

基于 PCA 特征优选和 AdaBoost 集成学习的齿轮箱油品状态识别方法^{*}

陈晓犇¹ 黄采伦¹ 赵延明^{1,2} 李智靖¹ 南茂元¹ 田勇军³

(1. 湖南科技大学信息与电气工程学院 湘潭 411100; 2. 机械设备健康维护湖南省重点实验室 湘潭 411100;
3. 湖南科技大学机电工程学院 湘潭 411100)

摘要:针对传统齿轮箱油液分析方法存在的精度低和泛化能力有限的问题,提出一种基于 PCA 特征优选和 AdaBoost 集成学习的齿轮箱油品状态识别方法。首先,通过箱型图和 smote 插值对油液多参量数据进行清洗以提高油液数据的质量;其次,采用 PCA 进行油品特征优选,获取有助于识别的油品特征优选子集,在有效融合油液多参量信息的同时,可显著降低模型运行的时间复杂度;然后,利用 BP 神经网络建立油品状态识别基本模型,引入 GWO 灰狼优化算法对模型进行优化,构建具有最优初始权值与阈值的弱分类器 GWO-BP,同时采取自适应提升 AdaBoost 算法组合多个弱分类器 GWO-BP,集成为较强鲁棒性的强分类器。最后利用实验进行验证和分析,实验结果表明,所提方法效果最优,平均识别率 $99.30 \pm 0.16\%$,平均用时 32.77 ± 1.27 s,能够快速高效、准确识别出齿轮箱润滑油油品状态,为实现在线齿轮箱的油品状态识别奠定了良好基础,具有重要的工程应用价值。

关键词:油液分析;数据清洗;特征优选;集成学习;油品状态识别

中图分类号: TP274.2; TN06 文献标识码: A 国家标准学科分类代码: 510.40

Gearbox oil status recognition method based on PCA feature optimization and adaboost ensemble learning

Chen Xiaoben¹ Huang Cailun¹ Zhao Yanming² Li Zhijing¹ Nan Maoyuan¹ Tian Yongjun³

(1. School of Information and Electrical Engineering, Hunan University of Science and Technology, Xiangtan 411100, China;
2. Hunan Provincial Key Laboratory of Mechanical Equipment Health Maintenance, Xiangtan 411100, China;
3. School of Mechanical and Electrical Engineering, Hunan University of Science and Technology, Xiangtan 411100, China)

Abstract: A gearbox oil state recognition method based on PCA feature optimization and AdaBoost ensemble learning is proposed to address the problems of low accuracy and limited generalization ability in traditional gearbox oil analysis methods. Firstly, the multi parameter oil data is cleaned using box plots and SMOTE interpolation to improve the quality of the oil data; Secondly, PCA is used for oil product feature optimization to obtain a subset of oil product feature optimization that is helpful for identification. While effectively integrating multi parameter information of oil, it can significantly reduce the time complexity of model operation; Then, a basic model for oil state recognition is established using BP neural network, and the GWO wolf pack optimization algorithm is introduced to optimize the model. A weak classifier GWO-BP with optimal initial weights and thresholds is constructed, and an adaptive boosting AdaBoost algorithm is adopted to combine multiple weak classifiers GWO-BP, integrating them into a strong classifier with strong robustness. Finally, the experimental data was applied for verification and analysis. The experimental results showed that the proposed method had the best performance, with an average recognition rate of $99.30 \pm 0.16\%$ and an average time of 32.77 ± 1.27 seconds. It could quickly, efficiently, and accurately identify the oil state of the gearbox lubricating oil, laying a good foundation for realizing online oil state recognition of gearboxes and having important engineering application value.

Keywords: oil analysis; data cleansing; feature optimization; ensemble learning; oil status identification

0 引言

齿轮箱作为传动系统关键部件,广泛用于航空、风电、汽车、农业等领域^[1-2]。齿轮箱因使用环境复杂极易出现故障,严重影响设备的安全性和可靠性,必须采取有效的运维措施。定期更换润滑油是一种有效的齿轮箱运维手段,但其效果依赖于维修人员的技术与经验、受限于对润滑油状态的实时监测和精准判断^[3]。因此,在齿轮箱的检修维护中,润滑系统状态监测方法的优化显得尤为重要。

智能化油液分析技术为齿轮箱的检修维护提供了新方向。通过对油液性能参数的精确分析,能够准确地判断齿轮箱的工作状态,预测潜在故障,为维修决策提供有力支持。在磨粒信息研究方面,文献[4]应用主成分分析法(principal component analysis, PCA)和支持向量机(support vector machines, SVM)实现了滤波后图像的特征层融合,提高了磨粒信息的综合利用率;文献[5]基于多光照图像采集与三维曲面重建等光度立体识别方法,实现了三维图像磨粒特征的识别;文献[6]应用均值偏移算法建立统计聚类模型,实现了在线监测磨粒的分类识别。在维修决策层面,文献[7]基于模糊逻辑的专家系统,开发了润滑油状态的智能决策方法;文献[8]基于改进卷积神经网络提取油液数据特征,能准确识别 TBM 变速箱的磨损状态;文献[9]通过 SVM 获取指标的显著性系数计算参数权重,简化了磨损度的指标,提高了风电齿轮箱的故障诊断效率。以上利用数据驱动的方式剖析齿轮箱油液状态,已取得了一定成效,但大多局限于单一油品指标监测或采用单一学习算法对润滑油液品质进行分析,其容易受学习的样本分布限制,导致模型泛化性能不佳,可靠性和精度存在不足。

集成学习能够有效地结合各个算法的互补信息,提高集成模型的鲁棒性。目前使用的集成分类方法主要包括 bagging^[10-11]、自适应提升(adaptive boosting, AdaBoost)^[12-13]、随机森林^[14-15]、随机子空间^[16]、梯度增强^[17]等。其中,AdaBoost 具有良好的理论基础,在实践中也具备较好的可扩展性。比如,文献[18]应用 AdaBoost 集成多个长短期记忆神经网络(long short-term memory network, LSTM),构建了时序预测集成优化模型,实践证明效果良好;文献[19]提出基于改进条件生成对抗网络的不平衡数据集成分类算法,显著提高了不平衡数据的分类精度;文献[20]优化集成学习算法结构,有效识别了核电站事故类型。本文根据齿轮箱检修维护需求,为实现准确、高效的齿轮箱油品状态识别,提出一种基于 PCA 特征优选和 AdaBoost 集成的齿轮箱油品状态识别方法。首先对采集的油液多参量数据进行清洗,采

用 PCA 对油品特征进行提取,获取有助于分类的油品特征优选子集;然后利用 BP 神经网络建立油品识别基本模型,采用灰狼优化算法(grey wolf optimizer, GWO)对 BP 初始权值与阈值进行搜索寻优,克服传统 BP 神经网络超参数选择困难、易陷入局部极小等缺点,同时采取自适应提升 AdaBoost 算法对模型进行优化,进一步提高模型的精度。

