

DOI: 10.19650/j.cnki.cjsi.J2514617

基于 QA-SSA-EARF 的调节阀多工况故障诊断方法*

罗柯达, 张登峰, 周 通, 王村松, 张泉灵

(南京工业大学智能制造研究院 南京 210009)

摘 要:调节阀在实际工作过程中需要运行于多种控制方式,使相同故障在不同控制工况下常呈现出差异的特征信息,导致基于单一工况数据训练的机器学习诊断模型难以泛化、性能下降。为此,提出了一种基于量子注意力麻雀搜索算法(QA-SSA)与弹性自适应随机森林(EARF)模型相结合的调节阀多工况故障分类诊断方法。所提 EARF 模型在自适应随机森林(ARF)模型基础上,通过引入两级决策机制、优化全局漂移检测器位置、设计局部剪枝策略,并动态调节决策树数量,减少 ARF 模型的计算量,提高建模效率与诊断精度,增强对工况变化的自适应能力。针对 EARF 模型超参数耦合难以优化的难题,设计了一种 QA-SSA 优化算法,通过在传统麻雀搜索算法(SSA)中引入量子行为与玻尔兹曼选择策略,提高了算法在高维超参数空间的搜索效率与鲁棒性。最后,利用实验室电动调节阀流体控制系统平台,分别在流量、压力、液位等 3 种控制工况下针对调节阀的 6 类故障进行了模拟试验验证。结果表明,所提出的 QA-SSA-EARF 模型方法对单一工况下的分类诊断准确率达到 97.47%,比优化后的随机森林(RF)模型和 ARF 模型分别提高了 9.65% 和 3.64%;多工况下的平均分类诊断准确率达到 93.12%,比其他两种模型方法分别提高了 2.59% 和 8.9%,充分证明了该方法在多工况故障诊断任务中的有效性与鲁棒性。

关键词:调节阀;故障诊断;麻雀搜索算法;自适应随机森林;超参数优化

中图分类号: TH86 TP206 文献标识码: A 国家标准学科分类代码: 460.40

An improved QA-SSA-EARF method for multi-operating condition fault diagnosis of control valves

Luo Keda, Zhang Dengfeng, Zhou Tong, Wang Cunsong, Zhang Quanling

(Institute of Intelligent Manufacturing, Nanjing Tech University, Nanjing 210009, China)

Abstract: In practical applications, control valves are often operated under multiple control modes, causing the same fault in valves to exhibit distinct characteristic information under different control conditions. This makes the machine learning based diagnostic models trained by a single-condition dataset difficult to generalize and leads to performance degradation. To address this problem, a multi-condition fault classification method is proposed in this article by combining the quantum attention-based sparrow search algorithm (QA-SSA) with an elastic adaptive random forest (EARF) model. The proposed EARF model is formulated on the adaptive random forest (ARF). By introducing a two-level decision mechanism, optimizing the placement of the global drift detector, implementing a local pruning strategy, and dynamically adjusting the number of trees, EARF reduces the computational cost of ARF, improves the modeling efficiency and diagnostic accuracy, and enhances the adaptability to variant operating conditions. To optimize the coupled hyperparameters of EARF model, QA-SSA optimization algorithm is designed by introducing quantum behavior and a Boltzmann selection strategy into the traditional SSA. The algorithm improves the searching efficiency and robustness in high-dimensional hyperparameter spaces. Finally, simulative experiments are implemented on the electric control valve fluid control system in our laboratory to evaluate the proposed method classifying six types of faults under flow, pressure, and level control conditions. The results show that the proposed QA-SSA-EARF method achieves a classification accuracy 97.47% under a single condition case, which is 9.65% and 3.64% higher than the optimized random forest and ARF models, respectively. In a multi-condition case, the average classification accuracy is 93.12% by the proposed method, which is 2.59% and 8.9% more than the other two methods. Therefore, the effectiveness and

收稿日期: 2025-10-23 Received Date: 2025-10-23

* 基金项目: 国家自然科学基金(62333010)、国家自然科学基金(62573223)项目资助

robustness of the proposed approach are verified for fault diagnosis tasks in multi-operating conditions.

Keywords: control valve; fault diagnosis; sparrow search algorithm; adaptive random forest; hyperparameter optimization

0 引言

调节阀作为过程工业中常用的末端执行元件,凭借结构简单、运行可靠与较高控制精度,广泛应用于石油、冶金、电力与市政工程等领域^[1-2]。在流体控制系统中,调节阀通过阀位驱动来实现对流量、压力、液位等目标量的调节,但在复杂工艺和长期运行环境下,不同的控制工况(如流量、压力或液位等控制模式),会改变阀门与控制系统的闭环动态特性和各监测信号数据的统计分布,造成相同的阀门物理故障(如泄漏、堵塞或磨损)在不同工况下往往呈现出截然不同的时域与频域特征,这使得以单一工况数据批量训练得到的诊断模型在其他工况下性能显著下降并产生误报或漏报,从而限制了诊断模型的工程可用性与可靠性^[3]。因此,提升调节阀在多工况下的故障诊断性能,对于保障装置安全、降低停机与维护成本具有迫切工程意义。

鉴于调节阀运行数据的多工况性与工况间的差异性表现^[4],理想的诊断方法不仅需要丰富的多工况训练样本,还需要能够随运行环境改变快速适应的新型学习机制。近年来在多工况故障诊断问题研究中,基于机器学习与深度学习的方法成为备受关注的领域^[5]。例如,乔卉卉等^[6]针对轴承故障诊断方法中采用单源信号难以提供完备故障信息的不足,提出一种基于注意力机制的多源信息融合网络模型;姜午恺等^[7]针对高维不平衡数据条件下的多工况故障诊断难题,提出一种基于动态自适应迁移的宽度学习方法,应用于污水处理过程的多工况故障诊断。此外,迁移学习已成为多工况下处理数据不平衡问题的有效手段。它利用源工况数据预训练模型,并将知识迁移至目标工况,能显著提升模型在目标工况下的诊断性能。例如,Li等^[8]提出了一种基于特征级和类级的多源域自适应迁移学习模型,用于不同工况下的过程故障诊断,并在多个工业案例上得到了验证。钱思宇等^[9]提出了一种基于卷积神经网络的领域适配模型,将该模型用于多种工况迁移场景的训练,解决了滚动轴承在复杂多工况下实现有效故障诊断的难题。这些工作表明,迁移学习在目标工况样本稀缺或样本分布差异时能提供有益先验信息,但是,上述方法多以一次性或静态的跨域迁移为主,缺乏对随时间演化的连续适应与抗遗忘能力,难以长期应对设备老化与工况变化所引起的判别特征转移^[10]。因此,越来越多的研究开始关注增量学习方法,以精准掌握其运行状态并提升运行可靠性。

增量学习方法能够在不依赖大量历史数据的情况

下,通过持续学习机制实现模型在线更新与自适应调整,既能更好地保持历史知识,又能学习新样本的特征,从而改善由工况差异引起的泛化问题。例如,Liu等^[11]提出了一种基于在线学习算法和增量更新的诊断系统,利用自适应随机森林(adaptive random forest, ARF)模型学习新数据的变化特征,预测印刷电路板是否发生故障;Zhou等^[12]提出了一种用于异常声音检测(anomaly sound detection, ASD)的师生增量学习方法,解决ASD模型的可扩展性问题;江灏等^[13]提出一种针对堆外中子探测器的时空动态检测模型(spatial-temporal dynamic detection model, STDDM),通过提取时序信号隐含关系,使模型能够自适应于反应堆运行工况的变化。这些研究均旨在保持历史知识的同时逐步吸纳新样本特征,以提升对多工况的鲁棒性。

