

一种面向工业过程的质量异常检测与故障量化评估方法

董洁^{1,2} 张伟^{1,2} 彭开香^{1,2} 马亮^{1,2}

摘要 质量异常检测 (Quality abnormality detection, QAD) 与故障量化评估 (Fault quantitative assessment, FQA) 作为工业过程监控的关键环节, 是故障诊断领域的研究热点. 本文提出了一种新的工业过程质量异常检测与故障量化评估方法. 首先, 采用弹性网络 (Elastic net, EN) 算法构建了质量相关的变量候选集, 借助典型相关分析 (Canonical correlation analysis, CCA) 构建了质量相关的特征向量, 并引入支持向量数据描述 (Support vector data description, SVDD) 实现质量异常检测. 其次, 从优化近邻点距离的角度提出了增强局部线性嵌入 (Enhanced local linear embedding, ELLE) 算法, 并提出了基于 CCA-ELLE 的质量异常故障量化评估方法. 最后, 通过田纳西-伊斯曼 (Tennessee-Eastman, TE) 过程进行仿真验证, 并与传统的方法进行对比分析, 实验结果验证了所提方法的优越性和有效性.

关键词 质量异常检测, 支持向量数据描述, 故障量化评估, 局部线性嵌入, 工业过程

引用格式 董洁, 张伟, 彭开香, 马亮. 一种面向工业过程的质量异常检测与故障量化评估方法. 自动化学报, 2022, 48(10): 2406-2415

DOI 10.16383/j.aas.c190880

A Novel Method of Quality Abnormality Detection and Fault Quantitative Assessment for Industrial Processes

DONG Jie^{1,2} ZHANG Wei^{1,2} PENG Kai-Xiang^{1,2} MA Liang^{1,2}

Abstract As the key link of industrial process monitoring, quality abnormality detection (QAD) and fault quantitative assessment (FQA) are the research hotspot in the field of fault diagnosis. In this paper, a novel framework of industrial process QAD and FQA is proposed. Firstly, the candidate set of quality related variables is constructed by elastic net (EN) algorithm. The quality related eigenvectors are constructed by canonical correlation analysis (CCA), and the support vector data description (SVDD) is introduced to detect quality anomaly. Secondly, an enhanced local linear embedding (ELLE) algorithm is proposed from the perspective of optimizing the distance between adjacent points, and a quantitative assessment method based on CCA-ELLE is proposed. Finally, the proposed framework is simulated and compared with some traditional methods by Tennessee-Eastman (TE) process simulation platform. The experimental results verify the superiority and effectiveness of the proposed method.

Key words Quality abnormality detection (QAD), support vector data description (SVDD), fault quantitative assessment (FQA), local linear embedding, industrial process

Citation Dong Jie, Zhang Wei, Peng Kai-Xiang, Ma Liang. A novel method of quality abnormality detection and fault quantitative assessment for industrial processes. *Acta Automatica Sinica*, 2022, 48(10): 2406-2415

质量异常的故障检测与量化评估技术是保证安全生产及获得可靠产品质量的有效手段, 是当前国

际过程控制领域的研究热点之一^[1]. 复杂工业过程的非线性、高维度特性给其应用带来了极大的挑战. 因此, 有必要建立准确可靠的监测模型质量异常的故障诊断, 为现场维护提供支持.

随着分布式控制系统 (Distributed control systems, DCS) 的广泛应用, 多元统计过程监测 (Multivariate statistical process monitoring, MSPM) 方法得到了广泛的关注^[2-3]. 主成分分析 (Principal component analysis, PCA) 作为 MSPM 的典型方法, 通过将高维空间分解为主元子空间和残差子空间, 进而采用 T^2 和平方预测误差 (Square predicted error, SPE) 统计量进行检测, 近年来广泛地应用到故障检测领域^[4-5]. 然而, 基于 PCA 的故障检

收稿日期 2019-12-24 录用日期 2020-04-06

Manuscript received December 24, 2019; accepted April 6, 2020

国家自然科学基金 (62273031, 61773053, 61873024), 中央高校基本科研业务费基金 (FRF-TP-19-049A1Z), 国家重点研发计划 (2017YFB0306403) 资助

Supported by National Natural Science Foundation of China (62273031, 61773053, 61873024), Fundamental Research Funds for the China Central Universities (FRF-TP-19-049A1Z), and National Key Research and Development Program of China (2017YFB0306403)

本文责任编辑 邓方

Recommended by Associate Editor DENG Fang

1. 北京科技大学自动化学院 北京 100083 2. 工业过程知识自动化教育部重点实验室 北京 100083

1. School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083 2. Key Laboratory of Knowledge Automation for Industrial Processes of Ministry of Education, Beijing 100083

测技术聚焦于故障是否发生, 而忽略了由过程变量引起的故障是否会影响最终产品的质量. 于是, 以偏最小二乘 (Partial least squares, PLS)、典型相关分析 (Canonical correlation analysis, CCA) 等为核心技术的质量异常检测 (Quality abnormality detection, QAD) 研究成果不断涌现. PLS 立足于两组变量之间的协方差最大化. 近年来基于 PLS 及其相关扩展模型的质量异常检测技术层出不穷. Zhou 等^[6] 将 PLS 提取过程变量中与质量相关空间的优势与主成分分析相结合, 构建了全潜结构投影 (Total-partial least squares, T-PLS) 模型, 实现了质量异常检测. Ding 等^[7] 针对静态过程, 提出了一种改进偏最小二乘 (Modified partial least squares, MPLS) 算法, 提高了分解后子空间可解释性和关键性能指标检测能力. Huang 等^[8] 通过核偏最小二乘法 (Kernel partial least squares, KPLS) 将主成分分析与质量异常检测相结合, 提高了质量监测的有效性. Wang 等^[9] 从模型鲁棒性的角度将正交信号校正 (Orthogonal signal correction, OSC) 与改进偏最小二乘 (Modified partial least square, MPLS) 方法相结合, 提高了模型对微小故障的敏感性, 增加了建模的复杂度, 但其检测性能与 MPLS 相当. Jiang 等^[10] 通过开发基于 MATLAB 的面向关键性能指标的故障检测工具箱, 对基于多变量统计分析的众多面向关键性能指标监测方法从建模原理、准确性、检测敏感性等多角度对比分析, 进一步验证了 MPLS 具有高计算效率, OSC-MPLS 具有良好的故障检测能力和灵敏性. 与 PLS 不同的是, CCA 通过最大化两组变量的相关系数来获得质量相关的投影矩阵, 取得了大量研究成果^[11]. Chen 等^[12] 将 CCA 应用于动态过程残差信号的构建, 提高了模型对动态过程的检测性能. Zhang 等^[13] 将 CCA 与分布式思想相结合, 改善了传统的质量异常检测的效果. 但是, 复杂工业过程的数据往往具有非线性、非高斯等特性, 传统的 MSPM 方法遭遇一定瓶颈. 支持向量数据描述 (Support vector data description, SVDD) 算法最早由 Tax 等^[14] 提出. 由于其对具有非线性、非高斯特性的数据具有很好的适应能力, 已经广泛地应用到复杂工业过程的故障检测领域. 为了提高模型的检测能力, 基于 SVDD 的故障检测方法不断涌现. Khediri 等^[15] 将核技术与 SVDD 相结合, 实现了多模态过程的非线性故障检测. Zhang 等^[16] 采用间隙度量进行数据预处理, 实现了流程工业的厂级监控. Zhu 等^[17] 采用粒子群优化算法对 SVDD 进行改进, 实现了检测模型的自适

应更新, 具有重要的现实意义.

