基于IWOA-LightGBM模型的矿用挖掘机发动机故障诊断研究*

顾清华[1,2,3],白书宇[1,2],王丹[1,3]

(1.西安建筑科技大学 管理学院, 陕西 西安 710055;
2.西安市智慧工业感知、计算与决策重点实验室, 陕西 西安 710055;
3.西安建筑科技大学 资源工程学院, 陕西 西安 710055)

摘要:针对矿用挖掘机发动机故障类别不均衡,导致故障诊断精度不高的问题,提出了一种改进的鲸鱼算法(WOA)优化轻量级梯度提升机(LightGBM)的矿用挖掘机发动机智能故障诊断方法。首先,利用递归特征交叉验证消除法(RFECV)对采集的挖掘机发动机故障数据的特征进行提取,删除不相关的特征。其次,采用Focal-Loss改进LightGBM的损失函数,提出一种改进的WOA对LightGBM的超参数寻优,构建新的诊断模型。最后,利用某矿山挖掘机发动机故障数据进行验证,并与常见的集成模型、调优框架和诊断算法进行对比分析。结果表明:所提出的矿用挖掘机发动机故障诊断模型IWOA-LightGBM的准确率和F_1分数分别为98.08%和98.53%,诊断性能较好,可为矿山机械设备的智能诊断提供参考。

关键词:矿用挖掘机;发动机;故障诊断;递归特征交叉验证消除法;轻量级梯度提升机;鲸鱼算法

中图分类号:TD422.2;TP277 **文献标识码**:A **文章编号**:1005-2763(2025)09-0184-08

Research on Fault Diagnosis of Mining Excavator Engine Based on IWOA-LightGBM Model

GU Qinghua[1,2,3], *BAI Shuyu*[1,2], *WANG Dan*[1,3]

(1.School of Management, Xi'an University of Architecture and Technology, Xi'an, Shaanxi 710055, China;
2.Xi'an Key Laboratory of Perception, Computing and Decision Making for Intelligent Industry, Xi'an, Shaanxi 710055, China;
3.School of Resources Engineering, Xi'an University of Architecture and Technology, Xi'an, Shaanxi 710055, China)

Abstract: Aiming at the problem of low fault diagnosis accuracy caused by unbalanced fault categories of mining excavator engine, an intelligent fault diagnosis method of mining excavator engine based on improved whale optimization algorithm (WOA) to optimize light gradient boosting machine (LightGBM) was proposed. Firstly, the recursive feature cross-validation elimination method (RFECV) was used to extract the features of the fault data of collected excavator engine, and the irrelevant features were deleted. Secondly, Focal-Loss was used to improve the loss function of LightGBM, and an improved WOA was proposed to optimize the hyperparameters of LightGBM, and a new diagnostic model was constructed. Finally, the engine fault data of a mine excavator was used for verification, and compared with common integrated models, tuning frameworks and diagnostic algorithms. The results show that the accuracy and F1 score of the proposed mining excavator engine fault diagnosis model IWOA-LightGBM are 98.08% and 98.53% respectively, and the diagnostic performance is good, which can provide reference for the intelligent diagnosis of mining machinery and equipment.

Key words: Mining excavator, Engine, Fault diagnosis, Recursive feature cross-validation elimination method, Light gradient boosting machine, Whale optimization algorithm

0 引言

随着智慧矿山建设的不断推进[1],矿山机械设备的诊断技术也逐渐向智能化和数字化方向发展。矿用挖掘机[2]作为典型的复杂分布式机电系统,具有结构层次复杂、分系统和零部件多等特点,如果发生故障时不能及时诊断,不仅会影响企业的经济效益,还会影响人身安全。因此,为了提高企业的效益和保证安全生产,需要对矿用挖掘机的故障进行精准诊断和分析,从而保证挖掘机的正常运行。目前

* **收稿日期**:2024-09-29
基金项目:国家自然科学基金项目(52074205,52374135)
作者简介:顾清华(1981—),男,山东诸城人,博士,教授,主要从事矿山智能科学与工程、故障诊断、智能优化算法研究工作。E-mail:qinghuagu@126.com
通信作者:白书宇(1999—),男,江苏徐州人,硕士研究生,主要从事矿山设备故障诊断研究。E-mail:baishuyu@xauat.edu.cn

对矿用挖掘机的故障诊断研究大多集中在液压系统[3]和减速器[4],而对发动机的诊断较少。在露天矿山复杂多变的工作环境中,矿用挖掘机的发动机一直处于高压、高负载的工作状态,"带病"工作现象较为普遍。虽然国内大型重工企业[5]上线了智能设备监控系统和诊断平台,但与国外相比智能化差距依然很大,很多故障仍然需要设备停机进行诊断,导致诊断效率和精度较低,因此寻找一种鲁棒性较好的辅助方法具有重要意义。

