

基于混合采样与 IHBA-SVM 的变压器故障辨识方法*

谢国民 王嘉良

(辽宁工程技术大学电气与控制工程学院 葫芦岛 125105)

摘要:针对变压器故障数据的不平衡性弱化故障分类能力的问题,提出混合采样与改进蜜獾算法(IHBA)优化支持向量机(SVM)的变压器故障诊断方法。首先采用K近邻去噪、K均值聚类(K-means)与合成少数类过采样(SMOTE)对数据进行混合采样处理,以缓解诊断结果向多数类的偏移;然后使用Tent映射、轮盘赌随机搜索机制和最优个体扰动策略对传统蜜獾算法(HBA)进行改进,并使用IHBA优化SVM参数,以进一步提升变压器故障辨识能力;最后对所提方法进行算例仿真,结果显示,相较于传统的变压器故障辨识方法,采用K近邻去噪、K-means、SMOTE混合采样与IHBA-SVM相结合的故障诊断模型获得了最高的宏F1和微F1值,分别达到0.877和0.886,表明提出模型不仅具有更高的整体分类能力,且更能兼顾对少数类故障的辨识。

关键词: 变压器;故障诊断;改进蜜獾算法;平衡数据集;混合采样

中图分类号: TM407;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 hybrid sampling and IHBA-SVM

Xie Guomin Wang Jialiang

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

Abstract: Aiming at the problem that the unbalance of transformer fault data weakens the ability of fault classification, a transformer fault diagnosis method based on hybrid sampling and improved honey badger algorithm (IHBA) and optimized support vector machine (SVM) is proposed. Firstly, K-nearest neighbor denoising, K-means and SMOTE are used for hybrid sampling of data to alleviate the shift of diagnosis results to the majority class. Then, the traditional honey badger algorithm (HBA) is improved by using tent mapping, roulette random search mechanism and optimal individual perturbation strategy, and the SVM parameters are optimized by IHBA to further improve the transformer fault identification ability. Finally, the simulation results of the proposed method show that, compared with the traditional transformer fault identification method, the fault diagnosis model combining K-Nearest Neighbor denoising, K-means, SMOTE hybrid sampling and IHBA-SVM obtains the highest macro $F1$ and micro $F1$ values, reaching 0.877 and 0.886 respectively, which indicates that the proposed model not only has higher overall classification ability, but also can better identify minority faults.

Keywords: transformer; fault diagnosis; improved honey badger algorithm; balanced data set; hybrid sampling

0 引言

变压器担负着电能的转换与传输的重要任务,其运行状况对电力系统的安稳运行至关重要^[1]。若不能在其发生故障时及时进行诊断,电力系统的安全则无法得到有效保证,在影响电能稳定供应的同时还会给国家带来

重大的经济损失并威胁人们的生命财产安全。因此,必须确保高效的故障诊断,从而保证电网的安全与稳定的电力供应。

分布均匀的训练数据是保障分类器具有较强泛化能力的关键^[2]。然而,变压器故障发生频率低,收集的不同故障类型的样本数目差距较大,所以最终获取到的多是各类数据不平衡的小样本数据集。在变压器故障样本类

别具有较高不平衡度时,分类器若以准确率最大化为目标易导致忽视少数类样本,导致对少数类样本的辨识能力变差。重采样方法能够有效缓解训练数据不平衡引起分类结果向多数类倾斜的问题,主要包括欠采样和过采样两方面,例如 K-means 欠采样^[3]、SMOTE^[4]、Borderline-Smote^[5]、SVM SMOTE^[6]等。然而单独使用欠采样易丢失数据集重要信息,不利于变压器故障的准确分类;单独使用过采样易造成噪声扩散并模糊类别边界,导致边界样本类别难以区分。

与此同时,随着人工智能理论应用的不断进步,各种机器学习方法被应用到变压器故障检测当中,如支持向量机^[7]、神经网络^[8]、随机森林^[9]、极限学习机^[10]等。其中,支持向量机具有相对出色的鲁棒性和泛化能力,且无需大量数据,与变压器数据的小样本特点相符,因此选用支持向量机作为变压器故障分类模型。然而分类器参数的选择对诊断结果至关重要^[11]。文献[12-13]分别用鲸鱼算法(whale optimization algorithm, WOA)和粒子群算法(particle swarm optimization, PSO)优化 SVM 参数以提高分类准确率,但这些算法存在易陷入局部最优和收敛速度慢的不足,影响分类的准确性。

为解决数据不平衡以及最佳分类器参数选取的问题,提出采用 K 近邻去噪与 K-means、SMOTE 混合采样以缓解失衡训练数据造成的分类偏差;然后,为提升蜜獾算法的寻优能力,采用 Tent 映射、轮盘赌随机搜索机制和最优个体扰动策略优化其搜索机制;最后,利用 IHBA 调优 SVM 的惩罚因子 c 和核函数参数 g ,以进一步提升变压器故障分类能力。通过实验将 IHBA-SVM 与 PSO-SVM、WOA-SVM、HBA-SVM 在训练集混合采样前后所得宏 $F1$ 、微 $F1$ 值进行比较,以验证本文方法的有效性。

1 数据平衡化方法

针对变压器数据不平衡导致少数类故障识别能力弱的问题,提出 K-means、SMOTE 和 K 近邻去噪结合的方法平衡数据,以加强变压器故障的辨识能力。

K-means 聚类欠采样将数据划分为 K 个簇并保留每个聚类簇的样本中心,实现欠采样,其中 K 值等于采样后样本数目。该算法采样考虑到了数据的分布特征,避免了随机欠采样的盲目性,在剔除多数类冗余信息的同时不会破坏原始样本的分布状态,保留了具有代表性的样本。