故障检测虽然能够通过报警提示质量异常, 但却不能给出故障根源、传播路径以及故障程度的量化信息. 近年来国内外众多学者对相关问题进行了大量研究^[18-19]. 但是, 这些研究专注于故障根源与传播路径辨识, 缺少对故障进行准确的量化评估, 轻则造成维护费用的增加, 重则会导致故障的蔓延和恶化. 因此, 不同于故障根源与传播路径辨识, 对质量异常故障进行合理评估能够为维护人员提供更好的决策支持. Luo 等^[20] 提出了一种采用两级判别指标的方案, 将故障划分为严重故障、轻微故障和无害故障, 实现了故障等级评估. Song 等^[21] 将过程变量空间划分为 3 个性能相关的子空间, 并在此基础上进行了故障的严重程度分析. Yang 等^[22] 在操作性能评估的过程中采用加权主成分分析 (Weighted principal component analysis, WPCA) 的方法针对系统不同的性能状态分别建立检测模型, 采用 T^2 统计量构建健康指标, 实现了对系统状态的等级划分. 以上方法虽然能够从检测的角度对故障进行分析, 但是并没有给出精确的量化评估结果. 为此, Guo 等^[23] 采用拉普拉斯特征映射 (Laplacian eigenmap, LE) 进行低维投影, 得到了更加准确的退化指标. Yan 等^[24] 在评估指标的构建过程中综合考虑不同操作条件, 实现了多工况的状态监测. Sun 等^[25] 在设备健康监测中采用了核稀疏表示的局部线性嵌入算法 (Kernel sparse representation-local linear embedding algorithm, KS-LLE) 对轴承的不同故障程度进行了量化评估, 得到了良好的评估效果. Atamuradov 等^[26] 采用基于特征融合的监测方法, 实现了电机系统的状态评估. 上述方法虽然能够实现量化评估, 但并未分辨出故障是否造成质量异常, 进而不能提供准确合理的维护建议.

针对上述问题, 本文提出了一种新的工业过程质量异常检测与故障量化评估 (Fault quantitative assessment, FQA) 方法. 对比已有的工作, 本文的主要贡献归纳如下: 1) 采用弹性网络 (Elastic net, EN) 算法构建了更加可靠的质量相关的变量候选集, 能够为后续的检测与评估提供支持; 2) 采用 CCA 算法强化了候选集变量与质量变量的关系, 并通过 SVDD 实现了质量异常检测; 3) 从优化近邻点距离的角度改进了局部线性嵌入 (Local linear embedding, LLE) 算法, 并提出了 CCA-ELLE 算法, 将质量异常的故障样本投影到二维空间进行量化分析, 实现了准确的故障量化评估.

1 基于 CCA-SVDD 的质量异常的故障检测方法

1.1 质量相关的变量候选集

复杂工业过程测量变量较多, 变量间存在多重共线性. 为了降低候选变量集的信息冗余, 需要对过程变量进行特征选择. 为了更好地结合质量相关信息, 考虑采用正则化的方法进行变量选择.

给定输入矩阵 $X \in \mathbf{R}^{n \times m}$, n 为样本个数, m 为变量维数, $Y \in \mathbf{R}^{n \times 1}$ 代表输出变量. $\mathbf{x} \in \mathbf{R}^{1 \times m}$, $\mathbf{y} \in \mathbf{R}^{1 \times 1}$ 为样本在 X 和 Y 中对应的元素. 标准的最小二乘算法的思想是最小化残差平方和

$$\hat{\beta} = \arg \min_{\forall \mathbf{x}, \mathbf{y} \in [X, Y]} \sum \{(\mathbf{y} - \mathbf{x}\beta)^2\} \quad (1)$$

式中, $\hat{\beta} = \{\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_m\}^T$ 为预估回归系数向量. 然而, 最小二乘算法在模型预测和解释性上存在不足. 岭回归在最小二乘算法的基础上使用岭约束来收缩系数, 改进思想如下:

$$\hat{\beta} = \arg \min_{\forall \mathbf{x}, \mathbf{y} \in [X, Y]} \sum \{(\mathbf{y} - \mathbf{x}\beta)^2 + \lambda \|\beta\|_2\} \quad (2)$$

式中, $\|\beta\|_2 = \sum_{j=1}^m \beta_j^2$, λ 为正则系数. 岭回归算法虽然能够进行系数收缩, 但是不能起到特征选择的作用. LASSO (Least absolute shrinkage and selection operator) 算法针对这一问题给出了改进的思路

$$\hat{\beta} = \arg \min_{\forall \mathbf{x}, \mathbf{y} \in [X, Y]} \sum \{(\mathbf{y} - \mathbf{x}\beta)^2 + \lambda \|\beta\|_1\} \quad (3)$$

式中, $\|\beta\|_1 = \sum_{j=1}^p |\beta_j|$. 然而, LASSO 算法在迭代过程中稳定性较差. 为了补偿 LASSO 的不足, EN 算法将岭回归方法和 LASSO 有机结合^[27]:

$$\hat{\beta} = \arg \min_{\forall \mathbf{x}, \mathbf{y} \in [X, Y]} \sum \{(\mathbf{y} - \mathbf{x}\beta)^2 + \lambda_1 |\beta|_1 + \lambda_2 |\beta|^2\} \quad (4)$$

式中, λ_1 和 λ_2 为非负参数. 令 $\eta = \lambda_2 / (\lambda_1 + \lambda_2)$, 则优化问题转化为

$$\begin{aligned} \hat{\beta} = \arg \min_{\forall \mathbf{x}, \mathbf{y} \in [X, Y]} \sum |\mathbf{y} - \mathbf{x}\beta|^2 \\ \text{s.t. } (1 - \eta) |\beta|_1 + \eta |\beta|^2 \leq s \end{aligned} \quad (5)$$

式中, s 决定 EN 的收缩程度. 当 $s \rightarrow 0$ 时全部的变量会被消除. 相反, 当 $s \rightarrow \infty$ 时算法逼近标准的最小二乘算法. 参数 η 可以调节两种惩罚项的权重,

当 $\eta = 1$ 时该算法为基本的岭回归算法, 当 $\eta = 0$ 时为基本的 LASSO 算法. 本文将过程变量作为 EN 算法的输入变量, 质量变量作为输出变量, 通过合理选择收缩参数 s , 构建质量相关的最优变量候选集.