基于人工智能、大数据、物联网等新一代信息技术的发展,以数据驱动为代表的诊断方式应运而生,其中机器学习[6]是以数据驱动诊断的核心技术,包括神经网络、支持向量机、深度学习。有学者采用改进的麻雀算法对核支持向量机的惩罚因子和权重进行参数寻优,构建 CSSA-F-SVM 诊断模型,并将其应用到矿用卡车发动机的故障诊断中,提高了矿卡发动机故障识别的可靠性和准确性[7]。CAI 等[8]使用小波阈值去噪和集成经验模态分解来处理振动信号,从分解的本征模态函数中提取敏感的特征值,将基于规则的算法与 BP 神经网络相结合,实现了较宽速度范围内的柴油发动机故障检测。刘弘毅等[9]针对传感器提取的压力信号,通过核鲁棒流形非负矩阵分解和融合特征结合来加工信号,输入到 CNN 中,显著提高了柴油机的诊断效率。上述方法诊断的精度确实有所提高,但 SVM、神经网路均属于单一分类器,容易出现过拟合和泛化能力弱的现象,基于深度学习的方法适用于数据量大、且存在训练时间长的问题,对于中小样本故障诊断的适用性不强。集成学习既能克服单一分类器的不足,又能在中小型数据的分类问题上具有优势,其中以轻量级梯度提升机(LightGBM)这个集成框架最为有效,但是 LightGBM 存在众多的超参数,存在不确定性。陈思勤等[10]提出了一种基于改进的灰狼优化算法优化的 LightGBM 磨煤机故障诊断模型,取得了更高的精度和泛化能力,也有学者使用 LightGBM 作为旋转森林的基学习器,在挖掘机的故障诊断中取得了不错的效果[11]。但上述改进方向均未考虑到数据分类的不平衡性和特征的自动选择性。

因此,针对以上问题,本文提出考虑到数据分类不平衡和特征自动选择的 LightGBM 框架的故障诊断模型。首先通过递归特征消除和交叉验证相结合的方法,自动完成特征筛选,解决特征难以筛选的问题;其次从两方面改进 LightGBM,一是将 Focal-loss 作为 LightGBM 的损失函数,解决分类不平衡的数据集,二是考虑到超参数对 LightGBM 的影响,采取改进的鲸鱼算法对其寻优,构建最终的 LightGBM 故障诊断模型。

1 理论基础

1.1 RFECV 特征提取

RFECV 是递归特征消除和交叉验证相结合的特征选择方法。该方法可以根据特征集自动选取特征数量,通过递归移除最不重要的特征来开展工作。在每一步中,RFECV 会移除一个特征,然后使用剩余的特征重新训练模型,以上过程会反复进行,直至达到自动确定的特征集个数为止。利用 RFECV 进行特征筛选时,需要选取合适的基模型以获取特征重要性和性能得分。随机森林[12](Random Forest)具有防止过拟合、处理高维数据、能够给出特征重要性等优点,因此,本文选取随机森林作为基模型,基于 RFECV 的特征筛选流程如图 1 所示。

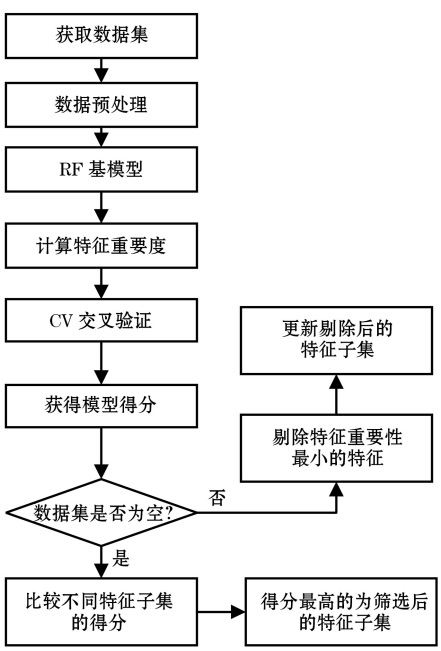


图 1 基于 RFECV 的特征筛选流程

Fig.1 Feature selection process based on RFECV

1.2 LightGBM

LightGBM 模型搭建于 KE 等[13]提出的梯度决策提升树算法(GBDT)框架之上,与极端提升树(XGBoost)、分类集成树(CatBoost)、随机森林(Randdom Forest, RF)相比,LightGBM 具有更快的训练速度、更低的内存消耗、更好的准确率,且支持分布式计算。一方面,LightGBM 通过直方图算法、单边梯度采样和互斥特征捆绑对样本和特征进行

