

基于 KPCA 融合 AdaBoost – IBOA – ELM 模型的 TE 过程故障诊断

赵文虎, 蔡生宏, 王 文

(新疆工业职业技术学院 机电工程系, 新疆 乌鲁木齐 830022)

摘要: 为了保障化工系统的安全运行和高质量生产, 准确判别化工过程的故障就显得尤为重要。针对田纳西 – 伊斯曼 (Tennessee Eastman, TE) 过程故障难以区分以及神经网络等方法在故障诊断中分类准确率较低、分类不稳定等问题, 提出一种优化改进极限学习机 (extreme learning machine, ELM) 的 TE 过程故障诊断模型。首先利用核主成分分析 (kernel principal components analysis, KPCA) 方法对 TE 过程数据进行降维和特征提取, 然后采用改进蝴蝶算法 (improved butterfly optimization algorithm, IBOA) 优化 ELM 的权值和阈值, 最后利用自适应提升 (adaptive boosting, AdaBoost) 算法对分类器进行集成, 完成故障分类。仿真结果表明, IBOA 比其他优化算法具有更好的寻优能力, 改进效果显著, AdaBoost – IBOA – ELM 模型能够对测试集中的不同故障进行准确分类, 最后的分类准确率高达 98.5%, 通过和其他网络对比, 进一步验证了模型的合理性和优越性。

关键词: 田纳西 – 伊斯曼过程; 核主成分分析; 改进蝴蝶算法; 极限学习机; 故障分类

中图分类号: TP183

文章编号: 1000 – 0682 (2024) 04 – 0102 – 08

文献标识码: A

DOI: 10.19950/j.cnki.CN61 – 1121/TH.2024.04.020

Fault diagnosis of TE process based on KPCA fusion AdaBoost – IBOA – ELM model

ZHAO Wenhui, CAI Shenghong, WANG Wen

(Department of Electrical and Power Engineering, Xinjiang Industrial Vocational and
Technical College, Xinjiang Urumqi 830022, China)

Abstract: In order to ensure the safe operation and high – quality production of chemical systems, it is particularly important to accurately identify the faults of chemical processes. In order to solve the problems of Tennessee Eastman (TE) process fault indistinguishability, neural network and other methods in fault diagnosis, such as low classification accuracy and unstable classification, a TE process fault diagnosis model with optimized and improved extreme learning machine (ELM) was proposed. Firstly, the kernel principal components analysis (KPCA) method was used to reduce the dimensionality and extract features of the TE process data, then the improved butterfly optimization algorithm (IBOA) was used to optimize the weights and thresholds of the ELM, and finally the adaptive boosting algorithm integrates the classifier to complete the fault classification. The simulation results show that IBOA has better optimization ability than other optimization algorithms, and the improvement effect is significant, and the AdaBoost – IBOA – ELM model can accurately classify different faults in the test set, and the final classification accuracy is as high as 98.5%, which further verifies the rationality and superiority of the model by comparing with other networks.

Keywords: Tennessee Eastman process; kernel principal component analysis; improved butterfly algorithm; adaptive boosting algorithm; fault classification

收稿日期: 2024 – 03 – 02

基金项目: 新疆乌鲁木齐市工业机器人系统操作员技能大师工作室项目

第一作者: 赵文虎 (1995 –), 男, 汉族, 甘肃金昌人, 硕士研究生, 助教, 研究方向为故障诊断、参数优化。

0 引言

随着化工生产过程的不断发展,流程也变得越来越复杂,发生事故的风险也變得越来越大。为了保证工业过程在运行中的稳定可靠性,就需要在事故发生前,有一个及时有效的监控,并且对产生的故障类型有一个明确的诊断。因此,在现代化的工业生产过程中,如何找到一种有效的方法保证工业过程的安全高效运行就成为了一个被广泛关注和值得研究的问题。如果化工过程出现意外,这就会造成非常严重的问题,包括经济损失、人员伤亡和环境污染。所以,准确判断故障且及时防止事故发生是最需要研究的问题^[1]。

目前,用在 TE 过程中的诊断方法有很多。针对数据的复杂程度,顾幸生等人^[2]提出的方法先利用局部近邻标准化来消除多模态特征,然后再利用方向熵加权核熵成分分析来进行降维提取特征。江升等^[3]将稀疏过滤和逻辑回归算法结合起来,用到 TE 过程中,表现出了很高的诊断准确率。蓝艇等人^[4]提出的误差生成策略能更好地体现数据本质,具有很大的优越性。索寒生等人^[5]将基于慢特征分析的方法用在 TE 过程中,具有较高的故障诊断准确率。针对化工故障数据复杂,不易区分等问题,吕鹏飞等人^[6]提出的改进核 Fisher 故障诊断方法,可以有效改进样本混叠现象,提升运算速度,提高了故障诊断精度。张成等人^[7]提出的故障检测策略能够有效诊断故障,也可以有效消除数据结构对 TE 过程故障诊断的影响。

