

基于独立元分析的制浆造纸 废水处理过程故障检测

杨冲¹ 宋留¹ 刘鸿斌^{1,2,*}

(1. 南京林业大学林业资源高效加工利用协同创新中心, 江苏南京, 210037;

2. 华南理工大学制浆造纸工程国家重点实验室, 广东广州, 510640)

摘要: 为及时、准确地做出故障诊断, 本课题采用独立元分析(ICA)和主成分分析(PCA)两种常用的多元统计分析方法对制浆造纸废水处理过程中的传感器故障进行检测并对诊断效果进行对比。结果表明, 对于制浆造纸废水数据中偏移和漂移两种故障, ICA模型的故障检测率分别为24%与54%, PCA模型的故障检测率分别为14%和42%, ICA模型的两种故障检测率均高于PCA模型, 但是两种模型均无法达到满意的检测效果; 对于完全失效故障, ICA和PCA模型的故障检测率均达到100%。

关键词: 制浆造纸废水处理过程; 故障检测; 主成分分析; 独立元分析

中图分类号: X793

文献标识码: A

DOI: 10.11981/j.issn.1000-6842.2019.01.66

随着现代工业过程的规模与复杂程度的提高, 工业过程的安全运行和产品质量得到更多的关注, 故障检测与分析技术也越来越不可或缺^[1]。

一般来说, 故障检测与分析可以分为3类: 基于解析模型、基于专家知识和基于数据驱动的故障诊断方法。基于解析模型的方法可以使系统根据特定的模型得出精确的结果, 但是随着系统复杂程度的提高, 解析模型构造的难度与成本也随之上升^[1]。基于专家知识的方法依赖于已存在的知识库与专家的经验知识以解决一些非结构化的问题, 适用于过程尚未成熟的领域, 但是完全依靠专家凭感觉、个人经验及简单仪表来判断和排除故障成为该方法的局限性^[2]。相对于前两种故障诊断方法, 基于数据驱动的方法没有以上限制, 该方法在不需知道系统精确解析模型的情况下可对系统运行数据进行采集、建模、分析, 从而完成系统的故障诊断。

制浆造纸废水的处理过程具有较强的时变性、耦合性、非线性以及易受外界干扰的特性。作为制浆造纸废水处理过程中使用最广泛的方法——活性污泥法主要包括了预处理、初次沉降、曝气、二次沉降和污泥回流5个部分(如图1所示)。活性污泥法处理制浆造纸废水的原理是利用在曝气池内的微生物群体的凝聚、吸附、氧化分解等作用去除制浆造纸废水中的

有机物。在每一部分处理工艺中都需要测量大量的在线参数, 用来保证出水水质达标以及完成控制过程的优化, 因此, 该方法对测量仪表的精密性有较高的要求。测量仪表精密度的提高增加了故障出现的概率, 结合制浆造纸废水处理过程的恶劣工作环境以及设备长期满负荷运行导致老化等因素, 故障检测结果的时效性与准确性需进一步提高^[3]。

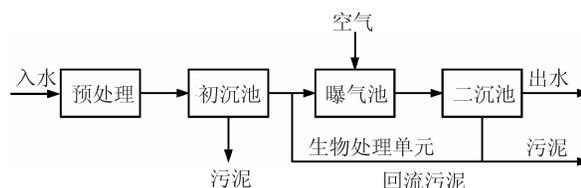


图1 制浆造纸废水活性污泥法处理流程图

制浆造纸废水处理复杂的过程特性使我们难以对其构建准确的数学模型与完备的专家知识库; 另一方面, 制浆造纸废水处理厂的长期满负荷运行使其积累了大量的过程历史数据, 因此采用数据驱动的方法对制浆造纸废水处理过程的故障进行诊断逐渐受到学术界和工业界的重视。制浆造纸废水处理领域中, 基于数据驱动的故障诊断方法是以机器学习与多元统计方法为主。机器学习方法的核心在于机器的自我学习、改善性能, 主要包括人工神经网络、贝叶斯和支持向量机等^[3]。多元统计方法的提出是区别于单变量统

收稿日期: 2017-09-30

基金项目: 制浆造纸工程国家重点实验室开放基金资助项目(201813, 201610); 南京林业大学高层次人才科研启动基金(163105996)。

作者简介: 杨冲, 男, 1994年生; 在读硕士研究生; 主要研究方向: 制浆造纸过程监测与控制。

*通信联系人: 刘鸿斌, 副教授; 主要研究方向: 复杂工业过程的建模、监测和控制; E-mail: hongbinliu@njfu.edu.cn。

计方法,主要利用多变量之间的相关性及内在的统计规律进行故障检测,适用于具有强相关性的制浆造纸废水处理系统^[4]。多元统计方法是将含多变量的高维数据通过多元投影的方式进行降维,在相对独立的低维空间构造统计量进行数据分析。

