

文章编号: 1001-0920(2010)06-0801-07

基于数据和知识的工业过程监视及故障诊断综述

刘 强¹, 柴天佑¹, 秦泗钊², 赵立杰^{1a,3}

(1. 东北大学 a. 流程工业综合自动化教育部重点实验室, b. 自动化研究中心, 沈阳 110819;
2. 南加州大学 化工系, 洛杉矶 90089; 3. 沈阳化工学院 信息工程学院, 沈阳 110032)

摘 要: 从复杂工业过程所可能具有的过程特性及数据存取过程中引入的数据特性分析出发, 综述了具有复杂数据特性的工业过程的多元统计监视方法, 并分别讨论了基于数据和基于知识方法进行故障诊断的优势、进展、适用范围及二者结合的可能. 最后探讨了这一领域中值得进一步研究的问题和可能的发展方向.

关键词: 多元统计过程监视; 基于数据的诊断; 基于知识的诊断; 工业过程

中图分类号: TP273

文献标识码: A

Progress of data-driven and knowledge-driven process monitoring and fault diagnosis for industry process

LIU Qiang¹, CHAI Tian-you¹, QIN S Joe², ZHAO Li-jie^{1a,3}

(1a. Key Laboratory of Process Industry Automation of Ministry of Education, 1b. Research Center of Automation, Northeastern University, Shenyang 110819, China; 2. The Mork Family Department of Chemical Engineering and Materials Science, University of Southern California, Los Angeles 90089, USA; 3. Information Engineering School, Shenyang Institute of Chemical Technology, Shenyang 110032, China; Correspondent: LIU Qiang, E-mail: qiang.liu1980@163.com)

Abstract: Based on the analysis of complex data characteristics due to the process characteristics or the data collection and storage problem, the developments of theory the researches on complex industry process multivariate statistical monitoring are reviewed. The advantages, development, applicable domain of the data-based and knowledge-based diagnosis methods are discussed. And the possibility of these two types of methods' combination are studied. Finally, some problems and their research tendencies in this field are presented.

Key words: Multivariate statistical process monitoring; Data-based fault diagnosis; Knowledge-based fault diagnosis; Industry process

1 引 言

复杂工业过程一旦发生事故, 可能对生产安全、效率或产品质量造成不良影响^[1], 这就对生产过程的监视及诊断提出了更高要求. 由于实际工业过程不确定因素多且过程复杂, 一般可包括非线性、时变、变量耦合、时间相关性、多模态、多时段、大规模、间歇等特性, 使其难以建立精确的过程模型及故障模型. 基于数据和知识的方法因可避免解析法要求详细、准确的定量数学模型的缺点, 已成为研究热点并取得了进展, 广泛用于工业过程的监视及诊断. 如田纳西伊斯曼过程^[2], 半导体制造^[3,4], 钢铁行业转炉、

连铸、轧制^[5,6]等.

目前, 过程工业主要采用如下方法解决监控、诊断、容错三方面问题: 1) 监控方面采用单变量过程控制 (SPC) 或多元统计过程监控 (MSPM) 方法; 2) 诊断方面包括: ① 定性的专家系统及动态趋势, ② 定量的多元统计方法、神经网络方法及聚类方法; 3) 容错是在诊断出故障的前提下, 基于症状、控制树、故障概率、过程历史操作等信息, 通过硬件冗余、控制器重新调整^[7,8]、传感器重构^[9]等方式实现.

虽然目前针对故障诊断问题已经建立了相对完整的理论体系, 但在过程工业的实际应用中仍有以下

收稿日期: 2009-05-06; 修回日期: 2009-08-31.

基金项目: 国家重点基础研究发展计划项目(2009CB320600); 高等学校学科创新引智计划项目(B08015); 教育部科学技术研究重大项目(308007); 国家 863 计划项目(2006AA040307).

作者简介: 刘强(1980—), 男, 辽宁法库人, 博士生, 从事复杂工业过程故障诊断、智能建模与控制的研究; 柴天佑(1947—), 男, 甘肃天水人, 中国工程院院士, 教授, 博士生导师, 从事自适应控制、智能解耦控制等研究.

