

基于数据驱动的间歇过程建模及故障监测

—— 质量控制角度

JIYU SHUJU QUDONG DE JIANXIE GUOCHENG JIANMO JI GUZHANG JIANCE
ZHILIANG KONGZHI JIAODU

常鹏 著



知识产权出版社

全国百佳图书出版单位

基于数据驱动的间歇过程建模 及故障监测：质量控制角度

常 鹏 著



知识产权出版社

全国百佳图书出版单位

图书在版编目 (CIP) 数据

基于数据驱动的间歇过程建模及故障监测：质量控制角度 / 常鹏著. —北京：知识产权出版社，2019.8

ISBN 978-7-5130-6069-1

I . ①基… II . ①常… ①过程统计—统计模型—研究 IV . ① O211.64

中国版本图书馆 CIP 数据核字 (2019) 第 012933 号

内容提要

随着先进的传感器及其对应测量技巧的应用及发展，在不断的生产过程中累积下了大量的数据，这样的数据不免会包含着没有被挖掘出来的信息。基于数据驱动的多元统计算法可以利用这些已有的数据对相应的过程实施过程监控、故障检测和诊断、质量预测及软测量，对实现相关工业过程的过程效果评估、安全生产、问题原因分析、过程优化以及在线控制有着不小的帮助。基于间歇工业生产过程对过程在线质量预测的需求，本书对间歇过程本身特性，包括生产规律及数据存在形式进行了分析，利用多元统计算法的优势，提出了一系列针对间歇过程的质量预测方法，为间歇过程的过程优化及安全生产尽了一份力量。

责任编辑：李娟

责任印制：孙婷婷

基于数据驱动的间歇过程建模及故障监测：质量控制角度 常鹏著

出版发行：知识产权出版社有限责任公司

网 址：<http://www.ipph.cn>

电 话：010-82004826

<http://www.laichushu.com>

社 址：北京市海淀区气象路 50 号院

邮 编：100081

责编电话：010-82000860 转 8689

责编邮箱：laichushu@cnipr.com

发行电话：010-82000860 转 8101

发行传真：010-82000893

印 刷：北京中献拓方科技发展有限公司

经 销：各大网上书店、新华书店及相关专业书店

开 本：787mm × 1092mm 1/16

印 张：13.5

版 次：2019 年 8 月第 1 版

印 次：2019 年 8 月第 1 次印刷

字 数：180 千字

定 价：68.00 元

ISBN 978-7-5130-6069-1

出版版权专有 侵权必究

如有印装质量问题，本社负责调换。



作者简介

常鹏，1981年出生，讲师。

主要从事间歇过程统计建模及监测等方面的研究。目前，作为主要参与者，参与国家自然科学基金项目1项、高等学校博士学科点专项科研基金1项。近三年，发表学术论文8篇，被SCI、EI收录5篇；获得国家发明专利1项。

责任编辑：李 娟

封面设计：陈 瑶

前 言

作为第二产业其中的一个子项，间歇生产过程是现代社会大生产当中比较有代表性的生产形式之一，因其具有比较好的灵活性、较高的产物附加值及较悠久的发展历史，被应用在诸如制药、食品生产、化工材料制备等领域。然而正因其具有的与连续生产过程所不同的特性，也会随之而来具有比较复杂的生产机理，多样的生产及工况以及多变的生产状态。

随着先进的传感器及其对应测量技巧的应用及发展，在不断的生产过程中累积下了大量的数据，这样的数据中不免会包含着没有被挖掘出来的信息。基于数据驱动的多元统计算法可以利用这些已有的数据对相应的过程实施过程监控、故障检测和诊断、质量预测及软测量的功能，对实现相关工业过程的过程效果评估、安全生产、问题原因分析、过程优化及在线控制有着不小的帮助。基于间歇工业生产过程对过程在线质量预测的需求，本书对间歇过程本身特性，包括生产规律及数据存在形式进行了分析，利用多元统计算法的优势，提出了一系列针对间歇过程的质量预测方法，为间歇过程的过程优化及安全生产尽了一份力量。主要研究内容如下：

