱洪波, 李秀广, 等. 基于贝叶斯网络和假设检验的变压器故障诊断[J]. *电力科学与技术学报*, 2021, 36(6): 20-27.
HE Ninghui, ZHU Hongbo, LI Xiuguang, et al. Transformer fault diagnosis based on Bayesian networks and hypothesis testing[J]. *Journal of Electric Power Science and Technology*, 2021, 36(6): 20-27.
- [10] 王艳, 李伟, 赵洪山, 等. 基于油中溶解气体分析的 DBN-SSAELM 变压器故障诊断方法[J]. *电力系统保护与控制*, 2023, 51(4): 32-42.
WANG Yan, LI Wei, ZHAO Hongshan, et al. Transformer DGA fault diagnosis method based on DBN-SSAELM[J]. *Power System Protection and Control*, 2023, 51(4): 32-42.
- [11] 谢国民, 刘东阳, 刘明. 多策略改进 MPA 算法与 HKELM 的变压器故障辨识[J]. *电子测量与仪器学报*, 2023, 37(4): 172-182.
XIE Guomin, LIU Dongyang, LIU Ming, Transformer fault identification based on multi-strategy improved MPA algorithm and HKELM[J]. *Journal of Electronic Measurement and Instrumentation*, 2023, 37(4): 172-182.
- [12] 赵平, 赵健, 吴姗, 等. 多变量相关性 Gentle Adaboost 的风机叶片覆冰故障诊断[J]. *控制工程*, 2023, 27(9): 1-7.
ZHAO Ping, ZHAO Jian, WU Shan, et al. Fault diagnosis of fan blade icing based on multiple variables correlation gentle Adaboost[J]. *Control Engineering*, 2023, 27(9): 1-7.
- [13] 商立群, 李洪波, 侯亚东, 等. 基于 VMD-ISSA-KELM 的短期光伏发电功率预测[J]. *电力系统保护与控制*, 2022, 50(21): 138-148.
SHANG Liqun, LI Hongbo, HOU Yadong, et al. Short-term photovoltaic power generation prediction based on VMD-ISSA-KELM[J]. *Power System Protection and Control*, 2022, 50(21): 138-148.
- [14] 王爽, 罗倩, 唐波, 等. 考虑样本类内不平衡的 CHPOA-DBN 变压器故障诊断方法[J]. *中国电力*, 2023, 56(10): 133-144.
WANG Shuang, LUO Qian, TANG Bo, et al. CHPOA-DBN transformer fault diagnosis method considering sample within-class imbalance[J]. *Electric Power*, 2023, 56(10): 133-144.
- [15] 李亮, 范瑾, 闫林, 等. 基于混合采样和支持向量机的变压器故障诊断[J]. *中国电力*, 2021, 54(12): 150-155.
LI Liang, FAN Jin, YAN Lin, et al. Transformer fault diagnosis based on hybrid sampling and support vector machines[J]. *Electric Power*, 2021, 54(12): 150-155.
- [16] 张宏杰, 陈贵凤, 闫宏伟, 等. 基于 SMOTE 与 Bayes 优化的 LSTM 网络变压器故障诊断[J]. *中国电力*, 2023, 56(10): 164-170.
ZHANG Hongjie, CHEN Guifeng, YAN Hongwei, et al. Fault diagnosis of LSTM network transformer based on SMOTE and Bayes optimization[J]. *Electric Power*, 2023, 56(10): 164-170.
- [17] 陈法法, 汤宝平, 苏祖强. 基于等距映射与加权 KNN 的旋转机械故障诊断[J]. *仪器仪表学报*, 2013, 34(1): 215-220.
CHEN Fafa, TANG Baoping, SU Zuqiang. Rotating machinery fault diagnosis based on isometric mapping and weighted KNN[J]. *Chinese Journal of Scientific Instrument*, 2013, 34(1): 215-220.
- [18] 甘团杰, 郑建涵, 张艳, 等. 基于领域自适应的变压器状态识别[J]. *智慧电力*, 2022, 50(2): 82-88.
GAN Tuanjie, ZHENG Jianhan, ZHANG Yan, et al. Transformer condition identification based on domain adaptation[J]. *Smart Power*, 2022, 50(2): 82-88.
- [19] 谭增强, 牛拥军, 李元昊, 等. 基于麻雀算法和深度极限学习机的 NO_x 预测研究[J]. *中国电机工程学报*, 2022, 42(增刊 1): 187-192.

