

在线识别间歇过程特性的新方法 及其在发酵过程中的应用¹⁾

傅春生

(南京化工学院化机系 210009)

王骥程

(浙江大学工控所 杭州 310027)

摘 要

本文运用模糊集^[1]和模式识别理论,提出了一种具有自学习功能的在线识别间歇过程特性的新方法,用该方法解决一个工业发酵过程菌体生长期的识别问题,取得了良好效果。

关键词: 间歇过程,动态特性,在线识别,模糊集。

1 定 义

间歇过程是工业生产中一类非常重要的动态过程,如化学反应,生物发酵,钢铁冶炼等常采用这类过程。众所周知,这类过程对象特性的描述往往要比对连续生产过程描述困难得多。

定义 1. 能够反映一种产品的工业生产过程动态特性的所有状态轨线的全体组成过程族,用 \mathcal{F} 表示。

定义 2. \mathcal{F} 中具有某些相似特性的状态轨线的全体组成过程类,用 P 表示。

定义 3. 将一生产过程状态轨线按某些属性分成若干段,其中每一段为一个模式,用 PA 表示。

定义 4. \mathcal{F} 中具有某些相同属性的模式的全体组成模式类,用 CPA 表示

定义 5. 一生产过程状态轨线按某些属性而分成的若干个有序模式为相关模式,用 RP 表示。

如图 1 所示, 3 维欧氏空间中的状态轨线 AB , 从起点 A 到终点 B 被人为地分成三段,即三个模式 $PA_i (i = 1, 2, 3)$ 。其中 PA_1 与 PA_2 为相关模式,并称 PA_1 是 PA_2 的上游相关模式,用 $RP[PA_2]^+$ 表示; PA_2 是 PA_1 的下游相关模式,用 $RP[PA_1]^-$ 表示。

