

基于改进 KFDA 和 RWn -SVM 的化工过程故障快速诊断

王斌 工程师 匡蕾 高级工程师

(江苏省安全生产科学研究院 化工安全与信息研究所, 江苏 南京 210009)

学科分类与代码: 6203070(安全系统工程) 中图分类号: X928.03 文献标志码: A

基金项目: 江苏省自然科学基金(SBK201222273)。

【摘要】 为提高复杂化工过程故障在线诊断水平, 提出一种基于改进的核费舍主元分析(KFDA)和鲁棒损失小波 n -支持向量机(RWn -SVM)集成的故障快速诊断方法。首先利用近邻边界法对 KFDA 进行监督降维, 扩大样本数据最近异类点和最远同类点的边界, 保证了样本投影后不同类别的数据不发生重叠。其次将参数 b 纳入支持向量机最优化问题的原问题求解中, 使回归估计函数表达式中不含参数 b , 从而简化了计算。设计的鲁棒损失函数和小波核函数能有效压制复杂数据样本中多种噪音和奇异点, 具有很强的鲁棒性和实用性。最后对经典的 Tennessee Eastman Process (TEP)化工过程进行实例分析, 结果表明改进方法的故障诊断速度和正确率均好于普通方法, 故障诊断结果能够反映化工过程的实际情况。

【关键词】 化工过程; 核主元分析; 支持向量机; 快速故障诊断

*本文已被《中国安全科学学报》录用(2013年23卷第7或8期)

A Fast Fault Diagnosis Method of Chemical Process Based on Improved Kernel Fisher Discriminant Analysis and Robust Wavelet n -Support Vector Machine

WANG Bin KUANG Lei

(Institute of Chemical Safety & Information Technology, Jiangsu Academy of Safety Science and Technology, Nanjing Jiangsu 210009, China)

Abstract: To improve the on-line fault diagnosis level of complex chemical process, an improved fast fault diagnosis method based on improved kernel Fisher discriminant analysis and robust wavelet n -support vector machine was proposed. Firstly, by using the neighbor boundary method for supervised dimensionality of KFDA, it can help to expand the boundary between nearest heterogeneous point and farthest congeneric point of sample, which to ensure the sample projection data of different classes do not overlap. Secondly, the final expression of regression estimation function does not contain the parameter b by bring b into the solution procedure of the original problem of support vector machine optimization problems, which simplifies the calculation. A robust loss function and the wavelet kernel function were designed to suppress various noise and singular points of complex data samples, which has very strong robustness and practicality. Finally, the example analysis was carried out to the classical Tennessee Eastman chemical process, and the results show that, compared with common fault diagnosis method, the improved method has better fault diagnosis speed and accuracy, which can reflect the actual situation of chemical process.

Key words: chemical process; kernel principal component analysis; support vector machine; fast fault diagnosis

0 引言

化工生产过程往往伴随着物理、化学、生化反应以及相变等多种复杂变化。为了生产的安全,通常都会配备相当数量的测量设备,各种传感器及显示、记录装置。如何从这些传感器及显示、记录装置提供的数据中提取有用的信息,进而辨识和分类生产过程中出现的故障,找出故障原因,使故障状态回到正常状态,对于确保生产的安全和高效,具有十分重要的意义。

目前,对于化工过程智能监测、故障诊断和分类的研究方法,主要包括基于解析模型的方法、基于定性知识的方法、基于信号处理的方法和基于数据驱动的方法^[1]。其中核多元统计分析方法具有不需要精确的数据模型,且可用于处理高维相关数据的特点,已经成为研究和应用最多的一种基于数据驱动的化工过程故障诊断方法。Scholkopf 等^[2]提出的核函数主元分析法能有效地提取故障特征,实现故障的检测。Cho^[3]对非线性系统中的故障识别进行了研究,但存在难以直接诊断故障源的问题。曹惠玲等^[4]提出将可靠性数据应用于民机故障诊断的计算模型,通过计算综合故障概率,实现了故障源的精确定位。邵辉等^[5]提出用关联维数的计算方法进行参数选择,可诊断出设备实测信号属于外圈故障的状态信号。Baudat 等^[6]提出先通过非线性核函数把输入空间映射到特征空间,然后在特征空间实现 FDA,找到判别特征向量进行分类的方法。但在采样数据较大时,由于核矩阵计算量大,导致诊断过程非常耗时,影响了实用性。薄翠梅等^[7]将 KPCA 与 PNN 结合对故障进行分类识别,分类正确率高,但对于故障诊断的小样本实际问题,其泛化能力还有待提高。马成正等^[8]结合故障模式影响分析方法和径向基函数神经网络对罐车进行故障诊断,也取得了较好的诊断效果。支持向量机(SVM)由于遵循了结构风险最小化原理,可将非线性问题转化为线性问题并能得到全局最优解,在小样本下具有良好的泛化能力,近年来在化工过程故障诊断领域得到了广泛的应用^[9]。将核主元分析方法和 SVM 结合能充分发挥各自的优势,提高故障诊断的实时性,适合于复杂化工过程的在线故障诊断^[10]。