1 PCA-GWO-BP-AdaBoost 识别模型

1.1 PGBA 模型结构

基于齿轮箱运维工作中对油品状态识别的实际需求,将齿轮箱油品分为正常、轻度变质、重度变质、严重变质等 4 个状态,提出图 1 所示的 PGBA 识别模型结构。由数据清洗模块(data cleaning module, DCM)、油品特征优选模块(oil product feature selection module, OPFSM)、GWO-BP 识别模块(gwo-bp recognition module, GBRM)、自适应提升集成模块(adaboost integrated module, AIM)4 个部分组成。DCM 采用箱型图和 smote 插值对输入的油液数据集依次进行缺失值及重复值的清洗、异常值的识别与修正,为后续油品特征优选提供良好的数据基础;OPFSM 能以较低的信息损失优选出有助于识别的油品特征优选子集;GBRM 对输入的油品特征优选子集进行初步处理,训练可识别正常、轻度变质、重度变质、严重变质四类油品状态的 BP 基识别器;AIM 将设置的 Q 个 GWO-BP 识别模块作为弱分类器进行训练,使其相互补充,集成为具有较强鲁棒性的强分类器,并最终输出油品状态的判断结果。

1.2 OPFSM

OPFSM 采用 PCA 算法^[21],基于方差度量原始信息的变异程度,以大于 95% 的信息保留率,优选出 k 个主成分为基,以投影方差最大化为目标,对原始数据进行线性变换,提取油品特征优选子集 U。具体实现步骤如下:

1) 计算清洗后数据集 \mathbf{X} 每个特征的样本均值 \bar{x}_m 和标准差 σ_m ,并对所有特征去中心化和标准化得到规范化矩阵 \mathbf{X}^* ,公式如下:

$$\left\{ \begin{array}{l} \bar{x}_m = \frac{\sum_{j=1}^d x_j}{d} \\ \sigma_m = \sqrt{\frac{\sum_{j=1}^d (x_j - \bar{x}_m)^2}{d-1}} \\ \hat{x}_m = \frac{x_m - \bar{x}_m}{\sigma_m} \end{array} \right. \quad (1)$$

式中: d, n 分别为数据集 \mathbf{X} 的样本数和特征数, $j = 1, 2, \dots, d$ 。

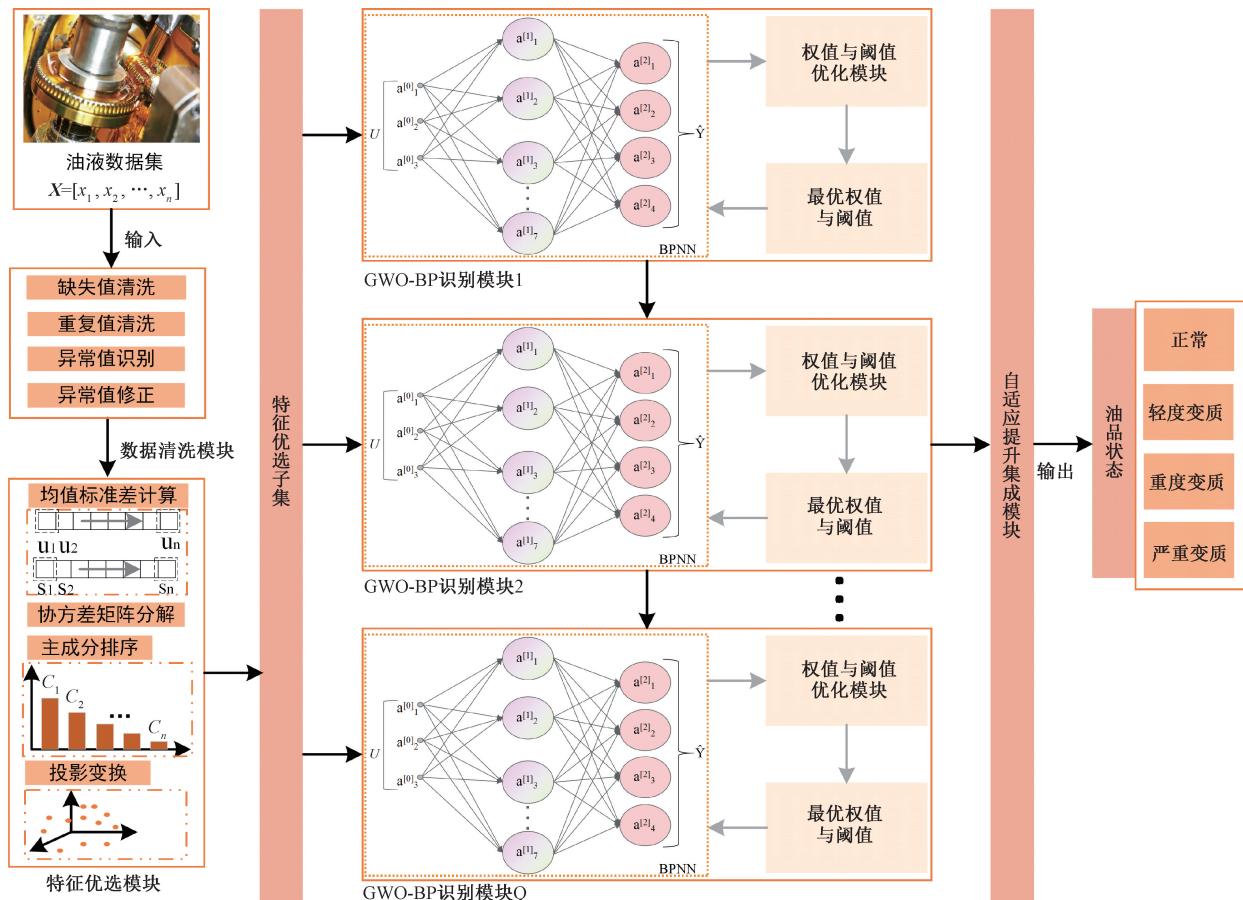


图 1 PGBA 模型结构

Fig. 1 PGBA Model structure

…, d 为样本序号, $m = 1, 2, \dots, n$ 为特征序号。

2) 将协方差矩阵 $R_{\text{cov}} = \frac{1}{n-1} X^* \text{Tr} X^*$ 正交对角化, 分

解为特征值 $\lambda = \{\lambda_1, \lambda_2, \dots, \lambda_n\}$ 和对应的特征向量 $\zeta = \{\zeta_1, \zeta_2, \dots, \zeta_n\}$ 。

3) 按照特征值从大到小对特征向量排序得到主成分组 $\phi = \{\phi_1, \phi_2, \dots, \phi_n\}$ 。

4) 计算主成分组的累计方差贡献率 θ_s , 选择前 k ($k < n$) 个主成分组成投影矩阵 D , 对 X^* 进行投影变换, 公式如下:

$$\theta_s = \frac{\sum_{s=1}^k \lambda_s}{\sum_{s=1}^n \lambda_s} \times 100\% \geq 95\% \quad (2)$$

$$U = DX^* \quad (3)$$

1.3 GBRM

GBRM 由 BP 基识别器、权值与阈值优化模块组成。基于 BP 神经网络结构简单、非线性映射能力强、自适应良好等优势, 将其用于油品状态的初步分类与识别; 权值

