



间歇过程统计建模 及故障监测研究

针对数据多阶段特性

常 鹏◎著

非外借



知识产权出版社

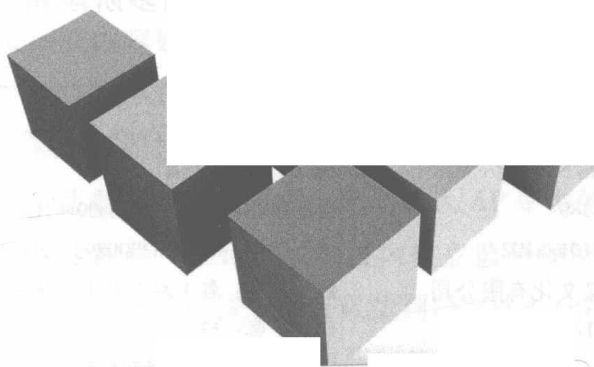
全国百佳图书出版单位

本书获得国家自然科学基金（61174109、61640312）、教育部数字社区、北京市轨道交通实验室和计算智能与智能系统北京重点实验室的资助

间歇过程统计建模 及故障监测研究

针对数据多阶段特性

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

间歇过程统计建模及故障监测研究: 针对数据多阶段特性/常鹏著.

—北京: 知识产权出版社, 2019. 5

ISBN 978 - 7 - 5130 - 6229 - 9

I. ①间… II. ①常… III. ①微生物—发酵—过程控制—研究

IV. ①TQ920. 6

中国版本图书馆 CIP 数据核字 (2019) 第 080026 号

内容提要

间歇过程是生物制药、精细化工和食品饮料行业中的主要生产方式, 但是也因其间歇式的特点, 存在着周期性批量生产、物料状态和操作参数呈现动态性、工艺控制要求高等特点。发酵过程是一种典型的间歇过程, 发酵过程关乎经济发展和人民生活水平的提高, 生物制药是国务院确立的七大战略性新兴产业之一, 在京津冀一体化中将起到重要的支撑作用。本书围绕生物发酵过程的批次不等长特性、动态特性和多阶段特性, 研究以往方法在进行监测时存在的问题, 通过建立高效高精度过程监测模型, 降低监测的误报率和漏报率, 保障运行安全, 做到及时捕捉发酵过程中各检测变量的变化, 若发现监测故障, 及时通知工作人员, 工作人员通过调整发酵环境或暂停生产, 尽可能提高产物质量、稳定生产或者减少损失, 进而减少能源消耗和资源浪费。研究成果一旦获得推广, 会极大地提高发酵过程生产的安全性, 减少事故的发生和资源的浪费, 创造较大的经济效益和社会效益。