问题有待研究: 1) 在工业过程的复杂过程特性及数据采集、传输、存储过程中引入复杂数据特性的前提下, 提出相应的改进过程监视方法以使多元统计方法得以应用且监视结果准确; 2) 由于进行完整故障诊断的工业需求, 多元统计方法的研究扩展到故障的辨识及诊断; 3) 重点进行过程故障(尤其是回路设定不当引起的)及尚未发生过故障的诊断研究。

2 多元统计监控

2.1 基本方法

主元分析(PCA)和部分最小二乘(PLS)是两种普遍应用的多元统计方法, 而独立元分析(ICA)因适于过程信息不满足正态分布的情况, 近年来受到了普遍关注。

PCA方法是将高维过程数据投影到正交的低维子空间, 并保留主要过程信息。在过程监视方面, PCA将观测空间划分成一个获取过程趋势的主元子空间和基本包含随机噪声的残差子空间, 通过两个空间的统计指标或综合指标^[10]是否超限来监视过程。与PCA的去相关不同, ICA将观测数据分解为统计上独立元的线性组合, 可提取测量值不服从高斯分布的部分, 更适于过程数据不满足正态分布的情况; 在假设潜在变量非高斯分布的前提下可找到非高斯元(ICs); 通常用 I^2 , I_e^2 和SPE图进行在线监控和故障识别, 并考虑了潜在变量的非高斯性。以上监视指标的控制限需通过Lee^[11]的核密度方法估计。

PCA和ICA方法都是典型的非因果关系方法; 而Wold提出的PLS则为因果关系方法, 通常用于监控过程操作引起的质量指标异常。该方法的出发点在于生产过程中不仅包括过程数据 \mathbf{X} , 而且包括与 \mathbf{X} 具有因果关系的产品质量数据 \mathbf{Y} 。PLS就是通过输入矩阵 \mathbf{X} 与输出矩阵 \mathbf{Y} 之间相互交换分解信息, 找到输入空间到输出空间的预测能力意义上的最优特征方向。

2.2 结合复杂数据特性的多元统计方法

实际工业过程一般包括: 动态、缓变、非线性、大规模、多尺度、多模态等过程特性, 而且, 从间断与连续生产的角度, 还可分为间歇过程和连续过程。因而以上过程特性对于采集到的数据会带来数据自相关、数据均值方差漂移、共线性、变量间非线性关系等数据特点, 同时过程操作还会引起数据的非正态; 另一方面, DCS进行数据采集存储过程中, 可能引起数据缺失点、离群点、不对齐等, 这些数据特性都使得基本的多元统计监视方法难以直接使用或结果难以解释(或不可靠), 易出现漏报或误报等问题。为此, 很多学者提出了改进的多元统计方法, 使之与过

程特性相适应。

2.2.1 动态及缓变特性

基本PCA方法适合监视稳态过程, 然而实际过程一般因储能环节的存在而具有动态特性, 因此若以较高采样速率采集, 数据就会具有明显的动态特性, 此时, 以“样本观测相互独立, 且潜在变量服从高斯分布”为假设前提条件的基本PCA方法就不再适用该类过程监视问题。为此, Kaurti等^[12]提出了将变量的延迟加入数据阵, 以构造时间滞后数据阵, 再对PCA加以改进, 从而将时间相关的线性关系提取出来。为使动态方法适于变量非高斯的情况, Lee等^[13]提出了动态ICA(DICA)方法, 通过ICA提取测量值不服从高斯分布的部分, 并使用时间延迟变量来增广矩阵, 提取自相关、互相关的独立信号。

此外, 实际生产过程中常因环境变化、设备磨损老化等引起正常的过程参数缓变, 这种情况下使用较长时间前建立的正常工况模型进行故障检测容易产生误报。为解决这类问题, Li等^[14]提出了迭代PCA(RPCA), 通过不断加入实时数据更新相关矩阵, 使算法可以跟踪正常的缓变过程。Manabu Kano等^[15]提出了Moving PCA方法, 通过监控过程变量相关结构的改变来实现过程缓变特性的跟踪。

2.2.2 非线性特性

虽然大部分生产过程的监控采用线性方法一般可取得较好的结果, 然而对于变量间具有强非线性关系的过程而言, 线性方法往往难以得到好的效果, 这就需要在模型中引入非线性。最直观的解决方法是将变量的非线性项作为新变量加入模型以形成广义PCA(GPCA)。但该方法要求了解非线性部分性质等先验知识, 因而限制了其应用。目前的解决方案主要包括: 基于神经网络的方法、基于核的方法和主曲面方法。神经网络和主曲面方法的共同特点是用非线性函数描述原始数据变量和得分向量之间的关系; 核PCA方法(KPCA)则是由一个非线性映射函数将输入空间投影到高维特征空间, 再在特征空间应用线性PCA算法。从过程监控角度, 基于核函数的方法无需复杂的非线性优化计算, 且可调参数少, 更适合在线监视, 已成为研究热点。另外, 对非线性过程监控而言, 核心问题是建立合适的统计指标、置信限并进行故障识别。如: 为解决置信限的确定问题, Lopes等^[16]提出了基于核概率密度函数的估计方法, 并定义非线性得分来进行在线监控; 为使KPCA方法适于解决非线性过程的故障识别难题, Choi等^[17]基于能量近似概念开发了统一指标用于故障检测, 并在此基础上提出了基于鲁棒重构误差

的故障识别方法.

