

基于马氏距离的改进核 Fisher 化工故障诊断研究

吕鹏飞¹ 闫云聚¹ 荔越¹

摘要 针对化工故障诊断数据存在非线性分布、数据类别复杂、数据量大且故障特征不易区分等问题,本文提出一种基于马氏距离的改进核 Fisher 故障诊断方法(Mahalanobis distance-based kernel Fisher discrimination, MKFD)。首先,针对数据非线性分布的特点,本文将核 Fisher 判别分析算法改进,改进后的算法可以有效解决原始样本在投影后出现的因类间距离差异过大、类内距离不够紧凑造成的样本混叠现象。除此之外,利用 Euclidean 距离对类间距做加权处理时,用组平均距离取代质心距离,提升了运算效率,降低了时间复杂度;其次,根据高斯径向基核函数(Radial basis function, RBF)在 MKFD 中所呈现出的诊断精度的规律,本文采用一种新的核参数选择方法:区间三分法,用以取代在实际应用中依靠经验的交叉验证法;最后,本文采用马氏距离对故障进行分类,基于田纳西伊—斯特曼过程(Tennessee-Eastman, TE)数据将本方法与其他改进核 Fisher 算法进行仿真验证对比。结果表明新提出 MKFD 算法不仅可以提高故障诊断的运算效率,也能有效提高诊断的精度。

关键词 核 Fisher, 故障诊断, 区间三分法, TE 过程, 优化

引用格式 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. 自动化学报, 2020, 46(11): 2379–2391

DOI 10.16383/j.aas.c190635

Research on Fault Diagnosis of Improved Kernel Fisher Based on Mahalanobis Distance in the Field of Chemical Industry

LV Peng-Fei¹ YAN Yun-Ju¹ LI Yue¹

Abstract Aiming at the problems of the non-linear distribution, complex category, large amount of fault diagnosis data in chemical industry and the difficulty of distinguishing fault features, a improved kernel Fisher fault diagnosis method based on Mahalanobis distance is proposed in this paper. Firstly, due to the data with non-linear property, a new improved kernel Fisher discriminant analysis method is proposed, which can effectively solve the sample aliasing phenomenon caused by large difference between classes and insufficient compact distance between classes after projection of original samples. In addition, using the Euclidean distance in class spacing, the group average distance is used to replace the center of mass distance, which improves the efficiency of operation and reduces the time complexity. Secondly, according to the rule of diagnostic accuracy presented by the (RBF) in Fisher discriminant analysis (MKFD), a new method, interval “three-point method”, of selecting nuclear parameters is proposed in this paper, which is used to replace the cross-validation method relying on experience in practical application. Finally, faults are classified based on Mahalanobis distance using Tennessee-Sterman process. The proposed method is compared with other improved kernel Fisher algorithm. The results show that (MKFD) can not only improve the calculation efficiency of fault diagnosis, but also improve the accuracy of diagnosis.

Key words Kernel Fisher, fault diagnosis, interval three-point method, Tennessee-Sterman process, optimization

Citation Lv Peng-Fei, Yan Yun-Ju, Li Yue. Research on fault diagnosis of improved kernel Fisher based on Mahalanobis distance in the field of chemical industry. *Acta Automatica Sinica*, 2020, 46(11): 2379–2391

随着科技的不断进步,现代化工业正发生着翻天覆地的变化,工厂操作的复杂度急剧增加。与此同时,操作过程中测量的变量更加具有复杂性、非线

性和巨量性,而这些测量变量的新性质导致了监测平台操作挑战性不断升级^[1]。在化工过程中,监测系统产生了大量的实测数据,对这些数据的有效利用来实现实时监控和故障诊断,为确保生产设备安全、降低维护成本、提高利润率提供了可靠的保证^[2]。因此,高效的故障诊断技术正在工业生产的发展中扮演着愈来愈重要的角色。

在化工生产过程中,需要建立监测平台对整个操作过程进行监控,并利用监测平台得到的数据进行故障诊断,从而确保化工产业生产设备安全,降低维护成本,提高利润率。化工过程的故障诊断问题可

收稿日期 2019-09-09 录用日期 2020-01-09

Manuscript received September 9, 2019; accepted January 9, 2020

西北工业大学硕士研究生创新创业种子基金(ZZ2019125),陕西省自然科学基金基础研究计划(2019JQ-564)资助

Supported by Graduate Innovative and Creative Seed Foundation of Northwest Polytechnical University (ZZ2019125), and Natural Science Basic Research Plan in Shanxi Province of China (2019JQ-564)

本文责任编辑 穆朝絮

Recommended by Associate Editor MU Chao-Xu

1. 西北工业大学 西安 710129

1. Northwest Polytechnical University, Xi'an 710129

以看作是对故障数据的分类问题,其中包括故障特征提取技术、模式识别技术以及故障分类技术.常见的特征提取方法有:主元分析法^[3](Principal component analysis, PCA)、偏最小二乘法^[4](Partial least squares, PLS)、核熵分析法^[5](Kernel entropy component analysis, KECA)、独立主元法^[6](Independent component analysis, ICA)等.然而,在面向化工的实际应用中,单纯的多元统计分析并没有取得良好的分类效果.这是由于化工数据具有很强的非线性,数据类别复杂、数据量大而且故障特征不易区分的特点^[7].基于以上问题, Mika 等^[8]将核函数引入 Fisher 判别分析算法,得到了一种新的核 Fisher 判别法.该方法在面对复杂的化工过程状态数据时,能有效地解决因非线性造成的分类困难问题,因此得到了广泛的应用.随后,国内外研究者对该算法进行了深入的研究并提出了许多改进措施.张曦等^[9]提出了基于核 Fisher 子空间特征提取的汽轮发电机组过程监控和故障诊断新方法,该方法将原始样本数据从低维非线性空间映射到高维线性空间,再利用线性 Fisher 判别分析算法提取原始样本数据的最优特征矢量,实现了对汽轮发电机组的过程监控.马立玲等^[10]采用 Euclidean 距离对 Fisher 判别准则中的类间距做加权,使得样本数据具有较好的投影效果,然后通过改进 K 近邻算法^[11]和马氏距离^[12-13]算法对数据进行分类,仿真验证结果表明改进后的核 Fisher 方法有效改善了因类间距较小而造成的投影数据混叠现象.

然而,已有的改进核 Fisher 判别分析方法中,仍存在以下问题:1)在投影过程中,类间距差异较大的类别之间出现投影混叠现象;2)因为类内比较分散,数据投影后会出现因类内距不够紧凑而出现的重叠现象;3)在引入高斯径向基核函数(Radial basis function, RBF)时,选取非最优的核参数会直接导致故障诊断性能的降低^[14].为解决以上问题,本文提出了一种基于马氏距离的改进核 Fisher 故障诊断方法(MKFD),该方法采用了区间三分迭代法选取核参数,利用组平均距离取代质心距离降低运算复杂度并通过加权改进类内的距离,有效改善了投影效果.最后,在田纳西伊—斯特曼过程平台上对该方法进行仿真试验验证,并与传统的核 Fisher 诊断方法(Kernel fisher discrimination, KFD)、基于质心距离对类间距进行加权处理的核 Fisher 判别分析(Centroid kernel fisher discriminate, CKFD)和基于 Fisher 判别分析的全局—局部保持投影算法(Fisher discriminant global-local preserving projection, FDGLPP)进行了对比.结果表明,本文提出的方法不仅提高了运算速度,同时也有效提高了故障诊断

的精度,具有很高的工程应用价值.

1 核 Fisher 判别分析

Fisher 判别分析是模式识别中一种非常重要的基于数据降维和分类的方法,是一种有监督的学习算法^[15-17].传统的 KFD 算法在故障诊断领域具有广泛的应用,但是当面对高维非线性数据时,其故障诊断的性能大大降低^[18].核函数的引入能够有效地解决这个问题,使其故障诊断的性能大大提升^[19].核 Fisher 算法的基本理论可表述为:

设在 P 维的原始空间中,样本点有 C 类,样本总数为 N ,记为 $X = \{X_1, X_2, \dots, X_C\}$,第 i ($i = 1, 2, \dots, C$) 个类 ω_i 包含 N_i 个样本记作 $X_i = \{\mathbf{x}_i^1, \mathbf{x}_i^2, \dots, \mathbf{x}_i^{N_i}\}$,其中,每一个向量表示的都是列向量.基于核 Fisher 判别函数先对原始样本数据进行非线性高维映射 $\phi: \mathbf{x} \in R^P \rightarrow \phi(\mathbf{x}) \in H^P$,在高维特征空间 H 中,第 i 类样本映射后的均值记为 \mathbf{m}_i ,所有样本点映射后的均值记为 \mathbf{m} .可得:

$$\mathbf{m}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \phi(\mathbf{x}_i^j) \quad (1)$$

$$\mathbf{m} = \frac{1}{N} \sum_{i=1}^C \sum_{j=1}^{N_i} \phi(\mathbf{x}_i^j) \quad (2)$$

在高维特征空间 H 中,经过非线性变换后的类内离散度 S_W^ϕ 和类间的离散度 S_B^ϕ 分别为:

$$S_W^\phi = \frac{1}{N} \sum_{i=1}^C \sum_{j=1}^{N_i} (\phi(\mathbf{x}_i^j) - \mathbf{m}_i)(\phi(\mathbf{x}_i^j) - \mathbf{m}_i)^T$$

$$S_B^\phi = \frac{1}{N} \sum_{i=1}^C N_i (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T \quad (3)$$

