

研究论文

一种提高微生物发酵单位的新优化控制策略

高学金, 王 普, 张亚庭, 张会清, 齐咏生, 关 伟
(北京工业大学电子信息与控制工程学院, 北京 100124)

摘要: 为了提高微生物发酵单位, 提出了支持向量机 (SVM) 与基于实数编码遗传算法 (RGA) 相互耦联的优化控制策略。为解决发酵参数之间的耦合问题, 进一步引入了模式的概念。SVM 建立微生物发酵过程的预估模型, RGA 以此模型为适应度函数计算最优控制模式。此策略应用于青霉素发酵过程的优化控制, 效价比没有实施优化控制策略的生产结果提高了 22.88%。

关键词: 微生物发酵; 优化控制; 建模; 支持向量机; 遗传算法

中图分类号: TP 273; Q 81

文献标识码: A

文章编号: 0438-1157 (2008) 06-1462-08

A new strategy of optimization control for increasing fermentation unit

GAO Xuejin, WANG Pu, ZHANG Yating, ZHANG Huiqing, QI Yongsheng, GUAN Wei

(School of Electronic Information and Control Engineering, Beijing University of Technology, Beijing 100124, China)

Abstract: To increase fermentation unit, the strategy of optimization control that combines the support vector machine (SVM) with genetic algorithm based on real coding (RGA) is proposed. To solve the coupling of fermentation parameters, the idea of pattern is also introduced. SVM establishes the prediction model for the microbial process, and RGA taking the model as fitness function calculates the optimal control pattern. The results show that the penicillin titer of fermentation process optimized is increased by 22.88% compared with that of fermentation process not optimized.

Key words: microbial fermentation; optimization control; modeling; support vector machine; genetic algorithm

引 言

发酵过程的参数值一般凭借工艺人员的经验确定, 以稳定生产为目标, 很少考虑优化操作, 带有很大的主观性和随意性。如果将此值作为单一给定值控制整个发酵过程, 就容易出现发酵单位低而且波动性大的情况, 主要是因为微生物生长和代谢产物生成阶段所需要的操作条件是不同的, 甚至在发酵过程的每个时刻, 发酵参数值都是不同的^[1]。微

生物发酵过程的优化控制就是寻找发酵参数的最优控制轨线, 使生产菌的生长和代谢变化沿着此轨线进行, 以较低的能量和物料消耗生产出更多的发酵产品。高成本和高能耗是微生物发酵生产的特征之一, 以青霉素发酵为例, 每生产 1 t 青霉素, 至少要消耗 10 t 葡萄糖, 其中 70% 主要用于维持。当今世界, 青霉素工业发酵水平已达 85000 U · ml⁻¹, 然而, 我国一些制药厂的发酵水平还徘徊在 60000 U · ml⁻¹ 左右^[2]。因此, 为了提高发酵单

2007-11-18 收到初稿, 2008-01-17 收到修改稿。

联系人及第一作者: 高学金 (1973—), 男, 博士, 讲师。

基金项目: 国家自然科学基金项目 (60704036); 北京工业大学青年基金项目 (97002011200701); 北京工业大学博士科研启动基金项目 (52002011200707)。

Received date: 2007-11-18.

Corresponding author: Dr. GAO Xuejin. **E-mail:** gaouxuejin@bjut.edu.cn

Foundation item: supported by the National Natural Science Foundation of China (60704036), the Science Fund for Young Scholars of Beijing University of Technology (97002011200701) and the Scientific Research Foundation for Doctors of Beijing University of Technology (52002011200707).

位和降低生产成本，优化控制微生物发酵过程就成为了一条重要途径。这主要集中在建模与优化控制方法的研究。

遗传算法 (GA) 基于多点搜索，所以能够同时计算多个解，并具有全局搜索能力^[3]。另外，计算机的并行计算能力也提高了搜索速度。在搜索过程中，GA 能自动获得和积累有关搜索空间的知识，并自适应地控制搜索过程以求得最优解。所以，GA 在函数优化、参数辨识、模式识别等领域得到了广泛应用^[4-5]。