采样,极大地降低模型的复杂度和内存占用率,另一方面,LightGBM 采用带深度限制的按叶生产策略,不仅降低了算法的误差,还加快了算法的迭代速度。

2 基于 IWOA-LightGBM 的故障诊断模型

2.1 Focal-Loss 损失函数

传统多分类问题的损失函数是交叉熵损失函数,交叉熵损失函数更多偏向于预测样本数多的类别,从而忽略少样本量的类别,会存在两个方面的问题,一是如果其中某个故障类型数据集多,在训练过程中会更多地关注这个故障类型多的数据集,从而会造成无意义的训练学习,二是单纯地关注某个故障数据集多的类型会导致整体模型的退化。而 Focal-Loss[14] 会增加困难样本的分类权重,降低简单样本的权重,让模型在训练过程中更多关注困难样本,即故障类别数据比较少的类型,从而完成一个更好的故障训练。

$$FL(p_t)=-\alpha(1-p_t)^{\gamma}\log p_t \tag{1}$$

式中:p_t 表示预测类别的概率;α 表示平衡焦距损失是从类别权重调节设置的,$\alpha\in(0,1)$;γ 为调制因子,表示难度权重调节,$\gamma\in(0,5)$,γ 为 2 的效果最佳[14]。

图 2 和图 3(曲线颜色标识见电子版)分别为不同 γ 取值和 α 取值对损失函数的影响,由图 2 可知,增加 γ 会不断增加负样本的损失函数的值。由图 3 可知,增加 α 会不断降低负样本的损失函数的值。

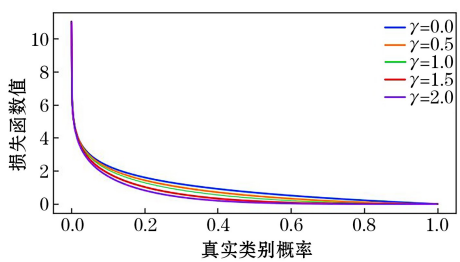


图 2 γ 对损失函数的影响

Fig.2 Effect of γ on the loss function

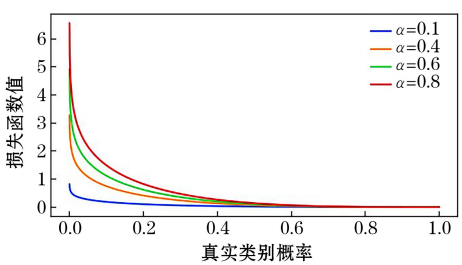


图 3 α 对损失函数的影响

Fig.3 Effect of α on the loss function

2.2 鲸鱼算法及其改进

鲸鱼算法(WOA)是 MIRJALILI 等[15] 于 2016 年提出的一种进化智能算法,相比于其他智能算法,具有求解流程简单、参数少等优点,已广泛应用在工程优化领域,但还面临着较差的局部和全局搜索能力、二者的不平衡和种群多样性低等问题,本文采用池化机制、迁移搜索策略、优先选择搜索策略、包围搜索策略等 4 种策略[16] 对鲸鱼算法进行改进,并提出一种改进的鲸鱼算法(IWOA)。

2.2.1 池化机制

池化机制是一种交叉运算准则,通过将当代种群中最好的解和最差的解聚合在一起,每当完成池达到设置比列的容量时,就会由新的解替换当前完成池的解,具体计算公式如下:

$$\boldsymbol{P}_i^t=\boldsymbol{B}_i^t\times\boldsymbol{X}_{\text{brnd}}^t+\bar{\boldsymbol{B}}_i^t\times\boldsymbol{X}_{\text{worst}}^t \tag{2}$$

假设给定的资源池为 $(P_1,P_2,\cdots,P_k)$,其中 $\boldsymbol{P}_i=(P_{i,1},P_{i,2},\cdots,P_{i,D})$ 表示第 i 代第 P 个鲸鱼个体,向量维度为 D,每一代结束前用公式(2)产生的,在这个公式(2)中,$\boldsymbol{X}_{\text{worst}}^t$ 是第 t 代中最差的解,$\boldsymbol{X}_{\text{brnd}}^t$ 使用公式(5)计算第 t 代中最佳头鲸的邻域内产生一个随机位置,$\boldsymbol{B}_i^t$ 是一个二元[0,1]向量,$\bar{\boldsymbol{B}}_i^t$ 和 $\boldsymbol{B}_i^t$ 对应位置非 0 即 1。