SMOTE 在每个少数类样本 x 及其近邻样本 x_n 间线性插值以合成新的少数类样本。相比于随机过采样直接复制数据实现类别平衡,该算法降低了过拟合的风险。生成新样本 x' 的表达式如式(1)所示。

$$x' = x + rand \times (x_n - x) \quad (1)$$

式中: $rand$ 为 0~1 之间的随机数。

但是,若数据集存在噪声则会对数据的约减和扩充造成不利影响,需要剔除噪声数据。首先判断训练集内各样本 K 个近邻的情况,若 K 个近邻完全没有其同类,则将该样本识别为噪声并剔除;同时,由于少数类样本数量较小,其珍贵的边界样本易被判断为噪声误删,因此应先对多数类进行欠采样处理,且少数类设定的近邻数 K 应高于多数类。根据多次实验结果测试,在 K 近邻降噪环节多数类和少数类近邻数分别设置为 3 和 5 时,分类效果最佳。

只对多数类进行欠采样处理易丢失重要信息,导致分类效果不理想;只对少数类进行过采样易受到残留噪声的影响,造成一定的噪声扩散并生成冗余样本。因此,以数据量适中的普通类样本数为目标,采用 K-means 和 SMOTE 对多数类和少数类分别进行约减和扩充,从而在保留更多多数类样本信息的同时缓解少数类噪声的扩散和冗余样本的合成,实现各类数据的平衡。数据平衡步骤如下:

1) 多数类样本的欠采样。在 K 近邻去噪方法中,少数类由于数据量较少,其边界样本易被误判为噪声。因此,应先对多数类样本进行约减处理,降低其数量占比。采用 K 近邻去噪方法剔除多数类样本的噪声,缓解其对后续采样过程的影响,然后利用 K-means 聚类欠采样方法缩减多数类数据规模。

2) 少数类样本的过采样。在对多数类样本规模进行约减后的基础上,利用 K 近邻去噪方法剔除少数类中的噪声,缓解过采样噪声扩散对分类效果的影响,然后使用 SMOTE 扩充少数类样本数,实现类别平衡。

2 改进 HBA 算法

2.1 基本 HBA 算法

通过模拟蜜獾挖掘和寻找蜂蜜的搜索行为,Hashim 等^[14]于 2021 年提出了蜜獾算法,其基本步骤为:

1) 种群初始化

蜜獾算法通过式(2)随机初始化蜜獾个体的位置。

$$x_i = lb_i + r_1 \times (ub_i - lb_i) \quad (2)$$

式中: x_i 为第 i 个蜜獾位置; ub_i 、 lb_i 分别表示搜索空间的上、下界, r_1 为 $[0, 1]$ 内的随机数。

2) 定义嗅觉强度

嗅觉强度 I_i 与蜜獾和猎物之间的距离 d_i 以及猎物的集中程度 S 有关,表达式为:

$$I_i = r_2 \times \frac{S}{4\pi d_i^2} \quad (3)$$

$$S = (x_i - x_{i+1})^2 \quad (4)$$

$$d_i = x_{prey} - x_i \quad (5)$$

式中: r_2 为 $[0, 1]$ 的随机数。

3) 更新密度因子

密度因子 α 的值随迭代次数的增加不断减少, 以确保从勘探到开采的平稳过渡, 表达式为:

$$\alpha = C \times \exp(-t/t_{\max}) \quad (6)$$

式中: t 为当前迭代次数; t_{\max} 为最大迭代次数, C 为大于等于 1 的常数(默认为 2)。

4) 更新个体位置

蜜獾位置的更新分为挖掘阶段和蜂蜜阶段。在挖掘阶段, 蜜獾在蜂巢周围搜索, 其位置的更新依靠嗅觉强度 I_i 、蜜獾和猎物之间的距离 d_i 以及密度因子 α 。挖掘阶段表达式为:

$$x_{\text{new}} = x_{\text{prey}} + F \times \beta \times I_i \times x_{\text{prey}} + F \times r_3 \times \alpha \times d_i \times \cos(2\pi r_4) \times [1 - \cos(2\pi r_5)] \quad (7)$$

式中: r_3, r_4, r_5 均为 $[0, 1]$ 内的随机数; $\beta \geq 1$ (默认为 6), 代表蜜獾获取食物的能力; x_{prey} 为猎物位置, 即最优个体的位置; F 为方向控制参数, 能够增强蜜獾搜索的遍历性, 表达式为:

$$F = \begin{cases} 1, & r_6 \leq 0.5 \\ -1, & r_6 > 0.5 \end{cases} \quad (8)$$

式中: r_6 为 $[0, 1]$ 内的随机数。

在蜂蜜阶段, 蜜獾跟随向导鸟直接向蜂巢移动, 蜜獾基于距离信息 d_i 在 x_{prey} 附近执行搜索。在该阶段, 搜索受密度因子 α 的影响。其搜索的数学表达式为:

$$x_{\text{new}} = x_{\text{prey}} + F \times r_7 \times \alpha \times d_i \quad (9)$$

式中: r_7 为 $[0, 1]$ 内的随机数, F 由式(8)确定。

2.2 HBA 算法改进策略

为进一步加强获取更优分类器参数的能力, 采用改进 Tent 映射、轮盘赌随机搜索机制和最优个体扰动策略对其寻优过程进行改进。

1) 改进 Tent 映射种群初始化

基本蜜獾算法随机生成每只蜜獾的初始位置, 这导致蜜獾个体在搜索空间中的均匀性、遍历性不足影响寻优性能。分布均匀的初始位置遍历性强, 能够较为彻底地搜索空间中的各个位置, 降低基本蜜獾算法因随机生成初始个体而错过全局最优解的风险, 有利于提升算法的寻优精度和效率^[15-16]。混沌映射因具有遍历性、随机性等特点, 可以加强初始种群的多样性, 其中 Tent 映射能够生成更为均匀的序列^[15], 为解决其易陷入小周期和不确定周期点的不足, 采用文献^[16]的方法在表达式中加入 r_8/n , 可得 Tent 映射表达式为:

$$y_{j+1}^i = \begin{cases} \mu y_j^i + r_8/n, & y_j^i < 0.5 \\ \mu(1 - y_j^i) + r_8/n, & y_j^i \geq 0.5 \end{cases} \quad (10)$$