责任编辑: 张水华

责任校对: 谷洋

封面设计: 邵建文 马倬麟

责任印制: 孙婷婷

间歇过程统计建模及故障监测研究——针对数据多阶段特性

常鹏著

出版发行: 知识产权出版社有限责任公司

社址: 北京市海淀区气象路50号院

责编电话: 010-82000860 转 8389

发行电话: 010-82000860 转 8101/8102

印刷: 北京九州迅驰传媒文化有限公司

开本: 720mm×1000mm 1/16

版次: 2019年5月第1版

字数: 280千字

ISBN 978 - 7 - 5130 - 6229 - 9

网 址: <http://www.ipph.cn>

邮 编: 100081

责编邮箱: miss.shuihua99@163.com

发行传真: 010-82000893/82005070/82000270

经 销: 各大网上书店、新华书店及相关专业书店

印 张: 17

印 次: 2019年5月第1次印刷

定 价: 69.00元

出版权专有 侵权必究

如有印装质量问题, 本社负责调换。

摘 要

间歇过程是生物制药、精细化工和食品饮料行业中的主要生产方式，但是因其间歇式的特点，存在着周期性批量生产、物料状态和操作参数呈现动态性、工艺控制要求高等特点。发酵过程是一种典型的间歇过程，发酵过程关乎经济发展和人民生活水平的提高，生物制药是国务院确立的七大战略性新兴产业之一，在京津冀一体化中将起到重要的支撑作用。本书围绕生物发酵过程的批次不等长特性、动态特性和多阶段特性，研究以往方法在进行监测时存在的问题，通过建立高效高精度过程监测模型，降低监测的误报率和漏报率，保障运行安全，做到及时捕捉发酵过程中各检测变量的变化，若发现监测故障，及时通知工作人员，工作人员通过调整发酵环境或暂停生产，尽可能地提高产物质量、稳定生产或者减少损失，进而减少能源消耗和资源浪费。研究成果一旦获得推广，会极大地提高发酵生产过程的安全性，减少事故的发生和资源的浪费，创造较大的经济效益和社会效益。

本书的主要研究内容如下：

(1) 研究一种基于 AP 聚类的阶段划分方法。

针对间歇过程的多阶段特性，采用 AP 聚类算法，此算法在进行过程阶段划分时无需过程的先验知识，通过将 S 准则引入 AP 聚类的迭代过程中，从而达到精确阶段划分的效果。在每个子阶段内建立监测模型更符合实际操作进程或过程的机理特性，同时阶段划分可以达到局部线性化的效果。

(2) 提出一种多变量自回归主元分析 (MAR-PCA) 算法。

间歇生产过程数据，由于系统本身存在时滞特性、闭环控制和扰动，大多数过程变量都呈现出动态特性，即不同时刻的采样之间时序相关，此时如果依然采用传统主元分析算法，那么得到的主元得分会时序自相关，甚至各主元间互相关，进一步造成故障的误报率增加。

(3) 研究基于信息传递的采样点阶段归属判断。

研究故障监测时新时刻采样点的最佳模型选择问题,引入信息度传递实现实时采样点的阶段归属判断,解决阶段不等长批次的最佳模型选择问题,做到新时刻采样点能落入对应的实际操作阶段,从而选取相对应阶段的监测模型实现实时样本点的监测。

(4) 提出子阶段自回归主元分析发酵过程故障监测方法。

将单变量过程的时序分析方法拓展到多变量情形,区别具有强动态性的过渡阶段及平稳的稳定阶段,对其分别建立自回归主元分析(Auto Regression - Principal Component Analysis, AR - PCA)模型以及多向主元分析(Multiway Principal Component Analysis, MPCA)模型,以消除过渡阶段的动态性,有效降低过程监测的误报率和漏报率。

(5) 大肠杆菌发酵现场试验研究。

将本书研究内容应用于实际生产过程,借助于大肠杆菌发酵实验检验所用研究方法的合理性及有效性。结果表明,本书所提出的方法较传统方法可有效降低故障的误报率和漏报率,有着更加可靠的监测性能,可以很好地指导操作人员及时发现并有效排除故障。

关键词: 发酵过程; 批次加权; 阶段归属; AR - PCA; 故障监测

Abstract

The fermentation process is the most promising branch of biological field, Bio-fermentation technology has played an increasingly important role in modern food, medicine and other high value-added processing industry. Bio fermentation industry will become one of the leading pillars of China's national economic development in the next few years. But the development of technology is a double-edged sword, many safety problems also highlight one by one, while the vigorous development of fermentation technology bring considerable changes for our production and life, this forcing people to pay more and more attention to the safety and reliability of the production process. Therefore, in order to improve the maintainability and security of fermentation process, and improve the quality of products, the production process is in urgent need for fault monitoring, capturing the change of each detection variables immediately, feeding the abnormal situation up the operator, making disposal timely, guaranteeing the continuity, stability and safety of fermentation process.

This topic analyzed multi stage characteristics and dynamic characteristics of fermentation process deeply, and for the defects of traditional methods for process monitoring, to study a novel online monitoring algorithm for fermentation process, to reduce the leaking alarm rate and nuisance alarm rate of process monitoring.

(1) Implementation of the batch weighted soft classifying based on Affinity Propagation Clustering.