在增量学习的诸多方法中,ARF 凭借其有效的漂移检测^[14-15]与增量学习机制,在工业故障诊断领域展现出很强的应用潜力。ARF 支持以逐样本或逐批次的方式进行模型更新,并集成了漂移检测与动态模型替换机制,能够适应工业环境的数据分布变化,同时保持历史知识的有效利用与复用。因此,在调节阀的多工况故障诊断研究中,可利用 ARF 模型的漂移检测机制识别不同工况下故障数据分布的差异变化,进而自适应调整模型结构,提升诊断性能,降低模型的维护需求。

然而,常规的 ARF 模型用于调节阀多工况故障诊断时,仍面临一些困难。例如,ARF 的增量学习与漂移检测机制需要消耗较高的计算与存储资源;漂移检测器的灵敏度受其部署粒度与位置影响,对某些工况之间细微特征差异的识别能力不足,难以有效降低数据类别不平衡带来的诊断误差;ARF 模型包含许多耦合超参数,其调优过程复杂,直接影响模型的性能与稳定性。

为解决上述问题,使诊断模型更好地适配调节阀多工况运行特性,本文提出一种面向多工况的量子注意力麻雀搜索算法-弹性自适应随机森林(quantum attention-based sparrow search algorithm-elastic adaptive random forest, QA-SSA-EARF)模型的故障诊断方法,主要贡献为:1)提出一种改进的弹性自适应随机森林(elastic adaptive random forest, EARF)模型。通过引入两级分类结构降低计算开销,并将漂移检测机制细化至决策树的叶节点级别以提升检测灵敏度;在模型训练过程中,通过分析漂移检测队列中各故障类别的召回率与误差指标,决定是否执行局部剪枝与节点重分裂;设计弹性森林规模调整机制,增强模型对类别分布不平衡的动态适应能力;2)针对 EARF 模型的多个超参数耦合优化问题,设计

一种基于量子行为扰动与玻尔兹曼选择策略改进的麻雀搜索算法 (quantum attention-based sparrow search algorithm, QA-SSA), 实现对多个耦合超参数的优化, 以提升诊断模型的鲁棒性与稳定性。由此, 本文提出的方法将 ARF 模型的漂移检测及增量学习能力与面向工况差异的模型结构改进有机结合, 为调节阀在多工况运行环境下的故障诊断提供更可靠的解决方案。最后, 利用实验室的智能电动调节阀流体控制实验平台开展多工况实验, 采集实际监控数据进行对比消融实验, 验证了所提方法在多工况环境下的适应性、学习能力及诊断性能优势。

1 EARF 故障诊断模型

1.1 EARF 改进模型

ARF 是以增量 Hoeffding 树^[16]为基学习器, 由若干棵决策树组成的集成模型。它通过在线 bagging 将样本子集随机分配到各棵树以维持集成多样性。在构建与分裂节点时, 算法采用随机特征子集策略引入随机子空间, 增强模型泛化能力。为应对概念漂移, 每棵树均设有独立漂移检测器 (如 ADWIN 检测器^[17], adaptive windowing), 持续监控其在线性能, 一旦检测到性能显著退化, 立即重新训练子树并进行替换, 实现局部自适应调整。最后通过集成所有决策树的预测结果以投票方式决定最终诊断结果。

然而, 调节阀在多工况条件下长期运行时, 监控数据绝大多数为正常数据, 故障数据呈现明显的类别不平衡。若直接应用 ARF 构建的诊断模型, 会因频繁对正常数据样本执行漂移检测造成计算资源浪费。同时, 基于树层级设置的漂移检测机制难以敏锐捕捉多工况间细微的故障特征差异, 易导致漂移响应滞后, 进而引发漏报。此外, 模型预设的固定森林规模不利于提高模型学习能力, 部分性能较弱的基学习器模型还会积累预测误差, 降低整体模型性能。因此, 基于上述缺点, 下文围绕诊断决策机制、漂移检测灵敏度调节以及森林规模动态调整 3 个方面来改进传统的 ARF 模型结构, 具体改进为:

1) 在 ARF 中引入两级决策机制来提高诊断效率。当数据样本输入 ARF 模型时, 森林首先进行一级决策的二分类 (正常/异常) 筛选, 排除正常样本, 只把计算资源集中于异常样本的处理。对于被判定为异常的数据样本, 其特征子集被随机分配到各决策树上, 森林再进行二级决策的多分类预测以识别故障类别, 同时, 各决策树基于分配的子集进行漂移检测并在必要时触发在线更新。上述两级决策操作均由各决策树独立给出分类结果。最后, 通过集成投票汇总, 作为整个 ARF 模型的诊断结果输出。这种两级机制既减轻了在线训练与漂移检测的计算负担, 又使漂移检测更聚焦于与故障相关的特征, 有助

于提升故障识别的准确性与响应速度。

2) 将传统 ARF 对整棵树的漂移检测, 细化到每颗树内部的分裂节点上, 既可保留树的有效结构又提高故障识别的敏感度。部署于分裂节点上的 ADWIN 检测器通过一个长度为 W 的循环队列, 缓存最新 W 个样本在该节点上的真实标签与预测标签对, 通过计算不同类别故障在该队列的召回率 Rec_c 与误差指标 e 来反映故障的漂移程度, 如式 (1) 所示。

$$\left\{ \begin{aligned} Rec_c &= \frac{\sum_{i=1}^W 1\{y_i = c \wedge \hat{y}_i = c\}}{\sum_{i=1}^W 1\{y_i = c\}} \\ e &= 1 - \frac{\sum_{c=1}^k (w_c \times Rec_c)}{\sum_{c=1}^k w_c} \end{aligned} \right. \quad (1)$$

式中: y_i 为样本真实故障标签; \hat{y}_i 为该样本在该节点下的故障预测标签; $1\{\cdot\}$ 为指示函数 (条件成立则为 1, 否则为 0); c 为故障类别; k 为故障类别总数; $w_c > 0$ 为设置的各故障类别的权重, 其配置原则是基于训练集中各故障类别样本出现的频率, 采用区间划分的方式确定。具体步骤为: (1) 统计训练集中每一个故障类别样本发生的频率 p_c ; (2) 将每种故障类别映射至 4 个预设的权重区间之一, 即 $(0, 0.25]$ 、 $(0.25, 0.5]$ 、 $(0.5, 0.75]$ 以及 $(0.75, 1]$, 同时确保所有权重之和为 1。映射过程遵循“频率越高, 权重区间越接近 0; 频率越低, 权重区间越接近 1”的准则, 这样在漂移度量中可赋予样本少的故障类别更高的权重上限, 增强对少样本故障类别的检测敏感度; (3) 各故障类别的具体权重数值可通过后文的寻优算法进一步确定。误差 $e \in [0, 1]$, 其值越大表明该节点分类性能越差。

3) 决策树模型利用 ADWIN 的内置算法将不同样本计算出的节点级分类误差序列 $\{e_t\}$ 自适应划分为两段, 计算它们的均值偏差 $|\mu_1 - \mu_2|$ 与 Hoeffding 界限^[18] (阈值 ε), 如式 (2) 所示。

$$\varepsilon = \sqrt{\frac{n_1 \times n_2}{2(n_1 + n_2)} \ln\left(\frac{4(n_1 + n_2)}{\delta}\right)} \quad (2)$$

式中: $\delta \in (10^{-4}, 10^{-3})$ 为漂移显著性水平; n_1 与 n_2 分别为当前序列被划分前后两段的样本个数。当均值偏差超过阈值 ε 时, 即判定该节点原有的分裂依据已失效, 触发局部剪枝和重分裂操作, 实现对节点分布漂移的精细化响应。通过将漂移检测器下沉至叶分裂节点的策略, 能更早地捕捉到调节阀在不同工况下的局部特征漂移, 提高稀有故障的保留率和诊断识别能力, 避免模型性能退化。

