

DOI: 10.19783/j.cnki.pspc.231094

基于改进 SMOTE 不平衡样本处理和 IHPO-DBN 的 变压器故障诊断方法研究

周莹^{1,2}, 吴伟丽^{1,2}

(1. 西安科技大学电气与控制工程学院, 陕西 西安 710054; 2. 西安市电气设备状态监测与
供电安全重点实验室, 陕西 西安 710054)

摘要: 针对由于变压器故障样本不平衡和故障模型陷入局部最优而导致的分类准确率低的问题, 提出了基于改进的合成少数类过采样技术和优化深度置信网络(deep belief network, DBN)的变压器故障诊断方法。首先采用聚类融合的 K-means 算法, 通过分簇和匹配的方式筛选出不稳定的少数类样本用以改进中心点合成少数类过采样技术(center point synthetic minority oversampling technique, CP-SMOTE)算法, 并对少数类样本进行扩增, 解决了变压器故障数据分布不平衡的问题。其次, 通过加入随机逆向学习和自适应惯性权重技术对猎食者优化算法进行改进, 并用改进后的算法对 DBN 的内部参数进行优化调整, 提高了模型精度。最后, 将不同数据预处理情况下以及不同数据规模下的变压器故障模型进行仿真对比。结果表明, 经过数据预处理和模型优化后的变压器故障识别准确率能够提高到 98%, 有效地解决了故障数据不平衡导致的分类精度低的问题。

关键词: 变压器故障诊断; 不平衡样本; K-means 聚类; 改进合成少数过采样; 改进猎食者优化

Transformer fault diagnosis method based on improved SMOTE unbalanced sample processing and IHPO-DBN

ZHOU Xuan^{1,2}, WU Weili^{1,2}

(1. College of Electrical and Control Engineering, Xi'an University of Science and Technology, Xi'an 710054, China;
2. Xi'an Key Laboratory of Electrical Equipment Condition Monitoring and Power Supply Security, Xi'an 710054, China)

Abstract: To solve the problem of low classification accuracy due to unbalanced transformer fault samples and a local optimal fault model, a transformer fault diagnosis method based on improved synthetic minority oversampling technique and an optimized deep belief network (DBN) is proposed. First, the K-means algorithm of clustering fusion is used to select the unstable minority class samples by clustering and matching to improve the center point synthetic minority oversampling technique (CP-SMOTE) algorithm and expand the minority class samples to solve the problem of unbalanced distribution of transformer fault data. Secondly, the algorithm of predator optimization is improved by adding stochastic reverse learning and adaptive inertial weight technology, and the internal parameters of deep confidence network are optimized and adjusted by the improved algorithm, which improves the accuracy of the model. Finally, the transformer fault models in different data preprocessing conditions and data scales are simulated and compared. The results show that the transformer fault identification accuracy can be increased to 98% after data preprocessing and model optimization. This effectively solves the problem of low classification accuracy caused by fault data imbalance.

This work is supported by the Science and Technology Project of State Grid Corporation of China (No. SGXJJCJ00KJJS2100582).

Key words: transformer fault diagnosis; unbalanced samples; K-means clustering; improved synthetic minority oversampling; improved predator optimization

基金项目: 国家电网公司科技项目资助(SGXJJCJ00KJJS2100582); 陕西省教育厅自然科学专项资助(17JK0503); 合肥市关键共性技术研发项目资助(2021GJ039)

0 引言

变压器作为电力系统互联和功率交换的枢纽设备^[1], 掌握其运行状态对保障电网稳定、经济、优质运行均具有重要意义^[2]。变压器发生故障后危害巨大, 会影响人们的日常生活、造成巨额的经济损失^[3]。

变压器内部溶解气体的组成成分及含量与其故障类型密切相关, 油中溶解气体分析(dissolved gas-in-oil analysis, DGA)是通过检测油浸式变压器油中的溶解气体(H_2 、 CH_4 、 C_2H_6 、 C_2H_4 、 C_2H_2)含量来分析变压器运行状态^[4], 因此 DGA 成为广泛应用于电力行业的变压器故障诊断方法^[5]。基于 DGA 的三比值法^[6]、Roger 比值^[7]等相关分析法虽简单, 但编码缺失可能会导致故障无法判断。近年来, 随着人工智能的发展, 研究者们以变压器油中溶解气体含量作为输入特征来研究变压器故障诊断的智能方法有人工神经网络、支持向量机^[8]、极限学习机等。文献[9-11]分别介绍了支持向量机、模糊逻辑、神经网络这几种模型, 并使用优化算法对它们进行改进, 提高了故障识别准确率。

虽然上述文献中的诊断模型最终获得了较高的准确率, 但却忽略了在现实情况下变压器各种故障产生的概率可能不相同的问题, 这会导致收集的不同类别的故障样本数据之间有明显的差距。

利用不平衡数据对模型进行训练, 会使分类模型偏向于多数类别, 从而增加了少数类的漏判率和误判率^[12]。使用合成少数类过采样技术(synthetic minority oversampling technique, SMOTE)算法对少数类样本进行扩增, 可以缓解样本不平衡的问题, 但如果仅用 SMOTE 处理, 就会存在数据分布边缘化和类间边界模糊等劣势^[13]。文献[14]提出了中心点合成少数类过采样技术(center point synthetic minority oversampling technique, CP-SMOTE)算法, 该算法首先从样本数据中寻找出几个中心点, 再通过少数类样本与中心点之间的线性插值生成新的样本, 然而此方法可能会由于新生成的样本加重了密集区样本的密集度而使得稀疏区样本在分类器进行分类时被忽略, 从而影响样本的正确分类, 因此本文在此基础上提出使用聚类融合的 K-means 技术先对少数类故障数据进行筛选, 选出处于边界的不稳定样本, 再对不稳定少数类样本使用 CP-SMOTE 算法进行扩增, 能够有效解决故障样本分布不均衡问题。

深度置信网络(deep belief network, DBN)是一种非参数多层神经网络, 在逐层训练受限玻尔兹曼机(restricted Boltzmann machine, RBM)后又能反向微调参数。DBN 具有两个优点: 一是具有多层次结构, 能够逐层提取特征, 得到高维表示; 二是可以

通过无监督学习方法自动学习数据分布, 具有鲁棒性, 易于推广^[15-16]。

猎食者优化(hunter-prey optimization, HPO)算法模拟了自然界中猎食者和猎物的行为, 对参数进行寻优, 具有收敛速度快、寻优能力强的特点, 与传统的人工智能算法相比具有更好的诊断能力。但由于 HPO 算法的初始种群是随机生成的, 这会使模型存在陷入局部最优和迭代收敛慢的问题, 因此利用随机逆向学习技术^[17]和自适应惯性权重技术^[18]对算法进行改进, 实现全局寻优和收敛加速。

