

文章编号: 1001-0920(2011)07-0967-06

一种改进的支持向量数据描述故障诊断方法

唐明珠¹, 王岳斌², 阳春华¹

(1. 中南大学 信息科学与工程学院, 长沙 410083; 2. 湖南理工学院 计算机学院, 湖南 岳阳 414006)

摘要: 针对故障诊断中故障类样本难于获取以及不均衡类问题, 提出了基于粒子群和滑动窗口的支持向量数据描述(M-SVDD)故障诊断方法. 该方法利用粒子群优化支持向量数据描述的核参数, 同时引入滑动窗口技术, 通过大窗口大小来控制故障诊断模型的训练样本数, 根据小窗口的预测误差变化动态调整大窗口的大小. 采用该方法对铜转炉吹炼过程进行故障诊断的实验结果表明, 该方法能有效抑制过拟合现象, 具有故障敏感性高、泛化能力强等特点.

关键词: 支持向量数据描述; 粒子群; 故障诊断; 不均衡类; 滑动窗口

中图分类号: TP18

文献标识码: A

Modified support vector data description for fault diagnosis

TANG Ming-zhu¹, WANG Yue-bin², YANG Chun-hua¹

(1. School of Information Science and Engineering, Central South University, Changsha 410083, China; 2. School of Computer, Hu'nan Institute of Science and Technology, Yueyang 414006, China. Correspondent: YANG Chun-hua, E-mail: ychh@mail.csu.edu.cn)

Abstract: Aiming at the difficulty of obtaining fault samples and class imbalanced problem, a modified support vector data description for fault diagnosis based on both particle swarm optimization and sliding windows(M-SVDD) is proposed in this paper. The kernel parameters of support vector data description are optimized by the particle swarm optimization. At the same time, the sliding window technique is introduced. The number of training samples for fault diagnosis model is controlled by a dynamic adjusted large window. The size of the large window is adjusted dynamically according to the changes of predicting error of the small window. M-SVDD is applied to the fault diagnosis of copper-converting smelting process. The experimental results show that M-SVDD can prevent effectively the phenomenon of over-fitting and has good fault sensitivity and generalization.

Key words: support vector data description; particle swarm optimization; fault diagnosis; class imbalanced; sliding window

1 引言

判断复杂工业过程是否发生故障是安全生产的重要步骤之一, 及时发现故障对于提高产品的产量和质量具有重要的意义. 对该过程采样时, 往往很难获取所有故障类别, 甚至有些故障样本无法获得. 因此, 在不考虑或较少考虑故障类样本的前提下, 研究如何利用比较容易获取的复杂工业过程正常类样本构造一个能准确判断复杂工业过程是否处于正常状态的诊断模型更具实用价值.

近年来, 基于数据驱动的故障诊断方法得到了广泛应用, 包括代价敏感支持向量机^[1]、代价敏感概

率神经网络^[2]和 K -最邻近(KNN)^[3]等方法. 代价敏感概率神经网络和代价敏感支持向量机在进行故障诊断时均假设样本集中至少有两类样本, 而在实际应用中往往类别不均衡, 甚至只能获取某一类样本, 如正常类样本. 因此, 这两种方法利用有限的有色冶金过程故障类样本和正常类样本建立的故障诊断模型应用于实际有色冶金过程时受到限制. 文献[3]提出的KNN算法用一类样本便可以进行故障诊断, 但在预测时存在过拟合现象.

Tax 等人在 SVM 的基础上提出了支持向量数据描述(SVDD)^[4], 在机械故障诊断、图像检测等方面得

收稿日期: 2010-03-25; 修回日期: 2010-06-11.

基金项目: 国家杰出青年科学基金项目(61025015); 国家自然科学基金项目(60874069); 国家863计划项目(2009AA04Z137); 中南大学优秀博士学位论文扶植项目.

作者简介: 唐明珠(1983-), 男, 博士生, 从事数据挖掘、机器学习等研究; 阳春华(1965-), 女, 教授, 博士生导师, 从事复杂工业过程建模与优化控制等研究.