2.2.2 迁移搜索策略

此搜索策略使用公式(3)随机分离鲸群的一部分,用于探索未知的空间,从而覆盖未访问的区域并改善探索。此外,分离的鲸群在一定程度上会增加整个种群的多样性,从而避免陷入局部最优解的状态,具体计算公式如下:

$$\boldsymbol{X}_i^{t+1}=\boldsymbol{X}_{\text{rnd}}^t-\boldsymbol{X}_{\text{brnd}}^t \tag{3}$$

$$\boldsymbol{X}_{\text{rnd}}^t=rand\times(\boldsymbol{\delta}_{\max}-\boldsymbol{\delta}_{\min})+\boldsymbol{\delta}_{\min} \tag{4}$$

$$\boldsymbol{X}_{\text{brnd}}^t=rand\times(\boldsymbol{\delta}_{\text{best_max}}-\boldsymbol{\delta}_{\text{best_min}})+\boldsymbol{\delta}_{\text{best_min}} \tag{5}$$

式中:$\boldsymbol{X}_i^{t+1}$ 表示第 $t+1$ 代中第 i 个鲸鱼的位置;$\boldsymbol{X}_{\text{rnd}}^t$ 为搜索空间中的任意一个位置;$\boldsymbol{\delta}_{\max}$ 和 $\boldsymbol{\delta}_{\min}$ 分别为搜索空间上界、下界构成的向量;$\boldsymbol{\delta}_{\text{best_max}}$ 和 $\boldsymbol{\delta}_{\text{best_min}}$ 分别为第 t 代最优个体的上界和下界;$\boldsymbol{X}_{\text{brnd}}^t$ 为头鲸附近的随机位置;$rand$ 为 0 到 1 的随机数。

2.2.3 优先选择搜索策略

此策略可以提升鲸群探索最佳值的能力,计算公式见式(6),其中,$\boldsymbol{P}_{\text{rnd1}}^t$ 和 $\boldsymbol{P}_{\text{rnd2}}^t$ 是第 t 代中随机从资源池中挑选出来的个体;$IT_{\max}$ 为最大迭代次数;A_i^t 为系数变量。C_i^t 由公式(7)定义,A_i^t 由公式(8)和公式(9)联合定义。

$$\boldsymbol{X}_i^{t+1}=\boldsymbol{X}_i^t+A_i^t\times(C_i^t\times\boldsymbol{P}_{\text{rnd1}}^t-\boldsymbol{P}_{\text{rnd2}}^t) \tag{6}$$

$$C_i^t=2\times rand \tag{7}$$

$$A_i^t = 2 \times a_i^t \times rand - a_i^t \tag{8}$$

$$a_i^t = 2 - t \times \frac{2}{IT_{\max}} \tag{9}$$

2.2.4 包围搜索策略

此策略呈现了头鲸在识别和包围猎物的状态，其中新的位置由公式(10)给出，扰动因子 $\boldsymbol{D}'^t$ 由公式(11)给出，$\boldsymbol{P}_{\text{rnd3}}^t$ 是从池化机制的资源池中随机选择的个体。

$$\boldsymbol{X}_i^{t+1} = \boldsymbol{X}_{\text{best}}^t - A_i^t \times \boldsymbol{D}'^t \tag{10}$$

$$\boldsymbol{D}'^t = \left| C_i^t \times \boldsymbol{X}_{\text{best}}^t - \boldsymbol{P}_{\text{rnd3}}^t \right| \tag{11}$$

2.3 基于 IWOA-LightGBM 的故障诊断流程

利用 Focal-Loss 替换交叉熵损失函数，由于 LightGBM 超参数[17]很多，存在很大的不确定性，为了提高诊断精度，本文采用 IWOA 对 LightGBM 的超参数进行优化，从而得到泛化性更强的故障诊断模型，具体的超参数信息见表 1。

表 1 LightGBM 超参数信息

Table 1 Hyperparameter information of LightGBM

超参数	寻优范围	说明
树深度	[3,8]	控制数的结构
学习率	[0.01,0.3]	提高准确性
树的个数	[50,150]	提高准确性
正则化因子	[0,100]	避免过拟合
特征提取比例	[0.5,1]	避免过拟合

