doi:10.3772/j.issn.1002-0470.2022.04.006

基于 MPA 优化 MKL-FSVDD 模型的聚合釜设备故障诊断^①

李国友②* 才士文③* 李东朔** 张新魁* 贾曜宇* 宁 泽*

(*燕山大学智能控制系统与智能装备教育部工程研究中心 秦皇岛 066004)(**电子科技大学光电科学与工程学院 成都 611731)

摘 要 针对化工流程工业数据具有强非线性、易受噪声影响和故障为多分类的问题,提 出一种基于海洋捕食者算法(MPA)优化多核学习-模糊支持向量机数据描述(MKL-FS-VDD)的故障诊断方法。利用 MKL 构建的多核函数,弥补单核函数的局限性,对非线性故 障数据分类具有较强的适应性;引入 MPA 对 MKL-FSVDD 模型的核参数进行高效寻优, 解决核参数选择难题。通过在 TE 数据平台上的对照实验,验证 MPA-MKL-FSVDD 模型 故障诊断的有效性能;最后将故障诊断模型应用于聚氯乙烯(PVC)聚合反应中,利用70 m³ 的聚合釜设备历史数据集进行仿真验证。结果表明该方法充分利用复杂样本集的数据信 息,并在参数寻优阶段快速、稳定获得最优解,保证了故障分类的效率和准确度。 关键词 故障诊断;海洋捕食者算法(MPA);多核学习(MKL);模糊隶属度;聚合釜

0 引言

化工生产流程通常为非流水型、不透明的状态, 聚氯乙烯(polyvinyl chloride, PVC)生产过程是一个 传质、传热的复杂化学工艺过程,其中聚合釜故障是 导致聚氯乙烯产品质量下降的主要原因之一。工业 生产过程中聚合釜设备会产生大量高维数据,工作 人员通过数据进行间接观察^[1-2],不仅加大了故障 诊断难度,而且诊断效果并不理想。相比简单系统 的解析模型,利用数据驱动模型,解决化工流程工艺 中复杂故障数据特征不易区分问题的优势显而易 见^[35]。

国内外针对聚合釜设备的故障诊断研究,早期 是使用神经网络的方法^[6]。为满足 PVC 生产过程中 聚合釜故障实时诊断和优化监测的要求,文献[7]提 出惯性权值动态调整粒子群优化算法优化自组织映 射(self-organizing map, SOM)神经网络结构参数的 故障诊断策略,结合聚合釜的工业现场历史数据的 仿真,证明其有效性。针对复杂化工生产过程中的 检测异常和故障诊断,文献[8]提出过程拓扑卷积 网络模型,减小对训练数据量和计算资源的依赖,提 高了故障诊断的准确率。但由于神经网络模型训练 过程需要大规模样本数据集支撑的要求,其应用场 合受到很多局限[9],相关学者相继引入支持向量机 (support vector machine, SVM)^[10]和支持向量机数 据描述(support vector data description, SVDD)^[11-12] 的机器学习算法,并以此为基础不断完善发展。我 国故障诊断研究起步晚,但发展比较快,故障诊断技 术在化工流程工业领域的研究与应用都取得了一定 的成效。文献[13]采用改进布谷鸟算法优化支持 向量机,进行聚合釜故障诊断的仿真研究,提高了 PVC 聚合过程中聚合釜故障的诊断率和预测率。 文献[14]针对聚合釜聚合过程故障特点和生产过 程的非线性、动态性特征,提出基于动态核主元分析 (dynamic kernel principal component analysis, DKP-CA)的故障诊断算法,以利于实际 PVC 聚合生产过

① 国家自然科学基金(F2012203111)和河北省高等学校科学技术研究青年基金(2011139)资助项目。

② 男,1972年生,博士,教授;研究方向:工业控制,故障诊断,机器视觉,智能控制;E-mail: lgyysu@163.com。

③ 通信作者, E-mail: 1669181224@qq.com。 (收稿日期:2021-05-08)

程的故障监测。

由于传统经典算法默认为输入的每个训练样本 以同等机会属于其所在类,存在忽略实际工艺流程 过程中不同训练样本对于其所属类的置信程度不 同[15-16],这使得模型出现在噪声干扰下分类效果不 佳的问题。文献[17]结合模糊理论思想,在SVDD 的基础上,为每个样本确立一个隶属度,提出模糊支 持向量数据描述(fuzzy support vector data description, FSVDD)算法,算法能够描述数据样本的重要 程度,处理噪声敏感问题。但是 FSVDD 依然存在 SVDD 在模型训练过程中核函数核参数的选择难 题^[18]。文献[19]利用 banana 数据集仿真验证了多 核分类器的数据描述能力明显优于任何一个单核分 类器,能够获得比较理想的分类精度。考虑分类器 训练准确率受其给定参数值的影响较大的问题,通 过元启发式参数寻优算法对分类器参数进行优化, 以有效提高故障诊断的效率和准确率。文献[20] 于2020年开发了一种元启发式海洋捕食者算法 (marine predators algorithm, MPA),在工程基准测试 中,通过对比实验证明了 MPA 相比于其他元启发式 算法更好的寻优性能,且计算成本较低。文献[21] 将 MPA 与 SVM 结合对滚动轴承故障进行类型识 别,可以得到较好的诊断效果同时避免模型陷入局 部最优,表明了寻优算法对分类器参数优化的良好 应用。

综上,本文构建了一种基于海洋捕食者算法优 化多核学习(multiple kernel learning, MKL)-模糊支 持向量机数据描述 MPA-MKL-FSVDD 故障数据分 类模型,可有效应用于 PVC 聚合反应聚合釜设备的 工业数据的故障诊断,减小 PVC 聚合釜设备故障诊 断难度,并实现保证聚合反应安全生产质量的目的。