For the multiphase property and slow time-varying characteristics inherent in the fermentation process, analyzing the relationship between stable phase and transition process deeply, on the basis of AP realize hard division for stage based on single batch, fusing multiple batches data by introducing Inverse Distance Weighted,

avoiding the limitation of a single batch as the input of AP cannot represent the stage characteristics of the entire production process, to achieve a reasonable division of the transition phase.

(2) Research on the stage attribution of real time sampling points based on information transmission.

Study on the selection of the optimal model of the new time sampling point for online monitoring, information transmission is introduced to determine the stage attribution of real time sampling points, and to solve the problem of optimal model selection for unequal length batch, realizing the new time sampling points can fall into the corresponding actual operation stage, and to select monitoring model corresponding to the stage to realize the monitoring of real-time sampling point.

(3) Extraction of sub-phase Auto Regression-Principal Component Analysis fault monitoring method for fermentation process.

The time series of single variable process analysis method is extended to the multivariate case, distinguishing the stable stage and transition process with strong dynamic property. After that AR-PCA and MPCA model was established for the transition phase and the stable phase respectively, while eliminating the dynamic of transition phase, can effectively reduce leaking alarm and false alarm.

(4) Field experiment study on the fermentation of Escherichia coli.

Applying the proposed method in this book to the actual production process, and to validate the rationality and validity of this method with the help of Escherichia coli fermentation experiment. The result indicated that this method can effectively reduce the leaking alarms and nuisance alarms than the traditional method, having more reliable monitoring performance, and can be a good practice guide for the operator to find and remedy fault in a timely and effective manner.

Keywords: Fermentation Process; Batch Weighted; Stage Attribution; AR-PCA; Online Monitoring

目 录

摘要	
Abstract	
第 1 章 绪论	1
1.1 本书研究背景及意义	1
1.2 发酵过程简介及特征分析	3
1.3 发酵过程的统计过程监测	10
1.4 本书的研究内容及章节安排	17
第 2 章 基于多阶段 MPCA 的间歇过程监测研究	21
2.1 引言	21
2.2 主元分析 (PCA)	22
2.3 多向主元分析 (MPCA)	25
2.4 基于改进 AP 聚类的间歇过程阶段划分方法研究	27
2.5 仿真验证与结果分析	33
2.6 本章小结	42
第 3 章 基于 MAR-PCA 的间歇过程监测研究	43
3.1 引言	43
3.2 动态性对过程监测的影响	44
3.3 基于 MAR-PCA 的间歇过程监测	45
3.4 MAR-PCA 算法步骤	48
3.5 数值实例仿真研究	50
3.6 本章小结	55
第 4 章 多阶段 MAR-PCA 在间歇过程监测中的应用研究	56
4.1 引言	56



4.2	多阶段 MAR - PCA 算法	56
4.3	基于多阶段 MAR - PCA 的间歇过程在线监测	59
4.4	仿真研究与结果分析	60
4.5	本章小结	65
第 5 章	基于仿射传播聚类的批次加权阶段软化分	66
5.1	引言	66
5.2	反距离加权	67
5.3	基于改进 AP 的阶段软化分	68
5.4	仿真研究	73
5.5	本章小结	78
第 6 章	基于信息传递的采样点阶段归属判断	80
6.1	引言	80
6.2	信息传递	81
6.3	采样点阶段归属的初步选择	83
6.4	采样点阶段归属的最终判定	84
6.5	仿真研究	86
6.6	本章小结	89
第 7 章	基于子阶段自回归主元分析的发酵过程在线监测	91
7.1	引言	91
7.2	主元分析与自回归模型	92
7.3	发酵过程子阶段监测模型的建立	95
7.4	子阶段 AR - PCA 在线监测	99
7.5	仿真研究	100
7.6	本章小结	110
第 8 章	基于 PDPSO 优化的 AP 聚类阶段划分	112
8.1	引言	112
8.2	AP 聚类算法	113
8.3	PDPSO 算法	113
8.4	基于 PDPSO 优化的 AP 聚类算法阶段划分	117
8.5	仿真研究	119