基于 IWOA-LightGBM 的故障诊断流程如图 4 所示，具体步骤如下：

(1) 输入带有标签诊断的数据集，以 RF 为基学习器，采用递归特征消除法和交叉验证对特征进行提取；

(2) 将 Focal-Loss 写入 LightGBM 模型，并初始化 IWOA 和 LightGBM 的基本参数；

(3) 采用 LightGBM 模型对样本训练集进行 5 折交叉验证，将准确率、F_1 分数等分类指标依次作为适应度函数，计算每个鲸体的适应度；

(4) 对适应度进行降序排列，找出适应度最大的个体位置并记为最优个体(最佳的超参数组合)，记录最佳适应度值；

(5) 按照公式(2)、公式(3)、公式(6)、公式(10)对每个鲸鱼个体的位置进行更新；

(6) 重新计算更新后的鲸鱼个体适应度，与上一代的最佳适应度进行比较，从而更新最优鲸鱼的位置和适应度；

(7) 判断是否满足终止条件，若不满足则返回步骤(4)，否则输出最优的超参数组合，寻优结束；

(8) 训练结束之后，将训练好的模型代入测试集中，从而输出最终的诊断结果。

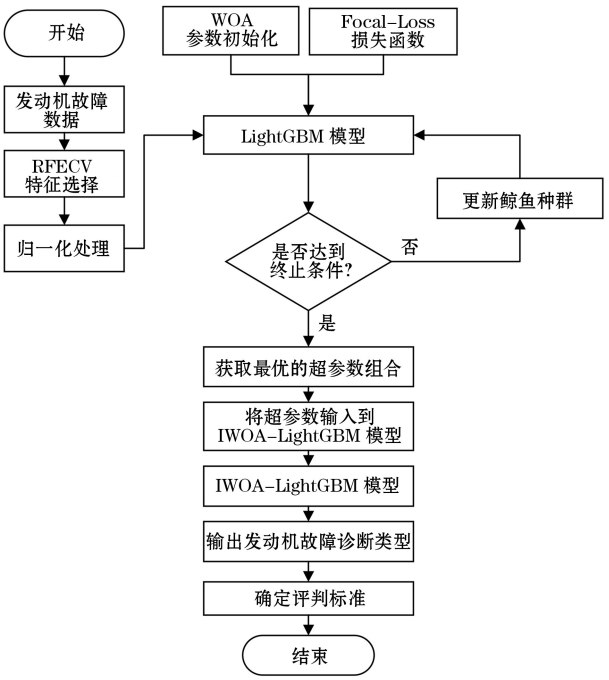


图 4 基于 IWOA-LightGBM 的故障诊断流程

Fig.4 Fault diagnosis process based on IWOA-LightGBM

3 仿真试验和结果分析

为了测试 IWOA-LightGBM 诊断模型的性能，本文以 WP12G460E310 发动机为研究对象，借助潍柴智多星路谱监测装置采集了 260 组 19 维的故障数据，并在 Python 3.9 环境下进行多次仿真试验，验证本文方法应用于矿用挖掘机发动机故障诊断的可行性。

3.1 评估指标

由于矿山作业环境的特殊性，矿用挖掘机发动机的故障类型主要为管道堵塞及滤芯等。本文采集了 7 种常见故障类型数据，分别为高压油泵故障、进气管路堵塞、增压器故障、空滤堵塞、中冷器故障、风扇故障、油门踏板故障，编号分别为 1，2，3，4，5，6，7(发动机正常用编号 0 表示)。将数据按照 7∶3 的比例划分为训练集和测试集，不均衡故障数据集见表 2。并采用 5 倍交叉验证的方法对训练集进行训练。为了更好地衡量分类器的性能，本文不仅考虑到精确率，还考虑了 F_1 分数、AUC，F_1 分数是考虑到召回率和精确率的综合性指标，而 AUC 是 ROC 曲线下的面积，上述指标的值越大，表示模型的性能越好，具体的计算公式如下：

$$\begin{cases} A=\dfrac{TP+TN}{TP+TN+FP+FN} \\ P=\dfrac{TP}{TP+FP} \\ R=\dfrac{TP}{TP+FN} \\ F_1=\dfrac{2\times P\times R}{P+R} \end{cases} \tag{12}$$

式中：TP 表示将故障样本正确预测成故障样本的个数；FP 表示将正常样本错误预测成故障样本的个数；TN 表示将故障样本错误预测成正常样本的个数；FN 表示将正常样本正确预测成正常样本的个数；P 和 R 分别表示精确率和召回率；A 表示准确率。

表 2 不均衡故障数据集

Table 2 Unbalanced fault data set

故障编号	故障类型	样本量/个
0	正常	37
1	高压油泵	30
2	进气管堵塞	47
3	增压器故障	23
4	空滤堵塞	37
5	中冷器故障	33
6	风扇故障	27
7	油门踏板	27

3.2 特征提取

由于采取的故障数据的特征为 19 维特征，为了合理地降维抓取重要信息以及加快模型的训练速度，采用 RFECV 算法对特征进行提取，选取重要的特征子集，故障特征个数和准确率之间的关系如图 5 所示。

