

基于 IWOA-stacking 集成学习模型的抽油机故障诊断方法

张涵钰¹ 王亚辉¹ 曾昭虎¹ 赵建慧¹

ZHANG Hanyu WANG Yahui ZENG ZhaoHu ZHAO Jianhui

摘要

针对抽油机故障诊断问题,提出一种基于振动特征和 IWOA-stacking 集成模型的抽油机故障诊断方法。首先,基于 stacking 算法构建了两层集成模型,模型第一层采用最小二乘支持向量机、随机森林、梯度提升模型和 XGBoost 模型作为基学习器,第二层采用多元线性回归作为元学习器;其次,基于改进鲸鱼优化算法对四个基学习器的超参数进行优化;最后,针对采集的振动数据优选用于故障诊断的时域特征指标,用超参数优化后的集成模型进行学习和诊断,并与 4 种基模型进行对比。实验结果表明,所提模型具有更高的识别效果且具有很好的泛化能力。

关键词

集成模型;鲸鱼优化算法;抽油机;故障诊断;振动分析

doi: 10.3969/j.issn.1672-9528.2024.02.044

0 引言

目前,通过安装电参、功图、压力等数字化采集设备,实现了抽油机生产状态的实时监测,而抽油机作为关键的生产单元,主要包括电机、减速箱、驴头、游梁、光杆等部件设备,这些设备的故障检测还需要依靠人工每天定期巡检来完成,检测手段还停留在人工到现场听、摸、看等经验判断的传统模式,很难满足超前感知快速预警的生产需求。国内外抽油机主要的诊断方法多数是基于示功图的诊断识别,主要有主动学习法^[1]、人工神经网络法^[2]、图形法^[3]、深度学习^[4-5]等方法。而振动信号是往复式机械设备工作状态的重要指标,包含了丰富的设备异常或故障信息。本文对抽油机关键部位的振动信号数据进行分析,提取特征向量作为故障样本建立合理、完善的标准故障诊断库,提出一种基于振动特征和 IWOA-stacking 集成模型进行抽油机故障诊断,取得了很好的应用效果。

1 stacking 集成学习模型原理

1.1 stacking 集成学习模型

集成学习^[6]通过结合多个学习器,采用某种结合策略将多个学习器的结果进行集成,构建一个强学习器以获得更好的学习性能,常用的方法有 bagging、boosting 和 stacking。bagging 与 boosting 这两种集成策略^[7]是采用的同一种基学习器,而 stacking^[8-9]通常都是组合多种不同的基学习器形成

的学习模型,通过对比发现 stacking 在很多领域的表现都优于投票法^[10],并且总是比最好的基算法效果更好。本文基于 stacking 算法构建两层集成模型如图 1 所示,模型第一层采用最小二乘支持向量机(LS-SVM)^[11]、随机森林(RF)^[12]、梯度提升树模型(GBDT)^[13]和 XGBoost 模型^[14]作为基学习器,第二层采用多元线性回归作为元学习器,用训练数据对第一层 4 个学习器进行学习,然后将多个学习器得到的预测结果拼接成特征矩阵,作为第二层多元线性回归学习器的训练集进行模型训练。选择不同模型作为 stacking 基学习器的主要目的在于通过各自算法的原理建立不同的预测模型,以从不同的空间和结构角度挖掘历史数据之间存在的内在联系。因此,为了在机器学习任务中取得更好的性能表现,选用差异度较大的算法集成起来可以获得鲁棒性更强的模型,并减小模型的过拟合风险。