1.2 CCA-SVDD 算法

为了增强过程变量与质量变量的关系, 同时降低数据的多重共线性, 采用 CCA 算法构建质量相关的特征矩阵. 首先, 获得一对投影向量 α 和 β , 使得随机变量 $\mathbf{u} = \alpha^T X$ 与 $\mathbf{v} = \beta^T Y$ 具有最大的相关系数 $\rho = \text{corr}(\mathbf{u}, \mathbf{v})$. 目标函数为

$$\arg \max_{\alpha, \beta} \frac{\alpha^T \times \frac{1}{n} XY^T \times \beta}{\sqrt{\alpha^T \times \frac{1}{n} XX^T \times \alpha \times \beta^T \times \frac{1}{n} YY^T \times \beta}} \quad (6)$$

令 $S_{xx} = \frac{1}{n} XX^T$, $S_{yy} = \frac{1}{n} YY^T$, $S_{xy} = \frac{1}{n} XY^T$, 分别表示 X 与 Y 的协方差矩阵及组间协方差矩阵. 通过拉格朗日乘法将优化问题转化为一般的特征值分解问题

$$\begin{cases} S_{xy} S_{yy}^{-1} S_{xy}^T \alpha = \lambda^2 S_{xx} \alpha \\ S_{xy}^T S_{xx}^{-1} S_{xy} \beta = \lambda^2 S_{yy} \beta \end{cases} \quad (7)$$

式中, 特征值 λ 为随机变量 \mathbf{u} 与 \mathbf{v} 的相关系数. 由于特征矩阵由过程变量构成, 因此只需计算投影向量 α .

$$S_{xx}^{-1} S_{xy} S_{yy}^{-1} S_{xy}^T \alpha = \lambda^2 \alpha \quad (8)$$

令 $M = S_{xx}^{-1} S_{xy} S_{yy}^{-1} S_{xy}^T$, 进行特征值分解

$$M = \Gamma \Sigma \Gamma^T \quad (9)$$

取最大的前 d 项特征值对应的特征向量作为投影向量

$$W_{\text{offline}} = \Gamma \times I(1:d) \quad (10)$$

其中, $I(1:d)$ 为 d 维单位矩阵, $W_{\text{offline}} \in \mathbf{R}^{p \times d}$, p 为变量维度, d 为投影变量的维度.

在线测试阶段, 测试样本 X_{test} 经过如下处理:

$$U_{\text{opt}} = X_{\text{test}} W_{\text{offline}} \quad (11)$$

得到最优典型相关变量集.

给定典型相关变量集 $U_{\text{opt}} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n]^T \in \mathbf{R}^{n \times p}$, 构建包含全部目标集合的最小超球面, 进而通过该超球面对测试样本进行分类. 为了提高模型的实用性, 引入松弛变量 ξ_i , 得到目标函数

$$\begin{aligned} \min R^2 + C \sum_{i=1}^n \xi_i \\ \text{s.t. } \|x_i - a\|^2 \leq R^2 + \xi_i, \xi_i \geq 0 \end{aligned} \quad (12)$$

式中, C 可以调节容错率, R 为超球体的半径, a 为超球体的中心. 通过拉格朗日乘子法并转化为对偶问题, 则目标函数转化为

$$\begin{aligned} \max L &= \sum_{i=1}^n \sigma_i (\mathbf{u}_i \times \mathbf{u}_i) - \sum_{i=1}^n \sum_{j=1}^n \sigma_i \sigma_j (\mathbf{u}_i \times \mathbf{u}_j) \\ \text{s.t. } &0 \leq \sigma_i \leq C, \sum_{i=1}^n \sigma_i = 1 \end{aligned} \quad (13)$$

式中, σ_i ($i = 1, \dots, n$) 为样本 \mathbf{u}_i 对应的拉格朗日乘子. $\sigma_i \neq 0$ 的样本称为支持向量. 内积的计算引入了核函数 $(\mathbf{u}_i, \mathbf{u}_j) = \exp(-\|\mathbf{u}_i - \mathbf{u}_j\|/\rho^2)$, 参数 ρ 可根据经验调节. 因此, 得到超球体中心 a 和半径 R

$$a = \sum_{\mathbf{u}_i \in U_{SV}} \sigma_i \mathbf{u}_i \quad (14)$$

$$R = \sqrt{(\mathbf{u}_k \mathbf{u}_k) - 2\Sigma_1 + \Sigma_2} \quad (15)$$

式中, U_{SV} 为包含全部支持向量的集合, \mathbf{u}_k 为任何满足 $0 \leq \sigma_k \leq C$ ($k \in \{1, \dots, n\}$) 的样本, $\Sigma_1 = \sum_{\mathbf{u}_i \in U_{SV}} \sigma_i (\mathbf{u}_i \mathbf{u}_i)$, $\Sigma_2 = \sum_{\mathbf{u}_i \in U_{SV}} \sum_{\mathbf{u}_j \in U_{SV}} \sigma_i \sigma_j (\mathbf{u}_i \mathbf{u}_j)$. 对于测试样本 \mathbf{z} , 将式 (15) 中的 \mathbf{u}_k 替换为 \mathbf{z} , 得到与超球面中心的距离 D .

于是, 测试样本是否为故障状态可以通过如下方式判别:

$$\begin{cases} D - R \leq 0, & \text{正常} \\ D - R > 0, & \text{故障} \end{cases} \quad (16)$$

2 基于 CCA-ELLE 的故障量化评估方法

2.1 增强 LLE (ELLE) 算法

给定测试样本的最优典型相关变量集 $U_{\text{test}} \in \mathbf{R}^{l \times d}$, 计算如下:

$$U_{\text{test}} = X_{\text{test}} W_{\text{offline}} \quad (17)$$

式中, l 代表测试样本数量. LLE 算法包括以下 3 个步骤^[24]:

步骤 1. 获取每个样本的近邻点集合. 样本的近邻点集合包含 k 项元素, 可以通过 k 近邻的方法得到.

步骤 2. 计算样本与其近邻点的连接权重 \mathbf{w} , 可以通过最小化重构误差

$$\begin{aligned} \min_w &\|U_i - U_{\text{test}} \mathbf{w}_i\|^2 \\ \text{s.t. } &\sum_{i=1}^l w_{ij} = 1 \end{aligned} \quad (18)$$

得到权重矩阵 $W \in \mathbf{R}^{l \times l}$, 其中, $U_i \in U_{\text{test}}$ 为选取的

样本近邻点, \mathbf{w}_i 为近邻点对应的权重向量. 在权重矩阵的取值过程中, 不属于近邻域的样本权重为零. 在接下来的运算过程中 W 保持不变.

