

JIYU SHUJU DE GUZHANG FENLI 基于数据的故障分离

张颖伟 著



東北大學出版社
Northeastern University Press

基于数据的故障分离

张颖伟 著

东北大学出版社

· 沈 阳 ·

© 张颖伟 2016

图书在版编目 (CIP) 数据

基于数据的故障分离 / 张颖伟著. — 沈阳: 东北
大学出版社, 2016. 11

ISBN 978-7-5517-1471-6

I. ①基… II. ①张… III. ①故障检测 IV. ①TB4

中国版本图书馆 CIP 数据核字(2016)第 290815 号

出 版 者: 东北大学出版社

地址: 沈阳市和平区文化路三号巷 11 号

邮编: 110819

电话: 024-83683655(总编室) 83687331(营销部)

传真: 024-83687332(总编室) 83680180(营销部)

网址: <http://www.neupress.com>

E-mail: neuph@neupress.com

印 刷 者: 沈阳航空发动机研究所印刷厂

发 行 者: 东北大学出版社

幅面尺寸: 170mm×240mm

印 张: 11.5

字 数: 225 千字

出版时间: 2016 年 11 月第 1 版

印刷时间: 2016 年 11 月第 1 次印刷

责任编辑: 刘乃义

封面设计: 刘江旸

责任校对: 文 浩

责任出版: 唐敏志

ISBN 978-7-5517-1471-6

定 价: 68.00 元

前 言

随着人们对工业品产能需求的日益扩大和对产品质量要求的日益提升，现代工业生产及制造过程变得结构越来越复杂、规模越来越庞大，通过信息技术和自动化技术来组织和进行生产活动已成为必然趋势。随着系统复杂性的提高和系统各组成部分耦合程度的加强，该类系统往往存在着许多不安全因素，并且其发生事故的严重性也在加强。对复杂系统的可靠性要求使得过程故障诊断技术成为近年来自动化领域内的研究热点。

随着科学技术水平的不断提高，现代生产设备向着大规模化、集成化、智能化和精密化的方向发展，生产系统对自动化程度的要求也越来越高。计算机技术和自动化技术在工业生产领域的广泛应用，不仅极大地降低了生产成本、减少了能耗，并且大幅提高了产品生产效率，为企业创造了可观的利润的同时也为国家带来了巨大的社会效益。但是，由于其系统结构和生产设备相对复杂，工业过程表现出高度的非线性、大范围的不确定性、易受干扰、强耦合性等特点，这给生产过程的管理和维护带来了极大的困难。如果一处出现的异常没有得到及时的处理，就可能引起连锁反应，将使得整个生产过程中断，给企业带来巨大的经济损失，情况严重的，甚至会造成人员伤亡。尤其是在化工、电力、冶金等具有高复杂性和高危险性的生产领域中，若生产过程中出现的异常未能及时排除，因此而导致的后果将更加严重。1984年12月，美国联合碳化物公司在印度博帕尔市的农药厂发生了毒气泄漏事故，造成2000多人死亡，20余万人中毒，被称为世界工业史上最大的恶性事故；1986年，苏联切尔诺贝利核电站的核泄漏事故，造成2000多人死亡，直接经济损失达30亿美元，对自然环境的影响空前巨大，被称为生态灾难，事故带来的严重后果至今还在延续中。2013年2月，贵阳市某化工厂车间发生爆炸事故，造成5人受伤。这些事故的发生都造成了一定程度的经济损失和人员伤亡。根据国内外许多资料显示，众多发达国家已经将过程故障诊断技术应用于工业生产中，不仅降低了事故发生率，还取得了可观的经济效益。由此可见，复杂工业生产过程的安全性、可靠性已成为保障经济效益与社会效益的一个关键因素，应该得到高度重视。

几十年的理论研究与实际应用表明,过程故障诊断技术为提高现代生产过程的安全性及可靠性提供了一种新的思路。过程故障诊断技术是一门综合性技术,具有很强的学科交叉性,其在近几十年的发展中取得了许多成果。现代工业生产过程中,设备运行过程的状态检测、系统的故障检测与诊断都可以用故障诊断技术来实现,因此对故障诊断方法的研究,在现代工业生产过程中具有深刻的理论价值。

本书主要介绍基于数据驱动的故障监测和诊断方法,重点是对 PCA、ICA 和 PLS 方法的改进,根据数据的不同特点,将不同的方法有效融合,进而给出有效的故障监测和诊断效果。