4) 引入弹性森林规模的动态调整机制。如式 (3) 所

示,该机制以ADWIN所计算的误差序列中各故障类别的总体误差作为触发依据,当其超过预设的模型误差阈值时,自动触发决策树数量的增减调整。

$$T_{\text{new}} = \begin{cases} T_0 + \Delta, & E_t > E_{\text{high}} \\ T_0 - \Delta, & E_t < E_{\text{low}} \end{cases} \quad (3)$$

式中: T_{new} 为森林中决策树的最新数量; T_0 为ARF模型决策树的原有数量; Δ 为树的增减数目,通常取为初始决

策树数量的1/10; E_t 为各故障类别的总体误差; E_{high} 与 E_{low} 分别为触发增树与减树的超参数阈值。通过引入该机制,在短期内能提高模型容量以快速学习不同工况下故障模式的差异;对长期运行而言,能剔除过时的树以节省资源并减少累积误差。

基于上述的模型改进技术,得到如图1所示的故障诊断EARF模型结构。

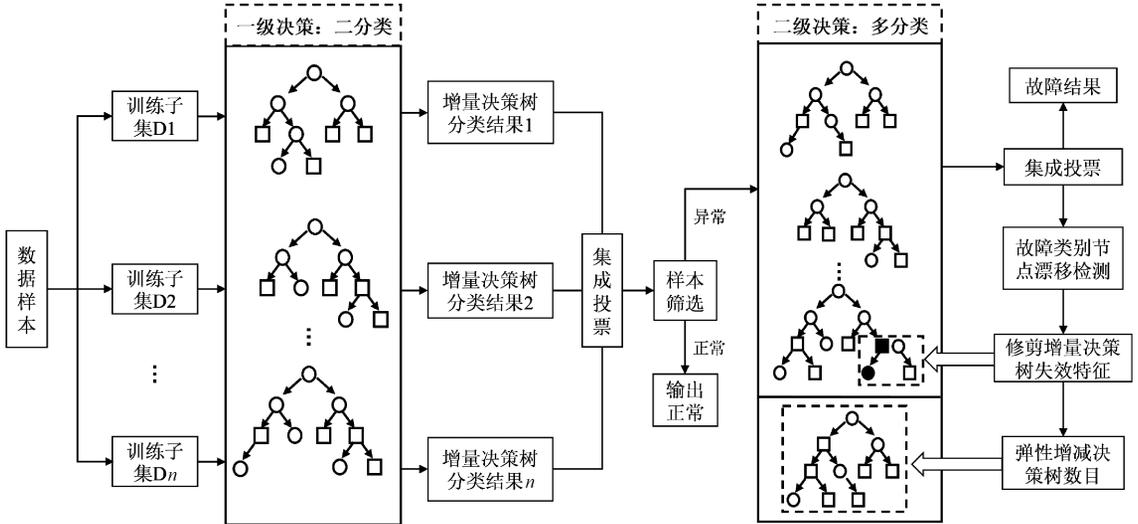


图1 EARF模型结构

Fig. 1 EARF model architecture

1.2 EARF超参数

改进后的EARF模型,涉及许多关键参数,如表1所示。这些参数分别来自于基础随机森林的超参数、专为适应数据流样本设计的超参数以及漂移检测器的超参数,它们之间存在显著的耦合效应,共同决定了诊断模型的故障分类识别性能。因此,在建模时需要确定其最优值。近年来,故障诊断领域广泛采用群体智能算法进行超参数优化^[19-20]。这类智能算法具有较高的全局搜索能力与快速的参数调优效率,在提高模型诊断精度和稳定性方面表现出良好的效果。因此,为了提升EARF模型的诊断性能,下文基于麻雀搜索算法(sparrow search algorithm, SSA),提出一种改进的QA-SSA算法,实现对EARF模型超参数的优化计算。

2 QA-SSA算法

SSA算法是薛建凯于2020年提出的一种模拟麻雀觅食行为的群体智能优化方法^[21]。该算法将候选解视为一个种群,通过“发现者—跟随者—警戒者”3类角色的协同更新策略迭代优化目标函数,在全局搜索与局部开发之间形成良好平衡,实现较高的收敛精度。然而,

表1 EARF模型的调优超参数

Table 1 Tunable hyperparameters of the EARF model

超参数	含义
T_0	初始决策树数量,影响模型容量与并行度
w_c	故障漂移检测中,对不同故障类型的召回赋予权重
δ	节点处漂移显著性水平,控制对误差变化的敏感度
W	检测器内滑窗长度,影响漂移检测与森林规模决策
$E_{\text{high}}/E_{\text{low}}$	分别为触发森林规模的“增”与“减”的界限阈值
H	决策树的最小分裂样本数,影响分裂频率与树深度

直接利用SSA对调节阈多工况故障诊断模型进行优化难以得到最优超参数组合,其原因有两点:1)发现者的指数衰减更新策略在复杂、高维的超参数空间中易导致种群迅速丧失多样性,从而陷入局部最优,难以提高某些故障类别的识别精度;2)SSA的搜索行为对当前个体位置和迭代轮次较为敏感,缺乏有效的概率导向机制以维持长期探索能力,因而在高噪声、分布多变的工况数据下,其收敛精度与鲁棒性均不高^[22]。因此,QA-SSA算法在保持SSA中跟随者与警戒者原有更新框架的基础上,在发现者中引入量子行为扰动与玻尔兹曼选择策略以增强群体多样性、改善收敛特性并提升在高维耦合超参数空间

中的寻优效果,进而可使用该算法找到EARF模型的最佳超参数组合以提高诊断性能。

2.1 SSA算法的改进

令优化问题的维度为 D ,算法中麻雀种群数量为 N ,每只麻雀个体位置用 $\mathbf{X}_i^t = (x_{i,1}^t, x_{i,2}^t, \dots, x_{i,D}^t)$ 表示,目标函数中第 i 个体适应度值为 f_i 。麻雀种群的发现者负责全局搜索食源任务,其位置更新如式(4)所示。

$$x_{i,d}^{t+1} = \begin{cases} x_{i,d}^t \times \exp\left(\frac{-i}{\kappa \times t_{\max}}\right), & R_2 < ST \\ x_{i,d}^t + Q \times L, & R_2 \geq ST \end{cases} \quad (4)$$

式中: t 为当前迭代次数; $x_{i,d}^t$ 为第 t 次迭代时第 i 只麻雀在第 d 维的位置; t_{\max} 为最大迭代次数; Q 为服从标准正态分布的随机数; κ 为 $(0,1]$ 的随机数; L 为 $1 \times d$ 维的全1矩阵; $R_2 \in [0,1]$ 为预警值; $ST \in [0,0.5]$ 为安全阈值。

虽然式(4)中的发现者位置更新策略能在“保留优势”与“追随最优”之间保持平衡,但它将所有参数分量与种群个体视为同等重要,无法将搜索资源集中在对性能影响更大的关键超参数子集上。此外,算法搜索步长随着迭代次数增加会急剧减小,使搜索效率急速下降,易陷入局部最优。为克服这些局限,借鉴文献[23]的思想,将发现者的位置更新策略改为当前位置 $x_{i,d}^t$ 与全局最优解 x_{best}^t 的线性凸组合,如式(5)所示。

$$x_{i,d}^{t+1} = (1 - \alpha) \times x_{i,d}^t + \alpha \times x_{\text{best}}^t \quad (5)$$