到了成功应用^[5-6]。但是, 由于引入了核函数, 其核参数的选择直接影响了 SVDD 的性能, 只有选择合适的核参数才能获得理想的效果。目前, 交叉验证是人们常用的核参数选择方法^[7], 但该方法耗时严重, 而且是经验上的最优值。文献 [8] 提出了一种改善支撑向量域描述性能的核优化算法, 但该算法要求数据服从高斯分布, 且各分量独立、同方差。因此, 上述两种核参数优化方法的应用受到限制。粒子群 (PSO) 优化算法在解决大规模非线性的连续问题中具有更高的优化效率和更好的优化结果, 在工业优化控制中得到了广泛应用^[9-10], 因而可用 PSO 来优化 SVDD 的核参数。另外, SVDD 很难应用大规模样本集, 文献 [11-12] 采用滑动窗口技术, 通过一个动态调整的大窗口来控制故障诊断模型中所包含的最近产生工况, 以反映工况的变化, 同时能控制样本集的规模。

本文利用粒子群优化支持向量数据描述的核参数, 通过滑动窗口技术中一个动态调整的大窗口来控制故障诊断模型中包含的最近产生工况, 以反映工况的变化, 根据小窗口的预测误差变化动态调整大窗口的大小, 提出了基于粒子群和滑动窗口的支持向量数据描述 (M-SVDD) 故障诊断方法。实验结果表明, 该方法具有良好的故障敏感性, 能有效防止过拟合现象。

2 支持向量数据描述

支持向量数据描述 (SVDD) 将需要描述的样本作为一个整体, 试图建立一个封闭而紧凑的区域 Ω , 使被描述的样本全部或尽可能多地包容在 Ω 内, 而非该类样本应没有或尽可能少地包含在 Ω 内^[4]。设训练样本集 X 中包含 n 个样本 $x_i, i = 1, 2, \dots, n, x_i \in R^d$, 设法找到一个以 a 为中心, 以 R 为半径的能包含所有训练样本的最小超球体。如果直接进行优化处理, 所得到的优化区域是一个超球体。为了使优化区域更紧凑, 采用核函数替代计算中的点积运算, 通过非线性映射函数 φ 实现低维空间到高维空间的映射, 然后求解在高维空间包含所有样本点的最小超球体。为增强支持向量数据域描述的鲁棒性, 引入松弛因子 $\xi_i \geq 0, i = 1, 2, \dots, n$ 。寻找一个满足 mercer 条件的核函数 $K(x_i \cdot x_j)$, 使得 $K(x_i \cdot x_j) = \langle \varphi(x_i), \varphi(x_j) \rangle$ 。这样, 优化问题可描述为

$$\begin{aligned} \min_R f(R, a) &= R^2 + C \sum_{i=1}^n \xi_i. \\ \text{s.t. } (\varphi(x_i) - a)(\varphi(x_i) - a)^T &\leq R^2 + \xi_i, \\ & i = 1, 2, \dots, n; \\ \xi_i &\geq 0, i = 1, 2, \dots, n. \end{aligned} \quad (1)$$

其中参数 C 用于控制最小超球体体积与数据误差之间的平衡, 称为 SVDD 的控制参数 C 。该问题的对偶

形式为

$$\begin{aligned} \max_{\alpha_i} L &= \sum_{i=1}^n \alpha_i K(x_i \cdot x_i) - \sum_{i,j} \alpha_i \alpha_j K(x_i \cdot x_j). \\ \text{s.t. } \sum_{i=1}^n \alpha_i &= 1, i = 1, 2, \dots, n; \\ 0 &\leq \alpha_i \leq C, i = 1, 2, \dots, n. \end{aligned} \quad (2)$$

其中参数 α_i 为拉格朗日乘子。解该优化问题可得到参数 α_i , 称 $\alpha_i > 0$ 对应的样本 x_i 为支持向量。由定义可知, R^2 为超球体中心 a 与边界上任何支持向量之间的距离, 即与拉格朗日乘子 $0 < \alpha_i < C$ 的支持向量 (样本) 之间的距离, 但不包含拉格朗日乘子为 $\alpha_i = C$ 的支持向量 (样本)。因此

$$\begin{aligned} R^2 &= K(x_k \cdot x_k) - 2 \sum_i \alpha_i K(x_i \cdot x_k) + \\ & \sum_{i,j} \alpha_i \alpha_j K(x_i \cdot x_j), \forall x_k \in \text{SV}_{<C}. \end{aligned} \quad (3)$$