本文主要进行了以下工作。

(1)为减小野点对分类边界影响、避免过拟合 情况,运用基于距离的模糊隶属度函数确定训练样 本的模糊隶属度,构造 FSVDD 模型。

(2)为充分利用复杂样本数据集的特征信息, 采用线性加权方式,构造多核学习函数,通过核权重 系数计算获得组合核,弥补单核函数描述数据信息 特征的局限性。 (3)针对故障诊断模型核参数选择问题,采用 MPA 智能搜索算法明确寻找模型最优参数,提高模 型寻优精度和故障诊断效率,同时避免陷入局部最 优。

(4)利用田纳西-伊斯曼(TE)数据集和 PVC 聚 合反应聚合釜设备工业历史数据集进行故障诊断的 实验仿真,对比故障诊断效果,以验证 MPA-MKL-FSVDD 分类模型的故障诊断性能。

1 MPA-MKL-FSVDD 算法介绍

1.1 FSVDD 算法基本原理

FSVDD 模型继承了 SVDD 模型的无需考虑数 据是否符合高斯分布、计算复杂度低、约束条件简 单、鲁棒性强、泛化能力良好等优点^[22],在标准 SVDD 模型的基础上,通过引入模糊系数,以隶属度 的形式报告每个样本的值,对异常点赋予较低的隶 属度值,减小野点对模型过拟合的影响,解决了 SVDD 对噪声数据敏感问题和大规模数据分类处理 能力欠佳的问题,提高模型的鲁棒性^[23]。

设目标数据 $X_{Nxdim} = \{x_1, x_2, \dots, x_N\}, N$ 为样本个数, dim 为样本维度。FSVDD 将目标数据 T 封装在一个半径为 R、中心为 a 的超球面中。对样本形成的特征向量为

$$\boldsymbol{\Phi}(x) = \{ \boldsymbol{\Phi}(x_1), \boldsymbol{\Phi}(x_2), \cdots, \boldsymbol{\Phi}(x_N) \}$$
(1)

加入模糊系数 s_i 的支持向量数据描述的优化问题可表示为

$$F(R, a, s_i, \xi_i) = R^2 + C \sum_{i=1}^{N} s_i \xi_i$$
(2)

约束条件为

 $\| \boldsymbol{\Phi}(x_{i}) - a \|^{2} \leq R^{2} + \xi_{i}, \xi_{i} \geq 0$ (3) 其中, *i* = 1,2,...,*N*, *N*为样本个数;松弛因子 $\xi_{i} \geq$ 0; *s_i*($\sigma < s_{i} < 1, \sigma > 0$), *s_i*在一定程度上减小了 ξ_{i} 影响,代人拉格朗日因子 α_{i}, β_{i} , 对构成的拉格朗日 函数求极值:

$$L(R, a, \xi_i, \alpha_i, \beta_i) = R^2 + C \sum_{i=1}^N s_i \xi_i - \sum_{i=1}^N \beta_i \xi_i - \sum_{i=1}^N \alpha_i (R^2 + s_i \xi_i - \| \boldsymbol{\Phi}(x_i) - a \|^2)$$
(4)

— 380 —

其中, $\alpha_i \ge 0$, $\beta_i \ge 0$ 。 对 $R_n \alpha_n \xi_i$ 偏微分求导,有: $\begin{cases} \sum_{i=1}^N \alpha_i = 1, \ 0 \le \alpha_i \le Cs_i \\ a = \sum_{i=1}^N \alpha_i \Phi(x_i) \\ C - \alpha_i - \beta_i = 0 \end{cases}$ (5)

对式(4)、式(5)优化,设定*i* = 1, 2,…,*N*; *j* = 1,2,…,*N*,将超球体问题转为对偶形式,即:

$$\max W = \sum_{i=1}^{N} \alpha_i (\boldsymbol{\Phi}(x_i), \boldsymbol{\Phi}(x_i))$$

-
$$\sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j \langle \boldsymbol{\Phi}(x_i), \boldsymbol{\Phi}(x_j) \rangle$$
(6)

根据 Mercer 定理:

$$K = K(x_i, x_j) = \langle \boldsymbol{\Phi}(x_i), \boldsymbol{\Phi}(x_j) \rangle$$
(7)
将高维特征空间内积运算转化为核函数运算:

$$\max W = \sum_{i=1}^{N} \alpha_i K(x_i, x_i) - \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j K \quad (8)$$

$$\Leftrightarrow \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j K = \boldsymbol{\alpha}^{\mathrm{T}} K \tilde{\boldsymbol{\alpha}}, \text{ s. t. } \boldsymbol{\alpha}^{\mathrm{T}} \mathbf{1} = 1; \boldsymbol{\alpha} \ge \mathbf{0}$$
(9)

根据 KKT 条件, 对所有满足 $0 \le \alpha_i \le C$ 的支持 向量样本 x_{ss} , 可得超球体半径为

$$R^{2} = \boldsymbol{K}(x_{sv}, x_{sv}) - 2 \sum_{i=1}^{N} \alpha_{i} \boldsymbol{K}(x_{i}, x_{sv}) + \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{K} \tilde{\boldsymbol{\alpha}}$$
(10)

测试样本 z 在特征空间中离球心 a 的距离为

$$f(z) = \| \boldsymbol{\Phi}(z) - a \|^{2}$$
$$= \boldsymbol{K}(z, z) - 2 \sum_{i=1}^{N} \alpha_{i} \boldsymbol{K}(x, z) + \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{K} \tilde{\boldsymbol{\alpha}}$$
(11)

其中,

$$\begin{cases} f(z) \leq R^2 & z \in W_{\text{target}} \\ f(z) \geq R^2 & z \in W_{\text{outlier}} \end{cases}$$
(12)

由式(11)和式(12)可知,通过比较样本点到球 心 a 的距离与超球体半径 R 的大小关系,可以判断 样本点是否属于目标类。因此,FSVDD 模型选用基 于距离的模糊隶属度函数计算模糊隶属度值,来描 述样本点属于目标类别的可能性。