定义 6. 两相关模式的连接点叫模式相关节点,用 $RJPP$ 表示。

1) 国家“七·五”攻关课题,同时得到江苏省自然科学基金资助。
本文于 1991 年 5 月 20 日收到

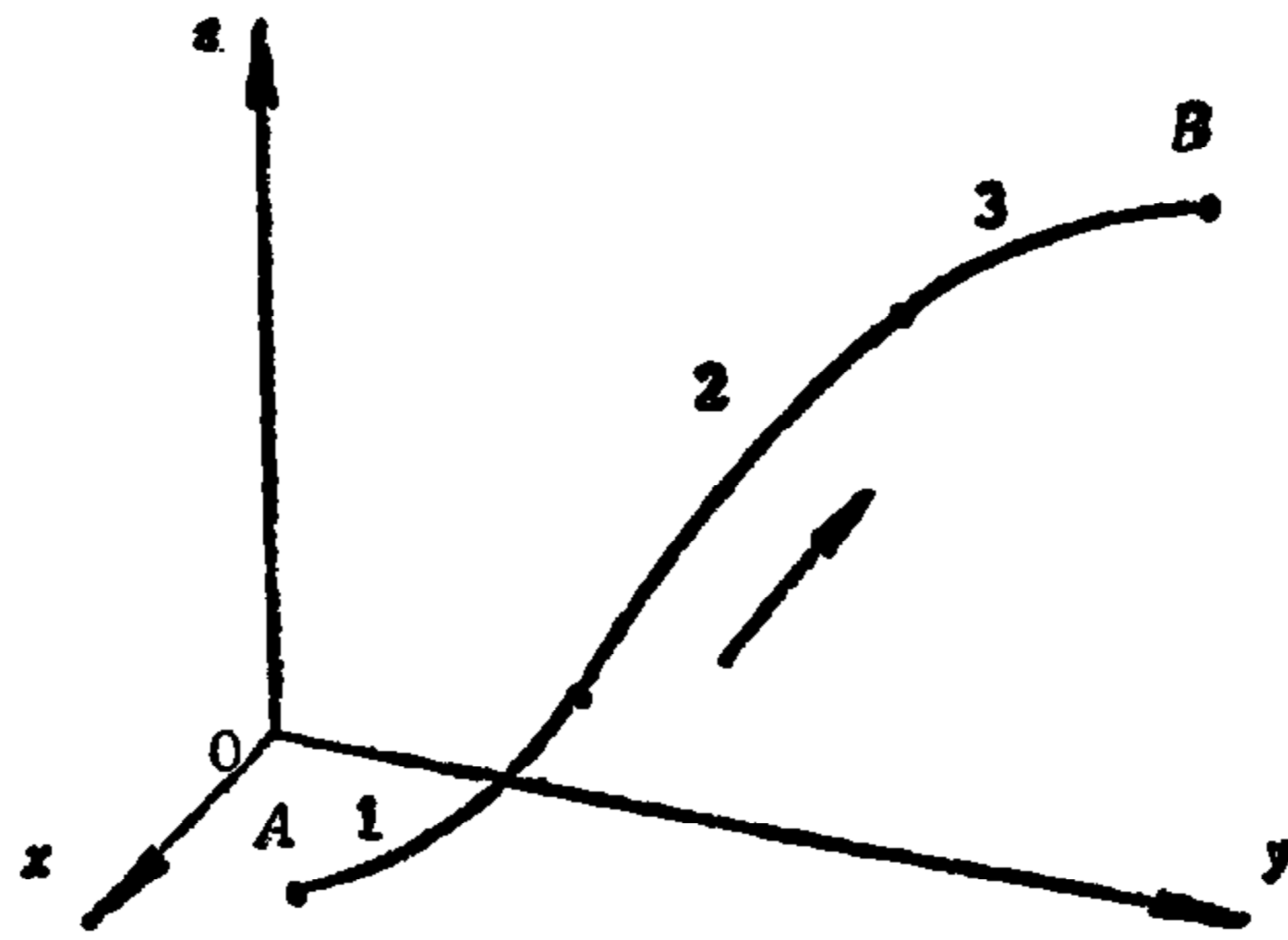


图 1 RP 和 RJPP 示意图

显然,模式相关节点是 n 维特征向量的函数。

设模式聚类中心和规一化处理后的聚类中心分别为 c_{ij} 和 \bar{c}_{ij} ($c_{ij}, \bar{c}_{ij} \in P_i, c_{ij}, \bar{c}_{ij} \in CPA^j$), 它们均是 n 维向量。

定义 7. 具有相同模式个数的生产过程两状态轨线 p 和 q 间的距离为 F ——距离, 它由 p 和 q 中各模式规一化的聚类中心差的 2 范数确定, 亦即

$$F_{p-q} = \sum_{j=1}^l \|\bar{c}_{pj} - \bar{c}_{qj}\|_2. \tag{1}$$

设向量 U_{ij} 为

$$\begin{aligned} U_{ij} &= [\bar{c}_{ij}, \bar{s}_{ij}^*, \bar{s}_{ij+1}^*] \\ &= [\bar{c}_{ij1}, \dots, \bar{c}_{ijn}, \bar{v}_{ij1}^*, \dots, \bar{v}_{ijn}^*, \bar{v}_{ij+1,1}^*, \dots, \bar{v}_{ij+1,n}^*] \\ &= [u_1, \dots, u_{3n}]. \end{aligned} \tag{2}$$

其中 \bar{s}_{ij}^* 和 \bar{s}_{ij+1}^* 分别为第 i 个过程类且属于第 j 种模式类中模式起点和终点规一化处理后的特征向量值(对于一过程状态轨线的任意中间模式, 它们均为模式相关节点的特征向量值), 且有 $u_k = \bar{c}_{ijk}; u_{k+n} = \bar{v}_{ijk}; u_{k+2n} = \bar{v}_{ij+1,k}^* (k = 1, \dots, n)$ 。

定义 8. 模式相关系数是同一模式类 CPA^j 中两模式 PA_x 和 PA_y 相关程度的度量, 用 RCP_{x-y} 表示, 其值确定为

$$RCP_{x-y} = 1 - \Sigma, \tag{3}$$

$$\Sigma = \text{Min} \left[\beta \frac{\|U_{Lj}^x - U_{Kj}^y\|_2}{\|U_{Lj}^x\|_2}, 1 \right]. \tag{4}$$