根据式 (3), 用任意一个支持向量可求出 R^2 。对于测试样本 z , 设

$$\begin{aligned} D(a, z) &= \\ & K(z \cdot z) - 2 \sum_i \alpha_i K(z \cdot x_i) + \sum_{i,j} \alpha_i \alpha_j K(x_i \cdot x_j), \\ & i = 1, 2, \dots, n, j = 1, 2, \dots, n. \end{aligned} \quad (4)$$

若 $D(a, z) \leq R^2$, 则 z 被判决为正常类样本; 否则, z 被判决为故障类样本。

由文献 [4] 可知, 在常见的 4 种核函数中, 高斯核函数在 SVDD 中能取得最好的分类效果。当选择高斯核函数

$$K_G(x, y) = \exp\left(-\frac{\|x - y\|^2}{\sigma^2}\right), \sigma > 0$$

时, 式 (2) 和 (4) 分别变为

$$\begin{aligned} L &= \sum_{i=1}^n \alpha_i K_G(x_i \cdot x_i) - \sum_{i,j} \alpha_i \alpha_j K_G(x_i \cdot x_j) = \\ & 1 - \sum_i \alpha_i^2 - \sum_{i \neq j} \alpha_i \alpha_j K_G(x_i \cdot x_j); \\ \text{s.t. } 0 &\leq \alpha_i \leq C, i = 1, 2, \dots, n, j = 1, 2, \dots, n. \end{aligned} \quad (5)$$

$$D(a, z) = K_G(z \cdot z) - 2 \sum_i \alpha_i K_G(z \cdot x_i) +$$

$$\begin{aligned} & \sum_{i,j} \alpha_i \alpha_j K_G(x_i \cdot x_j) = \\ & 1 - 2 \sum_i \alpha_i K_G(z \cdot x_i) + \sum_i \alpha_i^2 + \\ & \sum_{i \neq j} \alpha_i \alpha_j K_G(x_i \cdot x_j), \end{aligned}$$

$$i = 1, 2, \dots, n, j = 1, 2, \dots, n. \quad (6)$$

式 (6) 中只有一项 $2 \sum_i \alpha_i K_G(z \cdot x_i)$ 与测试样本 z 有关系。

3 线性权重粒子群算法

PSO初始化为一群随机粒子(随机解),然后通过迭代找到最优解^[10]. 假设 d 维搜索空间中的第 i 个粒子的位置和速度分别为 $S^i = (s_{i,1}, s_{i,2}, \dots, s_{i,d})$ 和 $V^i = (v_{i,1}, v_{i,2}, \dots, v_{i,d})$,在每一次迭代中,粒子通过跟踪2个最优化解来更新自己,第1个是粒子本身所找到的最优化解,即个体极值 p_{best} , $P^i = (p_{i,1}, p_{i,2}, \dots, p_{i,d})$;另一个是整个种群目前找到的最优解,即全局最优解 g_{best} . p_g 在找到这2个最优值后,粒子通过下面公式更新自己的速度和位置:

$$v_{i,j}^{(t+1)} = wv_{i,j}^{(t)} + c_1r_1(p_{i,j} - s_{i,j}^{(t)}) + c_2r_2(p_{g,j} - s_{i,j}^{(t)}); \quad (7)$$

$$s_{i,j}^{(t+1)} = s_{i,j}^{(t)} + v_{i,j}^{(t)}, j = 1, 2, \dots, d. \quad (8)$$

其中: w 为惯性权重因子, c_1 和 c_2 为正的学习因子, r_1 和 r_2 为0到1之间均匀分布的随机数.在PSO的可调参数中,惯性权重 w 是最重要的参数, t 和 $t+1$ 分别为当前迭代次数和下一次迭代次数.因为较大的惯性权重有利于跳出局部极小点,便于全局搜索,利于算法收敛,因此针对PSO算法容易早熟以及算法后期易在全局最优解附近产生振荡现象,可采用线性变化的权重,让惯性权重从最大值 w_{max} 线性减小到最小值 w_{min} . w 随算法迭代次数的变化公式为

$$w = w_{max} - \frac{t(w_{max} - w_{min})}{t_{max}}. \quad (9)$$

其中: w_{max} 和 w_{min} 分别表示 w 的最大值和最小值, t 和 t_{max} 分别为当前迭代次数和最大迭代次数.