在特征空间 H 中, Fisher 判别准则为:

$$J(\mathbf{w}_\phi) = \max \frac{\mathbf{w}_\phi^T S_B^\phi \mathbf{w}_\phi}{\mathbf{w}_\phi^T S_W^\phi \mathbf{w}_\phi} \quad (4)$$

式中: \mathbf{w}_ϕ 为任一非零列向量. Fisher 判别准则就是要通过优化式(4),找到最优特征矢量.由核函数的特点可知,非线性映射会导致式中的 \mathbf{w}_ϕ 无法被直接计算.此时最优判别矢量 \mathbf{w}_ϕ 可由 $\phi(\mathbf{x}_1), \phi(\mathbf{x}_2), \dots, \phi(\mathbf{x}_N)$ 进行线性表示,即:

$$\mathbf{w}_\phi = \sum_{k=1}^N \alpha_k \phi(\mathbf{x}_k) = \phi \boldsymbol{\alpha} \quad (5)$$

式中, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_N)^T$ 为一个列向量.

$$\begin{aligned} \mathbf{w}_\phi^\top \phi(\mathbf{x}_i) &= \boldsymbol{\alpha}^\top \phi^\top \phi(\mathbf{x}_i) = \\ & \boldsymbol{\alpha}^\top (\phi(\mathbf{x}_1), \phi(\mathbf{x}_2), \dots, \phi(\mathbf{x}_N))^\top \phi(\mathbf{x}_i) = \\ & \boldsymbol{\alpha}^\top (k(\mathbf{x}_1, \mathbf{x}_i), \dots, k(\mathbf{x}_N, \mathbf{x}_i))^\top = \\ & \boldsymbol{\alpha}^\top \boldsymbol{\varphi}(\mathbf{x}_i) \end{aligned} \quad (6)$$

式中, $\boldsymbol{\varphi}(\mathbf{x}_i) = (k(\mathbf{x}_1, \mathbf{x}_i), \dots, k(\mathbf{x}_N, \mathbf{x}_i))^\top$.

在高维特征空间 H 中, 训练集样本类均值向量与总体样本均值向量分别投影到 \mathbf{w}_ϕ 上有:

$$\begin{aligned} \mathbf{w}_\phi^\top \phi(\mathbf{x}_i) &= \boldsymbol{\alpha}^\top \phi^\top \frac{1}{N_i} \sum_{j=1}^{N_i} \phi(\mathbf{x}_i^j) = \boldsymbol{\alpha}^\top \boldsymbol{\mu}_i \\ \mathbf{w}_\phi^\top \phi(\mathbf{x}_0) &= \boldsymbol{\alpha}^\top \phi^\top \frac{1}{N} \sum_{g=1}^N \phi(\mathbf{x}_g) = \boldsymbol{\alpha}^\top \boldsymbol{\mu}_0 \end{aligned} \quad (7)$$

式中:

$$\begin{aligned} \boldsymbol{\mu}_i &= \frac{1}{N_i} \left(\sum_{j=1}^{N_i} \phi^\top(\mathbf{x}_1) \phi(\mathbf{x}_i^j), \dots, \sum_{j=1}^{N_i} \phi^\top(\mathbf{x}_N) \phi(\mathbf{x}_i^j) \right) = \\ & \frac{1}{N_i} \left(\sum_{j=1}^{N_i} k(\mathbf{x}_1, \mathbf{x}_i^j), \dots, \sum_{j=1}^{N_i} k(\mathbf{x}_N, \mathbf{x}_i^j) \right) \end{aligned} \quad (8)$$

$$\begin{aligned} \boldsymbol{\mu}_0 &= \frac{1}{N} \left(\sum_{g=1}^N \phi^\top(\mathbf{x}_1) \phi(\mathbf{x}_g), \dots, \sum_{g=1}^N \phi^\top(\mathbf{x}_N) \phi(\mathbf{x}_g) \right) = \\ & \frac{1}{N} \left(\sum_{g=1}^N k(\mathbf{x}_1, \mathbf{x}_g), \dots, \sum_{g=1}^N k(\mathbf{x}_N, \mathbf{x}_g) \right) \end{aligned} \quad (9)$$

在高维特征空间 H 中 Fisher 准则变为:

$$J(\mathbf{w}_\phi) = \max \frac{\mathbf{w}_\phi^\top S_B^\phi \mathbf{w}_\phi}{\mathbf{w}_\phi^\top S_W^\phi \mathbf{w}_\phi} = \max \frac{\boldsymbol{\alpha}^\top K_B^\phi \boldsymbol{\alpha}}{\boldsymbol{\alpha}^\top K_W^\phi \boldsymbol{\alpha}} \quad (10)$$

式中: K_W^ϕ 和 K_B^ϕ 的表达式为:

$$K_W^\phi = \frac{1}{N} \sum_{i=1}^C \sum_{j=1}^{N_i} (\boldsymbol{\varphi}(\mathbf{x}_i^j) - \boldsymbol{\mu}_i)(\boldsymbol{\varphi}(\mathbf{x}_i^j) - \boldsymbol{\mu}_i)^\top$$

$$K_B^\phi = \frac{1}{N} \sum_{i=1}^C N_i (\boldsymbol{\mu}_i - \boldsymbol{\mu}_0)(\boldsymbol{\mu}_i - \boldsymbol{\mu}_0)^\top$$

所以求解核 Fisher 最佳判别向量 \mathbf{w}_ϕ 的问题就转化为求解式 (10) 的 Fisher 判别函数达到最大值时最佳向量 $\boldsymbol{\alpha}$ 的问题. 求解 $\boldsymbol{\alpha}$ 问题等价于求解广义特征方程:

$$K_B^\phi \boldsymbol{\alpha} = \lambda K_W^\phi \boldsymbol{\alpha} \quad (11)$$

即 Fisher 判别准则转化为对 $(K_W^\phi)^{-1} K_B^\phi$ 的求解特征值和其所对应的特征向量的问题. 为解决非

奇异问题, 将式中 K_W^ϕ 替换为 $K_W^\phi + \mu I$, 其中 μ 为非常小的数值, 通常设为 10^{-7} , I 为单位矩阵^[20].

2 基于马氏距离改进的核 Fisher 故障诊断

核 Fisher 判别法在故障诊断领域应用广泛, 但是在处理大量非线性数据时, 仍存在问题. 第一, 不同类之间的类间距存在较大差异, 会存在数据混叠现象; 第二, 类内距离较分散, 数据投影后不紧凑, 会出现类别重叠现象; 第三, MKFD 中 RBF 核参数的快速选取问题. 针对这些问题, 本文提出了 MKFD 算法.

2.1 改进核 Fisher 判别分析

在许多情况下, 数据类别类间距存在很大差异性, 传统的类间距离散度的矩阵为:

$$S_B^\phi = \frac{1}{N} \sum_{i=1}^C N_i (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^\top \quad (12)$$

类间距差异的权重仅依据不同数据类别样本数在总体样本数中所占比进行计算, 但在实际应用中, 会出现部分数据类别间距较小, 另一部分数据类别间距较大的情况, 类间距较小的类信息被间距较大的类信息覆盖的情况, 从而降低故障诊断的性能^[21]. 为解决上述问题, 可在原先核 Fisher 判别中, 保证类内离散度不变, 采用 Euclidean 距离的函数对类间距进行加权, 重新定义映射后的类间离散度^[22]. 重新定义的类间离散度如下:

$$S_B^\phi = \frac{1}{N^2} \sum_{i=1}^{C-1} \sum_{j=i+1}^C N_i N_j W(d_{ij}) (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^\top \quad (13)$$

式中: d_{ij} 表示类 i 和类 j 的类间距离, \mathbf{m}_i 和 \mathbf{m} 分别代表类 i 和类 j 的平均值, 权重函数 $W(d_{ij})$ 为关于 d_{ij} 的一个函数. 式中:

$$k_{i1,i2} = \langle \phi(\mathbf{x}_i^{i1}), \phi(\mathbf{x}_i^{i2}) \rangle$$

$$k_{j1,j2} = \langle \phi(\mathbf{x}_j^{j1}), \phi(\mathbf{x}_j^{j2}) \rangle$$

$$k_{i1,j1} = \langle \phi(\mathbf{x}_i^{i1}), \phi(\mathbf{x}_j^{j1}) \rangle$$

此时类间距离 d_{ij} 选取的是质心距离^[23]. 然而质心距离并不能充分利用所有样本的信息, 所以该方法并不具有很好的代表性^[24]. 如果把核函数的计算个数作为衡量计算复杂度的标准, 则质心距离的计算复杂度为 $O(N_i^2 + N_j^2 + N_i N_j)$. 本文采取组平均距离作为类间距离 d_{ij} :

$$d_{ij} = \|\mathbf{m}_i - \mathbf{m}_j\| = \sqrt{\frac{\sum_{\mathbf{x}_i \in C_i, \mathbf{x}_j \in C_j} (\phi(\mathbf{x}_i) - \phi(\mathbf{x}_j))(\phi(\mathbf{x}_i) - \phi(\mathbf{x}_j))^T}{N_i N_j}} = \sqrt{\frac{\sum_{\mathbf{x}_i \in C_i} k_{ii} + \sum_{\mathbf{x}_j \in C_j} k_{jj} - 2 \sum_{\mathbf{x}_i \in C_i, \mathbf{x}_j \in C_j} k_{ij}}{N_i N_j}} \quad (14)$$