与阈值优化模块采用 GWO^[22] 算法强化 BP 基识别器的精度, 解决 BP 神经网络算法随机选取初始权值和阈值而导致的模型低精度问题, 具体实现步骤如下:

- 1) 初始化灰狼种群个数 P , 最大迭代次数 T 和随机向量 A, C 。
- 2) 随机初始化灰狼个体位置。
- 3) 计算每个灰狼的适应度值, 并依次保存适应度值最优的 α 狼、 β 狼、 γ 狼的位置信息。
- 4) 更新灰狼个体的位置。
- 5) 更新参数 A 和 C 。
- 6) 计算全部灰狼的适应度值, 并更新三匹头狼的最优位置。
- 7) 重复执行步骤 2)~6) 直到达到最大迭代次数 T , 输出最优权值 W 与阈值 B 。
- 8) 将 W, B 赋予 BP 神经网络进行训练。
- 9) 训练完成后, 计算 BP 基识别器的分类精度。

1.4 AIM

AIM 采用 AdaBoost 集成学习框架如图 2 所示, 在迭代训练过程中, 根据上一个 GBRM 的分类错误率调整样

本数据的权重分布,使得下一个 GBRM 训练时更关注分类错误率较大的样本,并根据每个 GBRM 弱分类器的分类误差进行自适应加权进一步提升模型精度。

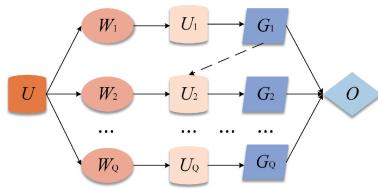


图 2 AIM 框架

Fig. 2 AIM framework

具体实现步骤如下:

1) 输入数据集 $U = \{u_1, u_2, \dots, u_{i'}\}$, 并设置弱分类器个数为 Q 、初始化样本权重 $w_{i'} = 1/N$ 。其中, $i' = 1, 2, \dots, N$ 为样本序号, N 为样本数。

2) 使用权重 $w_{i'}$ 的数据集 $U_{i'}$ 训练弱分类器 $G_{m'}(u)$ 其中, $m' = 1, 2, \dots, Q$ 为基学习器序号。

3) 计算弱分类器 $G_{m'}(u)$ 的分类错误率 $e_{m'}$:

$$e_{m'} = \frac{\sum_{i'=1}^N w_{i'} \prod [G_{m'} \neq y_{i'}]}{\sum_{i'=1}^N w_{i'}} \quad (4)$$

式中: $y_{i'}$ 为真实值。

4) 更新弱分类器 $G_{m'}(u)$ 的权重系数 $g_{m'}$:

$$g_{m'} = \log\left(\frac{1 - e_{m'}}{e_{m'}}\right) + \log(p - 1) \quad (5)$$

式中: p 为类别数。

5) 更新样本分布并归一化权重:

$$w_{i'} = w_{i'} \exp(g_{m'} \cdot \prod [G_{m'} \neq y_{i'}]) \quad (6)$$

6) 重复执行步骤 2)~5), 完成 Q 个 GBRM 弱分类器的训练。

7) 根据权重系数组合 GBRM 弱分类器, 构建最终的强分类器 O_{final} :

$$O_{final} = \operatorname{sgn} \sum_{m'=1}^Q g_{m'} G_{m'}(u) \quad (7)$$

8) 重复执行步骤 2)~7) 直到模型精度满足要求, 输出 O_{final} 。

2 评价指标与实验设置

2.1 油品监测指标选取

粘度是衡量油液内摩擦力的关键指标, 不仅反映了油液的使用性能, 更直接关系到其流动性; 粘度过高或过低都可能影响润滑效果, 甚至导致设备故障。密度作为

润滑油的一项基本物理性能指标, 直接反映了润滑油及侵入杂质的密集程度。水分即润滑油的含水量, 是评价润滑油性能优劣的关键性指标。水活性即润滑油中水分的相对含量, 直接关联着润滑油中水分的存在状态。温度既是保障齿轮箱系统正常运行的关键因素, 也是粘度、密度、水分、水活性的影响因素。结合设备生产运营的实际需要和油液各参量间的相互影响情况, 选取了油液的温度、粘度、密度、水分、水活性 5 个参量作为监测指标。

2.2 油品状态等级划分

齿轮箱油品状态等级的划分参考标准 SH/T0475 如表 1 所示。结合上节确定的监测指标, 考虑粘度作为润滑油的根本特性, 选择粘度作为换油标准, 可依次根据粘度进行区间划分。根据齿轮箱检修维护需求, 将齿轮箱油品分为: 正常(新油)、轻度变质(可正常工作, 对齿轮箱损耗影响较小)、重度变质(对齿轮箱损耗影响较大, 需更换)、严重变质(会造成齿轮箱损坏, 引发安全事故)等 4 种状态, 确定齿轮箱油品状态等级划分表如表 2 所示, 以此作为后续油品状态识别的分级标准。

表 1 齿轮箱换油指标

Table 1 Gearbox oil change indicators

分析项目	换油指标
运动粘度(100℃)变化率/%	>20
水分/%	1.0
酸性增加值/%	0.5
铁含量/%	0.5
正戊烷不溶物	2.0

表 2 齿轮箱油品状态等级划分表

Table 2 Classification of oil Status levels for gearboxes

油品状态	正常	轻度变质	重度变质	严重变质
划分等级	I	II	III	IV
运动粘度(100℃)变化率	0%~5%	5%~10%	10%~15%	15%~20%

2.3 样品制备与数据清洗

为了模拟齿轮箱正常运行时润滑油液的劣化情况, 搭建齿轮箱油液在线监测实验平台如图 3(a) 所示, 由齿轮箱、油液监测传感器、24 V 直流电源、RS485 通信模块、上位机组成。由油液监测传感器采集齿轮箱油液处于正常、轻度变质、重度变质、严重变质 4 种状态下(记为 I~IV 级)的温度、水活性、水分、粘度、密度数据, 各状态油样如图 3(b) 所示。油液监测传感器采集各参量参数如表 3 所示。

