

文章编号: 1001-0920(2006)10-1190-03

基于独立源分析的过程监测方法

杨英华, 吴英华, 陈晓波, 秦树凯
(东北大学 信息科学与工程学院, 沈阳 110004)

摘 要: 多元统计过程控制要求观测数据服从正态分布, 而实际的工业过程数据大都不满足正态分布条件。独立源分析(ICA)是近几年才发展起来的一种新的统计方法, 可以克服对数据分布的依赖性。对此, 以 ICA 算法为核心, 引入一种新型的过程监测方法, 应用 ICA 提取独立源, 利用 I^2 图, I_c^2 图和 SPE 图进行故障检测。最后以 3 水箱系统为例进行了实验研究, 取得了很好的效果。

关键词: 独立源分析(ICA); 过程监测; 故障检测

中图分类号: TP277 **文献标识码:** A

Process Monitoring Based on Independent Component Analysis Method

YANG Ying-hua, WU Ying-hua, CHEN Xiao-bo, QIN Shu-kai

(College of Information Science and Engineering, Northeastern University, Shenyang 110004, China. Correspondent: YANG Ying-hua, E-mail: yhyang@mail.neu.edu.cn)

Abstract: Multivariate statistical process control(MSPC) is based upon the assumption that the observed data must be subject to normal probability distribution, which sometimes can not be satisfied. Independent component analysis(ICA) is a recently developed method, which can overcome the need of the data distribution. A new method based on ICA is introduced to process monitoring. This method uses ICA to extract the independent components, and adopts I^2 , I_c^2 and SPE charts for fault detection. The simulation results of three water tank system show the effectiveness of this method.

Key words: Independent component analysis(ICA); Process monitoring; Fault detection

1 引 言

近年来, 基于主元分析方法(PCA)的过程性能监控和故障诊断技术受到了学术界和工业界的广泛重视, 并在化工生产过程中得到了成功的应用。人们采用 PCA 从过程观测数据中提取统计无关主元, 通过构造各种信息统计量对过程运行状况进行统计分析, 判断过程运行是否偏离了正常的操作区域, 并诊断引起状态偏移的原因, 其结论成立的前提是要求观测数据服从正态分布。然而, 实际的工业过程数据大都不满足正态分布条件, 传统的 PCA 必然导致过程性能分析不准和过程故障的误报、漏报^[1]。近几年

发展起来的独立源分析(ICA)方法能很好地解决这个问题。

ICA 与 PCA 最大的不同在于: PCA 仅利用了原始特征中的二阶信息, 也就是包含在协方差矩阵中的信息; 而 ICA 则利用了高阶统计信息, 正是高阶统计信息揭示了随机特征的统计分布。显然, 在随机特征为经典正态 2 高斯分布的情况下, PCA 比较合适, 但当不满足高斯分布的假设时, 可采用 ICA 方法。ICA 可以看作是 PCA 在非高斯情况下的推广, 它考虑的是原始特征的概率分布信息。其次, 我们发现 ICA 的子空间投影矢量不是相互正交的, 而 PCA 的投影矢量都是相互正交的, 如图 1 所示, 这

收稿日期: 2005-08-22; 修回日期: 2006-03-07

基金项目: 国家自然科学基金项目(60374003)

作者简介: 杨英华(1970—), 男, 辽宁辽阳人, 副教授, 博士, 从事复杂工业过程建模、过程监测及故障诊断的研究;
吴英华(1980—), 男, 长春人, 硕士, 从事复杂工业过程建模、过程监测及故障诊断的研究

说明独立特征提取的投影基将具有更大的通用性和适应能力

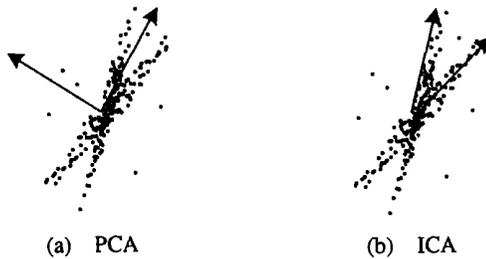


图 1 非正态高斯分布数据分析结果

2 独立源分析技术简介

ICA^[2]是近几年才发展起来的一种新的统计方法,其目的是将观察到的数据进行某种线性分解,使其分解成统计独立的成分。ICA 是从线性混合信号里恢复出一些基本的源信号的方法