其时间复杂度为 $O(N_i + N_j + N_i N_j)$, 对比发现, 采用组平均距离作为类间距离可以在保证精度相差不大的情况下大大提高运算效率. 除此之外, 定义

$$a(i) = \frac{d^2}{N_i + \sum_{j=1}^C d_{ij}^*}$$

$$d_{ij}^* = \sqrt{(\phi(\mathbf{x}_{i,j}) - \mathbf{m}_i)(\phi(\mathbf{x}_{i,j}) - \mathbf{m}_i)^T} = \sqrt{(\phi(\mathbf{x}_{i,j}) - \phi(\bar{\mathbf{x}}_i))(\phi(\mathbf{x}_{i,j}) - \phi(\bar{\mathbf{x}}_i))^T} \quad (15)$$

重新定义类内离散度矩阵

$$S_W^\phi = \frac{1}{N} \sum_{i=1}^C \sum_{j=1}^{N_i} a(i) (\phi(\mathbf{x}_i^j) - \mathbf{m}_i)(\phi(\mathbf{x}_i^j) - \mathbf{m}_i)^T \quad (16)$$

式中, K_W^ϕ 和 K_B^ϕ 的表达式变为:

$$\begin{cases} K_W^\phi = \frac{1}{N} \sum_{i=1}^C \sum_{j=1}^{N_i} a(i) (\varphi(\mathbf{x}_i) - \boldsymbol{\mu}_i)(\varphi(\mathbf{x}_i) - \boldsymbol{\mu}_i)^T \\ \varphi(\mathbf{x}_i) = (k(\mathbf{x}_1, \mathbf{x}_i), \dots, k(\mathbf{x}_N, \mathbf{x}_i))^T \end{cases} \quad (17)$$

$$\begin{cases} K_B^\phi = \frac{1}{N^2} \sum_{i=1}^{C-1} \sum_{j=i+1}^C N_i N_j W(d_{ij}) (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^T \\ \boldsymbol{\mu}_i = \frac{1}{N_i} \left(\sum_{j=1}^{N_i} k(\mathbf{x}_1, \mathbf{x}_i^j), \dots, \sum_{j=1}^{N_i} k(\mathbf{x}_N, \mathbf{x}_i^j) \right)^T \end{cases} \quad (18)$$

权重函数的表达式为 $W(d_{ij}) = f(d_{ij})$, 定义 $f(x) = \frac{1}{x^d}$ ($d \in \mathbb{N}$), 即表示权重与类间距离成反比; 对于参数 d , 通常选取 $3 \sim 10$ ^[25]. 求解 α 问题等价于求解广义特征方程:

$$\left(K_W^\phi \right)^{-1} K_B^\phi \alpha = \lambda \alpha \quad (19)$$

此时核 Fisher 判别准则转化为求解 $\left(K_W^\phi \right)^{-1} K_B^\phi$ 的特征值和特征向量的问题.

2.2 区间三分法确定高斯径向基 (RBF) 核参数

在改进的核 Fisher 判别分析方法中, 核函数及核参数的选取至关重要. 经文献 [26] 验证, RBF 核

函数具有良好的分类能力. 在本文中, 选择 RBF 核函数:

$$k(\mathbf{x}, \mathbf{x}_i) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}_i\|^2}{2\sigma^2}\right) \quad (20)$$

式中, 核参数 σ 的取值对 KFD 性能的优劣起着至关重要的作用. 目前常用的核参数选取方法有: 交叉验证法^[27]、经验选择法、梯度下降法^[28] 和 Bayesian 法等.

张小云等^[29] 从理论上证明了在支持向量机 (SVM) 中采用 RBF 核函数的性质, 当 $\sigma \rightarrow 0$ 时, 不管训练样本个数是多少, 如何标类, 高斯核 SVM 都能把它们正确的分开, 但对测试集样本的分类准确率差, 容易出现“过拟合”现象. 当 $\sigma \rightarrow \infty$ 时, 高斯核 SVM 对所有样本从理论上一视同仁, 认为没有学习推广能力. 基于以上结论, 本文对 MKFD 中 RBF 的性质同样进行两种理论假设并进行证明.

假设 1. 在 MKFD 中, 当 $\sigma \rightarrow 0$ 时, 所有的训练样本都能被正确分类. 但此时的 MKFD 没有学习推广能力.

证明. 在 MKFD 中, RBF 核函数为一个 $N \times N$ 的 Gram 矩阵

$$k = \begin{pmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & k(\mathbf{x}_1, \mathbf{x}_2) & \cdots & k(\mathbf{x}_1, \mathbf{x}_N) \\ k(\mathbf{x}_2, \mathbf{x}_1) & k(\mathbf{x}_2, \mathbf{x}_2) & \cdots & k(\mathbf{x}_2, \mathbf{x}_N) \\ \vdots & \vdots & \cdots & \vdots \\ k(\mathbf{x}_N, \mathbf{x}_1) & k(\mathbf{x}_N, \mathbf{x}_2) & \cdots & k(\mathbf{x}_N, \mathbf{x}_N) \end{pmatrix}$$

当 $\sigma \rightarrow 0$ 时, RBF 核函数为:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right) \rightarrow \begin{cases} 1, \mathbf{x}_i = \mathbf{x}_j \\ 0, \mathbf{x}_i \neq \mathbf{x}_j \end{cases}$$

此时, 矩阵 k 变成单位矩阵, 在做特征提取时, 将原问题转化为了求解矩阵 (19) 所对应的特征值和特征向量. 本文依据 TE 过程数据做了大量的数据实验, 从结果来看, 当 $\sigma \rightarrow 0$ 时, 矩阵最大特征值所对应的特征向量中各元素基本相等, 则对于 N 个训练样本, 判别函数作用后所得到的函数值为:

$$T_{csj,i}^* = \sum_{i=1}^N \alpha_0 k(\mathbf{x}_i, \mathbf{x}_j) = \alpha_0 \quad (21)$$

也就是说, 当 σ 取一个很小的数时, MKFD 可以将所有的训练集样本正确分类, 但是测试集所映射的函数值几乎相等, 使得测试样本的分类准确率很低, 此时 RBF 核函数无法分类, MKFD 没有学习推广的能力. \square

假设 2. 若在 MKFD 中 $\sigma \rightarrow \infty$, 此时 RBF 核函数的分类能力为 0, 即它将所有的样本点都归为同一类.

证明. 在 MKFD 中, 当 $\sigma \rightarrow \infty$ 时, $\exp\left(\frac{\|\mathbf{x}_{csj} - \mathbf{x}_i\|^2}{2\sigma^2}\right) \rightarrow 1, \forall \mathbf{x}_{csj}, \mathbf{x}_i$, 此时对于 q 个测试样本 \mathbf{x}_{csj} , 得到的高斯核 MKFD 的判别函数值为:

$$T_{csj}^* = \sum_{i=1}^q \alpha_i^* k(\mathbf{x}_{csj}, \mathbf{x}_i) = \sum_{i=1}^q \alpha_i^* \exp\left(-\frac{\|\mathbf{x}_{csj} - \mathbf{x}_i\|^2}{2\sigma^2}\right) = \sum_{i=1}^q \alpha_i^*$$

从上式的形式来看, 高斯核 MKFD 所得的判别函数为常函数: $T_{csj}^* = \sum_{i=1}^q \alpha_i^*$, 将所有类型的样本点都归结为同一类. \square

由以上证明说明 MKFD 的分类能力经历由低到高再到低的过程, MKFD 的分类准确率函数 $D(\sigma)$ 相当于一个凸函数, 即我们可以近似认为分类准确率函数 $D(\sigma)$ 是关于 σ 的先增后减的凸函数, 变化曲线近似于图 1. 当 $D(\sigma)$ 为一个连续凸函数时, 可以利用区间三分法迭代, 求取函数的极值问题.

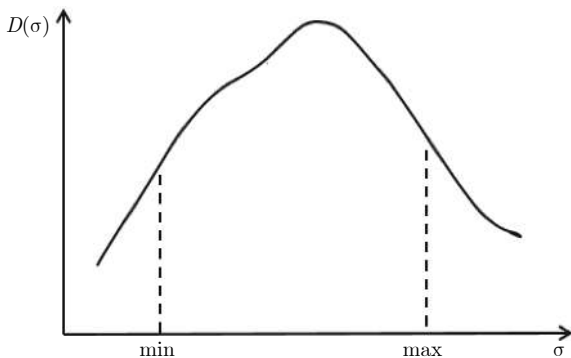


图 1 分类准确率变化图
Fig.1 Variation of classification accuracy

综上, MKFD 的分类能力会经历由低变高再变低的过程. 基于这种算法分类能力的规律, 受文献 [30] 启发本文给出区间三分法的基本步骤为:

1) 输入样本数据, 设置分类准确率条件, 选取核参数 σ 的范围 $[\min, \max]$;

2) 对区间进行迭代, 首先对区间三等分, 并将这四个等分点记为 X_1, X_2, X_3, X_4 , 分别计算四个 σ 取值所对应的分类准确率;

3) 如果四个 σ 取值所对应的分类准确率中有一个满足预设的分类准确率条件或者分类准确率不再改变, 终止迭代, 否则用 (4) 进行判断;

4) 记四个点中每两点间差商函数 A, B, C , 则 $A = D'(\xi_1), B = D'(\xi_2), C = D'(\xi_3)$, 其中 $\xi_1 < \xi_2 < \xi_3$. 分析 A, B, C 的符号, 如图 2 有以下四种;

5) a) $A, B, C > 0$, 如图 2(a) 所示, 极值点必定在

区间 $[X_2, X_4]$ 之间, 返回步骤 1). b) $A, B > 0, C < 0$, 如图 2(b) 所示, 由零点存在性定理可知极值点必定在区间 $[X_2, X_4]$ 之间, 再返回步骤 1). c) $A > 0, B, C < 0$, 如图 2(c) 所示, 由零点存在性定理可知极值点必定在区间 $[X_1, X_3]$ 内, 再返回步骤 1), d) $A, B, C < 0$, 如图 2(d) 所示, 极值点必定在区间 $[X_1, X_3]$ 之间, 返回步骤 (1);

6) 按步骤进行, 直到满足条件, 终止迭代, 得到最优 σ .

