

DOI:10.19651/j.cnki.emt.2211095

基于核熵成分分析的工业过程多类型故障诊断^{*}

李榕 申志 李元

(沈阳化工大学信息工程学院 沈阳 110142)

摘要:核熵成分分析(KECA)特征提取过程中只保留了数据的最大瑞丽熵(Renyi)信息,没有充分利用数据的类别信息。由于监督学习算法线性判别分析(LDA)能够有效提取特征中的类别信息,因此提出 KECA-LDA(KEDA)的特征提取方法。首先 KECA 依据最小 Renyi 熵损失策略对数据进行维数约简;然后在 KECA 特征空间使用 LDA 算法获取具有判别信息的低维特征并输入到支持向量机(SVM)分类器中,利用天牛须优化算法(BAS)得到最佳性能的 SVM 分类器,从而建立故障诊断模型。将 KEDA-BAS-SVM 方法应用于田纳西-伊斯曼化工过程(TE)进行仿真实验,结果表明:当采用基于距离测度的矩阵相似性优化确定 KEDA 中所选用的径向基函数(RBF)核参数时,相比 KECA 和 LDA 算法,KEDA 特征提取后多类型故障诊断准确率达到 99.7%,验证了 KEDA-BAS-SVM 在多类型故障诊断领域的优越性。

关键词:特征提取;多类型故障诊断;核熵成分分析;线性判别分析

中图分类号: TQ015 **文献标识码:** A **国家标准学科分类代码:** 520.2040

Multi-type fault diagnosis of industrial process based on KECA

Li Rong Shen Zhi Li Yuan

(College of Information Engineering, Shenyang University of Chemical Technology, Shenyang 110142, China)

Abstract: The kernel entropy component analysis(KECA) feature extraction process only retains the maximum Renyi entropy information of the data, but does not fully utilize the category information of the data. As a supervised learning algorithm, linear discriminant analysis(LDA) can effectively extract category information in features. Therefore, a feature extraction method of KECA-LDA (KEDA) was proposed. Firstly, KECA reduced the dimension of the data according to the minimum Renyi entropy loss strategy. Then, the LDA algorithm was used in the KECA feature space to obtain low-dimensional features with discriminative information and input them into the support vector machine (SVM) classifier. The best performance SVM was obtained by beetle antennae search (BAS) and to build a fault diagnosis model. The KEDA-BAS-SVM method was applied to the Tennessee eastman (TE) for simulation experiments. The results showed that When the matrix similarity optimization based on distance measure was used to determine that the kernel parameter of RBF selected in KECA, compared with the KECA and LDA algorithms, after using KEDA feature extraction, the accuracy of multi-type fault diagnosis reached 99.7%, which verified the superiority of KEDA-BAS-SVM in the field of multi-type fault diagnosis.

Keywords: feature extraction; multi-type fault diagnosis; kernel entropy component analysis; linear discriminant analysis

0 引言

工业生产过程复杂多变,生产过程中对于设备的安全性和可靠性要求也愈加严格^[1-4]。为了更多的捕捉过程信息,往往通过安装多样的传感器的方式,获取大量反映过程信息的数据。通过分析正常和不同故障运行模式下的数据,实现基于数据驱动的故障诊断^[5]。

工业过程数据具有维数高、高斯与非高斯共存、复杂性等特点,若直接对获取的原始数据进行故障诊断,则消耗计算机资源多、实时性较差,因此需要对降维方法进行研究。主成分分析(principal components analysis, PCA)^[6]是工业过程中最为经典的降维方法,能够处理高维、嘈杂和高度相关的数据。但是,当数据分布具有非高斯和非线性特征时,由于使用二阶统计量和线性假设,PCA 此时表现不佳。线

收稿日期:2022-08-17

^{*} 基金项目:国家自然科学基金(62273242,61673279)项目资助

性判别分析^[7](linear discriminant analysis, LDA)种监督学习算法,使高维数据投影到最佳鉴别坐标系空间以降低特征计算复杂性和抽取有效分类特征的效果。然而,与 PCA 相比, LDA 对于非高斯数据处理效果不好。在 2010 年, Jenssen^[8]提出一种新的降维方法—核熵成分分析方法。核熵成分分析(kernel entropy component analysis, KECA)^[9-12]以 Renyi 熵为信息衡量的指标,使用高阶统计量最大程度减少数据的熵值损失,提取数据中的非高斯信息。但是 KECA 存在降维过程中忽略数据样本类别信息以及核参数选择繁琐等问题。