4 M-SVDD

针对SVDD的核参数 σ 选择问题构造PSO的适应度函数.通过优化迭代计算核参数 σ ,可以提高SVDD的泛化性能和故障敏感性.对于未知样本,预测误差要尽可能地小.用预测误差作为适应度函数,对每一个粒子进行判断,最终选出的预测误差最小时所对应的粒子即为最优解

$$y^* = \min \frac{1}{N} \sum_{i=1}^N (y_i^{(t)} - y_i^{(p)})^2. \quad (10)$$

其中: $y_i^{(t)}$, $y_i^{(p)}$ 分别表示第 i 个样本真实的类别值(正常类取值为1,异常类取值为-1)和预测的类别值; N 表示训练集样本数; y^* 为适应度函数的最小值,即最优值.

为了将训练样本集规模控制在合理范围内,本文采用了滑动窗口技术.滑动窗口技术是指利用一个动态调整的大窗口控制预测模型中所包含的最近工况样本,以反映工况的改变,同时引入一个动态调整的小窗口来决定增加、减少或保持大窗口.当大窗口和小窗口中的预测误差一致增加时,表明过时的工况信

息增加了预测误差,因此应缩短2个窗口的长度.当大窗口和小窗口中的预测误差一致减少时,说明新的工况加强了原有模型的预测能力,因此应扩大2个窗口的长度.当大窗口和小窗口中的预测误差稳定时,表明新的工况和原有模型保持一致,此时应保持2个窗口的长度不变.当大窗口和小窗口中的预测误差不包含在上述任何一种情况时,为避免过早地忘掉旧的训练样本,可增加大窗口的大小,并保持小窗口的长度不变.

利用M-SVDD进行故障诊断可分为2步:首先利用PSO优化SVDD的核参数,利用滑动窗口技术中的大窗口控制样本的规模,以获得参数优化后的SVDD模型;然后将该模型用于测试新的工况样本类别.

M-SVDD算法描述如下:

算法1 M-SVDD算法.

输入参数: PSO算法参数,大窗口的大小 L_{win} ,小窗口的大小 S_{win} ,预测误差 $y_{L_{win}}^*$, $y_{S_{win}}^*$,样本集dataset,样本比例 p ;

输出参数: M-SVDD的故障敏感性和测试样本集的预测类别.

Step 1: 初始化PSO算法参数: 粒子数目 M ,粒子位置 S 和速度 V ,学习因子 c_1 和 c_2 ,最大权重 w_{max} 和最小权重 w_{min} ,最大迭代次数 t_{max} ,问题维数 D ,PSO的预设精度,大窗口大小 L_{win} ,小窗口大小 S_{win} ,样本集dataset,样本比例 p ,训练样本集trainingset = dataset * p ,测试样本集testset = dataset * (1 - p),预测误差temp $_{L_{win}} = y_{L_{win}}^*$, temp $_{S_{win}} = y_{S_{win}}^*$;

Step 2: While (测试样本集testset不为空){

Step 3: 根据文献[4]的方法,计算控制参数 C ;

Step 4: 调用PSO算法优化SVDD的核参数,返回训练样本集trainingset SVDD最优核参数;

Step 5: 将测试样本集testset中的采样时间最早的样本作为测试样本 z ;

Step 6: 用决策函数(6)对测试样本进行测试,判断该样本是否为故障类样本;

Step 7: 如果预测样本为故障类,则将该样本存入故障类样本库;如果预测样本为正常类,则将该样本存入正常类样本库;

Step 8: 将测试样本 z 添加到大窗口和小窗口,同时删除大窗口和小窗口的采样时间最早的样本;

Step 9: 计算大窗口和小窗口中的新的预测误差 $y_{L_{win}}^*$, $y_{S_{win}}^*$;

Step 10: 如果

$$y_{L_{win}}^* < \text{temp}_{L_{win}}, y_{S_{win}}^* < \text{temp}_{S_{win}},$$

则增大 $L_{\text{win}}, S_{\text{win}}$;

Step 11: 如果

$$y_{L_{\text{win}}}^* > \text{temp}_{L_{\text{win}}}, y_{S_{\text{win}}}^* > \text{temp}_{S_{\text{win}}},$$

则减小 $L_{\text{win}}, S_{\text{win}}$;

Step 12: 如果

$$y_{L_{\text{win}}}^* = \text{temp}_{L_{\text{win}}}, y_{S_{\text{win}}}^* = \text{temp}_{S_{\text{win}}},$$

则保持 $L_{\text{win}}, S_{\text{win}}$;