2.3 基于 MKFD 的故障诊断算法具体步骤

1) 将样本数据分为训练集和测试集. 对数据进行归一化处理, 预处理后的训练集和测试集用 X_{xlj} 和 X_{csj} 表示;

2) 采用区间迭代法确定 RBF 核参数, 求出训练集和测试集所对应的核函数集 ϵ_{xlj} 和 ϵ_{csj} ;

3) 基于 MKFD 算法, 求得训练集数据核类内离散度矩阵 K_W^ϕ 及核类间离散度矩阵 K_B^ϕ ;

4) 求解广义特征方程 $K_B^\phi \alpha = \lambda K_W^\phi \alpha$, 对得到方程对应的特征值和特征向量 α_{opt}^* ;

5) 将 ϵ_{xlj} 和 ϵ_{csj} 两类核数据集向所求得的最优判别向量 α_{opt}^* 方向投影, 得到两类数据集最优的核 Fisher 特征向量 V_{xlj}^* 和 V_{csj}^* ;

6) 计算最优的核 Fisher 特征向量之间的马氏距离 $D = \left\| V_{xlj}^* - V_{csj}^* \right\|$;

7) 验证故障类别.

3 故障诊断试验验证

TE 过程^[31] 是一个基于真实化工生产过程的模型. 它的数据在许多领域被广泛研究和应用, 比如优化控制、过程监控和故障诊断等. 为验证本文所提方法的有效性, 采用 TE 过程故障数据集进行仿真实验. TE 过程故障类型如表 1 所示. TE 过程是一个复杂的非线性大样本尺度系统, 全部过程数据均含有高斯噪声. 本文中, 每种故障类型分别选取 480 组和 960 组数据作为训练集和测试集, 每组数据有 52 个特征参数. 本文选取的故障类型为: 故障 3、4、5 和 7, 这些故障具有较高的非线性特征.

3.1 试验一: 区间三分法寻找最优核参数试验

选取故障 3、4、5 和 7 作为故障数据类型, 设置训练集共有样本 600 个, 每类故障样本有 150 个; 测试集共有样本 320 个, 每类故障样本有 80 个. 对上节选取的实验数据进行交叉验证处理, 首先采用 KFD 算法可得故障诊断准确率随核参数 σ 的变化规律如表 2 所示. 对表 2 中的数据绘制折线图, 如图 3

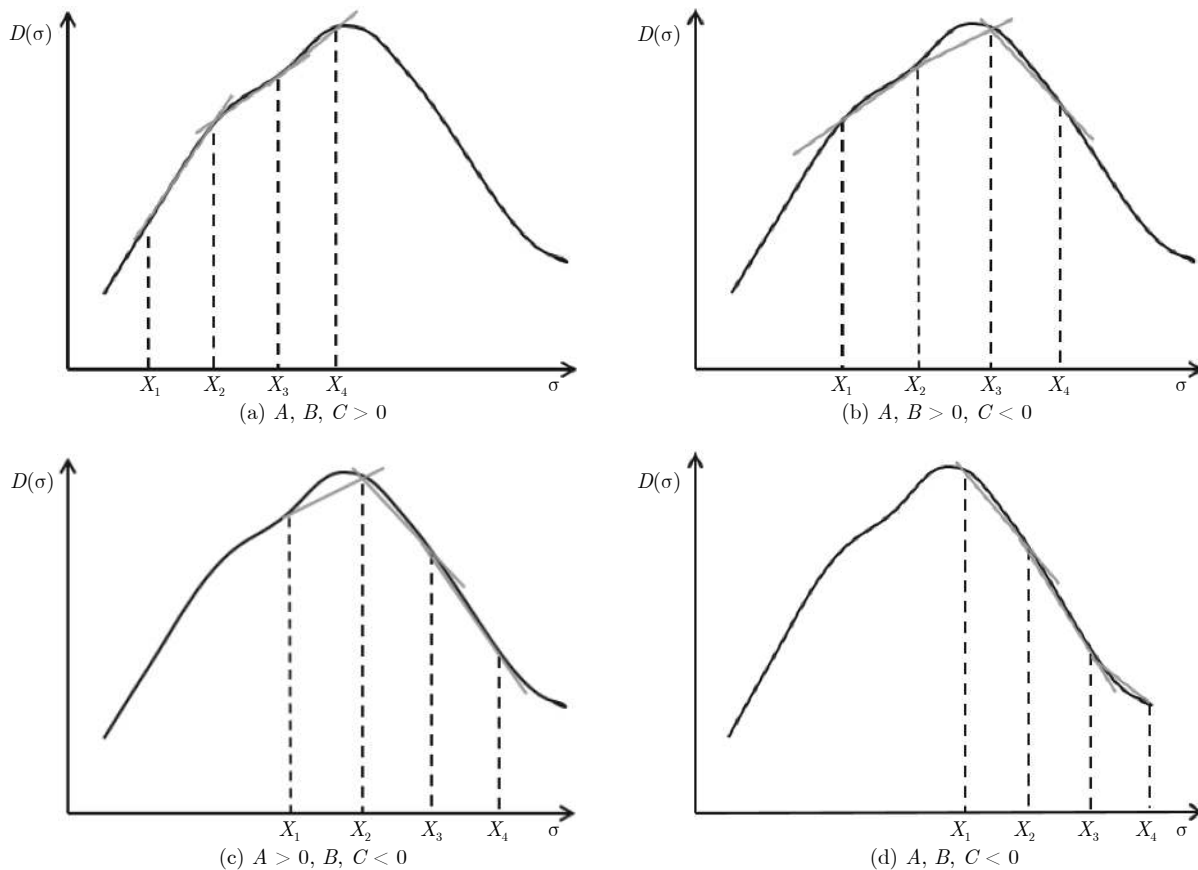


图 2 A, B, C 的符号

Fig.2 Symbols of A, B and C

表 1 故障类型描述

Table 1 Description of the selected fault sample sets

Fault Number	Fault description	Fault type
3	物料 D 的温度的异变	阶跃
4	反应器冷却水入口温度的异变	阶跃
5	冷凝器冷却水入口温度的异变	阶跃
7	物料 C 压力下降	阶跃

表 2 选取不同核参数 σ 下故障诊断的准确率 (KFD)

Table 2 The fault diagnosis accuracy based on different kernel parameter σ (KFD)

The value of the parameter σ	Test accuracy (%)	The value of the parameter σ	Test accuracy (%)
0.1	25	30	81.25
0.2	30.31	40	80.94
0.8	50	70	53.13
2	66.88	90	51.56
4	75.63	100	45
8	78.44	160	43.75
10	79.38	180	33.44

所示. 从图中可以清晰地看出故障诊断的准确率随

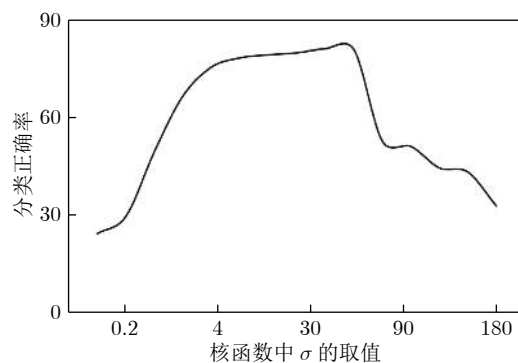


图 3 故障诊断准确率与核参数取值折线图

Fig.3 Line diagram of the fault diagnosis accuracy and kernel parameter

着核函数参数不断增大的变化规律: 首先随着参数的增加故障诊断的准确率迅速增加; 当参数达到 4 后, 分类准确率增长趋势变缓; 当参数超过 30 以后, 分类准确率随参数的增加而下降. 此时在核参数的选择上采用的是人工经验选择的方法, 效率较低并且有可能会遗漏最优解. 因此, 可以选择区间迭代法的方法取代人工经验选择法来求解最优核参数.

采用 KFD 算法对数据进行处理, 并利用区间三分法迭代求解最优核参数对应的故障诊断的准确率如表 3 所示. 由表 3 可知, 当区间迭代六次后, $\sigma = 30.9$, 分类准确率为 81.25%, 且相比第三次迭代中的 $\sigma = 30.3$, 分类准确率不再改变, 因此, 可以取 $\sigma = 30.9$ 或 $\sigma = 30.3$ 作为最优参数. 通过对比表 2 可得, 利用区间三分法求解出的最优核参数对应的分类准确率和交叉验证法得到的 $\sigma = 30$ 所对应的分类准确率相同, 但是三分法大大提高了运算效率, 并且避免了遗漏最优解的可能.

同时, 当核参数选取不同值时, 本文对 KFD 和 MKFD 两种算法的训练集和测试集的准确率进行了对比, 实验结果如表 4 所示.

分析表 4 可知, 两种算法的分类能力都是由低到高再到低的过程. 在 KFD 算法中, 当参数 σ 取 0.1 和取 1 的时候训练集的准确率在 100%, 但是测试集的准确率却只在 25% 和 50%; 相对应的, 在 MKFD 算法中, 参数 σ 取 0.1 和取 1 时也是同样的情况, 这与我们之前的理论证明一致, 即: 当 σ 趋近于 0 时, 核 Fisher 判别分析方法会出现严重的过拟合现象.