多类型故障诊断在工业过程中可认为是多分类问题。其中, SVM 因其泛化能力强、分类速度以及结果易解释而受到许多研究者的关注并成功应用于各个领域^[13-15]。但 SVM 性能也易受惩罚因子 C 和径向基核参数 δ 的影响。Jiang 等^[16]在 2017 年提出了天牛须优化算法(battle antennae search algorithm, BAS),由于 BAS 算法原理简单、易于编程实现,初始参数少等众多优势而使得在日前主流优化算法中极具竞争性。尽管 BAS 如同粒子群算法(particle swarm optimization, PSO)^[17]一样都有可能陷入局部极值的困境,但相比 PSO 算法, BAS 更容易跳出; BAS 不同于遗传算法(genetic algorithm, GA)^[18]算法有时需要二进制编码,从而引起计算量较大等问题;相比蝙蝠算法(bat algorithm, BA)^[19]来说, BAS 的效率更高并且复杂度极低。综上所述, BAS 算法更适合 SVM 参数寻优。

本文提出 KEDA 特征提取方法,具体描述为,原始特征空间转换到 KECA 空间,保留数据的分布信息,然后在 KECA 空间执行 LDA 算法,保留数据的类别信息,数据样本经 KEDA 特征提取后输入到 BAS-SVM 分类器中,建立故障诊断模型。对于 KECA 算法中核参数选择问题,引入基于距离测度的矩阵相似性度量方法对核参数进行寻优,该方法在确保高精度前提下能够大大减少计算量。本文对田纳西-伊斯曼化工过程(Tennessee-Eastman, TE)进行模拟实验,与 KECA 和 LDA 算法进行对比,以测试数据样本的诊断正确率为衡量准则,验证了 KEDA-BAS-SVM 诊断方法的有效性。

1 基础理论

1.1 KECA

定义数据集 $D = \{x_1, x_2, \dots, x_n\}$ 的概率密度函数 $p(x)$, 其 Renyi 熵 $H(p)$ 表示为:

$$H(p) = -\log\left(\int p^2(x) dx\right) \quad (1)$$

由上式可以看出, Renyi 熵估计取决于数据集的概率密度函数,不需要对数据的分布进行限定,算法的适用性更强。因为对数函数本身具有单调性,因此只需考虑:

$$\hat{V}(p) = \int p^2(x) dx \quad (2)$$

为了估计 $\hat{V}(p)$, 引入 Parzen 窗函数 $\hat{p}(x)$, 表示为:

$$\hat{p}(x) = \frac{1}{n} \sum_{x_i \in D} k_\gamma(x, x_i) \quad (3)$$

式中: n 表示样本数, $k_\gamma(x, x_i)$ 为 Mercer 核函数, γ 为核参数。

$\hat{V}(p)$ 用样本均值近似估计 $V(p)$:

$$\hat{V}(p) = \frac{1}{n} \sum_{x_i \in D} \hat{p}(x_i) = \frac{1}{n} \sum_{x_i \in D} \frac{1}{n} \sum_{x_{i'} \in D} K_\gamma(x_i, x_{i'}) = \frac{1}{n^2} \mathbf{I}^T \mathbf{K} \mathbf{I} \quad (4)$$

式中: \mathbf{I} 为 $n \times 1$ 的单位向量。对 $n \times n$ 的核矩阵 \mathbf{K} 特征分解:

$$\mathbf{K} = \mathbf{E} \mathbf{\Lambda} \mathbf{E}^T \quad (5)$$

式中: $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ 为特征值矩阵, $\mathbf{E} = (e_1, e_2, \dots, e_n)$ 为特征向量矩阵, 将式(5)代入式(4), 进一步简化得到:

$$\hat{V}(p) = \frac{1}{n^2} \sum_{i=1}^n (\sqrt{\lambda_i} e_i^T \mathbf{I})^2 = \frac{1}{n^2} \sum_{i=1}^n \zeta_i \quad (6)$$