确定样本与训练过的超球面中心的距离 d_i^{svdd},

即:

$$d_i^{\text{svdd}^2} = \boldsymbol{\Phi}(x_i) - a^2$$
$$= \boldsymbol{K}(x_i, x_i) - 2\sum_{j=1}^N \alpha_j \boldsymbol{K} + \alpha^{\mathrm{T}} \boldsymbol{K} \tilde{\boldsymbol{\alpha}} \qquad (13)$$

通过选定的隶属度函数给每个样本赋予权重 量,降低噪声或离群点对最优决策球面的影响。本 文采用降半高斯函数作为隶属度函数^[24],计算样本 对应隶属度为

$$s_{i}(d_{i}^{\text{svdd}}) = \begin{cases} 1 & d_{i}^{\text{svdd}} \leq R^{*} \\ \exp\left[-\frac{(d_{i}^{\text{svdd}} - R^{*})^{2}}{2\varphi^{2}}\right] & d_{i}^{\text{svdd}} > R^{*} \end{cases}$$
(14)

其中 R^* 为类半径值,设定带宽 $\varphi = 0.2R^*$;隶属度 越大,表示样本点距离中心的距离越近,即属于目标 样本的可能性越大,反之,属于野点的可能性越大。

1.2 多核学习算法

针对化工生产过程中工业数据具有的复杂非线 性和动态性特点,故障诊断算法对数据描述能力的 强弱,直接影响故障诊断效果。因核函数的选择与 数据信息特征密切相关,故选择合适的核函数核参 数,实现算法对数据的描述能力的增强,对保障分类 器性能具有重要意义^[25]。

对 FSVDD 模型利用多核学习算法,通过调整权 重将各特征的最优核函数组合^[26],将核选择问题转 化为选择组合系数的算法^[27]。这样既可以针对非 线性关系的数据选择合适的单一核函数,又可以针 对既存在线性关系又存在非线性关系的复杂数据集 选用多个基本核的组合代替单核,具有更强的适应 性。

鉴于多项式核函数良好的全局性能和高斯核函 数较强的局部性能,二者结合用以构建混合核函数 分类模型。

多核矩阵 **K** = **K**_{ij}, *i* = 1,2,…, *N*, *j* = 1,2,…, *N*; 核矩阵中的元素 **K**_{ij} 为

$$\mathbf{K}_{ij} = \mathbf{K}_{\text{muliti}}(x_i, x_j)$$
$$= m\mathbf{K}_{\text{ploy}}(x_i, x_j) + (1 - m)\mathbf{K}_{\text{gauss}}(x_i, x_j)$$
(15)

其中,

— 381 —

$$\begin{cases} \mathbf{K}_{\text{ploy}}(x_{i}, x_{j}) = [(x_{i}, x_{j}) + 1]^{p} \\ \mathbf{K}_{\text{gauss}}(x_{i}, x_{j}) = \exp\left(-\frac{\|x_{i} - x_{j}\|^{2}}{2\sigma^{2}}\right) \end{cases}$$
(16)

其中, m 为权值, p 为多项式次幂数, σ 为高斯核参数, 组合的多核函数由权值矩阵调节。

1.3 海洋捕食者算法基本原理

为进一步提高分类准确率、提高分类效率,多核 学习优化的 FSVDD 可以进一步利用寻优算法寻找 核函数的最优参数组合,提高模型训练速度;并且可 以通过对数据的不断学习,修正权重系数保证诊断 精度^[26]。2020 年 Faramarzi 等人^[20]开发的一种自 然启发式优化算法——海洋捕食者算法(MPA)遵 循在最佳觅食策略中自然支配的规则,Le'vy 和布朗 运动作为海洋捕食者选择的两种最佳觅食策略,同 时具备全局和局部搜索的能力,拓宽了 MPA 算法解 决优化问题的应用范围^[28-30]。

MPA 的算法步骤如下。

步骤1 初始化。

MPA 种群的随机初始化如下:

$$X_i = X_i^{\min} + (X_i^{\max} - X_i^{\min}) \times rand$$
(17)

其中, X_i^{max} 和 X_i^{min} 分别为控制变量的最大值和最小值边界, rand 是随机数 ($0 \le rand \le 1$), 初始总体的目标函数表示式如下:

$$F_i = obj(X_i) \tag{18}$$

步骤2 探索顶级捕食者。

首先,最适的解被指定为顶级捕食者,构建 Elite 矩阵,矩阵的数组根据猎物的位置信息监视猎 物的搜寻和发现:

$$Elite = \begin{bmatrix} X_{1,1}^{l} & X_{1,2}^{l} & \cdots & X_{1,d}^{l} \\ X_{2,1}^{l} & X_{2,2}^{l} & \cdots & X_{2,d}^{l} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n,1}^{l} & X_{n,2}^{l} & \cdots & X_{n,d}^{l} \end{bmatrix}_{n \times d}$$
(19)

其中, \vec{X} 代表顶层捕食者向量,将其复制 n 次,构造 Elite 矩阵, n 为搜索 agents 的数目, d 为维数。在每 次迭代结束时,如果顶级捕食者被更好的捕食者取 代, Elite 将被更新。

其次,种群排列在 Prey 猎物矩阵的矩阵中,捕 食者根据 Prey 矩阵更新自己的位置,最适者(捕食 者)构建 Elite 矩阵。

$$Prey = \begin{bmatrix} X_{1,1} & X_{1,2} & \cdots & X_{1,d} \\ X_{2,1} & X_{2,2} & \cdots & X_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n,1} & X_{n,1} & \cdots & X_{n,d} \end{bmatrix}_{n \times d}$$
(20)

其中, *X_{i,j}* 代表第*i* 个猎物的第*j* 位, 整个优化过程 主要与 *Elite*、*Prey* 两个矩阵直接相关。

步骤3 Le'vy飞行运动和布朗运动。

猎物和捕食者的位置根据猎物和捕食者的速度 比分为3个阶段进行更新。

阶段1为算法的探索阶段,采用高速比。猎物 和捕食者根据布朗运动进行更新,布朗运动用数学 表示为

当
$$T < \frac{1}{3}T_{\max}$$
时,有:
 $\overline{SZ_i} = \overline{R_{Br}} \otimes (\overline{Elite_i} - \overline{R_{Br}} \otimes \overline{Prey_i}), \quad (21)$
 $i = 1, \cdots, n$