由于化工过程的故障大都具有突发性,有时会在瞬间蔓延,导致安全事故。因此,化工过程故障诊断的实时性十分重要。本文首先利用近邻边界法(Neighbour Boundary Method, NBM)对 KFDA 方法进行监督降维,实现核主元的快速辨识和提取,并将提取到的有效特征作为支持向量机分类器的输入。然后构建一种结合鲁棒损失函数和小波核函数的支持向量机故障诊断模型,并在构建过程中对模型进行简化,减少了模型参数,缩短了模型的求解时间。最后设计一种基于改进的 KFDA 和 RW η -SVM 集成的多类故障分类算法,用于对故障进行诊断和分类,并应用到经典的 Tennessee Eastman Process (TEP)化工过程实时故障诊断中。

1 核特征向量提取

1.1 KFDA 核主元分析

KFDA 方法首先通过非线性映射函数把输入向量 \mathbf{x} 映射到高维特征空间 $\mathbf{j}(\mathbf{x})$, 然后在高维特征空间用线性 FDA 方法进行主元分析。设输入训练样本向量 $\mathbf{x} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_t\} \in \mathbf{R}^n$, t_j 为第 $j(j=1, 2, \dots, c)$ 类样本 c_j 的数量。样本向量 \mathbf{x} 经非线性映射 $\mathbf{j} : \mathbf{R}^n \rightarrow F$ 后的类间离散度矩阵和类内离散度矩阵分别为:

$$S_b = \sum_{j=1}^c (\bar{\mathbf{x}}^{(j)} - \bar{\mathbf{x}})(\bar{\mathbf{x}}^{(j)} - \bar{\mathbf{x}})^T \quad (1)$$

$$S_w = \sum_{j=1}^c \sum_{\mathbf{x} \in c_j} (\mathbf{j}(\mathbf{x}) - \bar{\mathbf{x}}^{(j)})(\mathbf{j}(\mathbf{x}) - \bar{\mathbf{x}}^{(j)})^T \quad (2)$$

其中， $\bar{\mathbf{x}}^{(j)}$ 为映射变换后的第 j 类样本的均值向量，即 $\bar{\mathbf{x}}^{(j)} = \frac{1}{t_j} \sum_{\mathbf{x} \in C_j} \mathbf{j}(\mathbf{x})$ ， $\bar{\mathbf{x}}$ 为映射

变换后所有样本均值向量，即 $\bar{\mathbf{x}} = \frac{1}{t} \sum_{l=1}^t \mathbf{x}_l$ ， c 为类别数。KFDA 的任务就是寻找最优核判别向量 \mathbf{m} ，使得：

$$\max \quad J(\mathbf{m}) = \mathbf{m}^T \mathbf{S}_b \mathbf{m} / \mathbf{m}^T \mathbf{S}_w \mathbf{m} \quad (3)$$

然后通过求解下式的特征向量所对应的前 d 个最大特征值 \mathbf{l} 可得最优投影判别向量 \mathbf{m} 。

$$\mathbf{S}_b \mathbf{m} = \mathbf{l} \mathbf{S}_w \mathbf{m} \quad (4)$$

1.2 改进 KFDA 核主元分析 (Improved Kernel Fisher Discriminant Analysis, IKFDA)

KFDA 方法通过核函数的引进，找到最优投影判别向量 \mathbf{m} ，解决了化工过程故障数据的非线性问题，但由于增加了投影后的数据空间维数，容易形成维数灾。同时，由于化工过程故障数据的复杂性，因此在对样本数据进行投影计算过程中，会出现部分异类数据的距离较近，导致分离效果不佳的情况，影响了故障诊断的效率和正确性。因此，有必要对 KFDA 方法做进一步改进。

对于样本数据集 $\mathbf{x} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_t\} \in \mathbf{R}^n$ 中第 j 类样本向量 $\mathbf{x}_l^{(j)}$ ($l=1, 2, \dots, \mathbf{K}, t$)，根据 NBM 法^[11]，令 $\Delta_k \in \mathbf{x}$ 为 $\mathbf{x}_l^{(j)}$ 同类别的 k 个近邻样本数据点集合，分别计算 $\mathbf{x}_l^{(j)}$ 邻域内与 $\mathbf{x}_l^{(j)}$ 同类别的距离最近的近邻点和不同类别的最近的近邻点，

$$\mathbf{r}_l = \{\mathbf{x} \mid \|\mathbf{x} - \mathbf{x}_l^{(j)}\| \geq \|\Omega - \mathbf{x}_l^{(j)}\|, \forall \Omega, \mathbf{x} \in \Delta_k\} \quad (5)$$

其中 \mathbf{x} 和 Ω 为和 $\mathbf{x}_l^{(j)}$ 同类别的样本数据向量。

$$\mathbf{s}_l = \{\mathbf{x} \mid \|\mathbf{x} - \mathbf{x}_l^{(j)}\| \leq \|\Omega - \mathbf{x}_l^{(j)}\|, \forall \Omega, \mathbf{x} \notin \Delta_k\} \quad (6)$$

其中 \mathbf{x} 和 Ω 为和 $\mathbf{x}_l^{(j)}$ 不同类别的样本数据向量。结合 NBM 法重新构造投影后的类间距离离散矩阵和类内距离离散矩阵：

$$\mathbf{S}_b = \sum_{l=1}^t [\mathbf{j}(\mathbf{x}_l) - \mathbf{j}(\mathbf{x}_l^{\mathbf{s}})] [\mathbf{j}(\mathbf{x}_l) - \mathbf{j}(\mathbf{x}_l^{\mathbf{s}})]^T \quad (7)$$

$$\mathbf{S}_w = \sum_{l=1}^t [\mathbf{j}(\mathbf{x}_l) - \mathbf{j}(\mathbf{x}_l^{\mathbf{r}})] [\mathbf{j}(\mathbf{x}_l) - \mathbf{j}(\mathbf{x}_l^{\mathbf{r}})]^T \quad (8)$$