2.2.3 大规模及多尺度特性

实际工业过程一般包括多个子系统及操作区域, 功能相对独立且包括复杂的耦合关系, 涉及较大规模的过程变量, 这样, 较远距离的变量进行降维时可能得到对诊断不利的结果. 这时, 可利用已有的过程知识对过程变量进行有意义地分块, 从而更好地检测及故障分析. 这些方法主要包括: 层次PCA(HPCA)^[18], 多块PCA(MBPCA)^[19]和一致PCA(CPCA)^[20]. 其区别在于: CPCA对块负载向量(p_b)和超负载向量(p_T)归一化; 而HPCA对块得分向量(t_b)和超得分向量(t_T)归一化. CPCA用超得分向量 $\underline{t}(t_T)$ 求残差; 而MBPCA用块得分向量(t_b)求残差. Qin和Westerhuis^[20]证明了CPCA与常规PCA的一致性, 从而表明可用常规PCA的得分计算CPCA的块得分向量和块残差, 且Qin通过定义CPCA的块SPE, 块 T^2 及其各自的置信限, 更利于直观理解各块及各变量的贡献, 进而定位故障.

实际生产过程可能包含多个频率特征, 不同位置不同性质的故障一般只发生于某个特定频段. 传统PCA方法将整个频率区间综合考虑, 从而影响了故障检测和分离的性能. 为此, Bakshi提出了多尺度PCA(MSPCA)方法, 将数据分频后再进行PCA建模. Misra等^[21]在MSPCA的基础上, 将基于贡献图的故障分离方法扩展到多尺度. 然而, 无论哪种多尺度方法, 频率划分机制目前尚无依据, 仍需根据经验确定.

2.2.4 多模态及多时段特性

工业过程的系统功能一般可通过子功能区域或子功能时段的划分来完成. 上面所述大规模过程的分块方法可视为解决不同区域局部空间内在联系的解决方案. 而某些典型的生产过程(尤其是某些间歇过程)受空间等因素制约, 需要在同一功能区域的不同时段内分别完成不同的功能, 使得各时段内变量间具有与其他时段相异的内在联系, 即具有多时段特性. 该特性与过程缓变特性不同, 难以通过RPCA作相应处理, 可采用不同子时段分别建立监视模型的方式或使用外部分析法来解决. 就第1种方式而言, Zhao等^[22]基于两个PCA模型的相似性评估, 提出一种基于过程知识划分多模态的多模型解决方案; 对于间歇过程的多时段特性可能引入变量非线性, Lu等^[23]提出一种子PCA建模方案, 在假设变量在各批次的同一时间点具有正态分布的前提下, 基于各时间片负荷的模糊聚类结果及过程知识来划分时段; 然而该假设使其难以适应批次间的缓慢变化, 为此, Zhao等^[24]采用一种两阶段ICA-

PCA子空间特征提取方法, 以ICA提取过程的正常缓变. 以上方法均通过典型CSTR过程的检验来表明其有效性. 就第2种解决方式而言, 考虑到过程操作模式的改变是通过调整与操作条件相关的设定值来实现的, Kano等^[25]和Takane等^[26]提出了外部分析法, 并将其用于CSTR过程的监视. 该方法将过程变量 $X = [H \ G]$ 分为外部变量 H 和主变量 G , 基于去除外部变量影响的主变量 $E = G - HC$ 进行多元统计监视, 以使之能够适应过程操作条件的改变. 其中: H 表征操作条件变化及可测扰动, HC 表示主变量受外部变量影响的部分.

2.2.5 间歇特性

间歇过程为一类典型过程, 成批次生产采集到的三维数据, 很难直接建模. 为此, 人们提出了多种间歇过程的故障检测方法, 其中包括: 三维数据二维展开的Multi-way PCA等; 三向模型的PARAFAC (Parallel Factor Analysis) 或Tucker3方法. 近年来, Chang等^[27]提出了MICA以更好地监控间歇过程中可能包含非随机模式的过程. Leo等^[28]比较了两种Unfolding方法和Tucker3方法, 总结出Unfolding MPCA对于全局的间歇变化敏感, Unfolding MPLS对于局部的间歇变化敏感, 而Tucker3方法则介于二者之间的结论. 另外, 为适应批次间的缓变行为, 文献[29]提出了自适应ICA方法, 首先根据有限的批数据开发多阶段的模型, 每个阶段用不同的sub-ICA模型近似; 然后采用自适应更新策略, 使算法能够适应批次间的缓变行为.