式中: $\alpha \in [0,1]$ 为混合权重(靠拢强度),用来在“保留当前解”与“向全局最优靠拢”之间做权衡,其优点在于保证了搜索的稳定性且易于分析。然而,权重 α 无法区分不同参数维度的重要性,可能会导致群体过早陷入局部最优。因此,进一步在每个维度上引入量子扰动因子^[24],使不同超参数根据其重要性获得差异化的搜索强度,且靠拢强度随迭代过程自动调整。令 $U(0,1)$ 为区间 $(0,1)$ 上的独立且分布均匀的随机变量, $\psi'_d(t)$ 为定义在第 d 维位置时第 t 次迭代的量子扰动因子,如式(6)所示。

$$\psi'_d(t) = (U(0,1))^{\frac{1}{T_q(t)}} \quad (6)$$

式中: T_q 为随迭代变化的调度参数,用于控制 ψ'_d 来决定每一维的缩放/靠拢强度, $T_q \in [0.1,5]$ 。当 T_q 接近于1时,控制算法在早期做大范围探索;随着 T_q 增大,指数 $1/T_q$ 变小,从而逐步增强算法向最优解的靠拢速度以利于后期精细搜索。用 ψ'_d 替换式(5)中的 α ,发现者的位置更新如式(7)所示。

$$x_{i,d}^{t+1} = (1 - \psi'_d) \times x_{i,d}^t + \psi'_d \times x_{\text{best}}^t \quad (7)$$

为进一步增强算法中发现者对全局适应度分布的感知能力并实现更平滑的探索-利用权衡,继而引入玻尔兹曼选择策略^[25],该策略中能量粒子权重计算式如式(8)所示。

$$p_i = \frac{\exp(-E_i/(T \times k_b))}{\sum_{j=1}^N \exp(-E_j/(T \times k_b))} \quad (8)$$

式中: E_i 为第 i 个粒子(在SSA算法中为第 i 只麻雀)的能量; k_b 为波尔曼兹常数; T 为热力学温度。在SSA算法中,将适应度值 f_i 视为能量,并借助该机制对候选解进行平滑的概率分配,在保持收敛速度的同时兼顾种群多样性。具体地,用当前麻雀个体的适应度值 f_i 替换式(8)中的 E_i ,取 $k_b = 1$,得到麻雀个体权重计算式(9)。

$$p_i = \frac{\exp(-\beta \times f_i)}{\sum_{j=1}^N \exp(-\beta \times f_j)} \quad (9)$$

式中: $\beta = 1/T$ 。当 β 较小时,各个解的随机性强,有利于探索;当 β 增大时,会放大优秀个体适应度解 f_i 对应的权重,有利于聚拢。由此在SSA算法中,实现了通过 f_i 的大小赋予个体不同的适应度权重。最后,将式(9)引入式(7)中右侧的第2项,与 ψ'_d 相结合,实现对发现者在第 d 维的混合更新,其改进后QA-SSA算法发现者的位置更新如式(10)所示。

$$x_{i,d}^{t+1} = (1 - \psi'_d) \times x_{i,d}^t + \psi'_d \times p_i \times x_{\text{best}}^t \quad (10)$$

式中: $(1 - \psi'_d) \times x_{i,d}^t$ 为残留项,用于保留个体在原维度上的信息以维持多样性;右侧第2项中的 ψ'_d 控制逐维靠拢强度, p_i 则依据个体适应度对全局最优解施加权重调整,从而实现维度聚焦与个体优选。

2.2 QA-SSA算法性能测试

为了验证QA-SSA算法的性能,分别使用蜣螂优化算法(dung beetle optimization, DBO)、粒子群优化算法(particle swarm optimization, PSO)、灰狼优化算法(grey wolf optimizer, GWO)、SSA与QA-SSA算法对表2的6种标准测试函数进行了对比实验。在实验中,用每次迭代得到的优化解代入测试函数,而测试函数计算的函数值视为该次迭代的适应度值。为对比各算法优化性能,使用适应度的平均值反映算法优化精度,用标准差来反映算法的稳定性。各测试函数的参数设置如表2所示。统一设定5种算法的种群数量为30个,最大迭代次数为500。各算法运行30遍后,得到的优化结果如图2和表3所示。

表2 测试函数的参数设置

Table 2 Test function parameter configuration

测试函数名称	维度	最优值	类型
$F_1(x)$: Ellipsoid 函数	30	0	单峰函数
$F_2(x)$: Matyas 函数	2	0	单峰函数
$F_3(x)$: Rotated Ackley 函数	30	0	多峰函数
$F_4(x)$: Schaffer F6 函数	2	0	多峰函数
$F_5(x)$: Weierstrass 函数	30	0	多峰函数
$F_6(x)$: Six-Hump Camelback 函数	2	-1.034 6	多峰函数

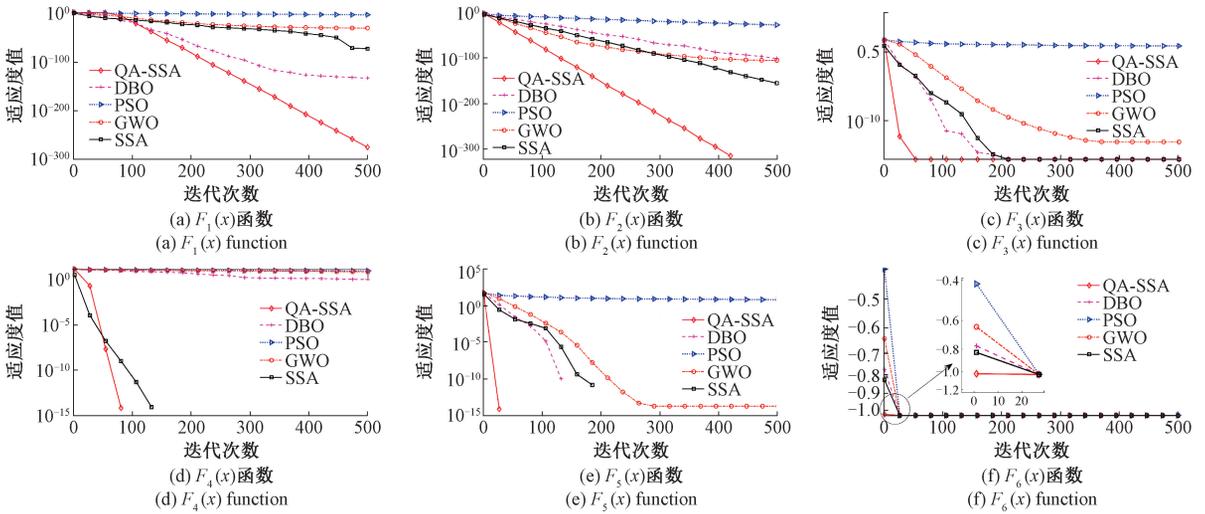


图 2 6 种测试函数的适应度曲线
Fig. 2 Fitness curves of six test functions

表 3 6 种测试函数的实验结果

Table 3 Experimental results on six benchmark functions

测试函数	指标	QA-SSA	SSA	PSO	GWO	DBO
$F_1(x)$	平均值	5.93×10^{-276}	1.42×10^{-27}	2.65×10^{-2}	1.23×10^{-10}	5.27×10^{-133}
	标准差	0	2.75×10^{-20}	2.65×10^{-2}	3.15×10^{-24}	8.27×10^{-101}
$F_2(x)$	平均值	5.40×10^{-316}	2.46×10^{-155}	3.22×10^{-26}	8.58×10^{-105}	5.47×10^{-101}
	标准差	0	1.08×10^{-154}	1.41×10^{-25}	4.61×10^{-104}	2.92×10^{-100}
$F_3(x)$	平均值	4.44×10^{-16}	4.44×10^{-16}	2.76	1.19×10^{-13}	4.44×10^{-16}
	标准差	0	0	3.08×10^{-1}	2.26×10^{-14}	0
$F_4(x)$	平均值	0	0	9.30	6.93	9.54×10^{-1}
	标准差	0	0	6.81×10^{-1}	1.99	1.81
$F_5(x)$	平均值	7.11×10^{-15}	1.43×10^{-11}	6.04	1.78×10^{-14}	9.82×10^{-11}
	标准差	0	0	1.15	5.02×10^{-15}	0
$F_6(x)$	平均值	-1.03	-1.03	-1.03	-1.03	-1.03
	标准差	0	0	0	0	0