 $\overline{Prey_i} = \overline{Prey_i} + P \cdot \overline{R} \otimes \overline{SZ_i}$

式中, $T 和 T_{max}$ 分别为当前迭代次数和最大迭代次数; $\overline{R_{Br}}$ 是一个基于正态分布表示布朗运动的包含随机数的矢量; $\overline{SZ_i}$ 是步长向量, P 是等于 0.5 的常数。

阶段 2 为间歇阶段,当捕食者的速度等于猎物的速度时,种群被分为两组。

第一组用于开发,另一组用于勘探,可以表示为 如下所述。

(1) 开发组
当
$$\frac{1}{3}T_{\max} \leq T \leq \frac{2}{3}T_{\max}$$
时,有:
 $\overline{SZ_i} = \overline{R_{Levy}} \otimes (\overline{Elite_i} - \overline{R_{Levy}} \otimes \overline{Prey_i}),$
 $i = 1, \dots, \frac{n}{2}$ (22)

$$Prey_{i} = Prey_{i} + P \cdot \overline{R} \otimes SZ_{i}$$
(2) 勘探组
$$\overline{SZ_{i}} = \overline{R_{Br}} \otimes (\overline{R_{Br}} \otimes \overline{Elite_{i}} - \overline{Prey_{i}}),$$

$$i = \frac{n}{2}, \cdots, n \qquad (23)$$

 $\overline{Prey_i} = \overline{Elite_i} + P \cdot CF \otimes \overline{SZ_i}$

R_{Levy} 是一个包含随机数的矢量, 基于 Le'vy 分布, 以 Le'vy 的方式模拟猎物的运动。CF 是一种用

— 382 —

于控制捕食者移动步长的自适应算子。

$$CF = \left(1 - \frac{T}{T_{\max}}\right)^{\left(2\frac{T}{T_{\max}}\right)}$$
(24)

阶段 3 为开发阶段,优化过程的最后迭代中,在 低速比(*v*=0.1)下,捕食者的最佳策略为 Le'vy。在 低速比或当捕食者比猎物移动得快时应用此阶段:

当
$$T > \frac{2}{3}T_{max}$$
时,有:
 $\overline{SZ_i} = \overline{R_{Levy}} \otimes (\overline{R_{Levy}} \otimes \overline{Elite_i} - \overline{Prey_i}), \quad (25)$
 $i = 1, \dots, n$

 $\overline{Prey_i} = \overline{Elite_i} + P \cdot CF \otimes \overline{SZ_i}$

 $R_{Levy} \otimes \overline{Elite_i}$ 在 *Elite* 位置中增加步长模拟捕食者的移动,帮助更新猎物的位置。

步骤4 Eddy 形成和 FADs 效应。

$$\overline{Prey_i} = \begin{cases} \overline{Prey_i} + CF[\overline{X_{\min}} + \overline{R} \otimes (\overline{X_{\max}} - \overline{X_{\min}})] \otimes \overline{U} \\ r \leq FADs \\ \overline{Prey_i} + [FADs(1 - r) + r](\overline{Prey_{r1}} - \overline{Prey_{r2}}) \\ r > FADs \end{cases}$$

(26)

其中, r 是 0 ~ 1 范围内的随机值; r1 和 r2 代表猎物 矩阵的随机指标; FADs 表示 FADs 概率, 等于 0.2; Ū 是一个二进制向量。

步骤5 海洋记忆。每更新捕食者位置和完成 FADs 效应后,计算相应适应度,与前一次迭代中的 解进行比较,选择最优适应度位置作为顶级捕食者 位置。

步骤6 重复步骤3~步骤5,直到满足终止条件,输出顶级捕食者位置坐标,以获得最优解。

海洋捕食者算法探索流程如图1所示。

2 MPA-MKL-FSVDD 集成算法

MPA-MKL-FSVDD 集成算法采用离线训练、在 线测试的工作模式。如图2所示,具体步骤如下。

(1) 离线建模阶段

步骤1 选取正常工况数据集和故障情况数据 集组成训练集样本。

步骤2 对数据进行标准化处理,FSVDD 引入 MKL确定核函数,采用 MPA 对FSVDD 的多核函数





图 2 基于数据的故障分类模型流程图

寻优核参数,构建训练过程的故障诊断 MPA-MKL-FSVDD 模型。

步骤3数据样本在模型中迭代更新参数,达 到最佳适应度终止,获得判断是否为目标类样本的 model 参数。

(2)在线测试阶段

步骤1 选取正常工况数据集和故障情况数据 集组成测试集样本。

步骤2 对数据进行标准化,根据离线建模阶段步骤3所得 model 参数赋值测试模型,同时计算测试样本在超球半径空间的样本距离 *f*(*z*)。

步骤3 将*f*(*z*)与最优超球体半径*R**相比较, 对数据集进行故障诊断,判断是否为目标类样本。

3 MPA-MKL-FSVDD 算法验证

实验仿真在仿真软件上进行,硬件配置处理器 为 Intel(R) Core(TM) i5-5200H CPU@2.20 GHz, Win10 64 位计算机。

3.1 田纳西-伊斯曼平台数据集选取

为验证本文所提算法性能,对 MPA-MKL-FS-VDD 故障诊断模型进行仿真实验,并将结果与 MKL-FSVDD、FSVDD 和 SVDD 模型进行比较。选用 田纳西-伊斯曼(TE)仿真平台过程数据集,该数据 集中训练样本和测试样本分别包含 20 种故障工况 和1种正常工况数据,每个样本信号包含41 个测量 变量和11 个控制变量,在化工过程的故障诊断领 域,是一个相对认可的公开数据集资源。与 TE 过 程中的其他故障相比,反应器单元中故障7(反应器 压力)相对正常状态的可分性较低^[31],故障9(反应 器温度)对测量和操作变量的影响很小^[32],难以进 行检测分类。鉴于需要对FSVDD 对噪声处理能力 和 MKL 描述数据特征效果进行验证,本案例研究选 择了故障7 和故障9 测试集数据,两故障分别除去样 本在8h前的160个正常数据,余下800组故障样本 构成故障数据集,定义故障7 为标签+1、故障9 为标 签-1。