步骤 3. 给定 \mathbf{y}_i 为 U_i 的低维空间投影, 通过最小化重构误差得到目标函数

$$J_W(\mathbf{y}) = \min \sum_{i=1}^l \left\| \mathbf{y}_i - \sum_{j=1}^l w_{ij} \mathbf{y}_j \right\|^2 \quad (19)$$

目标函数等价于

$$J_W(\mathbf{y}) = \min_{i,j=1}^l M_{ij} \mathbf{y}_i^T \mathbf{y}_j \quad (20)$$

式中, $M = (I - W)^T (I - W)$. I 为单位矩阵. 目标函数求解的过程中存在两个约束 $\sum_{i=1}^l \mathbf{y}_i = 0$ 和 $\sum_{i=1}^l \mathbf{y}_i \mathbf{y}_i^T / n = I$. 矩阵 M 的前 h 项特征向量构成了 U_{test} 的低维表示. 由于第 1 个特征向量对应的特征值接近于 0, 该特征向量被排除.

LLE 算法只考虑了重构误差, 计算权值的条件为近邻点之间保持线性关系, 也就是只考虑了近邻点间的拓扑结构. 在构建近邻点集合之后, 并没有考虑样本数据与其近邻点的距离信息. 因此, 样本点与其距离较小的近邻点的权重可能会比与较远距离近邻点的权重小. 这将会导致通过 LLE 建立的模型不准确, 并且会严重影响评估的准确性.

为此, 本文提出了增强 LLE (Enhanced LLE, ELLE) 算法. 建立如下目标函数

$$\min J = \min(\tau J_W + (1 - \tau) J_Q) \quad (21)$$

式中, J_W 是 LLE 算法的目标函数. J_Q 为反映近邻点与样本点距离的目标函数, τ 为调节两部分权重的参数. 矩阵化表示为

$$\begin{aligned} J_Q(Y) &= \min \sum_{i=1}^l \left\| \mathbf{y}_i - \sum_{j=1}^l Q_{ij} \mathbf{y}_j \right\|^2 = \\ &\min Y^T (I - Q)^T (I - Q) Y \end{aligned} \quad (22)$$

式中, $Q \in \mathbf{R}^{l \times l}$ 表示样本点与其近邻点距离的权重矩阵. 为了提高算法对非线性数据的处理能力, Q 矩阵元素的计算引入了核方法

$$Q_{ij} = \exp\left(-\frac{\|U_i - U_j\|}{\delta^2}\right) \quad (23)$$

其中, U_i 和 U_j 为不同样本点. 权重矩阵 Q 增强了样本点与其近邻点的距离. 由于距离的计算函数为减函数, 距离较小的近邻点对应的值相比于距离较

远的近邻点对应的值更大. 参数 δ 可以调节衰减率.

ELLE 算法的目标函数表示为

$$J = \tau J_W + (1 - \tau) J_Q = Y^T [\tau(I - W)^T(I - W) + (1 - \tau)(I - Q)^T(I - Q)]Y = Y^T \hat{M}Y \quad (24)$$

后续的计算过程与 LLE 算法相同.

2.2 量化评估指标的构建

与 LLE 算法类似, ELLE 算法会保留原始数据的拓扑结构, 低维数据保留着高维数据的局部几何结构. 定义正常状态的样本中心点与原点构成的方向向量为 $\vec{\gamma}_0$, 测试样本点与原点构成的方向向量为 $\vec{\gamma}_i$. 定义 $\theta_i = \vec{\gamma}_0 \times \vec{\gamma}_i$, 故障的严重程度指标定义为 $SI(i)$, 取值范围为 $[0^\circ, 360^\circ]$. 计算方法为

$$SI(i) = \begin{cases} \arccos(\vec{\gamma}_0, \vec{\gamma}_i), & \theta_i \geq 0 \\ 360^\circ - \arccos(\vec{\gamma}_0, \vec{\gamma}_i), & \theta_i < 0 \end{cases} \quad (25)$$

其中, $\arccos(\cdot, \cdot)$ 为两方向向量内积的反余弦函数, 此处计算的值为角度. 由于故障程度定义在二维空间, 指标计算要在 $[0^\circ, 360^\circ]$ 的范围内. 评估方法示意图如图 1 所示, 样本故障程度满足 $SI(2) > SI(1)$. 为了更直观地进行比较, 使评估结果更具有普遍意义, 同时提高评估结果的精度和可靠性, 采用每一种状态的样本中心点代表当前状态, 对评估结果进行归一化.

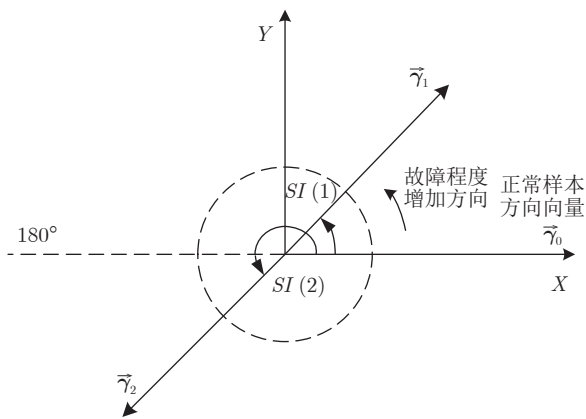


图 1 评估指标示意图

Fig. 1 Schematic diagram of evaluation indicator

2.3 质量异常检测与故障量化评估流程

本文提出的质量异常检测与故障量化评估框架包含离线建模和在线检测两部分.

2.3.1 离线阶段

1) 获取正常状态的过程数据, 并进行数据预处理 (z-score 标准化, 数据平滑处理).

2) 采用 EN 算法构建质量相关的变量候选集, 为之后的质量异常检测与故障量化评估提供支持.

3) 采用 CCA 算法获得投影矩阵 $W_{\text{offline}} \in \mathbf{R}^{p \times d}$ 和变量集 U_{opt} , 同时将 U_{opt} 送入 SVDD 模型中训练检测模型.

2.3.2 在线阶段

1) 获取测试数据 $X_{\text{test}} \in \mathbf{R}^{l \times p}$, 并进行数据预处理 (z-score 标准化, 数据平滑处理).

2) 送入已建立的 CCA 模型, 获得典型相关特征矩阵 $U_{\text{test}} \in \mathbf{R}^{l \times d}$, 进而送入 SVDD 模型进行判别. 根据式 (16) 的计算结果, 如果出现质量异常, 则进行下一步, 否则, 继续执行检测部分.

3) 采用 ELLE 进行投影, 得到二维空间表示的样本 $Y_{\text{test}} \in \mathbf{R}^{l \times 2}$, 进而获得表征测试样本的方向向量 $\vec{\gamma}_i$ 和表征正常样本的方向向量 $\vec{\gamma}_0$.

4) 计算测试样本间方向向量的夹角, 即故障的量化评估指标 $SI(i)$.

所提方案的流程图如图 2 所示.