8.6 本章小结	120
第9章 基于多阶段自回归主元分析的发酵过程监测	122
9.1 引言	122
9.2 主元分析与自回归模型	123
9.3 基于 AR 残差的 MPCA 模型	126
9.4 多阶段 AR - PCA 监测	127
9.5 多阶段 AR - PCA 监测模型的建立	128
9.6 仿真研究	129
9.7 大肠杆菌发酵现场实验与结果分析	141
9.8 本章小结	150
第10章 基于 KPCA - PCA 的多阶段间歇过程监控策略	151
10.1 引言	151
10.2 数据集的相似度理论	152
10.3 多阶段 KPCA - PCA 监控策略	156
10.4 仿真验证与应用研究	165
10.5 本章小结	181
第11章 基于 GMM - DPCA 的非高斯过程故障监控	182
11.1 引言	182
11.2 高斯混合模型 (GMM) 理论	183
11.3 基于 GMM - DPCA 的故障监控策略	185
11.4 基于 GMM - DPCA 监控策略的离线建模和新批次监控	190
11.5 应用研究	191
11.6 本章小结	200
第12章 基于 KECA 的间歇过程多阶段监测方法研究	201
12.1 引言	201
12.2 多阶段过程监测策略	203
12.3 构建多阶段的监测模型	206
12.4 算法验证	209
12.5 本章小结	219

第 13 章 间歇过程子阶段非高斯监测方法研究 220

13.1 引言 220

13.2 基于多阶段 KEICA 的间歇过程监测 222

13.3 算法验证 228

13.4 本章小结 241

第 14 章 总结与展望 242

14.1 总结 242

14.2 展望 244

参考文献 248

第1章 绪论

1.1 本书研究背景及意义

1.1.1 本书研究背景

《国民经济和社会发展第十二个五年规划纲要》明确指出，把生物产业作为重点培育的战略新兴产业之一。发酵产业是生物工程领域最具潜力的分支，生物发酵技术在现代食品、医药等高附加值产业中所发挥的作用越来越大。2014年7月，全国发酵工程技术工作委员会工作会议在京召开，会议强调了行业创新的重要性，要求把发酵产业发展的重点放在加强行业创新能力建设、推动行业淘汰落后产能以及节能减排上。中国轻工业联合会科技环保部主任于学军也指出，国家关于生物发酵行业技术改造的支持重点是安全和节能减排。经过“十二五”的建设，生物发酵产业正受到政府的高度重视，在未来几年内生物发酵产业将成为我国国民经济发展的主要支柱之一。

近年来，我国发酵产业技术创新能力不断增强，很多高尖端的研究成果和具有自主知识产权的发酵新技术产品层出不穷。中国生物发酵产业协会理事长石维忱指出，企业研发投入平均约占销售收入的4.3%，最高可达10%，明显高于其他食品企业的研发投入水准。另外，生物发酵企业取得的专利成果数目也在稳步增长，从而引领行业技术装备不断改善，产业规模持续壮大。这一结论在中国轻工业联合会副会长钱桂敬的数据介绍中得到印证：我国生物发酵产业是全球最大的板块，已逐步成为一个完备的现代生产体系，谷氨酸、赖氨酸、柠檬酸、麦芽糖浆、黄原胶、结晶葡萄糖、低聚异麦芽糖、葡萄糖酸钠、木糖醇、衣康酸等产品的产量和单一企业规模已达世界之最。与此同时，“京津冀”作为我国规划的主要经济区之一，被普遍认为是我国经济

发展的“第三极”，是城市布局最聚集、综合实力最强的地区之一，鉴于京津冀的重要经济战略地位和经济区建设的需求，在雄厚的技术实力背景下，涉及医药、食品等关乎经济发展和人民生活水平的发酵产业在此得到了长足发展，尤其是基因工程技术在发酵产业中的应用，使得制药产业等生命科学的生产方式发生了重大变革。

