

基于 RF 和改进蜜獾优化的牵引变压器在线故障预测^{*}迟青光¹ 蒋瑞培¹ 闫东旭²

(1. 大连交通大学电气工程学院 大连 116028; 2. 中车大连机车车辆有限公司机车开发部 大连 116028)

摘要: 为提高牵引变压器在线故障预测准确率,提出了一种基于随机森林特征优选和改进蜜獾优化算法的在线故障诊断方法。首先,运用 SMOTE 算法对数据集进行均衡化处理,进而通过无编码比值法对故障诊断样本进行扩充;其次,通过随机森林对特征向量集合进行重要度排序,分别输入极限学习机、支持向量机和长短期记忆神经网络,得到最佳基础模型和特征个数组合;然后,结合 Tent 混沌映射策略、改进控制因子和小孔成像策略对蜜獾优化算法进行改进,得到改进蜜獾优化算法;最后,利用改进蜜獾优化算法和最佳基础模型及特征个数相结合,有效解决了基础模型中超参数设置问题。实验结果表明,与其他优化算法比较,改进蜜獾优化算法在寻优能力、稳定性和收敛速度上均有明显提升;所提牵引变压器在线故障预测模型的故障诊断准确率为 96.05%,相比于 HBA-LSTM,准确率提高了 2.44%,验证了所提方法的有效性。

关键词: 牵引变压器;随机森林;SMOTE;蜜獾优化算法;长短期记忆神经网络

中图分类号: TN912.34 **文献标识码:** A **国家标准学科分类代码:** 510.20

Online fault prediction of traction transformer based on RF and improved honey badger optimization

Chi Qingguang¹ Jiang Ruipei¹ Yan Dongxu²

(1. School of Electrical Engineering, Dalian Jiaotong University, Dalian 116028, China;

2. Locomotive Development Department, CRRC Dalian Locomotive & Rolling Stock Co., Ltd., Dalian 116028, China)

Abstract: To improve the accuracy of online fault prediction for traction transformers, an online fault diagnosis method based on random forest feature optimization and improved honey badger optimization algorithm is proposed. Firstly, the SMOTE algorithm is used for data balancing processing, and the uncoded ratio method is adopted to expand the fault diagnosis; secondly, the feature vector set is ranked by importance using RF, and then input into the Extreme Learning Machine, Support Vector Machine, and Long Short Term Memory Neural Network to obtain the optimal combination of the base model and the number of features; then, the honey badger optimization algorithm was improved by combining Tent chaotic mapping strategy, improved control factor, and pinhole imaging strategy, and compared with other optimization algorithms to demonstrate its effectiveness in optimization ability, stability, and convergence speed; finally, by combining the improved honey badger optimization algorithm with the optimal base model and number of features, the problem of hyperparameter setting in the base model was effectively solved. The experimental results show that the fault diagnosis accuracy of the proposed online fault prediction model for traction transformers is 96.05%, which is 2.44% higher than that of HBA-LSTM, verifying the effectiveness of the proposed method.

Keywords: traction transformer; random forest; SMOTE; honey badger algorithm; long short term memory neural network

0 引言

近年来在轨道交通领域,随着城市化进程的加快和公

共交通系统的不断扩展,确保电网能够安全稳定地向用户供电变得尤为重要。在输变电过程中,变电站扮演着至关重要的角色,而其中的轨道交通牵引变压器作为核心组件

之一,其正常运行直接关系到整个电网及轨道交通系统的可靠性与效率。这种特殊类型的变压器不仅需要满足一般牵引变压器的要求,还要适应高频次启动、制动等复杂工况下的特殊需求,以保证列车运行的安全性和平稳性。因此,及时准确地发现并诊断出牵引变压器存在的任何问题,对于制定有效的维修策略来说至关重要^[1-3]。目前有多种方法可用于检测和识别变压器故障,比如局部放电测试^[4]、绝缘电阻测量^[5]以及油中溶解气体分析(dissolved gas analysis, DGA)^[6]。其中,DGA方法因其操作简便、数据完整易处理且不受外部电磁干扰等优点,在牵引变压器的状态监控与故障预测^[7]方面得到了广泛应用。

目前,基于牵引变压器故障诊断方法主要分为传统比值法^[8]和人工智能法^[9]。比值法通过设定阈值和边界对故障类型进行判断,具体简单高效的优点,但在识别某些故障时,该方法界限过于绝对,不能完全反映牵引变压器故障状态。随着人工智能技术的快速发展,文献[10]将改进引力搜索算法与与支持向量机(support vector machine, SVM)相结合,可有效提高故障诊断精度;文献[11]提出了一种将贝叶斯网络与黏菌优化算法相结合用于牵引变压器故障诊断,所提方法具有较高的故障诊断精度,训练集和测试集精度分别达到 98.2% 和 97.14%;文献[12]提出了一种将灰狼算法和最小二乘支持向量机结合用于牵引变压器故障诊断,所提方法准确率最优;文献[13]提出了一种改进灰狼优化算法与残差神经网络相结合的牵引变压器故障诊断方法,所提模型能有效解决变压器数据集不平衡问题,克服了深度学习模型训练不稳定的缺陷。尽管上述方法操作简便,却依赖于大量数据以供模型训练。在处理数据集不平衡的情况时,该法倾向于过度关注多数类样本,从而忽视少数类样本,这可能导致牵引变压器故障分类的失误。鉴于牵引变压器故障数据集的典型不平衡特性,必须采用采样技术来平衡数据集。常见采样方法包括欠采样、过采样和混合采样^[14]。欠采样通过减少多数类样本数据使其与少数类样本数量达到平衡;过采样则是增加少数类样本数据使数据平衡。文献[15]采用欠采样方法达到数据平衡,但数据规模较小,极易丢失重要信息;文献[16]通过合成少数类过采样技术(synthetic minority oversampling technique, SMOTE)算法对煤矿用变压器进行数据均衡化处理,具有改善过拟合问题的优点。由于牵引变压器数据集中样本数量较少,因此,采用过采样方法进行数据均衡处理。同时,上述模型均未能很好通过神经网络如长短期记忆神经网络(long short term memory, LSTM)等深度学习方法进行故障诊断,人为调整模型参数效率较低,无法有效保证模型效果,忽略了冗余特征对整体模型的影响。

鉴于此,本文提出了一种基于 RF 特征优选和改进蜜獾优化算法(improved honey badger algorithm, IHBA)的牵引变压器故障诊断方法。首先,通过 SMOTE 算法进行数据均衡处理,采用无编码比值法对故障诊断进行扩