为充分验证采用区间三分法的 MKFD 算法的适用性, 本文增加了 2 个仿真算例. 选用 UCI 机器学习数据库中 Ionosphere 和 Breast cancer 分类数据来做实验验证, Ionosphere 数据集包含 351 个样本, 34 个特征向量, 分为 “good” 和 “bad” 两个数据类别,

其中 “good” 类别含有 126 个样本, “bad” 类别含有 225 个样本; Breast cancer 数据集包含 638 个样本, 9 个特征向量, 分为 “benign” 和 “malignant” 两个数据类别, 其中 “benign” 类别包含 444 个样本, “malignant” 类别包含 239 个样本. 本算例中选取类别各 80 个作为样本的训练集和测试集, 仿真结果见表 5 和图 4.

按照仿真结果来看, 采用区间三分法的 MKFD 算法仍适用于 UCI 机器学习数据库中 Ionosphere 和 Breast cancer 分类数据, 其分类能力是由低变高再变低的过程, 对于 Ionosphere 和 Breast cancer 分类数据, 其分类准确率分别达到 92.8% 和 95.4%. 也可以充分说明该算法的有效性和适用性.

3.2 试验二: 数据诊断效率试验

为了进一步验证本文算法的优越性, 本节对四种故障诊断方法在运算效率上进行对比, 这四种方法分别为: KFD、CKFD、FDGLPP^[32] 和 MKFD. 目前在化工故障领域, 基于全局—局部保持投影算法备受关注, 这种算法不仅考虑数据的全局结构信息也考虑了数据的局部近邻结构信息, 在对数据进行特征提取时, 能大大降低信息量的损失. 将这几种算法与本文算法进行仿真对比, 实验验证结果见表 6 和表 7.

本文将 MKFD 算法和 FDGLPP 算法进行对比, 我们已知 MKFD 算法故障诊断准确率随核参数

表 3 利用区间三分法求解最优核参数 σ 对应的故障诊断的准确率 (KFD)

Table 3 The accuracy of fault diagnosis of optimal kernel parameter by using the interval three-part method (KFD)

迭代次数	对应区间	三分点 1		三分点 2		三分点 3		三分点 4	
		X_1	$D(X_1)$	X_2	$D(X_2)$	X_3	$D(X_3)$	X_3	$D(X_4)$
1	[1, 100]	1	50 %	34	79 %	67	51 %	100	45 %
2	[1, 67]	1	50 %	23	80 %	45	73.8 %	67	51 %
3	[1, 45]	1	50 %	15.7	79.4 %	30.3	81.25 %	45	73.8 %
4	[15.7, 45]	15.7	79.4 %	25.5	80 %	35.2	78.8 %	45	73.8 %
5	[15.7, 35.2]	15.7	79.4 %	22.2	80.3 %	28.7	80.4 %	35.2	78.8 %
6	[22.2, 35.2]	22.2	80.3 %	26.5	80 %	30.9	81.25 %	35.2	78.8 %

表 4 KFD 算法和 MKFD 算法中不同核参数的故障诊断结果

Table 4 The fault diagnosis with different kernel parameters in KFD algorithm and MKFD algorithm

The value of the parameter σ in KFD	Train accuracy (%)	Test accuracy (%)	The value of the parameter σ in MKFD	Train accuracy (%)	Test accuracy (%)
0.1	100	25	0.1	100	25
1	100	50	1	100	50
10	99.8	79.4	4	100	76.9
30	99.8	81.3	8	100	99.69
60	70.5	44.7	12	99.9	92.5
90	27.7	25.3	16	99.9	80.6

表 5 选取不同核参数 σ 下故障诊断的准确率 (按照区间三分法做纵向表)

Table 5 The fault diagnosis accuracy based on different kernel parameters σ (Make the longitudinal table according to the interval three-part method)

Ionosphere		Breast cancer	
The value of the parameter σ	Test accuracy (%)	The value of the parameter σ	Test accuracy (%)
1	78.9	1	31.7
34	91.6	149	95.1
49	92	223	94.9
56	92.4	248	95.4
59	92.8	297	95.4
63	92.8	334	95.4
67	92.4	346	95.4
68	92	445	94.6
78	90.8	667	94
100	86.1	1000	93.2

σ 的变化规律, 所以通过区间三分法迭代寻找最优核参数, 在第 4 步迭代时就可以找到最优 σ , 此时故障诊断率达 99.69%; 而 FDGLPP 算法, 在选取最优核参数时, 需要通过交叉验证法得到, 损耗时间较长, 故 MKFD 算法能有效提升找到最优核参数的效率.

通过分析表 8 可知: CKFD、FDGLPP 和 MKFD 三种算法分别在 $\sigma = 8, 10, 9$ 时, 故障诊断准确率达到最高值. 此时可以发现, FDGLPP 算法和 MKFD 算法的准确率都高达 99% 以上, 但是 FDGLPP

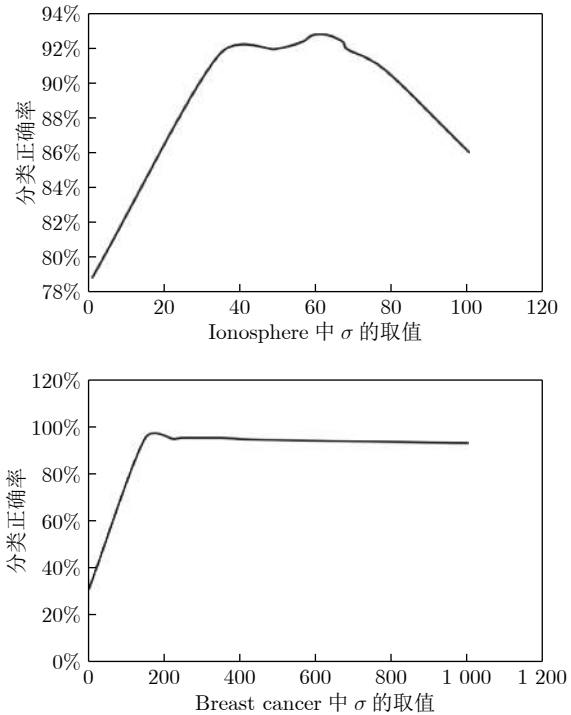


图 4 故障诊断准确率与核参数取值折线图
Fig.4 Line diagram of the fault diagnosis accuracy and kernel parameter

算法运行的时间是 MKFD 算法运行效率的两倍之多, 这也可以看出 MKFD 算法在运行效率上的优越性. 这两个实验通过对比可以发现, 本文的算法在提

表 6 区间三分法迭代求解最优核参数 σ (MKFD)

Table 6 The iterative solution of the optimal kernel parameters σ using interval partition method

迭代次数	对应区间	三分点 1		三分点 2		三分点 3		三分点 4	
		X_1	$D(X_1)$	X_2	$D(X_2)$	X_3	$D(X_3)$	X_3	$D(X_4)$
1	[1, 100]	1	50.9 %	34	60.6 %	67	57.5 %	100	58.1 %
2	[1, 67]	1	50.9 %	23	76.6 %	45	58.1 %	67	57.5 %
3	[1, 45]	1	50 %	15.7	96.3 %	30.3	63.8 %	45	58.1 %
4	[1, 30.3]	1	50 %	10.8	99.69 %	20.5	84.69 %	30.3	63.8 %
5	[1, 20.5]	1	50 %	7.5	99.38 %	14	97.81 %	20.5	84.69 %
6	[1, 14]	1	50 %	5.3	81.56 %	9.7	99.69 %	14	97.81 %

表 7 交叉验证法选取不同核参数 σ 下故障诊断的准确率 (FDGLPP)

Table 7 The fault diagnosis accuracy based on different kernel parameters σ by cross validation method

The value of the parameter σ	Test accuracy (%)	The value of the parameter σ	Test accuracy (%)	The value of the parameter σ	Test accuracy (%)
0.1	25	0.5	68.13	3	55.31
1	52.19	5	75.31	6	79.38
50	28.44	25	25.0	9	99.69
100	41.25	50	28.44	12	25.0
500	39.06	75	34.69	15	55.94
1000	38.75	95	40.0	18	25.0

升运算效率的同时并没有以损失准确率为代价, 这有效证明了本文算法的优越性.

3.3 试验三: 数据混叠的故障诊断试验

分别应用 KFD、CKFD、FDGLPP 以及 MKFD 四种算法对训练集和测试集样本进行特征

表 8 四种模型的故障诊断结果与运行时间
Table 8 Fault diagnosis results and running time of the four models

Model	Optimal value of parameter σ	Test accuracy (%)	Test time (s)
KFD	30	81.25	3.90072
CKFD	8	97.81	4.14769
FDGLPP	10	99.69	9.30612
MKFD	9	99.69	3.86806

投影, 并用马氏距离对样本进行故障分类, 特征投影结果如图 5、图 6、图 7 和图 8, 这四组图分别是在四种算法处理下, 训练集和测试集在三维主特征轴及二维主特征轴上的投影.

对训练集和测试集样本采用 KFD 投影, 得到的投影图如图 5 所示, 此时核参数取 30. 对训练集和测试集样本采用 CKFD 投影, 得到的投影图如图 6 所示, 此时核参数取 8, 权重函数取 3. 对训练集和测试集样本采用 FDGLPP 投影, 得到的投影图如图 7 所示, 此时核参数取 10. 同样对训练集和测试集样本采用 MKFD 投影, 得到的投影图如图 8 所示, 此时核参数取 8, 权重函数分别取 3 和 1.