3 仿真研究

3.1 TE 过程描述

TE (Tennessee-Eastman) 过程由 5 个主要单元: 反应器、冷凝器、压缩机、分离器、汽提塔组成; 存在 8 种成分, 其中, A, B, C, D, E 为反应物, G, H 为产物, F 为副产物^[28]. 本文的实验是在模态 1 的基础上完成的. 22 个过程变量作为过程监测的目标变量, 详细情况见表 1. TE 过程的 Simulink 仿真模型含有 28 中不同的故障类型, 每一种故障可以通过设置参数来决定是否注入. 对同一种类型的故障可以通过设置参数来调节注入故障的严重程度 (Fault severity, FS), FS 的调节范围为 $[0, 1]$, 代表过程从 $FS = 0$ 到 $FS = 1$ 的不同状态.

3.2 参数分析

本文提出的算法存在两个主要参数: 特征维数 d 和近邻点数量 k . 下面分别讨论各参数对算法评估效果的影响.

1) CCA 特征融合后的特征维数 d

较少的特征维数会破坏原始数据的近邻结构, 而维数较高则会引入冗余信息. 在参数分析过程中, 分别将 d 设置为 3, 4, 5, 6, 7. 仿真结果见图 3(a).

2) 近邻点数量 k

近邻点选择较少会破坏原始数据的近邻结构,

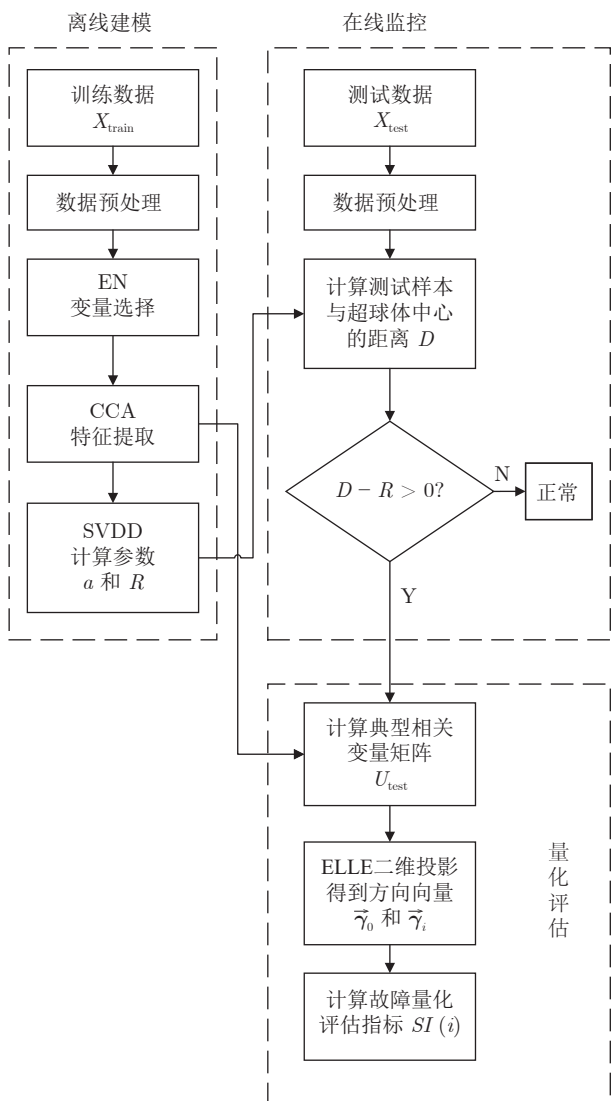
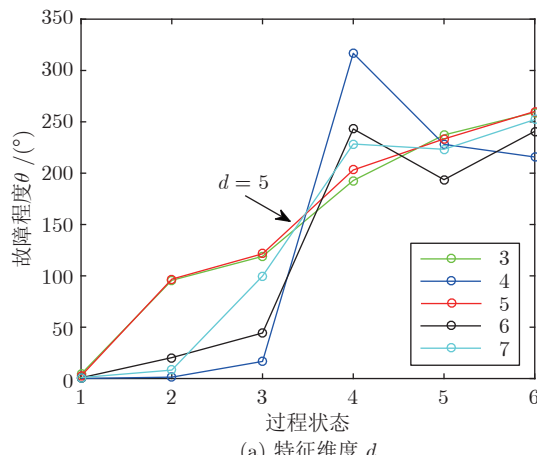


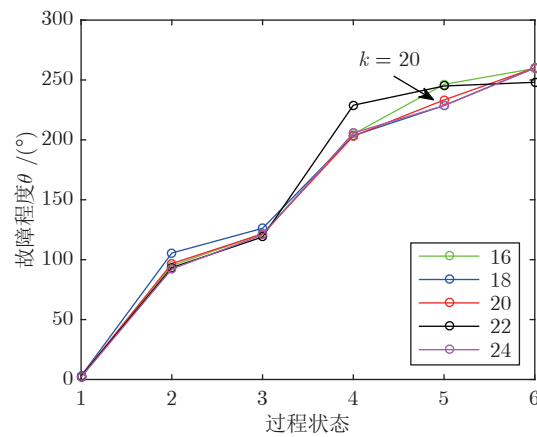
图 2 质量异常检测与故障量化评估流程图
Fig.2 Flowchart of QAD and FQE

表 1 TE 过程变量
Table 1 Process variables in the TE process

变量描述	变量描述
1) 物料 A 流量	12) 气/液分离器液位
2) 物料 D 流量	13) 气/液分离器压力
3) 物料 E 流量	14) 气/液分离器出口流量
4) 物料 C 流量	15) 汽提塔液位
5) 压缩机返回物料流量	16) 汽提塔压力
6) 反应器给料流量	17) 汽提塔塔底流量
7) 反应器压力	18) 汽提塔温度
8) 反应器液位	19) 汽提塔蒸汽流量
9) 反应器温度	20) 压缩机功率
10) 排空物料流量	21) 反应器冷却水出口温度
11) 气/液分离器温度	22) 冷凝器冷却水出口温度



(a) 特征维度 d
(a) Characteristic dimension d



(b) 近邻点数量 k
(b) Number of neighboring points k

图 3 参数分析

Fig.3 Analysis of parameters

近邻点选择较多则会忽略数据的局部特征. 在参数分析过程中, 分别将 k 设置为 16, 18, 20, 22, 24. 仿真结果如图 3(b).

图 3 中横坐标代表从 FS = 0 到 FS = 1 的 6 种不同的故障工况, 纵坐标代表量化评估结果. 为了进行准确的量化评估, 评估结果需要满足两个条件: 1) 不同过程状态的评估结果需符合实际故障程度; 2) 相邻过程状态的评估结果间隔均匀. 可以看出, 图 3(a) 中 $d = 5$ 的曲线和图 3(b) 中 $k = 20$ 的曲线均为最优情况. 因此本文的特征维数设置为 $d = 5$, 近邻点数量设置为 $k = 20$.