充^[17];其次通过 RF 对特征向量集合进行重要度排序,分别输入极限学习机(extreme learning machine, ELM)、SVM 和 LSTM 神经网络^[18-20],得到最佳基础模型和特征个数组合;然后,通过 IHBA 优化算法对最佳基础模型中相关超参数进行寻优,解决最佳基础模型中超参数的设置问题;最后经过标准测试函数和故障预测模型对比实验,验证了所提模型在牵引变压器故障诊断上的有效性。

1 基于 RF 特征优选和 IHBA 的牵引变压器故障预测模型

1.1 随机森林算法

RF 是一种基于集成学习的分类算法,它通过多棵决策树的训练与预测集成来提高分类性能。该方法采用 Bagging 技术,通过对原始样本进行随机重复抽样来构建多个重抽样样本集。每个重抽样样本集随后用于构建独立的决策树。最终,这些决策树被组合起来,通过多数投票机制确定最终的预测结果。在这个过程中,大约 1/3 的样本被保留为袋外数据(out of bag, OOB),这些数据未参与决策树的构建。利用 OOB 数据,可以计算特征的重要程度,例如通过平均精度下降指标,从而实现对特征的优化选择。

1.2 蜜獾优化算法

蜜獾优化算法(honey badger algorithm, HBA),受蜜獾寻找食物的挖掘、采蜜等行为启发。

1) 种群初始化

在设定的范围内随机初始化蜜獾的数量和位置,如式(1)所示。

$$x_i = lb_i + r_1 \times (ub_i - lb_i) \quad (1)$$

式中: r_1 表示(0,1)之间的随机数; x_i 表示候选个体的第*i*个个体的位置; lb_i 和 ub_i 分别表示搜索空间的上下限。

2) 定义强度

蜜獾的嗅觉强度受猎物气味强度及其与蜜獾之间的距离影响。猎物气味越浓烈,蜜獾的行动速度越快;气味较弱时,其运动速度则相对减慢,具体如式(2)所示。

$$\begin{cases} I_i = r_2 \times \frac{s}{4\pi d_i^2} \\ s = (x_i - x_{i+1})^2 \\ d_i = x_{prey} - x_i \end{cases} \quad (2)$$

式中: I_i 表示猎物的气味强度; r_2 表示(0,1)之间的随机数; s 表示源强度或猎物的集中强度; d_i 表示猎物与当前蜜獾个体之间的距离; x_{prey} 表示猎物的位置。

3) 更新密度因子

密度因子 α 为保障蜜獾在探索与开发过程中的平滑转换,采用式(3)所描述的方法对参数进行适时调整,旨在降低蜜獾在搜寻食物过程中由于时间推移所产生的不确定性影响。

$$\alpha = C \times \exp\left(\frac{-t}{t_{\max}}\right) \quad (3)$$

式中: t 表示当前迭代次数; t_{\max} 表示最大迭代次数; $C \geq 1$ (一般默认取 2)。

4) 挖掘阶段

在探寻猎物的过程中,蜜獾展现出类似心形的搜索轨迹,此类轨迹的数学描述可参照式(4)。

$$x_{\text{new}} = x_{\text{prey}} + F \times \beta \times I \times x_{\text{prey}} + F \times r_3 \times \alpha \times d_i \times |\cos(2\pi r_4) \times [1 - \cos(2\pi r_5)]| \quad (4)$$

式中: x_{new} 表示新蜜獾的位置; $\beta \geq 1$ (一般默认取 6) 表示蜜獾获取食物的能力; r_3, r_4, r_5 表示 (0,1) 之间的 3 个随机数; F 表示改变搜索方向的标志,具体如式(5)所示。

$$F = \begin{cases} 1, r_6 \leq 0.5 \\ -1, \text{其他} \end{cases} \quad (5)$$

式中: r_6 表示 (0,1) 之间的随机数。

5) 采蜜阶段

在寻觅蜜源的环节,蜜獾依据导蜜鸟的指引抵达蜂巢位置的过程,能够用式(6)来表示。

$$x_{\text{new}} = x_{\text{prey}} + F \times r_7 \times \alpha \times d_i \quad (6)$$

式中: x_{new} 表示更新后蜜獾的全局最优位置; F, α 和 d_i 分别通过式(5)、式(3)和式(2)计算; r_7 表示 (0,1) 之间的随机数。

1.3 改进蜜獾优化算法

由于常规的蜜獾寻优方法收敛缓慢,容易陷入局部极值,因此,本文提出了一种改进方法,在其基础上进行如下改进:

1) Tent 混沌映射策略

初始解的随机获取在蜜獾优化算法中存在一定局限性,难以实现个体在解空间中的均匀分布。为此,混沌映射策略被引入优化算法领域,其旨在通过分散种群分布,有效降低个体聚集现象。混沌映射的主要两类是 Logistic 混沌映射和 Tent 混沌映射,后者可以创建了一个更均匀的混沌序列,具有更快的收敛速度,具体如式(7)所示。

$$x_{i+1} = \begin{cases} 2x_i + R \times \frac{1}{N_{\text{dim}}}, 0 \leq x \leq 0.5 \\ 2(1 - x_i) + R \times \frac{1}{N_{\text{dim}}}, 0.5 \leq x \leq 1 \end{cases} \quad (7)$$

式中: N_{dim} 表示蜜獾优化算法中个体的维度; R 表示 (0,1) 之间的随机数。

2) 改进控制参数

一旦种群达到局部最优状态,就很难保持种群的多样性,从而导致寻优效果变差。为使公式(3)中的变化更加平滑,采用具有随机扰动条件下的控制参数,具体如式(8)和(9)所示。

$$\alpha = A \times (2 \times R_2 - 1) \quad (8)$$

$$A = [\cos(2 \times R_1) + 1] \times \left(1 - \frac{t}{t_{\max}}\right) \quad (9)$$

式中: R_1, R_2 表示 (0,1) 之间的 2 个不同的随机数。

3) 小孔成像策略

由式(4)和(6)可知,蜜獾搜寻猎物的过程主要受当前最优个体引导。若当前的最优个体仅为局部最优,将可能导致算法过早收敛于局部最优解。为规避此问题,可通过实施小孔成像策略,逐维作用于当前最优个体。此策略不仅能有效减少个体维度间的相互干扰,还能显著提升当前最优个体的多样性。一维数学模型可用公式(10)表示。

$$\frac{(a+b)/2 - x}{x^* + (a+b)/2} = \frac{l}{l'} \quad (10)$$

式中: a 和 b 分别表示投影图像的水平上界和下界; x 和 x^* 分别表示镜面两侧在水平平面的投影点; l 和 l' 分别表示镜面两侧投影点到水平平面的距离。