(1) 对传统的线性多向 PLS (MPLS) 方法进行深入分析和改进研究。分析了基于不同展开方式的传统 MPLS 方法实际应用中各自存在的优缺点, 在此基础上, 提出一种改进的 MPLS 方法, 用于间歇过程的故障监控、诊断及质量预测。首先该方法结合传统展开方式的优势, 把变量展开与批次展开相结合, 提高监控预测性能; 其次, 采用随时间更新主元协方差代替固定主元协方差, 改善 T2 监控图的监控性能; 最后给出一种随时间变化的贡献图计算方法。仿真实验结果表明, 改进方法确实能改善传统 MPLS 的监控性能及预测性能, 且具有一定的故障识别能力。

(2) 提出一种二维搜索相似样本的 JITL-MPLS 在线局部建模监测方法。分析了即时学习 (Just-in-Time Learning, JITL) 策略对发酵过程实行监测过程中存在的优缺点, 在此基础上, 提出一种兼顾采样时刻与相似度的即时学习策略。针对发酵过程批次内和批次间的动态性问题, 将采样时刻引入到相似样本搜索过程当中。利用过程的相似样本占比识别过程的动态性, 自适应地调整时间窗的窗长, 再结合累计相似贡献率进行建模样本选取, 构建 JITL-MPLS 模型。通过 Pensim 仿真平台产生的正常和故障数据验证了此方法的可行性与优越性。

(3) 针对间歇过程的多阶段局部模型选择问题, 提出一种基于快速的间歇过程质量预测 Just in Time Learning-Multiway Partial Least Squares (JITL-MPLS) 方法。该方法以 PLS 为主算法, 以沿变量展开的方式对间歇过程的三向数据进行展开, 采用模糊 C 均值聚类与即时学习算法对数据进行预聚类以避免数据量过多而导致的在线建模运算负荷的问题。考虑到部分间歇过程存在的潜在的大周期特性, 在在线训练模型构建阶段保持全部历史数据为候选数据的前提下提高了在线质量预测过程的实时性。经对青霉素仿真平台和大肠杆菌工业过程的应用, 相较于一般类似算法, 所提出的算法能够在相同情况下提高系统的运算

速度，同时能够保持算法对新参数较低的敏感度以提高算法自身的鲁棒性，维持算法的预测精度在相同的等级。

(4) 针对 MPLS 算法特别是局部 MPLS 算法在故障情况下缺乏样本重选机制的问题，提出了一种基于局部样本重选的过程故障下 MPLS 质量预测方法。如若在轻微故障下继续进行生产，传统局部算法的相似度量机制在遇到故障导致的变量偏移的情况时容易产生在线训练样本选择的误差从而导致模型精度的下降，所提出的方法考虑到这个问题，以过程监控贡献图幅值超限程度为判断标准、与离线阶段选择出的稳定变量为动态时间规整指标对历史数据进行建模数据重选，建立新的局部预测模型对过程进行报警的同时进行质量预测。通过对青霉素仿真平台及大肠杆菌发酵过程的测试说明相对于传统方法，所提出的算法能够一定程度上保持对过程在非严重故障下的质量预测精度。

(5) 对局部即时学习方法的在线训练数据库选择方式进行了分析，考虑到间歇过程生产不同阶段数据量不同的状况，提出了针对间歇数据的局部自适应 JITL-MPLS 方法。该策略结合了间歇过程特有的批次特性，与原有的相似性度量函数协同，通过动态地限制即时学习算法在一个合理的范围之内来控制在线训练数据库构建时的候选数据的数量，对每一次采样建立的模型进行调整。通过对青霉素的实验验证了所提出的算法相较于传统方法能够提高对质量数据的预测精度。之后，通过对实验结果的分析，提出了一种新的模型预测效果评价指标，相对于预测精度，新的评价指标着重衡量算法预测效果的稳定性。