其中, $\mu \in (0, 2]$ 为映射控制参数, 其值与混沌程度成正比, 取 $\mu = 2$; y_j^i 为当前映射结果; r_8 为 $[0, 1]$ 内的随

机数; $i = 1, 2, \dots, n$ 为种群规模; $j = 1, 2, \dots, d$ 为混沌变量序号。通过式(10)选取初始值后, 将混沌序列逆映射到搜索空间中, 表达式为:

$$x_j^i = lb_i + y_j^i \times (ub_i - lb_i) \quad (11)$$

2) 引入轮盘赌随机搜索机制

在蜜獾算法中, 蜜獾位置始终以 x_{prey} 为基准进行更新, 容易因种群聚集在最优个体周边而过早收敛。为加强蜜獾算法的全局寻优能力, 引入 WOA^[17] 中的随机搜索思想, 根据 A 的数值决定更新策略。当满足 $|A| \geq 1$ 时执行随机搜索策略, 另选取基准个体进行更新, 否则仍以 x_{prey} 为基准更新蜜獾位置, A 的表达式为:

$$A = 2 \times m \times r - m \quad (12)$$

式中: $m = 2 - 2 \times t/t_{\max}$, 其值从 2 线性降至 0, r 为 $(0, 1)$ 内的随机数。

由式(12)及 m 的表达式可知, $|A|$ 的数值总体呈线性下降的趋势, 在迭代前期, 算法多次执行随机搜索策略, 避免种群过早聚集, 增强蜜獾个体在搜索空间中的遍历性, 提高全局搜索能力。在此阶段, 采用文献^[18]的方法依据轮盘赌原则根据适应度值选取某一个体 x_{rand} 替代 x_{prey} 以避免劣解被多次选取。此时蜜獾执行挖掘阶段和蜂蜜阶段搜索的数学表达式变为:

$$x_{\text{new}} = x_{\text{rand}} + F \times \beta \times I_i \times x_{\text{rand}} + F \times r_3 \times \alpha \times d_i \times \cos(2\pi r_4) \times [1 - \cos(2\pi r_5)] \quad (13)$$

$$x_{\text{new}} = x_{\text{rand}} + F \times r_7 \times \alpha \times d_i \quad (14)$$

此时 d_i 的表达式相应变为:

$$d_i = x_{\text{rand}} - x_i \quad (15)$$

3) 最优个体扰动

随着迭代的进行, 个体不断收敛到最优解附近, 种群同化程度较迭代前期显著增加, 这会加剧算法局部停滞风险。为加强种群摆脱局部桎梏的能力, 提出利用准反射学习和小波变异策略对蜜獾种群中最优个体进行扰动。

(1) 准反射学习

与反向解相比, 准反射解靠近全局最优解的概率更高^[19]。准反射学习根据当前解与搜索空间的位置关系产生一个准反射解, 加强种群摆脱局部最优的能力, 准反射解计算公式为:

$$x'_{\text{prey}} = \text{rand}((lb + ub)/2, x_{\text{prey}}) \quad (16)$$

式中: 准反向解 x'_{prey} 表示 $(lb + ub)/2$ 与 x_{prey} 之间均匀分布的随机数。

引入准反射学习机制有利于更加充分地探索空间, 摆脱局部最优, 从而使算法收敛精度和速度得到提升。

(2) 小波变异策略

相比于高斯变异、柯西变异等扰动方法, 小波变异还具备微调能力^[20]。随着迭代次数的增加, 其通过控制小波函数中的伸缩参数来降低小波函数的振幅以约束扰动

的范围。小波变异的表达式为:

$$x'_{prey} = \begin{cases} x_{prey} + \delta \times (ub - x_{prey}), & \delta > 0 \\ x_{prey} + \delta \times (x_{prey} - lb), & \delta \leq 0 \end{cases} \quad (17)$$

$$\delta = \frac{1}{\sqrt{a}} \times \exp\left(-\left(\frac{\varphi}{a}\right)^2 / 2\right) \times \cos\left(5 \times \left(\frac{\varphi}{a}\right)\right) \quad (18)$$

$$a = \exp\left(-\ln(g) \times \left(1 - \frac{t}{t_{max}}\right)^{\zeta_{um}} + \ln(g)\right) \quad (19)$$

式中: δ 为小波函数值; a 为尺度参数; φ 在 $[-2.5a, 2.5a]$ 内随机生成; g 为 a 的上限; ζ_{um} 为形状参数。取 $g = 10\ 000$ 、 $\zeta_{um} = 2$ 。

为提升算法寻优能力, IHBA 以等概率交替执行准反射学习和小波变异策略, 扩展算法的搜索范围, 使算法能以更高概率探索到更优解, 提升寻优精度和速度。在最优个体扰动完成后依据贪婪原则在原个体和变异个体之间保留适应度值更佳者进入下次迭代。

2.3 算法寻优性能测试

为验证 IHBA 的寻优能力, 选取 Ackley 和 Penalized1 两个多峰函数对其进行测试, 并与 PSO、WOA 和 HBA 进行对比分析。Ackley 和 Penalized1 函数表面广泛分布着大量的极小点, 能够有效测试算法摆脱局部最优的能力以及寻优速度, 已被广泛应用于优化算法的性能测试中。测试函数如表 1 所示。