将式(10)扩展至多维空间,并按维应用于当前最优个体,具体如式(11)所示。

$$\frac{(a_{\text{prey}}^j + b_{\text{prey}}^j)/2 - x_{\text{prey}}^j}{x_{\text{prey}}^{j*} + (a_{\text{prey}}^j + b_{\text{prey}}^j)/2} = \frac{l}{l'} \quad (11)$$

设置比例因子 $l/l' = r$, 则当前最优个体的小孔成像如式(12)所示。

$$x_{\text{prey}}^{j*} = \frac{(a_{\text{prey}}^j + b_{\text{prey}}^j)}{2} + \frac{(a_{\text{prey}}^j + b_{\text{prey}}^j) - x_{\text{prey}}^j}{2 \times r} \quad (12)$$

式中: a_{prey}^j 和 b_{prey}^j 分别表示当前最优个体在第 j 个维度的下边界和上边界; x_{prey}^j 表示当前最优个体在第 j 个维度的对应值; r 表示 (-3,3) 之间随机数。

1.4 牵引变压器故障诊断模型

为提高牵引变压器故障诊断的准确率,本文提出了一种基于 RF 特征优选和改进蜜獾优化算法的牵引变压器故障诊断方法,如图 1 所示,整体流程的步骤具体如下:

步骤 1) 收集牵引变压器故障数据集,并进行缺失值处理和异常值处理,确定数据的完整性;

步骤 2) 对故障样本数据集中的不平衡故障类型,应用 SMOTE 算法进行样本扩充处理,并利用无编码比值法对故障特征进行进一步拓展。

步骤 3) 采用随机森林算法对初步故障特征向量集合中相关故障特征进行特征重要度计算,并进行排序;

步骤 4) 将训练集导入相关模型算法,初始化 IHBA 算法、ELM、SVM 和 LSTM 的相关参数;

步骤 5) 通过 Tent 混沌策略初始化种群,通过上述基础预测模型训练,计算相关适应度值,并确定食物源;

步骤 6) 通过改进控制参数优化搜索因子,通过小孔成像策略增强当前最优个体的多样性;

步骤 7) 评估是否达到迭代停止条件。若条件满足,输出当前最优个体,并确定最佳基础预测模型超参数,否则返回步骤 5);

步骤 8) 将测试集数据输入 IHBA 优化基础预测模型,设计相关对比实验,并对诊断结果进行分析。

2 数据扩充

牵引变压器属于典型不平衡数据,模型在进行分析判

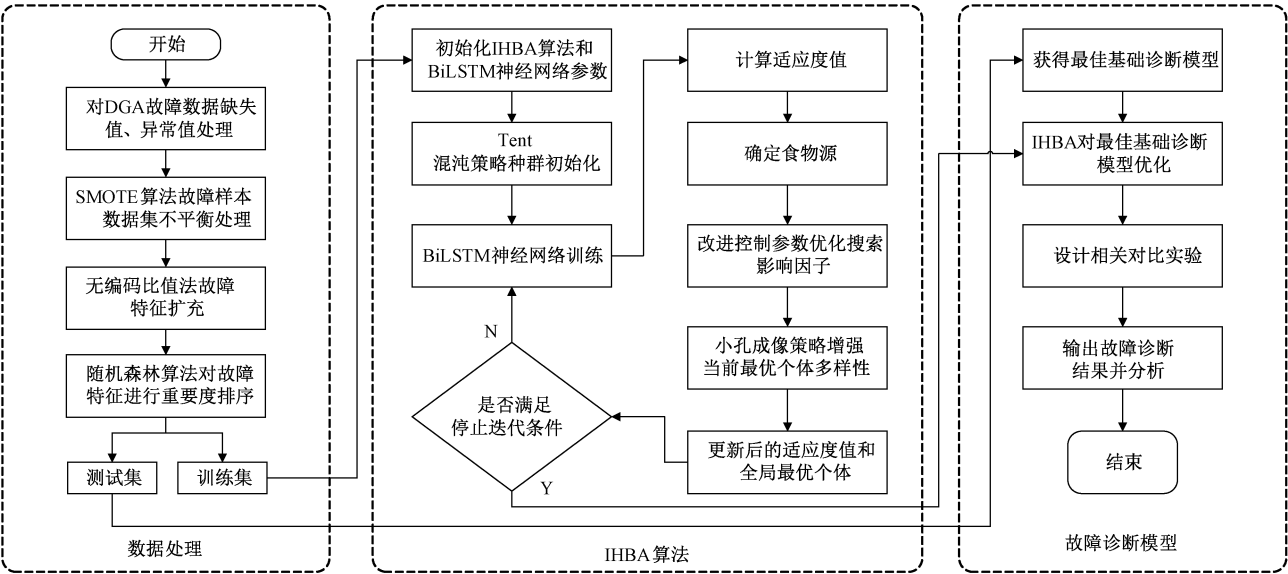


图 1 故障诊断流程图
Fig. 1 Fault diagnosis flowchart

断时极易偏向于多数类样本,忽略少数类样本,因此,需进行 SMOTE 样本扩充,以平衡不同类别样本。同时,工程现场通常选取的 DGA 气体无法有效表征牵引变压器的故障特征,需进行故障特征扩充。

2.1 SMOTE 样本扩充

SMOTE 算法是对随机过采样技术的一种优化。该算法并非通过简单的样本复制来实现过采样,而是针对非均衡数据集中少数类样本,通过线性随机插值生成新的合成样本,实现原始数据集的均衡化处理,进一步提升深度神经网络的识别准确率和推广性能。利用 SMOTE 算法进行样本扩充的具体步骤如下:

- 1) 在一个不平衡数据集中,从少数类样本中选取一个样本 x_i , 该样本作为根样本进行新样本的扩充;
- 2) 根据式(13)计算每种少数类故障类型中选为根样本的数据到相应少数类样本集中所有样本的距离 d , 得到每个根样本 k 个邻近的同类别样本:

$$d = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \tag{13}$$

式中: x_i 表示每个少数类样本集中的根样本; y_i 表示每个少数类样本集中除根样本以外的样本; m 表示少数类样本集中除根样本以外的样本数;

