

文章编号 1000-5013(2000)01-0071-05

遗传算法应用于分批发酵动力学 模型参数估算

陈宏文^① 方柏山^① 胡宗定^②

(① 华侨大学化工学院, 泉州 362011; ② 天津大学化工学院, 天津 300072)

摘要 由通用发酵动力学模型导出用于描述分批发酵特征的解析解, 把遗传算法应用于求解分批发酵动力学模型参数. 以赖氨酸分批发酵过程模拟为例, 表明用遗传算法可有效地解决赖氨酸分批发酵动力学模型这类复杂的非线性函数的参数估算问题. 用 POWELL 方法估算结果作比较后表明, 遗传算法能进一步提高赖氨酸分批发酵过程状态变量的计算值与实验值的吻合程度.

关键词 遗传算法, 发酵动力学, 参数估算

中图分类号 TQ 920.1

文献标识码 A

研究发酵动力学是实现发酵过程最优控制的前提条件, 也是研究发酵过程放大及从分批发酵过渡到流加发酵、连续发酵的理论基础. 动力学模型的建立, 离不开模型参数的估算. 对于显著非线性的发酵动力学模型而言, 传统的“点到点”的优化算法对于初始点的确定往往比较困难, 而初始点的选择又直接影响到算法的收敛. 近 20 年来, 一种成为研究热点的遗传算法^[1] (Genetic Algorithms, 简称 GA) 以其高效、自适应及益于全局搜索的优势, 在许多领域中得到应用. 在处理复杂的多参数非线性函数优化问题时, 该方法是在解空间中随机产生初始点, 再进行多点并行搜索, 从而能以较大的概率达到全局最优. 本文运用 GA 算法求解赖氨酸发酵动力学参数, 并与用 POWELL 方法求得的结果相比较^[1], 收到较好的效果.

1 常见分批发酵动力学模型的解析解

1.1 分批发酵动力学模型

1.1.1 微生物生长动力学模型 目前使用最广泛、形式比较简单的微生物生长动力学模型, 是由 Monod 于 1949 年提出的经验模型^[6]. 即

$$\frac{dx}{dt} = \mu_{\max} \frac{sx}{k_s + s}, \quad (1)$$

其中 $\frac{dx}{dt}$ 是微生物生长速度, μ_{\max} 是微生物最大比生长速度, x, s 分别为微生物浓度和限制性底

收稿日期 1999-04-19 作者简介 陈宏文(1969-), 女, 助教

基金项目 福建省自然科学基金和生物反应器工程国家重点实验室开放课题基金资助项目

1994-2024 China Academic Electronic Journal Publishing House. All rights reserved. <http://www.cnki.net>

物浓度, k_s 为 Monod 饱和常数.

1.1.2 产物形成动力学模型 较为通用的是由 Luedeking 和 Piret 于 1959 年描述乳酸发酵时所提出的数学模型^[6]. 即

$$\frac{dp}{dt} = \alpha \frac{dx}{dt} + \beta x, \quad (2)$$

其中 $\frac{dp}{dt}$ 是产物生长速度, α 和 β 皆为动力学模型参数. 当 $\alpha = 0, \beta = 0$, 式(2)称为生长关联型; 当 $\alpha = 0, \beta \neq 0$ 时, 式(2)称为非生长关联型; 而当 $\alpha \neq 0, \beta \neq 0$ 时, 式(2)则称为混合型.

1.1.3 底物消耗动力学模型 通过分批发酵过程中底物消耗的物料平衡原理, 可以建立如下方程式

$$-\frac{ds}{dt} = \frac{1}{Y_G} \frac{dx}{dt} + mx + \frac{1}{Y_P} \frac{dp}{dt}, \quad (3)$$

其中 $-\frac{ds}{dt}$ 是限制性底物的总消耗速度, m 是维持系数, Y_G 是最大的微生物生长得率系数, Y_P 是最大的产物生成得率系数.