针对 TE 过程复杂的故障类型、特征维数和上述所用到的方法,该文提出一种新的方法,首先利用 KPCA 方法的优势降低数据集的维数,因为传统的降维方法不能很好地处理非线性数据,然后通过 ELM 来进行故障诊断分类,由于极限学习机虽然具有训练速度快等优点,但是也存在稳定性差、分类精度不高等问题,对其诊断效果会有很大的影响,所以利用 IBOA 的优越性能对 ELM 的权值和阈值进行优化,最后将 IBOA - ELM 作为弱分类器进行 AdaBoost 集成,建立 AdaBoost - IBOA - ELM 诊断分类模型,通过实验验证该模型的分类性能,为化工过程的安全运行提供一个高效率和高质量的保证。

1 KPCA 算法

KPCA 是一种非线性方法^[8]。其主要原理是将输入样本 $X, X = \{x_1, x_2, \dots, x_N\}$ 通过非线性映射 j

映射到特征空间 F 变换成 $j(x_i)$ 。假设映射数据是零均值,则映射数据 $j(x)$ 的协方差矩阵为:

$$C_F = \frac{1}{N} \sum_{i=1}^N j(x_i) j(x_i)^T \quad (1)$$

协方差矩阵的特征值 l 和对应的特征向量 n 的等式可以表示为:

$$l n = C_F n \quad (2)$$

根据以上分析,映射 $j(x_k)$ 和等式(2)的内积可以表示为:

$$l(j(x_k) \cdot n) = (j(x_k) \cdot C_F n) \quad (3)$$

其中, $k = 1, 2, \dots, N$, 对应的特征值的特征向量 n 的矢量矩阵可以表示为:

$$n = \sum_{i=1}^N a_i j(x_i) \quad (4)$$

定义 $n \times n$ 的矩阵 $K, K_{ij} = \langle j(x_i), j(x_j) \rangle$, 通过以上等式,化简得到:

$$N l a = K a \quad (5)$$

列向量由 a_i 组成。

设 $l_1 \geq l_2 \geq \dots \geq l_N$ 为非零特征值,对应的特征向量为 a_1, a_2, \dots, a_N 。求得的归一化特征向量为

$$n_k = \sum_{i=1}^N a_i^k j(x_i)。$$

式 $\langle n_k, n_k \rangle = 1, k = 1, 2, \dots, p$ 变为:

$$l_k \langle a_k, a_k \rangle = 1 \quad (6)$$

主成分可以通过下式计算:

$$t_k = \langle n_k, j(x) \rangle = \sum_{i=1}^N a_i^k \langle j(x_i), j(x) \rangle \quad (7)$$

目前, KPCA 使用不同的核函数。在该文中,选择的是使用比较多的径向基函数(RBF),其中 s^2 是方差的平方。

$$K(x, y) = \exp\left(-\frac{\|x - y\|^2}{2s^2}\right) \quad (8)$$

2 AdaBoost - IBOA - ELM 模型原理

2.1 ELM 算法

ELM^[9]是一种单隐含层的前馈神经网络。具有结构简单、学习速度快等特点。由文献^[10-11]提出的改进极限学习机算法克服了训练时间长、无法到达全局最小值、学习率的选择敏感等缺点。原理介绍如下:

设输入权重矩阵为 $w = [w_1, w_2, \dots, w_l]^T$, 阈值矩阵为 $b = [b_1, b_2, \dots, b_l]^T$, 输出权值矩阵为 $\beta = [\beta_1, \beta_2, \dots, \beta_l]$, 其中 $w_i = [w_{i1}, w_{i2}, \dots, w_{im}]^T, \beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]$ 。

设激活函数为 $g(x)$, 则输出 $Y \in R^{m \times n}$ 的表达式

可以表示为:

$$Y = H\beta \quad (9)$$

式中: H 为隐含层输出矩阵。

$$H = \begin{bmatrix} g(w_1x_1 + b_1) & g(w_2x_1 + b_2) & \cdots & g(w_Nx_1 + b_N) \\ g(w_1x_2 + b_1) & g(w_2x_2 + b_2) & \cdots & g(w_Nx_2 + b_N) \\ \vdots & \vdots & \ddots & \vdots \\ g(w_1x_N + b_1) & g(w_2x_N + b_2) & \cdots & g(w_Nx_N + b_N) \end{bmatrix} \quad (10)$$

当 g 逼近零时,在 ELM 的训练过程中,只需要用下式来求输出层权值 β :

$$\min_{\beta} \| H\beta - Y \| \quad (11)$$

式中,最小二乘解可以表示为:

$$\beta = H^+ T \quad (12)$$

式中: H^+ 为输出矩阵 H 的 *Moore-Penrose* 广义逆矩阵。所以通过 ELM 求解最小二乘解就可以完成整个训练过程,网络的学习速度也将大大提高。

2.2 IBOA 算法

蝴蝶算法 (BOA) 是通过研究蝴蝶觅食行为,由 Arora^[12] 提出的。但是标准 BOA 算法存在易陷入局部最优和后期寻优速度低等问题,因此为了提高 BOA 算法的寻优性能,提出一种寻优效果较好的改进蝴蝶算法 (IBOA)。