- 3) 针对牵引变压器故障数据集的类别不平衡问题,计算其多数类与少数类之间的不平衡度,依据该不平衡度,按照式(14)确定相应的向上采样扩大倍数 n , 即从 x_i 的 k 个邻近样本中随机选取 n 个样本作为辅助样本记为 y_1, y_2, \dots, y_n 。

$$n = f(I_L) \tag{14}$$

式中: $f(\cdot)$ 表示四舍五入运算函数; I_L 表示不平衡程度;

- 4) 在牵引变压器少数类故障类型的根样本 x_i 和随机

选择的辅助样本 y_i 之间进行随机插值,具体如公式(15)所示,分别合成 n 个对应少数类故障样本 p_i , 为解决故障类别中少数样本的问题,本方法执行了样本的全方位扩展,具体而言,即对每一种罕见故障类别的五个维度的特征气体信息实施了扩展,确保了数据集的均衡性。

$$p_i = x_i + r_i(y_i - x_i) \tag{15}$$

式中: r_i 表示 $(0,1)$ 之间的随机数。

SMOTE 算法通过在少数类别样本与其最近邻样本之间的连线上执行随机插值来生成新样本,此过程本质上是线性插值的运用,遵循特定的数学逻辑以有计划地构建数据。这样的方法有效减少了盲目性和限制性,从而有助于缓解因随机过采样而引起的过拟合现象。

2.2 特征扩充

在工程实践中,常常把 H_2 、 CO 、 CO_2 、 CH_4 、 C_2H_4 、 C_2H_6 、 C_2H_2 这七类气体作为检测牵引变压器运行状况的关键指标。然而,未经筛选的特征参数使得 DGA 的数据分布广泛,且数据间的差异悬殊。所以,利用无编码比值法进行特征扩充。扩充后所有特征具体信息如表 1 所示。

在表 1 中, A 表示 CH_4 、 C_2H_4 、 C_2H_6 、 C_2H_2 四种气体之和; B 表示 CO 、 CO_2 两种气体之和。

3 算例分析

3.1 数据来源及预处理

本文所用数据源自某公司牵引变压器实测数据,共包含各类样本 708 个,涉及低温过热(低于 $150^\circ C$)、电弧放电、电弧放电兼过热、高温过热(高于 $700^\circ C$)、局部放电和变压器正常老化六种故障。牵引变压器样本数据分布如表 2 所示。

表 1 DGA 特征扩充

Table 1 DGA feature expansion

编号	特征信息	编号	特征信息
1	H ₂	10	C ₂ H ₂ /C ₂ H ₄
2	CO	11	C ₂ H ₄ /C ₂ H ₆
3	CO ₂	12	CH ₄ /A
4	CH ₄	13	C ₂ H ₄ /A
5	C ₂ H ₄	14	C ₂ H ₆ /A
6	C ₂ H ₆	15	C ₂ H ₂ /A
7	C ₂ H ₂	16	B
8	A	17	CO/B
9	CH ₄ /H ₂	18	CO ₂ /B

表 2 DGA 样本数据分布

Table 2 DGA sample data distribution

编号	运行状态	样本个数
1	低温过热(低于 150 ℃)	28
2	电弧放电	80
3	电弧放电兼过热	37
4	高温过热(高于 700 ℃)	52
5	局部放电	379
6	变压器正常老化	132

数据在传输过程中会出现通讯中断等问题,需对数据进行预处理,以避免缺失数据和异常数据对模型造成不必要的干扰。由于数据样本之间没有时间上逻辑关系,无法采用传统均值填充等方法对缺失数据和异常数据进行替换,因此,本文选择直接将缺失数据和异常数据进行剔除。

3.2 评价指标

本文选取常用的准确率作为评价指标,以便全面而深入地评估分类预测模型的性能,计算如式(16)所示。

$$L_{accuracy} = \frac{TP + TN}{TP + FN + FP + TN}$$

(16)

式中: $L_{accuracy}$ 表示分类预测模型的准确率; TP 表示将正样本预测为正样本的个数; TN 表示将正样本预测为负样本的个数; FN 表示将负样本预测为正样本的个数; FP 表示将负样本预测为负样本的个数。

3.3 样本平衡及特征扩充

由表 2 可知,不同故障类别所含数据样本存在较大差异,采用基于人工智能算法的分类模型对正负样本失衡的数据集进行分析时,故障诊断结果偏向于多数类,导致少数类样本的为解决牵引变压器故障数据集中样本分布不均所导致的故障诊断准确率偏低的问题,本研究运用了 SMOTE 算法对数据进行增强处理。执行该算法前后,对牵引变压器故障数据集的样本量及分布情况进行了对比分析。分布情况如图 2 所示,具体数据大小及占比如表 3 所示。

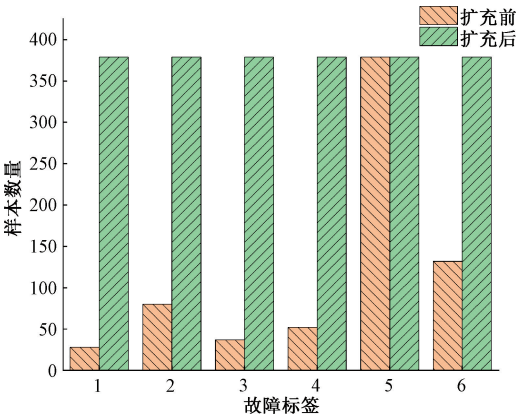


图 2 SMOTE 扩充前后数据对比

Fig. 2 Comparison of data before and after SMOTE expansion

表 3 SMOTE 扩充前后数据大小及占比

Table 3 Data size and proportion before and after SMOTE expansion

编号	SMOTE 扩充前		SMOTE 扩充后	
	数量	占比/%	数量	占比/%
1	28	3.95	379	16.67
2	80	11.30	379	16.67
3	37	5.23	379	16.67
4	52	7.34	379	16.67
5	379	53.53	379	16.67
6	132	18.64	379	16.67

由图 2 和表 3 可知,经 SMOTE 样本扩充后数据集中各类故障样本数据占比相同,有效处理故障样本数据不平衡问题。同时,采用无编码比值法对每个故障样本所含特征进行扩充,具体特征及编号如表 1 所示。