从图 2 的适应度曲线可以看出,QA-SSA 算法对所有测试函数收敛更快、最终的适应度值更优且曲线变化稳定,表现出更强的全局搜索能力与鲁棒性,从而能够保障高维超参数空间的种群多样性,并通过算法有效的概率导向机制来维持长期探索能力。表 3 结果显示,对 $F_1(x) \sim F_5(x)$ 函数 QA-SSA 算法的适应度平均值和标准差指标更优,仅对 $F_6(x)$ 函数各算法的性能相当,从而说明了所提算法具有更好的全局优化能力。

3 QA-SSA-EARF 诊断应用流程

基于 QA-SSA-EARF 模型的故障诊断方法与应用流

程如图 3 所示,其中包含两大模块:离线建模、在线诊断与模型更新。在离线建模阶段,首先收集调节阀历史多工况故障数据,完成必要的预处理,提取时域与频域的关键特征;然后利用这些历史数据样本,采用 QA-SSA 算法对 EARF 的超参数进行全局优化搜索计算,得到一组在测试集上表现优异的初始模型最优参数配置,以此得到初始化 EARF 模型。

在线诊断与模型更新阶段是该方法的核心应用流程。每当输入新待诊断数据样本,EARF 首先执行一级诊断,将其区分为“正常”或“异常”样本,从而排除大量正常样本,减轻后续漂移检测与模型重训练的负担。然后,被判定为异常的数据样本再进入二级诊断,进行精

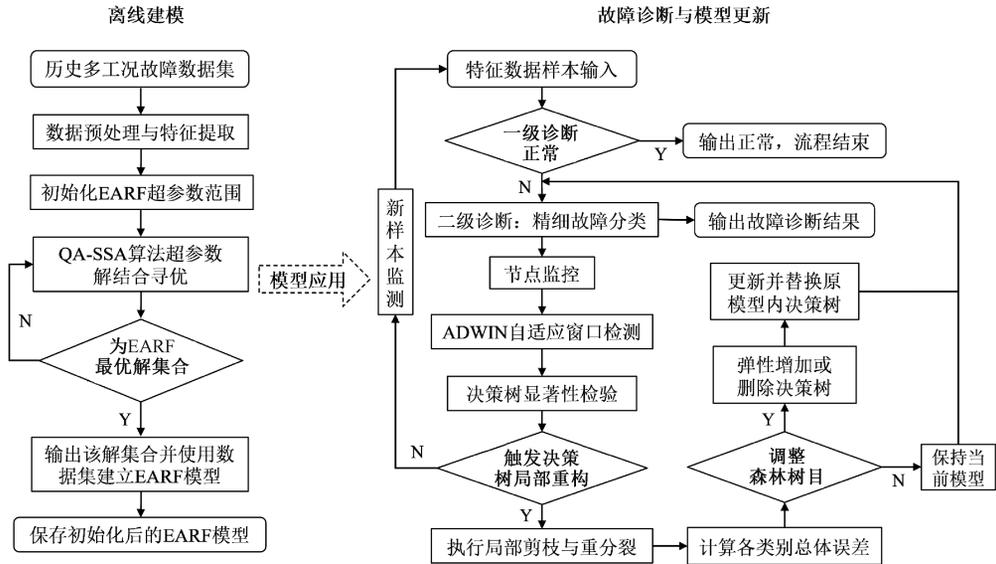


图 3 QA-SSA-EARF 模型构建与诊断应用流程

Fig. 3 Flowchart of EARF modeling and online diagnostic application

细的故障分类识别并输出诊断结果,同时该异常样本会被存入预设的缓冲队列,用于后续的漂移检测与模型子树训练。接着,在漂移检测环节,叶节点处的 ADWIN 检测器将对每一个故障类别的误差序列进行自适应窗口检测。通过计算误差序列的均值偏差 $|\mu_1 - \mu_2|$ 并与 Hoeffding 界限阈值 ϵ 比较,以决定是否在该叶子节点执行局部重分裂或剪枝替换。对于确认的局部概念漂移,模型主动对受影响子树进行局部修复,从而在不重建整棵树的前提下,快速重构该节点的判别规则。最后,在模型森林层面启用弹性树增减机制,依据模型的总体误差与预设阈值判断是否增加或减少森林规模,实现诊断准确率与泛化能力的协同提升。

4 调节阀故障诊断方法实验验证

4.1 数据来源

为验证所提出方法在调节阀故障诊断中的有效性,利用如图 4 所示的电动调节阀实验平台开展了模拟试验。该平台的直行程直通式单座电动调节阀,由 DN20 型电动执行器对阀杆开度进行精确操控。通过配备的智能控制仪表箱,可实现流量、压力及液位 3 种控制工况的切换,模拟调节阀多工况运行。实验在不同工况下设定的控制目标参数分别为:流量 $900 \text{ L}\cdot\text{h}^{-1}$ 、压力 18 kPa 和液位高度 15 cm 。3 种工况下各监测信号的变化范围如表 4 所示。实验平台可分别模拟调节阀的 6 种典型故障:阀口泄漏 I / II / III (通过调节阀门法兰模拟轻微、中度、严重 3 种程度的泄漏故障)、阀内堵塞 I / II (通过改

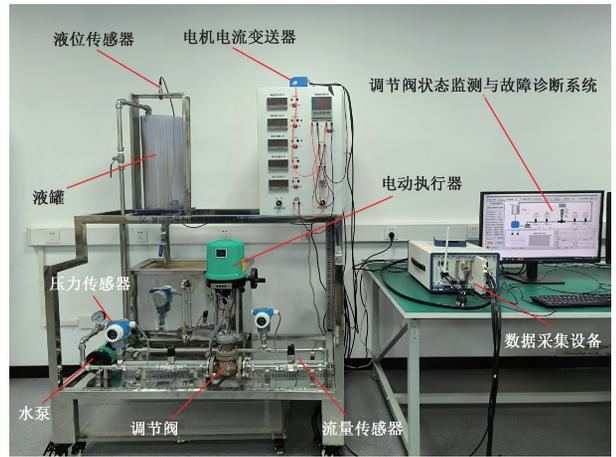


图 4 电动调节阀实验装置

Fig. 4 Experimental setup for electric control valve

表 4 实验过程中 3 种工况下各监测信号变化范围

Table 4 Observed ranges of monitored signals under three conditions during the experiment

控制工况	流量信号 变化区间 $/(L\cdot h^{-1})$	压力信号 变化区间 $/kPa$	液位信号 变化区间 $/cm$	执行器电 流变化区 间/ mA
流量控制	830~970	19~22	21~27	13~19
压力控制	780~850	17~19	17~21	8~13
液位控制	600~700	13~16	12~17	4~8

变阀内通道截面积分别为 30% 与 60% 模拟 2 种不同程度的堵塞故障)和阀体阀杆连接处外泄故障(通过打开阀

杆连接处螺栓模拟)。实验装置利用传感器和虚拟仪器平台实时测量采集调节阀运行过程的流量、压力、液位、电动执行器电流等参数信号数据。

本模拟实验设定每种控制工况下各类故障持续时长为 20 min。通过 NI PXIe-1092 主机与配套数据采集板卡,基于 LabVIEW 开发的调节阀状态监测系统完成数据采集,采样率为 100 Hz。在获取全工况数据集后,依次完成数据的离群点剔除、缺失值填补与平滑去噪等预处理操作,并采用长度为 400 个采样点(对应 4 s 时长)、无重叠的滑动窗口将原始时序数据集划分为样本段格式的数据样本序列。针对每个样本段,提取其均方根值、方差、标准差、峰值和能量等作为时域特征,提取其谱质心、谱展宽与谱偏度等作为频域特征。最后,将构建的特征样本集按 7:3 的比例分为训练