2.2.1 BOA 算法

在算法中,每一只蝴蝶都会产生不同的香味浓度,不同蝴蝶产生的香味浓度通过式 (13) 表示为:

$$f^i = cI^\alpha \quad (13)$$

其中: f_i 为蝴蝶感受到的香味强度; c 为感官因子,通常取 0.01; I 为刺激强度; α 为依赖于形态的幂指数,通常取 0.1。

蝴蝶搜索的过程分为全局搜索和局部搜索,当蝴蝶感受到香味时就会沿香味不断移动自己的位置,进行全局搜索。如果蝴蝶不能感受到香味,就会随机改变自己的位置,进行局部搜索。全局搜索公式如式 (14),局部搜索公式如式 (15):

$$x_i^{t+1} = x_i^t + (r^2 \times g^* - x_i^t) \times f_i \quad (14)$$

$$x_i^{t+1} = x_i^t + (r^2 \times x_k^t - x_i^t) \times f_i \quad (15)$$

其中: x_i^t 表示第 t 次迭代时蝴蝶 i 对应的位置; g^* 表示在这个过程中的最优解; r 是 $[0, 1]$ 之间的随机数。 x_j^t 表示第 t 次迭代时蝴蝶 j 对应的位置, x_k^t 表示第 t 次迭代时蝴蝶 k 对应的位置,它们都是在种群中随机选择的。

2.2.2 改进 BOA 算法

在 IBOA 中,首先利用反向学习策略使得所有

蝴蝶个体随机分布,使整个种群更多样化;然后通过自适应调节策略使算法的全局搜索和局部搜索能力得到平衡;最后为了避免算法陷入局部最优,通过 Tent 混沌局部搜索的策略,进行全面搜索,使得种群实现均匀的分布,还可以保证算法的收敛速度。在自适应调整策略中,权重设置如公式 (16) 所示:

$$w = \begin{cases} w_{\min} + \frac{w_{\max} - w_{\min} \times (f_i - f_{\text{avg}})}{f_{\text{avg}} - f_{\min}}, & f_i \leq f_{\text{avg}} \\ w_{\max}, & f_i > f_{\text{avg}} \end{cases} \quad (16)$$

其中: w_{\max} 为迭代初始阶段的惯性权重; w_{\min} 为迭代结束时的惯性权重; f_i 表示蝴蝶个体当前的适应度值; f_{avg} 和 f_{\min} 分别表示当前蝴蝶群体的平均适应度值和最小适应度值。经过反复试验,惯性权重 w_{\max} 和 w_{\min} 分别取为 1.0 和 0.6 时,算法的性能最优。式 (14) 的全局搜索公式就变为式 (17) 所示:

$$x_i^{t+1} = w \times x_i^t + (r^2 \times g^* - x_i^t) \times f_i \quad (17)$$

在混沌局部搜索中,最终搜索公式可以表示为:

$$x_i^k(d+1) = x_{\min}^i + \frac{x_{\max}^j - x_{\min}^j}{2} \times (2c_i^k(d+1) - 1) \quad (18)$$

其中: $i = 1, 2, \dots, n$; $k = 1, 2, \dots, K$, x_{\min}^j 和 x_{\max}^j 分别为决策变量 x 的第 j 维变量的上下界。

2.3 AdaBoost 算法

AdaBoost^[13] 是一种在 Boosting 算法的基础上进行改进的集成学习算法,通过组合多个弱分类器来构建一个强分类器。AdaBoost 的基本思想是通过迭代训练一系列弱分类器,在每次调整样本权重的训练过程中,对前一次分错的样本多加关注。具有简单易实现、高精度、可处理多类别问题和不易过拟合等优点。具体步骤如下:

步骤 1 对样本权重系数进行初始化, $w_i = 1/N$, $i = 1, 2, \dots, N$ 。

步骤 2 通过权重 w_i 和训练样本的学习得到弱分类器 f_m , $m = 1, 2, \dots, M$ 。

步骤 3 计算当前弱分类器的误差和权重。

$$\text{err}_m = E_m[1_{(y \neq f_m(x))}] \quad (19)$$

$$c_m = \log \frac{1 - \text{err}_m}{\text{err}_m} \quad (20)$$

步骤 4 根据式 (21) 更新样本权重。

$$w_i \leftarrow w_i \cdot \exp[c_m \cdot 1_{(y_i \neq f_m(x_i))}], i = 1, 2, \dots, N \quad (21)$$

步骤 5: 构建强分类器:

$$F(x) = \text{sign} \left[\sum_{m=1}^M c_m f_m(x) \right]$$

3 基于 AdaBoost – IBOA – ELM 的 TE 过程故障分类

对于 AdaBoost – IBOA – ELM 模型,首先以 ELM 为基础,针对 ELM 受随机参数影响易导致性能不稳定的问题,采用改进蝴蝶算法对 ELM 的权值和阈值