3.4 IHBA 有效性验证

为检验本文提出的 IHBA 优化算法的实际效能,特别挑选了 4 个具有代表性的基准测试函数进行综合评估。为确保算法的可靠性,其中 $F_1(x)$ 和 $F_2(x)$ 为单峰测试函数, $F_3(x)$ 和 $F_4(x)$ 为多峰测试函数。4 种基准函数的具体信息如表 4 所示。

将 IHBA 优化算法、蜣螂优化算法(dung beetle optimization, DBO)、HBA 和鲸鱼优化算法(whale optimization algorithm, WOA)针对四个标准评估函数实施了优化性能测试。这四个函数的相关参数,包括其维度、取值范围以及理想最优解,详见表 4 所列。所有参与比较的算法均设定为 500 次迭代。为了减少实验结果的随机性,每种算法针对每个标准函数独立重复实验 30 次,记录下各自的最优解、平均解以及解的标准偏差,并据此绘制出收敛走势图。4 个函数的收敛走势如图 3 所示,测试函数结果如表 5 所示。

由图 3 可知,本文提出的 IHBA 改进策略在加速收敛

表 4 测试函数
Table 4 Tests function

类型	函数名称	测试函数	维数 D	搜索范围	最优值
单峰测试函数	Schwefel 2.22	$F_1(x) = \sum_{i=1}^n x_i + \prod_{i=1}^n x_i $	30	$[-10, 10]$	0
	Quartic	$F_2(x) = \sum_{i=1}^n ix_i^4 + \text{random}[0,1)$	30	$[-1.28, 1.28]$	0
多峰测试函数	Rastrigin	$F_3(x) = \sum_{i=1}^n [x_i^2 - 10\cos(2\pi x_i)10]$	30	$[-5.12, 5.12]$	0
	Griewank	$F_4(x) = \frac{1}{4\,000} \sum_{i=1}^n x_i^2 - \prod_{i=1}^n \cos(\frac{x_i}{\sqrt{i}}) + 1$	30	$[-600, 600]$	0

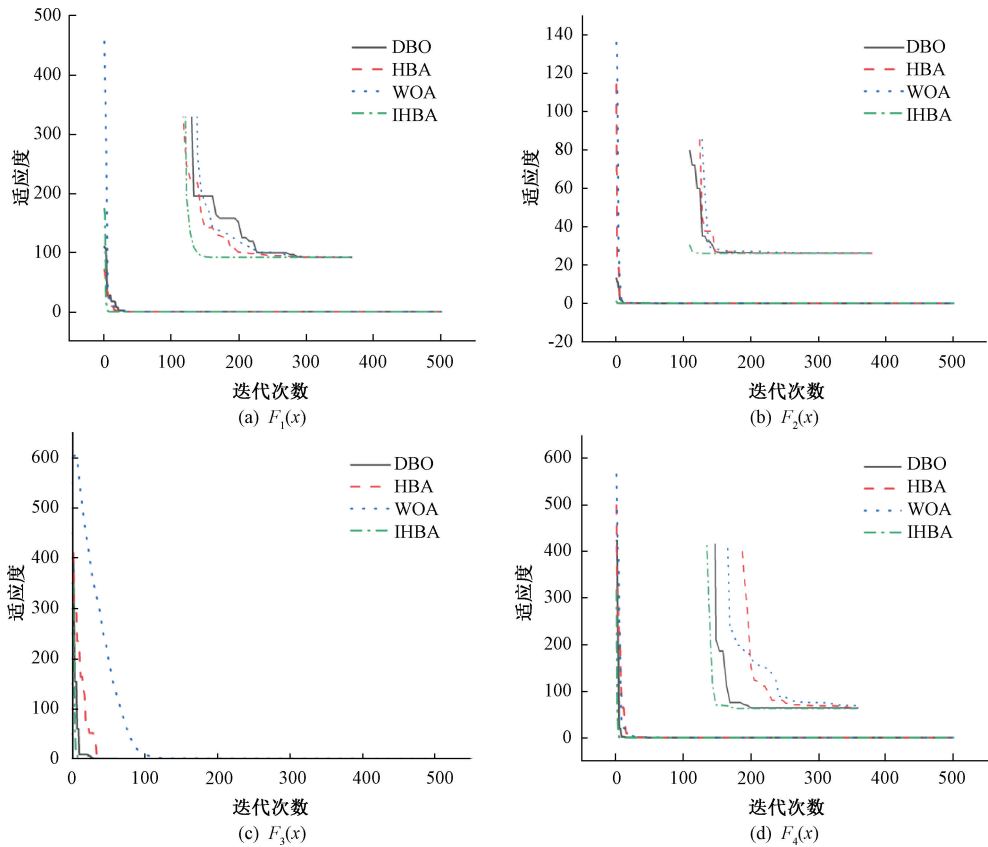


图 3 测试函数迭代曲线
Fig. 3 Test function iteration curve

进程及提升搜索精度方面,明显超越了现有的多种优化方法,表明 IHBA 算法所采用的多种改进策略不仅可以加快算法收敛速度,也可以提高寻优效率。同时,由表 5 可知, IHBA 在 4 个基准测试函数上均具有最好和最稳定的效果,表明 IHBA 具有显著的寻优性能和稳定性。

3.5 RF 特征优选及最佳基础模型确定

为剔除冗余特征对整体故障诊断模型的干扰,对牵引变压器故障样本数据所含特征进行 RF 特征优选,为避免偶然性的影响,采用对运行结果取平均值方法确定最终结果,具体结果如图 4 所示,排序结果如表 6 所示。

将上述故障特征按特征平均重要度排序依次输入 SVM、ELM 和 LSTM 中,为消除基础预测模型的波动,每个特征分别运行 10 次,以准确率为评价指标,得到平均准确率和最佳准确率,3 个基础模型的准确率结果如图 5 所示。

由图 5 可知,基础预测模型 ELM 的最佳特征个数均为 7、基础预测模型 SVM 的最佳特征个数均为 10、基础预测模型 LSTM 的最佳特征个数均为 10。在对比的模型中,ELM 模型展现了 71.90% 的最高平均准确度,并在最优情况下达到了 77.05% 的准确度;而 SVM 模型在平均准