2.2.6 混合特性

工业过程往往同时包含多种复杂特性, 为适应这种过程监控, 研究人员已取得了如下进展: 1) 对于具有非线性的动态过程, 文献[30]提出了动态核PCA(DKPCA)方法, 并给出了统一的附有SPE的复合 T^2 指标; 2) 对于具有非线性特性的间歇过程, Lee等^[31]提出了多向核PCA(MWKPCA)方法, 先将批数据展开, 再使用KPCA得到批过程的非线性特性; 3) 为监视具有缓变特性的间歇过程, 文献[32]在间歇过程处于正常操作范围内时, 将间歇数据加入正常历史数据库, 再基于修改的数据库建立新的MPCA模型, 从而利用不断更新的MPCA模型监视过程; 4) 针对具有非线性多尺度特性的过程, Choi等^[33]在对数据进行小波变换及逆小波变换重构的基础上, 使用KPCA捕捉变量间非线性关系并对数据降维, 从而降低了误报率和漏报率, 减小了故障检测延迟.

2.2.7 数据采集、传输、存储过程引入的特性

除了上述多种复杂过程特性会使采集过程的数

据具有相应特性,数据采集、传输、存储过程中的异常以及传感器自身漂移、不同传感器设置的不同采样速率、测量延迟等因素,均会使得数据可能具有缺失点、离群点等异常点以及漂移、测量不同步等多种数据特性,需要在建模时予以考虑。

建模时,模型参数受少数异常点的影响较大,需在建模前进行预处理,或在参数估计时使用鲁棒方法。对于缺失点的情况,可进行如下处理^[34]: 1) 用均值替代; 2) 删除含缺失值的样本; 3) 重构缺失值; 4) 极大似然估计和贝叶斯估计多重插补。离群点的情况可处理如下^[34]: 1) 3σ 离群点检测^[35]; 2) Hampel 辨识器; 3) 滑动窗口滤波器与 Hampel 辨识器相结合; 4) 基于 PCA; 5) 其他。

对于传感器漂移引起的数据漂移问题,可采用与过程缓变相同的处理方式,如 RPCA 处理; 对于关键过程变量及监视变量采用不同的采样速率问题,其处理方式可见文献^[36]; 对于同时记录不同位置的两个变量的采样时测量不同步问题,目前主要基于过程知识对其进行补偿。

3 基于数据和知识的故障诊断

复杂工业过程发生异常情况时,需要在上述改进的多元统计方法进行准确监视的基础上,通过以下方法进行有效诊断,为人们提供辅助诊断的有益信息或直接得到诊断结论。

3.1 基于数据的诊断

人们普遍认为,数据驱动方法从统计(如均值)、概率(如概率密度函数)、分类特征或频率(如小波分析)意义上监视系统状态,均将数据从测量空间变换到特征空间^[37],从而由特征空间得到故障特征。其中尤以多元统计方法为研究热点。该类方法通过将高维数据变换为低维来获得有用信息,更适于大系统的故障检测及从高维数据中提取特征。但该类方法一般难以利用过程及故障知识信息。目前普遍使用贡献图方法,如基于传感器有效指数(SVI)、基于重构的贡献(RBC)^[38,39]或基于角度的贡献(ABC)^[40]进行初步的故障识别与诊断。与基于解析模型及知识的方法相比较,基于多元统计的方法具有以下优势: 1) 多元统计方法认为贡献图对故障分析的指导同样是诊断,从而丰富了诊断的含义。2) 基于模型的诊断方法是先假设存在一个故障(如传感器、执行器故障),再找到故障,但一般故障发生前缺乏故障知识或深刻的过程知识;而多元统计方法可以在较少的先验知识前提下提供贡献图等指导信息,更适用于没有(或很少)故障知识情况下设备、来料、误操作及参数不匹配等过程的故障诊断。