第1章概述了故障监测与诊断技术,研究了国内外故障监测和诊断技术的发展和相关方法。

第2章提出了基于自适应核主元分析的过程监测。在实际的工业过程中,过程常具有时变的特性,由于传统的 KPCA 方法所建立的模型是非时变的、固定的模型,所以时变可能会导致系统误报警。针对这一特点,可以采用自适应核主元分析方法进行过程监测。自适应核主元分析方法可以实时地采集过程数据,实时地更新协方差矩阵和主元,进而更新核主元模型。常用的自适应核主元分析方法有递归核主元分析方法、滑动窗口核主元分析方法和指数加权核主元分析方法。自适应核主元分析算法有效地解决了传统方法不能实现模型实时更新的缺点,可以使过程监测更加有效。本章由张颖伟与张薇合著。

第3章提出了基于方向核偏最小二乘的过程监测方法,提出了改进的方向 PLS(DPLS)算法,然后将 DPLS 算法与核方法结合到一起运用到非线性的过程监测中去,由此推导出了核 DPLS 方法(DKPLS)。本章由张颖伟与张玲君合著。

第4章针对非线性、非高斯过程故障分离问题进行了研究,提出了在 KICA 方法基础上的基于故障特征方向的故障分离方法。其中介绍了基于独立元分析方法在独立元空间中故障信息丰富的特点、提出的思路 and 具体实现方法,将此方法应用于电熔镁炉实验数据的研究,并应用传统的贡献图和 KPCA 重构方法作为对比实验。结果显示,新方法在非高斯过程的故障特征方向的提取和基于故障特征方向的故障分离有较好的效果,能准确分离出新的故障数据的故障类型。证明了本章方法的正确性和有效性。

针对具有非线性、非高斯特性的多模式过程监测问题,从监测统计量角度提取公共和特殊信息,提出了一种基于改进的 KICA 的多模式核独立元分析方法(MKICA)并将其用于多模式过程监测。用此方法对田纳西(Tennessee Eastman Process)多模式过程进行了实验研究。此方法建立的公共模型能有效检测各模式下的故障,通用性较好;此方法建立的各模式的特殊模型对各模式

下的故障检测的灵敏性和有效性较好。此方法的提出,对非高斯多模式过程的故障检测领域的研究具有一定的创新意义。本章由张颖伟与杨楠合著。

第5章介绍了基于数据的过程监测及故障分离方法。提出了一种基于KPCA的故障重构方法,解决了非线性过程的故障分离问题。在离线建模阶段,该方法采用KPCA方法将故障数据空间分解为主元子空间和残差子空间,利用所得的负载方向对正常数据进行投影。利用PCA方法对投影的数据进行分析,通过比较各个方向上故障数据与正常数据的得分提取出引起统计量超限的故障方向,建立了故障重构模型。在线监测过程中,利用各故障重构模型依次对检测到的故障数据进行重构,只有当前故障所对应的故障模型能够正确地去掉数据中的故障信息,消除检测统计量超限报警现象,据此可确定故障类别,达到故障分离的目的。通过对电熔镁过程的数据建模以及其故障的检测和分离,验证了本章所提方法的有效性。

另外,本章又提出了一种基于KLSR的故障分离方法。该方法避免了对各类故障分别建模的烦琐过程,通过将不同类别的故障样本集投影到相应的回归目标,以实现故障分离。核函数方法的引入有效地解决了多类分类中不同类别的数据线性不可分的问题。此外,考虑到过多的训练样本会导致核矩阵存储空间与计算量的增加,在目标函数结构正则化中引入了 $L_{2,1}$ 范数,求解具有稀疏性的权值矩阵,并以此提取对建模作用较大的训练样本用以描述特征空间中的权值向量。通过应用于iris数据集以及电熔镁过程的仿真实验,可以看出所提方法具有良好的分离效果。本章由张颖伟与王正兵合著。

由于作者水平有限,书中难免有不妥之处,恳请各位专家和广大读者批评指正。

著 者

2016年5月

目 录

第 1 章 故障监测与诊断技术概述	1
1.1 多元统计过程故障监测概述	1
1.1.1 多元统计过程监测技术概述	1
1.1.2 多元统计过程监测方法概述	3
1.1.3 多元统计过程监测国内外研究现状	7
1.2 多元统计过程故障诊断概述	9
1.2.1 多元统计过程故障诊断技术概述	9
1.2.2 多元统计过程故障诊断方法概述	10
1.2.3 多元统计过程故障诊断国内外研究现状	13
本章参考文献	15
第 2 章 基于自适应核主元分析的过程监测	19
2.1 自适应核主元分析方法基础理论	21
2.1.1 基于滑动窗口机制的核主元分析	21
2.1.2 指数加权核主元分析方法	23
2.2 基于自适应核主元分析的过程监测	24
2.2.1 自适应核主元分析方法的建模过程	24
2.2.2 基于自适应核主元分析的在线过程监测	26
2.2.3 仿真研究与结果分析	26
2.3 基于特征空间损失函数的核主元分析方法	34
2.3.1 特征空间中的损失函数	34
2.3.2 基于特征空间损失函数的核主元分析方法	35
2.4 基于遗忘因子的核主元分析模型更新	36
2.5 迭代形式的核主元分析算法	38
2.5.1 迭代形式的核主元分析算法	38
2.5.2 加入惩罚因子的迭代核主元分析算法	39