但技术的发展是一柄双刃剑，发酵产业的繁荣在为我们的生产生活带来可观变化的同时，其存在的诸多安全性问题也逐一凸显，这迫使人们对生产过程的安全性和可靠性越来越重视。由于现有的发酵过程监测方式还不能及时发现故障，轻者导致产率低下、产品质量降低，重者造成生产失败，浪费大量原料，造成严重经济损失。更甚者，如果有不合格产品流向市场，将会对社会造成不可估量的危害。因此，为了提高发酵过程的安全性和可维护性，同时提高产物质量，急切需要对发酵过程进行异常监测，做到及时捕捉发酵过程各检测变量的变化，将异常情况反馈给操作人员和控制系统，以便做出及时处置，保障发酵过程的持续稳定安全运行。

1.1.2 本书研究意义

发酵过程属于典型的间歇过程，其与连续过程有着明显的区别，其中，生产产品变更与工艺操作条件的时常改变是发酵过程的正常活动方式。发酵过程不会一直处于某一稳定工作点，并且通常呈现时变特性、动态性和强非线性，其操作难度远大于连续过程。此外，产品质量极易受到原材料质量状态、设备状况、外部环境等不确定性因素的影响，难以实现在线测量。发酵过程的测量数据是三维形式（时间×变量×批次），这在很大程度上增加了过程监测的难度，使得传统适用于二维数据的监测方法无法得以应用。以上所面临的问题与工业过程迫切需求之间的矛盾，极大地推动着新兴技术的产生及发展。

传感器技术以及计算机技术的发展，使得发酵过程积累了丰富的历史生产数据。这些数据中蕴藏着大量的过程信息，如加以合理分析及应用便可以极大促进发酵过程的安全稳定，提高产品质量，减少事故发生。在过去的几十年中，基于数据驱动的多元统计过程监测（Multivariate Statistic Process Monitoring, MSPM）方法在发酵过程故障监测和诊断、质量预测等领域得到了广泛的关注。相对于故障诊断，MSPM涉及的范畴更广，其更注重对整个

过程生产流程的诊断,属于质量控制的范畴。它以提高系统运行过程的可靠性及安全性、保证产品质量为主要目的,是以过程监测、故障检测、故障识别、故障诊断、故障排除以及质量预测为核心的一门新兴边缘性学科。

本书针对发酵过程监测所面临的两个典型问题展开研究:①多阶段特性的存在使得传统对整个生产过程建立单一监测模型的方法无法在整个监测过程中获得较为理想的模型误差,最终导致模型在某些时段由于无法准确描述对应阶段的数据特性而出现大量的误报警和漏报警问题;②由于发酵过程是一个慢时变的过程,其阶段之间的切换并不是瞬时完成的快速转变过程,而是跟随时间渐变完成不同阶段的过渡转换,因此过渡过程中的过程变量会表现出强烈的相关性(包括自相关性和互相关性),而传统分阶段建模方式并没有针对该问题“有的放矢”,造成了稳定阶段监测效果良好,而在过渡阶段出现连续误报和漏报的问题。针对以上两个问题,本书通过研究阶段软划分算法,以实现在无需指定聚类个数的前提下完成稳定阶段划分和过渡阶段辨识;通过研究采样点阶段归属判断方法,以实现在线监测时实时采样点可以准确选择监测模型,保证模型误差最小化;通过研究子阶段建模方法,尤其是针对过渡阶段的建模方式,以实现过渡阶段内过程变量相关性的有效去除,保证良好的监测性能。

总之,对发酵过程建立统计过程监测模型,最终目的是在保证较低误报率和漏报率的前提下,快速准确检测到发酵过程中发生的异常工况,然后实时地依据报警指示提供给操作人员一定量的信息,指导操作员有目的地及时检修,进而排除异常,保障整个过程安全稳定运行,避免不必要事故的发生,而且还可以为发酵过程的优化以及发酵产物质量的改进提供必要的辅助和指导。这一技术的研究不仅有重要的理论意义,而且有着广阔的工业应用前景,是未来工业安全持续稳定生产的重要研究方向。建立完善且经过生产验证的发酵过程统计过程监测理论,必将极大推动整个发酵工业的长足发展。