进行寻优,得到最优的参数后,再通过 AdaBoost 算法对 IBOA – ELM 弱学习器进行加权集成,使 AdaBoost – IBOA – ELM 模型的精度得到大幅度的提高,最后将其用于 TE 过程的故障诊断分类中,具体算法流程图如图 1 所示。

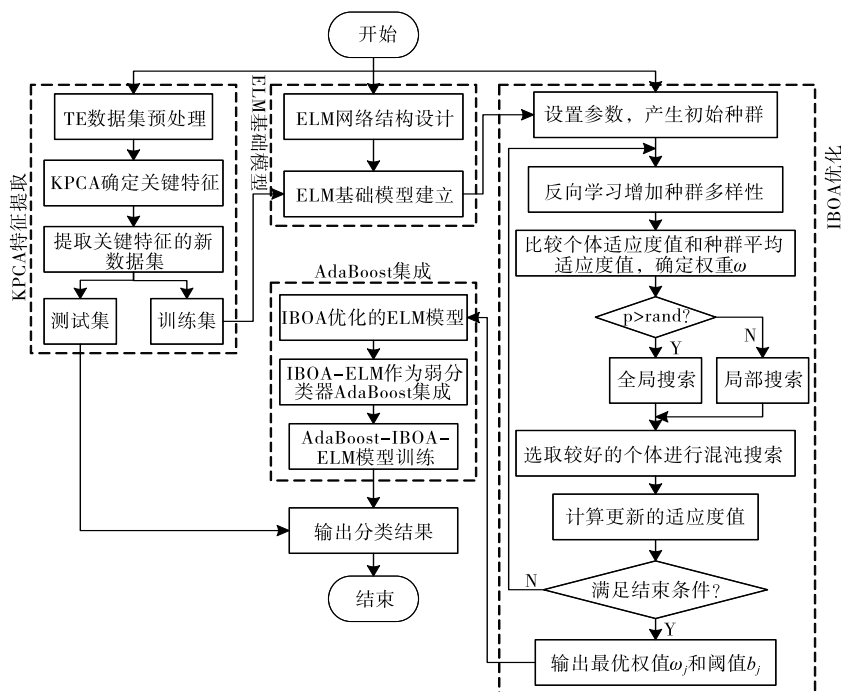


图 1 AdaBoost – IBOA – ELM 模型流程图

其中,在对 ELM 的权值和阈值进行寻优时,先将训练样本输入 ELM 进行学习,把适应度函数的最小化作为优化目标,不断优化参数,以实际值与预测值的均方根误差作为适应度函数,公式如下:

$$f(a_1) = \sqrt{\frac{\sum_{j=1}^{N_{\text{train}}} \left\| \sum_{i=1}^m \beta_i g(\mathbf{w}_i \times \mathbf{x}_j + b_i) - t_j \right\|^2}{m N_{\text{train}}}} \quad (22)$$

式中: f 为适应度函数; N_{train} 为训练样本数; m 为隐含层节点数。

4 实证分析

4.1 TE 过程数据集介绍

TE 过程^[14]是由 Downs 和 Vogel 提出的一个典型的复杂工业过程模型,已经在过程控制技术领域得到了广泛的应用^[15]。

TE 过程包含了由 41 个测量变量和 12 个控制变量组成的 21 种故障,21 种故障又分为 4 种类型,故障 1 ~ 7 呈现阶跃变化,故障 8 ~ 12 呈现随机变化,故障 13 和反应动力学有关,呈现出缓慢漂移的

特性,故障 14 ~ 15 是与粘滞阀有关的故障,故障 16 ~ 20 是未知故障,故障 21 为恒定位置的故障。

在整个数据集的采集过程中,采样时间均为 3 分钟。在采集训练集数据的过程中,仿真时间设置为 25 小时,引入故障的时间设置在 1 小时后,这种情况采集到总的样本数为 480。在测试集数据的采集过程中,仿真时间设置为 48 小时,引入故障的时间设置在 8 小时后,这种情况采集到总的样本数为 960,其中在这 960 个样本中,包含了 160 个正常样本和 800 个含有故障的样本。

4.2 KPCA 特征提取及故障识别

故障诊断的第一步就是识别故障,判断是否有故障发生,但是特征之间会存在信息重叠的情况,因此通过 KPCA 选取贡献率高的主成分代表原始特征信息,实现数据降维。先求解核矩阵 \mathbf{K} ,再确定主元的特征值和特征向量。根据文献[16]可知,RBF 函数中 σ^2 的大小对降维效果有很小的影响,选取 $\sigma^2 = 10\,000$,实验中累计贡献率设置为 85%,贡献率占比图如图 2 所示,当累计贡献率达到 85% 的时候,主元个数变为 17 个,可以使数据从最初的 52 个变量