3.3 实验结果

采用 EN 算法对正常状态的数据样本进行分析, 选取产物 G 的成分含量作为质量变量, 式 (5) 中的参数设置为 $\eta = 0.2$, 构建了质量相关的最优变量候选集 $X = \{x_1, x_2, x_3, x_4, x_7, x_{10}, x_{13}, x_{20}, x_{21}\}$. 为了验证所提方案, 本文选取 TE 过程标准故障类

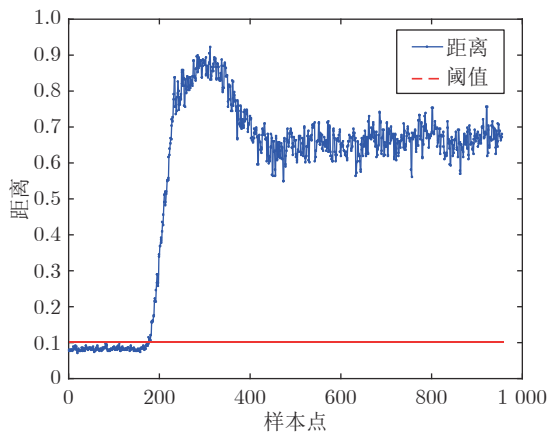
型 2 进行验证, 通过对 TE 过程注入不同程度的故障, 获得了相应的测试数据. 具体情况见表 2.

表 2 验证数据集

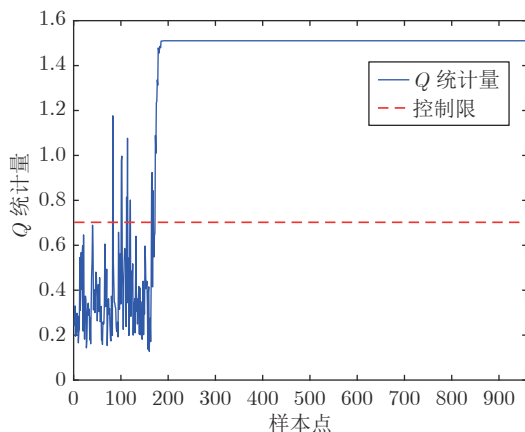
Table 2 Data sets used for validation

故障程度	样本数量
正常状态样本 (TE 标准数据)	500 (训练集)
故障数据 (TE 标准数据)	960 (测试集)
正常状态样本 (生成数据 FS = 0)	500 + 160 (训练集 + 测试集)
故障程度 1 (生成数据 FS = 0.2)	160 (测试集)
故障程度 2 (生成数据 FS = 0.4)	160 (测试集)
故障程度 3 (生成数据 FS = 0.6)	160 (测试集)
故障程度 4 (生成数据 FS = 0.8)	160 (测试集)
故障程度 5 (生成数据 FS = 1.0)	160 (测试集)

为了验证 CCA-SVDD 算法的有效性和优越性, 选择广泛应用于质量异常检测的 KPLS 算法进行对比, 控制限的置信度设置为 95%. 两种方法检测结果如图 4 所示.



(a) CCA-SVDD 检测结果
(a) CCA-SVDD-based detection result



(b) KPLS 检测结果
(b) KPLS-based detection result

图 4 两种方法的故障检测结果

Fig.4 Detection results of the two methods

从表 3 的对比结果可以看出, 传统的 KPLS 方法存在较高的误报率, 而本文提出的 CCA-SVDD 算法在保证高检测率与较低计算复杂度的前提下, 大大降低了误报率, 检测效果良好. 上述实验均在 MATLAB R2019b 平台上完成, 硬件配置为: i5 @1.6 GHz, RAM 6 GB.

表 3 两种方法的性能比较

Table 3 Comparison of the two methods

方法	误报率 (%)	检测率 (%)	建模运行时间 (s)
KPLS	3.75	98.375	0.307
CCA-SVDD	0	97.75	0.033

为了验证所提故障量化评估方案的有效性, 本文采用 TE 过程标准故障类型 2 进行验证, 式 (25) 的参数设置为 $\tau = 0.8$. 表 2 中的生成故障数据集对应的 FD 值满足递增的关系.

首先, 采用 TE 过程标准故障类型 2 进行验证. 将正常状态 (*) 与故障状态 (·) 的数据用两种标识符号来表示, 从图 5 中可以看出, 两种状态的样本点具有清晰的样本中心, 并且故障状态样本点的分布说明了过程从正常状态到故障状态是有向分布的, 而且样本点与原点构成的方向向量能够很好地反映状态的转变过程.

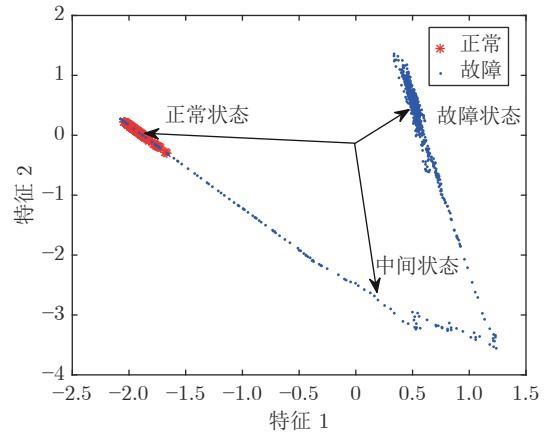


图 5 CCA-ELLE 二维投影

Fig.5 CCA-ELLE-based two-dimensional projection

图 6 是通过 TE 模型生成的不同程度故障数据对所提方案进行验证的结果. 图中不同的故障程度采用不同类型的标识符号表示. 可以看出, 随着故障程度的不断增加, 测试样本点与原点构成的方向向量具有连续变化的趋势, 并且随着数据的不断延伸, 不同类型的样本点可清晰地分开.