2.6	基于惩罚因子的自适应核主元分析过程监测	41
2.7	仿真研究与结果分析	42
2.8	本章小结	47
	本章参考文献	47
第3章	基于方向核偏最小二乘的过程监测方法	50
3.1	方向偏最小二乘算法的推导	51
3.1.1	PLS 残差与输出变量之间的相关性证明	51
3.1.2	相关部分的求取	53
3.1.3	残差剩余部分与输出变量之间的相关性证明	54
3.2	基于 DKPLS 的过程监测方法	54
3.2.1	核函数方法	55
3.2.2	基于 DKPLS 的过程监测	55
3.3	实验结果	59
3.3.1	电熔镁炉工作原理	59
3.3.2	实验结果分析	60
3.4	基于方向核偏最小二乘(DKPLS)的故障诊断方法	66
3.4.1	故障主元方向的选择方法	67
3.4.2	基于 DKPLS 的重构方法	68
3.5	基于方向核偏最小二乘(DKPLS)的故障重构诊断方法	73
3.6	仿真实验	76
3.6.1	故障主元方向的选择	76
3.6.2	故障的重构诊断	79
3.7	本章小结	83
	本章参考文献	84
第4章	基于故障特征方向的 KICA 故障分离方法	90
4.1	经典重构中的故障方向	91
4.2	独立元空间故障特征方向提取的仿真分析	92
4.3	改进的 KICA 故障分离原理	96
4.3.1	故障相关方向的提取	96
4.3.2	在线故障分离	100
4.4	仿真和结果分析	101
4.4.1	电熔镁炉的生产过程描述	101
4.4.2	仿真结果分析	103

4.5	多模式核独立元分析方法	107
4.5.1	公共模型和特殊模型分析	107
4.5.2	独立元空间公共模型和特殊模型的建立	109
4.5.3	残差空间的公共模型和特殊模型的建立	113
4.5.4	MKICA 在线监测	115
4.6	仿真研究与结果分析	118
4.6.1	田纳西过程介绍	118
4.6.2	田纳西数据仿真结果分析	119
4.7	本章小结	126
	本章参考文献	127

第5章 基于数据的过程监测及故障分离方法

5.1	基于 KPCA 的子空间划分方法	133
5.2	基于 KPCA 的故障重构方法	134
5.2.1	基于 T^2 统计量重构的故障特征方向的提取	134
5.2.2	基于 SPE 统计量重构的故障特征方向的提取	136
5.2.3	故障重构方法	137
5.3	基于 KPCA 方法的在线故障分离	138
5.4	仿真研究	140
5.4.1	电熔镁炉工作过程	140
5.4.2	仿真结果分析	140
5.4.3	故障特征方向的特性讨论	146
5.5	基于 KLSR 的故障分离方法	148
5.5.1	核最小二乘回归(KLSR)算法	149
5.5.2	基于 KLSR 的故障分离	152
5.6	基于 DS-KLSR 的故障分离方法	153
5.6.1	基于核最小二乘回归的数据提取算法	153
5.6.2	基于 DS-KLSR 的故障分离	159
5.7	仿真研究	160
5.7.1	算法分类特性分析	160
5.7.2	仿真结果分析	163
5.8	本章小结	167
	本章参考文献	167

第 1 章 故障监测与诊断技术概述

1.1 多元统计过程故障监测概述

1.1.1 多元统计过程监测技术概述

过程监测是一门以系统故障检测和诊断技术为基础发展起来的边缘性学科。过程监测方法的提出可以追溯到 1971 年,美国麻省理工学院的 Beard 博士首先提出了用“解析冗余”的概念代替物理冗余的故障检测和故障诊断的思想,开创了过程监测的先河,为过程监测系统理论奠定了基础。过程监测是一门目的明确、针对性强而且应用领域广泛的技术学科,其核心任务是故障检测与诊断(Fault Detection and Diagnosis, FDD)。故障检测是对过程的运行状态进行监视,一旦过程偏离了正常状态,则尽快报告故障的发生。故障诊断是对已知存在故障的过程判定故障的根源,一个有效的监测系统能帮助操作人员在故障发生后及时地采取正确的修复措施以阻止故障的传播和减少更大的损失。一般来说,流程工业的监测过程如图 1.1 所示。