由于 $\mathbf{j}(\mathbf{x}_l)$ 是样本数据 \mathbf{x}_l 到高维空间的映射，根据核函数相关性质，有

$\mathbf{m} = \sum_{l=1}^t a_l \mathbf{j}(\mathbf{x}_l)$ ，进一步可得：

$$\mathbf{m}^T \cdot \mathbf{j}(\mathbf{x}_p) = \sum_{l=1}^t a_l \mathbf{j}(\mathbf{x}_l)^T \mathbf{j}(\mathbf{x}_p) = \mathbf{a}^T \mathbf{K}_{l \times p} \mathbf{q}_{p \times 1} \quad (9)$$

其中 $\mathbf{a} = [a_1, a_2, \mathbf{K}, a_t]^T$, $p = 1, 2, \mathbf{K}, t$, \mathbf{q} 为所有元素均为 1 的向量或矩阵。核矩阵 $\mathbf{K}_{l \times p}$ 是以 $\mathbf{K}(\mathbf{x}_l, \mathbf{x}_p)$ 为元素的矩阵, $\mathbf{K}(\mathbf{x}_l, \mathbf{x}_p)$ 为 $\mathbf{j}(\mathbf{x}_l)$ 与 $\mathbf{j}(\mathbf{x}_p)$ 的内积, 即有 $\mathbf{K}(\mathbf{x}_l, \mathbf{x}_p) = \langle \mathbf{j}(\mathbf{x}_l), \mathbf{j}(\mathbf{x}_p) \rangle$ 。同理可知, $\mathbf{K}_{l \times \hat{s}}$ 是以 $\mathbf{K}(\mathbf{x}_l, \hat{\mathbf{x}}_p)$ 为元素的矩阵, $\mathbf{K}(\mathbf{x}_l, \hat{\mathbf{x}}_p) = \langle \mathbf{j}(\mathbf{x}_l), \mathbf{j}(\hat{\mathbf{x}}_p) \rangle$ 。

将式 (7) ~ 式 (9) 分别代入式 (1) 和式 (2), 可得

$$\mathbf{m}^T \hat{\mathbf{S}}_b \mathbf{m} = \mathbf{a}^T (\mathbf{K}_{l \times p} \mathbf{q}_{p \times 1} - \mathbf{K}_{l \times \hat{s}} \mathbf{q}_{\hat{s} \times 1}) (\mathbf{K}_{l \times p} \mathbf{q}_{p \times 1} - \mathbf{K}_{l \times \hat{s}} \mathbf{q}_{\hat{s} \times 1})^T \mathbf{a} \quad (10)$$

$$\mathbf{m}^T \hat{\mathbf{S}}_w \mathbf{m} = \mathbf{a}^T (\mathbf{K}_{l \times p} \mathbf{q}_{p \times 1} - \mathbf{K}_{l \times \hat{r}} \mathbf{q}_{\hat{r} \times 1}) (\mathbf{K}_{l \times p} \mathbf{q}_{p \times 1} - \mathbf{K}_{l \times \hat{r}} \mathbf{q}_{\hat{r} \times 1})^T \mathbf{a} \quad (11)$$

则新的基于 NBW 的 IKFDA 最优判别式为:

$$\max \quad J(\mathbf{a}) = \frac{\mathbf{m}^T \hat{\mathbf{S}}_b \mathbf{m}}{\mathbf{m}^T \hat{\mathbf{S}}_w \mathbf{m}} = \frac{\mathbf{a}^T (\mathbf{K}_{l \times p} \mathbf{q}_{p \times 1} - \mathbf{K}_{l \times \hat{s}} \mathbf{q}_{\hat{s} \times 1}) (\mathbf{K}_{l \times p} \mathbf{q}_{p \times 1} - \mathbf{K}_{l \times \hat{s}} \mathbf{q}_{\hat{s} \times 1})^T \mathbf{a}}{\mathbf{a}^T (\mathbf{K}_{l \times p} \mathbf{q}_{p \times 1} - \mathbf{K}_{l \times \hat{r}} \mathbf{q}_{\hat{r} \times 1}) (\mathbf{K}_{l \times p} \mathbf{q}_{p \times 1} - \mathbf{K}_{l \times \hat{r}} \mathbf{q}_{\hat{r} \times 1})^T \mathbf{a}} \quad (12)$$

$$\text{s.t.} \quad \mathbf{a}^T \mathbf{a} = \mathbf{I}$$

进一步, 通过求解 $\hat{\mathbf{S}}_b \mathbf{a} = \lambda \hat{\mathbf{S}}_w \mathbf{a}$ 中特征向量所对应的前 d 个最大特征值 $\lambda_1 \geq \lambda_2 \geq \mathbf{K} \geq \lambda_d$,

可得特征向量 $\mathbf{a} = (a_1, a_2, \dots, a_d)$ 。当有新的数据 \mathbf{x}_n 输入时, 只需将其做投影映射

$\mathbf{m}^T \cdot \mathbf{j}(\mathbf{x}_n) = \mathbf{a}^T \mathbf{K}_{l \times n} \mathbf{q}_{n \times 1}$, 而无需对整个样本数据 \mathbf{x} 重新计算。可见, 改进的 KFDA 方法大大减小了计算规模, 缩短了时间, 可应用性得到了提高。