测试函数	维度	搜索范围	最优解
Ackley	100	$[-32, 32]$	0
Penalized1	100	$[-50, 50]$	0

将 IHBA 算法和 HBA、WOA、PSO 进行寻优性能比较, 寻优过程如图 1 和 2 所示。

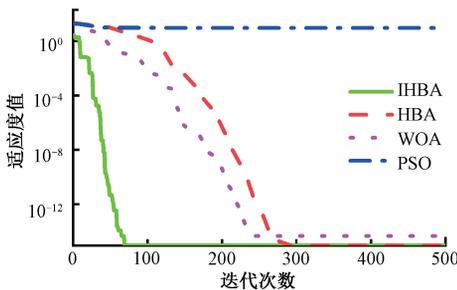


图 1 函数 Ackley 寻优对比曲线

Fig. 1 Function Ackley optimization comparison curve

如图 1 所示, 在 Ackley 函数测试中, PSO 曲线在迭代初期便陷入停滞; WOA 适应度曲线于第 238 次迭代不再下降, 与 HBA、IHBA 曲线仍有一段距离; HBA、IHBA 获得了相同的适应度, 但 IHBA 具有更快的收敛速度, 于第

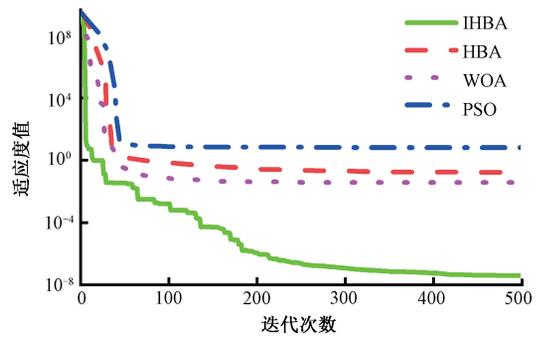


图 2 函数 Penalized1 寻优对比曲线

Fig. 2 Function Penalized1 optimization comparison curve

69 次迭代收敛, 快于 HBA 的 277 次。

如图 2 所示, 在 Penalized1 函数测试中, PSO、WOA、HBA 在迭代初期之后适应度曲线的下降极为缓慢。而 IHBA 仍能以较快的速度不断寻得更佳的适应度值, 且在迭代后期, 适应度曲线仍有小幅度的下降趋势, 说明其挣脱局部极值束缚的能力更强。

通过 Ackley 和 Penalized1 两个测试函数测试 4 种算法的寻优性能, 说明 IHBA 算法具有更佳的寻优能力。

3 基于平衡数据与 IHBA-SVM 的变压器故障诊断方法

3.1 变压器故障特征选择

当变压器处于不同状态时, 油中气体 H_2 、 CH_4 、 C_2H_6 、 C_2H_4 、 C_2H_2 含量会发生变化, 然而这 5 种气体含量较为分散, 直接输入模型会影响诊断效果。

无编码比值法能够准确表征变压器故障性质^[21], 包含 9 维数据特征, 分别为: CH_4/H_2 、 C_2H_2/C_2H_4 、 C_2H_4/C_2H_6 、 $C_2H_2/(C_1 + C_2)$ 、 $H_2/(H_2 + C_1 + C_2)$ 、 $C_2H_4/(C_1 + C_2)$ 、 $CH_4/(C_1 + C_2)$ 、 $C_2H_6/(C_1 + C_2)$ 、 $(CH_4 + C_2H_4)/(C_1 + C_2)$, 其中 C_1 为一阶碳氢化合物 CH_4 含量, C_2 为二阶碳氢化合物 C_2H_6 、 C_2H_4 、 C_2H_2 含量之和。采用上述 9 维无编码比值作为模型的输入。

3.2 变压器故障状态及编码

选取国家电网公司辽宁某电网提供的 565 组变压器状态数据作为实验样本数据, 将各类别数据按 6 : 2 : 2 的比例分为训练集, 验证集和测试集。样本类别划分及状态编号如表 2 所示。

3.3 分类结果的评价指标

训练数据不平衡时, 准确率会偏向于多数类, 单独使用准确率这一评价指标并不合理^[22], 因此还要选择能够反应少数类样本的评价指标。对于二分类问题, 将分类

表 2 实验数据集分类

Table 2 Classification of experimental data sets

运行状态	正常	低能放电	高能放电	局部放电	低温过热	中温过热	高温过热
编号	1	2	3	4	5	6	7
训练集	95	62	55	25	31	20	52
验证集	31	20	18	8	10	7	17
测试集	32	21	18	9	10	7	17

结果分为以下 4 种类型:

- 1) TP : 实际为正例, 被预测为正例的样本数。
- 2) FP : 实际为负例, 被预测为正例的样本数。
- 3) FN : 实际为正例, 被预测为负例的样本数。
- 4) TN : 实际为负例, 被预测为负例的样本数。

于是可分别得到查准率 $P = TP / (TP + FP)$, 查全率 $R = TP / (TP + FN)$, $F1 = 2 \cdot P \cdot R / (P + R)$ 。查准率和查全率分别反应分类器对样本的误判和漏判程度, $F1$ 值是前两者的调和平均数, 能够反应总体分类效果。

针对变压器多种状态的分类, 将其转化为多个二分类问题, 从而得到多分类评价指标宏 $F1(\text{macro_}F1)$ 和微 $F1(\text{micro_}F1)$ 来评价分类效果, 计算方式为:

- 1) 宏 $F1(\text{macro_}F1)$

$$\text{macro_}F1 = \frac{1}{N} \sum_{i=1}^N \frac{2 \cdot P_i \cdot R_i}{P_i + R_i} \quad (20)$$