GA 优化发酵参数，必需有可靠的、能够反映过程中各参数变化规律及其相互关系的模型^[6]。由于发酵过程的非线性、时变性和生物传感器的缺乏以及各参数之间的严重关联，经典系统理论几乎不能为复杂的发酵过程建立合适的模型^[7]。因此，许多学者采用人工神经网络 (ANN) 为发酵过程建模^[6-8]，但是基于经验风险最小化理论的 ANN，需要较大的样本数目，力求在样本点的训练误差最小化，因而不可避免地出现“过拟合”现象，降低了模型的泛化能力。并且，ANN 的结构设计依赖于设计者的先验知识和经验，缺乏一种理论指导^[9]。Vapnik^[10-11]提出的支持向量机 (SVM) 方法，采用结构风险最小化原则代替经验风险最小化原则，克服了 ANN 的固有缺陷，解决了小样本、非线性、高维数、局部极小点等实际问题，具有很强的泛化能力^[12]。

本文提出了支持向量机与基于实数编码遗传算法 (RGA) 相互耦联的优化控制策略，为解决发酵参数之间的耦合问题引入了模式的概念，即：将描述一个工况的一组参数作为一个整体来考虑。在 Matlab6.5 软件中，使用生产数据建立了基于 SVM 的青霉素发酵过程模型，评价了此模型的拟合精度和泛化能力。RGA 以此模型为适应度函数，计算得到了青霉素发酵过程的最优控制模式。以此控制模式控制发酵过程，并对发酵结果进行了分析和比较。

1 模型的结构

SVM 建立的微生物发酵过程模型属于“黑箱”模型，所谓“黑箱”模型是将发酵过程看作一个黑箱，不关心其内部变化，只考虑其输入变量和输出变量。输入变量是指影响发酵过程变化的关键因素，如：温度、pH 值、溶解氧、补料速率、基质

浓度、产物浓度。输出变量是指与建模目的息息相关的变量，如：产物浓度、基质浓度、菌体浓度。为了实现微生物发酵过程的优化控制，应建立预估模型。此模型通过 t 发酵时刻的状态估计 $t + \Delta t$ 发酵时刻的状态，属于动态模型，覆盖整个发酵过程。模型的结构如图 1 所示， T 代表发酵温度， F_s 代表补料速率， $P(t)$ 和 $S(t)$ 分别代表 t 时刻的产物浓度和基质浓度， $P(t + \Delta t)$ 、 $S(t + \Delta t)$ 和 $X(t + \Delta t)$ 分别代表 $t + \Delta t$ 时刻的产物浓度、基质浓度和菌体浓度。

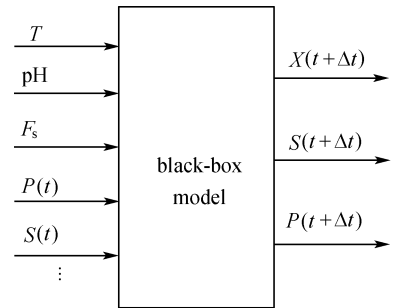


图 1 微生物发酵过程的黑箱模型结构

Fig. 1 Structure of black-box model

2 基于 SVM 的建模方法

基于 SVM 的微生物发酵过程建模方法如下。

Step1 依据发酵过程的特点和发酵过程模型的用途，分析影响发酵过程的因素。

Step2 从生产现场获取能够体现发酵过程特点和产率高的生产数据。为了开发出有效的基于 SVM 的模型，产生数据样本是非常关键的一步。包括原始数据的收集、数据分析、变量选择以及数据的预处理，只有经过这些步骤后才能对 SVM 进行有效的学习和训练。

Step3 将某批发酵数据按照采样间隔 Δt 分割成若干数据样本。采用 SVM 建立发酵过程模型不需要对样本进行归一化。设已知训练样本为 (x_1, y_1) , (x_2, y_2) , \dots , (x_n, y_n) ，其中， $x_i \in R^d$, $y_i \in R$, $i = 1, 2, \dots, n$ 。