其中, β 是定常系数。

定义 9. 特征空间 Ω 中任意坐标分量 v_k 的值 v_{k0} 属于 CPA^j 的程度称为 v_{k0} 对于 CPA^j 的分量隶属度, 用 $\mu_{CPA^j}^k(v_{k0})$ 表示。 Ω 中任意一点 $s_0 = [v_{10}, \dots, v_{n0}]$ 属于 CPA^j 的程度称为 s_0 对于 CPA^j 的隶属度, 用 $\bar{\mu}_{CPA^j}(s_0)$ 表示。 $\bar{\mu}_{CPA^j}(s_0)$ 与 $\mu_{CPA^j}^k(v_{k0}) (k = 1, \dots, n)$ 之间的关系为

$$\bar{\mu}_{CPA^j}(s_0) = w \mu_{CPA^j}^c, \tag{5}$$

$$\mu_{CPA^j}^c = [\mu_{CPA^j}^1(v_{10}), \dots, \mu_{CPA^j}^n(v_{n0})], \tag{6}$$

w 是 n 维权重向量。

2 模型化方法

间歇过程动态特性模式识别的模型化步骤如下:

1) 根据过程的历史资料, 确定识别过程动态特性的特征变量, 形成 n 维特征空间 Ω .

2) 根据过程在 Ω 中的历史资料, 建立 \mathcal{F} . 并分析 \mathcal{F} 中的所有状态轨线, 确定 l 个过程类 $P_i (i = 1, \dots, l)$ 和 m 个模式类 $CPA^j (j = 1, \dots, m)$.

3) 选择一种合适的方法求算各特征变量对于各过程类中模式类的聚类中心值, 形成一个 $l \times m \times n$ 的三维聚类中心数组 $M = (c_{ijk})$ 和规一化处理后的规一化聚类中心数组 $\bar{M} = (\bar{c}_{ijk}) (i = 1, \dots, l, j = 1, \dots, m, k = 1, \dots, n)$. 同时, 将全体模式的相关节点及所有过程的起点和终点的特征向量的样本值进行合成处理, 即确定各模式类中典型模式的相关节点和各过程类中典型过程的起点和终点, 并构成 $l \times m \times n$ 的三维数值矩阵 $R = (v_{ijk}^*)$ 和其元素的规一化数值矩阵 $\bar{R} = (\bar{v}_{ijk}^*) (i = 1, \dots, l, j = 1, \dots, m, k = 1, \dots, n)$.

4) 对于 Ω 上的 P_i 且属于 CPA^j , 确定合适的映射, 即

$$\mu_{CPA^j}^k(v_k, c_{ijk}): v_k \rightarrow [0, 1],$$

$$i = 1, \dots, l, j = 1, \dots, m, k = 1, \dots, n \quad (7)$$

和权重向量 w 的各分量.

3 在线识别方法

用计算机实现间歇过程动态特性(模式)在线识别的步骤如下:

1) 根据过程的初始条件, 设定当前过程 P_0 属于 P_i . 然后确定 \mathcal{F} 中属于 P_i 的 CPA^1 和 CPA^2 的聚类中心向量 c_{i1} 和 c_{i2} (在 M 数组中). 置 $j = 1$.

2) 特征变量的数据采集和处理. 判断: 过程结束否? 否, 进行下一步; 是, 转4).

3) 计算 $\mu_{CPA^j}^k, \mu_{CPA^{j+1}}^k, \bar{\mu}_{CPA^j}^k$ 和 $\bar{\mu}_{CPA^{j+1}}^k$. 比较: 若 $\bar{\mu}_{CPA^j}^k \geq \bar{\mu}_{CPA^{j+1}}^k$, 转2); 否则, 识别结果说明该过程已由 CPA^j 过渡到 CPA^{j+1} , 即已由 PA_0^j 过渡到 $RP[PA_0^{j+1}]^-$.

4) 用 PA_0^j 段所采集和处理的所有特征变量的数据计算聚类中心 c_{0j} 和 \bar{c}_{0j} . 判断: $j + 1 \geq m$? 否, 置 $j = j + 1$, 转下一步. 是, 判断: 该过程结束? 否, 转2); 是, 转7).