表 5 测试函数结果对比
Table 5 Comparison of test function results

函数	指标	IHBA	DBO	HBA	WOA
$F_1(x)$	最优值	$3.330\ 6\times10^{-226}$	$5.238\ 31\times10^{-87}$	$5.361\ 28\times10^{-76}$	$1.677\ 17\times10^{-59}$
	平均值	$8.278\ 2\times10^{-213}$	$1.451\ 5\times10^{-56}$	$4.639\ 94\times10^{-71}$	$6.088\ 45\times10^{-51}$
	标准差	0	$4.112\ 17\times10^{-56}$	$4.579\ 53\times10^{-71}$	$9.706\ 7\times10^{-51}$
$F_2(x)$	最优值	$6.120\ 9\times10^{-6}$	$7.504\ 83\times10^{-5}$	$2.615\ 85\times10^{-5}$	$2.559\ 49\times10^{-5}$
	平均值	$1.397\ 75\times10^{-4}$	$7.060\ 47\times10^{-4}$	$1.634\ 58\times10^{-4}$	$2.750\ 971\times10^{-3}$
	标准差	$1.711\ 22\times10^{-4}$	$9.538\ 07\times10^{-4}$	$2.351\ 05\times10^{-4}$	$3.722\ 485\times10^{-3}$
$F_3(x)$	最优值	0	0	0	0
	平均值	0	0.319 942 746	0	0
	标准差	0	0.907 223 27	0	0
$F_4(x)$	最优值	0	0	0	0
	平均值	0	0.006 401 682	0	0
	标准差	0	0.016 996 89	0	0

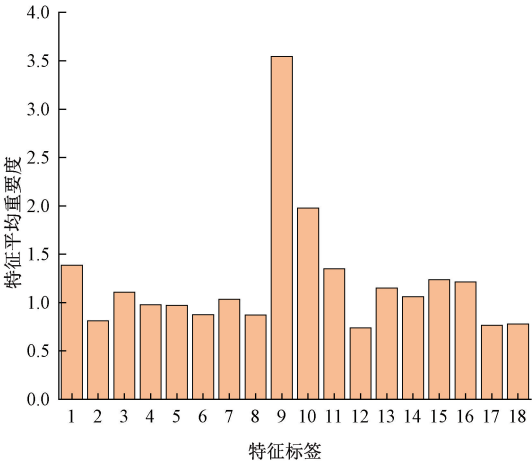


图 4 特征重要性程度

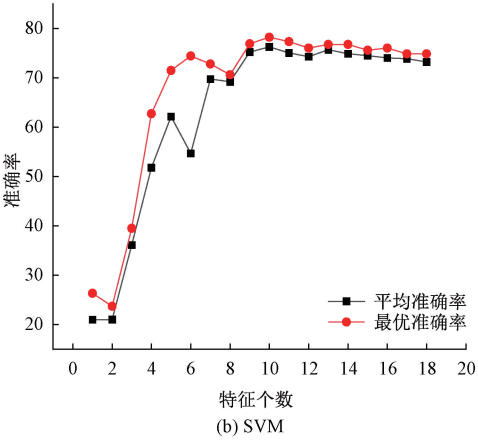
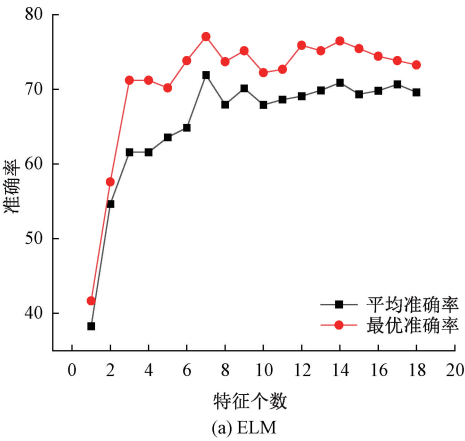
Fig. 4 Importance level of features

表 6 特征重要性程度及排序

Table 6 Importance and ranking of features

特征编号	平均重要度	排序	特征编号	平均重要度	排序
1	1.388 487 23	3	10	1.976 419 29	2
2	0.811 285 854	15	11	1.347 955 662	4
3	1.107 337 065	8	12	0.740 180 246	18
4	0.979 990 884	11	13	1.150 101 568	7
5	0.973 092 337	12	14	1.061 538 681	9
6	0.874 807 156	13	15	1.236 941 96	5
7	1.033 106 827	10	16	1.215 431 433	6
8	0.873 511 904	14	17	0.764 712 894	17
9	3.544 704 504	1	18	0.779 193 383	16

确度上以 76.27%领先,其最优准确度则为 78.22%;至于 LSTM 模型,其平均准确度达到 81.15%,在最佳表现下准确度高达 85.09%。因此,最佳基础预测模型为 LSTM,最佳特征个数为前十个,即特征编号为 9、10、1、11、15、16、13、3、14 和 7 的故障特征。