Step4 选择适当的不敏感损失函数 ϵ 和惩罚系数 C ；根据发酵过程的特点，选择适当的核函数 $K(x_i, x_j)$ ，并确定相应的核参数值。

Step5 按照样本的排列顺序，输入训练样本 (x_i, y_i) 。

Step6 构造并求解最优化问题

$$\max[W(a, a^*)] =$$

$$\max \left[-\epsilon \sum_{i=1}^n (\alpha_i^* + \alpha_i) + \sum_{i=1}^n y_i (\alpha_i^* - \alpha_i) - \frac{1}{2} \sum_{i,j=1}^n (\alpha_i^* - \alpha_i) \cdot (\alpha_j^* - \alpha_j) K(x_i, x_j) \right] \quad (1)$$

$$\text{s. t.} \quad \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0$$

$$0 \leq \alpha_i, \alpha_i^* \leq C, i = 1, 2, \dots, n \quad (2)$$

得到最优解 $\bar{\alpha} = (\bar{\alpha}_1, \bar{\alpha}_1^*, \dots, \bar{\alpha}_n, \bar{\alpha}_n^*)^T$ 。

Step7 构造决策函数

$$f(x) = \sum_{i=1}^n (\bar{\alpha}_i^* - \bar{\alpha}_i) K(x_i, x) + b^* \quad (3)$$

其中, b^* 计算方式^[13]如下:

选择位于开区间 $(0, C)$ 中的 $\bar{\alpha}_j$ 或 $\bar{\alpha}_k^*$, 若选到的是 $\bar{\alpha}_j$, 则

$$b = y_j - \sum_{i=1}^n (\bar{\alpha}_i^* - \bar{\alpha}_i) (x_i \cdot x_j) + \epsilon \quad (4)$$

若选到的是 $\bar{\alpha}_k^*$, 则

$$b = y_k - \sum_{i=1}^n (\bar{\alpha}_i^* - \bar{\alpha}_i) (x_i \cdot x_k) - \epsilon \quad (5)$$

按式 (4) 或式 (5) 求出所有支持向量的 b 值, 然后以它们的平均值作为 b^* 。

Step8 使用各样本点相对误差的均值 MER (mean error ration) 计算模型的拟合误差, 计算方法如式 (6)。如果拟合误差大于 ϵ 的设定值, 则转入 Step4, 重新选择 C 和核参数值; 否则转入 Step9。

$$\text{MER} = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \quad (6)$$

其中, n 为样本数目, y_i 、 \hat{y}_i 分别为实际值和估计值。

Step9 使用测试样本测试模型的泛化能力。同样, 模型的泛化能力也使用 MER 衡量。

3 SVM 与 RGA 耦联的优化控制策略

发酵参数之间相互耦合, 很难对单个参数进行优化。因此, 需要充分考虑参数之间的关系, 将描述一个工况的一组参数作为一个整体来考虑, 即为一个模式。

将描述发酵状况的工艺参数分为状态模式和控制模式。定义状态模式为 $\mathbf{x} = [x_1, x_2, \dots, x_m]$, 在优化过程中为常量; 定义控制模式为 $\mathbf{u} = [u_1, u_2, \dots, u_n]$, 在优化过程中为变量, 是需要优化的发酵参数; m 、 n 分别为状态模式和控制模式的维数; 定义优化目标为单变量 J , 一般为产物浓度或者菌体浓度。

定义寻找最优控制模式 $\mathbf{u}^*(t)$ 使目标函数 J

为最大的过程被称为发酵过程的优化控制, 即: 如果 $x_1 = \hat{x}_1, x_2 = \hat{x}_2, \dots, x_m = \hat{x}_m$ 并且 J 最大, 则 $u_1 = u_1^*, u_2 = u_2^*, \dots, u_n = u_n^*$ 。 $[u_1^*, u_2^*, \dots, u_n^*]$ 即为所求的最优控制模式。

3.1 优化控制问题的描述

在对数生长期, 只有控制最优的生长条件, 使菌体旺盛生长、繁殖, 才能获得较高的菌体浓度, 而最优的生长状态和较高的菌体浓度是更多地获得发酵产物的必要前提。在稳定期, 只有控制最优的产物生成条件, 使菌体充分地发挥固有的发酵性能, 才能获得最高的发酵水平。因此, 将微生物发酵过程划分为以菌体生长为主的对数生长期和以产物生成为主的稳定期两个阶段。在不同的发酵阶段, 应该设置不同的目标进行优化控制。

因此, 优化控制问题可以描述为: 求取最优控制模式并加以控制, 使得对数生长期的菌体浓度最大, 稳定期的产物浓度最大。

(1) 对数生长期

目标函数

$$\max X(t + \Delta t) \quad (7)$$

$$X(t + \Delta t) = f[X(t), P(t), S(t), \mathbf{u}(t)] \quad (8)$$

约束条件

$$\begin{cases} u_{\min}(t) \leq \mathbf{u}(t) \leq u_{\max}(t) \\ V(t_i) \leq V_{\max} \end{cases} \quad (9)$$