为了给 ICA 下一个严格的定义^[2],这里需要使用一个隐藏的统计变量模型

$$X = AS, \quad (1)$$

其中: 矩阵 $X = [x_1, x_2, \dots, x_m]^T$ 为包含 m 个变量的测量值, 矩阵 $S = [s_1, s_2, \dots, s_n]^T$ 为 n 个独立的非高斯源信号 X 的每一行是第 i 个测量值的有限时间序列; S 的每一行是第 j 个源信号的有限时间序列, 通常被称为独立源(IC)。 A 为混合矩阵

ICA 的目标是: 在只有测量值已知的情况下, 依据一定的目标函数和约束, 优化学习出 \hat{A} 和 \hat{S} 。该目标等价于找到一个分解矩阵 W , 满足 $\hat{S} = W X$, 其中 \hat{S} 是独立源 S 的估计值, W 是分解向量

通常在 ICA 中, 首先要对信号进行中心化和白化, 这样可以使 ICA 算法更简单更稳定。中心化即去均值, 使信号均值为 0; 白化是为了去除变量间的交叉相关, 设白化后的模型为

$$z(k) = B s(k), \quad (2)$$

其中 B 为正交矩阵。至此通过以上变换, 将问题大大简化了。要估计 B (或 W), 其中一种方法是: 先通过最大化 $s_i = (b_i)^T z$ 的非高斯指标来找到第一个分解向量 $(b_i)^T$, 然后按相同最大化方法依次估计分解向量 $(b_i)^T$, 并对分解向量进行互为正交约束。非高斯程度可以用统计峰度值或负熵等指标来衡量。本文使用 FastICA 方法求取独立源^[3]。

3 独立源分析的过程监测

3.1 独立源的降维及排序

在 ICA 所有的独立源中选取一些占主要地位的独立源在 ICA 监测中是很重要的, 占主要地位的独立源反映了产生所观察序列的随机系统的大量信息, 基于这些独立源所建立的模型在 ICA 监测中将

具有良好的鲁棒性, 同时也降低了分析的复杂性^[4]。为了选取这些独立源首先要对独立源进行排序

在 ICA 中, 独立源的排序方法在一些文献中已有介绍^[5,6]。本文将用 Euclidean 范数(L_2) 对分解矩阵 W 的行进行排序(W 的行对应 ICs), 因为这个方法既简单, 在 ICA 监测中也有很好的效果^[5]。

对 ICs 排完后, 为了得到较好的监测结果, 应选取合适的 ICs 数。ICs 数太多容易加大噪声并减弱过程的监测性能。假设 W 行的平方和越大对 S 的影响也越大, 于是数据维数的降低便可通过选取 W 的一些行来实现。这个方法是基于过程变量可以由基于主导地位的几个 ICs 来监控这一思想实现的^[6]。可以根据 W 每行 L_2 范数的百分比来选取 ICs 数

3.2 监测模型的建立

用 ICA 对生产过程进行监测, 首先要获得过程正常状态下的变量采样数据阵 X_{normal} 。假设变量数等于独立源数, 利用 FastICA 运算法则计算 W 和 S_{normal} ($S_{\text{normal}} = W X_{\text{normal}}$), B 以及 A 。通过对数据的白化及 FastICA 运算法则也可以得到。根据前文提到的 W 行的平方和越大对 S 的影响也越大, 于是对数据进行降维, 选取 W 的 a 行建立矩阵 W_d (主导部分), 剩下的行建立矩阵 W_e 。对应 W 行的选取, 同样可以建立 B_d, B_e 。对于生产过程中新采集到的一组数据 $x_{\text{new}}(k)$, 通过分解矩阵 W_d 和 W_e 可以得到独立源的估计值

$$\hat{s}_{\text{new}d}(k) = W_d x_{\text{new}}(k), \quad \hat{s}_{\text{new}e}(k) = W_e x_{\text{new}}(k).$$