变为 17 个变量。

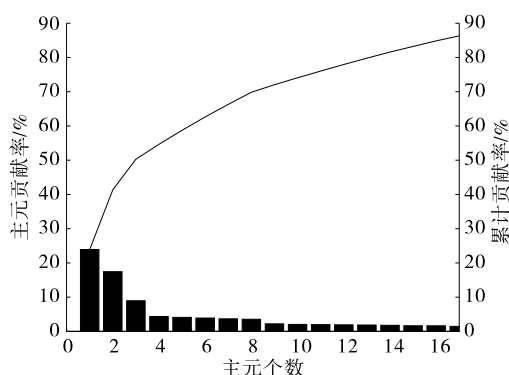


图2 贡献率占比图

根据降维后的正常工况数据得到 Hotelling 统计量(T^2)和平方预测误差(Squared Prediction Error, SPE)的控制限值。然后将待测数据的统计量与控制限值进行对比。如果待测数据的统计量超过控制限值,表明有故障发生。反之,就说明没有故障发生。

4.3 数据预处理

TE 过程故障诊断模型是以各故障特征作为输入参数,分类结果作为输出参数。由于这些故障特征之间的量纲和数值范围存在很大差异,对模型的影响很大,为了这些特征之间具有一定的可比性,通过式(23)将这些特征进行归一化,将所有数据缩放到区间 $[0,1]$ 内,以此来消除它们之间的量纲差异。

$$\bar{x}_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (23)$$

式中: x_i 表示原始数据; x_{\max} 表示原始数据所能取的最大值; x_{\min} 表示原始数据所能取的最小值; \bar{x}_i 表示对数据进行归一化后的结果。

4.4 网络结构设计

在 AdaBoost - IBOA - ELM 模型中,隐含层节点数的选择对模型精度至关重要,如果选择较多的节点个数,就容易导致过拟合现象的出现,使得最后的分类结果和真实情况出现偏差,不能反映其真实规律,严重影响分类精度,如果选择较少的节点个数,对复杂数据的学习就会变的很差,导致数据不能进行有效拟合,也会影响分类精度。因此,通过分类准确率的高低来确定最佳隐含层节点数,此时的正则化参数 λ 取为0.3,图3是分类准确率随着隐含层节点个数的变化曲线。图中显示在隐含层节点个数不断增加的同时分类准确率也在逐渐提高,虽然中间存在几次波动,但整体是呈上升趋势,但是在隐含层节点个数到达27的时候,继续增加,分类准确率都维持在同一个值,因此选择27作为模型的隐含层节点数。

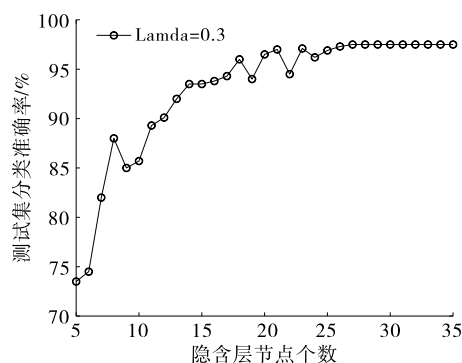


图3 分类准确率随隐含层节点数的变化曲线

4.5 参数设置及性能验证

基于4.2和4.4的分析,将通过 KPCA 提取的17维特征作为输入,分类结果作为输出,所以整个模型的输入神经元设为17,隐含层节点数取为27,输出神经元设为1。

设置 BOA 算法的参数为:蝴蝶的种群规模为30,最大迭代次数为500。IBOA 算法特有的3个参数为:最大混沌搜索次数30, $w_{\max}=0.9$, $w_{\min}=0.5$ 。为了验证 IBOA 算法的寻优性能,选择如式(24)所示的 Schaffer 低维函数和如式(25)所示的 Rosenbrock 高维函数来进行函数测试,搜索范围设为 $[-30,30]$ 。

$$f(x) = \sum_{i=1}^D ix_i^4 + \text{random}[0,1] \quad (24)$$

$$f(x) = \sum_{i=1}^D [100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2] \quad (25)$$

将 IBOA 与灰狼算法^[17](grey wolf optimizer, GWO)、BOA 算法、粒子群算法^[18](particle swarm optimization, PSO)、布谷鸟算法^[19](cuckoo search, CS)、蝙蝠算法^[20](bat algorithm, BA)和 BOA 算法进行对比,种群规模和最大迭代次数也分别设为30和500,PSO 算法中的两个学习因子为 $c_1 = c_2 = 2$,粒子最大速度和最小速度分别为6和-6。CS 算法中宿主鸟重新对鸟巢进行建立的概率为0.25。对比结果如图4和图5所示。