集与测试集,分别用于诊断模型的训练与性能评估。

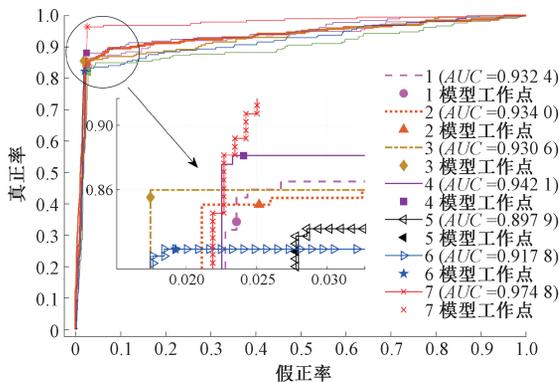
4.2 单一工况的验证

为检验所提出的诊断方法在单一工况下的性能,使用流量控制工况进行验证实验。根据前述的故障注入方案、数据采集与预处理流程以及统一的特征提取方法,构建了流量控制工况的样本数据集。通过建立基于随机森林(random forest, RF)、ARF、EARF、相应的经 QA-SSA 优化超参数后的 RF、ARF 与 EARF 等 6 种不同的诊断模型进行了对比消融实验。实验结果的性能评价指标如表 5 所示,经 QA-SSA 算法优化的后 3 种诊断模型的多分类受试者工作特征曲线(receiver operating characteristic curve, ROC 曲线)如图 5 所示,6 种模型用测试集得到的混淆矩阵如图 6 所示。

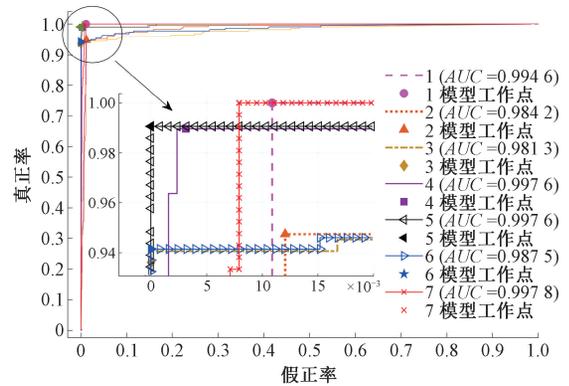
表 5 6 种诊断模型的评价指标结果

Table 5 Evaluation metrics for six models

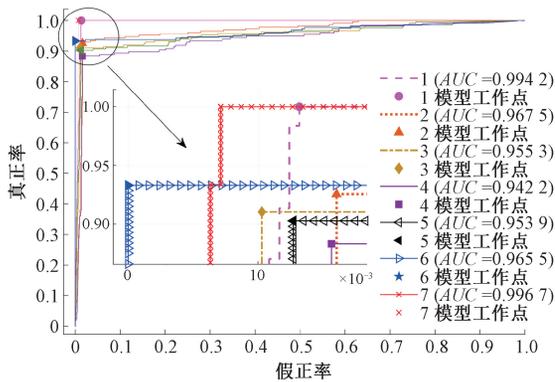
诊断模型	平均准确率/%	F1 分数/%	精确度/%	召回率/%	训练时间/s	单样本诊断平均时间/ms
RF	84.97	85.12	85.17	85.06	12.69	52.02
ARF	91.30	91.44	91.65	91.42	58.78	95.16
EARF	96.52	96.31	96.58	96.40	46.24	71.64
QA-SSA-RF	87.82	97.85	87.77	87.78	11.74	51.97
QA-SSA-ARF	93.83	93.83	93.96	93.83	54.63	93.21
QA-SSA-EARF	97.47	97.20	97.50	97.31	43.08	70.04



(a) RF模型ROC曲线
(a) RF model ROC curve



(c) EARF模型ROC曲线
(c) EARF model ROC curve



(b) ARF模型ROC曲线
(b) ARF model ROC curve

图 5 经 QA-SSA 优化后 3 种模型的 ROC 曲线

Fig. 5 ROC curves of three models by QA-SSA optimization

从表 5 结果与图 6 混淆矩阵可知,参数优化前 RF 模型对实验中测试集的诊断准确率低于 85%,模型训练时长为 12.69 s,单样本诊断需要 52.02 ms;而 ARF 因具备增量学习与漂移检测机制,能够有效学习故障相关敏感特征,诊断准确率比 RF 提升了 6.33%,但由于逐样本处理机制,其模型训练时间需要 58.78 s,约为静态 RF 模型训练时间的 5 倍,单样本诊断时间也更久,需要 95.16 ms;而 EARF 模型相较于未改进的 ARF 模型,其准

准确率提升了 5.22%,且因为两级诊断机制与局部剪枝策略改善了模型计算冗余的问题,使得模型的训练时长降低了 12.54 s,单样本诊断时间也降低了 23.52 ms。此外,在引入了 QA-SSA 算法对上述 3 种模型进行超参数调优后,其平均准确率比调优前分别提高了 2.85%、2.53%和 0.95%。同时,更优的模型超参数组合改善了模型结构,使各模型训练时长分别减少了 0.95、4.15 和

3.16 s,诊断时长分别减少 0.05、1.95 和 1.6 ms。图 5 展示的各诊断模型的 ROC 曲线中,每条曲线代表一种故障类别,曲线越靠近左上角,表明模型对该类故障的识别越准确。其中 EARF 模型的 ROC 曲线整体最贴近左上角,且各曲线下的面积(area under the curve, AUC)值最高,表明其诊断分类效果最好,从而验证了 QA-SSA-EARF 模型在调节阀单工况故障诊断中的优越性能。