Step 13: 如果新预测误差为其他情况, 则增加大窗口的大小 L_{win} , 并保持小窗口的长度 S_{win} 不变;

Step 14: 保存新的预测误差 $y_{L_{\text{win}}}^*, y_{S_{\text{win}}}^*$;

Step 15: 更新预测误差

$$\text{temp}_{L_{\text{win}}} = y_{L_{\text{win}}}^*, \text{temp}_{S_{\text{win}}} = y_{S_{\text{win}}}^*;$$

Step 16: 在测试样本集 testset 中删除测试样本 z ;

Step 17: 根据每一个测试样本的预测误差计算故障敏感性, 输出故障敏感性和测试样本的类别.

PSO 算法优化 SVDD 核参数算法描述如下:

算法 2 M-SVDD 核参数算法.

输入参数: PSO 算法的粒子数目 M , 粒子位置 S 和速度 V , 学习因子 c_1 和 c_2 , 最大权重 w_{max} 和最小权重 w_{min} , 最大迭代次数 t_{max} , 问题的维数 D , PSO 的预设精度, 控制参数 C ;

输出参数: SVDD 的预测误差和最优核参数 σ .

Step 1: 初始化 PSO 算法参数: 粒子数目 M , 粒子位置 S 和速度 V , 学习因子 c_1 和 c_2 , 最大权重 w_{max} 和最小权重 w_{min} , 最大迭代次数 t_{max} , 问题的维数 D , PSO 的预设精度, 控制参数 C ;

Step 2: 评价每个微粒的适应度, 将当前微粒的位置和适应值存储在各微粒的 p_{best} 中, 将所有 p_{best} 中适应值最优个体的位置和适应值存储于 g_{best} 中;

Step 3: 分别用式 (7) 和 (8) 更新粒子的速度和位移;

Step 4: 用式 (9) 更新权重;

Step 5: 对于每个微粒, 将其适应值与其经历过的最好位置进行比较, 如果较好, 则将其作为当前的最好位置, 比较当前所有 p_{best} 和 g_{best} 的值, 更新 g_{best} ;

Step 6: 如果满足预设的运行精度或最大的迭代次数 t_{max} , 则训练算法结束, 输出 SVDD 的最优核参数值和预测误差; 否则, 返回 Step 3 继续搜索.

5 工业应用

铜转炉吹炼过程是一个典型的复杂工业过程. 冰铜的吹炼造渣期分为 S1 期和 S2 期, 从风口向炉内熔体中鼓入空气或富氧空气. 通过空气的搅拌, 冰铜中的 FeS 被氧化, 生成 FeO 和 SO₂; FeO 再与添加的熔剂

中的 SiO₂ 进行造渣反应. 造渣反应一结束, 便暂停送风, 进行排渣操作. 在转炉吹炼过程中, 发生的反应几乎全部是放热反应, 为防止由于产生大量的热量而缩短耐火砖的寿命, 需向炉内加入包壳、块状烟灰和残极等冷料, 以控制炉内温度. 铜转炉吹炼过程是间歇式作业.

通过深入分析铜转炉吹炼过程机理可知: 造渣 S1 期的影响因素有 5 个, 分别为 S1 期入炉铜钼的总量, S1 期入炉铜钼的品位, S1 期熔剂投放总量, S1 期冷料投放总量, S1 期鼓风量; 造渣 S2 期的影响因素有 13 个, 分别为 S1 期入炉铜钼的总量, S1 期入炉铜钼的品位, S1 期熔剂投放的总量, S1 期冷料投放总量, S1 期鼓风量, S1 期出炉总渣量, S1 期出炉炉渣的含铜量, S1 期出炉炉渣的 SiO₂ 含量, S2 期新入炉的铜钼的总量, S2 期新入炉的铜钼的品位, S2 期新增熔剂投放总量, S2 期新增冷料投放总量, S2 期鼓风量. 将 S1 期的 5 个影响因素和 S2 期的 13 个影响因素分别作为 M-SVDD 的输入变量.