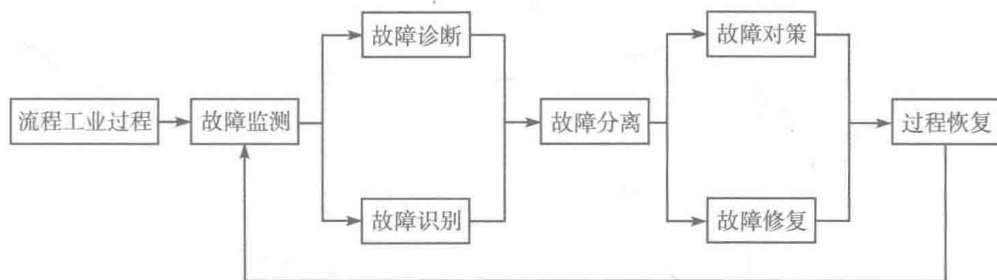


图 1.1 流程工业的监测过程示意图

Fig. 1.1 Diagram of the process industrial monitoring

FDD 的一些基本概念在众多研究中的描述不尽相同,本书综合了统计过

程监测方法的一些研究成果, 给出过程监测中的一些基本概念及定义。

(1) 故障(Fault)

故障是指过程中至少一个特性或变量相对于正常状态的不允许的偏离。可见, 故障指的是过程中发生的不正常现象。另外有些研究学者将“故障”定义为产生这种不正常现象的原因。本书中, 将导致过程异常事件的潜在原因称作基本事件(Basic Events)或故障源(Root Causes)。在工业生产中, 故障可以来自过程中的一些生产设备, 也可以来自其上的一些测量和控制装置(传感器和执行器)。就影响程度来说, 故障的出现可能会令相应部分的性能降低, 也可能令部分功能失灵, 甚至可能完全崩溃。根据故障来源的不同, 一般来说, 故障可以分为以下几类。

① 干扰参数故障: 此类故障是由于过程受到的随机扰动引起的, 也称为过程干扰故障。相对于过程自身而言, 它是一种外在故障, 如因随机干扰而使过程进料的浓度偏离正常值或反应器物料流量不稳定、环境温度的极端变化等。

② 过程参数故障: 此类故障是由于系统元部件功能失效或系统参数发生变化引起的, 如控制器失效、操作阀失灵、催化剂中毒以及热交换器结垢等。

③ 传感器或测量仪表故障: 此类故障是因传感器或者测量仪表功能失灵而导致的测量数据发生的偏差超出可接受的范围, 它会导致控制系统的性能迅速降低。

④ 执行器故障: 此类故障是指控制回路中用于执行控制命令的部件发生类似于恒增益变化的故障而不能正确地执行控制命令, 具体表现为执行器的设定值和它的实际输出之间的偏差。

(2) 故障检测(Fault Detection)

故障检测是指从过程可测量的信息或不可测变量的估计信息中提取出描述过程的特征信息, 以此来确定系统中是否有故障发生。

(3) 故障识别(Fault Identification)

故障识别是指故障发生时, 识别出与诊断故障最有关联的观测变量或观测变量子集, 又称为故障隔离(Fault Isolation)。

(4) 故障诊断(Fault Diagnosis)

故障诊断有广义和狭义之分。广义上的故障诊断通常指故障诊断这一领域, 包括故障的检测、识别、诊断与恢复; 狭义上指在上述步骤后, 确定出故障的原因, 甚至包括故障的类型、位置、量级和时间。故障诊断本质上是一类模式分类问题。

(5) 故障对策(Fault Strategy)

故障对策是指经过程故障的检测、识别与诊断后, 根据故障可能对系统带

来的不利影响来制订相应的应对策略,以确保过程能恢复至正常的运行状态。

1.1.2 多元统计过程监测方法概述

过程监测方法的核心在于利用过程先验知识对过程的正常操作状态(Normal Operating Conditions, NOC)和各种故障状态分别建立参考模型,作为分析过程状态是否存在故障以及判定故障类别的依据。国际故障诊断权威 Frank P. M. 把故障诊断(广义)的方法分为三类:基于数学模型的方法、基于信号处理的方法和基于知识的方法。按这种分类方式,过程监测领域的各类研究方法可以总结为如图 1.2 所示。