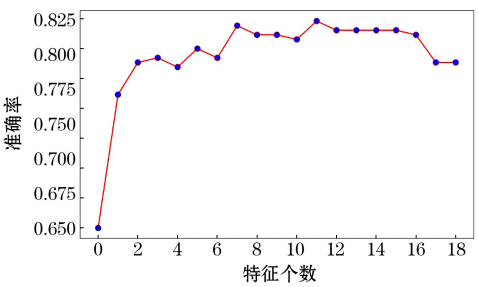


图 5 特征个数与准确率的关系

Fig.5 Relationship between the number of features and accuracy

由图 5 可知，当特征个数约为 12 时，诊断的准确率最高，此时对应的 RFECV 选择的特征重要性排序见表 3。

表 3 选取的特征重要度排序

Table 3 Ranking of selected feature importance

特征序号	特征名称	特征重要性
f_1	发动机转速	0.03
f_2	喷油量设定值	0.04
f_3	当前喷油量	0.06
f_7	冷却液温度	0.10
f_8	机油压力	0.03
f_9	机油温度	0.08
f_{11}	进气压力	0.11
f_{12}	进气温度	0.18
f_{14}	轨压设定值	0.14
f_{15}	轨压峰值	0.16
f_{18}	扭矩负荷率	0.04
f_{19}	输出扭矩	0.03

3.3 集成模型性能对比

为了验证用 IWOA 对 LightGBM 改进优化的有效性，将总体样本按照留一法划分为 70%的训练集和 30%的测试集，采取 5 折交叉验证法训练 70%的训练集，30%的数据用来验证，并将诊断结果与常用的集成树模型算法 XGBoost、CatBoost、Random Forest(RF)进行对比，诊断结果见表 4 和图 6 至图 9。

表 4 不同树模型各指标的评估结果

Table 4 Evaluation results of each index of different tree models %

评估指标	IWOA-LightGBM	IWOA-XGBoost	IWOA-CatBoost	IWOA-RF
A	98.07	94.23	90.38	84.61
P	98.61	93.85	91.65	87.29
R	98.61	94.99	92.48	87.92
F_1 分数	98.52	93.94	91.62	86.29
AUC	99.57	99.07	99.00	98.20

由表 4 可知，IWOA-LightGBM 模型整体的准确率和 F_1 分数最高，说明用 IWOA 优化改进 LightGBM 具有一定的有效性。

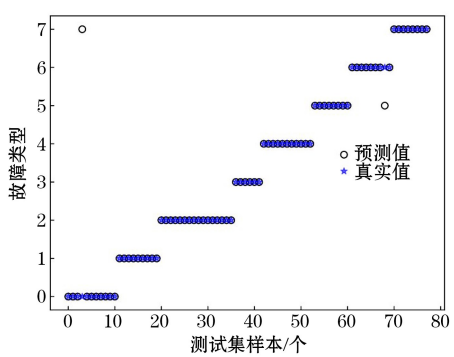


图 6 IWOA-LightGBM 故障诊断结果

Fig.6 Fault diagnosis results of IWOA-LightGBM

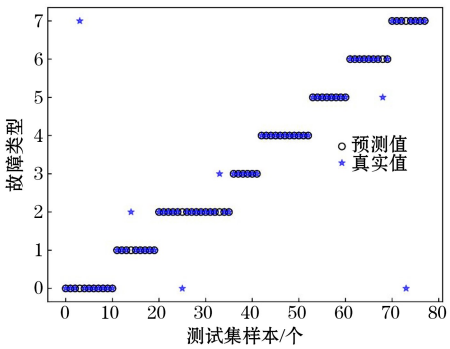


图 7　IWOA-XGBoost 故障诊断结果

Fig.7　Fault diagnosis results of IWOA-XGBoost

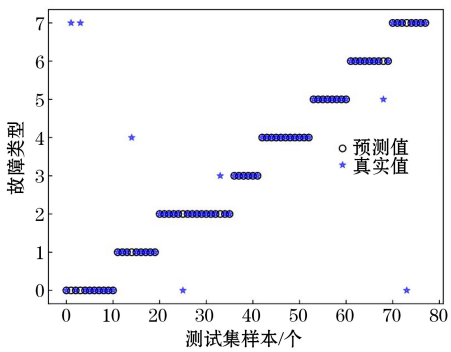


图 8　IWOA-CatBoost 故障诊断结果

Fig.8　Fault diagnosis results of IWOA-CatBoost

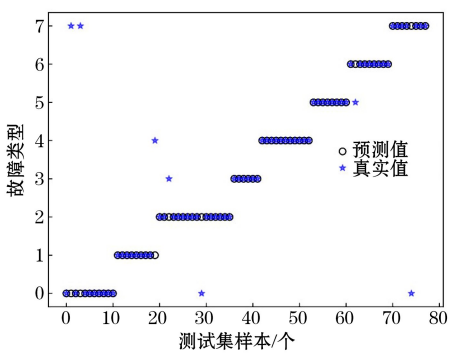


图 9　IWOA-RF 故障诊断结果

Fig.9　Fault diagnosis results of IWOA-RF

由图 6 可知，IWOA-LightGBM 对油门踏板故障和进气管堵塞的诊断准确率虽未达到 100%，但识别准确率均在 90%以上，此外对其他故障的识别准确率均能达到 100%。由图 7 可知，IWOA-XGBoost 对空滤堵塞的诊断准确率达到了 100%，并且对正常状态和油门踏板故障的诊断准确率有一定提升，但在进气管路堵塞、增压器故障和中冷器故障的识别上存在一些偏差。从图 8 和图 9 中可以看出，IWOA-CatBoost 和 IWOA-RF 针对所有故障的诊断准确率均未达到 100%，尤其对正常状态和油门踏板故障的诊断识别能力较弱。