式中: P_i 为类别 i 的查准率; R_i 为类别 i 的查全率; N 为类别数。各类别分别计算 TP_i 、 FP_i 和 FN_i , 进而计算 P_i 和 R_i , 得到宏 $F1$ 值。

- 2) 微 $F1(\text{micro_}F1)$

微 $F1$ 可通过微查准率 ($\text{micro_}P$) 和微查全率 ($\text{micro_}R$) 得到, 表达式如下:

$$\text{micro_}P = \bar{n}_{TP} / (\bar{n}_{TP} + \bar{n}_{FP}) \quad (21)$$

$$\text{micro_}R = \bar{n}_{TP} / (\bar{n}_{TP} + \bar{n}_{FN}) \quad (22)$$

$$\text{micro_}F1 = \frac{2 \cdot \text{micro_}P \cdot \text{micro_}R}{\text{micro_}P + \text{micro_}R} \quad (23)$$

式中: \bar{n}_{TP} 、 \bar{n}_{FP} 、 \bar{n}_{FN} 为各类别对应元素的均值, 其中微 $F1$ 值与准确率相等。为了能够兼顾对少数类故障辨识能力的评价并反应总体准确性, 同时选取宏 $F1$ 和微 $F1$ 作为分类效果的评价指标。

3.4 数据的平衡化处理

由表 2 数据可知, 局部放电、低温过热、中温过热训练集数据较少, 最大不平衡比达 4.75 : 1, 训练数据的不平衡会使分类结果向多数类倾斜, 因此, 利用 K 近邻去噪、K-means 和 SMOTE 混合采样平衡训练数据。少数类样本量较少, 在 K 近邻去噪算法中, 其边界样本易被错分为噪声, 故首先采用 K 近邻去噪和 K-means 约减多数类样本, 降低少数类周边的多数类样本比例; 然后, 利用 K 近邻去噪方法剔除少数类中的噪声, 以缓解过采样造成

的噪声扩散, 并使用 SMOTE 扩充少数类样本, 实现类别平衡。

3.5 基于 IHBA 优化的 SVM

为提升变压器故障分类能力, 采用 IHBA 对 SVM 的 c 、 g 参数进行优化。设置种群数为 30, 迭代 100 次, 核函数和惩罚因子的搜索范围为 (0, 100)。将平衡的训练数据输入 SVM, 利用改进蜜獾算法对 c 、 g 参数调优。

故障诊断流程如图 3 所示。

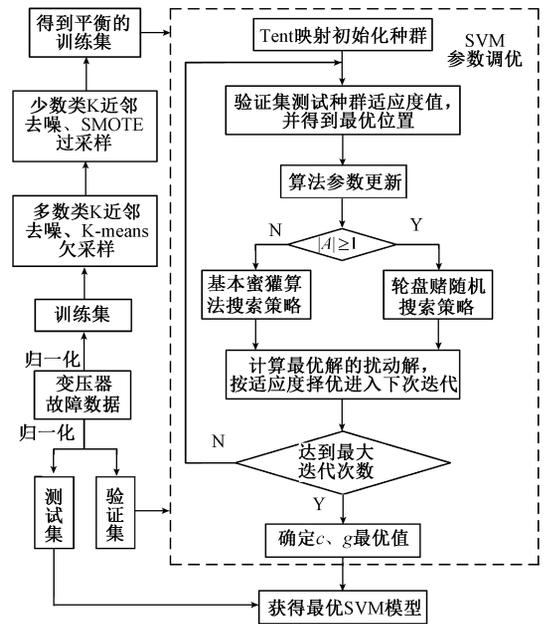


图 3 故障诊断流程

Fig. 3 Fault diagnosis process

4 故障诊断效果对比实验

4.1 分类效果评价指标

1) 首先测试在未对训练数据进行平衡化处理的情况下, 直接利用 PSO、WOA、HBA 和 IHBA 对支持向量机 c 、 g 参数进行寻优, 在测试集上得到的分类结果如图 4~7 所示。

通过图 4~7 可以看出: PSO-SVM、WOA-SVM 和 HBA-SVM 的宏 $F1$ 值分别为 0.720、0.741 和 0.744, 微 $F1$ 值分别为 0.737、0.763 和 0.754, 分类效果差, 尤其是对局部放电、低温过热和中温过热 3 个类别的辨识能力更不理想。IHBA 具有更强的寻优能力, 使 IHBA-SVM 得到了相对更高的宏 $F1$ 和微 $F1$ 值, 分别为 0.779 和 0.798, 但不平衡的训练数据造成其分类效果仍然比较局限。