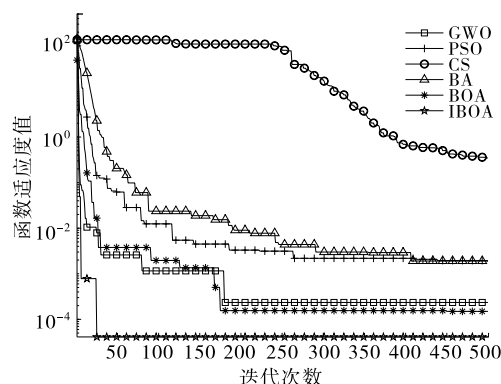
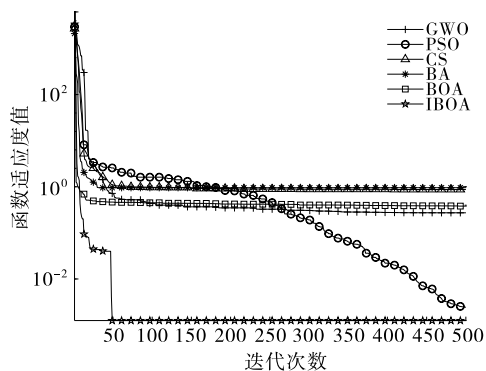


图4 Schaffer 函数下的适应度曲线对比



从图 4 和图 5 中可以看出,相比于其他 5 种算法,IBOA 算法迭代寻优的过程在 100 代内就可以完成,并且函数适应度值是最优的,而其他 5 种算法不能同时满足这两个条件,所以 IBOA 算法具有最快的收敛速度和最好的收敛精度,充分说明了 IBOA 算法的寻优效果最好。

该文选取其中的 17 种故障数据来验证 AdaBoost - IBOA - ELM 模型的分类性能,去除的 4 种故障类型分别是 3、9、15 和 21,因为故障 3、9、15 是不可识别的,故障 21 是常量。然后选择训练集和部分测试集来进行仿真实验,部分测试集由随机选取的每种故障的 200 组样本构成。这样选取的意义在于将故障准确分开的同时还能有效减少分类时间。需要注意的是,在分类标签中,0 对应正常情况,1 ~ 21 分别对应不同的故障类型。

在利用训练集对网络进行训练时,随着隐含层节点数的变化,训练集准确率和均方根误差发生相应的变化,如图6所示。

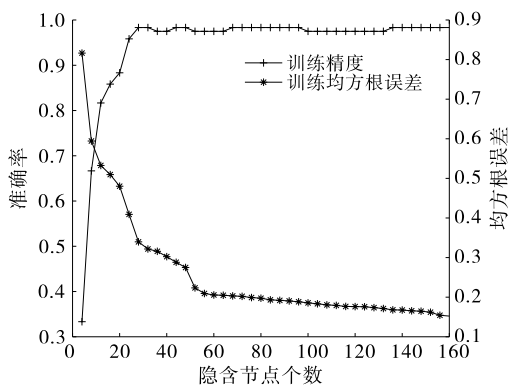


图 6 网络训练图

从图 6 中可以看出,随着隐含层节点数的增加,训练集准确率达到了 99.1%,均方根误差最小为 0.15,充分说明了训练是成功的。

在利用 AdaBoost - IBOA - ELM 模型对测试集的所有故障进行分类时,为了将分对和分错的样本

进行明确区分,18种工况的分类情况通过混淆矩阵表示,其中列是目标类别,反映数据的真实情况,行则是输出类别,反映的是其他数据错分到该类别的数目,每个小格内上方的数字代表样本数目,下方的百分数代表所占的比例,分类结果如图7所示。

[illegible]

图7 测试集混淆矩阵分类图

从图 7 中可以看出,整个测试集 18 种工况的分类情况都比较好,故障 16 和 19 错分样本最多,导致分类准确率最低,故障 6 和 17 全部分对,分类准确率达到 100%,最终的分类精度也高达 98.5%,充分说明了所提算法应用在 TE 过程故障诊断上具有很好的效果,能够准确将 18 种故障进行分类。

为了进一步说明 AdaBoost - IBOA - ELM 模型的广泛性和优越性,通过选择概率神经网络^[21] (probabilistic neural network, PNN)、BP 神经网络^[22] (back propagation, BP) 和支持向量机^[23] (support vector machine, SVM) 进行对比,其中, BP 的学习率取为 0.01; PNN 的平滑因子取为 25; SVM 的惩罚因子取为 1,核核函数半径取为 0.2。

为了说明不同模型的性能,对另外 3 种网络用相同的方法构成 AdaBoost - IBOA - BP、AdaBoost - IBOA - PNN 和 AdaBoost - IBOA - SVM 模型,再将 4 种模型 20 次的分类准确率进行比较,分类准确率对比如图 8 所示。

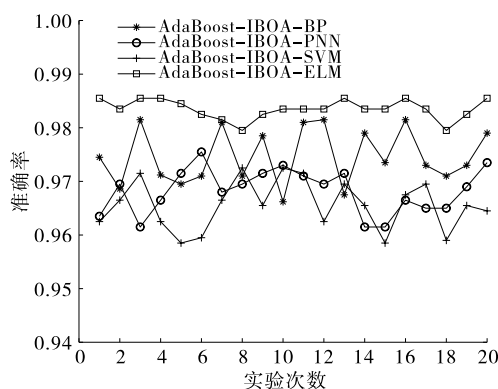


图 8 4 种模型的准确率对比图

从图 8 中可以发现, AdaBoost - IBOA - ELM 20 次的分类准确率明显高于其他三种模型, 并且 20 次的实验数据结果相差也不大, 足以说明 AdaBoost - IBOA - ELM 的稳定性也很好, 由此可见, AdaBoost - IBOA - ELM 模型是适用于 TE 过程的, 比其他 3 种模型性能更优, 分类效果也更好。