3.4　调参框架性能对比

超参数调优是优化模型的重要步骤，目前关于机器学习的超参数调优方法大致分为 3 种[18]，具体包括暴力搜索法、贝叶斯优化、元启发式算法。当面对参数众多且取值范围较大的情况时，暴力搜索法效率低下，此时随机搜索为更优选择。贝叶斯优化中使用最广的是以树结构(TPE)为代理模型的优化模型；而元启发式算法以遗传、粒子群等种群算法对超参数进行寻优。为了验证本文改进鲸鱼算法的超参数寻优方式，选取随机搜索(RD)、基于 TPE 的贝叶斯优化的调参框架进行对比，此外还与在诊断领域常用的支持向量机(SVM)和 BP 神经网络进行了对比。表 5 为每个模型在测试样本中对各故障类型诊断的准确率，表 6 为 5 种模型的诊断性能的对比结果。

表 5　不同模型对各故障的识别准确率

Table 5　Recognition accuracy of different models for various faults　%

故障编号	IWOA-LightGBM	RD-LightGBM	TPE-LightGBM	SVM	BP
0	100	88	88	88	88
1	100	86	100	86	100
2	100	100	100	100	100
3	100	100	100	100	100
4	100	100	100	100	82
5	90	80	90	80	100
6	100	80	80	80	88
7	90	86	86	75	100
平均准确率	97.5	90	93	89	95

由表 5 可知，相比于其他 4 种模型，IWOA-LightGBM 具有更高的平均识别准确率，除了在中冷器故障和油门踏板故障表现不佳，其余故障类型都能达到 100%的准确率。特别是在风扇故障、正常状态中，其他模型的准确率均在 80%～90%，而 IWOA-LightGBM 均能达到 100%的准确率。

表 6　不同模型的诊断精度

Table 6　Diagnostic accuracy of different models　%

诊断模型	*A*	*P*	F_1分数	*R*	*AUC*
IWOA-LightGBM	98.08	98.61	98.53	98.61	99.57
RD-LightGBM	88.46	90.70	89.94	89.87	97.45
TPE-LightGBM	92.31	93.87	93.04	92.90	98.07
SVM	86.54	89.31	88.29	88.53	98.06
BP	92.31	93.87	93.83	94.60	98.50

由表 6 可知，IWOA-LightGBM 故障诊断模型在所选的 5 个评估指标性能上均优于其他模型，相

比于表现较好的 TPE-LightGBM 和 BP 模型,诊断预测性能明显提升。因此,本文提出的 IWOA-LightGBM 故障诊断模型在多分类、不均衡故障样本诊断方面获得了不错的效果,既能够增强矿用挖掘机的发动机诊断模型的容错能力,又能提升效率和降低成本。

4 结论

本文针对矿用挖掘机的发动机实际故障数据较少且类别不均衡,导致故障诊断精度不高的问题,通过对 LightGBM 损失函数和鲸鱼算法进行改进,提出了基于 IWOA-LightGBM 的挖掘机发动机智能故障诊断方法,得出的主要结论如下。

(1) 利用 RFECV 算法对发动机故障数据进行特征提取,特征维度的减少可以有效缩短训练时间,并提升诊断精度。

(2) 利用改进的损失函数(Focal-Loss)和鲸鱼算法对 LightGBM 超参数寻优,提高了 LightGBM 的诊断能力,从而降低了因类别不均衡对诊断精度的影响。

(3) 本文提出的矿用挖掘机发动机故障诊断模型 IWOA-LightGBM 的诊断准确率高达 98.08%,其故障诊断准确率和诊断精度均优于常用的故障诊断模型 RD-LightGBM、TPE-LightGBM、SVM、BP,本文的模型具有一定的理论意义和应用价值。

参考文献(References):

[1] 顾清华,江松,李学现,等.人工智能背景下采矿系统工程发展现状与展望[J].金属矿山,2022(5):10-25.
GU Qinghua, JIANG Song, LI Xuexian, et al. Development status and prospect of mining system engineering under the background of artificial intelligence[J]. Metal Mine, 2022(5):10-25.

[2] 曹现刚,段雍,赵江滨,等.综采设备健康状态评估研究综述[J].工矿自动化,2023,49(9):23-35+97.
CAO Xian'gang, DUAN Yong, ZHAO Jiangbin, et al. Summary of research on health status assessment of fully mechanized mining equipment [J]. Journal of Mine Automation,2023,49(9):23-35+97.