与主元分析方法一样, 基于 IKFDA 主元分析也用 Hotelling T^2 统计量和 Q (SPE) 统计量进行过程监控, 其中 T^2 统计量的控制限可由 F 分布计算得到, Q 统计量的控制限可由近似分布得到。

1.3 小波核函数

小波变换就是通过母小波函数在时域内对信号进行一系列多尺度的变换, 通过母小波函数所进行的缩放和跳转操作, 可以在多尺度下提取到样本数据的局部信息。根据平移不变核理论, 平移不变核函数 $\mathbf{K}(\mathbf{x}_l, \mathbf{x}_l') = \mathbf{K}(\mathbf{x}_l - \mathbf{x}_l')$ 是一个允许支持向量核, 则满足平移不变核定理的小波核函数为:

$$\mathbf{K}(\mathbf{x}_l, \mathbf{x}_l') = \prod_{l=1}^t \mathbf{y} \left(\frac{\mathbf{x}_l - \mathbf{x}_l'}{\mathbf{g}_l} \right) \quad (13)$$

其中, $\mathbf{g} = (\mathbf{g}_1, \mathbf{g}_2, \mathbf{L}, \mathbf{g}_t)$ 为 t 维尺度因子列向量, $\mathbf{y}(\cdot)$ 为一个母小波。根据母小波函数, 可以得到相应的小波核函数。例如, Morlet 母小波函数为:

$$\mathbf{y}(\mathbf{x}) = \cos(\mathbf{r}_0 \mathbf{x}_l) \exp(-\|\mathbf{x}_l\|^2 / 2) \quad (14)$$

将式 (14) 代入式 (13), 可得 Morlet 小波核函数为:

$$\mathbf{K}(\mathbf{x}_l, \mathbf{x}'_l) = \prod_{l=1}^t \cos\left(r_0 \cdot \frac{\mathbf{x}_l - \mathbf{x}'_l}{g_l}\right) \exp\left(-\frac{\|\mathbf{x}_l - \mathbf{x}'_l\|^2}{2g_l^2}\right) \quad (15)$$

Mexican母小波函数为:

$$\mathbf{y}(\mathbf{x}) = (1 - \|\mathbf{x}_l\|^2) \exp(-\|\mathbf{x}_l\|^2/2) \quad (16)$$

将式 (16) 代入式 (13), 可得Mexican小波核函数为:

$$\mathbf{K}(\mathbf{x}_l, \mathbf{x}'_l) = \prod_{l=1}^t \left(1 - \frac{\|\mathbf{x}_l - \mathbf{x}'_l\|^2}{g_l^2}\right) \exp\left(-\frac{\|\mathbf{x}_l - \mathbf{x}'_l\|^2}{2g_l^2}\right) \quad (17)$$

2 改进鲁棒小波 n -支持向量机 (Improved Robust Wavelet n -Support Vector Machine, IRW n -SVM)

2.1 鲁棒损失函数

一般来说, 采集到的数据样本集的最优损失函数与样本集数据内在之间存在某种关系。例如, 以高斯函数作为损失函数, 则可以对正态分布的噪音起到较好的降噪效果。若以 Laplace 线性函数作为损失函数, 则可以对幅值较大的噪音和奇异点起到较好的降噪效果。考虑到化工现场采集到的数据样本的随机性和不平稳性, 本文设计一种新的鲁棒损失函数, 新的鲁棒损失函数分为三部分。第一部分用高斯函数作为损失函数, 用来抑制符合高斯特征的噪音; 第二部分为 e 不灵敏区, 在这个区域不惩罚小于 e 的偏差, 使向量机的解具有稀疏性; 第三部分选用 Laplace 损失函数, 用来抑制较大的噪音和一些奇异点。具体新的鲁棒损失函数表达式为:

$$R(\mathbf{x}) = \begin{cases} \frac{1}{2}(|\mathbf{x}| - e)^2 & e + d \geq |\mathbf{x}| > e \\ 0 & e \geq |\mathbf{x}| \\ d(|\mathbf{x}| - e) - \frac{1}{2}d^2 & |\mathbf{x}| > e + d \end{cases} \quad (18)$$

其中 $e \geq 0, d \geq 0$ 。

2.2 鲁棒小波 n -支持向量机

给定训练样本数据集为 $T = \{(\mathbf{x}_l, y_l) | l = 1, 2, \dots, \mathbf{K}, t\}$, 其中 $\mathbf{x}_l \in \mathbf{R}^n$ 为特征提取后的输入样本, y_l 代表不同类。分类的任务就是构造最优平面 $f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{j}(\mathbf{x}) \rangle + b$, 把属于不同类的向量 \mathbf{x}_l 分开, 其中 \mathbf{w} 为参数向量, b 为偏差量, $\mathbf{j}(\mathbf{g})$ 是将原始输入空间映射到一个高维 Hilbert 特征空间的核空间映射函数。采用式 (17) 的小波核函数和 (18) 的鲁棒损失函数, 得到鲁棒小波 n -支持向量机 (RW n -SVM), 从而将构造最优平面 $f(\mathbf{x})$ 函数估计问题转化为求解 RW n -SVM 问题:

$$\min J(\mathbf{w}, C, R_{emp}, b) = \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \cdot R_{emp} + \frac{1}{2} b^2 \quad (19)$$