图 7(a) 和图 7(b) 为两种方法的量化评估结果. 图 7(c) 和图 7(d) 为归一化的评估结果. 可以看出,

评估指标越接近 1, 故障状态偏离正常状态越大, 故障程度越高. 并且传统的 LLE 算法对于故障程

度 4 的评估结果与案例不符. 而本文改进的 CCA-ELLE 算法可以准确地量化评估不同程度的故障,

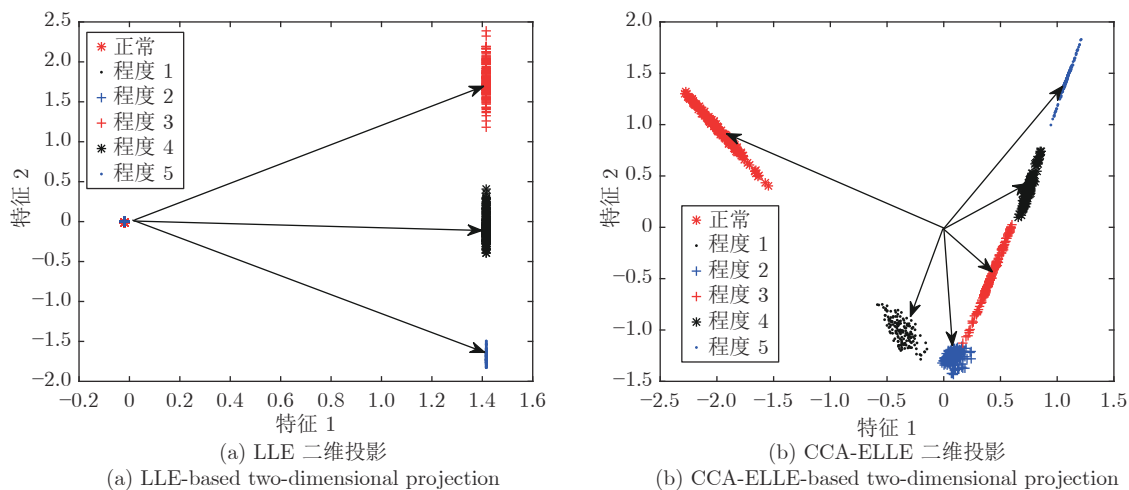


图 6 两种方法的二维投影

Fig.6 Two-dimensional projection results of the two methods

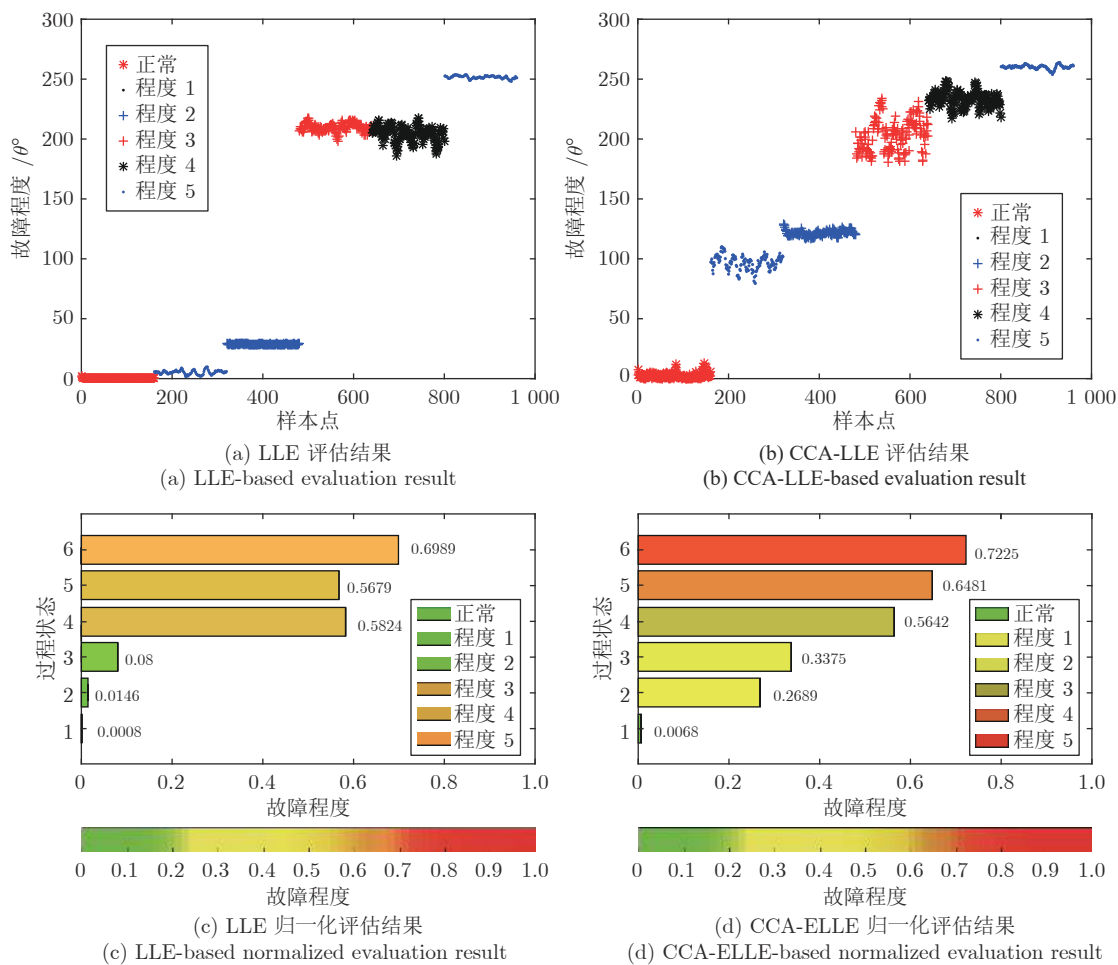


图 7 两种方法的量化评估结果

Fig.7 Evaluation results of the two methods

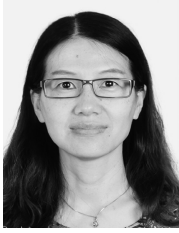
而且不同程度的故障样本点的间隔较均匀,与验证案例相符。

4 结束语

本文提出了一种新的工业过程质量异常检测与量化评估框架。首先,采用弹性网络算法筛选出可靠的质量相关的变量候选集;其次,提出了 CCA-SVDD 算法进行质量异常监测,并取得良好效果;再次,从优化近邻点距离的角度提出了 ELLE 算法,并与 CCA 相结合实现了质量异常故障的量化评估;最后,通过 TE 过程进行仿真验证,并与传统的方法进行对比分析,实验结果验证了所提方法的优越性和有效性。下一步的工作是将所提框架应用到多故障多种程度的质量异常检测与故障量化评估中。