5) 确定 PA_0^{j-1} 的 RJPP, 并得到 $\bar{s}_{0j-1}, \bar{s}_{0j}^+$ 和 U_{0j-1} .

6) 根据定义8计算 PA_0^{j-1} 与 PA_0^{j-1} 的 $RCP_{0 \sim i}^{j-1} (i = 1, \dots, l)$, 且在 CPA^j 中选择一个典型模式 \bar{PA}_u^j 使得 \bar{PA}_u^{j-1} 与 PA_0^{j-1} 的 $RCP_{0 \sim u}^{j-1}$ 满足下式:

$$\bar{PA}_u^j = \{\bar{PA}_i^j | \bar{PA}_u^{j-1} = RP[\bar{PA}_i^j]^+, RCP_{0 \sim u}^{j-1} = \max\{RCP_{0 \sim i}^{j-1}, \dots, RCP_{0 \sim m}^{j-1}\}\}. \quad (8)$$

7) 由 M 数组中查出 \bar{PA}_u^j 和 PA_u^{j+1} 的聚类中心向量 c_{uj} 和 c_{uj+1} , 令 $i = u$, 转2).

8) 按定义 7 计算 P_0 与 P_i 的 F —距离 $F_{0\sim i}(i = 1, \dots, l)$, 且取

$$F_{0\sim y}^{\min} = \min\{F_{0\sim 1}, \dots, F_{0\sim l}\}, 1 \leq y \leq l, \quad (9)$$

判断 $F_{0\sim y}^{\min} \leq \delta$ (常数)? 是, 转 9), 否则, 转 10).

9) 修改 P_y 所属的各典型模式特征变量的聚类中心 c_{yjk} , 规一化聚类中心 \bar{c}_{yjk} , RJPP v_{yjk}^* 和规一化 RJPP $\bar{v}_{yjk}^*(j = 1, \dots, m, k = 1, \dots, n)$, 使得修改后的值分别为

$$c_{yjk}^* = [\lambda c_{yjk} + (1 - \lambda)c_{0jk}]/2, \quad (10)$$

$$\bar{c}_{yjk}^* = [\lambda \bar{c}_{yjk} + (1 - \lambda)\bar{c}_{0jk}]/2, \quad (11)$$

$$v_{yjk}^* = [\lambda v_{yjk}^* + (1 - \lambda)v_{0jk}^*]/2, \quad (12)$$

$$\bar{v}_{yjk}^* = [\lambda \bar{v}_{yjk}^* + (1 - \lambda)\bar{v}_{0jk}^*]/2, \quad (13)$$

并将它们取代 $c_{yjk}, \bar{c}_{yjk}, v_{yjk}^*$ 和 \bar{v}_{yjk}^* 而分别存入 M, \bar{M}, R 和 \bar{R} 中. 上面式中的 $\lambda \in (0, 1)$ 为遗忘因子, 可根据需要确定其大小, 转 11).

10) 将 P_0 相应的 $c_{0jk}, \bar{c}_{0jk}, v_{0jk}^*$ 和 $\bar{v}_{0jk}^*(j = 1, \dots, m, k = 1, \dots, n)$ 作为一个新的过程类参数, 扩展并分别存入数组 M, \bar{M}, R 和 \bar{R} 中.

11) 停止.

4 工业应用实例

本节介绍运用上述方法实现工业发酵过程菌体生长期的直线识别与划分的应用情况.

4.1 模型化

根据杭州华东制药厂螺旋霉素发酵过程 23 批历史资料的分析, 选取了尾气 CO_2 浓度; CO_2 浓度变化率; 发酵液溶氧浓度和溶氧浓度变化率 4 个特征变量. 并将过程族 \mathcal{F} 分成具有强 CO_2 释放率特性 P_1 和具有弱 CO_2 释放率特性 P_2 的两个过程类. 又从概念上把每类过程分成 4 个模式, 以分别对应于过程的迟缓期 (Y_1); 菌体生长期和菌体代谢(抗生素)初期 (Y_2); 菌体代谢旺盛期 (Y_3) 和菌体自溶且代谢的减速期 (Y_4).