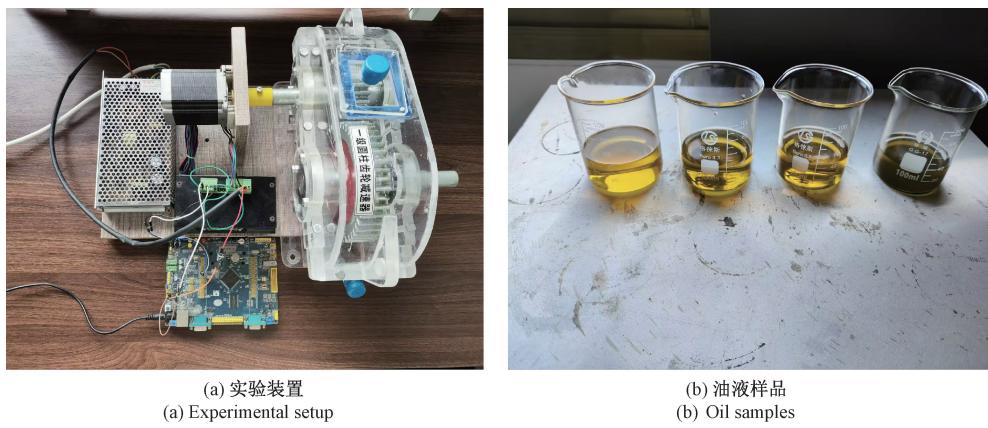


图 3 实验装置及样品
Fig. 3 Experimental setup and samples

表 3 各油液参数

Table 3 Various oil parameter

测量指标	测量范围	分辨率	单位
温度	0~100	0.1	℃
水活性	0~1	0.001	Aw
水分	0~30 000	1	mg/L
粘度	1~400	0.1	mPa·s
密度	600~1 250	0.1	kg/m ³

实验共采集 I~IV 级油液样本共 4 748 组, 删除原始油液数据集中的缺失值, 并删除处于同一温度下的水活性、水分、粘度、密度重复值。而原始油液数据中可能存在受传感器传输过程干扰和齿轮箱运行噪声影响产生的异常数据, 这往往难以直接观察出, 通过绘制油液数据箱形图, 能更加准确地了解油液数据分布特点和其中存在的异常值, 在 30 ℃~70 ℃ 时, I~IV 级油液的水活性、水分、粘度、密度分布如图 4~5 所示。

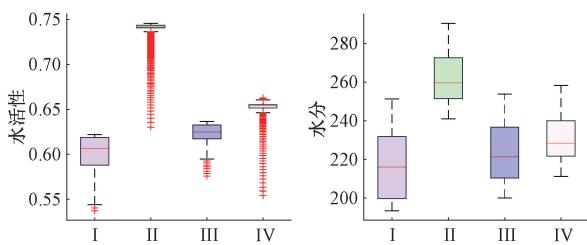


图 4 水活性、水分分布

Fig. 4 Water activity and water distribution

从图中可以看出, I~IV 级油液的水活性、密度分布中均有离群点, 根据箱形图剔除 I~IV 水活性、密度数据中的异常值, 并对 I~IV 水活性、密度进行 smote 插值均衡各类数据以提高后续油品状态识别的准确度。共得到 I~IV 级样本共 3 920 组, 将 I~IV 级样本分别标注为 1、

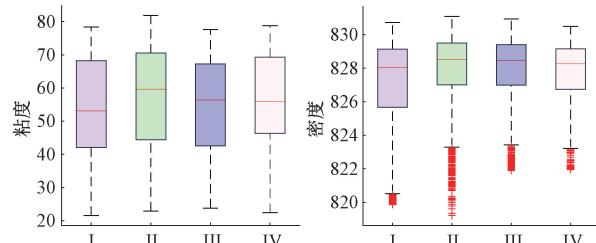


图 5 粘度、密度分布

Fig. 5 Viscosity and density distribution

2、3、4, 构建清洗数据集。

2.4 模型评价指标

模型性能主要通过准确率 A_{cc} 、精确率 P_{macro} 、召回率 R_{macro} 以及综合评价指标 F_1 值衡量, 公式如下:

$$A_{cc} = \frac{L}{M} \quad (8)$$

$$P_{macro} = \frac{\sum_{\xi=1}^4 TP_{\xi}}{4 \sum_{\xi=1}^4 TP_{\xi} + FP_{\xi}} \quad (9)$$

$$R_{macro} = \frac{\sum_{\xi=1}^4 TP_{\xi}}{4 \sum_{\xi=1}^4 TP_{\xi} + FN_{\xi}} \quad (10)$$

$$F_1 = \frac{1}{4} \sum_{\xi=1}^4 \frac{2P_{\xi} \times R_{\xi}}{P_{\xi} + R_{\xi}} \quad (11)$$

式中: M 为总样本数, L 为各类预测正确的样本数之和, 将第 ξ 类当作正类, 其余所有类统一为负类, TP_{ξ} 实际为正类预测为正类, FP_{ξ} 实际为负类预测为正类, FN_{ξ} 实际为正类预测为负类, P_{ξ} 、 R_{ξ} 分别为第 ξ 类按二分类方式计算的精确率、召回率。

3 实验结果与分析

3.1 油品特征优选

采用 PCA 方法对清洗后的油液多参量数据进行特征优选,得到 I~IV 级样本的各主成分贡献率如图 6 所示。根据累计方差贡献率 $>95\%$ 的原则,优选出前 3 个主成分 PC1~PC3 作为投影空间,对各指标进行投影,PCA 三维可视化结果如图 7 所示。可以观察到,I 级和 II~IV 级的样本在以 PC1~PC3 为基的空间中相互独立,II~IV 级样本之间具有相关性。

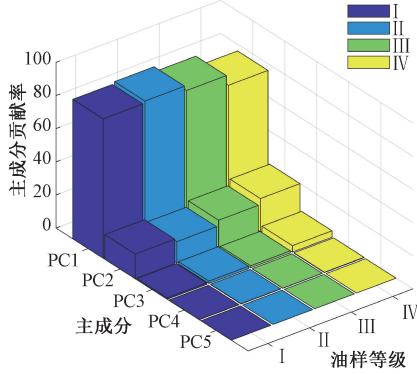


图 6 各主成分贡献率

Fig. 6 Contribution rate of each principal component

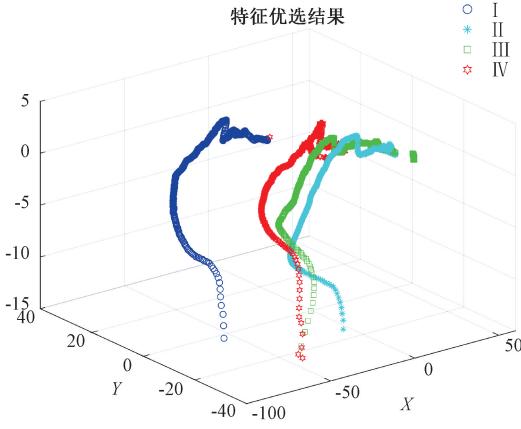


图 7 PCA 特征优选结果

Fig. 7 PCA feature optimization results

3.2 模型训练及参数确定

数据集划分为训练集 2 744 组、测试集 1 176 组,在 MATLAB 中构建 3 层 BP 神经网络进行训练,将 PCA 特征优选后的融合特征作为 BP 基识别器的输入,将油品状态值(分类值 1,2,3,4)作为识别模型的输出,通过文献[23]的经验公式及试凑法确定最佳隐含层节点数为 7。选择分类误差作为 GWO 的适应度函数,将 BP 网络