2) 对训练数据进行平衡化处理, 再利用各寻优算法调优支持向量机 c 、 g 参数, 得到的分类结果如图 8~

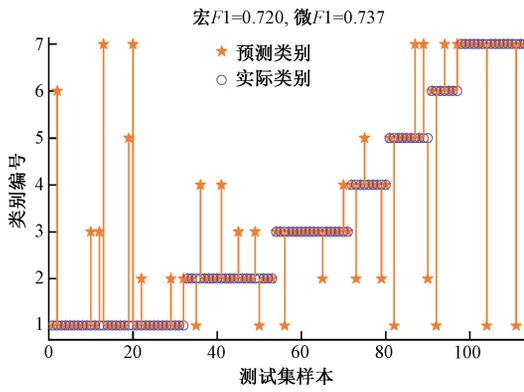


图 4 未进行数据平衡 PSO-SVM 诊断结果

Fig. 4 PSO-SVM diagnostic results without data balancing

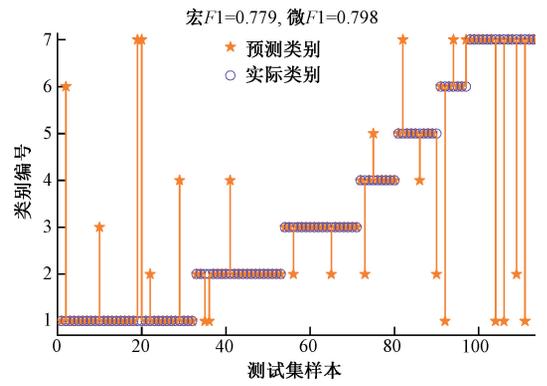


图 7 未进行数据平衡 IHBA-SVM 诊断结果

Fig. 7 IHBA-SVM diagnostic results without data balancing

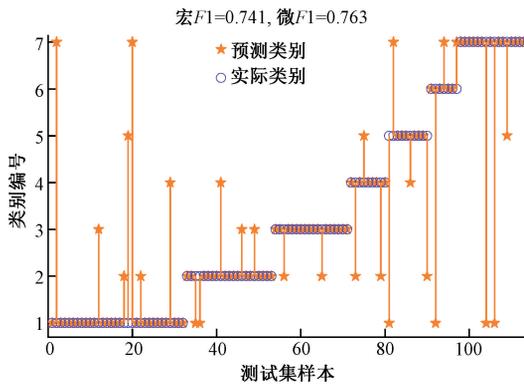


图 5 未进行数据平衡 WOA-SVM 诊断结果

Fig. 5 WOA-SVM diagnostic results without data balancing

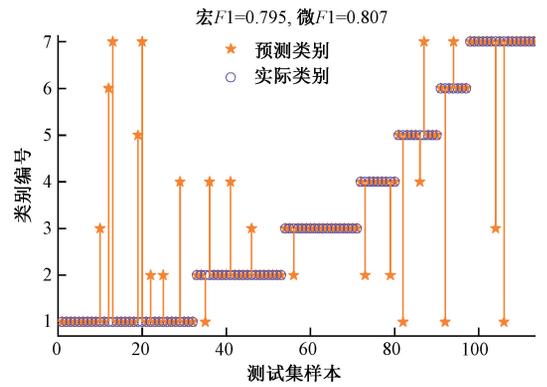


图 8 数据平衡后 PSO-SVM 诊断结果

Fig. 8 PSO-SVM diagnostic results after data balancing

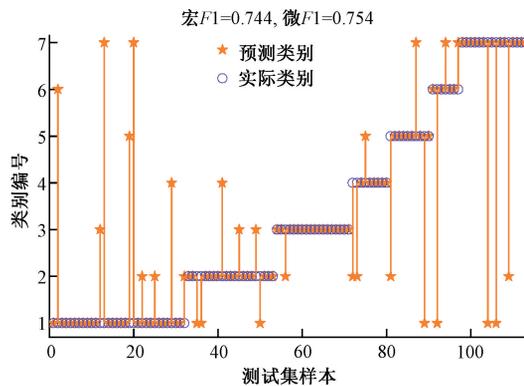


图 6 未进行数据平衡 HBA-SVM 诊断结果

Fig. 6 HBA-SVM diagnostic results without data balancing

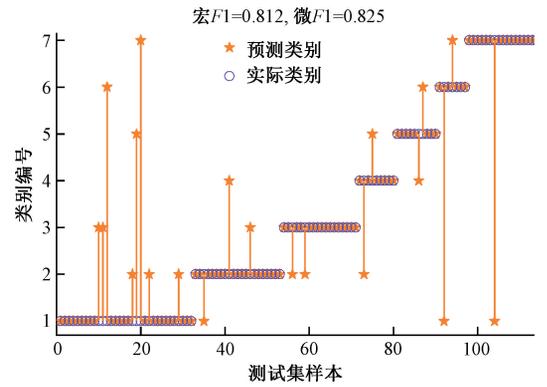


图 9 数据平衡后 WOA-SVM 诊断结果

Fig. 9 WOA-SVM diagnostic results after data balancing

11 所示。

通过图 8~11 可知,经过平衡化处理后的训练集数据对少数类样本的识别情况有明显改善,相比于 PSO-SVM、WOA-SVM 和 HBA-SVM, IHBA-SVM 模型得到的分类效果最好,宏 F1 和微 F1 值为 0.877 和 0.886,比数据平衡前分别提升了 9.8% 和 8.8%。

4.2 不同采样方式分类效果对比

为验证基于 K 近邻降噪、K-means 和 SMOTE 相结合的数据平衡方法的有效性,将其与:1) 数据未平衡;2) K-means 欠采样;3) SMOTE 过采样;4) Borderline-SMOTE;5) SVM SMOTE 过采样进行对比,各方法平衡数据得到的宏 F1 和微 F1 结果如表 3 所示。