根据过程原始数据, 运用模糊数学动态聚类分析方法, 对所划分的 CPA $_i$ 求取各特征变量 $v_{ijk}(i = 1, 2, j, k = 1, 2, 3, 4)$ 的聚类中心, 具体步骤如下:

1) 数据标定

经过多种方法分析比较, 选择了海明距离^[4]对上述 P_1, P_2 的 15 和 8 批的各批 30 组数据 \bar{v}_{ikp} 作标定, 即

$$r_{pq}^{(t)} = 1 - c \sum_{k=1}^4 |\bar{v}_{ikp} - \bar{v}_{ikq}|,$$

$$i = 1: t = 1, \dots, 15; i = 2: t = 1, \dots, 8; p, q = 1, \dots, 30.$$

其中常数 c 取作 0.08.

2) 动态聚类

通过上述的数据标定, 可以得到 23 个 30×30 阶的相似矩阵, 即

$$R^{(t_1)} = (r_{pq}^{(t_1)})_{30 \times 30}, t_1 = 1, \dots, 15, \quad (14)$$

$$R^{(t_2)} = (r_{pq}^{(t_2)})_{30 \times 30}, t_2 = 1, \dots, 8. \quad (15)$$

用平方法分别对 $R^{(1)}$ 和 $R^{(2)}$ 求传递闭包,得 $\tilde{R}^{(1)}$ 和 $\tilde{R}^{(2)}$. 再分别按 $\tilde{R}^{(1)}$ 和 $\tilde{R}^{(2)}$ 的 λ —截对其中元素进行动态聚类. 当 $\lambda = 0.965$ 时, 可将每一过程的历史数据分成 4 类, 即代表 4 种模式

3) 求标聚类中心

设论域 Ω 上第 k 个特征变量 v_k 的样本值属于 P_i 且处于 CPA_i^j 内的聚类中心值为 c_{ijk} ; v_k 的样本值属于 P_i 且处于 CPA_i^j 中的样本总数为 N_{ijk} ; 属于 P_i 且在 CPA_i^j 的 v_k 的样本值用 θ_{ijke} 表示 ($e = 1, \dots, N_{ijk}$), 则 c_{ijk} 取相关样本值的算术平均, 即

$$c_{ijk} = \frac{1}{N_{ijk}} \sum_{e=1}^{N_{ijk}} \theta_{ijke}, \quad i = 1, 2, j, k = 1, 2, 3, 4. \quad (16)$$

设属于 P_i 和 CPA_i^j 的表征样本模式端点的第 k 个特征变量的样本值为 $\theta_{ijk0}^{(i)}$ (起点) 和 $\theta_{ijkf}^{(i)}$ (终点), 且考虑到过程的连续性, 令 $\theta_{ijkf}^{(i)} = \theta_{ij+1,k0}^{(i)}$, 则典型模式相关节点确定为

$$v_{ijk}^* = \frac{1}{T_i} \sum_{i=1}^{T_i} \theta_{ijk0}^{(i)}, \quad j, k = 1, 2, 3, 4. \quad (17)$$

其中 $T_1 = 15; T_2 = 8$. 典型过程的终点为

$$v_{ijk}^* = \frac{1}{T_i} \sum_{i=1}^{T_i} \theta_{ijkf}^{(i)}, \quad k = 1, 2, 3, 4. \quad (18)$$

c_{ijk} 和 v_{ijk}^* 数值的规一化处理方式为: 令 c_{1jk} 和 v_{1jk}^* 均为 1, 即

$$\bar{c}_{1jk} = 1, \quad j, k = 1, 2, 3, 4. \quad (19)$$

$$\bar{v}_{1jk}^* = 1, \quad j = 2, 3, 4, k = 1, 2, 3, 4. \quad (20)$$

则 c_{2jk} 和 v_{2jk}^* 的规一化值为

$$\bar{c}_{2jk} = c_{2jk}/c_{1jk}, \quad j, k = 1, 2, 3, 4. \quad (21)$$

$$\bar{v}_{2jk}^* = v_{2jk}^*/v_{1jk}^*, \quad j = 2, 3, 4; k = 1, 2, 3, 4. \quad (22)$$