近年来,随着多元统计过程监测理论的发展,独立元分析法(Independent Component Analysis, ICA)获得了较大的关注。针对系统运行过程中检测变量间的相关性,ICA方法可生成一个模型,该模型假设数据变量由少数潜在变量线性组合而成,相关系数未知。这些潜在变量称为数据的独立元,相互独立且非高斯。ICA模型旨在提取潜在变量以及混合矩阵的相关系数并结合相应的统计量进行故障检测^[5]。主成分分析法(Principal Component Analysis, PCA)在流程工业领域也有广泛的应用。尽管ICA模型可看作PCA模型的有效延伸,但两者迥异。PCA模型旨在降维,将多个变量映射为少数几个主成分,这些主成分可由原始变量线性组合而成,并且包含着原始变量的大部分信息,为避免主成分包含的信息相互重叠,该方法要求各主成分互不相关^[6]。PCA模型在计算统计量控制限时,需要假设数据服从高斯分布,如果测量数据非高斯,那么该方法的故障诊断效果将受到影响^[7]。

结合制浆造纸废水处理过程的特点,本课题采用多元统计方法中的PCA与ICA两种方法对该过程数据进行故障检测与对比分析。

1 方法原理

1.1 PCA模型

基于PCA模型可以快速、有效地处理大量过程数据的特点,从20世纪80年代起,PCA模型开始被广泛应用于数据降维、图像处理、特征提取、模式识别以及过程控制等领域^[8]。

1.1.1 数据模型

假设 $x \in \mathbf{R}^m$ 表示含 m 个样本的向量,每个样本均有 N 个独立采样。由此构造数据矩阵 $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]^T \in \mathbf{R}^{N \times m}$ 并进行标准化处理。PCA模型对标准化后的矩阵 \mathbf{X} 进行分解^[9]:

$$\mathbf{X} = \mathbf{TP}^T + \tilde{\mathbf{X}} = \mathbf{TP}^T + \tilde{\mathbf{T}}\tilde{\mathbf{P}}^T = [\mathbf{T}\tilde{\mathbf{T}}][\mathbf{P}\tilde{\mathbf{P}}]^T \equiv \tilde{\mathbf{TP}}^T \quad (1)$$

式中, $\mathbf{P} \in \mathbf{R}^{m \times l}$ 为负载矩阵; $\mathbf{T} \in \mathbf{R}^{N \times l}$ 代表得分矩阵, l 代表主元个数; $\tilde{\mathbf{X}} = \tilde{\mathbf{T}}\tilde{\mathbf{P}}^T$ 为残差矩阵,且 $\tilde{\mathbf{T}} = [\mathbf{T}\tilde{\mathbf{T}}]$, $\tilde{\mathbf{P}} = [\mathbf{P}\tilde{\mathbf{P}}]$ 。标准化后样本 \mathbf{X} 的协方差矩阵估计值 \mathbf{S} 可表示为:

$$\mathbf{S} = \frac{1}{N-1} \mathbf{X}^T \mathbf{X} = \overline{\mathbf{P}} \overline{\mathbf{A}} \overline{\mathbf{P}}^T \quad (2)$$

则

$$\overline{\mathbf{A}} = \frac{1}{N-1} \overline{\mathbf{T}}^T \overline{\mathbf{T}} = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_m\} \quad (3)$$

式中, $\overline{\mathbf{A}} = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_m\}$ 为 \mathbf{S} 的 m 个特征值并按照降序排列,特征值对应的特征向量为 $\overline{\mathbf{P}}$ 的 m 个列向量,这些特征向量代表的信息量也按照降序排列, $\overline{\mathbf{P}}$ 为标准正交。这里采用累积贡献率来选取主元个数 l ,总方差中第 i 个主成分的方差所占的比例称为该主成分 p_i 的贡献率,主成分 p_i 的贡献率反映了第 i 个主成分包含的信息量大小。一般通过使前 l ($l \leq m$)个主成分的累计贡献率不低于某一数值(如80%)的方法来确定 l 的数值^[6]。

PCA模型建立之后,主元子空间(Principal Component Subspace, PCS)和残差子空间(Residual Subspace, RS)可分别表示为 $S_p = \text{span}\{\mathbf{P}\}$, $S_r = \text{span}\{\tilde{\mathbf{P}}\}$ 。则任意样本向量在PCS与RS上的投影可分别表示为^[9]:

$$\hat{\mathbf{x}} = \mathbf{PP}^T \mathbf{x} \equiv \mathbf{C} \mathbf{x} \in S_p \quad (4)$$

$$\tilde{\mathbf{x}} = \overline{\mathbf{P}}\overline{\mathbf{P}}^T \mathbf{x} = (\mathbf{I} - \mathbf{PP}^T) \mathbf{x} = (\mathbf{I} - \mathbf{C}) \mathbf{x} \in S_r \quad (5)$$