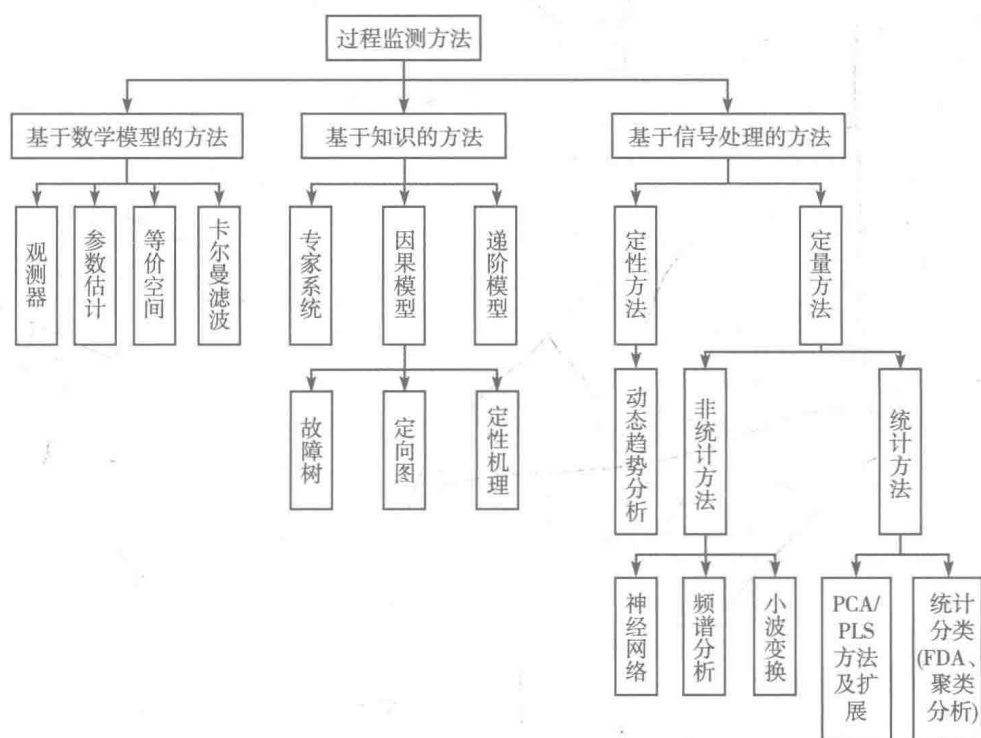


图 1.2 传统的过程监测方法的分类

Fig. 1.2 Traditional classification of process monitoring method

但在过程监测十几年的发展过程中,随着近年来统计监测理论与应用的发展,很多方法难以归入上述三种分类,如与统计学有关的方法,归为信号处理的范畴并不适当。另外,解析法、图形法等方法,将其归为基于知识的方法也不准确。由于过程监测方法的核心在于利用过程的先验知识对过程的正常操作状态和各种故障状态分别建立参考模型,作为分析过程状态是否存在故障以及判定故障类别的依据,因此根据所需的先验知识的不同, Venkatasubramanian 将过程监测方法分成三类:基于定量机理模型的方法、基于定性模型的方法和