本文利用改进的 CP-SMOTE 算法处理数据分布不均衡问题, 再用处理后的样本训练 DBN 模型, 同时使用改进的 HPO(improved hunter-prey optimization, IHPO)算法对 DBN 参数进行寻优, 最终确定将 IHPO-DBN 用作变压器故障分类的诊断模型。

1 基于改进 CP-SMOTE 的样本数据处理

文献[19]提出的 SMOTE 算法, 其思路是通过寻找样本的近邻集, 在样本点与其近邻集随机选择的样本连线上合成新的样本点, 使样本比例趋于平衡^[20]。CP-SMOTE 算法则是对传统 SMOTE 算法进行了改进, 有效地处理了噪声样本和边界模糊问题。本文提出的改进 CP-SMOTE 算法, 通过将聚类融合的 K-means 和 CP-SMOTE 相结合的方式对少数类样本进行分簇扩增, 不仅能处理多数类和少数类之间的类间不平衡问题, 还能处理少数类类内分布不均匀的问题。

1.1 基于聚类融合 K-means 改进的 CP-SMOTE 算法

CP-SMOTE 数据过采样算法的步骤如下。

- 1) 使用 K-means 算法找到少数类样本的 n 个区域和 i 个中心点 $\{O_1, O_2, \dots, O_i\}$ 。
- 2) 计算每个区域的中心点 O_i 到多数类样本的距离 D_i 。
- 3) 计算每个区域的少数类样本和与其对应的中心点的距离 d_i 。
- 4) 比较 D_i 和 d_i , 如果 $d_i < D_i$, 则在对应的少数类样本 X 和中心点 O_i 之间, 利用式(1)随机生成一个新的点。

$$X_{\text{new}} = \eta X + (1 - \eta)O_i \quad (1)$$

式中: X_{new} 为扩增后新生成的点; $\eta \in (0, 1)$ 。

按照以上步骤虽然能解决少数类和多数类之间的不平衡和边界模糊的问题, 但却没有考虑到少数类类内分布不均匀的情况。为此, 采用聚类融合的 K-means 算法加以改进, 其思路是先用 K-means 算法对样本数据进行分簇, 再用簇标记匹配法^[21-22]和簇类一致性系数^[23]对聚类结果中的每个数据进行匹配

和筛选, 最终筛选出位于边界的不稳定少数类样本。

聚类融合的 K-means 算法对变压器故障状态数据处理的具体步骤如下。

1) 聚类: 使用 K 个不同参数的 K-means 算法对变压器故障状态数据中的少数类样本进行聚类, 最终得到 K 个分类结果: $\pi_1, \pi_2, \dots, \pi_K$, 且每个聚类结果 $\pi_i = \{C_1^i, C_2^i, C_3^i, \dots, C_n^i\}$ 。

2) 簇标记匹配: 将聚类结果中的每一簇 C_j^i 转化为由二进制表示的 X_j^i 。利用匹配系数统计在不同聚类成员中的两个簇所共同含有的数据对象的个数。对所包含的相同数据对象最多的两个簇进行相同簇标记匹配。

3) 筛选不稳定少数类: 对于重新标记过程的结果, 计算每个数据点的聚类一致性系数, 它可以衡量样本对象在多次聚类中属于相同类的次数与总的聚类次数的比值。计算结果分为两种:

(1) 聚类一致性系数较小。表明该样本对象在不同的聚类结果中都被分配到不同的簇, 即该样本对象一般是处在簇边界的不稳定样本。

(2) 聚类一致性系数较大。表明该样本对象在不同的聚类结果中都被分配到相同的簇, 即该样本对象一般是处在簇内的稳定样本。

将聚类融合的 K-means 算法和 CP-SMOTE 算法结合即得到一种改进的 CP-SMOTE 算法, 改进的 CP-SMOTE 算法流程具体如下。

输入: 训练集 T , 聚类次数 k , 采样倍率 N , 新生成少数类样本数据集 T_{new} 。

输出: 新数据集 T_{train} 。

Step1: 使用 K-means 聚类方法对训练集 T 进行聚类。

Step2: 重复 k 次 Step1, 共计算出 k 个聚类结果。

Step3: 用簇标记匹配法匹配不同的聚类结果, 并进行重新标记, 最终得到 k 个新的聚类结果。

Step4: 计算所有少数类数据样本的簇类一致性系数 δ_{ci} 。

Step5: 将 δ_{ci} 较小的少数类数据样本定为边缘样本。

Step6: 选择不稳定边界少数类和包含不稳定边界少数类样本簇的簇中心作为 CP-SMOTE 算法中的采样点和采样中心点, 代入式(1), 最终生成由一组合成样本 X_{new} 组成的新的少数类样本数据集 T_{new} 。

Step7: 将 T_{new} 加入到 T 中产生新的训练数据集 T_{train} 。

1.2 基于改进 CP-SMOTE 油中溶解气体样本数据预处理

从 IEC TC10 数据库和变压器状态监测案例中共收集 1030 组变压器状态监测数据, 提供了 H_2 、 CH_4 、 C_2H_6 、 C_2H_4 、 C_2H_2 的气体浓度以及对应的变压器状态类型, 状态类型有 7 种, 分别是正常(F)、低能放电(D_1)、高能放电(D_2)、低温过热(T_1)、中温过热(T_2)、高温过热(T_3)和局部放电(PD)。具体数据分布见表 1, 其中 S^2 为每类少数类故障数据的方差。

表 1 变压器监测样本类型原始数据描述

Table 1 Raw data description of transformer monitoring

sample fault types				
故障类型	训练集个数	测试集个数	数据集总数	S^2
PD	78	34	112	2.99×10^{-4}
D_1	127	54	181	—
D_2	135	58	193	—
T_1	93	40	133	2.49×10^{-4}
T_2	109	47	156	1.35×10^{-4}
T_3	94	40	134	2.08×10^{-4}
F	84	37	121	1.89×10^{-4}

由表 1 可以看出, 方差的值越大, 说明数据类内分布越均匀。由表 1 可见, 在所获得的变压器监测数据中, 低能放电和高能放电这两种类型样本数据偏多。其他 5 种类型样本数据偏少, 将聚类次数 k 设置为 6, 聚类簇数 m 分别设置为 3、4、 \dots 、8, 在采样倍率 $N = 135$ 的情况下, 利用改进 CP-SMOTE 算法筛选出不稳定少数类样本, 处理后样本分布结果见表 2 所示, 其中 S_1^2 、 S_2^2 、 S_3^2 分别表示 3 种算法