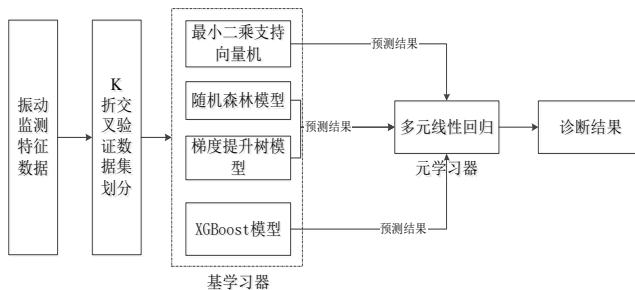


图 1 stacking 集成学习模型

1.2 stacking 集成学习模型训练算法

(1) 数据准备:使用 K 折交叉验证技术,将原始训练数据集分成 K 个子集,即将数据集划分成 K 份,然后依次将每份作为验证集,其余 K-1 份作为训练集。在本文中具体采

1. 大庆油田有限责任公司第五采油厂数字化运维中心
黑龙江大庆 163318
[基金项目]黑龙江省自然科学基金(LH2022F008)

用的是五折交叉验证法,即训练数据集被随机划分了五份,每次使用 4 份作为训练集,剩下一份作为验证集。

(2) 基学习器训练:将 4 份子集输入到第一层的各个基学习器中进行模型训练,以此得到多个训练好的基学习器。然后将剩余的一份数据集输入到训练好的基学习器中,得到多个基学习器的预测值,并将其组合,作为第二层元学习器的输入。

(3) 元学习器训练:将第一层中各个基学习器的输出作为新的训练集,输入到第二层的元学习器中,进行模型的训练。鉴于元学习器综合考虑了第一层各个基学习器中的输出,所以其预测性能较高。

通过以上 3 个步骤,stacking 集成学习模型能够在保持模型多样性的同时,充分利用不同模型的优势,从而提高整体预测的准确性和鲁棒性。

2 基于 IWOA 的集成学习模型超参数优化算法

本文提出的 stacking 模型在建模过程中使用了 4 个基学习器,不可避免地要对多个超参数进行优化调整,一种优秀的调参方法将直接决定各个基学习器的学习效果,进而影响 stacking 集成模型整体的识别性能。本研究提出一种改进鲸鱼优化算法(improve whale optimization algorithm, IWOA)对基学习器的超参数进行优化,以期在降低计算时间复杂度的同时,尽可能找到近乎全局最优的超参数组合,提升 stacking 集成学习模型的识别精度。

2.1 鲸鱼优化算法

鲸鱼优化算法^[15]模仿了座头鲸狩猎的策略,分为包围猎物、使用泡泡网进行捕食和随机捕食三个阶段。该算法的目标是通过模拟这些自然行为,以更智能、更高效的方式解决优化问题。

(1) 包围猎物

WOA 算法模拟该行为时,鲸鱼个体根据当前最优解的位置,调整自身位置朝向最优解的方向进行更新,以便更快地接近潜在的优秀解。用以下公式表示为:

$$D = |C \cdot X^*(t) - X(t)| \quad (1)$$

$$X(t+1) = X^*(t) - A \cdot D$$

式中: A 和 C 是系数向量; t 是当前迭代轮数; $X^*(t)$ 代表最优解的位置坐标, $X(t)$ 则是当前位置信息; D 表示两者之间的距离。向量 A 和 C 的公式为:

$$A = 2a \cdot \gamma - a \quad (2)$$

$$C = 2 \cdot \gamma$$

式中: a 在迭代过程中,从 2 线性递减到 0; γ 则是范围在 $[0,1]$ 的随机数。

(2) 泡泡网捕食

座头鲸围攻猎物时,同时缩小包围范围并沿着螺旋形路

径移动。可以使用数学模型来模拟这种泡泡网捕食行为中的缩小包围范围和螺旋更新位置的机制,公式为:

$$X(t+1) = \begin{cases} X^*(t) - A \cdot D & p < 0.5 \\ D^l \cdot e^{bl} \cdot \cos(2\pi l) + X^*(t) & p \geq 0.5 \end{cases} \quad (3)$$