图 5 可以看出, 通过 KFD 投影, 得到故障 3、4 和 7 的投影信息互相掩盖, 导致分类效果差. 图 6

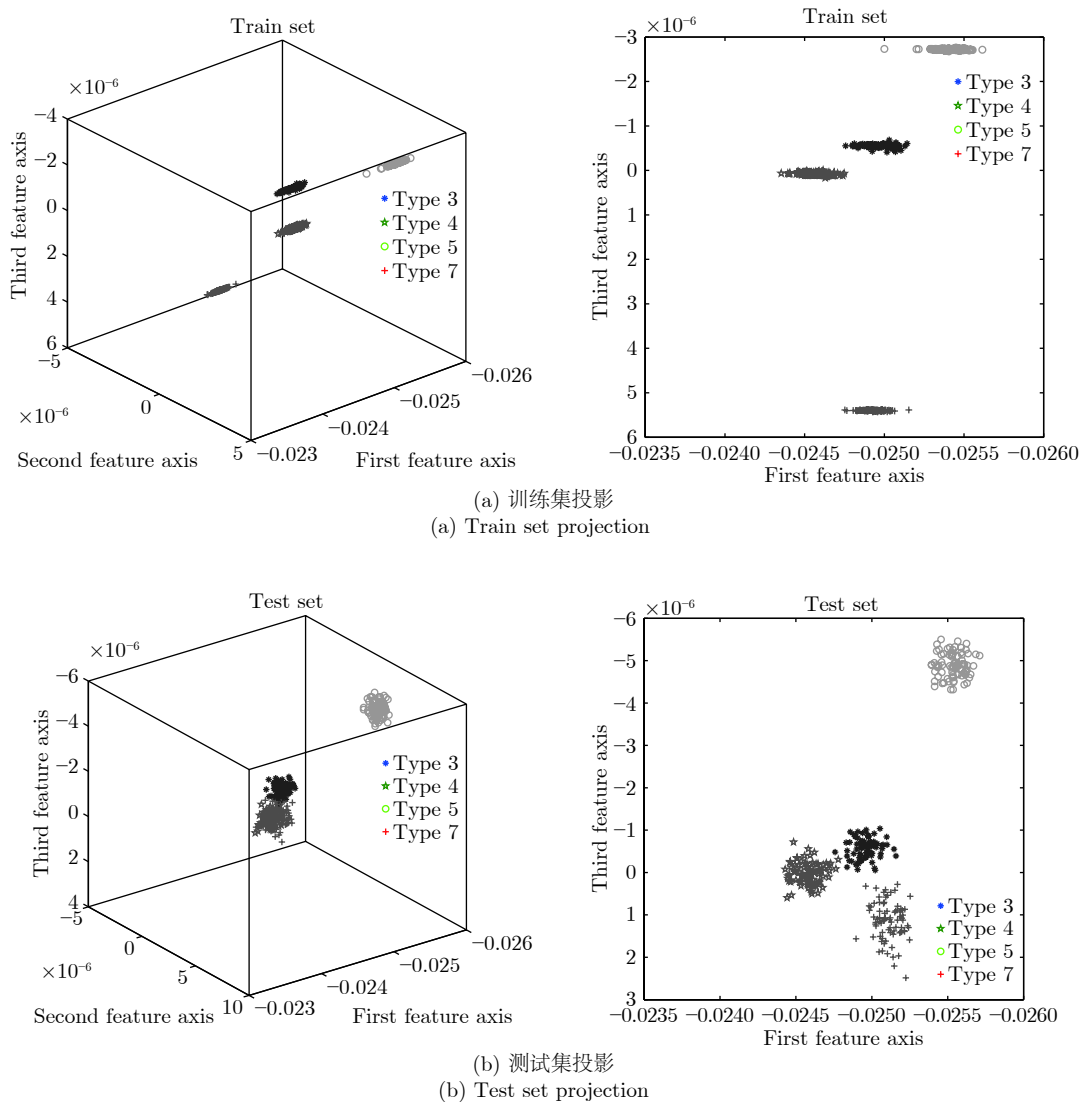


图 5 基于 KFD 算法的投影
Fig.5 Projection based on KFD algorithm

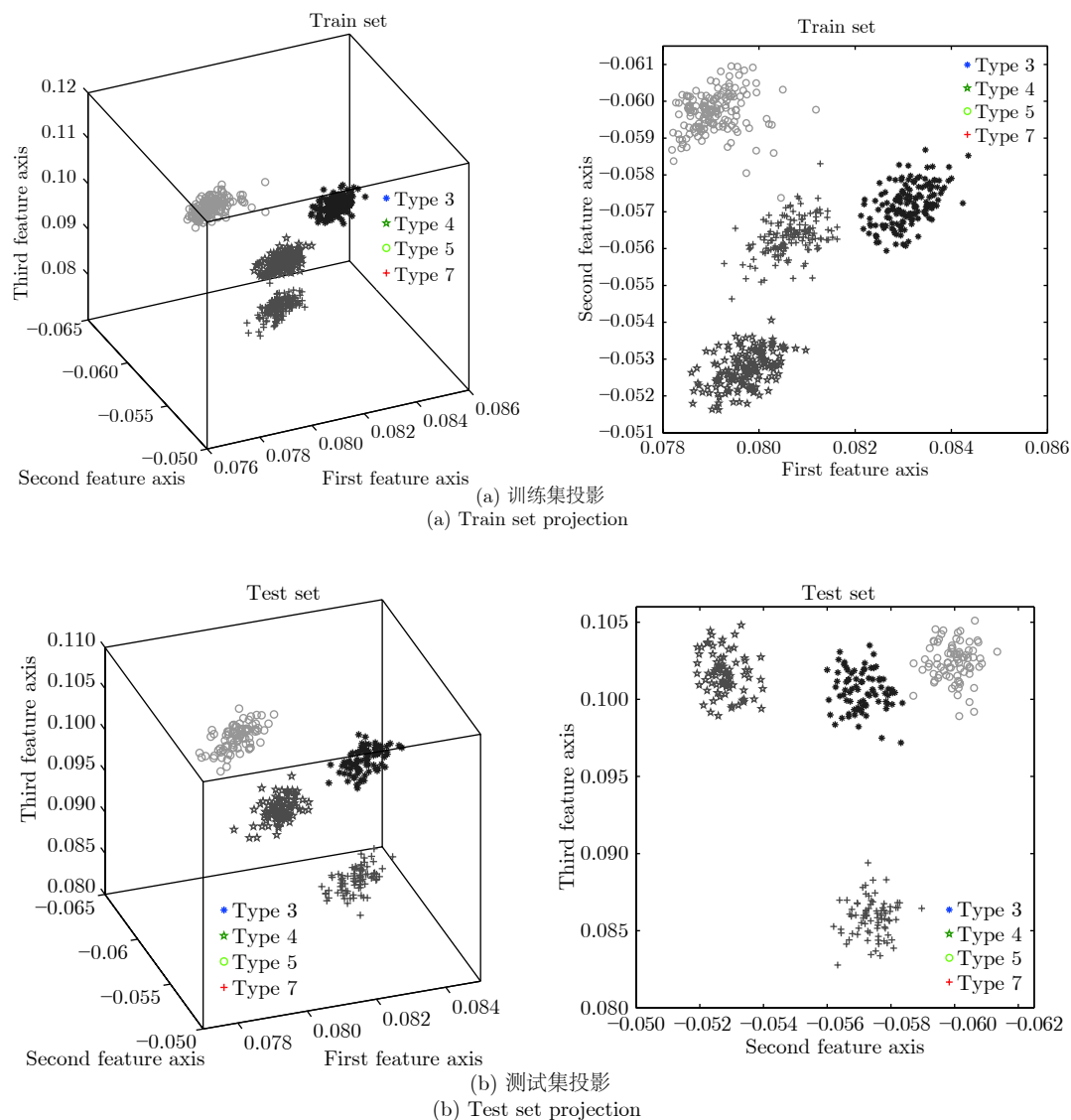


图6 基于CKFD算法的投影

Fig.6 Projection based on CKFD algorithm

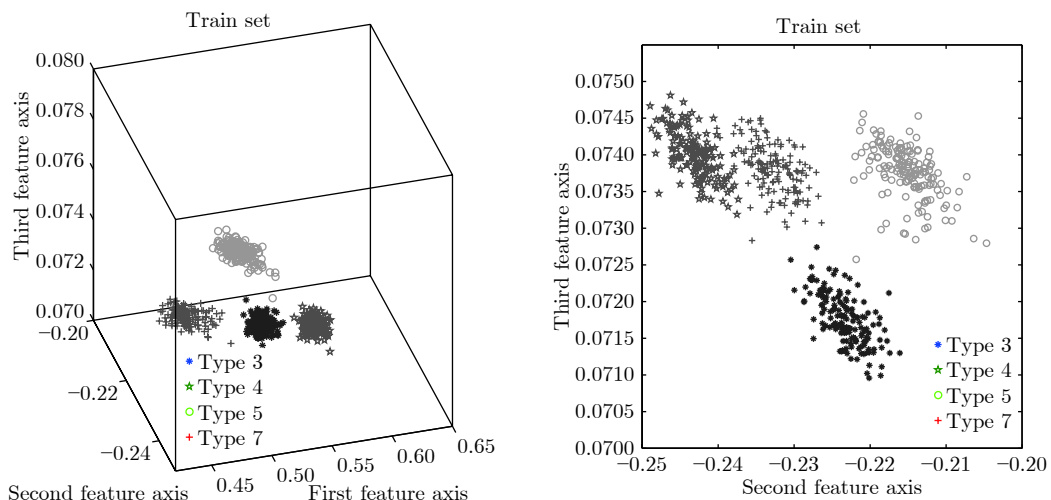
所示,通过CKFD投影,可以将三者区分开,增大两者之间的类间距,使得分类效果明显增强,可以达到97%以上,但是此时CKFD的运算效率明显低于KFD.在图7中通过FDGLPP投影,可以将这四类故障区分开,但是每一类故障投影的类内距离较分散,在数据量大的情况下,不利于分类准确.在图8中通过MKFD投影,缩小类内间距,用组平均距离代替质心距离作为改变类间距的权重函数,此时不仅能提升运算效率,还可以提升其准确率,此时的准确率为99.69%.