(6) 基于 MKPLS 的间歇过程监控与质量预测方法研究，改进的 MPLS 在性能上有所提高，但是这始终是线性方法，而大多数间歇过程都具有强非线性等特点，监控方面会出现大量误报漏报，且预测性能低。研究了基于核函数的核偏最小二乘 (KPLS) 方法，讨论了 KPLS 方法应用于间歇过程监控时存在的一些缺陷和问题。在此基础上，提出一种基于 FS 特征提取方法的 AT-MKPLS

方法。该方法采用 AT 方法对三维数据进行变换，保留了批次信息；对于核映射时数据量庞大无法运算的问题，采用 FS 特征提取的方法对样本进行简化，提取最简矩阵，然后采用 KPLS 方法提取过程的非线性信息；最后将该方法应用于一个数值非线性过程和间歇发酵过程，并与传统 MPLS 和改进 MPLS 方法进行比较，验证了该方法在监控及预测方面性能的有效性。

(7) 将核方法引入到即时学习策略中扩展核参数的取值范围，JITL 策略属于局部建模策略，能够在一定程度上减弱数据的非线性，但是在某些阶段非线性特征依然明显，对 JITL-MPLS 模型影响较大。本书将核方法 (Kernel Methods, KM) 引入到即时学习策略中，将非线性数据映射为线性数据，使得核空间中所建立的线性监测模型也能够有较高的精度。由于即时学习策略的相似样本以当前采样点为欧式空间的中心，而非离线建模的聚类中心，所以能够进行较好的局部线性化，从而扩展核参数的取值范围，降低了核参数的设置难度，最终取得了较好的故障监测效果。

(8) 提出一种基于残差更新判断的 JITL-MKPLS 监测方法，传统的即时学习策略属于“懒惰学习”，即每采集到一个数据将进行一次完整建模步骤，这样在监测比较稳定的阶段时会造成大量的计算浪费，影响系统响应其他操作。本书利用了多向核偏最小二乘 (Multi-way Kernel Partial Least Squares, MKPLS) 的回归特性，通过对比回归残差判断模型是否需要更新，从而减少了模型更新次数。模型更新判断机制在保证与传统的 JITL 建模精度相类似的前提下，大幅减少了计算量，扩展了本方法的应用范围。

目 录

第 1 章 绪 论	1
1.1 课题的研究背景及意义	1
1.2 发酵过程简介	2
1.3 发酵过程监测方法分类	6
1.4 基于统计性能的发酵过程监测方法研究现状	8
1.5 研究思路与内容安排	9
1.6 本章小结	11
第 2 章 基于改进的 MPLS 的间歇过程在线监控与质量预测	12
2.1 引言	12
2.2 偏最小二乘	13
2.3 改进的多向偏最小二乘方法	19
2.4 基于改进的 MPLS 的在线监控和质量预测	26
2.5 仿真研究	28
2.6 本章小结	38

第 3 章 基于二维搜索 JITL-MPLS 的发酵过程故障监测	39
3.1 引言	39
3.2 多向偏最小二乘方法	40
3.3 即时学习在线建模策略概述	48
3.4 基于二维搜索 JITL-MPLS 的故障监测	56
3.5 仿真研究	57
3.6 本章小结	66
第 4 章 基于 JITL-MKPLS 的发酵过程监测方法研究	67
4.1 引言	67
4.2 多向偏最小二乘方法	69
4.3 即时学习 (JITL) 局部模型选取方法	76
4.4 改进的局部 MPLS 质量预测策略	78
4.5 案例分析	83
4.6 本章小结	94
第 5 章 基于局部样本重选的过程故障下 MPLS 研究	96
5.1 引言	96
5.2 PLS 相关统计量简介	98
5.3 故障下的历史样本重选方法	100
5.4 案例分析	104
5.5 本章小结	113