为了进一步验证 AdaBoost 算法和 IBOA 算法对网络的不同优化性能, 通过分类准确率、均方根误差 (root mean square error, RMSE)、平均绝对误差 (mean absolute error, MAE) 和平均绝对百分误差 (mean absolute percentage error, MAPE) 来对 IBOA - ELM、AdaBoost - BOA - ELM 和 AdaBoost - IBOA - ELM 这 3 种模型的故障分类性能进行比较, RMSE 反映离散程度, MAE 准确反映实际误差的大小, MAPE 衡量准确性, 其中表格中所有数据都是取 20 次结果的平均值, 以此降低误差带来的影响, 四种指标的比较如表 1 所示。

表 1 3 种模型对比结果

指标	模型		
	AdaBoost - IBOA - ELM	AdaBoost - BOA - ELM	IBOA - ELM
分类准确率/%	98.527 7	96.083 3	94.388 9
均方根误差	0.017 4	0.065 9	0.139 1
平均绝对误差	0.072 1	0.120 9	0.225 0
平均绝对百分误差	1.64	4.66	8.32

从表 1 中可以看出, 3 种模型中, AdaBoost - IBOA - ELM 的分类准确率明显高于另外 2 种模型, AdaBoost - IBOA - ELM 模型的 RMSE、MAE 和 MAPE 这 3 种指标的数据也是最小的, 明显优于其他 2 种模型, 表现出了更好的分类效果。

最后将所提算法与另外 7 种算法在训练准确率、测试准确率、训练时间和测试时间进行比较, 结果如表 2 所示。

表 2 不同网络故障诊断准确率对比

网络	训练准确率/%	测试准确率/%	训练时间/s	测试时间/s
BP	89.24	88.45	5.06	3.68
SVM	86.27	84.94	8.66	3.31
LSTM	94.33	93.59	11.31	2.67
IBOA - CNN	96.89	95.36	25.14	5.14
PSO - ELM	94.32	93.11	15.96	4.18
BOA - ELM	93.64	92.39	13.30	3.42
IBOA - ELM	95.56	94.33	18.43	2.71
AdaBoost - IBOA - ELM	99.14	98.53	21.84	3.08

从表 2 中可以看出, 与其他 7 种网络相比较, AdaBoost - IBOA - ELM 在训练准确率和测试准确率上都是最高的, 同时也可以发现, 虽然通过优化算法改进网络可以提高分类准确率, 但是达不到想要的结果, 在加入 AdaBoost 算法后, 准确率得到了明显的提升, 进一步验证了所提方法对 TE 过程数据集分类的优势。

5 结束语

为了将 TE 过程中的故障准确分开, 该文提出一种 KPCA 和 AdaBoost - IBOA - ELM 相结合的 TE 化工过程故障诊断方法。首先, 利用 KPCA 处理原始数据, 一方面, 使原始数据的维数减少, 从而减少计算量; 另一方面, 提取各类故障的主要特征, 为分类模型的建立奠定基础。然后利用降维后待测数据的 T^2 统计量和 SPE 统计量进行故障识别。其次在用 IBOA 优化 ELM 的权值和阈值时, 改进的蝴蝶算法很好地平衡了全局搜索和局部搜索能力, 提高了 ELM 的训练速度和测试精度, 最后对 IBOA - ELM 分类器进行 AdaBoost 集成, 通过 TE 数据集进行验证, 相比优化过的 BP、PNN 和 SVM, 所提方法分类准确率最高, 说明了在浅层网络中优势明显, 与 AdaBoost - BOA - ELM 和 IBOA - ELM 的对比说明了 IBOA 和 AdaBoost 算法对 ELM 的优化是有效的。最后与 BP、SVM、LSTM、IBOA - CNN、PSO - ELM、BOA - ELM、IBOA - ELM 等网络的对比结果也显示出了在训练集和测试集上更高的分类准确率。然而, 该文也存在一些局限性, 比如如何提高 KPCA 提取特征时充分利用主成分空间中的信息。这一点在一定程度上也会影响最终的分类结果。

参考文献:

- [1] WANG N, LI H, WU F, et al. Fault diagnosis of complex chemical processes using feature fusion of a convolutional network [J]. Industrial And Engineering Chemistry