其中, $R_{emp} = v \cdot e + \frac{1}{t} \sum_{l \in P_1} \frac{1}{2} (\mathbf{x}_l^2 + \mathbf{x}_l^{*2}) + \frac{1}{t} \sum_{l \in P_2} d (\mathbf{x}_l + \mathbf{x}_l^*)$ 为经验风险, $C > 0$ 为惩罚系数, n

为支持向量的控制参数， \mathbf{e} 为管道大小的控制参数； $\mathbf{x}_l, \mathbf{x}_l^*$ ($l=1, \dots, t$) 为保证了约束满足的松弛变量； \mathbf{p}_1 是表示松弛变量落在 $0 < \mathbf{x}_l, \mathbf{x}_l^* < \mathbf{e}_d$ 区间内的样本集， \mathbf{p}_2 是表示松弛变量落在 $\mathbf{e}_d < \xi_l, \xi_l^*$ 区间内的样本集。式 (19) 的约束条件为：

$$\begin{aligned} e_1 &\geq 0, e_2 \geq 0 \\ \mathbf{x}_l, \mathbf{x}_l^* &\geq 0 \\ v &\in (0, 1] \\ \mathbf{e} &\geq 0, l=1, 2, \dots, t \end{aligned} \quad (20)$$

其中 $e_1 = \mathbf{e} + \mathbf{x}_l - y_l + \mathbf{w} \cdot \mathbf{x}_l + b \geq 0$ ， $e_2 = \mathbf{e} + \mathbf{x}_l^* + y_l - \mathbf{w} \cdot \mathbf{x}_l - b \geq 0$ 。

2.3 改进小波鲁棒 n -支持向量机

在上述向量机的回归函数的求解过程中，如何选择恰当的参数 \mathbf{b} 目前还没有有效的办法，其精确确定比较困难，这直接导致向量机模型参数在寻优过程中消耗了大量迭代计算时间，最终降低了整个模式分类的计算效率。为此，本文在建立最优化支持向量机模型的时候，将参数 \mathbf{b} 纳入最优化问题的原问题的求解中。

引入 Lagrange 乘子 $\mathbf{b}, \mathbf{h}_l, \mathbf{h}_l^*, \mathbf{a}_l, \mathbf{a}_l^* \geq 0$ ，根据目标函数式 (19) 和约束条件式 (20) 建立如下 Lagrange 函数：

$$L(\mathbf{w}, C, \mathbf{a}, \mathbf{a}^*, \mathbf{h}, \mathbf{h}^*, \mathbf{x}, \mathbf{x}^*, \mathbf{b}, \mathbf{e}, b) = J(\mathbf{w}, C, R_{emp}, b) - \mathbf{b}\mathbf{e} - \sum_{l=1}^t (\mathbf{h}_l \mathbf{x}_l + \mathbf{h}_l^* \mathbf{x}_l^* - \mathbf{a}_l e_1 - \mathbf{a}_l^* e_2) \quad (21)$$

分别对 $\mathbf{w}, \mathbf{b}, \mathbf{e}, \mathbf{x}_l, \mathbf{x}_l^*$ 求偏导数，并令它们等于 0，可得 $\mathbf{w} = \sum_{l=1}^t (\mathbf{a}_l - \mathbf{a}_l^*) \mathbf{x}_l$ ，

$\mathbf{b} = \sum_{l=1}^t (\mathbf{a}_l - \mathbf{a}_l^*)$ ，然后利用对偶原理和 KKT 条件，可得原问题的对偶最优问题，

$$\min J(\mathbf{a}, \mathbf{a}^*) = \frac{1}{2} [\mathbf{a}^*, \mathbf{a}]^T \begin{bmatrix} \mathbf{K}_{l \times p} + t \cdot \mathbf{q}_{t \times t} & -\mathbf{K}_{l \times p} \\ -\mathbf{K}_{l \times p} & \mathbf{K}_{l \times p} + t \cdot \mathbf{q}_{t \times t} \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ \mathbf{a}^* \end{bmatrix} \quad (22)$$

其中 $\mathbf{q}_{t \times t}$ 为 $t \times t$ 维单位矩阵， $\mathbf{K}_{l \times p}$ 是以 $\mathbf{K}(\mathbf{x}_l, \mathbf{x}_p)$ 为元素的矩阵， $\mathbf{K}(\mathbf{x}_l, \mathbf{x}_p) = \langle \mathbf{j}(\mathbf{x}_l), \mathbf{j}(\mathbf{x}_p) \rangle$ 。为加快上式求解速度，可用智能算法（遗传算法，粒子群算法等）对参数 (C, v, a) 进行辨识，所得最优解 \mathbf{a} 和 \mathbf{a}^* 代入回归估计函数 $f(\mathbf{x})$ ，核函数为式 (17) 所示 Mexican 小波核函数，可得到支持向量机的输出：

$$f(\mathbf{x}) = \sum_{l=1}^t (a_l - a_l^*) \left(\prod_{p=1}^l \left(1 - \frac{\|x_p - x_p'\|^2}{g_l^2} \right) \exp \left(-\frac{\|x_p - x_p'\|^2}{2g_l^2} \right) + 1 \right) \quad (23)$$