以某铜冶炼厂 3 号 PS 铜转炉为对象, 从生产现场采集的样本集分布如表 1 所示. 表 1 中: d_{s1} 和 d_{s2} 分别表示 2009 年 1~6 月某铜冶炼厂 3 号 PS 铜转炉吹炼过程 S1 期样本集和 S2 期样本集, 其中正常类样本数分别为 454 和 641, 2 个样本集都无故障类样本; d_{s3} 和 d_{s4} 分别表示 2009 年 7~12 月某铜冶炼厂 3 号 PS 铜转炉吹炼过程 S1 期样本集和 S2 期样本集, 其中正常类样本数分别为 548 和 736, 人工产生故障类样本数分别为 128 和 134. $|d_i|$ 表示第 i 个样本集的正常类样本数; p 表示训练样本数所占对应样本集中正常类样本数的比例, $p \in [0, 1]$. $|d_i|p$ 表示训练集的样本数, $|d_i|(1-p)$ 与故障类样本数之和为测试集样本数. 对 4 个样本集中的所有样本进行归一化处理.

表 1 4 个样本集训练集样本数与测试集样本数分布

样本集名	正常类样本数	训练集样本数	测试样本数	
			正常类	故障类
d_{s1}	454	$ d_1 p$	$ d_1 (1-p)$	0
d_{s2}	641	$ d_2 p$	$ d_2 (1-p)$	0
d_{s3}	548	$ d_3 p$	$ d_3 (1-p)$	128
d_{s4}	736	$ d_4 p$	$ d_4 (1-p)$	134

以 p 的步长为 0.1 设定 M-SVDD 故障诊断模型的训练样本数, 然后设定线性递减权重 PSO 种群参数 $M = 25$, $c_1 = c_2 = 2$, $w_{\text{max}} = 0.9$, $w_{\text{min}} = 0.4$, $v_{\text{max}} = 1$, $t_{\text{max}} = 50$, $D = 1$. 由文献 [4] 优化控制参数 C 理论可知, 样本集 d_{s1} , d_{s2} , d_{s3} 和 d_{s4} 对应的控制参数 C 值分别为 0.028 2, 0.015 6, 0.014 8, 0.011 5. 滑动窗口技术中大窗口大小初值

$$L_{\text{win}} = \frac{1}{2}|d_i|p,$$

小窗口大小初值

$$S_{win} = \frac{1}{4}|d_i|p.$$

通过线性递减 PSO 算法,可获得 M-SVDD 最优核参数如表 2 所示. 表 2 中: p 表示训练样本集所占正常类样本集的比例, σ_{d_i} 分别表示不同比例 p 样本集 d_{s_i} 的 M-SVDD 最优核参数.

表 2 4 个训练样本集 10 个 p 值下 M-SVDD 的最优核参数

p	σ_{d_1}	σ_{d_2}	σ_{d_3}	σ_{d_4}
0.1	20.7094	11.0684	13.2791	9.1125
0.2	20.4363	9.8598	12.1162	9.1315
0.3	20.3381	9.5647	9.1315	9.7271
0.4	14.3086	14.5013	11.5520	11.5520
0.5	10.1141	5.1850	12.8977	11.7057
0.6	17.8387	7.3114	9.7271	12.1162
0.7	16.7641	13.9130	13.8117	12.8977
0.8	5.5253	12.6210	11.7057	13.2791
0.9	10.8941	13.2141	9.1125	13.8117

利用表 2 获得 SVDD 的核参数, 分别对测试样本集 $|d_1|(1-p)$ 和 $|d_2|(1-p)$ 进行测试, 测试结果如图 1 和图 2 所示.

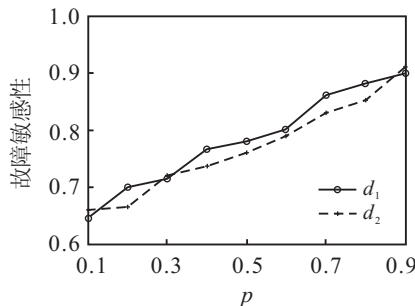


图 1 M-SVDD 在样本集 d_1 和 d_2 上故障敏感性比较

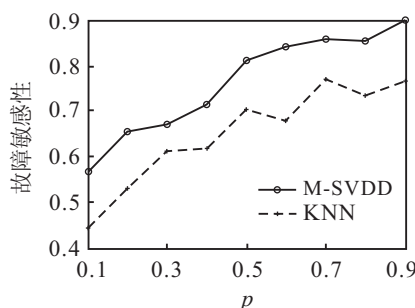


图 2 M-SVDD 与 KNN 在样本集 d_3 上故障敏感性比较

由图 1 可知, M-SVDD 分别随着 S1 期和 S2 期样本集 d_1 和 d_2 的训练集所占整个样本集比例 p 的增加, 故障敏感度越来越高, 原因是训练样本数增加有利于 M-SVDD 训练更充分, 有更多有用的样本信息.