1.2 解析解

式(1)~式(3)系多元的非线性微分方程组. 经积分, 可求解得到 $x-s$, $p-s$ 和 $t-s$ 的数学关系式. 即

$$x = A \left(B \ln \frac{s+B}{s_0+B} + s_0 - s \right) + x_0, \quad (4)$$

$$p = C \left(D \ln \frac{s+B}{s_0+B} + s_0 - s \right) + p_0, \quad (5)$$

$$t = b_1 \ln(s + b_2) + b_3 \ln[-(s^2 + b_4 s + b_5)] + b_6 \ln\left(-\frac{s+b_7}{s+b_8}\right) + b_9, \quad (6)$$

其中 A, B, C, D 和 $b_1 \sim b_9$ 是与动力学参数 $\mu_{\max}, k_s, \alpha, \beta, Y_G, Y_P$ 和 m 有关的复合参数^[6]. 据此, 由分批发酵过程中所测得的 $s \sim t$ 和 $p \sim t$ 数据, 便可确定 $\mu_{\max}, k_s, \alpha, \beta, Y_G, Y_P$ 和 m 等 7 个参数. 方柏山等人曾以 *Lorynebacterium glutamicum* 分批发酵生产赖氨酸的实验数据为例^[6](表 1), 用 POWELL 法进行参数估算, 取得了较好的结果.

表 1 赖氨酸分批发酵实验值^[6]

t/h	0	10	20	30	40	50	60
$s/g \cdot L^{-1}$	200	192	180	133	104	55.4	22.2
$p/g \cdot L^{-1}$	0	0.83	9.17	16.7	29.6	40.6	43.9
$x/g \cdot L^{-1}$	10.0	12.2	15.5	20.0	23.9	25.7	26.1

为了探讨应用 GA 算法进行参数估算的可行性和先进性, 我们以式(7)为目标函数, 即

$$g(s) = \sum_{j=1}^3 \sum_{i=1}^n (y_{ij})^2, \quad (7)$$

其中 y_{ij} 表示 x, p, t 在不同 s 浓度下的实验值与计算值的离差平方和. 为求目标函数的极小值, 对赖氨酸分批发酵动力学模型参数采用 GA 算法进行估算.

2 GA 算法的应用

2.1 编码

首先根据已有的知识和经验确定 7 个待估算参数的取值范围. 因 Y_G 和 Y_P 为最大的微生物生长得率和最大的产物生长得率, 其值不超过 1, 故设 $Y_G \in [0.01, 1]$, $Y_P \in [0.01, 1]$. 参考有关的文献报道^[1], 为了尽可能既扩大 μ_{\max} , k_s , α , β , m 等 5 个参数的寻优范围, 又避免盲目搜索以及减少计算量, 可以设 $\mu_{\max} \in [0.001, 1]$, $k_s \in [2, 200]$, $m \in [0.001, 10]$, $\alpha \in [0.1, 100]$, $\beta \in [0.01, 1]$.

编码采用二进制. 按照多参数编码方法, 取每个子串长度 l 均为 11, 则所连成的一个染色体(个体)长度 L 为 $77(11 \times 7 = 77)$. 每个参数实数值 $X(X \in [U_{\min}, U_{\max}])$ 与其二进制解码整数值 $X'(X' \in [0, 2^l - 1])$ 的对应关系为

$$X = U_{\min} + \frac{U_{\max} - U_{\min}}{2^l - 1} X' = U_{\min} + a X', \tag{8}$$

其中 $a = \frac{U_{\max} - U_{\min}}{2^l - 1}$ 为各参数的精度. 当 $l = 11$ 时, 各参数的精度如表 2 所示.

表 2 子串长度 $l = 11$ 时各参数的精度

参数	μ_{\max}	k_s	m	α	β	Y_G	Y_P
a	4.88×10^{-4}	9.70×10^{-2}	4.88×10^{-3}	4.88×10^{-2}	4.88×10^{-4}	4.88×10^{-4}	4.84×10^{-4}

2.2 初始化种群

种群规模要保持适当. 若过小, 其所含的信息太少, 不能发挥 GA 算法的效力. 若过大, 计算量随之增加, 收敛时间延长. 针对这个问题, 我们随机产生 40 个个体. 该群体代表了优化问题的一些可能解的集合.