综上,本文提出的MKFD算法,对于故障样本的投影效果具有很大改善,提升了故障诊断的精度,同时,在MKFD算法中所提出的寻找最优核参数的区间三分法也可以大大提升算法的运行效率.

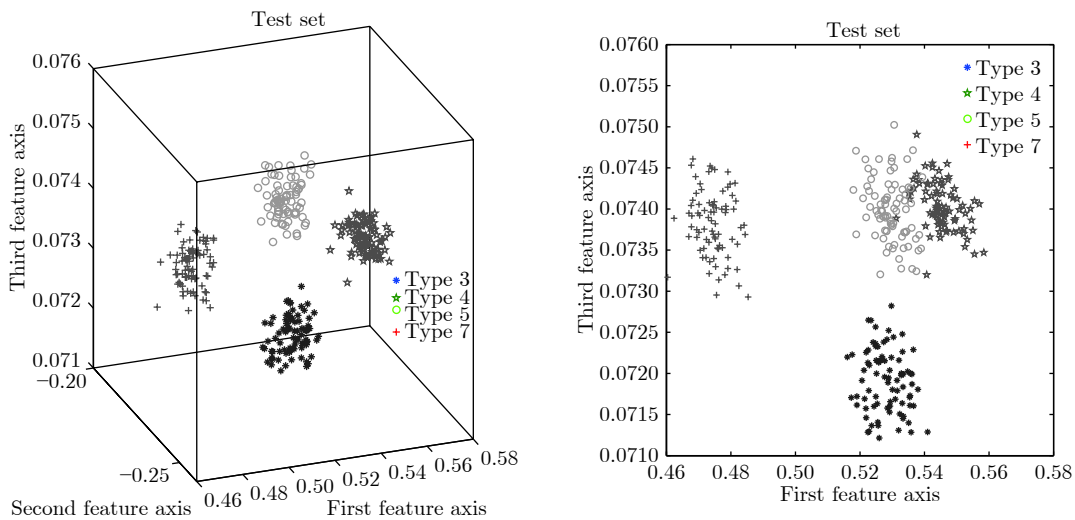
4 结论

复杂的化工过程产生了大量的状态数据,这些数据具有大量、非线性的特点.引入KFD算法进行特征提取,可以有效解决数据非线性问题,提升算法效率.但KFD算法也有一些缺点,本文针对故障数据分类效果差及无法快速确定最优核参数这两个缺点,提出了MKFD算法.

在MKFD算法中,选用区间三分法来克服核参数的低效选取问题,该方法不仅可以解决依靠经验选择参数的劣势,还可以避免遗漏最优解,可以有效提升核参数的选取效率.该方法基于TE过程数据做仿真验证,并与之前常用的交叉验证法做对比,结果表明:在最优参数精度范围内,本文所提方法加快了参数选择速度;同时,运用马氏距离对类间和类内