表 2 数据处理后变压器故障类型数据集的描述

Table 2 Description of transformer fault type data sets after data processing

故障类型	SMOTE 处理后	CP-SMOTE 处理后	改进 CP-SMOTE 处理后	S_1^2	S_2^2	S_3^2
PD	146	128	125	2.98×10^{-4}	2.78×10^{-4}	3.34×10^{-4}
D_1	127	127	127	—	—	—
D_2	135	135	135	—	—	—
T_1	140	139	131	2.44×10^{-4}	2.21×10^{-4}	3.24×10^{-4}
T_2	139	141	140	1.36×10^{-4}	1.17×10^{-4}	1.52×10^{-4}
T_3	141	137	132	2.07×10^{-4}	2.05×10^{-4}	2.22×10^{-4}
F	143	130	127	1.88×10^{-4}	1.69×10^{-4}	1.95×10^{-4}

处理后少数类故障数据的方差。为便于进行对比，将利用 SMOTE 算法、CP-SMOTE 算法处理后的结果也一并列出。

表 2 中，不平衡数据在经过算法处理后，每类训练数据之间达到平衡，且经过改进 CP-SMOTE 算法处理后的少数类的方差值最大，说明此时少数类类内分布最均匀。

为了更直观地展示样本均衡处理结果，以变压器状态监测数据训练集中的高温过热(少数类)和高能放电(多数类)为例，绘制出改进 CP-SMOTE 处理后的样本分布图，为便于进行对比，将原始数据分布、利用 SMOTE 过采样和 CP-SMOTE 过采样处理后的样本分布也一并给出，如图 1 所示。

将图 1 中红色圆框所圈部分进行局部放大，可以看出在扩增少数类样本数据时，SMOTE 算法明

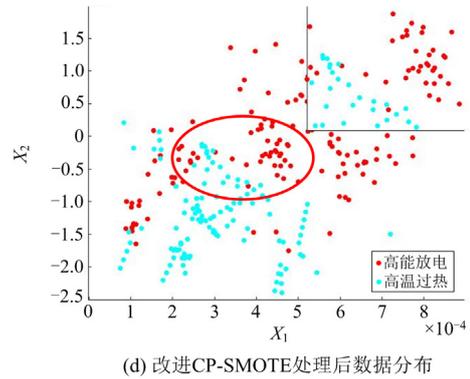
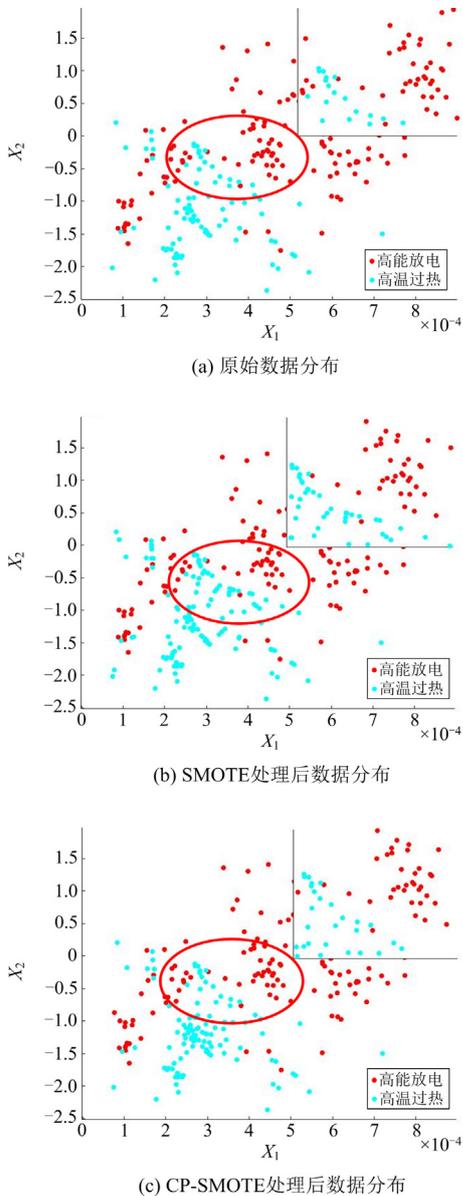


图 1 4 种情况下高能过热与高能放电数据分布图
Fig. 1 Partial discharge and high energy discharge data distribution in four cases

显会更多地对处于多数类和少数类边界之间的样本进行插值，这会使样本模糊，诊断模型不能够正确地辨别边界样本的类型，最终影响故障类型识别的准确率。而 CP-SMOTE 算法虽然避免了上述问题，但其采用的仅对少数类类内的某一个中心点与该中心点周围样本进行插值的方法，忽略了样本分布不均匀的情况，导致处于稀疏区域的少数类样本被漏判，同样会影响故障识别准确率。

本文先使用聚类融合的 K-means 算法对少数类样本进行筛选，选出 4 簇拥有不稳定样本的区域，以每一簇的聚类中心作为采样中心，每一簇对应的不稳定样本作为采样点进行簇内的 CP-SMOTE 处理，最终得到如图 1(d)所示的样本分布图，可以看出与图 1(c)相比，图 1(d)中少数类类内分布更均匀，说明本文提出的改进 CP-SMOTE 算法既能避免类间数据不平衡、边界模糊状况的出现，又能处理少数类类内分布不均匀的问题，更有利于数据样本作为后续故障诊断模型的输入特征向量进行故障诊断。

2 基于 IHPO-DBN 的变压器故障诊断模型

2.1 DBN 原理

文献[24]提出的深度置信网络是一种深度学习模型。该模型由多个 RBM 堆叠而成，并对每层 RBM 进行无监督贪婪训练，优化网络的连接权重和偏置，从而有效提取数据的特征。

RBM 的隐含层 h 和可视层 v 两层之间的神经元相连接，但同一层的神经元不连接。两者的状态向量可表示为 $\mathbf{h}=[h_1, h_2, \dots, h_{n_h}]$ 和 $\mathbf{v}=[v_1, v_2, \dots, v_{n_v}]$ ， W_{ij} 是第 i 个可视层神经元之间的连接权值。 $\mathbf{a}=[a_1, a_2, \dots, a_{n_v}]$ 和 $\mathbf{b}=[b_1, b_2, \dots, b_{n_h}]$ 分别表示隐含层和可视层的偏置向量。

当堆叠的所有 RBM 训练完毕后,使用 BP 神经网络进行反向微调,从而达到全局最优。利用 DBN 提取变压器油中溶解气体数据特征的训练过程就是求解模型参数 θ 的过程。