因为 S_p 和 S_r 相互正交,则 $\hat{\mathbf{x}}^T \tilde{\mathbf{x}} = 0$,且 $\mathbf{x} = \hat{\mathbf{x}} + \tilde{\mathbf{x}}$ 。

1.1.2 故障检测

一般来说,基于PCA模型的故障检测方法采用 T^2 统计量和SPE统计量进行故障检测,分别定义为^[9]:

$$T^2 = \mathbf{t}^T \mathbf{A}^{-1} \mathbf{t} = \mathbf{x}^T \mathbf{P} \mathbf{A}^{-1} \mathbf{P}^T \mathbf{x} \quad (6)$$

$$\text{SPE} \equiv \|\tilde{\mathbf{x}}\|^2 = \|(\mathbf{I} - \mathbf{PP}^T) \mathbf{x}\|^2 \quad (7)$$

T^2 的控制限 T_α^2 利用 F 分布确定:

$$\frac{N(N-l)}{l(N^2-1)} T^2 \sim F_{l, N-l} \quad (8)$$

$$T_\alpha^2 \equiv \frac{l(N^2-1)}{N(N-l)} F_{l, N-l; \alpha} \quad (9)$$

式中, F 分布自由度为 l 和 $N-l$, l 为主成分个数, N 为采样次数, α 为显著性水平。SPE的控制限 δ_α^2 利用 χ^2 分布确定:

$$\delta_\alpha^2 = g \chi_{h; \alpha}^2 \quad (10)$$

式中, $g = \theta_2 / \theta_1$, $h = \theta_1^2 / \theta_2$, θ_i 表示为:

$$\theta_i = \sum_{j=l+1}^m \lambda_j^i, \quad i = 1, 2, 3 \quad (11)$$

如果任意统计指标的值超出各自的控制限,则认为过程出现了故障。

1.2 ICA模型

ICA模型用来提取变量、测量值或者信号间潜在

的独立元 (Independent Components, ICs), 这种方法最初被用来解决未知信息资源的分离问题, 包括恢复与未知矩阵进行线性组合后的独立信号 (如不同的声音、音乐或者噪音信号等)^[10]。

1.2.1 数据模型

ICA 模型旨在基于观测数据 \mathbf{X} 通过线性组合寻求独立、非高斯的潜在变量 \mathbf{S} 以及未知矩阵 \mathbf{A} , 数据测量模型如下^[5]:

$$\mathbf{X} = \mathbf{AS} + \mathbf{E} \quad (12)$$

$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbf{R}^{d \times n}$ 为观测数据矩阵, $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m] \in \mathbf{R}^{d \times m}$ 是未知的混合矩阵, $\mathbf{S} = [s_1, s_2, \dots, s_n] \in \mathbf{R}^{m \times n} (m \leq d)$ 是独立元矩阵, $\mathbf{E} \in \mathbf{R}^{d \times n}$ 是残差矩阵, n 为样本的数量。一般认为 \mathbf{X} 和 \mathbf{S} 同维, 即 $m = d$, 此时矩阵 \mathbf{E} 为零矩阵, 则 ICA 模型可转化为寻求一个混合矩阵 \mathbf{W} 以满足重构矩阵 $\hat{\mathbf{S}}$ 各元素间最大的独立性, $\hat{\mathbf{S}}$ 可表示为:

$$\hat{\mathbf{S}} = \mathbf{WX} \quad (13)$$

为方便数学表述, 定义独立元具有单位方差, 即 $\mathbf{E}\{\mathbf{ss}^T\} = \mathbf{I}_m$ 。通常求解 ICA 模型第一步是对观测数据 \mathbf{X} 进行白化处理, 使观测向量 \mathbf{x} 经线性变换表示为:

$$\mathbf{z} = \mathbf{Qx} \quad (14)$$

经过观测数据的白化处理, 使 \mathbf{z} 的各变量间线性无关且具有单位方差, 即 $\mathbf{E}\{\mathbf{zz}^T\} = \mathbf{I}_m$ 。则可得到

$$\mathbf{z} = \mathbf{Qx} = \mathbf{QAs} = \mathbf{Bs} \quad (15)$$

这里定义 $\mathbf{B} = \mathbf{QA}$, 根据式(16) 得出 \mathbf{B} 为正交矩阵。

$$\mathbf{E}\{\mathbf{zz}^T\} = \mathbf{BE}\{\mathbf{ss}^T\}\mathbf{B}^T = \mathbf{BB}^T = \mathbf{I}_m \quad (16)$$

由于正交矩阵的限制, ICA 模型问题得到简化, 这里不必再寻求满秩矩阵 \mathbf{A} , 转而寻求正交矩阵 \mathbf{B} 。根据公式(15), 重构向量 $\hat{\mathbf{s}}$ 可表示为:

$$\hat{\mathbf{s}} = \mathbf{B}^T \mathbf{z} = \mathbf{B}^T \mathbf{Qx} \quad (17)$$