贡献图是目前普遍采用的故障辨识工具,其通过各变量对统计指标的贡献来找到受故障影响最大的异常变量。重构诊断的基本思想是:通过沿定义的故障方向抑制故障对统计指标的影响来估计正常状态量,可最有效地抑制故障影响的方向即为故障方向,幅度为故障大小。该方法首先由 Wise 和 Ricker^[41]用于传感器重构,随后扩展到由经验指定或由历史故障数据确定的任意方向上的过程故障的重构^[42]。从应用角度看, Qin 不仅给出了故障的可重构及可识别条件,且指出了重构法的两种使用方式: 1) 无先验知识的情况下,使用重构方法得到故障重构指数(FRI)以识别故障,再由人依过程知识解释结果; 2) 由先验知识或历史故障确定的故障方向信息直接通过重构方法进行诊断。对于第 1 种方式, Carlos 等通过仿真实例分析了贡献图可能导致错误的诊断,给出了采用传统贡献图及基于重构的贡献进行正确诊断的必要条件,并通过蒙特卡洛仿真表明了 RBC 方法比传统贡献图方法诊断正确率更高;对于重构的第 2 种诊断方式, Qin 从数据要求、历史信息的利用、未发生过故障的处理能力这 3 个角度,总结了其与贡献图方法的区别。因此,对于生产过程中典型的、重大的、具有明确故障方向且满足可重构条件的故障,可采用重构的第 2 种方式诊断;而对于未发生过的及较少故障先验知识(或历史故障数据)的故障,则可通过贡献图(或 RBC)结合过程知识^[43,44]辅助分析故障。

采用多元统计方法进行故障诊断除以上两种主要方式外,还可以基于数据的分类特征进行诊断。如基于得分分布图的数据分类特征的方法,以及判别分析(如费舍尔判别式分析^[45])和聚类分析(如模糊 C 均值聚类)这两种方式。

3.2 基于知识的诊断

基于知识(知识被定义为过程输入输出、不正常模式、故障特征、操作约束、评价等)的诊断方法因能将过程知识尤其是故障知识与相关推论结合起来而适合于故障诊断。主要包括因果分析、专家系统、模式识别等 3 类方法。

1) 因果分析: 基于故障症状关系的因果模型,包括符号定向图 SDG^[46]、症状树 STM 方法等;

2) 专家系统: 用来模拟专家诊断故障时的推理,作为解释器,有基于规则的^[47]、基于案例的方法等;

3) 模式识别: 利用数据模式和故障类之间的关系进行诊断,如贝叶斯分类器、神经网络分类,通过输入故障症状和输出故障原因进行诊断。

以上各类基于知识的诊断方法都可归结为通过历史的操作及过程理解获得的事实、规则、启发信息进行诊断。从搜索方式角度,可分为与正常操作集不

匹配以及与已知异常症状匹配两种^[48]。该类方法的优势在于无需过程的详细数学模型, 且使用过程中可方便加入过程知识及故障知识, 擅长建立故障特征空间及故障类空间的关系, 更适于故障诊断且结论易于理解。但构造一个大系统的故障模型需要付出巨大的努力。目前该类方法大多应用于输入、输出和状态数相对较小的系统。

近年来, 基于知识的方法用于过程故障尤其是与质量预报相关的异常工况的诊断已成为热点, 这是由于复杂工业过程可能发生多种故障, 一般并非传感器和执行器的故障, 而是由工艺操作、控制、设备三方面协调不当而引起的。如给定的控制系统设定值不合适或运行条件变化与操作不当等因素造成异常工况, 如果不及时处理, 会使控制系统性能变坏, 甚至造成系统瘫痪。这类异常工况属于过程故障, 难以采用传统的故障诊断方法来诊断以及相应的容错控制方法来消除。而专家系统则可模拟人类专家在诊断故障时的推理, 作为解释器, 执行故障诊断并推测结果以诊断这类异常工况问题。如文献[49]采用过程参量作为工况描述的案例推理故障诊断系统, 用于竖炉焙烧过程的异常工况预报; 文献[47]将现有的故障诊断方法和规则推理与案例推理等智能方法相结合, 以规则推理诊断故障工况, 并建立以调整控制回路设定值为手段的容错控制方法, 使得系统在发生异常工况时能尽快恢复到正常工况范围, 并使系统逐渐远离异常工况, 将指标的实际值控制在目标值范围内。

对于故障产生原因的诊断, 目前各种方法的应用仍然有限, 主要依靠人来完成。

3.3 数据和知识相结合的诊断

从以上两类故障诊断方法可以看出: 1) 多元统计监控方法本身只是统计工具, 不包含用于诊断的过程知识。虽然贡献图可提供故障指导, 重构法可用于诊断典型故障, 但二者仍非普遍意义上的诊断方法, 主要用于定位故障区域以及辅助人来诊断。2) 基于知识的方法虽适于诊断, 但不适于解决大规模的问题, 且现有基于知识的诊断方法一般都使用单变量故障检测机制(精度和灵敏度都不够), 没有利用实时数据的统计信息。