由式(6)可得, Renyi 熵可以表示为 n 个分量的累积, 各个分量都同时包含特征值与特征向量。

数据集 D 通过非线性 φ 映射到高维空间表示为 $\Phi = [\varphi(x_1), \varphi(x_2), \dots, \varphi(x_n)]$ 。由于 KECA 可以被看作是最大限度地保留 Renyi 熵估计的主轴在包含原始数据最多信息的情况下构造的子空间, 即将数据映射到由 k 个 KPCA 主轴张成的子空间 U_k 上, 对 ζ_i 值从大到小排序, 选取前 k 个特征值和特征向量, 转换后的数据表示为:

$$R^{keca} = U_k \Phi = \mathbf{\Lambda}_k^{0.5} \mathbf{E}_k^T \quad (7)$$

$$\min_{\lambda_1, e_1, \dots, \lambda_n, e_n} \{\hat{V}(p) - \hat{V}_k(p)\}$$

式中: $\hat{V}_k(p)$ 是选取前 k 个特征向量估计的 Renyi 熵, $\mathbf{\Lambda}_k$ 为前 k 个特征值, \mathbf{E}_k 为前 k 个特征向量。

对于新的测试样本 x_{new} 在特征空间上的投影可表示为:

$$R_{new}^{keca} = U_k \Phi(x_{new}) = \mathbf{\Lambda}_k^{-0.5} \mathbf{E}_k^T \mathbf{K}_{new}(x, x_{new}) \quad (8)$$

1.2 KEDA

在 KECA 特征空间执行 LDA 算法, 找到一个最佳的特征集, 从而提高诊断准确率。计算类间离散度 \mathbf{S}_b^{keca} 与类内离散度矩阵 \mathbf{S}_w^{keca} , 表达式如下:

$$\mathbf{S}_b^{keca} = \sum_{i=1}^c n_i (m_i^{keca} - m_0^{keca})(m_i^{keca} - m_0^{keca})^T \quad (9)$$

$$\mathbf{S}_w^{keca} = \sum_{i=1}^c \sum_{j=1}^{n_i} (r_{ij}^{keca} - m_i^{keca})(r_{ij}^{keca} - m_i^{keca})^T \quad (10)$$

式中: c 为类别数, n_i 为第 i 类群的数据样本个数, m_i^{keca} 为第 i 类群的均值向量, r_{ij} 为第 i 个类群第 j 个数据样本。 m_0^{keca} 表示所有数据样本的均值向量。Fisher 准则如下:

$$\max J(\mathbf{w}^{keda}) = \frac{\mathbf{w}^{keda \top} \mathbf{S}_b^{keda} \mathbf{w}^{keda}}{\mathbf{w}^{keda \top} \mathbf{S}_w^{keda} \mathbf{w}^{keda}} \quad (11)$$

投影矩阵 $\{\mathbf{w}_i^{keda}, i = 1, 2, \dots, t\}$ 是对应于 $\lambda^{keda} \mathbf{S}_w^{keda} \mathbf{w}^{keda} = \mathbf{S}_b^{keda} \mathbf{w}^{keda}$ 的 t 个最大特征值对应的特征向量, 变量 t 也即为 KEDA 降维后数据的维数。因此, 经 KEDA 转换后的数据表示为:

$$\mathbf{X}_{keda} = \mathbf{w}^{keda \top} \mathbf{R}^{keda} = \mathbf{w}^{keda \top} \mathbf{A}_k^{0.5} \mathbf{E}_k^T \quad (12)$$

测试数据经 KEDA 转换后的数据表示为:

$$\mathbf{Y}_{keda} = \mathbf{w}^{keda \top} \mathbf{R}_{new}^{keda} = \mathbf{w}^{keda \top} \mathbf{A}_k^{0.5} \mathbf{E}_k^T \mathbf{K}_{new}(x, x_{new}) \quad (13)$$

核参数确定方法

由于 RBF 无论对大样本还是小样本都有比较好的性能, 而且相对于多项式核函数优化参数要少, 是较为理想的分类依据函数^[20]。因此本文选用 RBF, 表示为:

$$\mathbf{K}(x_i, x_j) = \exp \frac{\|x_i - x_j\|^2}{2\gamma^2} \quad (14)$$

定义 \mathbf{K} 为实际 Gram 矩阵, 核参数 γ 待定。

根据实际数据样本核矩阵, 构造理想 Gram 矩阵 \mathbf{G} , 理想状态下, 即 x_i, x_j 完全相同时, 向量内积为 1; 否则为 0, 表达式为:

$$\mathbf{G} = \begin{cases} 1, & L(x_i) = L(x_j) \\ 0, & L(x_i) \neq L(x_j) \end{cases} \quad (15)$$