(2) 稳定期

目标函数

$$\max P(t + \Delta t) \quad (10)$$

$$P(t + \Delta t) = f[P(t), X(t), S(t), \mathbf{u}(t)] \quad (11)$$

约束条件

$$\begin{cases} u_{\min}(t) \leq \mathbf{u}(t) \leq u_{\max}(t) \\ V(t_i) \leq V_{\max} \end{cases} \quad (12)$$

其中, $X(t)$ 、 $P(t)$ 和 $S(t)$ 构成 t 时刻的状态模式; $\mathbf{u}(t)$ 是控制模式, 主要包括温度、压力、pH 值和补料速率等发酵参数; $u_{\min}(t)$ 代表这些参数在 t 时刻的最小值; $u_{\max}(t)$ 代表这些参数在 t 时刻的最大值; t_f 、 V_f 、 V_{\max} 分别为发酵终止时刻、发酵终止时的发酵液体积和发酵罐的最大体积。

3.2 RGA

编码是设计遗传算法的一个关键步骤, 编码方式不仅直接决定了个体染色体的排列形式和个体从搜索空间的基因型到解空间的表现型的解码方法, 而且也影响到交叉算子、变异算子等的运算方法, 很大程度上影响了遗传算法的运算效率, 而且与遗

传算法的收敛性能有密切的关系^[14]。Holland^[3]运用模式定理分析编码机制，建议采用二进制编码。二进制编码方式具有编码、解码简单，对交叉、变异等遗传操作易于实现，但是它不便于反映问题的特定知识，对于连续优化问题局部搜索能力较差，对于高维、高精度优化问题往往由于个体串位太大而造成计算量过大及过早收敛^[15]。

基于实数编码的遗传算法是对连续参数优化问题的自然描述，不存在编码和解码过程，能够大大提高解的精度和收敛速度；便于进行大空间搜索；既可以克服二进制编码引起的海明悬崖（Hamming Cliffs）问题又具有微调功能；便于和其他搜索技术相结合等。Michalewicz^[14]通过实验比较了实数编码和二进制编码的性能，结果表明实数编码的收敛速度和精度均好于二进制编码。

所以，本文采用基于实数编码遗传算法（RGA）。

3.3 优化控制策略

采用 SVM 与 RGA 相互耦联的方法优化控制发酵过程，如图 2 所示。对数生长期的目标函数是使菌体浓度最大，即： $\max X(t + \Delta t)$ 。稳定期的目标函数是使发酵单位最大，即： $\max P(t + \Delta t)$ 。利用生产现场的历史数据，采用 SVM 方法建立发酵过程模型，将此模型作为 RGA 的适应度函数。

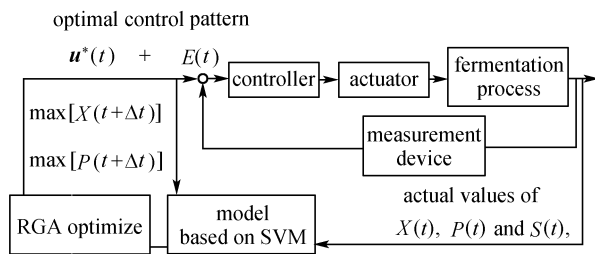


图 2 发酵过程的优化控制框图

Fig. 2 Block diagram of optimization control

在发酵过程的 t 时刻，根据生产经验，确定样本所属的发酵阶段；将输入样本输入相应的 SVM 模型；RGA 计算出使目标函数最大的最优控制模式 $u^*(t)$ ；将 $u^*(t)$ 给定控制器作为给定值，然后对整个发酵过程进行控制。在生化室内，离线测量菌体浓度 $X(t)$ 、产物浓度 $P(t)$ 和基质浓度 $S(t)$ ，测量每个样本所需要的时间约为 Δt 。 t 时刻泛指发酵过程的每个时刻。

SVM 与 RGA 相互耦联的优化控制策略是依据一批产率高的历史数据，将其初始条件作为实际

生产的初始条件，求取最优控制模式 $u^*(t)$ 的方法如下。

Step1 从生产现场获取能够体现发酵过程特点和产率高的历史数据。根据建立发酵过程模型的要求，对原始数据进行分析、变量选择以及预处理。数据的预处理主要指将补料累积量换算成补料瞬时量，即补料速率。