第 6 章 针对间歇数据的局部自适应 JITL-MPLS 方法研究	114
6.1 引言	114
6.2 针对间歇过程的自适应 JITL 局部样本选择	117
6.3 案例分析	125
6.4 一种新预测效果度量指标设计：预测标准差	133
6.5 本章小结	138
第 7 章 基于 JITL-MKPLS 的发酵过程监测方法研究	140
7.1 引言	140
7.2 核空间中的故障监测方法介绍	141
7.3 核空间中的即时学习监测方法研究	145
7.4 青霉素发酵过程仿真	153
7.5 本章小结	161
第 8 章 即时学习在线建模方法的模型更新机制研究	162
8.1 引言	162
8.2 即时学习策略中模型更新机制研究	163
8.3 基于改进即时学习策略的发酵过程在线故障监测	168
8.4 仿真研究	169
8.5 本章小结	178
第 9 章 大肠杆菌发酵过程现场数据分析与讨论	180
9.1 大肠杆菌发酵过程简介	180

9.2 大肠杆菌发酵过程的监测结果及讨论·····	181
9.3 本章小结·····	192
第 10 章 总结与展望·····	193
10.1 总结·····	193
10.2 展望·····	194
参考文献·····	196

第1章 绪论

1.1 课题的研究背景及意义

随着生产规模不断扩大、产品种类逐渐增多、复杂度日益增高，人们对生产过程的安全性越发关注。发酵过程因具有高灵活性、高附加值等特点，已经成为现代流程工业的重要组成部分，广泛应用于医药、食品、香料等生产领域中。发酵过程一般持续时间较长，温度较高，压力较大，生产过程中一旦发生事故就会对工作人员、厂区环境、生产设备造成极大伤害。在现今的生产过程中不可避免的突发事件或故障的处理仍需要人工介入^[1]。当过程发生异常时，如果不能提早发现就会耽误故障处理的时间，造成经济损失，甚至引发严重的安全事故^[2,3]，所以需要一种准确的故障监测方法以及及时发现早期故障。

目前，以主成分分析（Principal Component Analysis, PCA）^[4,5]和偏最小二乘（Partial Least Square, PLS）^[6]等为主的数据驱动多元统计建模方法在发酵工业生产过程中得到了广泛的关注。这些多元统计过程监测方法可以实现将含有噪声和共线变量的高维度原始数据降维到低维度主元空间中，较好地解决了现有发酵过程数据存在的原始数据维度较高、过程变量之间存在耦合关系及生

产数据含有噪声等问题。而在众多多元统计数据驱动建模方法中，偏最小二乘凭借其独特的优势得到了广泛的关注与研究。偏最小二乘不但可对系统中的数据进行分解和提取特征，而且所提取的特征对质量变量具有很强的解释性，特别适用于普遍存在质量变量的发酵过程。同时，偏最小二乘法可以有效解决由于受历史数据样本数量的限制而导致的样本数目少于变量个数的问题。但是偏最小二乘方法属于全局静态线性建模方法，而大多数发酵过程具有非线性和动态性的特征，所以全局偏最小二乘方法会产生大量的误报和漏报现象。

本书将在线建模即时学习方法应用到发酵过程故障监测中，有效提高了在正常生产过程中故障监测的准确率，保证了工作人员采取措施排除故障的时效性，从而降低生产成本，提高产品质量。

1.2 发酵过程简介

发酵过程特性极其复杂，表现出强烈的批次变化特性和时变特性，最终导致产物质量波动较大，过程异常和故障延迟发现。一旦觉察到故障发生，一般情况下已不可逆转，势必造成原材料的浪费和设备的空转。许多情形下，发酵过程的产物效价或浓度只要能够稳定在一定范围之内，就可以认为该批次的发酵是成功的。发酵产品质量与下游工段的产品精制纯化操作息息相关，产品质量的波动会直接增加下游工段的负担和成本。因此，快速准确地监测到发酵过程运行中出现的各类异常工况并加以识别排除，已成为发酵过程控制领域的迫切需求。由于测量装置完备，我们可以方便地获取多变量的过程监测数据，因此结合有效的数据处理方法，从海量数据中挖掘出对表达过程运行状态信息有用的情报，便成为发酵过程监测的出发点。