根据式(13)和式(17), 矩阵 \mathbf{B} 与 \mathbf{W} 间的关系可表示为:

$$\mathbf{W} = \mathbf{B}^T \mathbf{Q} \quad (18)$$

根据 $\hat{\mathbf{s}}_i = (\mathbf{b}_i)^T \mathbf{z}$, 为确保第 i 个独立元素的最大非高斯性, \mathbf{B} 矩阵每列向量 \mathbf{b}_i 确定前均进行初始化处理。Hyvärinen 和 Oja 在文献[11]中根据中心极限定理提出非高斯性代表着独立性, 而峰度和负熵可以用来度量非高斯性。基于对负熵的近似估计, Hyvärinen 在文献[12]中提出对于矩阵 \mathbf{B} 的算法。计算出矩阵 \mathbf{B} 后, 结合式(17)和式(18)可分别得出重构的独立元矩阵 $\hat{\mathbf{S}}$ 和混合矩阵 \mathbf{W} 。

获取独立元后, 需选取其中一部分作为主元以提高 ICA 模型的鲁棒性并降低分析过程的复杂性。选

取主元之前, 需对 ICs 进行排序。文献[13]中根据混合矩阵 \mathbf{W} 中各行的 Euclidean 范数 (L_2) 大小, 将其对应的 ICs 进行排序, 该方法简单有效。确定 ICs 的排序问题之后, Lee 等^[5]采取绘图的方式来确定主元个数。混合矩阵 \mathbf{W} 中各行 Euclidean 范数 (L_2) 所占的比例如图2所示。横坐标代表 ICs 的序号, 纵坐标代表对应混合矩阵 \mathbf{W} 中每行 L_2 所占比例, 该图中的混合矩阵 \mathbf{W} 根据正常情况下的废水数据 (2.1 部分) 所得。由图2可知, 后3个独立元对应的 L_2 比例明显小于其他的独立元对应的 L_2 比例, 因此, 图2中 ICs 所建立的 ICA 模型包含5个主元。

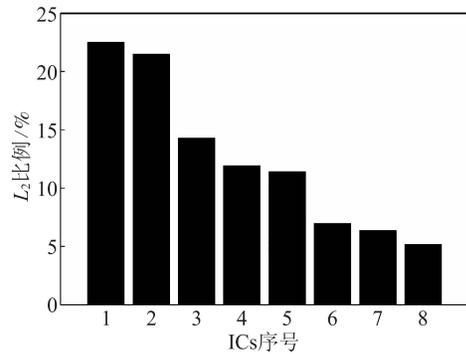


图2 混合矩阵 \mathbf{W} 中各行 Euclidean 范数 (L_2) 所占比例

1.2.2 故障检测

根据 1.2.1 确定主元个数 (设为 l) 后, 记 \mathbf{W}_d 为 \mathbf{W} 的主元部分, 其由 l 个具有最大 Euclidean 范数 (L_2) 的行向量构成, 混合矩阵 \mathbf{W} 中剩余部分为残差部分 \mathbf{W}_e , 根据式(13)可得, $\hat{\mathbf{s}}_d = \mathbf{W}_d \mathbf{x}$, $\hat{\mathbf{s}}_e = \mathbf{W}_e \mathbf{x}$, 据式(18)得: $\mathbf{B}_d = (\mathbf{W}_d \mathbf{Q}^{-1})^T$ ^[5]。

这里, ICA 模型的故障检测采用3种统计量作为监控指标:

$$I^2 = \hat{\mathbf{s}}_d^T \hat{\mathbf{s}}_d \quad (19)$$

$$I_e^2 = \hat{\mathbf{s}}_e^T \hat{\mathbf{s}}_e \quad (20)$$

$$\text{SPE} = \mathbf{e}^T \mathbf{e} = (\mathbf{x} - \hat{\mathbf{x}})^T (\mathbf{x} - \hat{\mathbf{x}}) \quad (21)$$

式中, $\hat{\mathbf{x}} = \mathbf{Q}^{-1} \mathbf{B}_d \hat{\mathbf{s}}_d$ 。以上3种统计量的控制限可通过核密度估计获取:

$$f(x) = \frac{1}{nh} \sum_{i=1}^n K\left\{\frac{x - x_i}{h}\right\} \quad (22)$$

式中, x_i 为数据集中的观测值, h 为平滑参数, K 为核函数, n 为样本数, h 的取值可参考文献[14]。当 x 点达95%的密度函数值时, x 为3种统计量的控制限。