Step2 将发酵数据按照采样间隔 Δt 分割成若干数据样本。将样本划分为两类：训练样本和测试样本。

Step3 采用 SVM 方法建立发酵过程模型，采用 MER 计算拟合误差。

Step4 采用测试样本测试模型的泛化能力，使用 MER 计算预测误差。如果预测误差不满足要求，就返回 Step3 重新建模。

Step5 将 t 时刻的离线测量值 $X(t)$ 、 $P(t)$ 和 $S(t)$ 以及 $u(t)$ 作为输入样本，判断此样本所属的发酵阶段，同时能够确定出此阶段所采用的目标函数。

Step6 RGA 对待寻优控制模式 $u(t)$ 进行实数编码，生成初始群体。

收敛条件：当种群进化稳定，即进化种群中最大适应度值和平均适应度值的差小于某个常数时，算法终止运算。为了防止迭代永不停止，规定最大进化代数为 500。

Step7 SVM 模型为适应度函数，所以适应度函数被定义为

$$F = f[x(t), u(t)] \quad (13)$$

其中， $x(t)$ 为状态模式， $u(t)$ 为控制模式。计算个体的适应度值，评价是否满足收敛条件。

Step8 如果满足收敛条件，输出结果 $u^*(t)$ ；否则继续执行 Step9。

Step9 RGA 执行遗传操作，生成新一代群体，并转向 Step7。

特别需要指出的是： $u^*(t)$ 每隔 Δt 计算一次并作为新的控制系统给定值，在这里忽略了 RGA 求取 $u^*(t)$ 的时间。

4 实验研究

采用青霉素发酵过程验证 SVM 与 RGA 相互耦联的优化控制策略的有效性。

4.1 青霉素发酵过程

青霉素大规模生产过程中，除了大量繁殖菌丝

体外，主要是生产青霉素。生产菌种产黄青霉 (*P. chryso-genum*) 利用葡萄糖和氨，经由 α -氨基己二酸、半胱氨酸和缬氨酸组成的三肽，环化生成异青霉素 N，再与苯乙酸或苯氧乙酸进行转酰基反应，产生青霉素 G 或青霉素 V^[2]。

河北某制药公司青霉素发酵生产系统如图 3 所示。该发酵系统采用 SIEMENS 的 PLC 控制系统，可实时检测温度、pH 值、通气量、搅拌转速、压力、溶解氧等参数，同时包括温度、搅拌转速、进出气流量、pH 值、补料等控制回路。



图 3 青霉素发酵系统

Fig. 3 Penicillin fermentation system

培养基组分主要包括：葡萄糖、花生饼粉、麸质水、尿素、硝酸铵、硫代硫酸钠、苯乙酰胺和硫酸钙。按照发酵工艺提供的培养基配方，进行补料分批发酵生产。发酵周期约 220 h，每隔 4 h 取样一次，检测项目为葡萄糖浓度和效价。

4.2 青霉素发酵过程模型概述

青霉素是青霉素产生菌的次级代谢产物。青霉素发酵属于半连续发酵方式，是一个非常复杂的多变量耦合的生物化学过程，具有高度的非线性、时变性和不确定性，很难给出一个精确的数学模型^[16]。影响青霉素发酵产率的因素很多，主要有温度、pH 值、基质浓度、溶氧饱和度、菌体浓度等。通过对河北某制药公司提供的生产数据的分析，青霉素发酵效价预估模型的输入变量采用对效价影响大的因素，分别是 t 发酵时刻的补糖速率、补氨速率、补苯乙酸速率、补硫铵速率、温度、pH 值、空气流量、搅拌电流和效价。输出变量是 $t+\Delta t$ 发酵时刻的效价， Δt 取 4 h。

4.3 青霉素发酵过程建模

为了体现支持向量机具有“小样本”的特性，

本文只采用 2 批生产数据，每批都表示一个完整的发酵过程。A 批数据作为训练数据，从第一次生化室取样开始，每隔 4 h 获得一个样本，每个样本包含了此时刻的温度、补糖速率、pH 值、效价等 9 个参数值，共计 55 个样本。B 批数据作为测试数据，共计 54 个样本。为反映发酵过程的非线性特点，采用径向基核函数，因此 SVM 的参数主要有不敏感系数 ϵ 、惩罚系数 C 和宽度系数 σ 。