式中: L 为类别判定函数, $L(x_i) = L(x_j)$ 表示 x_i 与 x_j 属于同类样本, $L(x_i) \neq L(x_j)$ 表示 x_i 与 x_j 属于异类样本。当 \mathbf{K} 和 \mathbf{G} 越相似, 此时 γ 更有利于数据分类。引入基于距离测度的矩阵相似性度量方法^[21] 来确定 KECA 算法中核参数 γ , 计算公式为:

$$M(\mathbf{K}', \mathbf{G}') = (\mathbf{K}' - \mathbf{G}')_{1 \times n^2} (\mathbf{K}' - \mathbf{G}')_{n^2 \times 1}^T \quad (16)$$

式中: \mathbf{K}' 表示矩阵 \mathbf{K} 转换的 n^2 维向量; \mathbf{G}' 表示矩阵 \mathbf{G} 转换的 n^2 维向量。通过求取 M 的极小值确定最优核参数 γ 。

1.3 支持向量机

1) 基本原理

KEDA 对数据样本处理后, 将包含原始训练数据的类别信息的低维特征表示 X_{keda} 输入到 SVM 模型进行训练获得判别分类函数。训练好的 SVM 模型对测试数据低维特征表示 Y_{keda} 进行有效分类。数据集表示为 $X_{keda} = \{(x_{keda}^1, y_1), (x_{keda}^2, y_2), \dots, (x_{keda}^n, y_n)\}, y_i \in \{-1, +1\}$, 在该空间中找到一个最大分离超平面, 将不同类别数据进行分类。超平面定义如下:

$$\mathbf{w}^T x_{keda} + b = 0 \quad (17)$$

式中: \mathbf{w} 为法向量, b 为位移项。为了提高 SVM 的容错率, 引入惩罚因子 C 和松弛变量 ζ , SVM 的基本模型表示为:

$$\begin{aligned} \min_{\mathbf{w}, b, \zeta} & \left(\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \zeta_i \right) \\ \text{s. t. } & y_i (\mathbf{w}^T x_{keda}^i + b) \geq 1 - \zeta_i, \\ & \zeta_i \geq 0, i = 1, 2, \dots, n \end{aligned} \quad (18)$$

为了求解上式, 将其转换为对偶问题, 引入 Lagrange 乘法法对其求解, 该问题的 Lagrange 函数表示为:

$$L(\mathbf{w}, b, \alpha, \zeta, \mu) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \zeta_i + \sum_{i=1}^n \alpha_i (1 - \zeta_i - y_i (\mathbf{w}^T x_{keda}^i + b)) - \sum_{i=1}^n \mu_i \zeta_i \quad (19)$$

其中, $\alpha_i \geq 0, \mu_i \geq 0$ 为拉格朗日乘子。

实际工业生产过程中数据往往呈现非线性分布, 为了对数据更好地分类, 将数据样本通过非线性映射至高维空间, 但直接定义映射函数较为困难, 并且计算量较大。因此直接引入核函数避免了一系列繁琐步骤, 相应的 SVM 模型的判别函数表示为:

$$f(x_{keda}) = \text{sign} \left(\sum_{i,j=1}^n \alpha_i y_i \mathbf{K}(x_{keda}^i, x_{keda}^j) + b \right) \quad (20)$$

2) 参数更新

惩罚因子 C 权衡了经验风险和结构风险, 径向基核参数 δ 决定了样本划分的精细程度, 两者在一定程度上影响着 SVM 的分类性能^[22-23], 因此对该参数的优化选取是十分必要, 得到最优解 C 和 δ 将能够大大提高 SVM 分类准确率。

BAS 搜索流程简要描述如下:

(1) 在维度 d 中建立天牛须的随机朝向并进行归一化;

$$dir = \frac{\text{rands}(d, 1)}{\|\text{rands}(d, 1)\|} \quad (21)$$

(2) 天牛须左触须 h_l 和右触须 h_r 的位置表示如下, h_s 为天牛 s 时刻所在位置, d_s^0 为 s 时刻两触须间距离;

$$\begin{cases} h_l = h_s + d_s^0 \cdot \frac{dir}{2} \\ h_r = h_s - d_s^0 \cdot \frac{dir}{2} \end{cases} \quad (22)$$

(3) 根据左右两触须对食物浓度感知程度, 判断天牛下一步位置。

$$h_{s+1} = h_s - step_s \cdot dir \cdot \text{sign}(f_1(h_l) - f_1(h_r)) \quad (23)$$