其中 $x_p \in \mathbf{x}$ 为输入样本数据的第 p 个分量, x_p' 为输入训练样本的第 p 个分量。从上述分析可知, 通过在新模型中减少一个对偶问题的约束条件, 使参数 \mathbf{b} 将不出现在最终的决策回归函数中, 这样整个求解计算过程就无需对参数 \mathbf{b} 进行辨识, 从而加快了计算速度, 提高了计算效率。经过这个模型转换过程, 使简化后的决策回归函数中不出现参数 \mathbf{b} 。如此在最优函数的寻优过程中, 将省略对参数 \mathbf{b} 的估计和计算, 简化了计算流程, 提高模式分类的计算速度和效率。

2.4 基于 IKFDA 和 IRWn-SVM 集成的多类分类算法

根据上述结果, 建立基于 IKFDA 和 IRWn-SVM 集成的多类分类算法:

- 1) 获取化工过程正常运行状态下的样本数据, 用 IKFDA 方法离线提取正常样本数据的非线性特征核主元;
- 2) 计算正常样本数据的 T^2 或 Q , 并计算相应的控制限 (对应于 99% 置信限);
- 3) 获取故障状态下的数据作为训练样本, 用 IKFDA 方法离线提取故障样本数据的非线性特征核主元, 并用该核主元离线训练 IRWn-SVM, 获得 IRWn-SVM 模型, 对 IRWn-SVM 求解, 建立相应的多 IRWn-SVM 分类器;
- 4) 采集实时数据作为测试数据样本, 用 IKFDA 方法在线提取实时样本数据的非线性特征核主元;
- 5) 计算测试数据样本的 T^2 或 Q ;
- 6) 判断测试数据 T^2 或 Q 的控制限是否超过正常数据的控制限。若超过, 则将测试数据样本的核主元送入到多 IRWn-SVM 分类器, 进行故障诊断和分类;
- 7) 输出故障信息。

3 应用实例

田纳西-伊士曼过程 (Tennessee-Eastman Process, TEP) 是一个典型的复杂多变量化工生产过程, 具有变量众多、强耦合、非线性且具有不确定性, 是化工过程监测和控制领域的 Benchmark。本文以 TEP 故障诊断问题为背景, 对所研究的方法进行验证。TEP 的流程是从四种反应物中生产两种产品, 另外还有一种添加物和一种副产品, 生产过程包括 5 个单元: 反应器、冷凝器、压缩机、分离器 (闪蒸汽) 和汽提塔, 生产过程涉及 8 种成分, 52 个观测变量, 其中 41 个为测量变量 (XMEAS1~XMEAS41), 11 个为控制变量 (XMV1~XMV12), 21 类故障 (IDV1~IDV15)。仿真程序在 Matlab R2007a 平台中编制, 运行环境为 Core(TM) i3-2120 CPU 3.30GHz、内存 2G 个人计算机。

3.1 核主元分析

数据采样间隔为 3 分钟，总共运行 48 个小时，共获得 960 个数据。图 1 为部分样本在 48 个小时期间采集到的观测值的曲线图，(a)和(b)分别为采集到的测量变量 XMEAS8 和 XMEAS10 的观测值变化曲线。

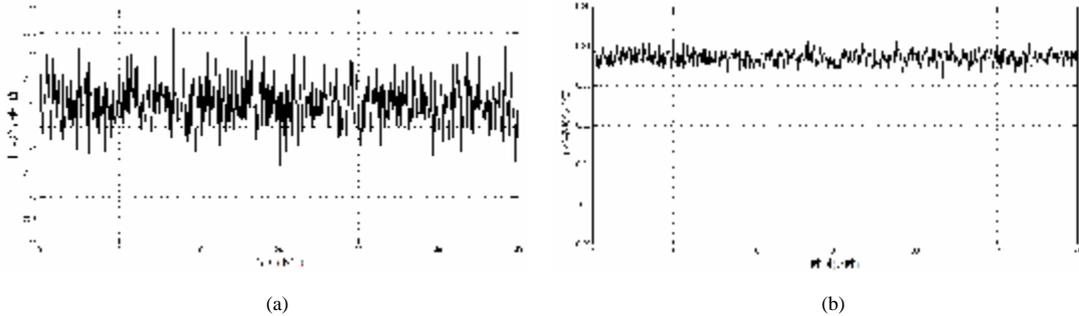


图 1 部分样本数据观测值变化曲线

Fig. 1 Trend graph for part of samples from the real TEP process data

TEP 的故障类型分为 3 种，分别为 0、1、2，其中 0 表示正常工况，1 和 2 分别为故障 1 和故障 2。小波核函数采用式(15)的 Morlet 小波核函数形式，采用离线计算方式通过 IKFDA 模型对不同样本组进行多次计算，确定小波核函数的核参数和主元数量。当核参数 $g = 4$ 时，IKFDA 可以达到较好的降维效果。表 1 显示了用 IKFDA 模型分别对两类故障数据进行特征提取并按特征值排序选取关键变量的步骤，其中 t-检验的拒绝域取 10%。