式中: $D = |X^*(t) - X(t)|$, p 为 $[0,1]$ 随机概率, l 为 $[-1,1]$ 随机概率。

(3) 随机捕食

随机捕食与泡泡网捕食不同的地方在于,这个阶段不是根据当前最优解,而是基于随机选择的鲸鱼个体作为参考点来更新其他鲸鱼的位置。以下是随机捕食的数学模型。

$$D = |C \cdot X_{rand}(t) - X(t)| \quad (4)$$

$$X(t+1) = X_{rand}(t) - A \cdot D$$

式中: $X_{rand}(t)$ 是一个随机位置向量,在随机捕食中设定了系数变量 $|A| \geq 1$,用以提高算法的搜索和捕食能力,帮助寻找更适合的目标。

2.2 改进的鲸鱼优化算法

2.2.1 基于佳点集理论的种群初始化

WOA 算法开始进化时通过随机生成个体开始寻优的,如果个体均匀分布在解空间或是个别解在最优解的周围,则有利于提升寻优性能和算法效率。佳点集^[16]理论表明,在 s 维欧氏空间单位立方体上进行近似计算函数的积分时,用 n 个佳点构成的加权和,其误差比较小,优于采用任何其他 n 个点,具体初始化方法如下。

设 G_s 是 s 维欧氏空间中的单位立方体,如果 $r \in G_s$,形为 $p_n(k) = \{\{r_1^{(n)} \times k\}, \{r_2^{(n)} \times k\}, \dots, \{r_s^{(n)} \times k\}\}, 1 \leq k \leq n\}$,其偏差 $\varphi(n)$ 满足 $\varphi(n) = C(r, \varepsilon) n^{-1+\varepsilon}$,其中: $C(r, \varepsilon)$ 是只与 r 和 $\varepsilon(\varepsilon > 0)$ 有关的常数,则称 $p_n(k)$ 为佳点集, r 为佳点。一般情况下,取 $r = \{2\cos(2\pi k/p), 1 \leq k \leq s\}$, p 是满足 $(p-s)/2 \geq s$ 的最小素数。

2.2.2 基于六度空间理论的自适应分组方式

本文在算法迭代初期先将整个群体进行分组,在不断迭代进化过程中,随着逐渐寻优到最优解附近合并成一个群体加速寻优能力。六度空间理论表明,任何两个陌生人之间的联系不会超过六个人。基于此,本文提出了以下具体的分组和信息交流方式。

(1) 算法初始阶段,按整个种群分成 $6 \times m$ 个子群加强算法局部寻优能力,每个子群里至少 6 个个体,其中 m 是大于等于 1 的整数,故如果种群个数少于 36 个,则本文算法不执行该自适应分组方法,按整个种群进行文中算法的后续进化操作。

(2) 每次迭代之后,在每一个分组里利用锦标赛选择算法选出 1 个个体,通过计算选出个体之间的距离,从其他分组里选出与该分组选出个体距离最小的 5 个个体进行相关联用于信息交换。

(3) 计算每个分组里适应度的变异系数, 子群变异系数连续 n (文中取 3) 次低于所有分组变异系数平均值的种群, 则该子群的个体平均分配到其他子群, 直到合并成一个种群增加算法全局寻优性能。

2.2.3 基于分组关联个体的最优个体更新

社会中的个体通常表现出趋同性, 即他们的行为受到群体共识的影响, 将基于六度空间理论分组得到的 5 个个体, 用“群体位置质心”替代子群平均知识, 简化计算, 即取所有个体位置的平均值作为质心, 对分组内最优个体按式 (5) 进行更新计算适应度, 如果优于之前的位置则替换。

$$x_i(t) = \frac{\sum_{j=1}^{num_i=5} x_{ij}}{5} \quad (5)$$
$$x_{i,gbest}(t) = \omega \times x_{i,gbest}(t) + rand \times (x_i(t) - x_{i,gbest}(t))$$