式中: $step_s$ 为 s 时刻步长, f_1 为适应度函数。

(4) 步长更新与两触须间距离更新。

$$\begin{aligned} step_{s+1} &= step_s \cdot \eta \\ d_0^{s+1} &= \frac{step_{s+1}}{\alpha} \end{aligned} \quad (24)$$

式中: α 为常数, 设置为 5; η 为步长衰减系数, 设置为 0.95。

(5) 满足最大迭代次数时, 程序停止, 输出最优解。

2 KEDA-BAS-SVM 多类型故障诊断

基于 KEDA-BAS-SVM 多类型故障诊断分为两部分: 离线建模和故障诊断。诊断流程如图 1 所示。主要步骤如下。

离线建模:

1) 给定训练数据 D 以及相对应的数据标签, 使用正常数据的均值和标准差对其进行标准化处理;

2) 根据式(12)提取原始数据的低维嵌入 X_{keda} ;

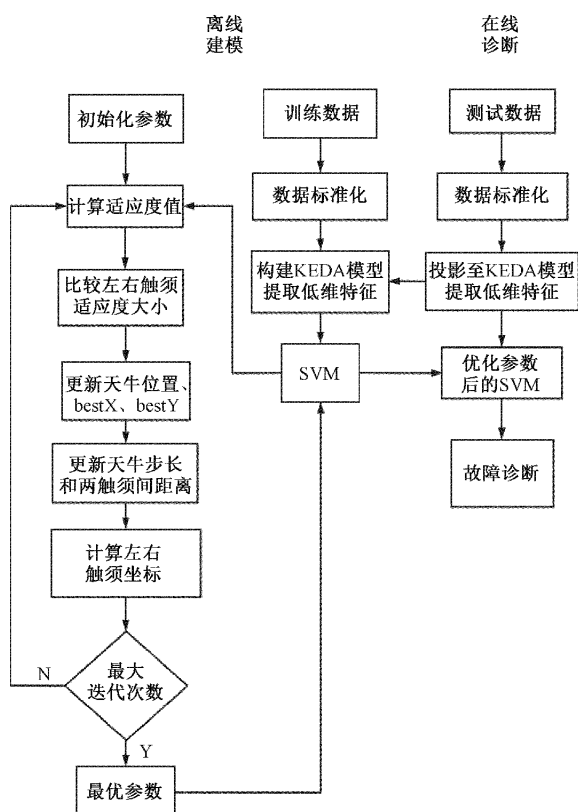


图 1 基于 KEDA-BAS-SVM 多类型故障诊断流程

3) 以训练数据集五折交叉验证平均准确率作为适应度函数;

4) 在二维平面内确定天牛初始位置保存在 $bestX$ 中, 计算适应度值保存在 $bestY$ 中; 同时, 建立天牛须的随机朝

向并进行归一化;

5) 根据式(22)计算天牛须左右坐标值以及左右须适应度值并与 $bestY$ 比较, 将更优值重新保存在 $bestY$ 中, 更新 $bestX$;

6) 步长与两触须间距离根据式(24)更新;

7) 迭代过程, 达到所设定的最大迭代次数时停止循环, 进入步骤 8), 否则返回步骤 5);

8) 输出最优参数, 最优 $bestY$ 对应的 $bestX$ 即最优的参数 C 和 δ , 以此建立最佳性能的 SVM 故障诊断模型;

在线诊断:

1) 给定测试数据 x_{new} , 使用建模正常数据样本的均值和标准差进行标准化处理;

2) 测试数据 x_{new} 向 KEDA 模型投影, 根据式(13)获取测试数据的低维特征表示 Y_{keda} ;

3) 将 Y_{keda} 输入 SVM 诊断模型中, 得到多类型故障诊断结果。

3 TE 仿真实验

TE 过程是由美国 Eastman 化学公司的 Downs 和 Vogel 共同开发的 TE Benchmark 仿真模拟平台, 近年来被广泛的应用于工业生产故障检测与诊断^[24-25]。TE 工业流程主要包括 5 个操作单元: 反应器、冷凝器、循环压缩机、气液分离器和解析塔, 整个过程中共涉及 8 种物料成分, 分别为主要参加反应的气体进料 A、C、D、E; 惰性不可溶进料 B; 反应副产品 F 以及反应液态主产物 G 和 H。该平台总共预设了 21 种故障, 其中, IDV1~IDV7 为阶跃故障, IDV8~IDV12 为随机变化故障, IDV13 为慢偏移故障, IDV14~IDV15 为堵塞干扰故障, IDV16~IDV21 为未知型故障。工艺流程如图 2 所示。具体反应过程如下:

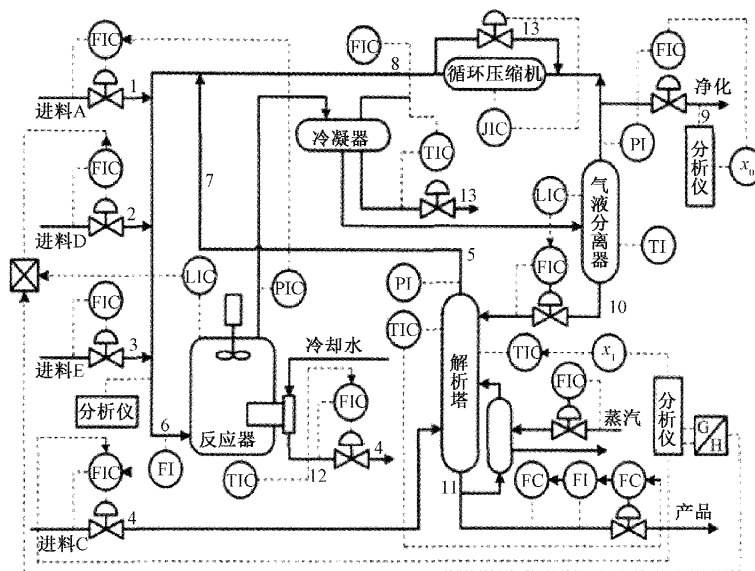
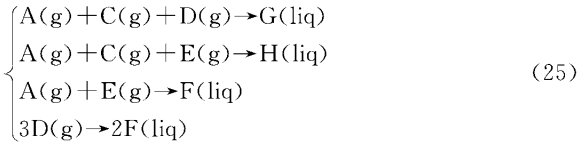


图 2 TE 工业流程



式中:用 g 和 liq 表示气体和液体,其中惰性不可溶进料 B 为催化剂。

3.1 样本选取

本文以 TE 过程正常数据、故障 1、故障 4、故障 7 和故障 12 为例,故障具体描述如表 1 所示。在训练集中,正常数据样本是在 25 h 运行仿真下获得,样本总数为 500;故障数据样本是在 24 h 运行仿真下获得,样本总数为 480。本文选取训练集为正常工况 400×52 ,故障数据 $4 \times 400 \times 52$,其中 4 是 4 类故障类型,400 是数据样本个数,52 是变量数。在测试集中,正常数据样本是在 48 h 运行仿真下获得,采样间隔时间为 3 min,样本总数为 960;故障数据样本是在 48 h 运行仿真下获得,故障在 8 h 的时候引入,样本总数为 960,其中,前 160 个数据样本为正常数据样本。本文选取测试集为正常工况 400×52 ,故障数据 $4 \times 400 \times 52$ 。

表 1 样本运行模式

故障模式	故障描述	类型
正常数据	—	—
故障 1	物料 A/C 进料比改变,物料 B 含量不变	阶跃
故障 4	反应器冷却入口温度改变	阶跃
故障 7	物料 C 压力损失	阶跃
故障 12	冷凝器冷却水入口温度改变	随机变化

3.2 实验过程

1)软件仿真环境为 MATLAB 2018b;

2)数据收集与预处理。对训练集和测试集进行标准化处理;

3)数据特征提取。使用 KEDA 方法来提取低维数据特征并消除冗余信息,以提高故障诊断正确率。其中,KECA 方法中核熵主成分由 95% 累计熵贡献率确定 $k = 23$; KEDA 中确定数据维数为 $t = 4$; 给定 γ 范围为 $2^{-8} \times 2^u (u = 0, 1, \dots, 10)$, 通过 1.2 节 1) 基于距离测度的矩阵相似性度量方法确定最优 γ 值为 4;

4)确定最优 SVM 模型。训练数据经 KEDA 特征提取后输入到 SVM 分类模型,使用 BAS 算法和五折交叉验证确定最佳的 SVM 模型。其中,BAS 算法中进化代数设置为 100; SVM 模型中 C 和 δ 的搜索范围分别设置为 $[0.0001, 300]$ 和 $[0.0001, 10]$ 。确定最佳惩罚参数为 $C = 182.9025$,核参数 $\delta = 10$ 。