3.2 FSVDD 噪声数据的处理

FSVDD 模型利用模糊系数 s_i 对训练样本赋予 一个权重,降低噪声或离群点对最优决策球面的影 响。FSVDD 模型训练建立故障诊断检测分类模型, 需要目标类样本建立样本库^[31],示例将选取故障 7 和故障 9 两类样本各 200,样本标签 +1 作为目标样 本,借助点到中心的距离 d_i^{svdd} 计算样本模糊隶属 度。样本对应模糊隶属度越大,表示样本点为目标 样本的可能性越大,而对隶属度非常小的样本视为 野点。

将 FSVDD 对目标样本的决策边界曲线可视化 处理,如图 3(a)所示,其中菱形 \diamond 为在决策边界曲 线上的支持向量;星形 \diamond 为避免过拟合而忽略的支 持向量;圆点 \circ 为训练样本中判断为目标类的样本; 方块 \Box 为非目标类样本。根据基于距离的模糊隶属 度函数得到的模糊系数,反映出的是隶属度值大小, 由式(14)可知,当 $d_i^{svdd} \leq R^*$ 时该样本点属于目标 样本;当 $d_i^{svdd} \geq 1.4375R^*$ 时,此时隶属度小于0.05,



图 3 FSVDD 模型分类可视化图

将此类样本点视作野点剔除,对分类面的影响几乎 可以忽略不计,但是可以提高分类速度。可视化处 理将隶属度值映射成等高线式图,如图 3(b)所示, 以形成的目标类超球体球心(标值 0.8 等高线包围 的中心位置)为球心,自球心向外周,隶属度值逐渐 减小,标值 1.15 等高线圈外的区域表示隶属度小于 0.05 的样本所在的区域,即为避免模型过拟合而忽 略的野点落于此范围。综上,由 FSVDD 模型分类可 视化图对目标类样本决策边界的分类拟合效果,可 知 FSVDD 模型在避免模型过拟合、有效处理噪声数 据方面的可行性。

3.3 FSVDD 决策边界的 MKL 优化

MKL-FSVDD 模型训练过程寻求一种使外围球 在核空间中半径最小的核组合。为验证模型应用性 能,本示例随机选取数据集中两个属性生成一个 2D 的目标类,对算法的分类性能进行了可视化展示。 FSVDD 模型在加入多核学习函数后,随着核权重在 多项式核和高斯核间的比重不断修正,决策边界表 现出寻优过程的训练效果如图 4 所示。由 MKL-FS-VDD 边界支持向量的决策曲线可视化效果图可以 观察出,MKL 的引入使得模型对目标类形成拟合比 较精准的决策边界,相对于单核 FSVDD 模型,可以 获得更好的分类效果,验证了 MKL-FSVDD 模型数 据分类的有效性。



图 4 MKL-FSVDD 模型边界决策曲线修正过程可视化图

3.4 MPA-MKL-FSVDD 实验及分析

为验证 MPA-MKL-FSVDD 故障诊断的能力,利用故障7 和故障9 数据样本,通过与 MKL-FSVDD、

单核 FSVDD、SVDD 模型的对照实验,验证 MPA-MKL-FSVDD 模型故障诊断的效率、准确性以及抗 噪能力。

MKL-FSVDD 模型训练中多核函数包含高斯径 向基核和多项式两类核函数,程序调试中对参数进 行初始设置^[17]:高斯径向基核带宽 g 取值范围为 [0.01,100];多项式核幂指数 p 选值范围为[1,2, 3,4];惩罚因子 c 取值范围为[0.01,200]。设最大 迭代次数为100。

数据预处理后按7:3比例分为训练样本和测试 样本,取10次实验的平均值作为结果(见表1)。对 比可知:

(1) FSVDD 和 SVDD 在数据集的分类精度上 差异并不显著,但 FSVDD 相对于 SVDD 的训练时间 和分类时间有所降低。

(2) MKL-FSVDD 相对于单核 FSVDD 训练时间 增加了 25.21% 左右,但是故障诊断精度明显提高。

(3)因多核学习过程中寻找最优核参数权值组 合的过程耗时较长,针对此问题引入的 MPA 对模型 参数进行全局优化, MPA-MKL-FSVDD 相对于单核 FSVDD 训练时间仅增加了 12.09% 左右,改进了故 障诊断效率。

另一方面,鉴于在实际的工程应用中模型训练 是前置的,在使用时采用的是训练完好的模型,算法 训练时间的增加不会影响生产线效率^[33]。所以综 合表1中各模型故障诊断分类性能参量的对比分 析,可以验证 MPA-MKL-FSVDD 算法在故障诊断准 确率和效率方面的优势性能。

4 MPA-MKL-FSVDD 故障分类模型应用

4.1 PVC 聚合反应工艺流程分析

PVC 生产过程是一个传质、传热的复杂化学反 应工业过程,监测变量存在种类多、纯滞后、强耦合、 参数广等特性。反应过程的安全性是决定产品质量 的关键,其中聚合釜故障是导致聚氯乙烯产品质量 下降的主要原因之一^[34],本文以某化工集团公司 PVC 生产过程聚合反应中70 m³ 的聚合釜为例,图5 所示为该化工集团PVC聚合釜设备流程示意图。

| ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ | | | | | | | | | | |
|---------------------------------------|--------|--------|-----------|---------------|--|--|--|--|--|--|
| 数据集 | 训练时间/s | | | | | | | | | |
| 样本数 | SVDD | FSVDD | MKL-FSVDD | MPA-MKL-FSVDD | | | | | | |
| 400 | 2.814 | 2.783 | 3.476 | 2.967 | | | | | | |
| 1000 | 9.592 | 9.545 | 11.585 | 10.833 | | | | | | |
| 1600 | 21.316 | 21.147 | 27.359 | 24.572 | | | | | | |
| | | | | | | | | | | |
| 数据集 | 训练时间/s | | | | | | | | | |
| 样本数 | SVDD | FSVDD | MKL-FSVDD | MPA-MKL-FSVDD | | | | | | |
| 400 | 2.107 | 2.105 | 2.013 | 1.998 | | | | | | |
| 1000 | 6.937 | 6.721 | 6.546 | 6.532 | | | | | | |
| 1600 | 13.591 | 13.263 | 12.978 | 12.945 | | | | | | |
| | | | | | | | | | | |
| 数据集 | 准确率/% | | | | | | | | | |
| 样本数 | SVDD | FSVDD | MKL-FSVDD | MPA-MKL-FSVDD | | | | | | |
| 400 | 57.53 | 57.51 | 59.48 | 60.77 | | | | | | |
| 1000 | 64.78 | 64.77 | 69.19 | 70.26 | | | | | | |
| 1600 | 76.04 | 76.05 | 83.33 | 84.37 | | | | | | |









其中聚合釜搅拌系统在搅拌转速、搅拌桨叶形式及 内件设置等3个方面配置满足悬浮法生产聚氯乙烯 工艺的需要;聚合釜传热系统由釜体传热系统、内冷 挡板传热系统、体外冷凝器传热系统 3 部分组 成^[35]。

PVC 聚合过程中,将各种原料和添加剂加入反 应釜中,冷却水注入反应釜的夹套和挡板内,聚合釜 进行搅拌加热,当氯乙烯(VCM)的转化率达到要求 值并稳定后产生压降,压力降到设定值后反应终止, 得到最终产物。汽提工艺分离出剩余浆料中的氯乙 烯单体(vingl chloride monomer, VCM),余下浆料送 入干燥工艺^[36]。

使用聚合设备运行监控中的各种变量的历史数据,根据聚合釜相关参数如釜内温度、釜内压力、搅拌电流、搅拌速度、注入水流量、密封水流量、夹套水流量、档板水流量、冷却水进口温度、夹套水出口温度、档板水出口温度等相关变量^[7],可确定聚合釜设备聚合反应故障的类型。表2示例了聚合釜监控变量包含的部分参数。

表 2 聚合釜参数示例

| 参数 | 单位 | 参数范围 | 故障类型 |
|--------|--------------|-----------|------|
| 搅拌速度 | r∕min | 62.5 ~98 | 电机故障 |
| 搅拌电流 | А | 125 ~ 175 | 电机故障 |
| 釜温度系统 | $^{\circ}$ C | 45 ~65 | 中轴故障 |
| 机械密封压力 | MPa | 0.8~1.5 | 组件故障 |

4.2 故障数据集选取

为验证本文所提算法在实际工艺中的应用性 能,选用 PVC 聚合釜工业历史数据库存储的数据, 分类整理选出具有均匀性和代表性的 500 组数据作 为训练和测试样本集,其中包括聚氯乙烯聚合釜正 常聚合过程下的 100 组数据和典型故障情况下的 400 组故障数据样本。数据预处理后按 3:1 比例分 为训练样本和测试样本,输入故障分类模型进行验 证。

对应于聚合釜工业数据集定义训练过程的标签 类别如下:聚合釜正常工况数据样本定义标签为0, 聚合釜电机出现故障定义标签为1,聚合釜减速机 出现故障定义标签为2,聚合釜机封中轴故障定义 标签为3,聚合釜组件故障诊断定义标签为4。

4.3 优化 MKL-FSVDD 实验及分析

为验证 MPA-MKL-FSVDD 故障诊断分类模型 的性能,另选取哈里斯鹰算法(Harris hawks optimization, HHO)、鲸鱼算法(whale optimization algorithm, WOA)、粒子群算法(particle swarm optimization, PSO)3种元启发式算法,与本文优化的 MKL-FSVDD 故障分类模型结合并对其优化效果进行对 比分析。

实验参数同 3.4 节初始化设置,利用相同的数 据集、设定相同的迭代次数上限及模型参数的搜索 范围,搜索代理数量为 20。MKL-FSVDD 模型分别 与 MPA、HHO、WOA、PSO 算法结合,优化的效果由 其最佳适应度曲线和平均适应度变化曲线体现,其 中程序中适应度值的大小即为适应度函数值大小, 适应度值曲线呈上升趋势。

4种算法适应度寻优的过程如图 6、图 7、图 8 和图9的子图(a)所示。根据适应度曲线变化趋势 可知,优化过程中,MPA 算法的寻优速度最快,最佳 适应度在第3代时达到最佳,其次是 WOA 算法(第 10代),HHO 算法(第16代),最后是 PSO 算法(第 20代之后)。且 MPA 算法平均适应度在第 13代时 就完成收敛,达到最佳适应度值并趋于稳定状态,此 间段其他算法的平均适应度曲线仍处于波动状态, 收敛速度明显低于 MPA 算法,证明 MPA 算法具有 较强的搜索能力。且 HHO、WOA、PSO 优化算法寻 优过程中适应度曲线均存在明显的折线式跳跃状态 的收敛过程,其中 PSO 算法更是出现了多次层断式 阶跃状态,说明 PSO 算法搜索陷入局部收敛的概率 较大,易导致寻优效果的不稳定性。相较之下,MPA 算法适应度收敛曲线的收敛趋势近似圆滑曲线状 态,体现了 MPA 参数寻优过程中比较好的稳定性。

图 6、图 7、图 8 和图 9 的子图(b)所示为 MKL-FSVDD 模型利用 MPA、HHO、WOA、PSO 算法在目标空间内搜索,取得的最佳适应度得分情况,从图中可以看到,MPA 优化的 MKL-FSVDD 故障分类模型的最佳适应度得分最高为 99.67%,寻优精度值最佳。