1.2.1 发酵工业生产过程的特点

发酵通常指利用微生物在有氧或者无氧环境下，于特定设备中通过控制一定的外部条件来生产制备生物体本身、直接代谢产物或次级代谢产物的过程。目前，发酵工程在工业化生产中也得到了广泛的应用，其中主要是利用特定微生物获取特定的工业产品。相比连续生产过程，发酵生产过程具有工作点频繁、变量数值阶段性明显、操作严格、模型复杂等特点^[7,8]。

鉴于发酵过程的复杂性，需在该过程中监测发酵过程变量是否与历史正常批次一致，当发生区别时，尽快报警，人工介入处理，以保障整个生产过程的正常运转。同时可将影响发酵过程生产水平的因素，如温度、压力、pH、通风量等由传感器感知并反馈给操作人员，通过分析数据来指导改进操作过程。因此，要实现发酵过程的生产监测，一个基本的前提是了解发酵的整体特性及过程数据的特点。

1.2.2 发酵过程三维数据的预处理问题

发酵过程属于典型的间歇过程，而数据构成的三维特性^[9,10]是间歇过程所共有的特征。这是由于发酵过程的生产是在同一发酵罐中不同时段内重复性的实现，每一次生产结束时收集发酵产物，然后清空发酵罐中的废液，继而重复以往的发醇生产过程。历史生产数据含有三个维度，即生产批次、过程变量、采样时刻。由于传统多元统计模型输入为二维矩阵，而发酵过程以批次为单位进行生产，因此需要对三维原始数据进行预处理。

目前，常用的处理方法是将三维矩阵进行切片操作展开成二维矩阵^[11]，按不同方向展开成二维数据的形式后再进行统计分析。在对三维数据矩阵处理的方法中沿批次方向展开和沿变量方向展开应用最为广泛^[11]，下面将分别进行介绍。

(1) 沿批次方向展开。

沿三维矩阵生产批次的方向展开的方式最先由 Wold 提出^[12]，之后 Nomikos 和 MacGregor^[13,14] 对其进行研究并将多向主成分分析和多向偏最小二乘应用于间歇过程监测领域，此方法保留了历史数据沿批次方向的信息。如式 1-1 所示。

$$\mathbf{X} = \begin{bmatrix} x_{1,1}^1 & x_{2,1}^1 & \cdots & x_{j,1}^1 & \cdots & \cdots & x_{1,K}^1 & x_{2,K}^1 & \cdots & x_{j,K}^1 \\ x_{1,1}^2 & x_{2,1}^2 & \cdots & x_{j,1}^2 & \cdots & \cdots & x_{1,K}^2 & x_{2,K}^2 & \cdots & x_{j,K}^2 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \ddots & \vdots & \vdots & \ddots & \vdots \\ x_{1,1}^l & x_{2,1}^l & \cdots & x_{j,1}^l & \cdots & \cdots & x_{1,K}^l & x_{2,K}^l & \cdots & x_{j,K}^l \end{bmatrix} \quad (1-1)$$

(2) 沿变量方向展开。

Wold 提出沿三维历史数据变量方向将矩阵展开的多向主成分分析方法^[15]。展开后矩阵的每一行代表一个时刻的采样点，变量为该采样点的维度。如式 1-2 所示。

$$\mathbf{X} = \begin{bmatrix} x_{1,1}^1 & x_{1,2}^1 & \cdots & x_{1,J}^1 \\ x_{2,1}^2 & x_{2,2}^2 & \cdots & x_{2,J}^2 \\ \vdots & \vdots & \ddots & \vdots \\ x_{1,1}^l & x_{K,2}^l & \cdots & x_{K,J}^l \\ \vdots & \vdots & \ddots & \vdots \\ x_{K,1}^1 & x_{K,2}^1 & \cdots & x_{K,J}^1 \\ x_{K,1}^2 & x_{K,2}^2 & \cdots & x_{K,J}^2 \\ \vdots & \vdots & \ddots & \vdots \\ x_{K,1}^l & x_{K,2}^l & \cdots & x_{K,J}^l \end{bmatrix} \quad (1-2)$$