1.2 发酵过程简介及特征分析

1.2.1 发酵过程简介

发酵有时也会写作“醱酵”,其定义依使用场合的不同而有所差别。其通

通常情况下是指人们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。在认识发酵以前，人类很早就已经开始接触发酵这一类的生化反应，图 1-1 所示为古埃及人利用发酵方式制作食用面包的过程。现如今，发酵技术已经在食品、制药及工业化工等领域中得到了广泛的使用，对其基本过程的研究属于生物工程以及发酵工程的范畴，对其机理以及过程监测的研究是当下学术界的研究热点，并且一直在持续。近些年来，发酵工程又结合基因工程和计算机技术进入了一个全新的发展时期。

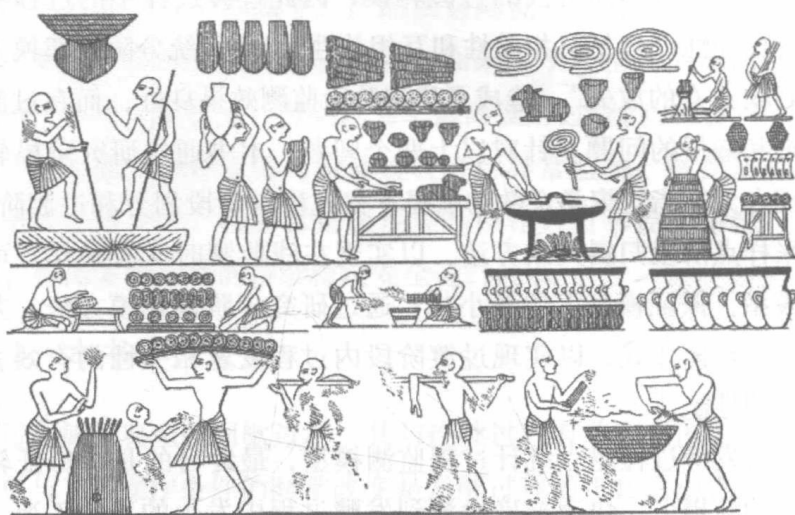


图 1-1 古埃及人制作面包的发酵过程

Fig. 1-1 The fermentation process of making bread of the ancient Egyptians

一般情况下，发酵过程是将糖源转化生成酸、气体或者醇的生物代谢过程，这一过程通常主要发生于酵母和细菌中，当然也有发生于缺氧肌肉细胞中的特殊情形，如乳酸发酵。发酵的学科范畴通常被称为酶学。图 1-2 所示为发酵过程的流程示意图。

目前，发酵工程在工业化生产中得到了广泛的应用，其中主要是利用特定微生物获取特定的工业产品。工业化发酵过程通过设定特定微生物的生长环境对其加以培养，使得该微生物被需求的特定理想代谢产物得以富集，最终通过一定的分离手段加以提取。这些代谢产物包括微生物、有机酸、氨基酸、核酸、酶制剂以及抗生素等。