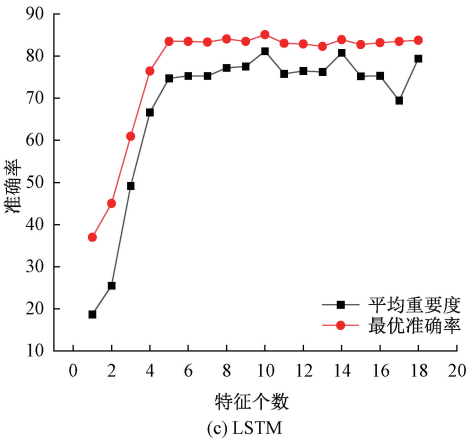


图 5 基础模型的准确率曲线

Fig. 5 Accuracy curve of the basic model

3.6 故障诊断结果分析

为验证 IHBA 在优化 LSTM 神经网络上的有效性,分别将其与 DBO、HBA 和 WOA 进行对比,迭代曲线如图 6 所示,参数优化结果如表 7 所示。

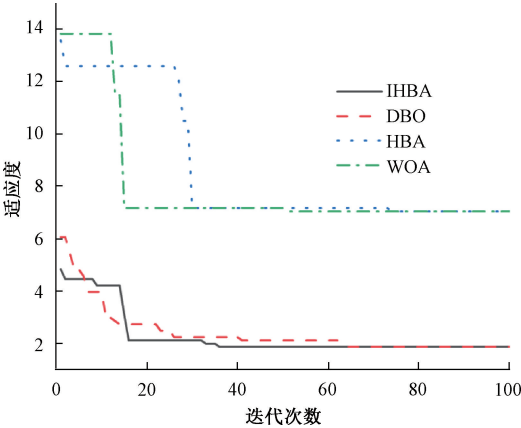
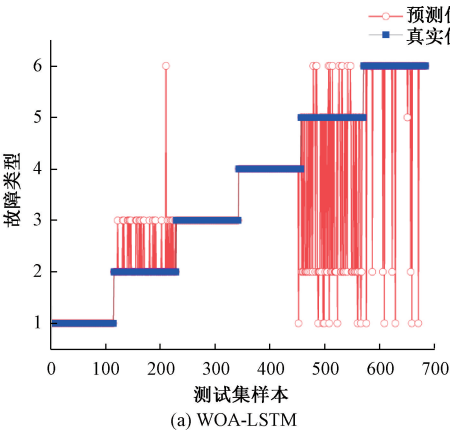
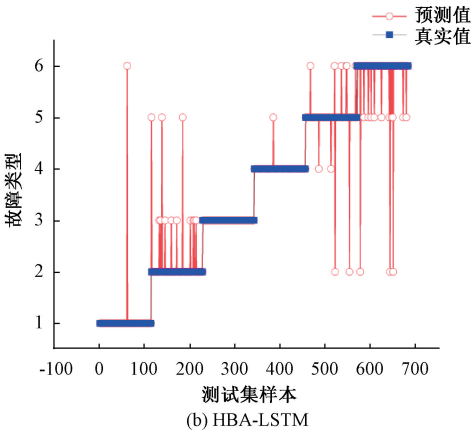


图 6 优化 LSTM 迭代曲线

Fig. 6 Optimize LSTM iteration curve



(a) WOA-LSTM



(b) HBA-LSTM

表 7 LSTM 超参数优化结果

Table 7 LSTM hyperparameter optimization results

优化算法	初始学习率	隐藏层 神经元个数	L2 正则化 系数
IHBA	0.066 6	445	1.06×10^{-6}
DBO	0.096 1	1 000	1.83×10^{-6}
HBA	0.057 9	614	2.04×10^{-6}
WOA	0.099 1	938	1.01×10^{-6}

由图 6 可知,在优化 LSTM 神经网络上,相比于 HBA 等优化算法,IHBA 可以更快且更准确地找到最小值,有效跳出来局部最优解,验证了所用改进策略的有效性。

为确认提出故障检测模型的准确度,本研究将训练数据集逐一导入 WOA-LSTM、HBA-LSTM、DBO-LSTM 以及 IHBA-LSTM 中进行对照研究,诊断成效如图 7 所示,详细数据如表 8 所示。

表 8 准确率结果

Table 8 Accuracy results

故障诊断模型	准确率/%
WOA-LSTM	85.23
HBA-LSTM	94.44
DBO-LSTM	93.86
IHBA-LSTM	96.05

由图 7 和表 8 可知,相比于所选组合模型 WOA-LSTM 和 DBO-LSTM,IHBA-LSTM 准确率分别提高了 11.26%和 2.33%,表明选择恰当的模型参数有助于进一步提高模型性能;相比于 HBA-LSTM,效率提升了 2.44%,证明了实施优化手段的成效,显著增强了故障诊断系统的准确度。此外,在众多故障诊断算法中,运用 IHBA-LSTM 算法的故障诊断系统,在辨别编号为 2、5、6 的故障类别方面展现出更为出色的识别能力。

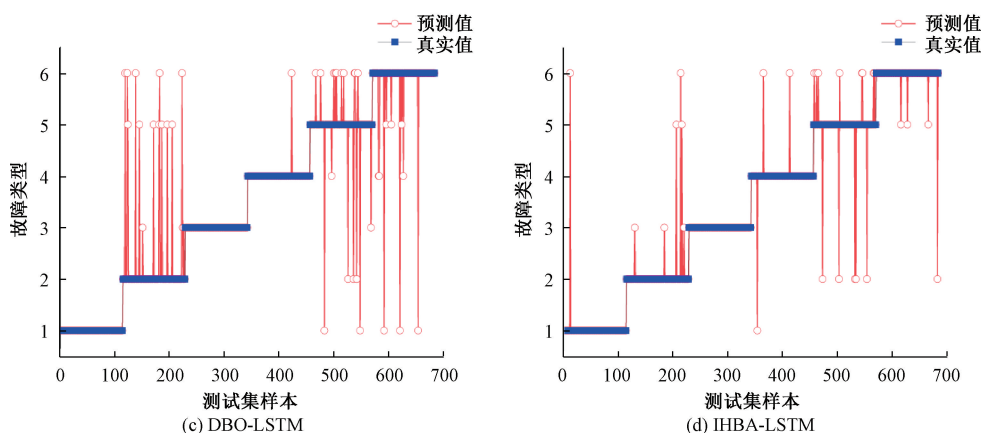


图 7 诊断结果

Fig. 7 Diagnostic results

4 结 论

本文提出了一种基于 RF 特征优选和 IHBA 的牵引变压器故障诊断,有效区分了牵引变压器的不同故障状态,得到如下结果:通过 SMOTE 算法进行故障数据样本扩充和无编码比值法进行故障数据特征扩充,可有效解决所用故障数据样本不均衡和 DGA 数据过于分散且数据量差异较大的问题,为后续模型分析提供了数据基础;本文所探讨的 IHBA 算法,在经过 4 个标准测试函数的对比分析后,展现出其与传统的 DBO、HBA 及 WOA 算法的不同性能特点,表明 IHBA 算法具有更好地寻优性能、稳定性和更快地收敛速度,验证了所提改进策略的有效性;通过 RF 进行特征优选和最佳基础预测模型判断,避免了冗余特征对预测模型的干扰,降低了模型复杂度,得到最佳基础预测模型为 LSTM。相比于输入全部特征,显著增强了故障检测的准确度;构筑以 IHBA 为核心的故障检测框架,并将其与 WOA-LSTM、HBA-LSTM 及 DBO-LSTM 算法进行比较分析。结果显示,该框架在辨别故障类别上展现出显著优势,实现了 96.05% 的诊断准确率,这一指标优于对比的各组模型。之后工作将研究区分实际情况下不同程度故障的诊断方法,设计出更为精准的诊断模型,以便处理更为复杂的故障情况。