References

- Peng Kai-Xiang, Ma Liang, Zhang Kai. Review of quality-related fault detection and diagnosis techniques in complex industrial processes. *Acta Automatica Sinica*, 2017, **43**(3): 349–365 (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. *自动化学报*, 2017, **43**(3): 349–365)
- Krzanowski W J. A user's guide to principal components. *Journal of the Royal Statistical Society: Series A*, 1992, **155**(1): 173–174
- Qin S J. Survey on data-driven industrial process monitoring and diagnosis. *Annual Reviews in Control*, 2012, **36**(2): 220–234
- Ma Jie, Li Gang, Chen Mo. Fault prediction method of rotating machinery based on nonlinear fault reconstruction. *Acta Automatica Sinica*, 2014, **40**(9): 2045–2049 (马洁, 李钢, 陈默. 基于非线性故障重构的旋转机械故障预测方法. *自动化学报*, 2014, **40**(9): 2045–2049)
- Wang Jing, Hu Yi, Shi Hong-Bo. Intermittent process fault detection based on GMM. *Acta Automatica Sinica*, 2015, **41**(5): 899–905 (王静, 胡益, 侍洪波. 基于 GMM 的间歇过程故障检测. *自动化学报*, 2015, **41**(5): 899–905)
- Zhou D H, Li G, Qin S J. Total projection to latent structures for process monitoring. *AIChE Journal*, 2010, **56**(1): 168–178
- Ding S X, Yin S, Peng K X, Hao H Y, Shen B. A novel scheme for key performance indicator prediction and diagnosis with application to an industrial hot strip mill. *IEEE Transactions on Industrial Informatics*, 2013, **9**(4): 2239–2247
- Huang J P, Yan X F. Quality-driven principal component analysis combined with kernel least squares for multivariate statistical process monitoring. *IEEE Transactions on Control System Technology*, 2019, **27**(6): 2688–2695
- Wang G, Yin S. Quality-related fault detection approach based on orthogonal signal correction and modified PLS. *IEEE Transactions on Industrial Informatics*, 2015, **11**(2): 398–405
- Jiang Y C, Yin S. Recent advances in key-performance-indicator oriented prognosis and diagnosis with a MATLAB toolbox: DB-KIT. *IEEE Transactions on Industrial Informatics*, 2019, **15**(5): 2849–2858
- Shang Lin-Yuan, Tian Xue-Min, Cao Yu-Ping, Cai Lian-Fang. Performance monitoring and diagnosis of MPC based on PLS cross product matrix non-similarity analysis. *Acta Automatica Sinica*, 2017, **43**(2): 271–279 (尚林源, 田学民, 曹玉苹, 蔡连芳. 基于 PLS 交叉积矩阵非相似度分析的 MPC 性能监控与诊断. *自动化学报*, 2017, **43**(2): 271–279)
- Chen Z, Ding S X, Zhang K. Canonical correlation analysis-based fault detection methods with application to alumina evaporation process. *Control Engineering Practice*, 2016, **46**: 51–58
- Zhang K, Peng K X, Ding S X, Chen Z W, Yang X. A correlation-based distributed fault detection method and its application to a hot tandem rolling mill process. *IEEE Transactions on Industrial Electronics*, 2013, **9**(4): 2226–2238
- Tax D M J, Duin R P W. Support vector domain description. *Pattern Recognition Letters*, 1999, **20**(11): 1191–1199
- Khediri I B, Weihs C, Limam M. Kernel k-means clustering based local support vector domain description fault detection of multimodal processes. *Expert Systems with Applications*, 2012, **39**(2): 2166–2171
- Zhang C F, Peng K X, Dong J. A novel plant-wide process monitoring framework based on distributed Gap-SVDD with adaptive radius. *Neurocomputing*, 2019, **350**: 1–12
- Zhu K D, Mei F, Zheng J Y. Adaptive fault diagnosis of HVCBs based on P-SVDD and P-KFCM. *Neurocomputing*, 2017, **240**: 127–136
- Zhu Q Q, Qin S J. Supervised diagnosis of quality and process faults with canonical correlation analysis. *Industrial and Engineering Chemistry Research*, 2019, **58**(26): 11213–11223
- Ma L, Dong J, Peng K X. A novel key performance indicator oriented hierarchical monitoring and propagation path identification framework for complex industrial processes. *ISA Transactions*, 2020, **96**(1): 1–13
- Luo L J, Lovelett R J, Ogunnaike B A. Hierarchical monitoring of industrial processes for fault detection, fault grade evaluation, and fault diagnosis. *AIChE Journal*, 2017, **63**(7): 2781–2795
- Song B, Zhou X G, Shi H B, Yang T. Performance indicator oriented concurrent subspace process monitoring method. *IEEE Transactions on Industrial Electronics*, 2018, **65**(2): 1508–1571
- Yang T, Shi H B, Song B, Tan S. Operating performance assessment and non-optimal cause identification for chemical process. *Canadian Journal of Chemical Engineering*, 2017, **97**(1): 1475–1487
- Guo S, Sun Y F, Wu F Z, Li Y H. Integrating laplacian eigenmaps feature space conversion into deep neural network for equipment condition assessment. *Automatic Control and Computer Sciences*, 2018, **52**(6): 465–475
- Yan H, Liu K B, Zhang X, Shi J J. Multiple sensor data fusion for degradation modeling and prognostics under multiple operational conditions. *IEEE Transactions on Reliability*, 2016, **65**(3): 1416–1426
- Sun C, Wang P, Yan R Q, Gao R X, Chen X F. Machine health monitoring based on locally linear embedding with kernel sparse representation for neighborhood optimization bearing fault diagnosis. *Mechanical System and Signal Processing*, 2019, **114**: 25–34
- Atamuradov V, Medjaher K, Camci F, Dersin P, Zerhouni N. Railway point machine prognostics based on feature fusion and health state assessment. *IEEE Transactions on Instrumentation and Measurement*, 2019, **68**(8): 2691–2704
- Zhang F, Sun K, Wu X L. A novel variable selection algorithm for multi-layer perceptron with elastic net. *Neurocomputing*, 2019, **361**: 110–118
- Downs J, Vogel E F. A plant-wide industrial process control problem. *Computers and Chemical Engineering*, 1993, **17**(3): 245–255



董洁 北京科技大学自动化学院教授. 2007 年获得北京科技大学控制科学与工程博士学位. 主要研究方向为智能控制理论与应用, 过程监控与故障诊断, 复杂系统建模与控制.

E-mail: dongjie@ies.ustb.edu.cn

(DONG Jie Professor at the School of Automation and Electrical Engineering, University of Science and Technology Beijing. She received her Ph.D. degree in control science and engineering from University of Science and Technology Beijing in 2007. Her research interest covers intelligent control theory and application, process monitoring and fault diagnosis, and complex system modeling and control.)



彭开香 北京科技大学自动化学院教授. 2007 年获得北京科技大学控制科学与工程博士学位. 主要研究方向为复杂工业系统故障诊断与容错控制. 本文通信作者.

E-mail: kaixiang@ustb.edu.cn

(PENG Kai-Xiang Professor at the School of Automation and Electrical Engineering, University of Science and Technology Beijing. He received his Ph.D. degree in control science and engineering from University of Science and Technology Beijing in 2007. His research interest covers fault diagnosis and fault-tolerant control for complex industrial system. Corresponding author of this paper.)



张伟 北京科技大学自动化学院硕士研究生. 主要研究方向为数据驱动的故障诊断与容错控制.

E-mail: zw719228639@163.com

(ZHANG Wei Master student at University of Science and Technology Beijing. His research interest covers data-based fault diagnosis and fault-tolerant control.)



马亮 北京科技大学自动化学院副教授. 2019 年获得北京科技大学控制理论与控制工程博士学位. 主要研究方向为数据驱动的故障诊断与容错控制.

E-mail: mlypplover@sina.com

(MA Liang Associate professor at the School of Automation and Electrical Engineering, University of Science and Technology Beijing. He received his Ph.D. degree in control science and engineering from University of Science and Technology Beijing in 2019. His research interest covers data-based fault diagnosis and fault-tolerant control.)