DBN 分类模型用在非线性数据中,随机初始化可视层神经元和隐含层神经元之间的连接权重 W_{ij} 与神经元偏置,虽然通过迭代次数的增加连接权重值会接近最优解,但也有可能会在迭代更新过程中使 DBN 参数陷入局部最优状态^[25],因此在 DBN 模型训练之前采用改进猎食者算法对其结构参数和输出参数进行寻优,可以有效地提高参数寻优过程中的抗振荡能力和特征提取能力。

2.2 IHPO 优化算法优化 DBN

文献[26]提出的猎食者优化算法是一种新型智能优化算法,它通过模拟动物猎食的过程对模型进行寻优。该算法用于搜索猎食者机制的数学模型由方程(2)给出,猎物位置更新的数学模型由方程(3)中给出。

$$x_{i,j}(t+1) = x_{i,j}(t) + 0.5[(2CZP_{\text{pos}(j)} - x_{x,j}(t)) + (2(1-C)Z\mu(j) - x_{i,j}(t))] \quad (2)$$

$$x_{i,j}(t+1) = T_{\text{pos}(j)} + CZ\cos(2\pi R_4) \times [T_{\text{pos}(j)} - x_{i,j}(t)] \quad (3)$$

式中: $x_{i,j}(t)$ 是当前猎食者的位置; $x_{i,j}(t+1)$ 是猎食者在后续迭代中的位置; C 是探索和开发之间的平衡参数,其值在迭代过程中逐渐降低; Z 是可调变量; $P_{\text{pos}(j)}$ 是猎物的位置; μ 是所有位置的平均值; $T_{\text{pos}(j)}$ 是全局最优位置; R_4 是随机值, $R_4 \in [-1,1]$ 。

2.2.1 随机逆向学习

由于 HPO 算法的初始种群是随机生成的,这会导致种群陷入局部最优的境况。因此,本文将随机逆向学习技术应用于初始种群,增加目标区域种群的多样性,提高算法在全局寻优中的性能。

文献[27]提出的随机逆向学习是一种逆向学习的改进方法,随机逆向学习的思路是在群体寻优过程中,根据当前解产生一个逆向解,并比较两个解之间的目标函数值,然后选择最优解以便进行下一次迭代,其计算公式为

$$x_{\text{rand}} = b_{\text{max}} + b_{\text{min}} - rx_{\text{best}} \quad (4)$$

式中: x_{rand} 是随机逆向解; r 是[0,1]之间的任意数; x_{best} 是当前最优解; b_{max} 和 b_{min} 分别为解的上界和下界,并且 $x_{\text{rand}} \in [b_{\text{min}}, b_{\text{max}}]$ 。

2.2.2 自适应惯性权重

惯性权重是提高算法收敛性的一类重要的控制参数。对算法进行计算时,在迭代前期,希望收敛因子受个体极值影响大,使种群快速找到相对较好

的位置;在迭代中期,一部分种群向全局最优值靠近,一部分种群继续搜寻更好的位置;在迭代后期,所有种群向全局最优值靠近,使得所有种群得到更好的收敛^[28]。然而标准 HPO 中的线性收敛因子没有办法满足上述要求,因此使用非线性的自适应惯性权重,这类方法使得算法在迭代阶段既能使全局平衡,又能提高最优解的收敛速度。

非线性自适应惯性权重动态变化曲线如图 2 所示。

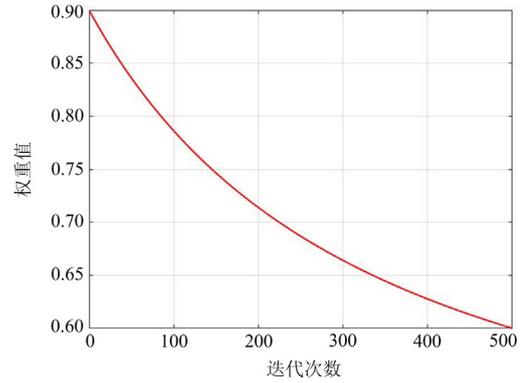


图 2 非线性自适应惯性权重动态曲线

Fig. 2 Nonlinear adaptive dynamic curve of inertia weight

将凹函数的非线性递减惯性权重应用于 HPO 算法,其公式为

$$\omega = \omega_{\min} \left(\frac{\omega_{\max}}{\omega_{\min}} \right)^{\frac{1}{1 + \frac{ct}{t_{\max}}}} \quad (5)$$

式中: ω 是权重因子; ω_{\min} 和 ω_{\max} 分别是权重调整参数的最小值和最大值,一般分别取 0.4 和 0.9; c 是调整参数; t 是当前迭代次数; t_{\max} 是整个算法的最大迭代次数。将式(3)代入式(5)中,最终得到式(6),它能改进猎物和迭代过程的拟合,并且加速算法向最优解收敛。

$$x_{i,j}(t+1) = \omega T_{\text{pos}(j)} + CZ\cos(2\pi R_4) \times [T_{\text{pos}(j)} - x_{i,j}(t)] \quad (6)$$

经过上述两种技术改进后,对提出的 IHPO 算法进行算法性能测试,详见附录 A。最终结果表明,同 PSO 算法、SSA 算法和 HPO 算法相比,改进之后的 IHPO 算法的全局搜索能力得到提升,速度和精度都要高于其他算法。