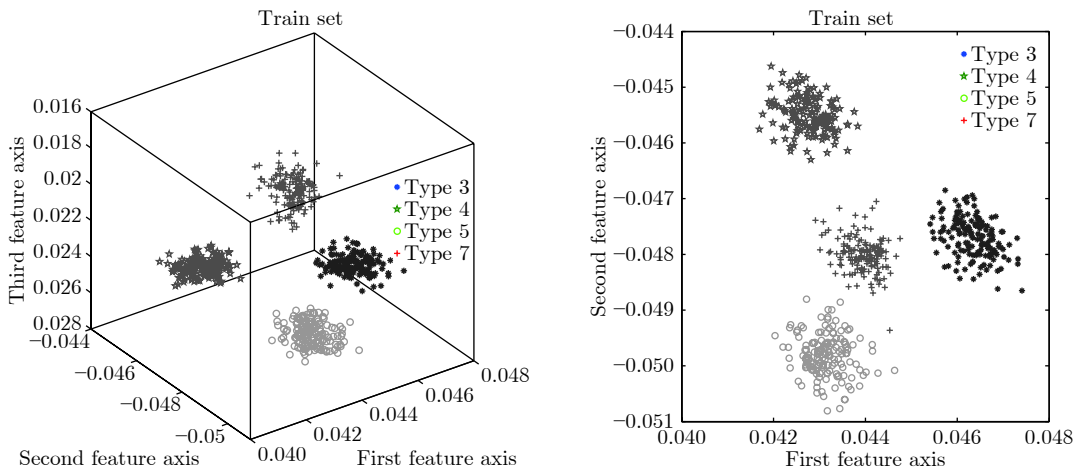
(a) 训练集投影
(a) Train set projection



(b) 测试集投影
(b) Test set projection

图 7 基于 FDGLPP 算法的投影

Fig.7 Projection based on FDGLPP algorithm



(a) 训练集投影
(a) Train set projection

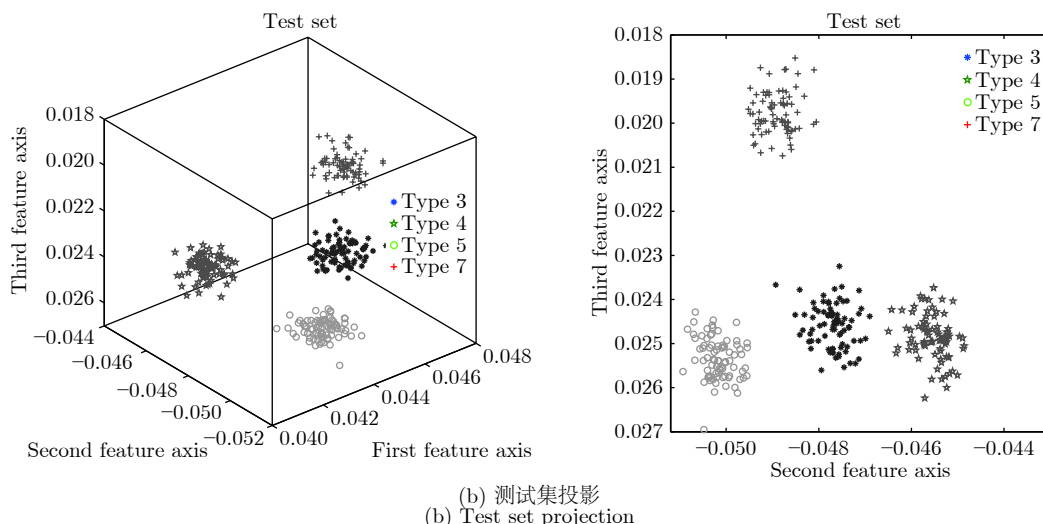


图8 基于 MKFD 算法的投影

Fig.8 Projection based on MKFD algorithm

距离进行加权,不仅可以最大限度地分离投影向量的类间距离,还可以使得其类内距离更紧凑,显著提升了故障诊断算法的运行效率。

在本文提及的方法中,数据量大所带来的矩阵非奇异的情况可以进一步深入研究,除此之外,并行运算,实时监控,提取最大信息量也是需要进一步深入研究方向。

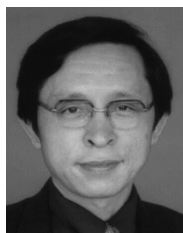
References

- Zhang Ni, Che Li-Zhi, Wu Xiao-Jin. Present situation and prospect of data-driven based fault diagnosis technique. *Computer Science*, 2017, **44**(S1): 47-52
(张妮, 车立志, 吴小进. 基于数据驱动的故障诊断技术研究现状及展望. *计算机科学*, 2017, **44**(S1): 47-52)
- Guo Yi-Fan, Tang Jia-Yin. A review of the development of life prediction and fault diagnosis technology based on machine learning algorithm. *Computer Measurement & Control*, 2019, **27**(3): 13-19
(郭一帆, 唐家银. 基于机器学习算法的寿命预测与故障诊断技术的发展综述. *计算机测量与控制*, 2019, **27**(3): 13-19)
- Zhang Y, Zhang Y, Zhang J, et al. Collaborative representation cascade for single-image super-resolution. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2017: 1-16
- Kock N, Hadaya P. Minimum sample size estimation in PLS - SEM: The inverse square root and gamma - exponential methods. *Information Systems Journal*, 2018, **28**(1): 227-261
- Maisinger K, Hobson M P, Lasenby A N. A maximum entropy method for reconstructing interferometer maps of fluctuations in the cosmic microwave background radiation. *Monthly Notices of the Royal Astronomical Society*, 2018, **290**(2): 313-326
- Hwang J S, Noh J T, Lee S H, et al. Experimental verification of modal identification of a high-rise building using independent component analysis. *International Journal of Concrete Structures and Materials*, 2019, **13**(1): 4
- Hu Zhi-Xin. Research on Chemical Fault Diagnosis Methods Based on Deep Learning [Master thesis], Hangzhou Dianzi University, China 2018
(胡志新. 基于深度学习的化工故障诊断方法研究. 杭州电子科技大学, 中国, 2018)
- Mika S, Ratsch G, Weston J, et al. Fisher discriminant analysis with kernel. In: *Proceedings of the Neural Networks for Signal Processing IX*, the 1999 IEEE Signal Processing Society Workshop. Madison, WI, USA: IEEE, 1999: 41-48
- Zhang Xi, Zhao Xu, Liu Zhen-Ya, et al. Process monitoring and fault diagnosis of turbine generator unit based on feature extraction in kernel Fisher subspace. *Proceedings of the CSEE*, 2007, (20): 1-6
(张曦, 赵旭, 刘振亚, 等. 基于核 Fisher 子空间特征提取的汽轮发电机组过程监控与故障诊断. *中国电机工程学报*, 2007, (20): 1-6)
- Ma Li-Ling, Xu Fa-Fu, Wang Jun-Zheng. A fault diagnosis method based on improved kernel Fisher. *CIESC Journal*, 2017, **68**(3): 1041-1048
(马立玲, 徐发富, 王军政. 一种基于改进核 Fisher 的故障诊断方法. *化工学报*, 2017, **68**(3): 1041-1048)
- Chikr-Elmezouar Z, Almanjahie I M, Laksaci A, et al. FDA: strong consistency of the kNN local linear estimation of the functional conditional density and mode. *Journal of Nonparametric Statistics*, 2019, **31**(1): 175-195
- Li Pu-Huang, Li Min, Fan Xin-Nan, et al. Hyperspectral anomaly detection algorithm based on iterative analysis with relative density. *Journal of Image and Graphics*, 2018, **23**(2): 219-228
(李普煌, 李敏, 范新南, 等. 迭代分析相对密度的高光谱异常检测. *中国图象图形学报*, 2018, **23**(2): 219-228)
- Zhao Zhong-Gai, Liu Fei. Application research of statistical monitoring index based on Mahalanobis distance. *Acta Automatica Sinica*, 2008, **34**(4): 493-495
(赵忠盖, 刘飞. 基于马氏距离统计监控指标的应用研究. *自动化学报*, 2008, **34**(4): 493-495)
- Wei Fen. Research on Key Technologies of Intelligent Fault Diagnosis Based on Multi-sensor Information Fusion for Rotating machinery [Ph. D. dissertation]. Harbin Institute of Technology, China, 2018
(卫芬. 旋转机械多传感器信息融合智能故障诊断关键技术研究. 哈尔滨工业大学, 中国, 2018)
- Du Wei, Fang Li-Qing, Qi Zi-Yuan. Neighborhood adaptive semi-supervised local Fisher discriminant analysis algorithm. *Application Research of Computers*, 2019, **36**(1): 105-108
(杜伟, 房立清, 齐子元. 一种邻域自适应半监督局部 Fisher 判别分析算法. *计算机应用研究*, 2019, **36**(1): 105-108)
- Yang Wu-Yi, Liang Wei, Xin Le, et al. Subspace semi-supervised Fisher discriminant analysis. *Acta Automatica Sinica*, 2009, **35**(12): 1513-1519
(杨武夷, 梁伟, 辛乐, 等. 子空间半监督 Fisher 判别分析. *自动化学报*, 2009, **35**(12): 1513-1519)
- Zheng Jian-Wei, Wang Wan-Liang, Yao Xiao-Min, et al. Face recognition using tensor local Fisher discriminant analysis. *Acta Automatica Sinica*, 2012, **38**(9): 1485-1495

- (郑建伟, 王万良, 姚晓敏, 等. 张量局部 Fisher 判别分析的人脸识别. 自动化学报, 2012, **38**(9): 1485–1495)
- 18 Wen Ting-Xin, Yu Feng-E. Research on prediction of coal spontaneous combustion based on KPCA-Fisher discriminant analysis. *Mining Safety & Environmental Protection*, 2018, **45**(2): 49–53
(温廷新, 于凤娥. 基于 KPCA-Fisher 判别分析的煤炭自燃预测研究. 矿业安全与环保, 2018, **45**(2): 49–53)
- 19 Wen T, Jia Y, Huang D, et al. Feature extraction of electronic nose signals using QPSO-based multiple KFDA signal processing. *Sensors*, 2018, **18**(2): 388
- 20 Liu Ting-Rui, Chang Lin. Quasi-steady response and LLTR control of a wind turbine blade with bending-torsion coupled. *Journal of Vibration and Shock*, 2018, **37**(13): 123–129
(刘廷瑞, 常林. 弯扭耦合风力机叶片的准稳态响应及 LLTR 控制. 振动与冲击, 2018, **37**(13): 123–129)
- 21 Shi H, Liu J, Wu Y, et al. Fault diagnosis of nonlinear and large-scale processes using novel modified kernel Fisher discriminant analysis approach. *International Journal of Systems Science*, 2016, **47**(5): 1–15
- 22 Guo Jin-Yu, Han Jian-Bin, Li Yuan, et al. Fault diagnosis of complex chemical process based on local Fisher discriminant analysis. *Application Research of Computers*, 2018, **35**(4): 1122–1125
(郭金玉, 韩建斌, 李元, 等. 基于局部 Fisher 判别分析的复杂化工过程故障诊断. 计算机应用研究, 2018, **35**(4): 1122–1125)
- 23 Lin Xin-Chuan, You Gui-Rong. Study on optimization of indoor location algorithm based on iBeacon. *Journal of Shanxi University of Technology (Natural Science Edition)*, 2017, **33**(3): 67–73
(林信川, 游贵荣. 基于 iBeacon 的室内定位算法优化研究. 陕西理工大学学报 (自然科学版), 2017, **33**(3): 67–73)
- 24 Liu C, Sun Y. The research and application of learning program in adaptive learning system. *Applied Mechanics & Materials*, 2014, **347-350**: 3109–3113
- 25 Meng Ya-Hui. A method of mechanical fault classification based on optimal feature subset and K-nearest neighbor using mahalanobis distance. *Machinery Design & Manufacture*, 2017, (7): 104–108
(孟亚辉. 基于最优特征集和马氏距离 KNN 分类的机械故障分类方法研究. 机械设计与制造, 2017, (7): 104–108)
- 26 Lin Sheng-Liang, Liu Zhi. Parameter selection in SVM with RBF kernel function. *Journal of Zhejiang University of Technology*, 2007, **35**(2): 163–167
(林升梁, 刘志. 基于 RBF 核函数的支持向量机参数选择. 浙江工业大学学报, 2007, **35**(2): 163–167)
- 27 Wei Guo, Liu Jian, Sun Jin-Wei, et al. Study on nonlinear multifunctional sensor signal reconstruction method based on LS-SVM. *Acta Automatica Sinica*, 2008, **34**(8): 869–875
(魏国, 刘剑, 孙金玮, 等. 基于 LS-SVM 的非线性多功能传感器信号重构方法研究. 自动化学报, 2008, **34**(8): 869–875)
- 28 Huang Xin-Han, Du Ke-Lin, Wang Min, et al. A simulation investigation on dynamic assembly process based on impedance control. *Acta Automatica Sinica*, 2000, **26**(2): 169–175
(黄心汉, 杜克林, 王敏, 等. 基于阻抗控制的动态装配过程仿真研究. 自动化学报, 2000, **26**(2): 169–175)
- 29 Zhang Xiao-Yun, Liu Yun-Cai. Performance analysis of support vector machines with Gauss kernel. *Computer Engineering*, 2003, **29**(8): 22–25
(张小云, 刘允才. 高斯核支撑向量机的性能分析. 计算机工程, 2003, **29**(8): 22–25)
- 30 Lawn S D, Kerkhoff A D, Burton R, et al. Diagnostic accuracy, incremental yield and prognostic value of determine TB-LAM for routine diagnostic testing for tuberculosis in HIV-infected patients requiring acute hospital admission in South Africa: A prospective cohort. *Bmc Medicine*, 2017, **15**(1): 67
- 31 Peng Ze-Dong, Feng Yi-Ping, Yang Sheng-Lan, et al. Design and implementation of extended TE process simulation system based on multi-agent. *Journal of Chemical Engineering of Chinese Universities*, 2019, **33**(5): 1195–1205
(彭泽栋, 冯毅萍, 杨胜蓝, 等. 基于多智能体的 TE 过程扩展仿真系统设计与实现. 高校化学工程学报, 2019, **33**(5): 1195–1205)
- 32 Tang Q, Chai Y, Xu J, Fang X, Tang Q, Chai Y, Qu J, et al. Industrial process monitoring based on Fisher discriminant global-local pre-serving projection. *Journal of Process Control*, 2019, **81**: 76–86



吕鹏飞 西北工业大学硕士研究生. 主要研究方向为机器学习与故障诊断.
E-mail: 13571807486@163.com
(LV Peng-Fei Master student at the School of Mechanics, Civil Engineering and Architecture, Northwestern Polytechnical University. Her research interest covers machine learning and fault diagnosis.)



闫云聚 西北工业大学力学与土木建筑学院教授. 主要研究方向为信息融合与故障智能诊断系统. 本文通信作者.
E-mail: yjyan_2895@nwpu.edu.cn
(YAN Yun-Ju Professor at the School of Mechanics, Civil Engineering and Architecture, Northwestern Polytechnical University. His research interest covers information fusion and fault intelligent diagnosis. Corresponding author of this paper.)



荔越 西北工业大学力学与土木建筑学院博士研究生. 主要研究方向为故障诊断与机器学习.
E-mail: christli@mail.nwpu.edu.cn
(LI Yue Ph. D. candidate at the School of Mechanics, Civil Engineering and Architecture, Northwestern Polytechnical University. His research interest covers machine learning and fault diagnosis.)