基于数据驱动的方法。按这种分类方式，过程监测方法可以总结为如图 1.3 所示。

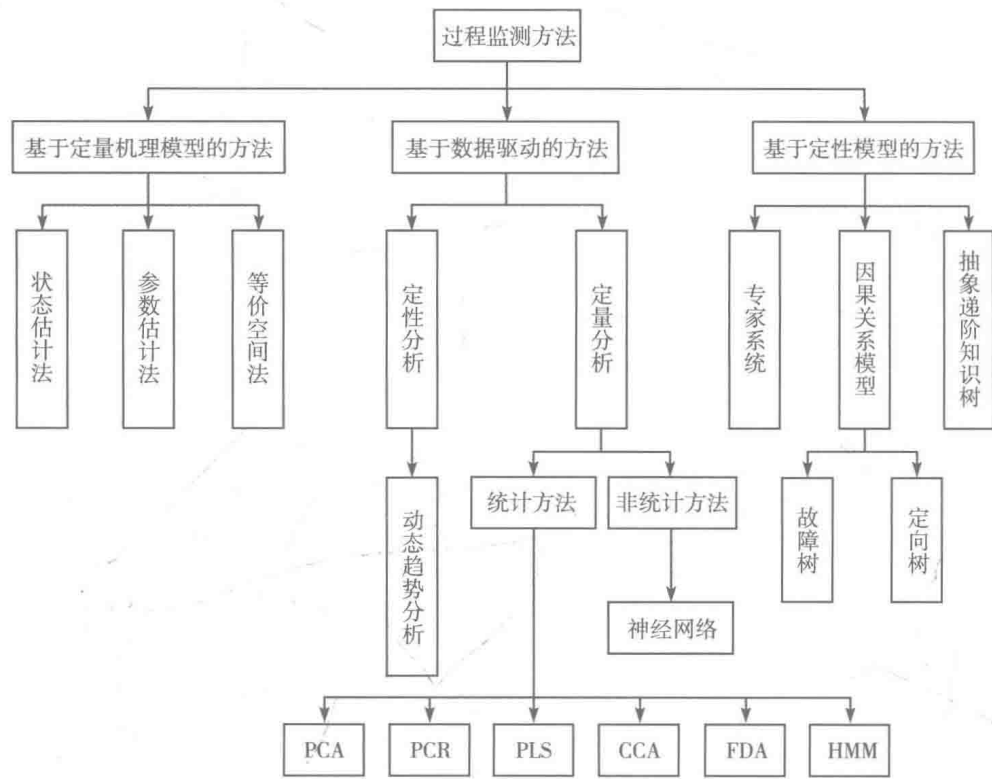


图 1.3 根据先验知识的过程监测分类方法

Fig. 1.3 Classification of process monitoring method based on prior knowledge

(1) 基于定量机理模型的方法

基于定量机理模型的方法又称为解析模型方法，也是最早被研究和应用的方法，此方法要求对被监测对象建立精确的数学模型，它所需要的先验知识是过程的机理模型。美国麻省理工学院的 Beard 于 1971 年提出了使用“解析冗余”代替传统的软件冗余，用于容错控制系统的设计，这为基于解析模型的故障诊断方法奠定了基础。

典型的基于解析模型的过程监测方法需要依据过程机理建立输入-输出之间的数学模型，并且利用观测器或者变量间的解析冗余关系进行残差序列构造，进一步采取措施进行故障信息的增强和非故障信息（如模型的随机干扰等）的抑制，最终分析残差序列以实现故障检测、故障诊断、故障识别及故障分离。基于解析模型的方法中常见的有参数估计方法、基于观测器的状态估计方法以及等价空间（Parity Space）方法等。基于解析模型的过程监测方法示意图如图 1.4 所示。

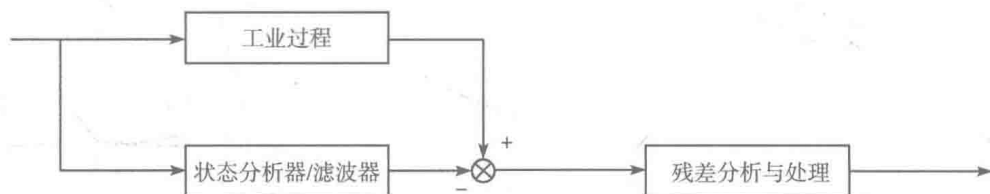


图 1.4 基于解析模型的过程监测方法

Fig. 1.4 Process monitoring method based on analytical model

如果过程故障与模型参数的变化相联系，而且能够得到适当的数学模型，那么参数估计方法就是比较适合的方法。通过标准参数估计可以得到模型参数而不需测量，接着根据基本原理建立模型，该模型能让过程中有物理意义的参数和模型参数直接相关。如果过程故障与传感器、执行器或者不可测的状态变量的变化相关时，那么基于观测器的方法就比较适合，这种方法比较适用于检测和分离加性故障。为了使状态空间方程的状态具有物理意义，就需要得到从基本原理导出的精确的数学模型。不可测状态可以利用伦博格 (Luenberger) 观测器或者卡尔曼滤波器对可测的输入输出变量进行重构而得出。

(2) 基于定性模型的方法

基于定性模型的方法又称为基于知识的方法，基于知识的方法不需要精确的过程数学模型，而是通过抽象递阶知识树 (Abstraction Hierarchy of Process Knowledge)、因果关系模型 (如故障树和符号定向图) 和专家系统等模型，将过程中各单元之间的连接关系、故障传播模式等过程知识定性地描述出来。在系统故障出现后，通过推理、演绎和模式识别完成故障的定位和诊断工作。这些技术都是基于定性模型的，它们可以通过系统的因果模型、专家知识和系统的详细描述或者故障症状举例来获得，比较适合非线性系统和复杂的大型系统。