3.3 实验结果分析

测试数据经 KEDA 特征提取后 BAS-SVM 的分类结果图如图 3 所示,该图横坐标为 5 种故障类型的数据样本

个数,纵坐标为正常数据、1、4、7、12 代表 TE 过程 5 种数据类型,对正常数据以及 4 类故障联合诊断,可以清晰的看到对于故障 1、故障 4、故障 7 以及故障 12 均保持很好的诊断结果,对于正常数据,小部分样本被错误划分到故障 1 和故障 12 中,诊断效果相比其他故障类型的诊断结果较差。

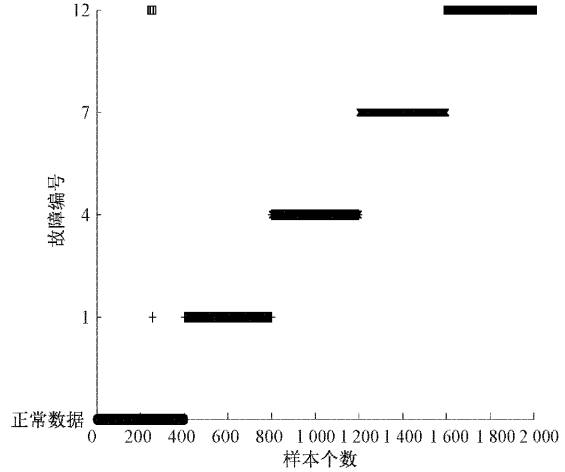


图 3 5 种故障类型数据诊断结果图

为了验证 KEDA 方法在故障诊断领域的优越性,分别采用 KECA-BAS-SVM、LDA-BAS-SVM 和 KEDA-BAS-SVM 对 TE 过程的 5 类故障类型测试数据样本进行故障诊断。不同特征提取方法对应的各故障类型数据样本诊断正确率如表 2 所示。

表 2 不同特征提取方法的诊断准确率

故障模式	KECA-BAS-SVM	LDA-BAS-SVM	KEDA-BAS-SVM
数据维数/%	23	4	4
正常数据/%	94.75	88.75	98.75
故障 1/%	100	100	100
故障 4/%	100	100	100
故障 7/%	96.5	100	100
故障 12/%	100	52.25	100
平均准确率/%	98.25	88.2	99.75
BAS-SVM 运行时间/s	58.235	25.645	27.959

从表 2 可以看出,LDA 的诊断准确率低于其他两种方法。KECA 的诊断正确率高于 LDA 的诊断正确率,KECA 最大程度减少数据 Renyi 熵损失,但降维后数据维数仍然较高,BAS-SVM 的运行时间相比 LDA 和 KEDA 的运行时间较长。KEDA 对于正常数据、故障 7、故障 12 的诊断准确率均有不同程度的提高,而且,数据经 KEDA 特征提取后,数据维数相比 KECA 明显减少,使得计算机存储容量更少,BAS-SVM 的运行时间相比 KECA 也减少 50% 左右,KEDA-BAS-SVM 在高诊断准确率前提下又能快速诊

断出结果,验证了 KEDA-BSA-SVM 方法在故障诊断领域的优越性。

4 结 论

本文提出一种 KEDA-BAS-SVM 的故障诊断方法。KEDA 在降低数据维度、保留数据最大 Renyi 熵的同时还保留了最具有判别力的类别信息,提高故障诊断的准确率。采用 BAS 与五折交叉验证搜索建立最佳超参数的 SVM 模型。通过仿真实验并与 KECA-BAS-SVM 和 LDA-BAS-SVM 方法对比。结果表明,数据经 KEDA 特征提取后,不仅诊断准确率相比 KECA 与 LDA 分别提高了 1.5% 和 11.55%,而且 BAS-SVM 运行时间最短。KEDA-BAS-SVM 目前不能对故障发生的位置进行判断,下一步将在故障定位方面进行研究。