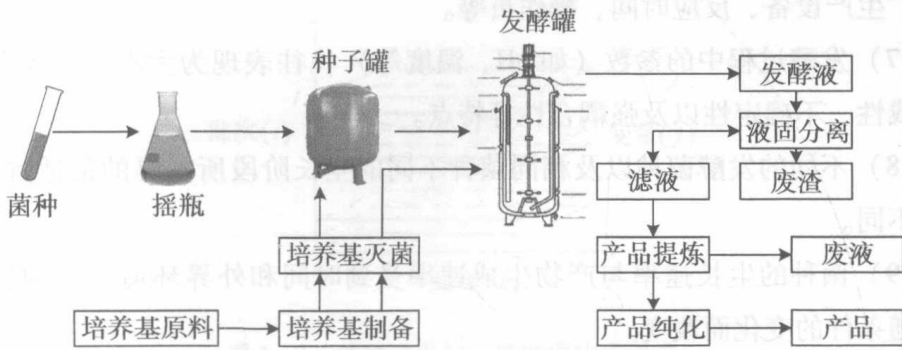


图 1-2 发酵过程的流程示意图

Fig. 1-2 The flow chart of fermentation process

不同于通常所见的化学反应过程，发酵过程有所独有的微生物学属性，主要有以下几个方面：

(1) 发酵过程是若干生物种群的生命代谢过程，该过程不可逆转。通常情况下，按照微生物的代谢和生长规律，完整的发酵过程包含四个阶段：迟滞期（即调整期）、对数生长期（即生长旺盛期）、稳定期（即平衡期）以及死亡期（即衰退期）。

(2) 发酵过程对生物体的生长状态以及外部环境的要求较为苛刻。一般情况下，要求使用“纯种”状态的微生物，所使用的培养基、发酵罐以及配套周边设备在发酵开始前必须进行严格的灭菌操作，后续的生产过程也必须严格控制在无杂菌状态下进行。

(3) 发酵过程中，发酵结果极易受到培养基成分改变的影响，甚至会导致发酵结果的不确定性，这就是发酵过程所表现出的混沌现象。

(4) 参与发酵微生物的初级和次级代谢在时间上会交织在一起，而生长期和生长期隶属于两个截然不同的生长阶段，次级代谢产物在微生物的生长繁殖速率降低乃至停止时才会开始合成，与微生物本身的生长并不同步。

(5) 相当数量的生物化学反应方式尚处于定性解释阶段，目前仍无法定量描述。

(6) 在外界培养条件适宜的情况下，菌体生长速率和产物生成速率以及菌体数量和产物量之间不具有明显的线性关系，基于一定机理所建立的产物预测模型并不能很好地匹配实际产量。这是由多方面的原因导致的，如季节、

原料、生产设备、反应时间、操作员等。

(7) 发酵过程中的参数(如 pH、温度等)往往表现为大滞后、大惯性、强非线性、不确定性以及强耦合性等特点。

(8) 不同的发酵菌种以及相同菌种不同的生长阶段所需要的最适宜发酵条件不同。

(9) 菌种的生长速率与产物生成速率受到时间和外界环境条件的影响,会伴随条件的变化而改变。

(10) 一些与产物相关联的生物参量,如菌体浓度、产物浓度、葡萄糖浓度等,难以实现在线测量。

单从控制的角度而言,发酵过程监测的目的是在发酵工艺条件不变和能耗平稳的前提下,保障整个生产过程的正常运行,以此确保发酵工业的整体经济效益。发酵过程中有许多影响发酵过程生产水平的因素,如温度、压力、pH、通风量等,这些都和发酵过程的机理有关,它们当中的大多数都可以由传感器感知并反馈给操作人员记录下来,以便加以分析来指导改进过程的操作。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以及过程数据的特点。

1.2.2 发酵过程数据的三维特性

发酵过程属于典型的间歇过程,而数据构成的三维特性是间歇过程所共有的特征。因此,发酵过程同一般性的间歇过程一样,三维结构的数据是其固有的存在。这是由于发酵过程的生产是在同一发酵罐不同的时间段内重复地实现,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。因此在记录过程生产数据的时候,伴随生产的持续我们可以按照事先设定的采样间隔在每一时刻采集到多个变量的数据,构成连续过程中常见的二维形式的数据矩阵 $X (K \times J)$, 其中, K 为每一批次过程中的采样点个数, J 为过程变量的个数。因为发酵过程的生产不可能像连续过程一样永远持续下去,所以每一个操作批次结束时我们都可以记录到这样的二维矩阵,周而复始,我们便可以累积得到发酵过程具有代表性特征的三维数据矩阵 $X (I \times J \times K)$, 其中 I 为发酵过程的生产批次数量,如图 1-3 所示。