单纯的基于知识或基于数据的方法不能进行完整的故障诊断, 且诊断效率和准确性均有待于提高。目前已有学者考虑二者的综合, 如图1所示。

通过多元统计方法监视过程, 监控到异常情况时触发基于知识的系统(KBS)^[50]。KBS基于过程行为(MSPM产生的统计信息、贡献图)、过程知识等模拟专家进行故障诊断, 即将多元统计方法得到的低维定性、定量(或模糊化)信息送入KBS, 结合过程知

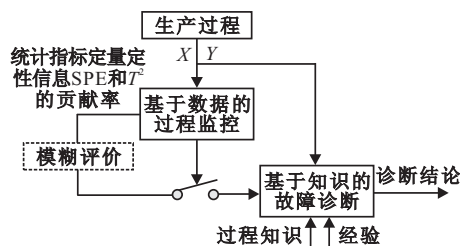


图1 统计监控与知识联合诊断图

识和经验进行故障诊断。如文献[51]用KBS自动解释MSPM图; [52]将MSPC信息送入SPG方法进行诊断。但这两种方式都是将数据统计信息转换为简单的符号信息用于诊断, 忽略了偏离程度等定量信息。对此, [53]提出了概率因果关系(PCEG)方法, 使用多元统计得到的SPE和 T^2 的变量贡献率的定量信息在概率框架内诊断故障。

4 结 论

尽管多元统计方法及基于知识的监控和诊断方法已在很多工业过程尤其是化工过程取得了良好的应用, 为进一步改进监控和诊断的性能, 以满足工业应用的需要, 仍有一些问题有待于深入研究。

(1) 多元统计方法信息用于诊断

目前已有研究通过数据和知识方法的综合进行故障诊断, 而考虑到基于数据方法可得到定量信息, 基于知识方法多为定性推理, 故需要探讨如何将基于数据的不具有明确物理意义的主元及统计指标等定性、定量信息用于基于知识的定性推理。

(2) 诊断与人的交互作用

故障诊断系统的建立以及实施过程中, 由于人对整个过程以及故障的多样性和多变性的了解更深入, 人始终发挥着至关重要的作用。例如: 1) 模型建立过程中由人对变量进行选择; 2) 结合过程知识对结果进行解释。因此需要系统地考虑人在故障诊断系统中的作用。目前很多过程监控系统将人的作用包含进去, 以各种故障诊断方法为工具获得故障信息, 然后由人进行最终的诊断。例如目前故障诊断应用比较发达的医疗领域, 最终的诊断结论仍需要医生的确认, 而未实现完全自动化。从工业应用角度, 目前条件下, 一方面需要研究建立融入过程知识的过程数据建模方法; 另一方面需加强与专家的交互, 并注重人的维护功能方法的研究, 以发挥人解释故障的优势。

(3) 混合多种复杂特性过程的监控

目前, 对于系统复杂过程的监控主要集中于单一特性, 而实际工业过程混合了多种复杂特性(时变、动态、非线性、间歇、分频、大规模等), 应进一步将各种方法综合使用, 以适应复杂的生产过程。例如对于大规模非线性过程的监视, 若采用核方法提取变

量间的非线性关系,则需考虑非线性函数将输入映射到高维特征空间对实时计算带来的难度,此时可采用分块方法降低问题的求解难度;再如,考虑过程缓变的大规模过程的监视,需要定义并求取分块情况下不断更新统计指标及其置信限。

(4) 混合多种复杂特性过程的诊断

考虑到复杂的数据特性,基于多元统计的诊断方法用于建模或求取贡献率时,应进行相应处理,如: 1) 考虑到分频特性,数据分频后建模,再计算变量对某个频率的贡献,以诊断发生于不同频段的故障; 2) 定义变量对分块统计指标的贡献率,以进行大规模过程的故障诊断; 3) 对于多模式(或多时段)问题的故障诊断,需要对各个模式(或各时段)分别建模。

另外,考虑到过程复杂特性引入的诊断复杂性,需要处理好在线/离线诊断的关系: 1) 对于有快速容错动作需求的诊断问题,需采用在线诊断方法,有待研究如何保证多次发生相同故障时,由数据和经验提取的故障特征的一致性及其与故障原因关联的准确性; 2) 过程的一般监视可采用在线监视、离线诊断的方式解决。

(5) 多个层次的诊断

生产过程的控制包括底层回路控制、过程控制、RTO 实时优化、计划调度等多个层次。各层次自身及层次间协调均有异常的可能,而目前的方法局限于底层和过程控制级,因此有待于向上层扩展的研究。