[3] 莫明慧,黄玉森.基于 ACA-BP 神经网络的挖掘机液压系统故障诊断研究[J].工程建设与设计,2020(13):137-140.
MO Minghui, HUANG Yusen. Research on fault diagnosis of excavator hydraulic system based on ACA-BP neural network[J]. Construction & Design for engineering, 2020(13):137-140.

[4] RAPOSO H, FARINHA J T, FONSECA I, et al. Predicting condition based on oil analysis: a case study[J]. Tribology International,2019,135:65-74.

[5] 连春平.大型矿用挖掘机智能化现状及发展趋势[J].露天采矿技术,2023,38(5):70-73.
LIAN Chunping. Current situation and development trend of intelligence large-scale mining excavators [J]. Opencast Mining Technology,2023,38(5):70-73.

[6] 李晴,顾清华,江松,等.基于知识图谱的近十年智能矿山研究热点及态势演化[J].采矿技术,2023,23(4):233-238.
LI Qing, GU Qinghua, JIANG Song, et al. Research hotspots and trend evolution of intelligent mines in the past ten years based on knowledge graph [J]. Mining Technology,2023,23(4):233-238.

[7] 顾清华,王培培,李学现,等.基于 CSSA-F-SVM 模型的矿用卡车发动机智能故障诊断[J].有色金属(矿山部分),2023,75(1):1-8.
GU Qinghua, WANG Peipei, LI Xuexian, et al. Intelligent fault diagnosis of mining truck engine based on CSSA-F-SVM model[J]. Nonferrous metals (Mining Section),2023,75(1):1-8.

[8] CAI B P, SUN X T, WANG J X, et al. Fault detection and diagnostic method of diesel engine by combining rule-based algorithm and BNs/BPNNs[J]. Journal of Manufacturing Systems,2020,57:148-157.

[9] 刘弘毅,王瑞,吴贯锋,等.基于核鲁棒流形非负矩阵分解和融合特征的柴油机故障诊断[J].计算机科学,2023,50(增刊1):884-891.
LIU Hongyi, WANG Rui, WU Guanfeng, et al. Dieselengine fault diagnosis based on kernel robust manifold nonnegative matrix factorization and fusion features [J]. Computer Science,2023,50(Suppl.1):884-891.

[10] 陈思勤,周浩豪,茅大钧.基于改进 GWO-LightGBM 的磨煤机故障预警方法研究[J].自动化仪表,2024,45(2):106-110+115.
CHEN Siqin, ZHOU Haohao, MAO Dajun. Research on fault early warning method based on improved GWO-LightGBM coal mills [J]. Process Automation Instrucmentation,2024,45(2):106-110+115.

[11] 顾清华,孙文静,李学现.基于 IFOA-RotGBM 的矿用挖掘机发动机故障诊断[J].金属矿山,2023(9):156-163.
GU Qinghua, SUN Wenjing, LI Xuexian. Fault diagnosis of mining excavator engine based on IFOA-RotGBM[J]. Metal Mine,2023(9):156-163.

[12] 顾清华,王晨曦,王倩,等.考虑卡车尾气排放因素的露天矿粉尘质量浓度预测[J].矿业研究与开发,2024,44(4):161-167.
GU Qinghua, WANG Chenxi, WANG Qian, et al. Prediction of dust mass concentration in open-pit mines considering truck exhaust emission factors [J]. Mining Research and Development,2024,44(4):161-167.

[13] KE Guolin, MNNG Qi, FINLEY Thoma, et al. LightGBM: a highly efficient gradient boosting decision tree [C]//Neural Information Processing Systems. State of Michigan: Curran and Associates,INC.,2017.

[14] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss or dense object detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence,2017,42(2):2999-3007.

[15] MIRJALILI S, LEWIS A. The whale optimization algorithm[J]. Advances in Engineering Software,2016,95:51-67.

[16] NADIMLIMI-SHAHRAKI M H, ZAMANI H, MIRJALILI S. Enhanced whale optimization algorithm for medical feature selection: a COVID-19 case study[J]. Computers in Biology and Medicine,2022,148:105858.

[17] OMOTEHINWA T O, OYEWOLA D O, DADA E G. A light gradient-boosting machine algorithm with tree-structured parzen estimator for breast cancer diagnosis[J]. Healthcare Analytics,2023,4:100218.

[18] 李海霞,宋丹蕾,孔佳宁,等.传统机器学习模型的超参数优化技术评估[J].计算机科学,2024,51(8):242-255.
LI Haixia, SONG Danlei, KONG Jianing, et al. Evaluation of hyperparameter optimization techniques for traditional machine learning models[J]. Computer Science, 2024, 51(8):242-255.