2.2.3 IHPO 优化 DBN 流程

将随机逆向学习技术和自适应惯性权重技术引入到 HPO 算法中,得到 IHPO 算法,并用该方法对 DBN 内部参数进行寻优。

首先采用 IHPO 算法对 DBN 模型中神经元之间的连接权重和偏置值进行全局寻优,使得 DBN 的

内部参数接近最优解，将最优解作为 DBN 内部参数的初始值进行迭代训练^[29]。

用 DBN 输出的均方差作为 IHPO 的适应度函数，如式(7)所示。

$$F = \frac{1}{M} \sum_i^M (Y_i - y_i)^2 \quad (7)$$

式中： F 为 DBN 网络输出的均方差值； M 为训练样本数； Y_i 为训练样本的期望输出值； y_i 为训练样本的实际输出值。

利用 IHPO 对 DBN 结构中的 RBM 逐层进行训练，其具体步骤如下。

Step1: 初始化种群。随机生成搜索空间中初始种群的每一个成员。

Step2: 参数初始化。初始化 HPO 算法的参数，设置种群大小 p ，最大迭代次数 t_{max} ，维度 D 。

Step3: 计算适应度的值。计算个体与种群的初始适应度值，计算出个体最优位置和全局最优位置。

Step4: 猎食者搜索。利用式(2)和式(3)分别对猎食者和猎物的位置进行更新，并用式(5)替换惯性权重公式。

Step5: 种群选择。将迭代后猎食者的位置或猎物的位置与最优个体位置作为下一代的适应度值，并与由式(4)计算出的当前的逆向解进行对比，选择最优解用于下次迭代。

Step6: 判断终止条件。判断算法的当前迭代次数 t 是否达到最大迭代次数 t_{max} ，如果是，则输出期望误差最小的个体参数，否则返回 Step4。

Step7: 将 IHPO 搜索到的最优参数作为 RBM 的网络参数，逐层对 DBN 参数进行优化，确定 DBN 的网络结构参数。

图 3 为 IHPO 优化 DBN 的流程图。

3 基于数据处理的 IHPO-DBN 变压器故障诊断模型

采用 IHPO-DBN 模型进行不平衡样本下的变压器故障诊断。具体过程如下。

1) 预处理及划分数据集。对变压器 DGA 故障数据集进行归一化处理，并将故障样本按照 7:3 的比例划分为训练集和测试集。

2) 样本平衡化。对于训练集中每个少数类中的数据样本，通过聚类融合的 K-means 对数据样本进行聚类及区分，再通过 CP-SMOTE 对少数类样本进行合成，生成平衡化的变压器故障样本数据集。

3) 变压器故障诊断。将经过平衡化处理后的变压器故障样本数据集作为故障诊断模型的输入，通

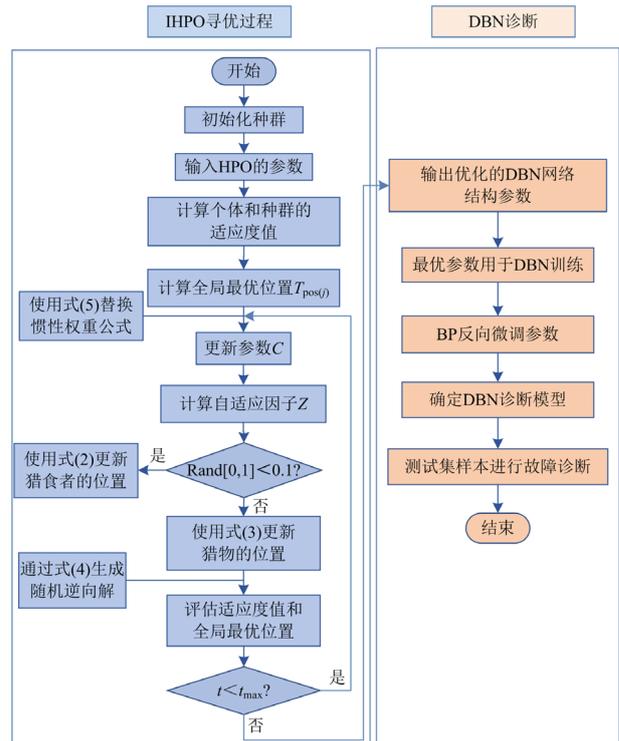


图 3 IHPO 优化 DBN 流程图

Fig. 3 Flow chart of IHPO optimizing DBN

过 IHPO-DBN 诊断模型进行故障诊断，输出故障诊断结果。

图 4 为基于 IHPO-DBN 的变压器故障诊断流程图，其中 C_1 和 C_2 分别表示一价与二价碳氢化合物的气体含量之和。从图 4 中可以看出，本文提出的故障诊断模型包含变压器状态监测数据预处理、DBN 特征提取、HPO 优化改进、IHPO 参数优化和 DBN 故障诊断这几个阶段。

4 算例分析

为了验证所提出的改进 CP-SMOTE 算法对数据的预处理有利于变压器监测数据的分类，本文利用 Matlab R2021a 进行仿真测试，并且以 IHPO-DBN 作为测试的诊断模型，网络模型的初始参数设置如表 3 所示。

在原始的不平衡数据下，利用 IHPO-DBN 模型进行故障诊断，结果如图 5 所示。由图 5 可见，最终测试结果的准确率为 91.6129%，其中，局部放电、低能放电、高能放电、低温过热、中温过热的准确率分别为 85.3%、96.4%、96.6%、92.5%、95.7%，高温过热的准确率为 82.1%，正常的准确率为 86.5%。造成局部放电、低温过热、中温过热、高温过热、正常这 5 类监测数据的测试准确率较低的

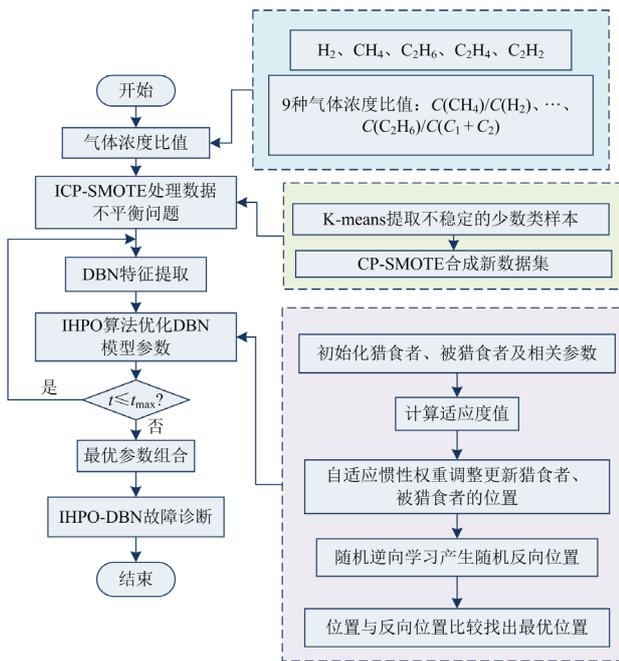


图 4 基于 IHPO-DBN 的变压器故障诊断流程图

Fig. 4 Transformer fault diagnosis flow chart based on IHPO-DBN

表 3 模型参数设置值

Table 3 Model parameter setting values

参数	数值	参数	数值
网络层数	3	Batch-size	30
RBM 个数	3	DBN 动量	0.3
DBN 学习率	0.001	—	—

原因是样本数较少导致在故障模型诊断时被漏判、误判。

采用 SMOTE 算法处理不平衡数据后, 利用 IHPO-DBN 模型进行故障诊断, 结果如图 6 所示。由图 6 可见, 该测试结果的准确率为 92.5806%, 其中局部放电、低能放电、高能放电、低温过热、中温过热、高温过热、正常的准确率分别为 85.3%、96.4%、96.6%、85%、95.7%、92.3%、91.9%。虽然每类少数类的数据量都被扩增到与多数类的的数据量相近, 但是少数类局部放电、低温过热、中温过热、高温过热、正常的准确率仍然不高, 甚至低温过热的准确率与原始数据情况相比降低了 7.5%, 这是由于多数类和少数类的类间边界模糊导致的。