建立统计量 $I^2(k)$, $\text{SPE}(k)$, $I_e^2(k)$ 分别为

$$I^2(k) = \hat{s}_{\text{new}d}(k)^T \hat{s}_{\text{new}d}(k), \quad (3)$$

$$\text{SPE}(k) = e(k)^T e(k) = (x(k) - \hat{x}(k))^T (x(k) - \hat{x}(k)), \quad (4)$$

$$I_e^2(k) = \hat{s}_{\text{new}e}(k)^T \hat{s}_{\text{new}e}(k). \quad (5)$$

I^2 统计量表征了模型内部变化的测度, SPE 表征了模型外部变化的测度, 而利用 $\hat{s}_{\text{new}e}(k)$ 建立的 I_e^2 统计量提供了另一种监测的工具, I_e^2 往往较其他两个更加有效, 因为 I_e^2 能够补偿由于占主导地位的 ICs 数目选取不合适所造成的错误

在 ICA 监测中, 经过独立源分析后所得到的信号通常不服从正态分布。此外, 过程中往往存在着许多独立的非高斯信源。因此, 在这种情况下, 统计量控制限就不能通过一个特定的近似分布来确定, 而需采用其他的分类方法确定控制限, 本文采用核密度估计方法来确定 3 个统计量的控制限^[7]。首先通过正常数据建立统计量的样本, 然后按照核密度估计来估计统计量的密度函数, 再按 99% 的置信区间确定控制限, 画出监测图

至此, 我们得到了完整的故障监测过程

4 实验结果及研究

4.1 3 水箱系统结构简介

3 水箱实验系统总体结构如图 2 所示。水箱主体由 3 个圆柱形容器、一个回收水槽及多个阀门和连接件构成。3 水箱系统是一个强非线性、高耦合、双入双出对象，为研究非线性过程监测提供了一个很好的模型。

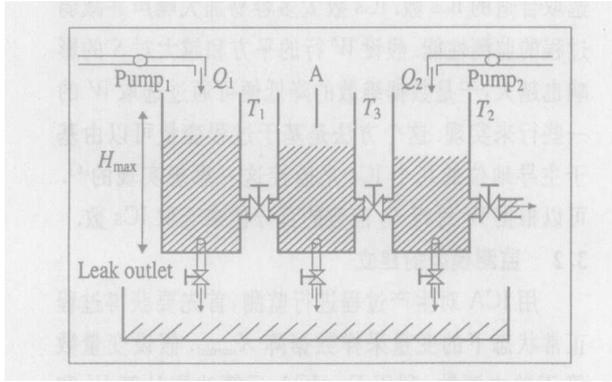


图 2 3 水箱实验装置系统结构

4.2 实验结果

选取前一时间刻 3 水箱液位以及水泵的流量作为过程变量，当前时刻水箱 T_1 及水箱 T_2 的液位作为质量变量。在该水箱无漏水、传感器工作正常时，3 水箱从初始值（液位均为 0）至到达并维持设定值之间的过程采集 400 个样本数据作为正常数据 $X(Q_1, Q_2, h_1, h_2, h_3)$ 。过程变量 X 即两个控制水泵的流量及 3 个水箱的液位值。

用 FastICA 进行独立源分析，求出 W ，并按照 W 的每行 L_2 范数对独立源进行排序，如图 3 所示。选取两个独立源作为占主要地位的独立源，按照前文方法进行分析。对正常数据求取 $I^2(k)$, $SPE(k)$, $I_c^2(k)$ ，进行核密度分析，按照 99% 置信区间确定控制限。

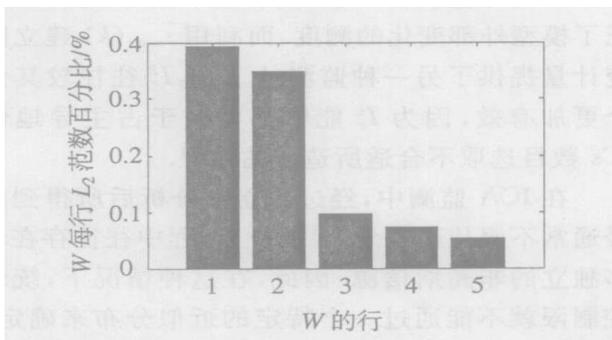


图 3 分解矩阵 (W) 每行 L_2 范数百分比图 (排序后)