至此, 可按第 2 节中步骤 4) 确定隶属函数形式如下:

$$\mu_{CPA_i^j}^k(v_k, c_{ijk}) = 1 - \xi \left(\frac{\tilde{v}_k - c_{ijk}}{\Delta_{ijk}} \right)^2, \quad i = 1, 2, j, k = 1, 2, 3, 4. \quad (23)$$

其中, ξ 为常数, 这里取为 0.8; \tilde{v}_k 为第 k 个特征变量的测量值; Δ_{ijk} 由下式定义:

$$\Delta_{ijk} \triangleq \max_e \{(\theta_{ijke} - c_{ijk})\}, \quad (24)$$

其中, θ_{ijke} ($e = 1, \dots, N_{ijk}$) 如上所述是属于 P_i 且在 CPA_i^j 中的第 k 个特征变量的样本值.

4.2 模型的实施及结果

将上节所建立的螺旋霉素间歇生产过程菌体生长期的在线模糊识别模型, 按第 3 节介绍的实施方案, 用 BASIC 语言在 IBM PC/XT 机上编制了实时系统软件. 经过几批生产过程系统软件的在线调试后, 对于该模型实时连续运行考核的三批结果列于表 1 中. 从表中可见, 模型识别预报的第一个生长期(Y_1 , 过程的迟缓期)单位时间产生的发酵总亿均为零, 即菌丝不代谢螺旋霉素; 模型识别预报的第二个生长期(Y_2 , 菌体的主要生长期和

其代谢初期)单位时间产生的发酵总亿较少;而模型识别预报的第三个生长期(Y_3 , 菌丝代谢的旺盛期)单位时间产生的发酵总亿明显增加。由此说明,该模型从性能上能够满足菌体的不同生长期(代表过程的不同模式)在线识别、预报的要求。需要说明的是用于考核该模型的三批过程,由于生产调度上的原因,均未使过程达到菌体生长的第4个生长期(Y_4)就放罐了。

表1 菌体生长期识别模型的在线考核结果

批 号	发酵时间 (hr.)	模型预报	每小时产出总亿 (r/hr.)
1	0~7.5	I	0
	7.5~53.5	II	0.198
	53.5~138	III	0.371
2	0~8	I	0
	8~62.9	II	0.194
	62.9~136	III	0.475
3	0~4	I	0
	4~58.3	II	0.249
	58.3~132	III	0.585

5 结语

本文提出的一种间歇生产过程动态特性的模糊模式识别方法,概括起来有以下特点:

1) 这是一种以工业生产过程为主要对象的识别其动态特性的模糊模式识别方法,具有很强的在线可行性。它的实现无疑将有助于提高间歇生产过程监督控制的质量及性能。

2) 由于这种方法的实施方案中引入了学习功能(如第3节步骤8)~10)所述),故对于最初所建立的识别模型的精度要求不高。由此方便了建模和提高了识别模型的自适应性。

3) 在本方法的实施方案中,由于采用了启发式搜索模式聚类中心的机制(如第三节步骤5)~7)所述),从而可提高该模型在线识别的精度。

参 考 文 献

- [1] 吴望名等,应用模糊集方法。北京师范大学出版社,1985。

A NEW METHOD FOR ON-LINE RECOGNIZING BATCH PROCESS CHARACTERISTICS AND ITS APPLICATION FOR A FERMENTATION PROCESS

FU CHUNSHENG WANG JICHENG

(Dept. of Chem. Machinery, Nanjing Institute of Chem. Tech. 210009)

WANG JICHENG

(Institute of Industrial Process Control, Zhejiang Univ., Hangzhou, 310027)