对比可知,4 种算法中最优解质量最佳的是 MPA 算法,其次是 WOA 算法和 HHO 算法,最后是 PSO 算法。



图 6 MPA-MKL-FSVDD 模型寻优效果图









通过 MPA、HHO、WOA、PSO 算法优化 MKL-FS-VDD 模型后得到对应的寻优参数、准确率及运行时 间结果如表 3 所示,4 种模型故障诊断结果如图 10 所示。对比可知, MPA 优化的 MKL-FSVDD模型故 障分类的准确率最高,为92.80%,其次是 WOA 算 法和 HHO 算法,最后是 PSO 算法。MPA-MKL-FS-VDD 模型运行时间最短,仅需 48 s,且该模型训练速 度最快。HHO-MKL-FSVDD 模型训练速度较快,为 52 s,其次是 WOA-MKL-FSVDD 模型,最后是 PSO-MKL-FSVDD 模型。综上可知, MPA 算法优化的 MKL-FSVDD 故障分类模型在训练精度和测试精度 上可以得到较佳的分类效果,模型分类的运行速度 相对较快,在基于数据参数下聚氯乙烯聚合工艺的 故障诊断中可满足对故障分类效率和质量的要求。

表 3 寻优参数及模型性能比较

| 诊断方法 | 最佳参数 c | 最佳参数 g | 训练准确率 | 测试准确率 | 运行时间 |
|---------------|--------|--------|--------|-------|------|
| MPA-MKL-FSVDD | 96.728 | 0.314 | 99.67% | 92.8% | 48 s |
| HHO-MKL-FSVDD | 29.173 | 0.646 | 99.33% | 91.2% | 52 s |
| WOA-MKL-FSVDD | 87.851 | 0.395 | 99.43% | 92.0% | 55 s |
| PSO-MKL-FSVDD | 44.573 | 0.323 | 98.33% | 86.4% | 71 s |



图 10 MKL-FSVDD 结合 4 种寻优算法故障诊断结果比较

5 结论

本文以处理噪声干扰较好的 FSVDD 模型为基础,通过引入 MKL 算法保留复杂数据集的原始特征,再结合 MPA 寻优算法优化 MKL-FSVDD 模型,

构造 MPA-MKL-FSVDD 故障诊断模型。通过应用 TE 平台故障数据集,与 SVDD、FSVDD、MKL-FSVDD 故障诊断模型进行对照实验,验证了所提算法具有 较高的诊断效率和精度。通过利用 PVC 聚合反应 聚合釜设备的工业历史数据,对 MPA、HHO、WOA、 PSO 算法优化的 MKL-FSVDD 模型进行故障分类仿 — 389 — 真,实验对比各算法的优化效果,结果表明 MPA 对 模型参数全局优化、精准寻优的显著优势,说明本文 选用的 MPA 优化 MKL-FSVDD 模型具有较精准的 故障分类能力以及在聚合釜设备故障诊断中较强的 鲁棒性和泛化性能,达到了减小 PVC 聚合釜设备故 障诊断难度和提高聚合反应安全性的应用效果,该 模型为其相关化工过程的故障诊断及其他流程工艺 的生产过程建模、智能控制提供了新的优化方案。

鉴于实验选用了工业历史数据库中具有均匀性 特征的数据集进行仿真验证,考虑实际工业数据库 中故障类别不平衡数据问题的存在,接下来将以处 理不平衡样本故障诊断问题为方向进行深入研究。

参考文献

- [1]赵金萍,熊君星,刘建胜.基于故障数据的设备运行 可靠性分析与评估[J].高技术通讯,2017,27(4): 359-363
- [2] 邓佳伟,邓晓刚,曹玉苹,等.基于加权统计局部核 主元分析的非线性化工过程微小故障诊断方法[J]. 化工学报,2019,70(7):2594-2605
- [3] TRITTENBACH H, BOHM K, ASSENT I. Active learning of SVDD hyperparameter values [C] // 2020 IEEE 7th International Conferenceon Data Science and Advanced Analytics (DSAA), Sydney, Australia, 2020: 109-117
- [4]刘强,卓洁,朗自强,等.数据驱动的工业过程监控与自优化研究展望[J].自动化学报,2018,44(11): 26-38
- [5] 徐静,王振雷,王昕. 基于非线性动态全局局部保留 投影算法的化工过程故障检测[J]. 化工学报,2020, 71(12):5655-5663
- [6]周家驹,许志宏,石乐明,等.人工神经网络方法在 材料性能预报及化工过程故障诊断中的应用[J].化 工冶金,1993,14(1):57-62
- [7] WANG J S, LI S X, GAO J. SOM neural network fault diagnosis method of polymerization kettle equipment optimized by improved PSO algorithm [J]. Scientific World Journal, 2014, 2014(7): 1-12
- [8] WU D, ZHAO J. Process topology convolutional network model for chemical process fault diagnosis [J]. Process Safety and Environmental Protection, 2021, 150: 93-109
- [9] CHANG J, WANG G C, WANG Y, et al. Based on genetic algorithm and BP network to fault diagnosis of fermentation process [C] // 2008 Chinese Control and Decision Conference, Yantai, China, 2008; 578-581
- [10] WEN S, WANG J S, GAO J, et al. Fault diagnosis strat-- 390 ---

egy of polymerization kettle equipment based on support vector machine and cuckoo search algorithm [J]. *Engineering Letters*, 2017, 25(4): 2306-2314