2 仿真实验与讨论

2.1 制浆造纸废水处理过程数据

制浆造纸废水数据采取广东东莞某造纸厂的好氧

段废水监测数据, 该制浆造纸废水数据包括 8 个废水变量, 每个变量包括 170 个测量值。制浆造纸废水处理过程数据如图 3 所示, 左侧纵坐标分别对应进水化学需氧量 (COD_{inf})、出水化学需氧量 (COD_{eff})、进水悬浮固形物 (SS_{inf})、出水悬浮固形物 (SS_{eff}); 右侧纵坐标分别对应溶解氧量 (DO)、流量 (Q)、温度 (T)、pH 值^[15]。170 个测量数据中, 将前 100 个样本作为训练集, 后 70 个样本作为测试集。

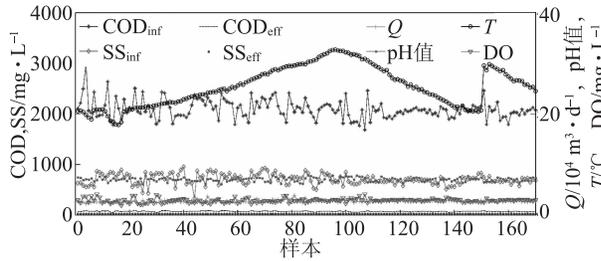


图 3 制浆造纸废水处理过程数据

2.2 故障数据构建

针对制浆造纸废水处理所面临的复杂工作环境并结合传感器在数据监测中起到的重要作用, 本课题考虑构建以下 3 种传感器故障: 偏移故障、漂移故障和完全失效故障^[16]。3 种故障数据的构建如表 1 所示。在表 1 中, 对变量 COD_{inf} 加入均值的 20% 作为偏移故障, 对变量 pH 值加入以 0.05 为系数的漂移故障, 将 SS_{eff} 的数据改至 30, 作为完全失效故障。如图 4 所示, 在测试集后 50 组数据中加入故障信息。

表 1 3 种故障数据的构建

故障类型	偏移	漂移	完全失效
故障构建公式	$COD_{inf}(t) + 418$	$pH(t) + 0.05t$	$SS_{eff}(t) = 30$

2.3 故障检测

PCA 和 ICA 两种方法模型对 3 种故障类型的检测结果如表 2 和表 3 及图 5 ~ 图 7 所示。所有图中的虚线代表 95% 的置信度控制限, 超出控制限的统计量代表出现了故障, 根据测试集后 50 组数据中超出

样本控制限的检测点个数所占的比例计算故障检测率, 对于每种故障对应的多个统计量, 选择故障检测率最高的统计量对应结果作为该故障的检测率。

由表 2 和表 3 及图 5 ~ 图 7 可知, 对于偏移故障, PCA 模型的 SPE 统计量与 ICA 模型的 I^2 统计量从测试集样本点 21 处开始均有明显的上升, 但是两者的检测效果不佳; PCA 模型的偏移故障检测率为 14%, ICA 模型的偏移故障检测率为 24%。对于漂移故障, PCA 模型的 T^2 统计量与 ICA 模型的 I_e^2 、SPE 统计量均表现出明显的线性增长趋势; 相对而言, ICA 模型的漂移故障检测率达到 54%, 检测效果优于 PCA 模型。对于完全失效故障, PCA 模型与 ICA 模型的各个统计量在故障出现后的第 1 个样本开始均有明显变化, 且 PCA 的 SPE 统计量和 ICA 的 I^2 和 SPE 统计量对应的完全失效故障检测率均达到 100%。

为了优化故障检测的效率, 结合变量贡献图与统计量的值进行故障检测可高效地检测出故障所属的变量类型。以 ICA 模型对于图 7(c) 所检测出的完全失效故障点为例, 对位于测试集第 70 个样本点进行故障的变量贡献值检测, 图 8(a) ~ 图 8(c) 分别显示了