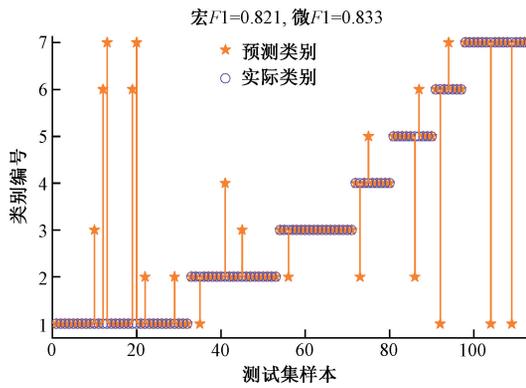


图 10 数据平衡后 HBA-SVM 诊断结果

Fig. 10 HBA-SVM diagnostic results after data balancing

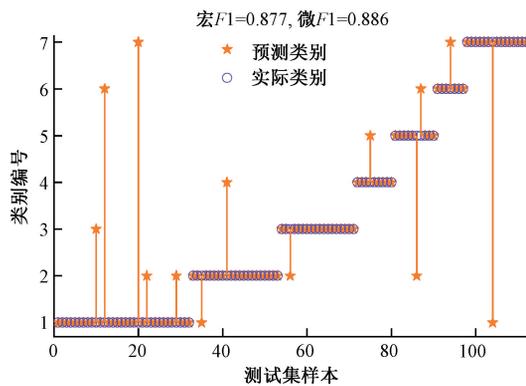


图 11 数据平衡后 IHBA-SVM 诊断结果

Fig. 11 IHBA-SVM diagnostic results after data balancing

表 3 不同平衡数据方法结果对比

Table 3 Comparison of results of different balanced data methods

平衡数据方法	宏 F1 值	微 F1 值
未平衡	0.779	0.798
K-means 欠采样	0.827	0.807
SMOTE	0.796	0.807
Borderline-SMOTE	0.813	0.825
SVM SMOTE	0.829	0.833
K 近邻去噪, K-means, SMOTE 混合采样	0.877	0.886

由对比试验表 3 可知,使用本文方法平衡训练数据,分类结果更佳。这是由于算法在欠、过采样前先通过 K 近邻降噪缓解噪声扩散及边界模糊对后续采样过程的不利影响,进而采用 K-means 和 SMOTE 平衡训练数据,既避免了随机欠采样的盲目性,同时又剔除更少的多数类样本,使分类器能从样本中学习到更多有效的多数类信息,且过采样减少了噪声和冗余样本的生成,获得了更好的分类效果。

5 结论

针对变压器故障样本失衡造成少数类故障识别能力弱的问题,本文提出一种混合采样与 IHBA-SVM 的变压器故障辨识方法,结论如下:

1) 针对变压器数据不平衡的问题,通过 K 近邻降噪以及 K-means 与 SMOTE 相结合的数据平衡化处理,剔除噪声数据;同时合理选择了多数类样本,提高了模型对变压器不平衡数据的分类能力。

2) 为提升蜜獾算法的寻优能力,采用 Tent 映射、轮盘赌随机搜索机制和最优个体扰动策略对其优化,增强了蜜獾种群在搜索空间的遍历性,有效提高了蜜獾算法的寻优精度和速度。通过与 PSO、WOA、HBA 等寻优能力对比,证明了改进蜜獾算法的优越性。

3) 经对比实验分析,提出的基于 K 近邻降噪、K-means 和 SMOTE 的混合采样与 IHBA-SVM 的变压器故障辨识模型的宏 F1 和微 F1 值分别达到 0.877 和 0.886,相较于 PSO-SVM、WOA-SVM 和 HBA-SVM,宏 F1 分别提升了 8.2%、6.5% 和 5.6%,微 F1 分别提升了 7.9%、6.1% 和 5.3%;相较于未进行数据平衡、K-means、SMOTE、Borderline-SMOTE 和 SVM SMOTE 平衡数据的 IHBA-SVM 的诊断结果,宏 F1 分别提升了 9.8%、5%、8.1%、6.4% 和 4.8%,微 F1 分别提升了 8.8%、7.9%、7.9%、6.1% 和 5.3%,表明模型泛化能力较强,具有较高的故障辨识能力。

参考文献

[1] 王雨虹, 王志中, 付华, 等. 多策略改进麻雀算法与 BiLSTM 的变压器故障诊断研究[J]. 仪器仪表学报, 2022, 43(3): 87-97.
WANG Y H, WANG ZH ZH, FU H, et al. Research on transformer fault diagnosis based on the improved multi-strategy sparrow algorithm and BiLSTM [J]. Chinese Journal of Scientific Instrument, 2022, 43 (3): 87-97.

[2] 葛磊蛟, 廖文龙, 王煜森, 等. 数据不足条件下基于改进自动编码器的变压器故障数据增强方法[J]. 电工技术学报, 2021, 36(S1): 84-94.
GE L J, LIAO W L, WANG Y S, et al. Data augmentation method for transformer fault based on improved auto-encoder under the condition of insufficient data[J]. Transactions of China Electrotechnical Society, 2021, 36 (S1): 84-94.

[3] 武森, 刘露, 卢丹. 基于聚类欠采样的集成不平衡数据分类算法[J]. 工程科学学报, 2017, 39(8): 1244-1253.
WU S, LIU L, LU D. Imbalanced data ensemble