- [11] TAX D M, DUIN R P W. Support vector domain description [J]. Machine Learning, 2004, 54(1): 45-66
- [12] 张汉元,田学民. 基于 KSFDA-SVDD 的非线性过程故 障检测方法[J]. 化工学报, 2016, 67(3): 827-832
- [13] 高杰. 基于数据的聚合间歇过程故障诊断方法研究
 [D]. 鞍山:辽宁科技大学电子与信息工程学院, 2014:42-57
- [14] 高淑芝, 赵娜. 基于 DKPCA 的聚合釜故障诊断研究 [J]. 沈阳化工大学学报, 2015(2):84-88,91
- [15] 王敏光, 王喆. 基于类心距离的模糊支持向量数据描述[J]. 计算机科学, 2016, 43(5): 230-233,242
- [16] 刘忠宝,赵文娟.面向大规模数据的模糊支持向量数据描述[J].广西大学学报(自然科学版),2012,37
 (6):1254-1260
- [17] 孙飞. 基于 FVMD-MPE 和 FSVDD 的运行可靠性评估 方法[D]. 秦皇岛: 燕山大学机械工程学院, 2018: 47-54
- [18] LOOSLI G, ABOUBACAR H. Using SVDD in simple MKL for 3D-shapes filtering [EB/OL]. https://arxiv. org/pdf/1712.02658.pdf:arXiv, (2017-12-07), [2021-05-08]
- [19] 卢明,刘黎辉,吴亮红. 多核支持向量数据描述分类 方法研究[J]. 计算机工程与应用, 2016, 52(18): 68-73
- [20] FARAMARZI A, HEIDARINEJAD M, MIRJALILI S, et al. Marine predators algorithm: a nature-inspired metaheuristic[J]. Expert Systems with Applications, 2020, 152(8): 113377-113405
- [21] CHEN X, QI X, WANG Z, et al. Fault diagnosis of rolling bearing using marine predators algorithm-based support vector machine and topology learning and out-of-sample embedding[J]. *Measurement*, 2021, 176(7): 109-116
- [22] 秦亮,周绍磊,史贤俊,等.基于模糊机会约束 SVDD 的故障诊断方法[J].系统工程与电子技术,2012,34 (12):2554-2558
- [23] CHEN D, KWONG S, QIANG H, et al. Geometrical interpretation and applications of membership functions with fuzzy rough sets[J]. *Fuzzy Sets and Systems*, 2012, 193: 122-135
- [24] 温祥西, 孟相如, 马志强, 等. 基于局部投影降噪和 FSVDD 的网络流量异常检测[J]. 计算机应用研究, 2013, 30(5): 1523-1526
- [25] TAO X M, LI Q, REN C, et al. Affinity and class probability based fuzzy support vector machine for imbalanced

datasets[J]. Neural Networks, 2020, 122(2): 289-307

- [26] 曹惠玲. 多核学习 SVM 在发动机故障诊断上的应用 [J]. 机械设计与制造, 2019, 339(5): 8-11,16
- [27] CHEN L, CHEN D, WANG H. Alignment based kernel selection for multi-label learning [J]. Neural Processing Letters, 2019, 49(3): 1157-1177
- [28] EBEED M, ALHEJJI A, KAMEL S, et al. Solving the optimal reactive power dispatch using marine predators algorithm considering the uncertainties in load and wind-solar generation systems [J]. *Energies*, 2020, 13 (17): 4316-4335
- [29] SATTAR M, SUMAITI A A, ALI H, et al. Marine predators algorithm for parameters estimation of photovoltaic modules considering various weather conditions[J]. Neural Computing and Applications, 2021, 32(7): 1-21
- [30] ELAZIZ M A, SHEHABELDEEN T A, ELSHEIKH A H, et al. Utilization of random vector functional link integrated with marine predators algorithm for tensile behavior prediction of dissimilar friction stir welded aluminum alloy joints[J]. Journal of Materials Research and Technology,

2020, 9(5): 11370-11381

- [31] CHEN G, GE Z. SVM-tree and SVM-forest algorithms for imbalanced fault classification in industrial processes[J]. IFAC Journal of Systems and Control, 2019, 8: 1-10
- [32] 张清清. 基于孪生 LSTM 的 TE 过程故障诊断方法研究[D]. 成都:电子科技大学自动化工程学院, 2020: 12-15
- [33] 王建国,陈锴,张文兴,等. 鲸鱼优化多核支持向量 机的滚动轴承故障诊断[J]. 电子设计工程,2021,29 (3):31-35
- [34] 张磊. 建模与优化相集成的聚氯乙烯生产过程优化控制方法研究[D]. 鞍山:辽宁科技大学电子与信息工程学院, 2020: 2-7,12
- [35] 费鹏程. 粗糙集与神经网络在聚合釜故障诊断中的方 法研究[D]. 沈阳:东北大学信息科学与工程学院, 2010:41-51
- [36] 高淑芝. 聚氯乙烯生产过程智能控制方法及应用研究 [D]. 沈阳:东北大学信息科学与工程学院, 2012: 11-19,44

Polymerization kettle equipment fault diagnosis based on MPA optimized MKL-FSVDD model

LI Guoyou*, CAI Shiwen*, LI Dongshuo**, ZHANG Xinkui*, JIA Yaoyu*, NING Ze*

(*Engineering Research Center of the Ministry of Education for Intelligent Control System

and Intelligent Equipment, Yanshan University, Qinhuangdao 066004)

(** School of Optoelectronic Science and Engineering, University of Electronic Science

and Technology of China, Chengdu 611731)

Abstract

In order to solve the problems that the data of chemical process industry has strong non-linearity, easy to be affected by noise and the fault is multi-fault classification, a fault diagnosis method based on marine predators algorithm (MPA) optimized multi-kernel learning and fuzzy support vector machine data description (MKL-FSVDD) is proposed. The multi-kernel function constructed by MKL makes up for the limitation of single kernel function, and has strong adaptability to nonlinear fault data classification. MPA is introduced to optimize the kernel parameters of MKL-FSVDD model efficiently, and the problem of kernel parameter selection can be solved. The effective performance of MPA-MKL-FSVDD model for fault diagnosis is verified through the control experiment on the TE data platform. Finally, the algorithm is applied to the polyvinyl chloride (PVC) polymerization reaction, and the historical data set of the 70 m³ polymerization kettle is used for simulation verification. The results show that the proposed method makes full use of the data information of the complex sample set, and the optimal solution can be obtained quickly and stably in the parameter optimization stage, which guarantes the efficiency and accuracy of fault classification.

Key words: fault diagnosis, marine predators algorithm (MPA), multiple kernel learning (MKL), fuzzy membership degree, polymerization kettle