表 2 PCA 模型故障检测率 %

故障	PCA	
	SPE	T^2
偏移	14	0
漂移	0	42
完全失效	100	16

表 3 ICA 模型故障检测率 %

故障	ICA		
	I^2	I_e^2	SPE
偏移	24	2	2
漂移	0	54	52
完全失效	100	88	100

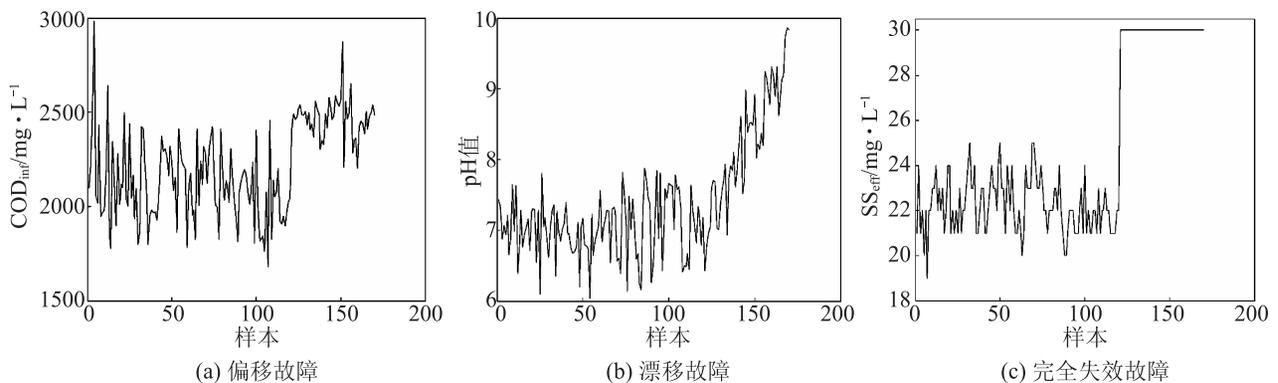


图 4 制浆造纸废水处理过程的故障构建

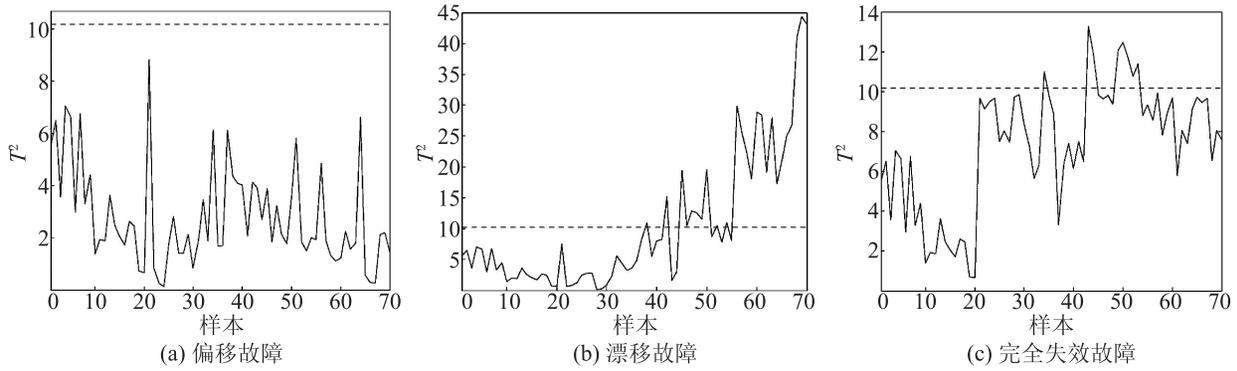


图5 基于 T^2 的 PCA 模型检测结果

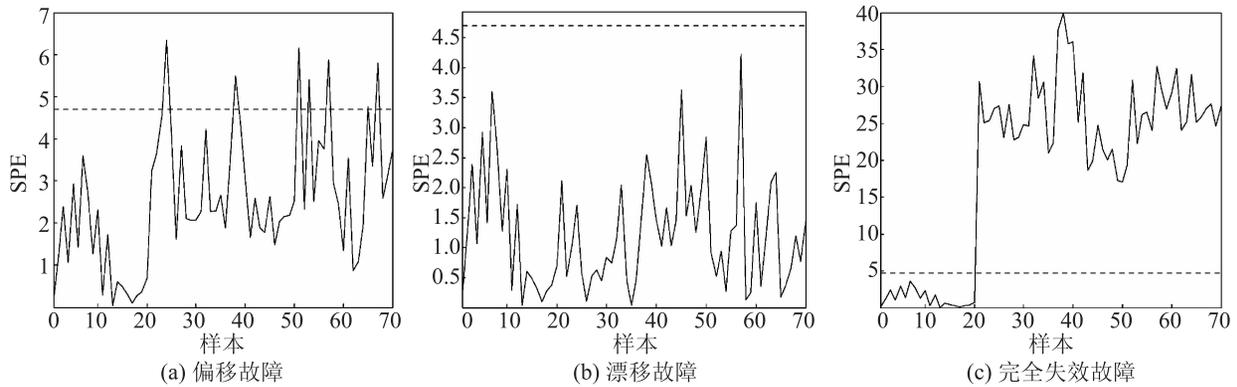


图6 基于 SPE 的 PCA 模型检测结果

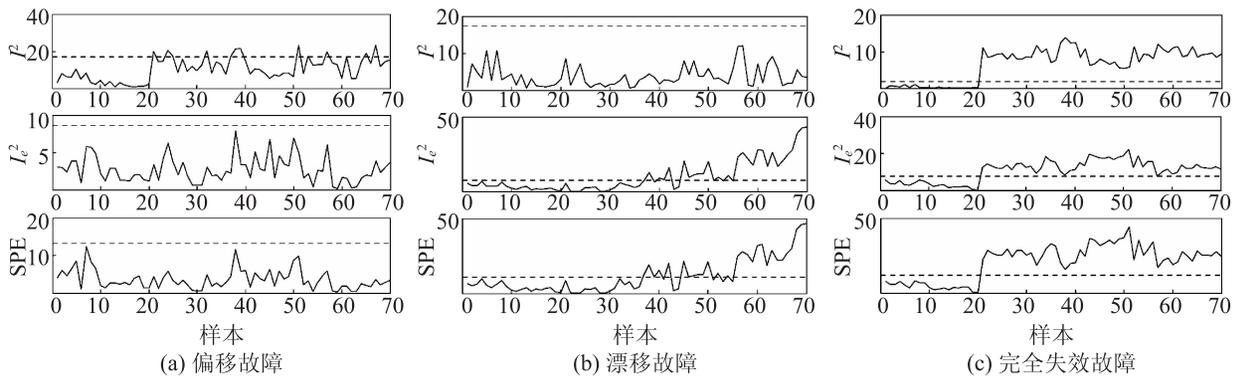


图7 ICA 模型检测结果

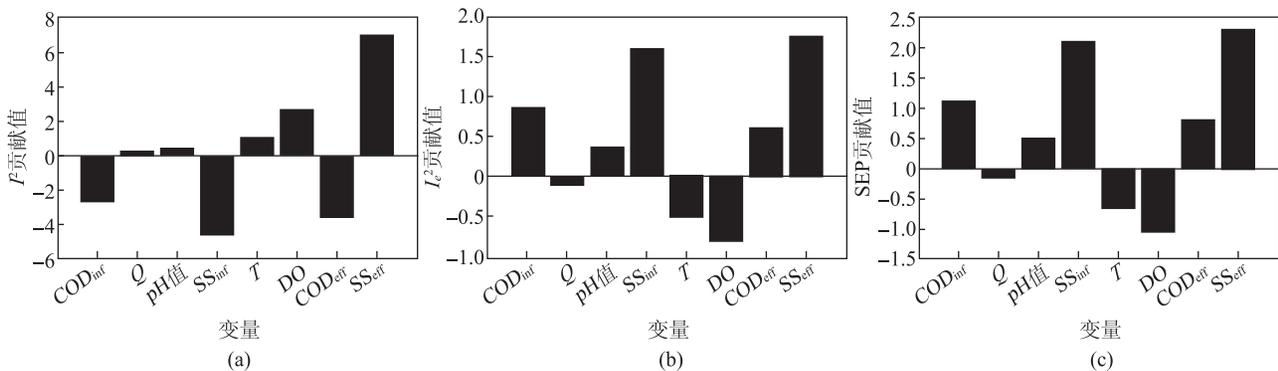


图8 统计量 F^2 (a)、 I_e^2 (b) 和 SPE (c) 对测试集第70个样本点完全失效故障的变量贡献图

统计量 I^2 、 I_e^2 和 SPE 对于故障点的变量贡献值, 结果显示, 变量 SS_{inf} 和变量 SS_{eff} 对完全失效故障的贡献值最高, 变量 COD_{inf} 、变量 COD_{eff} 以及变量 DO 的贡献值次之, 其余变量贡献值均很低。由此, 可判断变量 SS_{inf} 和变量 SS_{eff} 与完全失效故障有较强的相关性, 根据以上信息可有效缩小故障排查范围, 提高故障诊断效率。