2.3 适应度计算

由于所求解的问题是目标函数 $g(s)$ 的最小化, 故取适应度函数. 即

$$f(s) = \begin{cases} C_{\max} - g(s) & (g(s) < C_{\max}), \\ 0 & (g(s) \geq C_{\max}), \end{cases}$$

其中 $C_{\max} = 10\,000$ 为进化过程中得到的较大目标函数值. 这样 $f(s)$ 的变化范围在 $0 \sim 10\,000$ 之间.

2.4 选择(复制)

采用转轮法, 即适应度越高的个体, 从父代中被选中的概率越大. 在运算过程中, 为了增加高适应度个体被选中的机会, 加快进化过程, 我们引入期望值判断方法. 即每个个体在下一代生存的期望数 M 为

$$M = f_i / \bar{f} = f_i / (\sum f_i / n). \tag{9}$$

设 $C = 0.5$, 若 $M - C > 0$ 且符合转轮法选择判断, 则该个体被选中, 否则被淘汰. 显然, C 的取值不能太大, 否则容易陷入局部极值.

2.5 交叉

采用单点交叉方式. 交叉概率 P_c 为 0.65, 取得较好的结果.

2.6 变异

单点变异的概率 P_m 为 0.05 时, 结果较好.

2.7 收敛判断

以前后两代的串集适应度平均值之比趋于 1 为准.

3 计算结果和讨论

3.1 结果

通过表 3 和图 1、2 进行计算结果的分析. 取种群大小为 40, 最大代数为 100, 子串长度

表 3 利用 GA 和 POWELL 算法得到的参数估算值

算法	μ_{\max}/h^{-1}	$k_s/\text{g} \cdot \text{L}^{-1}$	m/h^{-1}	$\alpha/\text{g} \cdot \text{g}^{-1}$	β/h^{-1}	$Y_C/\text{g} \cdot \text{g}^{-1}$	$Y_P/\text{g} \cdot \text{g}^{-1}$	$g(s)$
GA	0.032 9	29.4	0.118	1.58	0.016 2	0.720	0.919	202.36
POWELL	0.021 7	19.6	0.092	1.66	0.014 3	0.428	0.985	246.62

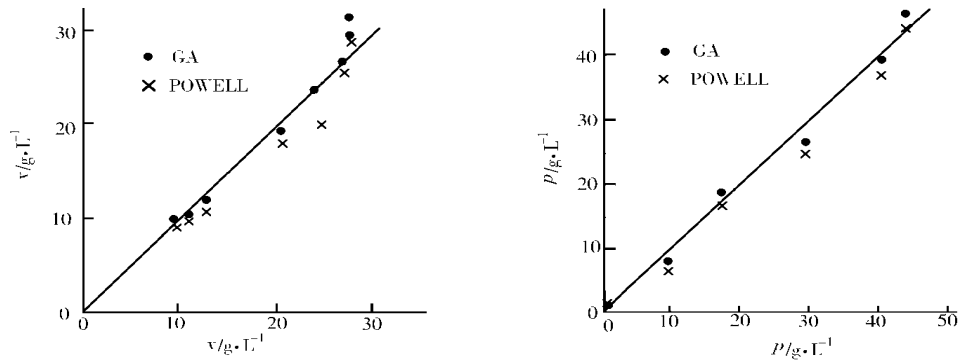


图 1 两种算法所得到的 x, p 计算值与其实验值的比较

为 $l=7$, 每个个体串长 $L=77$, 交叉概率 $P_c=0.65$, 变异概率 $P_m=0.05$. 用 C 语言编程, 在 Pentium133 微机上进行参数估算. 当 $gen=0$ 时, 初始串集的平均值为 667.8; 经过 37 次迭代, 可上升到 9 758.2. 此时, 前后两代串集适应度平均值为 1.007. 相应的参数估算值及适应度值见表 3; 相应由两种算法得到的 x, p 和 t 计算值与对应实验值的偏差程度, 如图 1、2 所示.