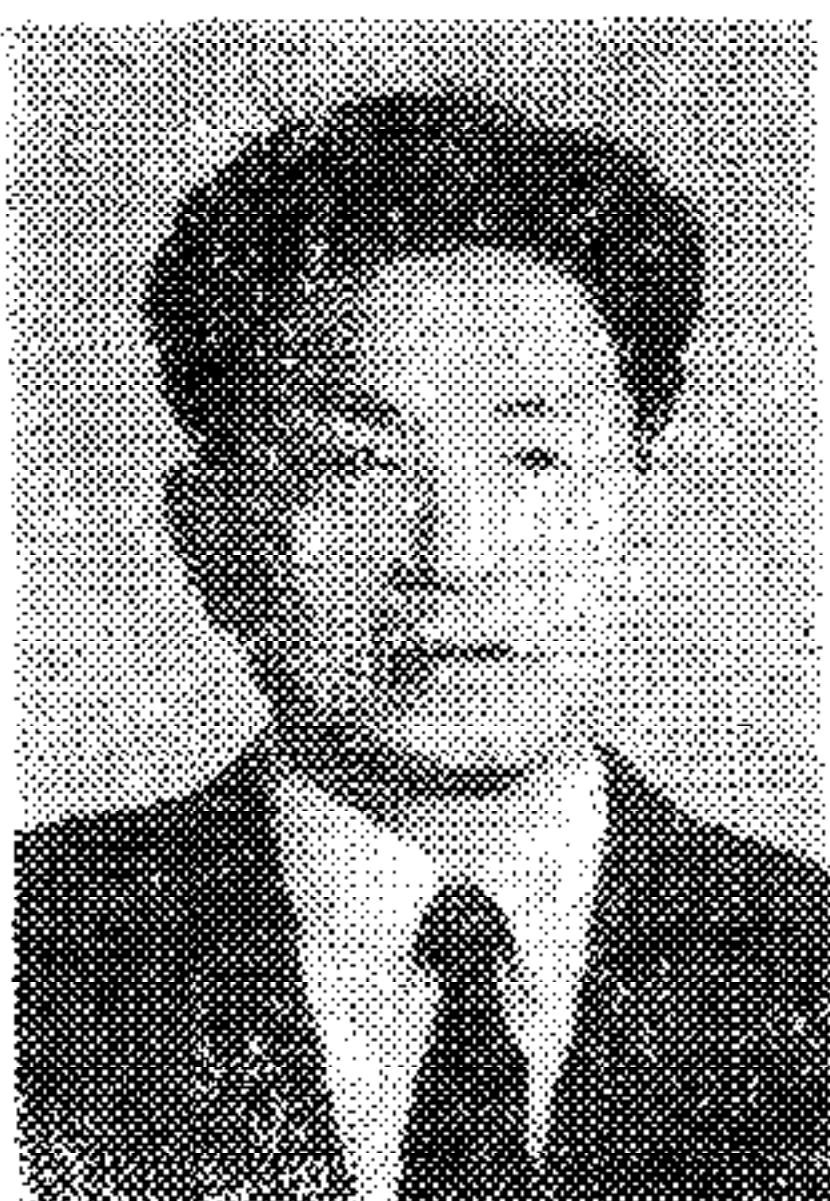
ABSTRACT

A new modelling method with self-learning function for on-line recognizing batch process characteristics is proposed by using fuzzy set theory and pattern recognition techniques. This method is applied to the recognition problem of cell growth phases for an actual industrial fermentation process.

Key words: batch process; dynamic characteristic; on-line recognition; fuzzy set; learning.



傅春生 1956年3月生。1982年在华东化工学院获学士学位，1990年在浙江大学获博士学位。1982年2月至1985年8月及1990年4月至1992年1月工作于南京化工学院。现在西班牙巴塞罗那自治大学做博士后。研究方向：模糊模型和控制，专家系统，神经网络，模式识别和生化过程控制。



王驥程 生于1928年。1951年在浙江大学获学士学位，1953年获硕士学位，现为浙江大学工业控制技术研究所教授、博士生导师，并兼任杭州应用工程技术学院院长。长期从事教学和科研工作，出版了多种教学用书，其中《化工过程控制工程》获化工部优秀教材奖。研究领域涉及过程动态学，生化过程的建模，计算机优化控制，专家系统在生产中的应用等。已发表论文140余篇。