为更好地说明 M-SVDD 算法性能, 将 M-SVDD 与一类分类的 KNN 算法^[3]在样本集 d_{s3} 和 d_{s4} 上进行对比实验. 对不同 p 值下 M-SVDD 和 KNN 的故障敏感性进行了比较. 不同的 p 值在 d_3 和 d_4 上所对应的 M-SVDD 上最优核参数 σ 如表 2 所示.

利用表 2 获得的 M-SVDD 的核参数 σ_{d_3} 对测试集 $|d_3|(1-p)$ 和 d_{s3} 故障类样本集进行测试, 实验结果如图 2 所示. 用 M-SVDD 核参数 σ_{d_4} 对测试集 $|d_4|(1-p)$ 和 d_{s4} 故障类样本进行测试, 实验结果如图 3 所示. 由图 2 和图 3 可知: M-SVDD 在训练样本集所占整个样本集比例 p 比较小的情况下, 表现出了较好的故障敏感性; M-SVDD 和 KNN 随着 p 值的增加, 其故障敏感性提高, 但由于 KNN 在训练集上有过拟合现象, 在测试集上表现不佳, 而 M-SVDD 是对样本集边界进行描述, 其在测试集上具有良好的故障敏感性.

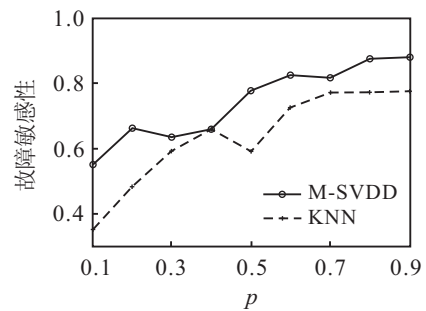


图 3 M-SVDD 与 KNN 在样本集 d_4 上故障敏感性比较

滑动窗口技术中大窗口的大小是影响 M-SVDD 故障敏感性的关键参数. 故障敏感性是指 M-SVDD 正确诊断的正类样本比例. 故障敏感性定义为 $Sensitivity = tp/pos$. 其中: tp 是指使用 M-SVDD 模型将原本属于正类的样本正确地诊断为正类数, pos 为总的正类样本数. 大窗口大小用于控制 M-SVDD 故障诊断模型中的训练样本数 pos , 根据小窗口的预测误差变化动态调整大窗口的大小. 大窗口越大, M-SVDD 故障诊断模型训练越充分, M-SVDD 的故障敏感度越高; 大窗口越小, M-SVDD 故障诊断模型训练越不充分, M-SVDD 故障敏感度越小. 从图 2 和图 3 可知, 对于某一个样本集而言, 随着 p 值增大, 大窗口的大小 $L_{win} = \frac{1}{2}|d_i|p$ 和小窗口的大小 $S_{win} = \frac{1}{4}|d_i|p$ 相应增大, M-SVDD 的故障敏感度不断地迅速增加, 但增加到一定程度后, 故障敏感度增加幅度减小.

故障诊断的响应时间是评价故障诊断系统的关键参数. 在故障诊断系统中, 故障诊断的响应时间越短, 越能满足 M-SVDD 实时应用需求. M-SVDD 通过滑动窗口技术中的大窗口大小控制故障诊断模型的训练样本数, 这样可以有效地减少该方法的训练时间. M-SVDD 通过训练时获得的支持向量对未知故障样本进行诊断. 一般情况下, 支持向量数远小于训练样本数^[4], 这样可以有效地减少对未知故障样本诊断的时间. 在实验中, M-SVDD 在 4 个样本集上的平均训练时间分别为 0.18 s, 0.25 s, 0.21 s, 0.29 s, 对一个未知故障样本的平均诊断时间分别为 0.15 s, 0.26 s, 0.23 s, 0.3 s. 因此, M-SVDD 能够满足故障诊断实时应用的需求.