3.2 讨论

(1) 研究结果表明, GA 算法可用于分批发酵动力学模型的参数估算. 从表 3 可以看出, 与 POWELL 算法相比, 目标函数值 $g(s)$ 从 246.80 下降到 202.36. 由图 1、2 可见, 基于 GA 算法的数学模型的计算值与实验值的吻合度较好. (2) 每一次运算, 虽然只有 40 个个体参与, 但 GA 算法的隐并行性处理方式, 使得每一次的运算隐含地处理了 40^3 个模式⁶⁾. 又由于每一次运算的初始串集都是在一定的范围随机产生的, 不同于盲目的随机搜索. 因此, GA 算法搜索面很广, 寻优速度快. 通过对不同初始串集的多次实验, 各动力学参数收敛

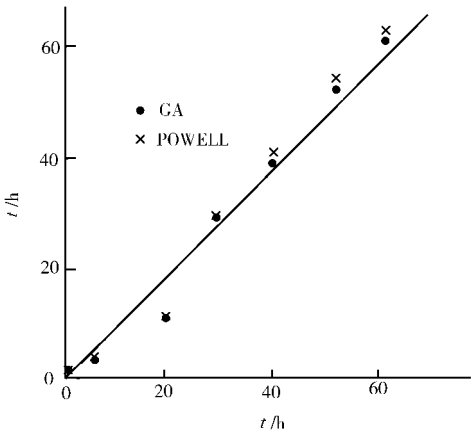


图 2 两种算法所得到的 t 计算值与其实验值的比较

值与上述结果基本一致. 可以认为, 我们所得到的结果可能以较大的概率接近全局最优解⁶⁾. (3) 在计算适应度时, 我们曾尝试加入适应度函数转换, 但在多次运算过程中, 没有发现超常个体. 引入适应度函数转换, 反而减弱了个体之间的竞争水平, 使运算过早收敛, 串集整体性

能无法提高. 而去掉适应度函数转换后, 很快就获得较满意的结果. 因此可以认为, 是否引入适应度函数转换以及选用什么样的转换方式, 都要视具体问题而定.

参 考 文 献

- 1 Holland J H, 李英毅. 基因算法[J]. 科学, 1992, 18(4): 24 ~ 31
- 2 方柏山, 林金清. 分批发酵动力学模型参数的估算[J]. 生物工程学报, 1992, 8(3): 283 ~ 287
- 3 Monod J. The Growth of Bacterial Culture[J]. Ann. Rev. Microbiol. 1949. (3): 371 ~ 372
- 4 Luedeking R, Piret E L. A Kinetic Study of the Lactic Acid Fermentation Batch Process at Controlled pH [J]. J. Biochem. Microbiol. Tech. Eng., 1959, (1): 431 ~ 459
- 5 Atkinson B, Mauituna F. Biochemical engineering and biotechnology handbook[M]. 1st ed. England: Machilla Publisher Ltd, 1983. 287 ~ 288
- 6 孙艳丰, 王众托. 遗传算法在优化问题中的应用研究进展[J]. 控制与决策, 1996, 11(4): 425 ~ 431

Application of Genetic Algorithm to Estimating Parameters from Kinetic Model of Lysine Batch Fermentation

Chen Hongwen^① Fang Baishan^① Hu Zongding^②

(^① College of Chem. Eng., Huaqiao Univ., 362011, Quanzhou;

^② College of Chem. Eng., Tianjin Univ., 300072, Tianjin)

Abstract By using analytical solution derived from general kinetic model of fermentation for describing characteristic of batch fermentation, a study is made on the application of genetic algorithm to estimating parameters from kinetic model of batch fermentation. Taking the simulation of batch fermentation process of lysine as example, useful results are obtained. As shown by the results, genetic algorithm is capable of effectively estimating such kind of parameters with complex nonlinear function from kinetic model of lysine batch fermentation. As compared with estimating results gotten by POWELL method, the application of genetic algorithm will further increase the coincidence of calculated values and experimental values of state variables from lysine batch fermentation process.

Keywords genetic algorithm, kinetics of fermentation, parameter estimation