参考文献

- [1] 彭开香,马亮,张凯.复杂工业过程质量相关的故障检测与诊断技术综述[J].自动化学报,2017,43(3):349-365.
- [2] 姚羽曼,罗文嘉,戴一阳.数据驱动方法在化工过程故障诊断中的研究进展[J].化工进展,2021,40(4):1755-1764.
- [3] 鲍宇,程硕,王靖涛.基于深度学习的化工过程故障检测与诊断研究综述[J].化学工业与工程,2022,39(2):9-22.
- [4] 张强,刘坤,雷国正,等.化工机械的障诊断与控制[J].能源技术与管理,2021,46(5):113-115.
- [5] 钱锋,杜文莉,钟伟民,等.石油和化工行业智能优化制造若干问题及挑战[J].自动化学报,2017,43(6):893-901.
- [6] 卞华棋,王育辉,王栋,等.基于 PCA 的空战机动高阶重构与评估[J].电子测量与仪器学报,2020,34(6):191-197.
- [7] 李嘉莹,赵丽,边琰,等.基于 LDA 和 KNN 的下肢运动想象脑电信号分类研究[J].国外电子测量技术,2021,40(1):9-14.
- [8] JENSSEN R. Kernel entropy component analysis[J]. IEEE Trans Pattern Anal Mach Intell, 2010, 32(5):847-860.
- [9] 李颖,张国林.互信息和核熵成分分析的油中溶解气体浓度建模[J].山东大学学报(理学版),2022,57(7):43-52.
- [10] 江风云,唐勇波.基于核熵成分分析的油中溶解气体浓度预测[J].控制工程,2020,27(8):1419-1424.
- [11] 刘春菊,刘春玲,李召.基于核熵成分分析结合独立元分析的故障检测方法[J].仪表技术与传感器,2016(9):94-97.
- [12] 潘泓,朱亚平,夏思宇,等.基于上下文信息和核熵成分分析的目标分类算法[J].电子学报,2016,44(3):580-586.
- [13] 周建民,陈超,涂文兵,等.红外热波技术、有限元与 SVM 相结合的复合材料分层缺陷检测方法[J].仪器仪表学报,2020,41(3):29-38.
- [14] 冯瑞杰,陈争光,衣淑娟.基于贝叶斯优化的 SVM 玉米品种鉴别研究[J].光谱学与光谱分析,2022,42(6):1698-1703.
- [15] 许敏.隐空间特征增强自标记半监督 SVM 分类新方法[J].统计与决策,2022,38(7):11-15.
- [16] JIANG X, LI S. BAS: Beetle antennae search algorithm for optimization problems[J]. Sciedu Press, 2018(1),DOI:10.5430/IJRC.V1N1P1.
- [17] 李周姿,冯跃,林卓胜,等.基于改进 PSO 算法优化 SVM 模型的面色识别[J].计算机仿真,2022,39(4):241-247.
- [18] 黄林生,阮超,黄文江,等.基于 GF-1 遥感影像和 relief-mRMR-GASVM 模型的小麦白粉病监测[J].农业工程学报,2018,34(15):167-175,314.
- [19] 罗源睿,王海瑞.基于蝙蝠算法优化支持向量机的滚动轴承故障诊断[J].化工自动化及仪表,2021,48(3):248-252,263.
- [20] 宋晖,薛云,张良均.基于 SVM 分类问题的核函数选择仿真研究[J].计算机与现代化,2011(8):133-136.
- [21] 李晓宇,张新峰,沈兰荪,等.一种确定径向基核函数参数的方法[J].电子学报 2005,33:2459-2463.
- [22] 王二化,刘颖.基于 PCA 和改进型 SVM 的齿轮裂纹故障诊断方法[J].机械设计与研究,2021,37(2):83-87,91.
- [23] 徐可,陈宗海,张陈斌,等.基于经验模态分解和支持向量机的滚动轴承故障诊断[J].控制理论与应用,2019,36(6):915-922.
- [24] ADELIM M, MAZINA A H. High efficiency fault-detection and fault-tolerant control approach in Tennessee Eastman process via fuzzy-based neural network representation [J]. Complex & Intelligent Systems, 2020:199-212.
- [25] ISURU A, UGUGAMA K V, GERNAEY M A, et al. A novel use for an old problem: The tennessee eastman challenge process as an activating teaching tool [J]. Education for Chemical Engineers, 2020, 30(C), DOI: 10.1016/j.ece.2019.09.002.

作者简介

李榕,硕士研究生,主要研究方向为基于数据驱动的故障诊断与检测。

E-mail:18409168093@163.com

申志,硕士研究生,主要研究方向为基于数据的故障诊断。

E-mail:shenzhi9696@163.com

李元,博士,教授,主要研究方向为基于数据驱动的故障诊断与检测。

E-mail:Li-yuan@mail.tsinghua.edu.cn