模型结构为

$$y(t+\Delta t) = f(y(t), x_1(t), x_2(t), \dots, x_8(t)) \quad (14)$$

其中， $y(t)$ 代表 t 时刻效价； $x_1(t) \sim x_8(t)$ 分别代表其他 8 个模型输入变量； $y(t+\Delta t)$ 代表模型输出，即 $t+\Delta t$ 时刻的效价。

在 Matlab 6.5 软件中，使用 Gunn 编写的 Matlab 支持向量机工具箱^[17]以及自编程序实现青霉素发酵过程建模和测试。根据经验选择 SVM 参数，经过多次训练和测试后，参数为 $\sigma=25000$ 、 $C=8000$ 和 $\epsilon=1000$ 。A 批数据训练的结果如图 4 所示，支持向量数是 29，训练相对误差的均值 MER 是 0.02558。用训练好的模型来预估 B 批数据的实验结果如图 5 所示，预测相对误差的均值 MER 是 0.023591。从图 5 可以看出模型的预测能力很强，预测值与实际测量值基本吻合。从图 4 和图 5 可以看出，SVM 方法所建立的模型在测试数据上的表现略高于在训练数据上的表现。所以，SVM 方法适合为发酵过程建立模型。

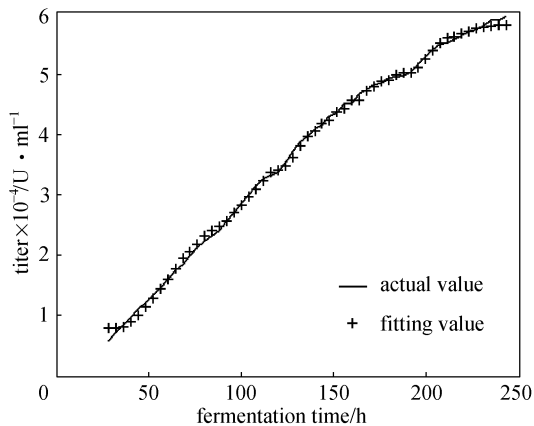


图 4 A 批数据训练结果

Fig. 4 Training results with batch A data

4.4 青霉素发酵过程的优化控制

以上面建立的模型为基础，采用 RGA 优化控制青霉素发酵过程。需要优化的发酵参数包括：补糖速率、补氨速率、补苯乙酸速率、补硫铵速率、

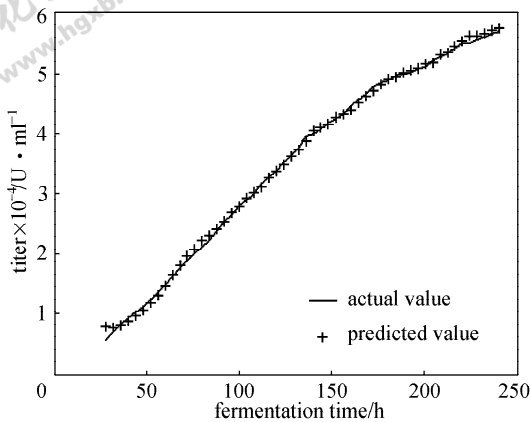


图 5 B 批数据预估结果

Fig. 5 Testing results with batch B data

温度、pH 值、空气流量、搅拌电流。这些参数的变化范围分别为： $50 \sim 200 \text{ kg} \cdot \text{h}^{-1}$ ； $0 \sim 50 \text{ L} \cdot \text{h}^{-1}$ ； $0 \sim 50 \text{ kg} \cdot \text{h}^{-1}$ ； $0 \sim 200 \text{ L} \cdot \text{h}^{-1}$ ； $25 \sim 29^\circ\text{C}$ ； $5.5 \sim 7.0$ ； $6000 \sim 8000 \text{ m}^3 \cdot \text{h}^{-1}$ ； $350 \sim 500 \text{ A}$ 。