TAN Zengqiang, NIU Yongjun, LI Yuanhao, et al. Research on NOx prediction of thermal power plant based on SSA-DELMM[J]. Proceedings of the CSEE, 2022, 42(S1): 187-192.

[20] 王春明, 朱永利. 基于NSGA2优化正则极限学习机的变压器油色谱故障诊断[J]. 高压电器, 2020, 56(9): 210-215.

WANG Chunming, ZHU Yongli. Transformer oil chromatograph fault diagnosis based on regularized extreme learning machine optimized by NSGA2[J]. High Voltage Apparatus, 2020, 56(9): 210-215.

[21] XU X, ZHANG T, QIU Z, et al. A method based on NGO-HKELM for the autonomous diagnosis of semiconductor power switch open-circuit faults in three-phase grid-connected photovoltaic inverters[J]. Sustainability, 2023, 15(12): 9588-9613.

[22] KARI T, HE Z, ROUZI A, et al. Power transformer fault diagnosis using random forest and optimized kernel extreme learning machine[J]. Intelligent Automation and Soft Computing, 2023, 37(1): 691-705.

[23] 章涛, 陈勇旗, 廖紫洋, 等. 一种北方苍鹰参数优化的VMD-MRE轴承故障诊断方法[J]. 机械科学与技术, 2023, 26(4): 1-8.

ZHANG Tao, CHEN Yongqi, LIAO Ziyang, et al. A VMD-MRE bearing fault diagnosis method for northern goshawk parameter optimization[J]. Mechanical Science and Technology, 2023, 26(4): 1-8.

[24] MAHMOUD A, SEHIEMY R A, HASSANIEN H M, SAAD B, et al. Photovoltaic model parameters identification using northern Goshawk optimization algorithm[J]. Energy, 2023, 262: 125522.

[25] XIANZHI H, GANG H, et al. An enhanced northern Goshawk optimization algorithm and its application in practical optimization problems[J]. Mathematics, 2022, 10(22): 4383-4390.

[26] 何静, 高见, 张昌凡. 集成自适应变异混沌松鼠搜索和LSTM算法的RUL预测方法及应用[J]. 电子测量与仪器学报, 2023, 37(5): 88-97.

HE Jing, GAO Jian, ZHANG Changfan. Prediction of RUL and application of integrated squirrel search algorithm with adaptive mutation chaos for LSTM[J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(5): 88-97.

[27] CHUNYU P, NAN W, CHAOFENG Y. Rebalance weights AdaBoost-SVM model for imbalanced data[J]. Computational Intelligence and Neuroscience, 2023, 26(1): 1055-1081.

[28] 杨国华, 张鸿皓, 郑豪丰, 等. 基于相似日聚类和IHGWO-WNN-AdaBoost模型的短期光伏功率预测[J]. 高电压技术, 2021, 47(4): 1185-1194.

YANG Guohua, ZHANG Honghao, ZHENG Haofeng, et al. Short-term photovoltaic power forecasting based on similar weather clustering and IHGWO-WNN-AdaBoost model[J]. High Voltage Engineering, 2021, 47(4): 1185-1194.

[29] 李可军, 徐延顺, 魏本刚, 等. 基于PSO-HKELM的变压器顶层油温预测模型[J]. 高电压技术, 2018, 44(8): 2501-2508.

LI Kejun, XU Yanshun, WEI Bengang, et al. Prediction model for top oil temperature of transformer based on hybrid kernel extreme learning machine trained and optimized by particle swarm optimization[J]. High Voltage Engineering, 2018, 44(8): 2501-2508.

收稿日期: 2023-08-16; 修回日期: 2023-11-06

作者简介:

谢国民(1969—), 男, 博士, 教授, 研究方向为电力系统信息检测与故障诊断; E-mail: lngdxgm@163.com

江海洋(1999—), 男, 通信作者, 硕士研究生, 研究方向为电力系统信息检测与故障诊断。E-mail: haiyang598@qq.com

(编辑 姜新丽)