- classification based on cluster-based under-sampling algorithm [J]. Chinese Journal of Engineering, 2017, 39(8): 1244-1253.
- [4] 谢桦, 陈俊星, 赵宇明, 等. 基于 SMOTE 和决策树算法的电力变压器状态评估知识获取方法[J]. 电力自动化设备, 2020, 40(2): 137-142,1.
- XIE H, CHEN J X, ZHAO Y M, et al. Knowledge acquisition method of power transformer condition assessment based on SMOTE and decision tree algorithm [J]. Electric Power Automation Equipment, 2020, 40 (2): 137-142,1.
- [5] 韩笑, 王新迎, 韩帅, 等. 基于不均衡数据集学习的大型电力变压器状态评价方法[J]. 电网技术, 2021, 45(1): 107-114.
- HAN X, WANG X Y, HAN SH, et al. Ensemble learning method for large-scale power transformer status evaluation based on imbalanced data[J]. Power System Technology, 2021, 45 (1): 107-114.
- [6] 刘云鹏, 和家慧, 许自强, 等. 基于 SVM SMOTE 的电力变压器故障样本均衡化方法[J]. 高电压技术, 2020, 46(7): 2522-2529.
- LIU Y P, HE J H, XU Z Q, et al. Equalization method of power transformer fault sample based on SVM SMOTE [J]. High Voltage Engineering, 2020, 46 (7): 2522-2529.
- [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. Study of transformer fault diagnosis based on improved sparrow search algorithm optimized support vector machine[J]. Journal of Electronic Measurement and Instrumentation, 2021, 35(3): 123-129.
- [8] ZHOU Y C, YANG X H, TAO L Y, et al. Transformer fault diagnosis model based on improved gray wolf optimizer and probabilistic neural network[J]. Energies, 2021:14.
- [9] 曹伟嘉, 杨留方, 徐天奇, 等. 基于精英反策略麻雀搜索优化随机森林的变压器故障诊断[J]. 国外电子测量技术, 2022, 41(2): 138-143.
- CAO W J, YANG L F, XU T Q, et al. Transformer fault diagnosis based on counter strategy sparrow search algorithm optimized random forest [J]. Foreign Electronic Measurement Technology, 2022, 41 (2): 138-143.
- [10] 吴杰康, 覃炜梅, 梁浩浩, 等. 基于自适应极限学习机的变压器故障识别方法[J]. 电力自动化设备, 2019, 39(10): 181-186.
- WU J K, QIN W M, LIANG H H, et al. Transformer fault identification method based on self-adaptive extreme learning machine [J]. Electric Power Automation Equipment, 2019, 39 (10): 181-186.
- [11] 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 Y B, FAN J, ZHANG Y, 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.
- [12] 安国庆, 史哲文, 马世峰, 等. 基于 RF 特征优选的 WOA-SVM 变压器故障诊断[J]. 高压电器, 2022, 58(2): 171-178.
- AN G Q, SHI ZH W, MA SH F, et al. Fault diagnosis of WOA-SVM transformer based on RF feature optimization[J]. High Voltage Apparatus, 2022, 58 (2): 171-178.
- [13] 黄新波, 王享, 田毅, 等. 基于 PSO-ELM 融合动态加权 AdaBoost 的变压器故障诊断方法[J]. 高压电器, 2020, 56(5): 39-46.
- HUANG X B, WANG X, TIAN Y, et al. Transformer fault diagnosis algorithm based on PSO-ELM fusion dynamically weighted AdaBoost [J]. High Voltage Apparatus, 2020, 56 (5): 39-46.
- [14] HASHIM F A, HOUSSEIN E H, HUSSAIN K, et al. Honey badger algorithm: New metaheuristic algorithm for solving optimization problems [J]. Mathematics and Computers in Simulation, 2021(2) : 84-110.
- [15] 张达敏, 陈忠云, 辛梓芸, 等. 基于疯狂自适应的樽海鞘群算法 [J]. 控制与决策, 2020, 35 (9): 2112-2120.
- ZHANG D M, CHEN ZH Y, XIN Z Y, et al. Salp swarm algorithm based on craziness and adaptive [J]. Control and Decision, 2020, 35 (9): 2112-2120.
- [16] 张文军, 富立友. 含运行协调性并改进差分进化算法的机组组合[J]. 电力系统及其自动化学报, 2022, 34(4): 135-142.
- ZHANG W J, FU L Y. Unit commitment with operation coordination and improved differential evolution [J]. Proceedings of the CSU-EPSA, 2022, 34 (4): 135-142.
- [17] MIRJALILI S, LEWIS A. The whale optimization algorithm[J]. Advances in Engineering Software, 2016, 95: 51-67.
- [18] 刘景森, 郑智远, 李煜. 一种交互演化改进鲸鱼算法及其收敛性分析[J]. 控制与决策, 2023, 38 (1): 75-83.
- LIU J S, ZHENG ZH Y, LI Y. An interactive

evolutionary improved whale algorithm and its convergence analysis [J]. Control and decision making, 2023, 38(1): 75-83.

- [19] FAN Q, CHEN Z J, XIA Z H. A novel quasi-reflected Harris hawks optimization algorithm for global optimization problems [J]. Soft Computing, 2020, 24(19): 14825-14843.
- [20] 李苗苗. 蝙蝠算法的改进及其在参数估计中的应用[D]. 西安:西安理工大学, 2021.
LI M M. Improvement and application in parameter estimation of bat algorithm [D]. Xi'an: Xi'an University of Technology, 2021.
- [21] 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.
- [22] 李敏波, 董伟伟. 面向不平衡数据集的汽车零部件质量预测[J]. 中国机械工程, 2022, 33(1): 88-96.
LI M B, DONG W W. Quality prediction of automotive parts for imbalanced datasets [J]. China Mechanical Engineering, 2022, 33(1): 88-96.

作者简介



谢国民, 1991 年于辽宁工业大学获得学士学位, 2003 年于辽宁工程技术大学获得硕士学位, 2012 年于辽宁工程技术大学获得博士学位, 现为辽宁工程技术大学副教授, 主要研究方向为电力系统故障诊断与预测控制。

E-mail: lngdxgm@163.com

Xie Guomin received his B. Sc. degree from Liaoning University of Technology in 1991, M. Sc. degree from Liaoning Technical University in 2003 and Ph. D. degree from Liaoning Technical University in 2012, respectively. Now he is an associate professor in Liaoning Technical University. His main research interests include power system fault diagnosis and predictive control.



王嘉良 (通信作者), 2020 年于辽宁工程技术大学获得学士学位, 现为辽宁工程技术大学研究生, 主要研究方向为电力系统故障诊断与预测控制。

E-mail: 1346984868@qq.com

Wang Jialiang (Corresponding author) received his B. Sc. degree from Liaoning Technical University in 2020. Now he is a M. Sc. candidate of Liaoning Technical University. His main research interests include power system fault diagnosis and predictive control.