各神经元的初始权值与阈值编码为实数串,作为 GWO 中的个体,输出最优 BP 初始权值与阈值给 BP 基识别器训练,建立 GWO-BP 弱分类器。

弱分类器的叠加效果决定最终强分类器的分类能力,因此弱分类器数量的确定标准为强分类器在测试集的准确率。设置 GWO-BP 弱分类器开始数量为 1,随后在 5~50 之间取步长 5,依次进行实验,强分类器在训练集、测试集上的准确率表现如图 8 所示。

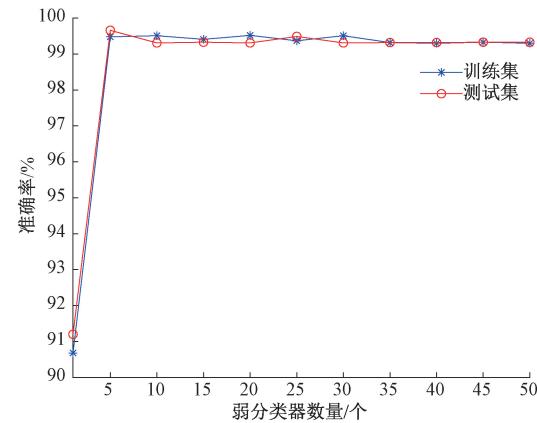


图 8 准确率表现

Fig. 8 Accuracy performance

初始阶段强分类器在测试集上的准确率随弱分类器数量增加呈现较为明显的震荡变化,在弱分类器数量达到 10 之后准确率稳定在 99.3% 左右。因此,选择 10 个弱分类器进行自适应加权集成,获得最终的 PGBA 模型,其参数如表 4 所示。

表 4 模型参数

Table 4 Model parameters

模型	参数		
PGBA 模型	输入层节点数:3	输出层节点数:4	隐含层节点数:7
	BP 迭代次数:100	BP 学习率:0.1	目标误差: 1×10^{-6}
	GWO 种群数:10	GWO 最大迭代次数:10	GWO 优化节点总数:60
	GWO 优化目标参数下限:-1	GWO 优化目标参数下限:1	弱分类器个数:10

GWO 优化 BP 过程如图 9 所示,在训练集和测试集识别结果分别如图 10~11 所示。AdaBoost 集成模型在训练集、测试集上的准确率分别为 99.49%、99.31%,可见分类效果良好。

3.3 各模型性能对比

为了验证所提模型的优越性,分别训练了 SVM、KNN、BN、RBF、BP、GA-BP、PSO-BP、GWO-BP、GWO-BP-AdaBoost 模型用于油品状态识别,使用清洗后的油液数据对 10 种模型算法进行测试,重复 10 次实验取平均值,

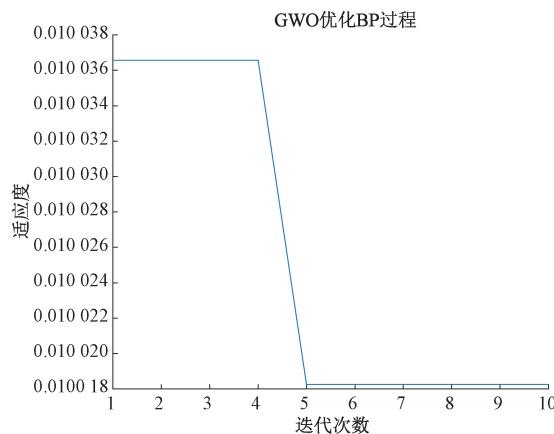


图 9 GWO 优化 BP 过程

Fig. 9 GWO optimization BP process

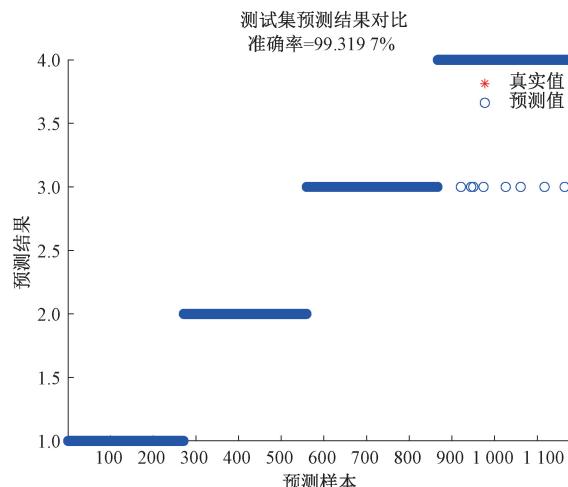


图 11 测试集样本识别结果

Fig. 11 Test set sample recognition results

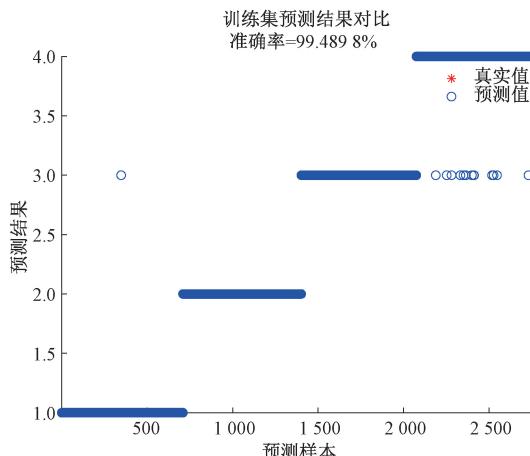


图 10 训练集样本识别结果

Fig. 10 Training set sample recognition

具体结果如表 5 所示。

1) GWO 优化有效性验证

GWO-BP 相比 BP、GA-BP、PSO-BP, 准确率、精准率、召回率、 F_1 值均表现优异, 相比 GWO 和 PSO 寻优, GWO 具有较强的全局搜索能力, 能找到更高质量的全局最优解, 对 BP 识别精度的提升作用显著。此外, 从算法效率表现来看, GWO-BP 比 PSO-BP 表现出较快的收敛速度, 用时更短, 与 GA-BP 用时相当。综合来看, GWO 优化 BP 初始权值与阈值的策略最优。

2) Adaboost 集成有效性验证

GWO-BP-AdaBoost 相比 GWO-BP, 虽然算法的时间复杂度有所增加, 但准确率、精准率、召回率、 F_1 值均得到极大提升, AdaBoost 集成多个 GWO-BP 弱分类器, 并