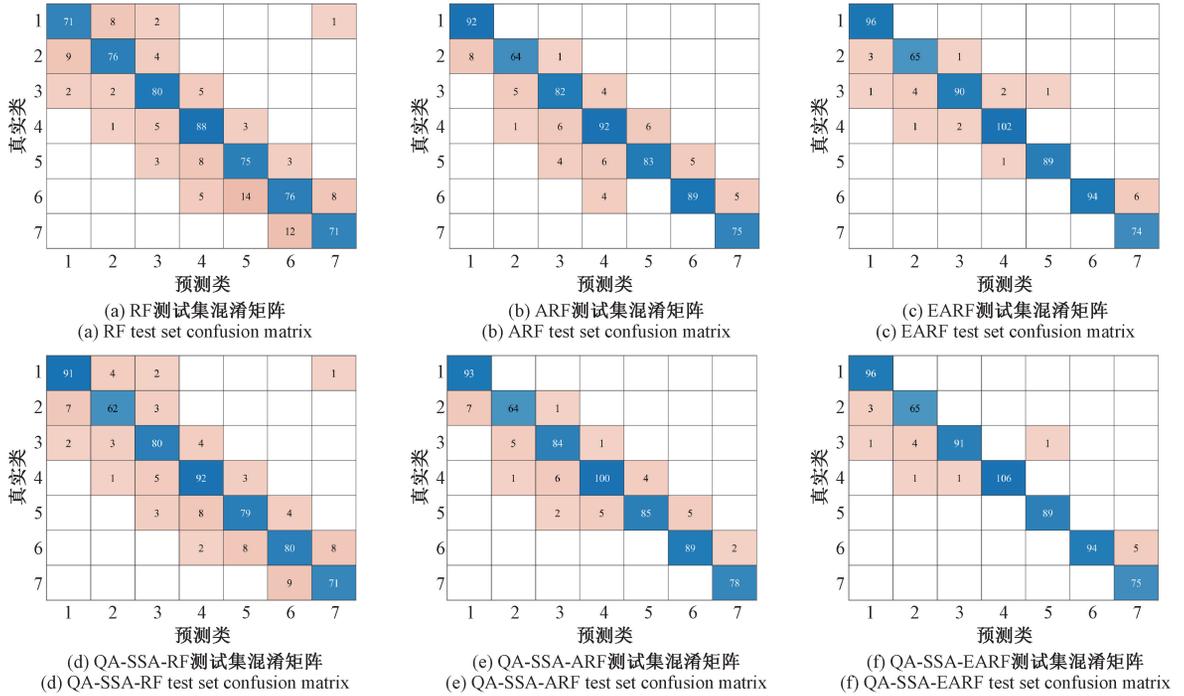


图 6 6 种模型的混淆矩阵对比

Fig. 6 Comparison of confusion matrices for six models

4.3 多工况验证

为验证所提出方法在多工况条件下的泛化能力,在相同设置下,模拟实验将调节阀依次切换为压力控制与液位控制两种工况,并依据前述相同的故障注入、数据采集与预处理流程构建相应的特征样本集。然后,将来自

这两种新工况的数据集分别输入基于流量控制工况训练得到的经 QA-SSA 优化后的 RF、ARF 与 EARF 诊断模型,进行多工况分类诊断比较。实验结果得到的各模型在两种工况下的性能指标如表 6 所示,各模型的预测诊断结果如图 7 所示。

表 6 不同工况下各诊断模型性能评价指标结果

Table 6 Metric results of diagnostic models under different operating conditions

工况	诊断模型	平均准确率/%	F1 分数/%	精确率/%	召回率/%	单样本平均诊断时长/ms
压力控制	QA-SSA-RF	84.24	84.25	84.55	84.23	51.88
	QA-SSA-ARF	90.54	90.59	90.82	90.54	93.24
	QA-SSA-EARF	93.12	93.14	93.27	93.13	70.19
液位控制	QA-SSA-RF	79.14	79.30	80.12	79.04	51.91
	QA-SSA-ARF	86.29	86.42	86.91	86.52	93.18
	QA-SSA-EARF	91.43	91.44	91.78	91.42	70.05

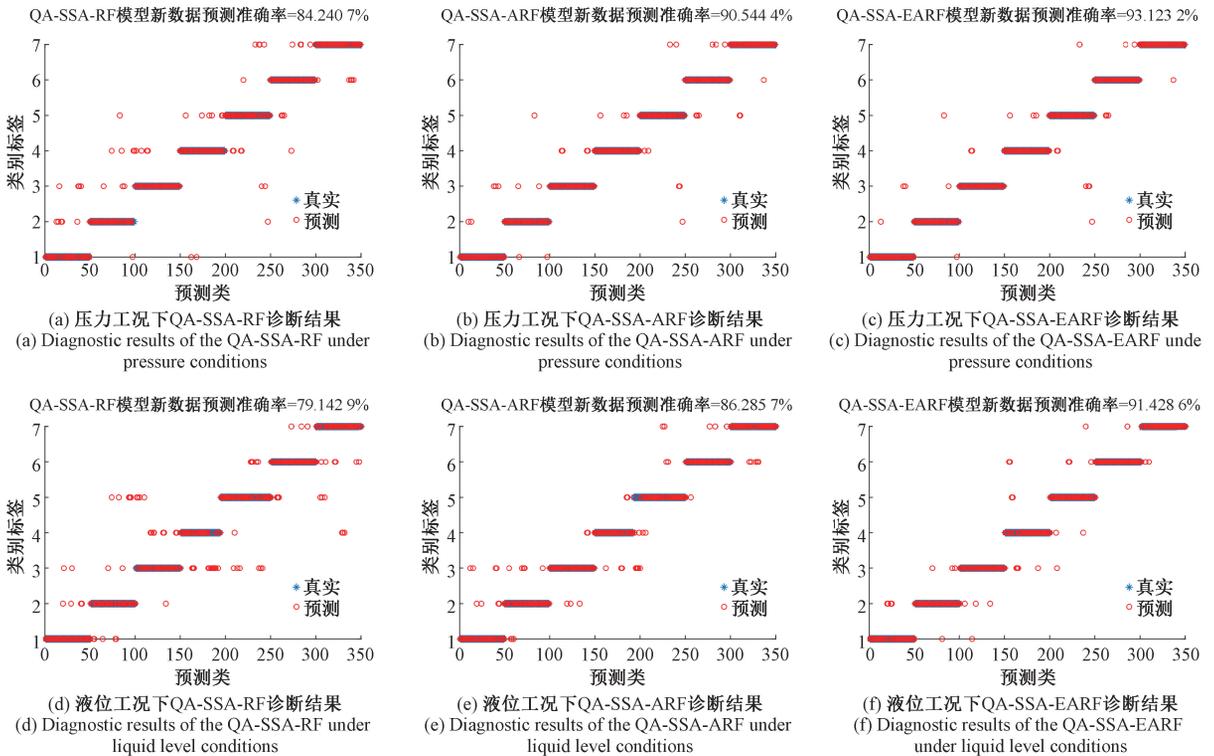


图7 多工况下分类诊断平均准确率

Fig. 7 Average accuracy of fault classification under multiple operating conditions

从表6所示的结果可知,在压力控制与液位控制工况下,由于故障特征分布发生变化,导致各诊断模型对不同故障的诊断准确率均有下降。然而,提出的QA-SSA-EARF模型在多工况条件下,平均诊断准确率均超过91%,与同样经QA-SSA优化的RF和ARF相比,EARF模型在压力工况下诊断准确率分别提升了88.8%和2.58%;在液位工况下则分别提升了12.29%与5.14%,效果显著。该结果表明,本文提出的EARF结构在输入其他工况数据样本后,模型内部的增量决策树能够通过节点内子特征集合的分类准确率下降快速检测到故障相关的分布漂移,并据此触发局部结构的重新分裂并用新分支替换失效子树。这一自适应替换机制使模型能迅速重建受影响区域的判别规则,增强了对故障特征的特征能力,从而提升了识别准确率,有更优的多工况泛化性能。而从图7可以直观看出3种模型在压力和液位控制工况下的诊断结果分布对比;对于QA-SSA-EARF模型方法,分类诊断结果的数据点在各类别区间与真实结果的标记高度重合,错误识别的样本明显最少;相比之下,经QA-SSA优化的RF与ARF存在多个区间较为集中的错误识别簇。综上研究表明:1) QA-SSA优化算法有助于各模型提升对故障的识别性能;2) 错误识别样本的减少说明改进的EARF模型结构能更好地捕捉到了不同工况

下的故障特征变化,从而快速更新模型提高识别的准确性。

5 结论

针对调节阀在多工况条件下故障诊断易受工况差异与数据分布变化影响的问题,本研究提出了一种将QA-SSA与EARF相结合的新型多工况故障分类诊断方法。通过对ARF模型结构设计两级分类策略,动态调整决策树数量,优化漂移检测器部署位置,以及引入基于滑动窗口误差的节点级局部剪枝策略,构建EARF模型,增强了模型对局部与全局工况变化的响应能力与鲁棒性;提出的改进QA-SSA算法,有效解决了EARF超参数耦合的寻优难题,显著提升了超参数搜索的效率与稳定性。最终通过单一工况和多工况模拟对比实验验证,表明了QA-SSA-EARF模型方法,相较于经过优化的RF与ARF模型,在故障诊断中更具优势。

本文方法研究了多种稳态工况下的分类诊断问题,对实际中更复杂的动态多工况及多工况切换过程的故障诊断问题仍有待进一步研究。同时,如何解决复杂环境干扰下具有强噪声、参数剧烈波动的数据处理问题,提高诊断模型的鲁棒性与泛化能力也需要深入研究。此外,