随着信号采集存储技术逐渐应用于现代工业过程,不断涌现具有各自过程特点的过程监视、过程运行可靠性、改善产品质量的新需求,以及亟待解决的以过程监视和诊断为目的的数据挖掘和知识发现问题。如此广阔的研究空间,将更有利于基于数据和知识的过程监视及故障诊断方法的推广以及进一步发展,并最终推动整个过程工业安全、节能、高质、高效、环保地生产。

参考文献(References)

- [1] Kano M, Nakagawa Y. Data-based process monitoring, process control, and quality improvement: Recent developments and applications in steel industry[J]. *Computers & Chemical Engineering*, 2008, 32(1/2): 12-24.
- [2] Kano M, Nagao K, Hasebe S, et al. Comparison of multivariate statistical process monitoring methods with applications to the Eastman challenge problem[J]. *Computers & Chemical Engineering*, 2002, 26(2): 161-174.
- [3] Qin S J, Valle S, Piovoso M J. On unifying multiblock analysis with application to decentralized process monitoring[J]. *J of Chemometrics*, 2001, 15(9): 715-742.
- [4] Qin S J, Cherry G, Good R, et al. Semiconductor manufacturing process control and monitoring: A fab-wide framework[J]. *J of Process Control*, 2006, 16(3): 179-191.
- [5] Zhang Y, Dudzic M S. Online monitoring of steel casting processes using multivariate statistical technologies: From continuous to transitional operations[J]. *J of Process Control*, 2006, 16(8): 819-829.
- [6] Undey C, Tatara E, Cinar A. Real-time batch process supervision by integrated knowledge-based systems and multivariate statistical methods[J]. *Engineering Applications of Artificial Intelligence*, 2003, 16(5/6): 555-566.
- [7] Desborough L, Harris T. Performance assessment measures for univariate feedback control[J]. *J of Chemical Engineering*, 1992, 70(6): 262-268.
- [8] Kesavan P, Lee J H. Diagnostic tools for multivariable model-based control systems[J]. *Industrial Engineering Chemistry Research*, 1993, 36(7): 2725-2738.
- [9] Dunia R, Qin S J, Edgar T F, et al. Identification of faulty sensors using principal component analysis[J]. *AIChE J*, 1996, 42(10): 2797-2812.
- [10] Qin S J. Statistical process monitoring: Basics and beyond[J]. *J of Chemometrics*, 2003, 17(8/9): 480-502.
- [11] Lee J M, Yoo C K, Lee I B. Statistical process monitoring with independent component analysis[J]. *J of Process Control*, 2004, 14(5): 467-485.
- [12] Kourti T, MacGregor J F. Multivariate SPC methods for process and product monitoring[J]. *J of Quality Technology*, 1996, 28(4): 409-428.
- [13] Lee J M, Yoo C K, Lee I B. Statistical monitoring of dynamic processes based on dynamic independent component analysis[J]. *Chemical Engineering Science*, 2004, 59(14): 2995-3006.
- [14] Li W H, Yue H, Cervantes S V, et al. Recursive PCA for adaptive process monitoring[J]. *J of Process Control*, 2000, 10(5): 471-486.
- [15] Kano M, Hasebe S, Hashimoto I, et al. A new multivariate statistical process monitoring method using principal component analysis[J]. *Computers & Chemical Engineering*, 2001, 25(7/8): 1103-1113.
- [16] Lopes J A, Menezes J C. Multivariate monitoring of fermentation processes with non-linear modelling methods[J]. *Analytica Chimica Acta*, 2004, 515(1): 101-108.
- [17] Choi S W, Lee C, Lee J M, et al. Fault detection and identification of nonlinear processes based on kernel PCA[J]. *Chemometrics and Intelligent Laboratory Systems*, 2005, 75(1): 55-67.