一般来说，基于知识的方法适合于有大量生产经验和专家知识可利用的场合，其诊断能力较好，但正是由于其对知识的依赖性，导致该类方法的通用性比较差。不过，随着人工智能和知识表示等领域的发展，基于知识的过程监测方法有可能在不久的将来取得较大的进展。

(3) 基于数据驱动的方法

基于数据驱动的过程监测方法的研究与应用起始于 20 世纪 90 年代。一方面，集散控制系统 (Distributed Control System)、智能化仪表以及现场总线技术的发展带动了其在工业过程中的广泛应用。大量包含过程运行状态信息的过程数据被采集并且存储，但这些数据并没有得到有效的利用，“数据丰富，信息匮乏”成为亟待解决的问题之一。另一方面，随着计算机技术和数据库技术在 20 世纪 90 年代的迅猛发展，可靠的数据存储技术和廉价的计算资源给工业数据分析提供了可能性。数据挖掘相关算法和理论也不断涌现。

工业界逐渐意识到,在当今市场竞争日趋激烈、环境保护要求日益严格的社会背景下,工业企业必须降低成本、提高企业竞争力,为此必须从现有的过程数据中提取有用的信息,使之对生产安全与产品质量控制产生指导和预测作用。因此,工业过程性能检测的研究已经成为过程控制领域研究的热点之一。近年来,美国、西欧等发达国家已投入大量的人力和物力,关注该领域的研究进展,期望通过生产数据分析来揭示、反映过程的本质,为提高产品质量提供有用信息,把数据资源的拥有优势转化为生产效益和产品质量优势。

和基于知识的方法相同,基于数据驱动的过程监测方法同样不需要精确的过程数学模型,而是以工业过程中采集的海量数据为基础,通过多种方法将高维数据进行降维处理成为低维数据,从中提取出有用信息,对生产过程产生指导作用。

按数据分析的方法不同,基于数据驱动的方法又可以分为定性和定量两种。定性的基于数据驱动的过程监测方法通常指的是数据的动态趋势分析,是现代时间序列分析方法在过程监测领域的扩展。定量的基于数据驱动的过程监测方法又可以分为非统计方法和统计方法两种:非统计方法包括频谱分析、神经网络方法和小波变换等;统计方法指的是多变量统计过程监控(Multivariate Statistical Process Monitoring, MSPM),最常用的有主元分析(Principal Component Analysis, PCA)、主元回归(Principal Component Regression, PCR)、偏最小二乘(Partial Least Square, PLS)、正则相关分析(Canonical Correlation Analysis, CCA)以及Fisher判别式分析(Fisher Discriminant Analysis, FDA)等。

基于数据驱动的方法只需要利用过程数据,从而得到数据的特征向量,该特征向量保留了原始数据的特征信息,去除了冗余信息,非常适合于工业过程的故障检测与诊断,具有较强的通用性。然而,MacGregor在其研究中指出了运用该类方法在处理过程数据时应注意以下几个方面。

① 数据维数(Data dimension)。由于分布式控制系统(Distributed Control System, DCS)和计算机技术的普遍应用,大量高维的过程数据被实时采集并存储,但在实际生产过程中,只对少数的几个关键变量进行监控,未被监控的数据中所包含的有用信息则往往被丢弃。

② 数据质量(Data quality)。在实际工业过程中,观测变量通常会受到各类噪声源的影响,还有一些情况,如传感器故障,可能导致数据丢失。噪声源影响和数据丢失对过程数据的影响会导致提取信息的不可靠,并且信息难以解释。

③ 数据共线性(Data collinearity)。工业过程包含了众多的变量,但这并不意味着过程本质上是高维的,与此相反,大多数工业过程的观测变量之间都是广泛地线性相关的,也就是说,这些众多的变量可以由更少维数的数据来描

述。而这使得传统的统计方法难以奏效,因为它们假定变量之间是相互独立的,即用于过程监测时效果不佳。

④ 数据非线性。现今的工业过程往往具有非线性的属性,所以不适宜用线性函数来描述变量之间的关系,以防止破坏监测系统的监测效果,因此,必须考虑过程的非线性特性。

⑤ 数据的时变特性及过程的多工况性。由于设备状况、原料性质、市场需求等情况的不同,工业生产过程通常在多个稳态操作点下运行,而且不同稳态操作点下的生产负荷的差异很大。统计过程监控系统应该具备区分操作模式改变与过程故障的能力。

⑥ 数据的动态特性。大多数工业过程的观测变量之间具有自相关性,而且测量的过程数据又具有时序相关性,也就是说,当前时刻的测量值与先前的测量值不独立。Negiz 等人的研究成果表明,数据的动态特性对统计量的统计特性有很大影响,进而对过程的监测效果产生重大影响。