表 1 IKFDA 特征向量选取结果

Table 1 The result of eigenvector selection with IKFDA

步骤	IKFDA 特征向量选取	分类正确率	均值差	10%拒绝域
1	44, 50, 47, 1	故障 1 0.983 078 13	0.004 287	0.002 438
		故障 2 0.945 484 38	0.044 353	0.015 333
2	19, 50, 47, 1, 10, 44	故障 1 0.980 681 25	0.046 957	0.005 249 6
		故障 2 0.902 151 46	0.010 437	0.023 126 2
3	44, 50, 47, 1, 10	故障 1 0.984 677 08	-0.000 84	0.013 970 4
		故障 2 0.944 755 21	0.001 61	0.006 672 5
4	44, 47, 1	故障 1 0.985 395 83	0.000 879	0.013 944 9
		故障 2 0.968 708 33	0.023 454	0.006 162 3

由表 1 结果，进一步计算累积方差贡献率并以累积方差贡献率 $\geq 90\%$ 为标准，确定核主元数为 6 (分别为 XMEAS19, XMV50、XMV47、XMEAS1、XMEAS10 和 XMV44)，相应的控制限为 21.46，对应于 99% 的置信限。以故障 1 状态为例，在故障 1 状态下，获得 960 组测试数据用于过程在

线故障诊断，图 2 对比了本文 IKFDA 方法与 KFDA 方法用于在线故障 1 诊断的过程，图中虚线表示控制限。

从图 2 可以看出，基于 IKFDA 的化工过程监控要优于传统的 KFDA 方法。在控制限确定的情况下，IKFDA 方法的误判率较低，基本上达到 100% 的识别率。相反，KFDA 方法的误判率较高，有时还有无法判断故障的情况。因此，本文提出的基于 NBM 和小波核函数的 IKFDA 模型可有效对 TEP 过程进行故

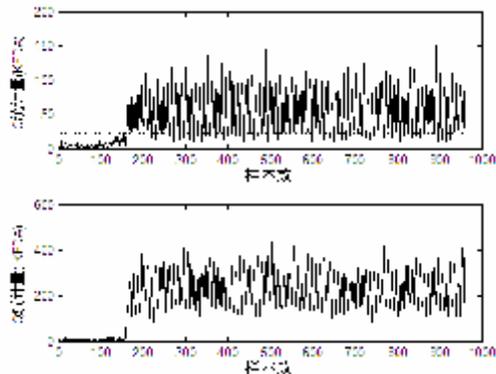


图 2 故障 1 的 Q 统计量过程监控图 (P165, P124)

Fig. 2 Process monitoring graph based on Q statistic of fault 1

障诊断。根据各类型样本数据的统计量变化，判断生产过程是否出现故障，如果出现故障，可进一步通过支持向量机对故障进行识别分类。

3.2 多类 RWn -SVM 故障诊断

每个变量在每一种状态下各取 30 组数据样本，经 IKFDA 主元分析后得到 180 组数据样本用于支持向量机训练和测试，180 组训练数据样本中 6 个核主元的样本数分别为 30。支持向量机的分类方式为“1 对多”算法，为此需建立 3 个向量机，分别为 SVM1~SVM3。其中用于 SVM1 训练的样本为 30 个 (1~30)，测试样本 29 个 (31~59)；用于 SVM2 训练的样本为 36 个 (60~95)，测试样本为 45 个 (96~130)；用于 SVM3 训练的样本为 24 个 (131~154)，测试样本为 26 个 (155~180)。选择式 (16) 的 Mexican 母小波函数为核函数，经对比分析，当 $C = 10000$ ， $n = 0.9$ ， $g = 4$ 时能达到较好的分类效果。图 3 显示了对测试变量的分类效果。

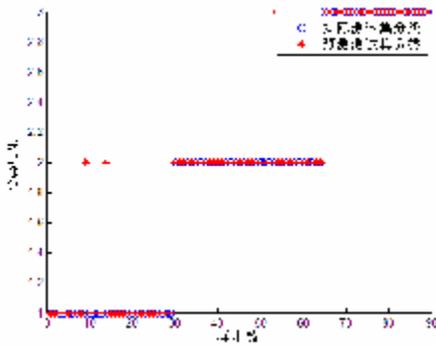


图 3 对不同变量的故障分类效果

Fig. 3 The effect of fault classification to different variables

从图 3 可以看出，整个分类过程只有 3 个点的预测分类结果和实际分类不同，分类的正确率达到了 98.33%。为了进一步验证本文提出的基于 IKFDA 的 RWn -SVM 的有效性，利用 TEP 过程已知故障的检验样本进行验证，每种状态取 100 个样本，共形成 300 个测试样本，同时用标准 RWn -SVM 和本文提出的 RWn -SVM 方法进行测试和对比，结果如表 2 所示：

由表 2 可以看出，本文提出的基于 IKFDA 和 $IRWn$ -SVM 集成方法的诊断结果与普通的基于 KFDA 的 RWn -SVM 相比较，本文提出的方法具有更高的故障识别率，其

诊断性能也优于标准 RWn -SVM。另外，本文提出的基于 IKFDA 和 $IRWn$ -SVM 的诊断方法在故障诊断的计算时间上，也比普通 RWn -SVM 的计算时间短，基本上，在识别率大致相同的情况

表 2 故障诊断分类结果

Table 2 The result of fault diagnosis and classification

故障类型	检验样本数量	分类结果(100%)		诊断结果对比(诊断时间/s)	
		训练结果	测试结果	KFDA- RWn -SVM	IKFDA- $IRWn$ -SVM
故障 0	100	100	100	90(30.2)	100(15.7)
故障 1	100	96.3	95	90(35.6)	95(18.4)
故障 2	100	93.3	85	75(35.9)	85(18.8)