采用 CP-SMOTE 算法处理不平衡数据后, 利用 IHPO-DBN 模型进行故障诊断, 结果如图 7 所示。由图 7 可见, 该测试结果的准确率为 96.129%, 其中局部放电、低能放电、高能放电、低温过热、中温过热、高温过热、正常的准确率分别为 91.2%、

98.2%、98.3%、92.5%、97.9%、94.9%、97.3%。与 SMOTE 处理后相比, 虽然每一类监测数据的准确率都有所提高, 但是局部放电、低温过热、高温过热这 3 类与其他几类相比其准确率仍然比较低, 这是由于其类内分布不均匀导致的。

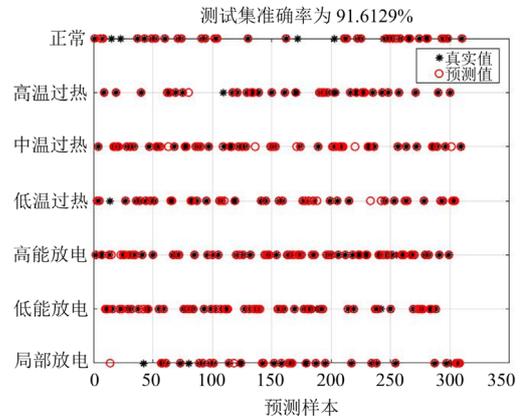


图 5 原始数据集下的模型诊断结果

Fig. 5 Model diagnosis results under the raw data set

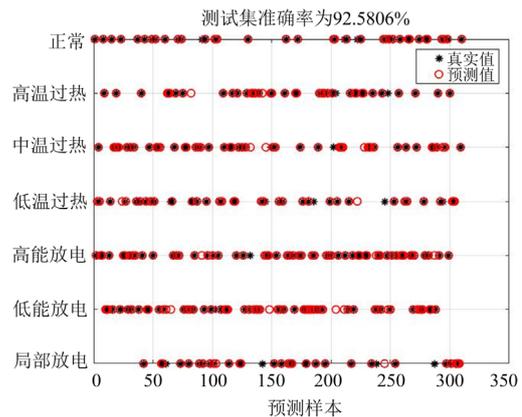


图 6 SMOTE 处理下的模型诊断结果

Fig. 6 Model diagnosis results under SMOTE

采用改进 CP-SMOTE 算法处理不平衡数据后, 利用 IHPO-DBN 模型进行故障诊断, 结果如图 8 所示。由图 8 可见, 该测试结果的准确率为 98.0645%, 其中局部放电、低能放电、高能放电、低温过热、中温过热、高温过热、正常的准确率分别为 97.1%、100%、98.3%、97.5%、97.9%、97.4%、97.3%。发现变压器监测样本的每一类诊断准确率都维持在 98% 上下, 说明本文提出的改进 CP-SMOTE 算法与 SMOTE 算法、CP-SMOTE 算法相比, 能同时解决数据样本不平衡、样本类间模糊和类内分布不均匀的问题。

通过仿真得到不同平衡化方法下的不同故障诊断模型的准确率, 如表 4 所示。

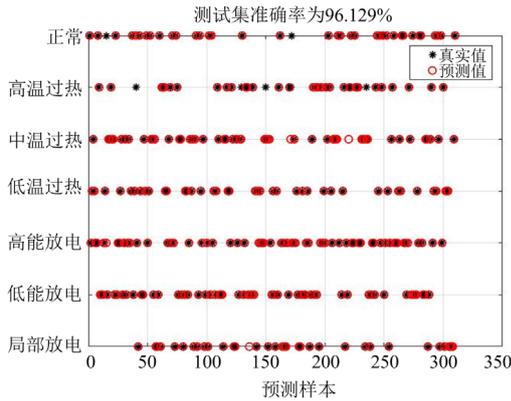


图 7 CP-SMOTE 处理下的模型诊断结果

Fig. 7 Model diagnosis results under CP-SMOTE

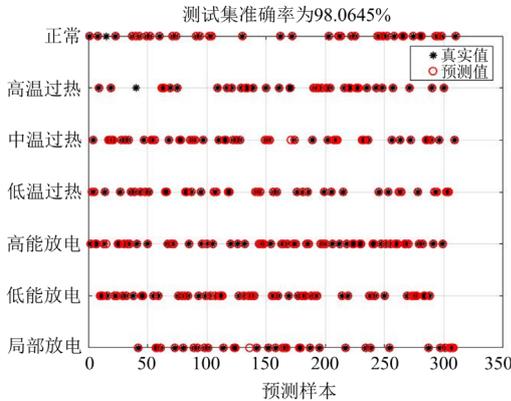


图 8 改进 CP-SMOTE 处理下的模型诊断结果

Fig. 8 Model diagnosis results under improved CP-SMOTE

表 4 不同平衡化方法下的 3 种模型准确率对比

Table 4 Comparison of accuracy of three models with different balancing methods

平衡化方法	准确率/%		
	PSO-DBN	SSA-DBN	IHPO-DBN
原始数据集	81.5287	87.8981	91.6129
SMOTE	84.0764	88.8535	92.5806
CP-SMOTE	89.0323	93.2258	96.129
改进 CP-SMOTE	90.6452	95.1613	98.0645

由表 4 可见，在不平衡样本预处理改进方面，本文提出的改进 CP-SMOTE 算法对于 PSO-DBN、SSA-DBN 以及 IHPO-DBN 模型的故障辨识准确率的提高都有明显的效果，在故障模型的优化方面，IHPO-DBN 模型在任何数据处理下其故障诊断结果都能够比 PSO-DBN 模型和 SSA-DBN 模型好，说明本文所提方法在诊断数据的处理和诊断模型的改进方面都存在较大的优势。

为了验证所提出模型的适应性，选取另一个样本量为 610 组的变压器监测数据进行仿真对比，判

断不同训练集情况下的模型对变压器故障诊断结果的影响，仿真结果如表 5 所示。

表 5 不同训练集下的变压器故障诊断结果

Table 5 Transformer diagnosis results under different training sets

方法	训练集	测试集	错分数	准确率/%
未处理 IHPO-DBN	300	310	35	88.7821
SMOTE IHPO-DBN	420	310	32	90.0621
CP-SMOTE IHPO-DBN	420	310	14	95.4839
改进 CP-SMOTE IHPO-DBN	400	310	11	96.4516