2.2.4 基于反向学习的随机捕食个体更新

反向学习算法^[17]具有较快的学习速度和更强的优化能力。可以在随机捕食阶段先计算个体的反向解, 并按照公式 (6) 计算反向解的个体 $X'(t+1)$, 再与公式 (4) 计算的个体 $X(t+1)$ 进行比较, 并取最优解作为个体最终位置。

$$X'(t) = a + b - X(t)$$
$$D = |C \cdot X_{rand}(t) - X'(t)| \quad (6)$$
$$X'(t+1) = X_{rand}(t) - A \cdot D$$

式中: a 和 b 分别为个体寻优空间中的下界和上界。

3 实验仿真

3.1 超参数寻优范围及 IWO 运行参数设置

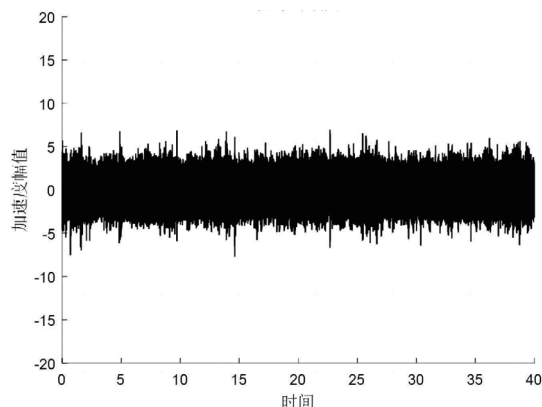
文中建立的 stacking 集成学习基模型采用 LS-SVM、RF、GBDT 和 XGBoost, 其中 LS-SVM 的参数包括正则化参数 γ 、核宽度 σ ; RF 的超参数包括学习器数量 $n_estimators$ 、随机选择特征数目 $max_features$ 、树的最大深度 max_depth ; GBDT 的超参数包括学习器数量 $n_estimators$ 、学习率 $learning_rate$ 、树的最大深度 max_depth ; XGBoost 的超参数包括学习器数量 $n_estimators$ 、学习速率 $learning_rate$ 、树的最大深度 max_depth 和最小叶子节点样本权重 min_child_weight 。采用 IWOA 算法对所提模型进行超参数优化, 设置 IWOA 算法的迭代寻优次数为 200, 初始化群体个体数目为 30, 超参数寻优取值范围如表 1 所示。

表 1 各基模型超参数寻优范围

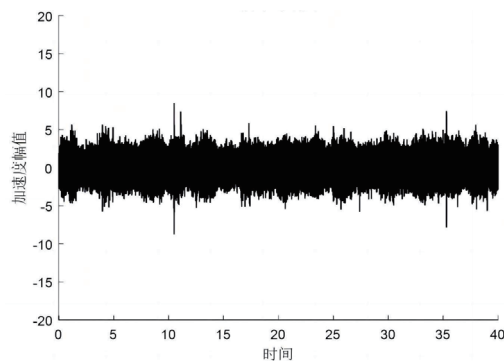
超参数	说明	取值范围
γ	正则化参数	[0,1000]
σ	RBF 核参数	[0,10]
$n_estimators$	学习器数量	[5,100]
$max_features$	随机选择特征数目	[2,10]
max_depth	树的最大深度	[1,10]
$learning_rate$	学习率	[0.005,0.1]
min_child_weight	最小叶子节点样本权重	[1,5]