表 5 各模型性能对比

Table 5 Performance comparison of various models

识别模型	运行时间/s	准确率/%	精准率/%	召回率/%	F_1 /%
SVM	1.66±0.03	65.09±0.83	78.31±2.01	65.09±0.83	59.34±0.75
KNN	1.29±0.04	65.26±0.89	67.51±1.34	65.31±1.15	64.77±1.26
BN	2.06±0.04	70.3±1.35	74.8±0.48	70.42±1.28	70.33±1.81
RBF	5.59±0.33	72.74±1.24	72.14±1.36	72.74±1.24	71.28±1.11
BP	3.57±0.22	78.89±1.87	83.84±3.22	78.89±1.87	75.21±4.33
GA-BP	58.01±0.57	85.21±2.40	89.98±0.96	85.21±2.40	83.81±3.06
PSO-BP	68.77±1.35	82.62±2.74	88.87±1.10	82.63±2.75	80.27±3.75
GWO-BP	62.13±1.25	91.96±1.73	93.33±1.21	91.8±1.76	91.75±1.78
GWO-BP-AdaBoost	71.36±1.38	95.31±0.83	95.53±0.89	95.30±0.70	95.31±0.86
PAGB	32.77±1.27	99.30±0.16	99.33±0.13	99.33±0.15	99.24±0.16

在迭代中更关注易分错的样本, 对 GWO-BP 识别精度的提升作用显著。

3) PCA 特征优选有效性验证

PGAB 相比 GWO-BP-AdaBoost, 准确率、精准率、召

回率、 F_1 值分别提高了 3.99%、3.8%、4.03%、3.93%, 运行时间极大减少, 进一步分析原因可知, 前期的数据清洗虽然清除了一些明显的异常值, 却难以完全清除噪声, 在 PCA 进行特征优选的过程中, 最小的特征值所对应的特

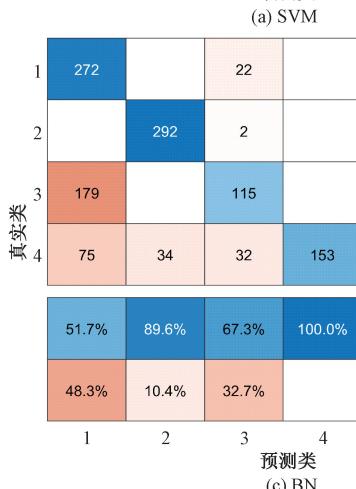
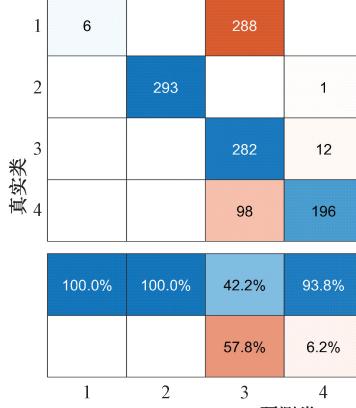
征向量往往与噪声有关,将它们舍弃在一定程度上削减了噪声的影响,而 AdaBoost 分类器对噪声敏感,PCA 在保留主要特征的同时抑制了噪声特征的影响,对 AdaBoost 识别精度和效率的提升具有显著作用。

4) 常见模型对比验证

结合表 5 可知,PGBA 集成模型准确率高达 99.30%,较其他 8 种单一分类器 SVM、KNN、BN、RBF、BP、GA-BP、PSO-BP、GWO-BP 准确率显著提高,分别提高了 34.21%、34.04%、29%、26.56%、20.41%、14.09%、16.68%、7.34%,说明了 PCA 特征优选、GWO 优化 BP 和 AdaBoost 集成策略的有效性。

SVM、KNN、BN、RBF、BP 运行时间均在 6 s 以内,但准确率均低于 80%。优化后的 BP 模型准确率均有所提升,但运行时间均高于 50 s。PGBA 模型的运行时间为 32.77 ± 1.27 s,显著优于 GA-BP、PSO-BP、GWO-BP,体现了 PCA 在提取高维度数据特征方面所具有的优良性能,不仅最大限度保留了数据信息,同时降低了模型的时间复杂度。

5) 分类效果对比验证

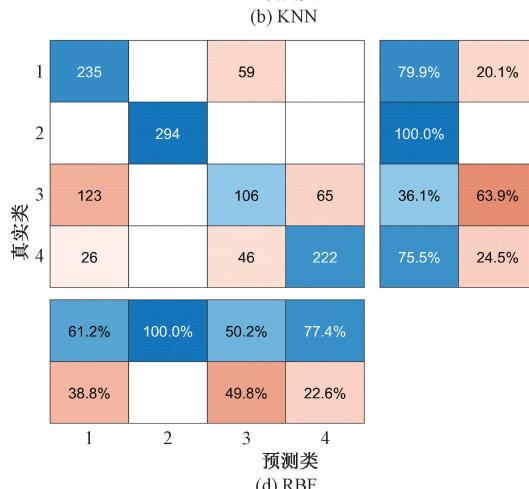
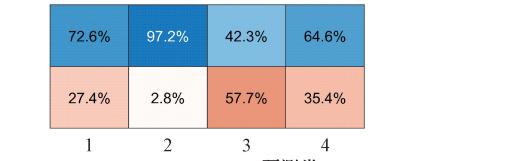
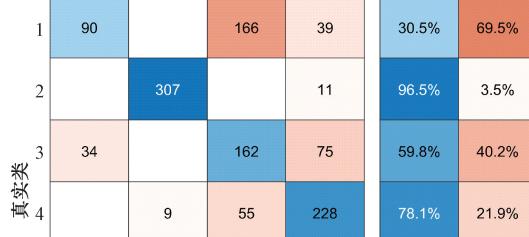


2.0%	98.0%
99.7%	0.3%
95.9%	4.1%
66.7%	33.3%

72.6%	97.2%	42.3%	64.6%
27.4%	2.8%	57.7%	35.4%

为了进一步观察各模型对不同等级油样的分类效果,进行了各个模型的混淆矩阵计算,如图 12(a)~(j) 所示,其中混淆矩阵对角线位置的数据为各类预测正确的样本数。由图 12 可知,PGBA 集成学习模型表现出良好的分类性能,识别结果最佳,对等级 I~III 样品的识别率均达到 100%,IV 级样品中有 8 个被误判为 III 级样品。

另外,对 I 级样品的识别率较分类器 SVM、KNN、BN、RBF、PSO-BP、GWO-BP-AdaBoost 分别提高了 98%、69.5%、7.5%、20.1%、22.8%、4.4%,对 II 级样品的识别率较分类器 SVM、BN 分别提高了 0.3%、3.5%,对 III 级样品的识别率较分类器 SVM、KNN、BN、RBF、BP、PSO-BP、GWO-BP、GWO-BP-AdaBoost 分别提高了 4.1%、40.2%、60.9%、63.9%、85.7%、31%、34.9%、12.2%,对 IV 级样品的识别率较其他 8 种分类器 (a)~(e)、(g)~(i) 分别提高了 30.7%、19.3%、45.4%、21.9%、0.5%、0.1%、0.2%、0.5%,表明基于 PCA 特征优选和 AdaBoost 算法的集成分类器能够较好地实现对齿轮箱润滑油油品状态的检测。