⑦ 数据的非正态特性。传统的统计过程监测方法中都假设观测变量是相互独立并且满足正态分布的,但在实际工业生产过程中,观测数据的分布情况往往不满足正态分布,因此,采用传统的过程监测方法将会对监测效果产生较大影响,降低系统可靠性。

1.1.3 多元统计过程监测国内外研究现状

一般说来,统计过程监测大致可以分为单变量统计过程监测和多变量统计过程监测。单变量统计过程监测包括 Shewhart 控制图、累积和 (Cumulative Sum, CUSUM) 图、移动平均 (Moving Average, MA) 图以及指数加权移动平均 (Exponential Weighted Moving Average, EWMA) 等。但随着工业过程规模的逐渐扩大及测量技术的发展,单变量过程监测方法越来越显示出其局限性,因此,各类单变量控制图得到了扩展,如多元累积和 (Multivariate CUSUM, MCUSUM) 图、多元指数加权移动平均 (Multivariate EWMA, MEWMA) 图等。为了更加有效地监测出过程存在的异常状态,由 Kresta 等、Piovoso 等及 MacGregor 等提出的多变量统计过程监测方法的故障检测和诊断方法得到了迅猛发展。

多元统计过程监测 (Multivariate Statistical Process Monitoring, MSPM) 又被称为多元统计过程控制 (Multivariate Statistical Process Control, MSPC), 其采用多元投影降维的方法处理过程变量的观测数据,实现统计过程监测。其基本思想是将大量测量变量张成的高维空间投影到维数相对较少的模型空间上,这样得到的新的特征变量不仅能够解决变量间相关性严重、原始数据空间维数过大以及众多的未知干扰等问题,并且不损失原测量数据有价值的信息。原测量变

量的线性组合被称为主元变量或隐变量(Latent Variable), 主元变量或隐变量张成模型空间, 投影算法实现了用较少维数的模型空间来描述整个过程的主要特征的目的。

传统的 MSPM 方法包括基于 PCA 和 PLS 的过程监测方法, 利用这些方法进行过程监测时, 对过程变量做出如下假设: 各个过程变量均服从高斯分布; 各变量之间的关系是线性的; 过程运行在单一稳定操作模式下, 参数不会随着时间变化而改变; 数据采样服从独立性条件。目前多变量统计方法已在北美和欧洲的工业生产中得到了广泛应用, 与多变量统计监测相关的商业软件业逐渐增多。常见的方法包括主元分析、主元回归、Fisher 判别分析、偏最小二乘、典型相关分析、独立元分析等。其中, 最基本的方法就是主元分析方法。

主元分析(PCA)方法是 Pearson 于 1901 年在研究如何对空间中的点进行直线和平面的最佳拟合时最早提出的。有学者认为, 在系统响应和方差分析方面, 主元分析比系统建模具有更大的价值。随后, Hotelling 对主元分析方法进行了改进, 由此产生了目前被人们广泛应用的主元分析方法。

作为一种典型的 MSPM 方法, 主元分析方法同样利用投影技术, 利用少量的独立变量表示大量相关变量的动态信息, 从而对原始的高维数据进行降维处理。主元分析方法建立了主元子空间和残差子空间这两个空间的概念, 将过程监测数据向量投影到这两个正交的子空间上, 并且分别在这两个子空间上建立相应的统计量, 利用这些统计量对数据进行假设检验以对过程运行状态进行判断。

在过去的几十年里, 众多学者对主元分析方法进行了大量的研究并不断改进算法, 目前对主元分析方法的改进已经有了长足的发展, 比如非线性主元分析(Nonlinear Principal Component Analysis, NPCA)、多尺度主元分析(Multi-scale Principal Component Analysis, MPCA)、动态主元分析(Dynamic Principal Component Analysis, DPCA)、核主元分析(Kernel Principal Component Analysis, KPCA)以及针对间歇过程的多向主元分析(Multi-way Principal Component Analysis)等。

在这些方法中, KPCA 是解决过程变量的非线性问题的一种非常好的方法。它的基本思想是: 首先将低维输入空间中各变量之间的非线性关系通过非线性映射映射到高维特征空间中, 然后在高维特征空间中进行线性分析。其中非线性映射的具体形式并不要求取, 只需得到特征空间中的内积即可, 而内积可以由一维的非线性函数表示。KPCA 方法本质上是在高维特征空间实现线性 PCA, 因此很容易理解和应用用于过程监控。一些学者的研究已经显示了 KPCA 方法在故障检测方面的作用。