参考文献

- [1] BASHIR T, WANG H, TAHIR M, et al. Wind and solar power forecasting based on hybrid CNN-ABiLSTM, CNN-transformer-MLP models [J]. Renewable Energy, 2025, 239:122055-122055.
- [2] 赵剑, 张哲, 李召端, 等. 三端口 CLLC 固态变压器的设计与优化[J]. 电工技术学报, 2024, 39(23): 7542-7553.
ZHAO J, ZHANG ZH, LI ZH D, et al. Design and optimization of three-port CLLC solid-state transformer[J]. Transactions of China Electrotechnical Society, 2024, 39(23): 7542-7553.
- [3] ZHANG X, FENG Z, MU W. Reformer: Reparameterized kernel lightweight transformer for grape disease segmentation [J]. Expert Systems with Applications, 2025, 265, 125757.
- [4] 周亚中, 何怡刚, 邢致恺, 等. 基于 IDBO-ARIMA 的电力变压器振动信号预测[J]. 电子测量与仪器学报, 2023, 37(8):11-20.
ZHOU Y ZH, HE Y G, XING ZH K, et al. Power transformer vibration signal prediction based on IDBO-ARIMA[J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(8):11-20.
- [5] 汪繁荣, 李州. 基于 EBWO-SVM 的变压器故障诊断研究[J]. 电子测量技术, 2024, 47(10):101-107.
WANG F R, LI ZH. Research on transformer fault diagnosis based on EBWO-SVM [J]. Electronic Measurement Technology, 2024, 47(10):101-107.
- [6] 韩文文, 谭政, 张昆, 等. 基于多传感器的干式变压器故障诊断系统设计[J]. 国外电子测量技术, 2024, 43(9):173-181.
HAN W W, TAN ZH, ZHANG K, et al. Design of fault diagnosis system of dry type transformer based on multi sensor[J]. Foreign Electronic Measurement Technology, 2024, 43(9):173-181.
- [7] 侯赛, 成润坤, 刘达. 基于二次采样和集成学习方法的变压器故障预测[J]. 智慧电力, 2024, 52(7):40-47.
HOU S, CHENG R K, LIU D. Transformer fault forecast based on re-sampling and integrated learning approach[J]. Smart Power, 2024, 52(7):40-47.
- [8] 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.
- [9] AMIN Z, KASSYMZHOMART K, AKHAVAN S B, et al. Transformer fault prognosis using deep recurrent neural network over vibration signals[J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70.
- [10] 咸日常, 范慧芳, 李飞, 等. 基于改进 GSA-SVM 模型的电力变压器故障诊断[J]. 智慧电力, 2022, 50(6):50-56.
- XIAN R CH, FAN H F, LI F, et al. Power transformer fault diagnosis based on improved GSA-SVM model[J]. Smart Power, 2022, 50(6):50-56.
- [11] 全兆景, 兰孟月, 荆利菲. 改进贝叶斯网络在变压器故障诊断中的应用[J]. 电子科技, 2024, 37(5):47-53,70.
- TONG ZH J, LAN M Y, JING L F. Research on transformer fault diagnosis based on improved bayesian network [J]. Electronic Science and Technology, 2024, 37(5):47-53,70.
- [12] 李云谔, 咸日常, 张海强, 等. 基于改进灰狼算法与最小二乘支持向量机耦合的电力变压器故障诊断方法[J]. 电网技术, 2023, 47(4):1470-1478.
- LI Y H, XIAN R CH, ZHANG H Q, et al. Fault diagnosis for power transformers based on improved gery wolf algorithm coupled with least squares support vector machine[J]. Power System Technology, 2023, 47(4):1470-1478.
- [13] 宋辉, 苑龙祥, 郭双权. 基于数据增强和特征注意力机制的灰狼优化算法-优化残差神经网络变压器故障诊断方法[J]. 现代电力, 2024, 41(2):392-400.
- SONG H, YUAN L X, GUO SH Q. GWO-ResNet power transformer fault diagnosis method based on data augmentation and feature attention mechanism [J]. Modern Electric Power, 2024, 41(2):392-400.
- [14] 周玉, 岳学震, 刘星, 等. 不平衡数据集的自然邻域超球面过采样方法[J]. 哈尔滨工业大学学报, 2024, 56(12):81-95.
- ZHOU Y, YUE X ZH, LIU X, et al. A natural neighborhood hypersphere oversampling method for imbalanced data sets[J]. Journal of Harbin Institute of Technology, 2024, 56(12):81-95.
- [15] DAS B, KRISHNAN N C, COOK D J. RACOG and WRACOG: Two probabilistic oversampling techniques[J]. IEEE Transactions on Knowledge and Data Engineering, 2015, 27(1):222-234.
- [16] 张化昭, 朱建武, 邱日强, 等. 基于 SMOTE-SSA-CNN 的煤矿用变压器 DGA 故障诊断方法[J]. 煤矿机械, 2024, 45(12):172-176.
- ZHANG H ZH, ZHU J W, QIU R Q, et al. DGA fault diagnosis method for coal mine transformer based on SMOTE-SSA-CNN [J]. Coal Mine Machinery, 2024, 45(12):172-176.
- [17] 相晨萌, 闫鹏, 赵海涛, 等. 基于无编码比值法的天然酯绝缘油变压器故障诊断方法研究[J]. 河北电力技术, 2022, 41(2):62-66.
- XIANG CH M, YAN P, ZHAO H T, et al. Research on diagnosis method if natural ester insulating oil transformer based on non-code ratio method [J]. Hebei Electric Power, 2022, 41(2):62-66.
- [18] XIE H, XIAO D, MAO Z. Exploration of grade distribution in iron mines based on rough set extreme learning machine and multispectral [J]. Expert Systems with Applications, 2025, 265,125938.
- [19] ZHAO Y, YANG Z, HOU Z, et al. Deep learning based on image analysis for refrigerant charging and leakage detection in building heat pump[J]. Energy & Buildings, 2025, 328, 115157.
- [20] 王妍, 王新发, 王延峰, 等. 基于改进天鹰优化算法优化 LSTM 的滚动轴承故障诊断方法[J]. 振动与冲击, 2024, 43(23):144-154.
- WANG Y, WANG X F, WANG Y F, et al. A rolling bearing fault diagnosis method based on improved aquila optimization algorithm to optimize LSTM[J]. Journal of Vibration and Shock, 2024, 43 (23): 144-154.

作者简介

迟青光, 博士, 副教授, 主要研究方向为牵引变压器、牵引电机优化设计及智能运维。

E-mail: 75826540@qq.com

蒋瑞培, 硕士研究生, 主要研究方向为故障诊断。

E-mail: 2077553114@qq.com

闫东旭(通信作者), 工程师, 主要研究方向为牵引变压器的故障检测与诊断。

E-mail: 013500016045@crrecg.cc