改善 QA-SSA 算法在高维参数优化过程中“维数灾难”带来的影响,也是后续关注的内容之一。上述问题的深入研究,将进一步增强诊断方法的工程适用性和可靠性。

参考文献

- [1] 陈林,王兴松,张逸芳,等. 阀门故障诊断技术综述[J]. 流体机械, 2015, 43(9): 36-42.
CHEN L, WANG X S, ZHANG Y F, et al. Survey of valve fault diagnostic technology[J]. Fluid Machinery, 2015, 43(9): 36-42.
- [2] 郝洪涛,王凯,张炳建,等. 多尺度特征自适应融合的气动控制阀故障诊断[J]. 仪器仪表学报, 2023, 44(10): 167-178.
HAO H T, WANG K, ZHANG B J, et al. Fault diagnosis of pneumatic control valves with multi-scale features adaptive fusion[J]. Chinese Journal of Scientific Instrument, 2023, 44(10): 167-178.
- [3] VAN K T, HUYNH T H, DAI T T, et al. Real-time fault detection algorithms for industrial process control valve[C]. 2022 6th International Conference on Computing, Communication, Control and Automation, IEEE, 2022: 1-7.
- [4] SINGH D, ALIYU A M, CHARLTON M, et al. Local multiphase flow characteristics of a severe-service control valve[J]. Journal of Petroleum Science and Engineering, 2020, 195: 107557.
- [5] HAN P Y, LIU Z Y, HE X, et al. Multi-condition fault diagnosis of dynamic systems: A survey, insights, and prospects[J]. IEEE Transactions on Automation Science and Engineering, 2025, 22: 15728-15744.
- [6] 乔卉卉,赵二贤,郝如江,等. 基于注意力机制与多源信息融合的变工况轴承故障诊断[J]. 仪器仪表学报, 2024, 45(9): 120-130.
QIAO H H, ZHAO ER X, HAO R J, et al. Attention mechanism and multi-source information fusion-based method for bearing fault diagnosis under variable operating conditions [J]. Chinese Journal of Scientific Instrument, 2024, 45(9): 120-130.
- [7] 姜午恺,周平. 非平衡数据下基于自适应迁移与宽度学习的多工况故障诊断[J]. 中国科学:信息科学, 2023, 53(5): 946-971.
JIANG W K, ZHOU P. Multi-condition fault diagnosis based on adaptive transfer and broad learning under imbalanced data [J]. Scientia Sinica (Informationis), 2023, 53(5): 946-971.
- [8] LI SH J, YU J B. A multisource domain adaptation network for process fault diagnosis under different working conditions[J]. IEEE Transactions on Industrial Electronics, 2022, 70(6): 6272-6283.
- [9] 钱思宇,秦东晨,陈江义,等. 基于卷积神经网络的领域适配模型的多工况迁移的轴承故障诊断[J]. 振动与冲击, 2022, 41(24): 192-200.
QIAN S Y, QIN D CH, CHEN J Y, et al. Bearing fault diagnosis based on a domain adaptation model of convolutional neural network under multiple working conditions[J]. Journal of Vibration and Shock, 2022, 41(24): 192-200.
- [10] PARISI G I, KEMKER R, PART J L, et al. Continual lifelong learning with neural networks: A review [J]. Neural Networks, 2019, 113: 54-71.
- [11] LIU M Y, YE F M, LI X, et al. Board-level functional fault identification using streaming data[J]. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2020, 40(9): 1920-1933.
- [12] ZHOU H, WANG K P, LI H, et al. Anomaly sound detection of industrial devices by using teacher-student incremental continual learning[J]. Measurement Science and Technology, 2024, 35(5): 056107.
- [13] 江灏,叶铭新,林蔚青,等. 基于时空动态检测的核电厂堆外中子探测器故障检测方法[J]. 仪器仪表学报, 2024, 45(9): 131-144.
JIANG H, YE M X, LIN W Q, et al. Fault detection for ex-core neutron detectors in nuclear power plants using the spatial-temporal dynamic detection model[J]. Chinese Journal of Scientific Instrument, 2024, 45(9): 131-144.
- [14] LU J, LIU AN J, DONG F, et al. Learning under concept drift: A review[J]. IEEE Transactions on Knowledge and Data Engineering, 2018, 31(12): 2346-2363.
- [15] GOMES H M, BIFET A, READ J, et al. Adaptive random forests for evolving data stream classification[J]. Machine Learning, 2017, 106(9): 1469-1495.
- [16] MUALLEM A, SHETTY S, PAN J W, et al. Hoeffding tree algorithms for anomaly detection in streaming datasets: A survey[J]. Journal of Information Security, 2017, 8(4): 339-361.
- [17] LI J T, GUO J Z, GUO CH, et al. Online anomaly detection for nuclear power plants via hybrid concept drift[J]. Energies, 2025, 18(17): 4491.
- [18] KUMAR A, KAUR P, SHARMA P. A survey on Hoeffding tree stream data classification algorithms[J]. Career Point University, Hamirpur-Research Journal, 2015, 1(2): 28-32.
- [19] 李斌,杨润,舒洋. 改进麻雀搜索算法在 PMSM 匝间短路中应用研究[J]. 电子测量与仪器学报, 2024, 38(7): 224-235.

- LI B, YANG R, SHU Y. Application of improved sparrow search algorithm in PMSM inter-turn short-circuit [J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(7): 224-235.
- [20] 周玉, 房倩, 裴泽宣, 等. 基于改进麻雀搜索算法优化 LSTM 的滚动轴承故障诊断[J]. 工程科学与技术, 2024, 56(2): 289-298.
- ZHOU Y, FANG Q, PEI Z X, et al. Fault diagnosis of rolling bearing based on improved sparrow search algorithm optimized LSTM [J]. Advanced Engineering Sciences, 2024, 56(2): 289-298.
- [21] XUE J K, SHEN B. A novel swarm intelligence optimization approach: Sparrow search algorithm [J]. Systems Science & Control Engineering, 2020, 8(1): 22-34.
- [22] YUE Y G, CAO L, LU D W, et al. Review and empirical analysis of sparrow search algorithm [J]. Artificial Intelligence Review, 2023, 56(10): 10867-10919.
- [23] YANG X X, QIU Y H. Research on improving gray wolf algorithm based on multi-strategy fusion [J]. IEEE Access, 2023, 11: 66135-66149.
- [24] CUI X R, LUO Q F, ZHOU Y Q, et al. Quantum-inspired moth-flame optimizer with enhanced local search strategy for cluster analysis[J]. Frontiers in Bioengineering and Biotechnology, 2022, 10: 908356.

- [25] LEE C Y. Entropy-Boltzmann selection in the genetic algorithms [J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 2003, 33(1): 138-149.

作者简介



罗柯达, 2018 年于湖南城市学院获得学士学位, 现为南京工业大学硕士研究生, 主要研究方向为故障诊断。

E-mail: 3023454813@qq.com

Luo Keda received his B. Sc. degree from Hunan City University in 2018. He is currently a master's student at Nanjing Tech University. His main research interest includes fault diagnosis.



张登峰(通信作者), 2003 年于南京理工大学获得博士学位, 现为南京工业大学教授、博士生导师, 主要研究方向为动态系统的故障诊断与容错控制理论、工业人工智能应用技术。

E-mail: zhdf@njtech.edu.cn

Zhang Dengfeng (Corresponding author) received his Ph. D. degree from Nanjing University of Science and Technology in 2003. He is currently a professor and Ph. D. advisor at Nanjing Tech University. His main research interests include fault diagnosis and fault-tolerant control theory for dynamic systems, and AI technology for industrial applications.