- [18] Wold S, Kettaneh N, Tjessem K. Hierarchical multiblock PLS and PC models for easier model interpretation and as an alternative to variable selection[J]. *J of Chemometrics*, 1996, 10(5/6): 463-482.
- [19] Cheng G, McAvoy T. Multi-block predictive monitoring of continuous processes[C]. *IFAC Symposium on Advanced Control of Chemical Processes*. Banff, 1997: 73-77.
- [20] Westerhuis J A, Kourti T, MacGregor J F. Analysis of multiblock and hierarchical PCA and PLS models[J]. *J of Chemometrics*, 1998, 12(5): 301-321.
- [21] Misra M, Yue H, Qin S J, et al. Multivariate process monitoring and fault diagnosis by multi-scale PCA[J]. *Computers & Chemical Engineering*, 2002, 26(9): 1281-1293.
- [22] Zhao S J, Zhang J, Xu Y M. Monitoring of processes with multiple operating modes through multiple principle component analysis models[J]. *Industrial Engineering Chemistry Research*, 2004, 43 (22): 7025-7035.
- [23] Lu N Y, Gao F R. Sub-PCA modeling and on-line monitoring strategy for batch processes[J]. *AIChE J*, 2004, 50(1): 255-259.
- [24] Zhao C H, Wang F L, Gao F R, et al. Enhanced process comprehension and statistical analysis for slow-varying batch processes[J]. *Industrial Engineering Chemistry Research*, 2008, 47(24): 9996-10008.
- [25] Kano M, Hasebe S, Hashimoto I, et al. Evolution of multivariate statistical process control: Application of independent component analysis and external analysis[J]. *Computers & Chemical Engineering*, 2004, 28(6/7): 1157-1166.
- [26] Takanel Y, Hunter M A. Constrained principal component analysis: A comprehensive theory[J]. *Applicable Algebra in Engineering, Communication and Computing*, 2001, 12(5): 391-419.
- [27] Yoo C K, Lee J M, Vanrolleghem P A, et al. On-line monitoring of batch processes using multiway independent component analysis[J]. *Chemometrics and Intelligent Laboratory Systems*, 2004, 71(2): 151-163.
- [28] Chiang L H, Leardi R, Pell R J, et al. Industrial experiences with multivariate statistical analysis of batch process data[J]. *Chemometrics and Intelligent Laboratory Systems*, 2006, 81(2): 109-119.
- [29] Zhao C H, Wang F L, Mao Z Z, et al. Adaptive monitoring based on independent component analysis for multiphase batch processes with limited modeling data[J]. *Industrial Engineering Chemistry Research*, 2008, 47(9): 3104-3113.
- [30] Choi S W, Lee I B. Nonlinear dynamic process monitoring based on dynamic kernel PCA[J]. *Chemical Engineering Science*, 2004, 59(24): 5897-5908.
- [31] Lee J M, Yoo C K, Lee I B. Fault detection of batch processes using multiway kernel principal component analysis[J]. *Computers & Chemical Engineering*, 2004, 28(9): 1837-1847.
- [32] Lee J M, Yoo C K, Lee I B. On-line batch process monitoring using a consecutively updated multiway principal component analysis model[J]. *Computers & Chemical Engineering*, 2003, 27(12): 1903-1912.
- [33] Choi S W, Morris J L, Lee I B. Nonlinear multiscale modelling for fault detection and identification[J]. *Chemical Engineering Science*, 2008, 63(8): 2252-2266.
- [34] Kadlec P, Gabrys B, Strandt S. Data-driven soft sensors in the process industry[J]. *Computers & Chemical Engineering*, 2009, 33(4): 795-814.
- [35] Lin B, Recke B, Knudsen J, et al. A systematic approach for soft sensor development[J]. *Computers & Chemical Engineering*, 2007, 31(5): 419-425.
- [36] Ding F, Chen T. Modeling and identification for multirate systems[J]. *Acta Automatica Sinica*, 2005, 31(1): 105-122.
- [37] Venkatasubramanian V, Rengaswamy R, Kavuri S N, et al. A review of process fault detection and diagnosis, Part III: Process history based methods[J]. *Computers & Chemical Engineering*, 2003, 27(3): 327-346.
- [38] Alcalá C, Qin S J. Reconstruction-based contribution for process monitoring[C]. *Proc of 17th IFAC World Congress*. Seoul, 2008: 1593-1600.
- [39] Alcalá C, Qin S J. Reconstruction-based contribution for process monitoring[J]. *Automatica*, 2009, 45(7): 1593-1600.
- [40] Yoon S, MacGregor J F. Fault diagnosis with multivariate statistical models, Part I: Using steady state fault signatures[J]. *J of Process Control*, 2001, 11(4): 387-400.
- [41] Wise B M, Ricker N L. Recent advances in multivariate statistical process control: Improving robustness and sensitivity[C]. *Proc of IFAC ADCHEM Symp*. Toulouse, 1991: 125-130.
- [42] Dunia R, Qin S J. Subspace approach to multidimensional fault identification and reconstruction[J]. *AIChE J*, 1998, 44(8): 1813-1831.
- [43] Miller P, Swanson R E, Heckler C F. Contribution plots: A missing link in multivariate quality control[J]. *Int J of Applied Mathematics Computer Science*, 1998, 8(4): 775-792.
- [44] Chen J G, Bandoni J A, Romagnoli J A. Robust PCA and normal region in multivariable statistical process monitoring[J]. *AIChE J*, 1996, 42(12): 3563-3566.