6 结 论

本文针对故障诊断中故障类样本难以获取以及不平衡类问题,提出了基于粒子群和滑动窗口的支持向量数据描述(M-SVDD)故障诊断方法。

用支持向量数据描述模型进行故障诊断时,模型的核参数选择非常关键。利用线性粒子群算法优化支持向量数据描述的核参数,可以提高支持向量数据描述的泛化能力和故障敏感性。采用滑动窗口技术,通过一个动态调整的大窗口来控制故障诊断模型中所包含的最近产生工况,以反映工况的变化和控制训练样本集规模,根据小窗口的预测误差变化动态调整大窗口的大小。

以典型有色冶金过程——铜转炉吹炼过程为例,深入分析了该过程机理,选择该过程中合适的输入条件和过程参数作为M-SVDD的输入参数,减少了噪声变量干扰。

实验结果表明,与KNN相比,M-SVDD能有效地提高故障敏感性和泛化能力,防止过拟合现象。M-SVDD仅依靠正常类样本,便可以对铜转炉吹炼过程进行故障诊断。该方法在故障诊断领域具有重要的推广价值和广泛的应用前景。

参考文献(References)

- [1] 郑恩辉,李平,宋执环. 代价敏感支持向量机[J]. 控制与决策, 2006, 21(4): 473-476.
(Zheng E H, Li P, Song Z H. Cost sensitive support vector machines[J]. Control and Decision, 2006, 21(4): 473-476.)
- [2] 唐明珠,阳春华,桂卫华. 代价敏感概率神经网络及其在故障诊断中的应用[J]. 控制与决策, 2010, 25(7): 1074-1078.
(Tang M Z, Yang C H, Gui W H. Cost-sensitive probabilistic neural network and its application in the fault diagnosis[J]. Control and Decision, 2010, 25(7): 1074-1078.)
- [3] Angiulli F. Condensed nearest neighbor data domain description[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2007, 29(10): 1746-1758.
- [4] Tax D M J, Duin R P W. Support vector data description[J]. Machine Learning, 2004, 54(1): 45-66.
- [5] Zhang Y, Liu X D, Xie F D, et al. Fault classifier of rotating machinery based on weighted support vector data description[J]. Expert Systems with Applications, 2009, 36(3): 7928-7932.
- [6] Bovolo F, Camps V G, Bruzzone L. A support vector domain method for change detection in multitemporal images[J]. Pattern Recognition Letters, 2010, 31(10): 1148-1154.
- [7] Lee K Y, Kim D W, Lee D, et al. Improving support vector data description using local density degree[J]. Pattern Recognition, 2005, 38(10): 1768-1771.
- [8] 赵峰,张军英,刘敬. 一种改善支撑向量域描述性能的核优化算法[J]. 自动化学报, 2008, 34(9): 1122-1127.
(Zhao F, Zhang J Y, Liu J. An optimizing kernel algorithm for improving the performance of support vector domain description[J]. Acta Automatic Sinica, 2008, 34(9): 1122-1127.)
- [9] Del V Y, Venayagamoorthy G K, Mohagheghi S, et al. Particle swarm optimization: Basic concepts, variants and applications in power systems[J]. IEEE Trans on Evolutionary Computation, 2008, 12(2): 171-195.
- [10] Chang Y P, Ko C N. A PSO method with nonlinear time-varying evolution based on neural network for design of optimal harmonic filters[J]. Expert Systems with Applications, 2009, 36(3): 6809-6816.
- [11] Tanbeer S K, Ahmed C F, Jeong B, et al. Sliding window-based frequent pattern mining over data streams[J]. Information Sciences, 2009, 179(22): 3843-3865.
- [12] Li H F, Lee S Y. Mining frequent itemsets over data streams using efficient window sliding techniques[J]. Expert Systems with Applications, 2009, 36(2): 1466-1477.
- [18] Zhang Qingfu, Sun Jianyong, Edward Tsang, et al. Hybrid estimation of distribution algorithm for global optimization[J]. Engineering Computations, 2004, 21(1): 91-107.
- [19] 王湘中,喻寿益. 适用于高维优化问题的改进进化策略[J]. 控制理论与应用, 2005, 23(1): 148-151.
(Wang X Z, Yu S Y. Improved evolution strategies for high-dimensional optimization[J]. Control Theory & Application, 2003, 16(4): 390-396.)
- [20] 江中央. 正交遗传算法及其应用研究[D]. 长沙: 中南大学信息科学与工程学院, 2008.
(Jiang Z Y. Research on orthogonal genetic algorithm and its applications[D]. Changsha: School of Information Science and Engineering, Central South University, 2008.)

(上接第966页)