模拟故障，首先在水箱液位 h_1 和 h_2 平稳状态下，采样时间 70 时刻打开水箱 T_1 的泄漏阀，模拟水箱 T_1 泄漏故障。在线监测结果如图 4 所示。

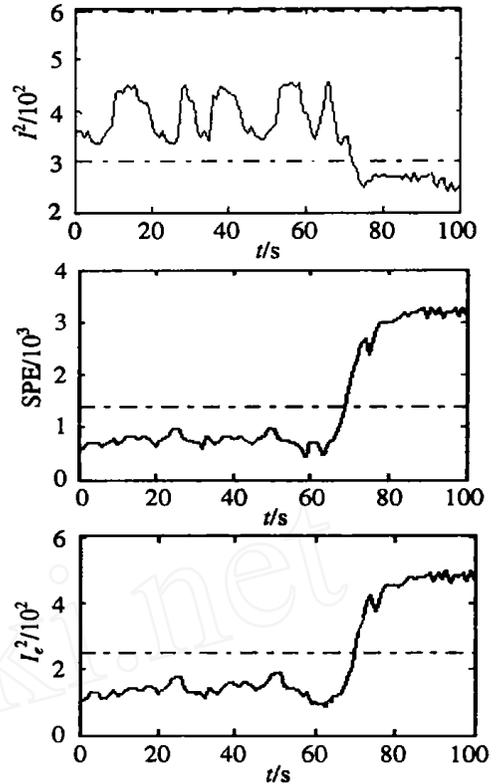


图 4 水箱 T_1 泄漏时 I^2 , SPE , I_c^2 监测图

同理，模拟水箱 T_2 泄漏故障，在采样时刻 80 打开水箱 T_2 的泄漏阀，在线监测结果如图 5 所示。很明显，使用本文的监测方法取得了很好的监测效果。

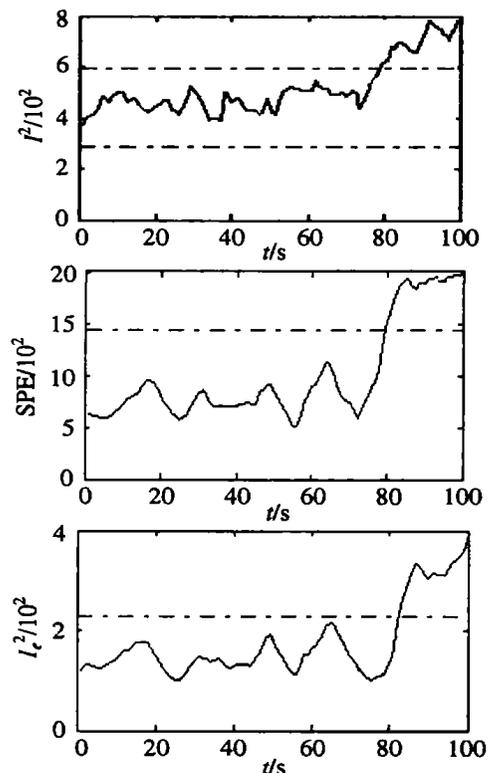


图 5 水箱 T_2 泄漏时 I^2 , SPE , I_c^2 监测图

(下转第 1196 页)

列的目标值 从表中数据对比可以看出,本文提出的IPSO 算法明显优于文献[7]和文献[8],与文献[9]相当,说明IPSO 算法对于实际优化问题是有效的

表3 反应动力学参数优化结果对比

方法	a	b	c	A	E_a	SSE
文献[7]	0.88	0.41	0.34	100	46.200	0.2494
文献[8]	0.8181	0.4750	0.3276	70.4	45.1532	0.2225
文献[9]	0.8067	0.4425	0.3336	71.06	46.502	0.2177
IPSO	0.8081	0.4444	0.3239	63.5465	45.626	0.217685