3.2 振动数据特征参数选取

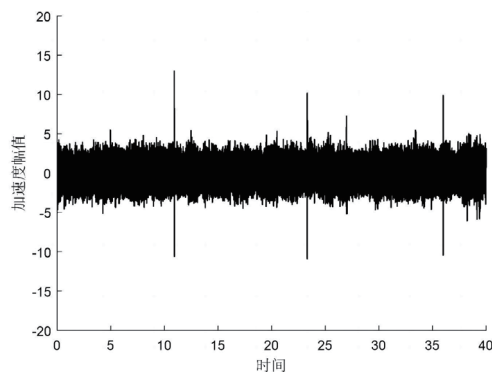
测试数据来自某区块 15 口井, 其中六型 10 口、十型 4 口、十四型 1 口, 分别采集了抽油机的减速器、基座、曲柄连杆、中轴、尾轴 5 个部位的振动数据。以加速度为信号输入源, 选取均值、标准差、方差、峰峰值、方根幅值、平均幅值、有效值、峰值、波形指标、峰值指标、脉冲指标、裕度指标、偏斜度及峭度等 14 个指标进行计算分析, 分析各个测点处抽油机振动测试的重要指标在不同工况下变化范围。以减速器为例确定减速器敏感测点位置—减速器垂直方向, 汇总减速器垂直方向测点归一前时域指标, 优选用于故障诊断的时域特征指标为峰峰值、峰值、峰值指标、脉冲指标、裕度指标、峭度。减速器螺栓松动时, 频域图对比如图 2, 减速器垂直方向时域指标对比如图 3 所示。