由表 5 可以看出，当分别选取不同规模的数据训练时，经改进 CP-SMOTE 数据处理后，IHPO-DBN 方法对故障诊断的准确率为 96.4516%，仍保持相对较高的准确率，表明所提方法对不同规模的样本均具有较好的辨识性。

5 结论

针对变压器故障状态样本多数类和少数类分布不平衡和诊断精度较低的问题，在利用改进 CP-SMOTE 算法对数据进行预处理后，提出了一种基于 IHPO-DBN 的变压器故障诊断模型，并通过与 SMOTE、CP-SMOTE 两种数据预处理算法以及和 PSO-DBN、SSA-DBN 模型进行对比，得到以下结论。

1) 与原始数据集的每类训练集样本数量相比，改进 CP-SMOTE 算法使少数类样本分别扩增至 125、131、140、132、127，实现了与多数类的类间样本平衡。再通过方差 S^2 与 SMOTE 算法、CP-SMOTE 算法进行对比，发现改进的 CP-SMOTE 算法方差值总是大于其他两种算法的方差值，说明本文提出的算法还能够解决少数类类内样本分布不均匀的问题。

2) 对各种变压器监测数据下的故障模型进行诊断，发现相比于 PSO-DBN、SSA-DBN 模型，在不同的平衡化方法下，所提出的 IHPO-DBN 模型的准确率分别提升了 10.0842%、3.7148%、8.5042%、3.7271%、7.0967%、2.9032%、7.4103%、2.9032%。验证了 IHPO 对 DBN 变压器故障诊断模型优化的优越性。

3) 610 组样本在经过各种数据处理后的变压器故障诊断训练测试的准确率与 1030 组数据下的相比仅分别降低了 2.8308%、2.5185%、0.6451%、1.6129%，且仍然是 IHPO-DBN 模型训练下准确率最高，说明无论在何种数据规模下，改进的 CP-SMOTE 算法与基于 IHPO-DBN 的变压器故障诊断模型都适用。

附录 A

为了测试 IHPO 算法的寻优能力，本文选取单

峰 Sphere 函数和多峰 Rastrigin 函数分别对粒子群算法 PSO、麻雀优化算法 SSA、猎食者优化算法 HPO 及改进猎食者优化 IHPO 算法进行对比分析。

单峰函数 Sphere 测试函数公式为

$$f_1(x) = \sum_{i=1}^{30} x_i^2 \quad (A1)$$

多峰函数 Rastrigin 测试函数公式为

$$f_2(x) = \sum_{i=1}^{30} [x_i^2 - 10 \cos(2\pi x_i) + 10] \quad (A2)$$

式(A1)中, 自变量 x_i 的取值范围为 $[-100, 100]$, 最优解为 0; 式(A2)中, 自变量 x_i 的取值范围为 $[-5.12, 5.12]$, 最优解为 0。两种测试函数中每一个算法的迭代次数皆为 100, 其仿真结果图如图 A1 所示。

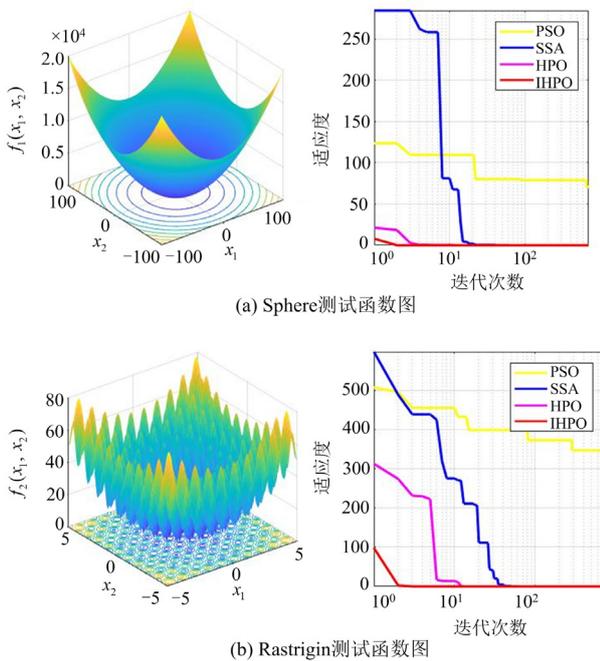


图 A1 测试函数图

Fig. A1 Test function diagram

由图 A1 可以看出, 在 Sphere 函数和 Rastrigin 函数测试的过程中, PSO 算法的收敛速度最慢且不能够收敛到最优适应度值 0; HPO 分别在第 6 次和第 13 次迭代收敛至 0, 且与 PSO 和 SSA 相比收敛速度明显快于两者; 而 IHPO 的收敛速度皆在第 2 次迭代时就找到了全局最优解。

参考文献

- [1] 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].
- [2] 李平, 胡根铭. 基于数据增强型一维改进卷积神经网络的变压器故障诊断方法[J]. 电网技术, 2023, 47(7): 2957-2967.
LI Ping, HU Genming. Transformer fault diagnosis based on data enhanced one-dimensional improved convolutional neural network[J]. Power System Technology, 2023, 47(7): 2957-2967.
- [3] 张国治, 陈康, 方荣行, 等. 基于 DGA 与鲸鱼算法优化 LogitBoost-决策树的变压器故障诊断方法[J]. 电力系统保护与控制, 2023, 51(7): 63-72.
ZHANG Guozhi, CHEN Kang, FANG Rongxing, et al. Transformer fault diagnosis based on DGA and a whale algorithm optimizing a LogitBoost-decision tree[J]. Power System Protection and Control, 2023, 51(7): 63-72.
- [4] XIANG C M, HUANG Z Y, LI J, et al. Graphic approaches for faults diagnosis for Camellia insulating liquid filled transformers based on dissolved gas analysis[J]. Dielectrics and Electrical Insulation, 2018, 25(5): 1897-1903.
- [5] LI J, ZHANG Q G, WANG K, et al. Optimal dissolved gas ratios selected by genetic algorithm for power transformer fault diagnosis based on support vector machine[J]. IEEE Transactions on Dielectrics and Electrical Insulation, 2016, 23(2): 1198-1206.
- [6] QI B, WANG Y, ZHANG P, et al. A novel self-decision fault diagnosis model based on state-oriented correction for power transformer[J]. Dielectrics and Electrical Insulation, 2020, 27(6): 1778-1786.
- [7] JONGVILAIKASEM K, PATTANADECH N. The comparison of DGA interpretation techniques application for actual failure transformer inspections including experience from power plants in Thailand[J]. International Journal on Electrical Engineering and Informatics, 2022, 14(1): 224-233.
- [8] 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 the improved bald eagle search algorithm[J]. Power System Protection and Control, 2023, 51(8): 118-126.
- [9] 方涛, 钱晔, 郭灿杰, 等. 基于天牛须搜索优化支持向量机的变压器故障诊断研究[J]. 电力系统保护与控制, 2020, 48(20): 90-96.
FANG Tao, QIAN Ye, GUO Canjie, et al. Research on transformer fault diagnosis based on a beetle antennae search optimized support vector machine[J]. Power System Protection and Control, 2020, 48(20): 90-96.
- [10] SU Q, MI C, LAI L L, et al. A fuzzy dissolved gas analysis method for the diagnosis of multiple incipient faults in a transformer[J]. IEEE Transactions on Power Systems, 2000, 15(2): 593-598.