- Research, 2021, 60(5): 2232–2248.
- [2] 顾幸生, 周冰倩. 基于 LNS – DEWKECA 算法的多模态工业过程故障检测 [J]. 控制与决策, 2020, 35(8): 90–97.
- [3] 江升, 旷天亮, 李秀喜. 基于稀疏过滤特征学习的化工过程故障检测方法 [J]. 化工学报, 2019, 70(12): 4698–4709.
- [4] 蓝艇, 朱莹, 俞海珍, 等. 基于缺失数据的误差生成策略及其在故障检测中的应用 [J]. 控制与决策, 2020, 35(2): 143–149.
- [5] 索寒生, 蒋白桦, 宫向阳, 等. 基于 SFA 的工业过程质量相关的在线故障检测 [J]. 控制工程, 2019, 26(6): 1222–1227.
- [6] 吕鹏飞, 闫云聚, 荔越. 基于马氏距离的改进核 Fisher 化工故障诊断研究 [J]. 自动化学报, 2020, 46(11): 143–155.
- [7] 张成, 高宪文, 李元. 基于 k 近邻主元得分差分的故障检测策略 [J]. 自动化学报, 2020, 46(10): 229–238.
- [8] A J M L, B C K Y, A S W C, et al. Nonlinear process monitoring using kernel principal component analysis [J]. Chemical Engineering Science, 2004, 59(1): 223–234.
- [9] HUANG G B, ZHU Q Y, SIEW C K. Extreme learning machine: a new learning scheme of feedforward neural networks. In: IEEE International Joint Conference on Neural Networks, 2004. Proceedings. Budapest: IEEE, 2004, 985–990.
- [10] 崔鹏宇, 王泽勇, 邱春蓉, 等. 基于多尺度排列熵与双核极限学习机的滚动轴承故障诊断方法 [J]. 电子测量与仪器学报, 2019, 33(05): 142–147.
- [11] YU H, YUAN K, LI W, et al. Improved butterfly optimizer – configured extreme learning machine for fault diagnosis [J]. Complexity, 2021, 2021(4): 1–17.
- [12] ARORA S, SINGH S. Butterfly optimization algorithm: a novel approach for global optimization [J]. Soft Computing, 2019, 23(3): 715–734.
- [13] 王晓东, 盛庆博, 孙立群, 等. 基于 AdaBoost 算法的光伏电站中长期发电预测 [J]. 工业仪表与自动化装置, 2023(02): 65–69.
- [14] DOWNS J J, VOGEL E F. A plant – wide industrial process control problem [J]. Computers and Chemical Engineering, 1993, 17(3): 245–255.
- [15] 陈晓, 王瑞璇, 王晶, 等. 基于混合型判别分析的工业过程监控及故障诊断 [J]. 自动化学报, 2020, 46(8): 64–78.
- [16] 毛志勇, 黄春娟, 路世昌, 等. 基于 KPCA – MPSO – ELM 的矿井突水水源判别模型 [J]. 中国安全科学学报, 2018, 28(8): 111–116.
- [17] SENTHIL T, DEEPIKA J, NITHYA R. Detection and classification of breast cancer using improved grey wolf algorithm [J]. IOP Conference Series Materials Science and Engineering, 2021, 1084(1): 012023.
- [18] 刘洪基. 基于混沌 PSO 的大数据智能加权 K 均值聚类算法 [J]. 计算机应用与软件, 2022, 39(04): 311–319.
- [19] 高文欣, 刘升, 肖子雅, 等. 柯西变异和自应权重优化的蝴蝶算法 [J]. 计算机工程与应用, 2020, 56(15): 43–50.
- [20] 杨如民, 雷倩, 余成波. 基于蝙蝠算法 – 随机森林分类算法和脉搏波的精神疲劳检测算法 [J]. 科学技术与工程, 2022, 22(04): 1495–1501.
- [21] 李辉, 李欣同, 贾嵘, 等. 基于分形和概率神经网络的水电机组故障诊断 [J]. 水力发电学报, 2019, 38(3): 92–100.
- [22] ZHANG N, LI Y, YANG X, et al. Bearing fault diagnosis based on BP neural network and transfer learning [J]. Journal of Physics: Conference Series, 2021, 1881(2): 022084.
- [23] 杨泰春, 陶建峰, 覃程锦, 等. 采用支持向量机的非对称阀控液压缸模型预测控制 [J]. 西安交通大学学报, 2020, 54(1): 93–100.

(上接第 101 页)

- [7] 周琳琳, 刘庆胜. LNG 储罐压力控制系统设计[J]. 石油工程建设, 2016, 40(5): 39-41.
- [8] 贾保印, 纪明磊, 刘禹岑. LNG 储罐安全阀选型探讨[J]. 山东化工, 2017, 5(46): 117-119.
- [9] BS EN 1473-2016: Installation and equipment for liquefied natural gas - design of onshore installations[S]. Brussels. 2016.
- [10] ISO 28300-2008: Petroleum, petrochemical and natural gas storage tanks[S]. Switzerland. 2008.
- [11] 全国石油天然气标准化技术委员会. 液化天然气设备与安装陆上装置设计: GB/T 22724-2022[S]. 2022.
- [12] 全国石油天然气标准化技术委员会. 液化天然气(LNG)生产、储存和装运: GB/T 20368-2021[S]. 2021.