3 结论

本课题分别采用主成分分析 (PCA) 模型与独立元分析 (ICA) 模型对制浆造纸废水处理过程进行故障诊断。考虑到 PCA 模型控制限的设置对于测量数据高斯性的要求, 结合制浆造纸废水处理过程中数据呈非高斯分布的特点, PCA 模型的限制性相比于 ICA 模型较为明显。相比于 PCA 模型采用的二阶统计信息, ICA 模型的优势在于采用 I^2 、 I_e^2 、SPE 3 种统计量进行故障检测。其中统计量 I^2 替代了 PCA 模型中 T^2 统计量, ICA 模型在确定主元个数时所排除掉的独立元矩阵可用于构建统计量 I_e^2 , 以此来弥补主元个数选取不当引起的故障检测能力下降的缺点。此外, ICA 模型在故障诊断中要求潜变量之间统计独立, 没有对潜变量正交性与高斯性的限制。

制浆造纸废水处理过程的故障检测结果表明, ICA 模型的整体故障检测率要高于 PCA 模型, 但优势并不明显。制浆造纸废水处理过程有着较强时变性, 随着时间的改变, 变量的均值与方差会随之改变。而常规的多元统计过程监测方法 (如 PCA 和 ICA 方法) 属于静态的统计监测方法, 因此在废水数据故障检测中可能造成较大的误报与漏报现象, 造成故障检测效果不佳。鉴于此, 未来考虑将动态 ICA 方法应用到制浆造纸废水处理过程监测领域。