- [11] 赵文清, 严海, 周震东, 等. 基于残差 BP 神经网络的变压器故障诊断[J]. 电力自动化设备, 2020, 40(2): 143-148.
ZHAO Wenqing, YAN Hai, ZHOU Zhendong, et al. Fault diagnosis of transformer based on residual BP neural network[J]. Electric Power Automation Equipment, 2020, 40(2): 143-148.
- [12] 刘云鹏, 许自强, 和家慧, 等. 基于条件式 Wasserstein 生成对抗网络的电力变压器故障样本增强技术[J]. 电网技术, 2020, 44(4): 1505-1513.
LIU Yunpeng, XU Ziqiang, HE Jiahui, et al. Data augmentation method for power transformer fault diagnosis based on conditional Wasserstein generative adversarial network[J]. Power System Technology, 2020, 44(4): 1505-1513.
- [13] 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 Yunpeng, HE Jiahui, XU Ziqiang, et al. Equalization method of power transformer fault sample based on SVM SMOTE[J]. High Voltage Engineering, 2020, 46(7): 2522-2529.
- [14] BAO Y, SIBO Y. Two novel SMOTE methods for solving imbalanced classification problems[J]. IEEE Access, 2023, 11: 5816-5823.
- [15] WU S, ZHENG L, HU W, et al. Improved deep belief network and model interpretation method for power system transient stability assessment[J]. Journal of Modern Power Systems and Clean Energy, 2020, 8(1): 27-37.
- [16] XUE L, NINGYU L, CHUANG C, et al. Attention mechanism based multi-scale feature extraction of bearing fault diagnosis[J]. Journal of Systems Engineering and Electronics, 2023, 34(5): 1359-1367.
- [17] 冯增喜, 何鑫, 崔巍, 等. 混合随机反向学习和高斯变异的混沌松鼠搜索算法[J]. 计算机集成制造系统, 2023, 29(2): 604-615.
FENG Zengxi, HE Xin, CUI Wei, et al. Hybrid random opposition-based learning and Gaussian mutation of chaotic squirrel search algorithm[J]. Computer Integrated Manufacturing Systems, 2023, 29(2): 604-615.
- [18] PRAKASH M, CHINNAMUTHU P. Techno-economic modeling of integrated renewable energy system using adaptive inertia weight based PSO[J]. Journal of Scientific & Industrial Research, 2020, (7): 647-652.
- [19] CHAWLA N V, BOWYER K W, HALL L O, et al. Smote: synthetic minority over-sampling technique[J]. Artificial Intelligence and Research, 2002, 16(1): 321-357.
- [20] ZHOU Z, TANG W. Clusterer ensemble[J]. Knowledge Based Systems, 2006, 19(1): 77-83.
- [21] 陈俊丰, 郑中团. WKMeans 与 SMOTE 结合的不平衡数据过采样方法[J]. 计算机工程与应用, 2021, 57(23): 106-112.
CHEN Junfeng, ZHENG Zhongtuan. Over-sampling method on imbalanced data based on WKMeans and SMOTE[J]. Computer Engineering and Applications, 2021, 57(23): 106-112.
- [22] 李海英, 沈益涛, 罗雨航. 基于改进深度置信网络的智能电网暂态安全状态感知[J]. 电力系统保护与控制, 2022, 50(5): 171-177.
LI Haiying, SHEN Yitao, LUO Yuhang. Transient security situational awareness of smart grids based on an improved deep belief network[J]. Power System Protection and Control, 2022, 50(5): 171-177.
- [23] TOPCH, ALEXANDER I P. Adaptive clustering ensembles[C]// Proceedings of the 17th International Conference on Pattern Recognition, August 26, 2004, Cambridge, UK: 272-275.
- [24] HINTON E G. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313: 504-507.
- [25] 贾鹤鸣, 刘宇翔, 刘庆鑫, 等. 融合随机反向学习的黏菌与算术混合优化算法[J]. 计算机科学与探索, 2022, 16(5): 1182-1192.
JIA Heming, LIU Yuxiang, LIU Qingxin, et al. Hybrid algorithm of slime mould algorithm and arithmetic optimization algorithm based on random opposition-based learning[J]. Journal of Frontiers of Computer Science & Technology, 2022, 16(5): 1182-1192.
- [26] NARUE I, KEYNIA F, MOLAHOSSEINI A S. Hunterprey optimization: algorithm and applications[J]. Soft Computing, 2022, 26(3): 1279-1314.
- [27] LONG W, JIAO J J, LIANG X M, et al. A random opposition-based learning grey wolf optimizer[J]. IEEE Access, 2019, 7: 113810-113825.
- [28] 张渊博, 邹德旋, 张春韵, 等. 自适应惯性权重的粒子群优化算法[J]. 计算机仿真, 2023, 40(4): 350-357.
ZHANG Yuanbo, ZOU Dexuan, ZHANG Chunyun, et al. Particle swarm optimization algorithm with adaptive inertial weight[J]. Computer Simulation, 2023, 40(4): 350-357.
- [29] 李黄曼. 基于深度学习的电力变压器故障诊断研究[D]. 西安: 陕西科技大学, 2021.
LI Huangman. Research on fault diagnosis of power transformer based on deep learning[D]. Xi'an: Shaanxi University of Science and Technology, 2021.

收稿日期: 2023-08-24; 修回日期: 2023-12-11

作者简介:

周莹(1999—), 女, 通信作者, 硕士研究生, 研究方向为电力系统设备故障诊断; E-mail: 1036507637@qq.com

吴伟丽(1978—), 女, 博士研究生, 副教授, 研究方向为电力系统运行与控制、大电网完全控制与灾变防治等。E-mail: wwlxm@163.com

(编辑 周金梅)