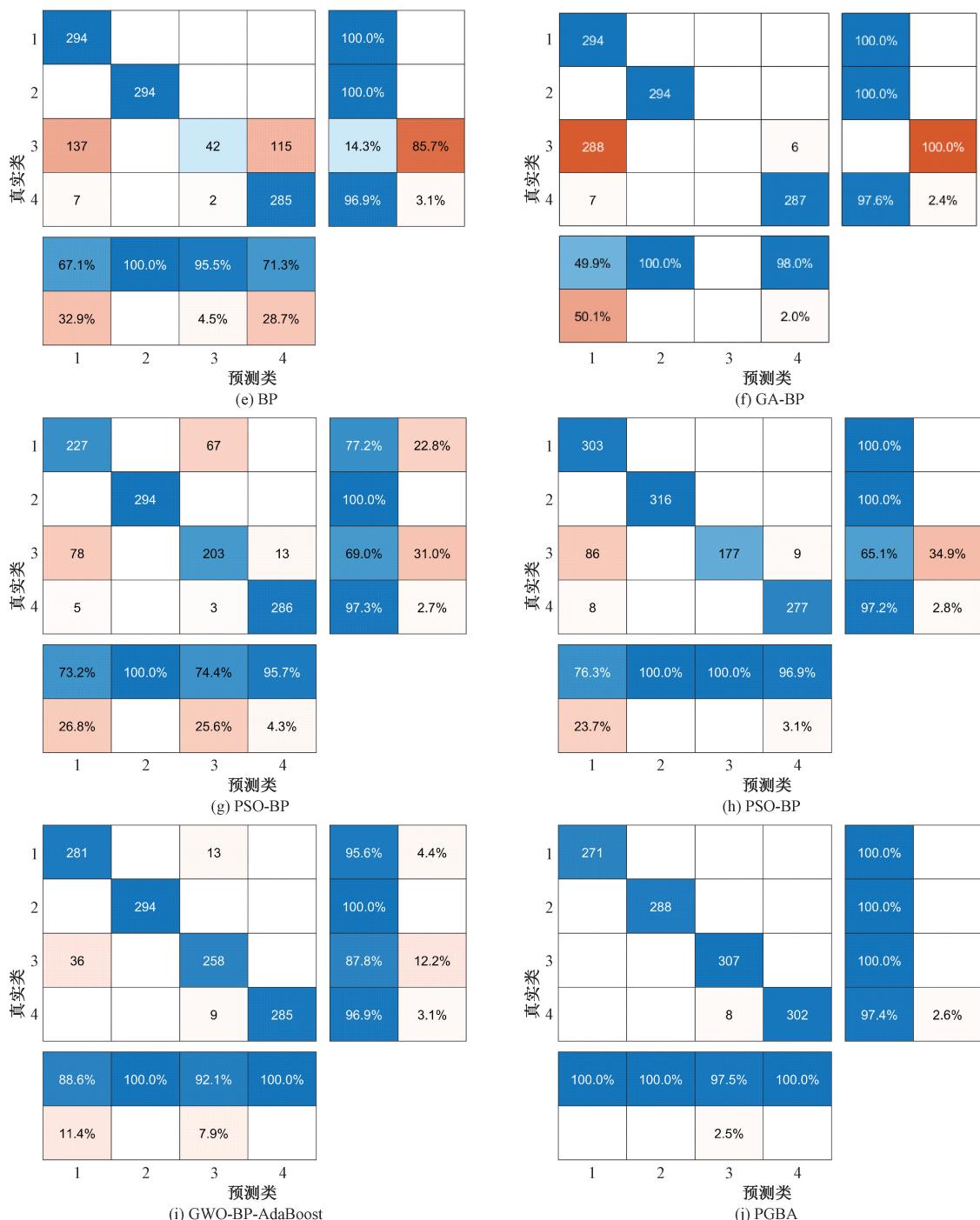


图 12 不同分类器建模结果混淆矩阵

Fig. 12 Confusion matrix of modeling results for different classifiers

4 结 论

针对齿轮箱检修维护的实际需求,提出一种基于 PCA 特征优选和 AdaBoost 集成的齿轮箱油品状态识别方法。该方法通过箱型图和 smote 插值对油液多参量数

据进行清洗提高油液数据的质量;采用 PCA 对油品特征进行提取,获取有助于分类的油品特征优选子集;利用 BP 神经网络建立油品识别基本模型,采用灰狼优化算法对 BP 的初始权值与阈值进行搜索寻优,克服传统 BP 神经网络超参数选择困难、易陷入局部极小等缺点,同时采取自适应提升 AdaBoost 算法对模型进行优化,进一步提

高模型的精度。实验结果表明,与其他9种模型相比,PGBA模型效果最优,平均识别率 $99.30\pm0.16\%$,平均用时 32.77 ± 1.27 s,在优化集成后能以更小的时间代价获得更高的精度,能够有效地识别齿轮箱润滑油油品状态,具有较高的实用价值,为实现在线齿轮箱的油品状态识别奠定了良好基础。提升模型的运行效率与泛化性,以及扩大模型在其他关键设备中的应用将是未来的工作重点。