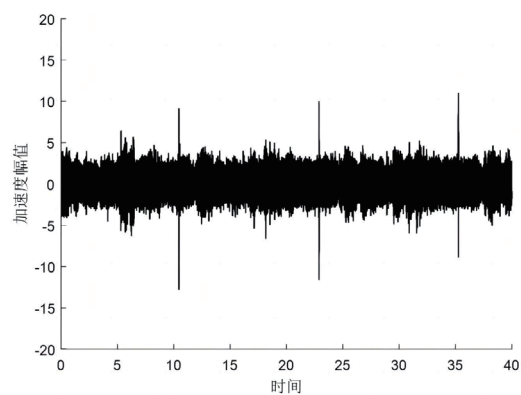
(a) 减速器正常工况



(b) 减速器螺栓左前一颗松动



(c) 减速器螺栓左前两颗松动



(d) 减速器螺栓左前两颗+右前一颗松动

图2 减速器螺栓松动时、频域图对比图

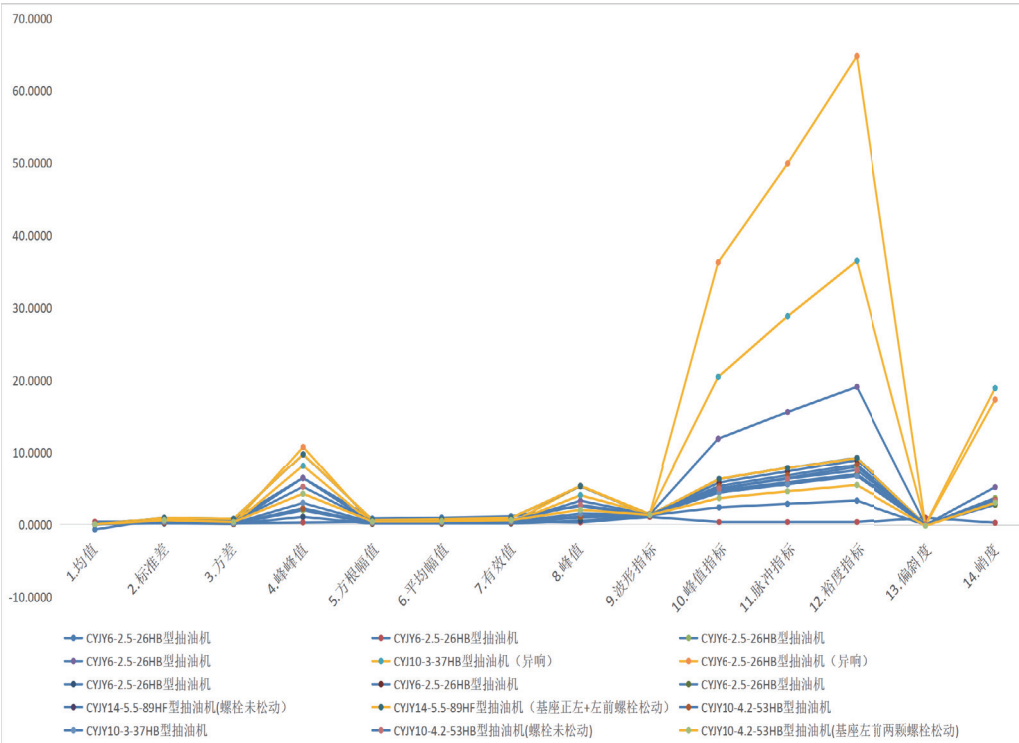


图3 减速器垂直方向时域指标对比

3.3 多模型对比分析

依据3.2选取的特征参数，分别采用LS-SVM、RF、GBDT、XGBoost、IWOA-stacking5种模型对故障数据集进行训练和预测，预测结果对比如表2所示。

从表2的对比可以看出IWOA-stacking要比LS-SVM、RF、GBDT、XGBoost的预测效果高。这是由于前四种模型的参数设置相对固定，而通过IWO算法优化后得到的超参数，可以使stacking集成学习模型能够充分利用不同模型的优势，从而提高整体预测的准确性和鲁棒性。同时在表2中可以看出尾轴正常及尾轴螺栓松动的预测效果不是很好，分析实测数据可知，由于螺栓松动得不是很大，螺栓未松动和松动的振动信号较相似造成识别效果不高，但所提模型相对于其他模型还是得到了较好的预测效果。

4 结语

振动信号蕴含了丰富的机械设备异常或故障工况的信息，本文利用振动信号的主要频谱特征向量作为故障样本建立故障诊断库，利用所提的IWOA-stacking集成模型对抽油机设备故障进行智能预警，既有利于保证抽油机设备的可靠性和安全性，又能减少因巡检产生的人员安全风险，提升了油田开发关键设备的数字化监测管理水平。

参考文献：

[1] 汪敏,周磊,闵帆,等.抽油机故障诊断的分布驱动主动学习算法[J].南京航空航天大学学报,2022,54(3):517-527.

[2] 董巧玲.基于改进神经网络的抽油机故障智能诊断研究[J].西安石油大学学报(自然科学版),2022,37(6):124-132.

[3] 李翔宇,袁春华,高宪文.抽油井示功图诊断分析方法研究[J].沈阳理工大学学报,2021,40(1):88-94.

表2 各模型预测结果对比

故障诊断类型	特征值选取	LS-SVM	RF	GBDT	XGBoost	IWOA-stacking
抽油机基座正常及地脚螺栓松动	有效值、峰值指标、脉冲指标、裕度指标、峭度	88.00%	88.34%	89.19%	88.00%	96.70%
减速器正常及减速器基座螺栓松动	均值、有效值、峰值、偏斜度、峭度	80.97%	79.85%	81.88%	82.85%	95.73%
减速器正常及减速器异常响动	有效值、峰值指标、脉冲指标、裕度指标、峭度	88.43%	89.85%	88.13%	89.67%	94.97%
曲柄销正常及曲柄销螺栓松动	有效值、峰值指标、脉冲指标、裕度指标、峭度	80.43%	81.55%	80.64%	82.56%	92.98%
尾轴正常及尾轴螺栓松动	均值、方差、有效值、偏斜度、峭度	78.12%	79.14%	78.90%	79.41%	90.58%
中轴正常及中轴螺栓松动	有效值、峰值、裕度指标、偏斜度、峭度	88.30%	89.87%	88.55%	89.69%	94.33%

5G+ 警用无人机在智慧安防领域的应用研究

胡剑飞¹ 丁 宁¹

HU Jianfei DING Ning

摘 要

5G 作为新一代的移动通信技术,不但具备更高速率、更低时延、更大连接及更高可靠性特性,还推动了人工智能和移动边缘计算技术的发展。凭借 5G 网络的大带宽、低时延、广连接等特点,结合人工智能、边缘计算等技术,5G 警用无人机实现了与指挥调度系统的无缝集成,以及与前端各类安全防控场景的精准对接。5G 技术与警用无人机技术的结合,有力推动了警用无人机在智慧安防领域的创新应用,实现智能监控、智能预警、自动化应对和大数据分析,以构建起立体化智慧安防体系,更好地实现预警和快速响应,为警务工作的开展提供必要的技术手段。