下，本文提出的 $IRWn$ -SVM 的训练速度比标准 RWn -SVM 的训练速度快 2-3 倍，这得益于本文在建模过程中，分别对 KFDA 和 RWn -SVM 模型进行了优化，特别是在样本数量比较大时，计算速度的优势尤为明显，这对实时在线故障诊断是非常关键的，也正是本文所提出方法的关键所在。

4 结论

本文针对化工过程变量多，相互关系复杂，不易进行在线实时故障诊断的实际问题，提出基于改进的 KFDA 核主元分析和结合鲁棒损失函数和小波核函数的多支持向量机的故障诊断模型，模型具有以下特点：

- 1) 结合了核函数主元分析与多支持向量机技术各自的优点，所建故障诊断模型具有适

应小样本、学习能力强的特点。

2) 对 KFDA 方法和支持向量机模型进行了改进, 降低核主元分析过程中样本数据的维数, 减少了求解支持向量机过程所需估算参数的数量, 缩短了故障诊断计算时间, 利于实际使用。

3) 对样本数据的测试和预测结果显示, 本文所建立的模型具有泛化能力好, 抗噪性强的特点, 能快速反应化工生产过程故障的出现, 在化工过程的监控领域具有较强的使用前景和应用价值。

参考文献

- [1] 于春梅,张洪才. 工业过程故障诊断方法的研究进展[J]. 合肥工业大学学报, 2008,31(1): 13-16.
YU Chun-mei, ZHANG Hong-cai. Advances of fault diagnosis methods in industrial processes[J]. Journal of Hefei University of Technology, 2008,31(1): 13-16.
- [2] Scholkope R, Smola A J, Muller K. Nonlinear Component Analysis as a Kernel Eigenvalue Problem[J]. Neural Computation, 1998(10): 1299-1319.
- [3] Choi S W, Lee C, Lee J M. Fault Detection and Identification of Nonlinear Processes Based on Kernel PCA[J]. Chemometrics and Intelligent Laboratory Systems, 2005, 75(1):55-67.
- [4] 曹惠玲,赵明,曲春刚,薛鹏,王凯. 可靠性数据在民机故障诊断中的应用研究[J]. 中国安全科学学报, 2011,21(10): 95-101.
CAO Hui-ling, ZHAO Ming, QU Chun-gang, XUE Peng. Research on Application of Reliability Data in Fault Diagnosis of Civil Aircraft [J]. China Safety Science Journal, 2011,21(10):95-101.
- [5] 邵辉,赵庆贤,李燕. 关联维数在机械设备故障诊断中的应用研究[J]. 中国安全科学学报, 2006,16(3):129-134.
HAO Hui, ZHAO Qing-xian, LI Yan. Study and Application of Correlation Dimension in Fault Diagnosis of Machinery[J]. China Safety Science Journal, 2006,16(3):129-134.
- [6] Baudat G, Anouar F. Generalized Discriminant Analysis Using a Kernel Approach[J]. Neural Computation, 2000, 12(10):2385-2404.
- [7] 薄翠梅,王执铨,张广明. 基于KPCA-PNN的复杂工业过程集成故障辨识方法[J]. 信息与控制, 2009, 38(1):98-109.
BO Cui-mei, WANG Zhi-quan, ZHANG Guang-ming. An Integrated Fault Identification Method Based on KPCA-PNN for Complex Industrial Process[J]. Information and Control, 2009, 38(1):98-109.
- [8] 马成正,王洪德. 基于FMEA与RBF神经网络的LPG汽车罐车储罐系统故障诊断[J]. 中国安全科学学报, 2011,21(1):99-104.
MA Cheng-zheng, WANG Hong-de. Fault Diagnosis of LPG Tank Car Based on FMEA and RBF Neural Network[J]. China Safety Science Journal, 2011,21(1):99-104.
- [9] 于德介,陈淼峰,程军圣. 基于AR模型和支持向量机的转子系统故障诊断方法[J]. 系统工程理论与实践, 2007, 27(5):152-157.
Yu De-jie, Chen Miao-feng, Cheng Jun-sheng. Fault Diagnosis Approach for Rotor System Based on AR Model and Support Vector Machine[J]. Systems Engineering-Theory & Practice, 2007, 27(5):152-157.
- [10] 刘爱伦,袁小艳,俞金寿. 基于KPCA-SVC的复杂过程故障诊断[J]. 仪器仪表学报, 2007, 28(5):868-872.
Liu Ai-lun, Yuan Xiao-yan, Yu Jin-shou. Fault Diagnosis of Complex Chemical Process Based on KPCA-SVC[J]. Chinese Journal of Scientific Instrument, 2007, 28(5):868-872.
- [11] 魏莱,王守觉,徐菲菲,王睿智. 近邻边界Fisher判别分析[J]. 电子与信息学报, 2009, 31(3):509-513.
Wei Lai, Wang Shou-jue, Xu Fei-fei. Neighborhood Margin Fisher Discriminant Analysis[J]. Journal of



作者简介：王斌（1974-），男，博士，主要从事自动控制工程与优化理论方面的研究。E-mail: wbsky@163.com.



作者简介：匡蕾（1970-），女，高级工程师，江苏省安全生产科学研究院化工安全与信息研究所所长，主要从事化工安全与信息化理论方面的研究工作。E-mail: kuangleinj@163.com.