参考文献

- [1] 唐浩,谭建军,李浩,等.应用滑动轴承的风电齿轮箱行星轮系动力学建模及解耦方法[J].中国机械工程,2024,35(4):591-601.
TANG H, TAN J J, LI H, et al. Dynamic modeling and decoupling method of planetary gear train of wind power gearbox using sliding bearing [J]. China Mechanical Engineering,2024,35(4):591-601.
- [2] 苏舟,石娟娟,关云辉,等.频域特征驱动的车辆旋转部件灰色故障诊断[J].振动.测试与诊断,2024,44(3):514-522,619-620.
SU ZH, SHI J J, GUAN Y H, et al. Grey fault diagnosis of vehicle rotating parts driven by frequency domain characteristics [J]. Vibration Testing and Diagnosis,2024,44(3):514-522,619-620.
- [3] 姜陈,傅双波,高原.动车组齿轮箱润滑油更换周期优化[J].润滑与密封,2023,48(5):179-184.
JIANG CH, FU SH B, GAO Y. Optimization of lubricating oil replacement cycle for gearbox of EMU [J]. Lubrication and Sealing,2023,48(5):179-184.
- [4] XU B, WEN G R, SU Y, et al. Application of multi-level information fusion for wear particle recognition of ferrographic images [J]. Optics and Precision Engineering, 2018, 26(6):1551-1560.
- [5] WANG SH, WU T H, YANG L F, et al. Three-dimensional reconstruction of wear particle surface based on photometric stereo [J]. Measurement, 2019, 133: 350-360.
- [6] PENG Y P, WU T H, WANG SH, et al. Wear state identification using dynamic features of wear debris for on-line purpose [J]. Wear, 2017, 376: 1885-1891.
- [7] MARKOVA L V. Intelligent method for monitoring the state of lubricating oil [J]. Journal of Friction and Wear, 2016, 37: 308-314.
- [8] 顾伟红,毛梦薇.改进CNN在TBM变速箱磨损状态识别中的应用[J].隧道建设,2023,43(11):1853-1861.
GU W H, MAO M W. Application of improved CNN in TBM gearbox wear state recognition [J]. Tunnel Construction,2023,43(11):1853-1861.
- [9] 徐启圣,王俊,白琨,等.数据驱动的风电齿轮箱油液监测磨损度量化及其参数权重研究[J].润滑与密封,2020,45(11):125-129.
XU Q SH, WANG J, BAI K, et al. Study on data driven oil monitoring wear degree quantification and parameter weight of wind power gearbox [J]. Lubrication and Sealing,2020,45(11):125-129.
- [10] CHI Y J, PORIKLI F. Classification and boosting with multiple collaborative representations [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2013,36(8): 1519-1531.
- [11] 马森财,赵荣珍,吴耀春.云模型和集成分类结合的故障数据不平衡学习[J].振动.测试与诊断,2023,43(6):1114-1120,1243.
MA S C, ZHAO R ZH, WU Y CH. Fault data imbalance learning based on cloud model and ensemble classification [J]. Vibration. Testing and Diagnostics,2023,43(6):1114-1120,1243.
- [12] 潘礼正,王顺超,丁亿,等.基于全视域特征表征与ELM-Adaboost方法的情绪辨识研究[J].仪器仪表学报,2022,43(11):162-171.
PAN L ZH, WANG SH CH, DING Y, et al. Research on emotion recognition based on full-sight feature representation and ELM-Adaboost method [J]. Chinese Journal of Scientific Instrument,2022,43(11):162-171.
- [13] 陈博桓,王馨雨,许学彬,等.基于优化ELM的光纤连接器表面自识别降噪技术[J].电子测量与仪器学报,2022,36(4):169-178.
CHEN B H, WANG X Y, XU X B, et al. Surface self-identification and noise reduction technology of optical fiber connector based on optimized ELM [J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(4):169-178.
- [14] 常梦容,王海瑞,肖杨.mRMR特征筛选和随机森林的故障诊断方法研究[J].电子测量与仪器学报,2022,36(3):175-183.
CHANG M R, WANG H R, XIAO Y. Research on fault diagnosis method of mRMR feature screening and random forest [J]. Journal of Electronic Measurement and Instrumentation,2022,36(3):175-183.
- [15] 李志刚,胡德安,纪勇,等.阿尔茨海默病多病程自适应筛查模型的构建[J].仪器仪表学报,2023,44(1):182-189.
LI ZH G, HU D AN, JI Y, et al. Construction of a multi-course adaptive screening model for Alzheimer's disease [J]. Chinese Journal of Scientific Instrument, 2023, 44(1):182-189.
- [16] NIETO P J G, GARCÍA - GONZALO E, ARBAT G, et al. Construction of a multi-course adaptive screening model for Alzheimer's disease [J]. Chinese Journal of Scientific Instrument, 2023, 44(1):182-189.

- al. Pressure drop modelling in sand filters in micro-irrigation using gradient boosted regression trees [J]. Biosystems engineering, 2018, 171: 41-51.
- [17] 虞瑤, 杨锦, 邱洁. 随机子空间和 MRF 的高光谱遥感集成学习算法 [J]. 测绘科学, 2022, 47 (10): 161-167, 175.
- YU Y, YANG J, QIU J. Hyperspectral remote sensing ensemble learning algorithm for stochastic subspace and MRF [J]. Science of Surveying and Mapping, 2022, 47 (10): 161-167, 175.
- [18] 崔靖奇, 吴顺川, 程海勇, 等. 滇中引水软岩隧洞围岩位移时序预测 [J]. 清华大学学报 (自然科学版), 2024, 64 (7): 1215-1225.
- GUI J Q, WU SH CH, CHENG H Y, et al. Temporal series prediction of surrounding rock displacement of soft rock tunnel in central Yunnan [J]. Journal of Tsinghua University (Natural Science Edition), 2024, 64 (7): 1215-1225.
- [19] 刘宁, 朱波, 荆晓娜, 等. 一种基于改进 CGAN 的不平衡数据集分类算法 [J]. 小型微型计算机系统, 2023, 44 (9): 1918-1924.
- LIU N, ZHU B, JING X N, et al. An integrated classification algorithm for unbalanced data based on improved CGAN [J]. Journal of Chinese Computer Systems, 2023, 44 (9): 1918-1924.
- [20] 李翔宇, 程坤, 谭思超, 等. 基于 Adaboost 算法的核电站故障诊断方法 [J]. 核动力工程, 2022, 43 (4): 118-125.
- LI X Y, CHENG K, TAN S CH, et al. Nuclear power plant fault diagnosis method based on Adaboost algorithm [J]. Nuclear Power Engineering, 2022, 43 (4): 118-125.
- [21] 夏加富, 叶奕君, 郭嘉俊, 等. 基于支持向量机与特征降维的直流断路器机械故障诊断技术研究 [J]. 高压电器, 2024, 60 (2): 51-61.
- XIA J F, YE Y J, GUO J J, et al. Research on mechanical fault diagnosis technology of DC circuit breaker based on support vector machine and feature dimensionality reduction [J]. High Voltage Electrical Apparatus, 2024, 60 (2): 51-61.
- [22] SHAHEEN M A M, HASANIEN H M, ALKUHAYLI A. A novel hybrid GWO-PSO optimization technique for optimal reactive power dispatch problem solution [J]. Ain Shams Engineering Journal, 2021, 12 (1): 621-630.
- [23] 闫鹏程, 尚松行, 张超银, 等. 改进 BP 神经网络算法对煤矿水源的分类研究 [J]. 光谱学与光谱分析, 2021, 41 (7): 2288-2293.
- YAN P CH, SHANG S X, ZHANG CH Y, et al. Research on the classification of coal mine water source by improved BP neural network algorithm [J]. Spectroscopy and Spectral Analysis, 2021, 41 (7): 2288-2293.

作者简介



陈晓犇, 2020 年于安徽建筑大学城市建设学院获得学士学位, 2024 年于湖南科技大学获得硕士学位, 主要研究方向为智能检测技术与自动化装置。

E-mail: 1663973240@qq.com

Chen Xiaoben received B. Sc. Degree from Anhui Jianzhu University Urban Construction in 2020, and received M. Sc. from Hunan University of Science and Technology in 2024. Her main research interests include intelligent detection technology and automation devices.



黄采伦 (通信作者), 1990 年于湘潭矿业学院获得学士学位, 2002 年于湖南大学获得硕士学位, 2007 年于中南大学获得博士学位, 现为湖南科技大学教授, 主要研究方向为智能检测技术与自动化装置、地雷探测、设备状态监测与故障诊断、矿区水害防治。

E-mail: 806313410@qq.com

Huang Cailun (Corresponding Author) received B. Sc. Degree from Xiangtan Mining College in 1990, received M. Sc. from Hunan University in 2002, and received Ph. D. from Central South University in 2007. Now he is a professor of Hunan University of Science and Technology. His main research interests include the research of intelligent detection technology and automation devices, landmine detection, equipment status monitoring and fault diagnosis, mining area water hazard prevention and control.