关键词

5G 网络; AI; MEC; 警用无人机; 智慧安防

doi: 10.3969/j.issn.1672-9528.2024.02.045

0 引言

当前我国正处于社会转型的关键阶段,各社会阶层之间的利益诉求日益多元化。城市规模不断扩大,城市结构日益复杂,犯罪手段日趋专业化,犯罪方式日趋隐蔽。同时,随着经济和社会的发展,人们对公共安全、治安防控的要求也越来越高。这些情况都给公安机关的工作开展带来了

新的考验,向科技要警力已经成为安防领域的大势所趋。^[1]在此背景下,构建新型智慧安防体系成为在信息化条件下,公安机关维护国家安全,维护社会秩序,保护公民的人身财产安全,保护公共财产,预防、制止和惩治违法犯罪活动的重要保障。

5G 技术的引入,推动了警用无人机在智慧安防领域的创新应用。5G+ 警用无人机以其灵活机动、快速隐蔽、安全高效以及便于融入警务工作体系等特点^[2],有效地弥补了传

1. 中通服咨询设计研究院有限公司 江苏南京 210019

[4] 魏航信,张青.基于深度学习的示功图多混合故障诊断[J].机电工程技术,2022,51(3):112-116.

[5] 商敬量.基于深度学习的游梁式抽油机故障诊断研究[D].大庆:东北石油大学,2023.

[6] 周志华.机器学习[M].北京:清华大学出版社,2016.

[7] 刘倩倩.基于 stacking 模型的房屋租赁价格回归算法研究[D].阜阳:阜阳师范大学,2023.

[8] 王辉,李昌刚.stacking 集成学习方法在销售预测中的应用[J].计算机应用与软件,2020,37(8):85-90.

[9] 王飞,黄涛,杨晔.基于 stacking 多模型融合的 IGBT 器件寿命的机器学习预测算法研究[J].计算机科学,2022,49(S1):784-789.

[10] SIGLETOS G, PALIOURAS G, SPYROPOULOS C D, et al. Combining information extraction systems using voting and stacked generalization[J].Journal of machine learning research, 2005, 6(3): 1751-1782.

[11] 张仲华,赵福媛,郭钧枫,等.柯西自适应回溯搜索与最小二乘支持向量机的集成预测模型[J].计算机应用,2022,42(6):1829-1836.

[12] 姚鹏飞,涂亚楠,王瑞红.基于随机森林算法拖拉机齿轮箱故障诊断研究[J].农机化研究,2024,46(3):246-251.

[13] 周晓敏,郝勇凯,丛文韬,等.基于梯度提升决策树模型的冷连轧机颤振研究[J].振动与冲击,2021,40(13):154-158.

[14] 闫星宇,顾汉明,肖逸飞,等.XGBoost 算法在致密砂岩气储层测井解释中的应用[J].石油地球物理勘探,2019,54(2):447-455+241.

[15] 许德刚,王再庆,郭奕欣,等.鲸鱼优化算法研究综述[J].计算机应用研究,2023,40(2):328-336.

[16] 郝芄斐,池瑞,屈志坚,等.求解铁路物流配送中心选址问题的改进灰狼优化算法[J].计算机应用,2021,41(10):2905-2911.

[17] 夏学文,刘经南,高柯夫,等.具备反向学习和局部学习能力的粒子群算法[J].计算机学报,2015,38(7):1397-1407.

【作者简介】

张涵钰(1991—),女,黑龙江兰西人,中级工程师,研究方向:信息工程。

(收稿日期:2023-11-27)