5 结 论

本文提出的基于模式优选思想改进的粒子群优化算法能有效地搜索到全局最优解,通过函数测试表明,该算法的优化效率和优化性能较PSO 有显著的提高 应用改进的粒子群优化算法进行反应动力学参数的估计,得到的优化结果较文献报道的结果更好

参考文献(References)

- [1] Kennedy J, Eberhart R. Particle Swarm Optimization [A]. *IEEE Int Conf on Neural Networks* [C]. Perth, 1995: 1942-1948
- [2] Eberhart R, Kennedy J. A New Optimizer Using Particle Swarm Theory [A]. *Proc of the 6th Int Symposium on Micro Machine and Human Science* [C]. Vagoya, 1995: 39-43
- [3] Shi Y, Eberhart R. A Modified Particle Swarm Optimization [A]. *IEEE World Congress on Computational*

- Intelligence* [C]. Anchorage, 1998: 69-73
- [4] Clerc M. The Swarm and the Queen: Towards a Deterministic and Adaptive Particle Swarm Optimization [A]. *Proc of the Congress on Evolutionary Computation* [C]. Washington DC, 1999: 1951-1957
- [5] Shi Y, Eberhart R. C. Fuzzy Adaptive Particle Swarm Optimization [A]. *Proc of the Congress on Evolutionary Computation* [C]. Seoul, 2001: 101-106
- [6] 陈国良,王煦法,庄镇泉. *遗传算法及其应用* [M]. 北京:人民邮电出版社,1996
(Chen G L, Wang X F, Zhuang Z Q. *Genetic Algorithm and Its Application* [M]. Beijing: People Posts and Telecom Press, 1996)
- [7] Li R, Savage P E, Szmukler D. 2-chlorophenol Oxidation in Supercritical Water: Global Kinetics and Reaction Products [J]. *AIChE J*, 1993, 39(1): 178-187
- [8] 颜学峰,陈德钊,胡上序,等. 混沌遗传算法估计反应动力学参数 [J]. *化工学报*, 2002, 53(8): 810-814
(Yan X F, Chen D Z, Hu S X, et al. Estimation of Kinetic Parameters Using Chaos Genetic Algorithms [J]. *J Chem Ind and Eng*, 2002, 53(8): 810-814)
- [9] 张兵,陈德钊,饶骏. 优选策略支持的进化规划估计反应动力学参数 [J]. *高校化学工程学报*, 2004, 18(5): 638-642
(Zhang B, Chen D Z, Rao J. Estimation of Kinetic Parameters by Using Eugenic Evolution Programming [J]. *J of Chemical Engineering of Chinese University*, 2004, 18(5): 638-642)

(上接第1192页)

5 结 语

本文引入一种基于独立源分析(ICA)的新型统计方法应用于过程监测,采用了 I^2 图, I_e^2 图和SPE图作为故障监测的在线监测图 最后以3水箱系统监测为例,对此方法进行了应用研究,取得了很好的应用效果

参考文献(References)

- [1] Ypma A, Tax D, Duin R. Robust Machine Fault Detection with Independent Component Analysis and Support Vector Data Description [A]. *Proc of the 1999 IEEE Signal Processing Society Workshop* [C]. Madison, 1999: 67-76
- [2] Comon P. Independent Component Analysis — A New Concept? [J]. *Signal Processing*, 1994, 36(3): 287-314

- [3] Hyvärinen A, Oja E. Independent Component Analysis: Algorithms and Applications [J]. *Neural Networks*, 2002, 13(4): 411-430
- [4] Li R F, Wang X Z. Dimension Reduction of Process Dynamic Trends Using Independent Component Analysis [J]. *Computers and Chemical Engineering*, 2002, 26(3): 467-473
- [5] Cardoso J F, Soulomica A. Blind Beamforming for Non-Gaussian Signals [J]. *IEEE Proc F*, 1993, 140(6): 362-370
- [6] Cheung Y M, Xu L. Independent Component Ordering in ICA Time Series Analysis [J]. *Neuro Computing*, 2001, 41(1): 145-152
- [7] Wand M P, Jones M C. *Kernel Smoothing* [M]. London: Chapman and Hall, 1995