参 考 文 献

- [1] Ge Z, Song Z, Gao F. Review of recent research on data-based process monitoring [J]. *Industrial & Engineering Chemistry Research*, 2013, 52(10): 3543.
- [2] Liu T L, Shen W H. A review of application of fault diagnostic expert system in wastewater treatment [J]. *Paper Science & Technology*, 2011, 30(2): 75.
刘天龙, 沈文浩. 污水处理过程中故障诊断专家系统的应用 [J]. *造纸科学与技术*, 2011, 30(2): 75.
- [3] Huang D P, Qiu Y, Liu Y Q, et al. Review of data-driven fault diagnosis and prognosis for wastewater treatment [J]. *Journal of South China University of Technology (Natural Science Edition)*, 2015, 43(3): 111.
黄道平, 邱禹, 刘乙奇, 等. 面向污水处理的数据驱动故障诊断及预测方法综述 [J]. *华南理工大学学报(自然科学版)*, 2015, 43(3): 111.
- [4] Zhou D H, Hu Y Y. Fault diagnosis techniques for dynamic systems [J]. *Acta Automatic Sinica*, 2009, 35(6): 748.
周东华, 胡艳艳. 动态系统的故障诊断技术 [J]. *自动化学报*, 2009, 35(6): 748.
- [5] Lee J M, Yoo C K, Lee I B. Statistical process monitoring with independent component analysis [J]. *Journal of Process Control*, 2004, 14(5): 467.
- [6] Xie Z H. *Matlab statistical analysis and application of 40 case studies* [M]. Beijing: Beihang University Press, 2010.
谢中华. *MATLAB 统计分析与应用: 40 个案例分析* [M]. 北京: 北京航空航天大学出版社, 2010.
- [7] Qin S J. Survey on data-driven industrial process monitoring and diagnosis [J]. *Annual Reviews in Control*, 2012, 36(2): 220.
- [8] Shen Y, Ding S X, Haghani A, et al. A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process [J]. *Journal of Process Control*, 2012, 22(9): 1567.
- [9] Qin S J. Statistical process monitoring: basics and beyond [J]. *Journal of Chemometrics*, 2003, 17(8/9): 480.
- [10] Olsson G. ICA and me—A subjective review [J]. *Water Research*, 2012, 46(6): 1585.
- [11] Hyvärinen A, Oja E. Independent component analysis: algorithms and applications [J]. *Neural Networks*, 2000, 13(5): 411.
- [12] Hyvärinen A. Fast and robust fixed-point algorithms for independent component analysis [J]. *IEEE Transactions on Neural Networks*, 1999, 10(3): 626.
- [13] Cardoso F, Souloumiac A. Blind beamforming for non Gaussian signals [J]. *Radar & Signal Processing IEE Proceedings F*, 1993, 140(6): 362.
- [14] Silverman B W. *Density estimation for statistics and data analysis* [M]. London: Chapman & Hall, 1986.
- [15] Liu Y Y, Li X Y, Zhang G R, et al. Sensor fault detection of papermaking wastewater treatment processes based on multivariate statistical analysis [J]. *China Pulp & Paper Industry*, 2017, 38(8): 41.
刘耀瑶, 李祥宇, 张光锐, 等. 基于多元统计分析的造纸废水处理过程传感器故障检测 [J]. *中华纸业*, 2017, 38(8): 41.
- [16] Dunia R, Qin S J, Edgar T F, et al. Identification of faulty sensors using principal component analysis [J]. *AIChE Journal*, 1996, 42(10): 2797.

Fault Detection of Papermaking Wastewater Treatment Process Based on Independent Component Analysis

YANG Chong¹ SONG Liu¹ LIU Hongbin^{1,2,*}

(1. *Co-Innovation Center of Efficient Processing and Utilization of Forest Resources, Nanjing Forestry University,*

Nanjing, Jiangsu Province, 210037; 2. State Key Laboratory of Pulp and Paper Engineering,

South China University of Technology, Guangzhou, Guangdong Province, 510640)

(* E-mail: hongbinliu@njfu.edu.cn)

Abstract: To monitor and control papermaking wastewater treatment process (WWTP) effectively, two common methods of multivariate statistical analysis named independent component analysis (ICA) and principal component analysis (PCA) were used to detect the sensor faults in a papermaking WWTP. The results showed that the detection rates of the bias and drifting faults using ICA were 24% and 54%, respectively. Meanwhile, the bias and drifting faults detection rates of PCA were 14% and 42%. The fault detection rates of ICA were higher than those of PCA, but neither of the two methods achieved satisfactory result of detecting the bias and drifting faults. Concerning the complete failure fault, both the fault detection rates of the two methods were 100%.

Keywords: papermaking wastewater treatment process; fault detection; principal component analysis; independent component analysis

(责任编辑:杨 艳)