由于青霉素发酵过程的对数生长期和稳定期相互耦联，在对数生长期，菌体就开始合成青霉素，在稳定期，菌体仍在生长和繁殖。所以，目标函数取一个，即：使产物浓度最大。

在 Matlab6.5 软件中，以 B 批发酵生产的初始条件为优化控制的初始点，计算最优发酵控制模式。青霉素发酵属于半连续发酵方式，在发酵过程中会间歇放掉部分发酵液。这样，只要能够遵守补料速率的取值范围，就不用考虑发酵液的体积。

RGA 参数设置：实数编码的精度为 10^{-2} ；种群规模 $N=100$ ；采用算术交叉算子，每代执行交叉操作的次数为 4；采用多重非均匀变异，其参数包括每代执行变异操作的次数、最大进化代数 and 系统参数，它们的值分别为 6、100 和 2。收敛条件：当种群进化稳定，即进化种群中最优适应度值和平均适应度值的差小于 100 时，算法终止运算。为了防止迭代永不停止，规定最大进化代数为 500。

为了节约篇幅，只列出了部分最优发酵参数曲线图。整个发酵过程的最优 pH 值如图 6 所示。青霉素发酵的 pH 一般控制在 $5.5 \sim 7.0$ ，由于青霉素在碱性条件下不稳定，容易加速其水解，因此，应尽量避免 pH 超过 7.0。通过采用下列手段控制发酵过程的 pH：①当 pH 高于最适 pH 时，可以通过补加糖和生理酸性物质（如硫酸铵等无机氮源）降低 pH；②当 pH 低于最适 pH 时，可以通过补加氨水或者尿素，也可提高通气量，促进有机

酸氧化来提高 pH；③可利用自动加入酸或碱的方法，使发酵液 pH 维持在 $5.5 \sim 7.0$ ，以提高青霉素产量。用补糖控制 pH 要比用酸或碱调节的效果好。

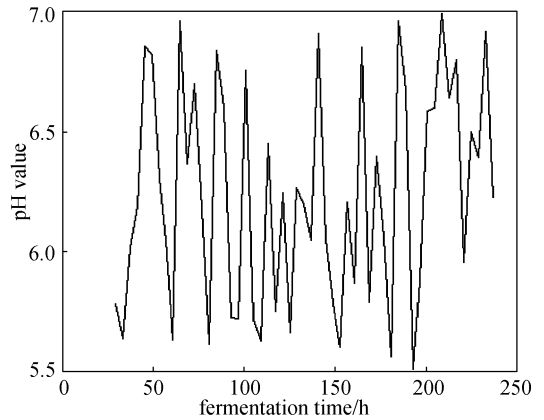


图 6 最优 pH 值轨线

Fig. 6 Optimal trajectory of pH value

最优补糖速率如图 7 所示。采用连续流加葡萄糖的方法实现，避免了葡萄糖的分解对产物产生的阻遏作用，也使发酵液 pH 值较稳定，有利于青霉素的合成。

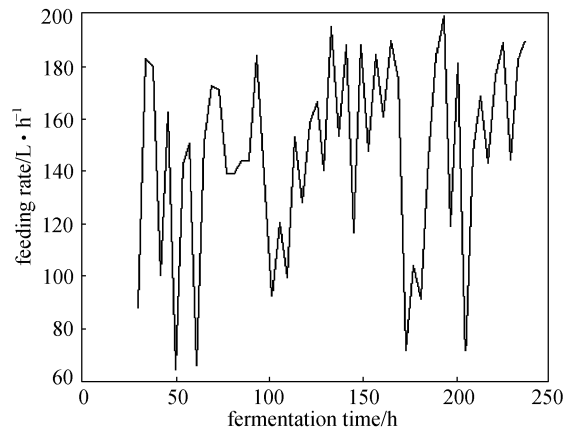


图 7 最优补糖速率轨线

Fig. 7 Optimal trajectory of feeding sugar rate

通过对 pH 和补糖速率的控制的分析得出，发酵参数之间是相互耦合的，一个控制作用能够引起多个参数的变化。例如，补糖不但能调节 pH 值，而且还能补充碳源；改变通气量不但能调节溶解氧，而且还能影响 pH 值。所以，本文将需要优化的发酵参数构成控制模式同时进行优化，这样就能够避免各个参数之间的相互耦合，也就没有必要再去判断某一控制作用到底影响了哪一个发酵参数。

如图 8 所示，以最后一个寻优时刻为例，

RGA 在第一代就找到了最大适应度值, 经过两代后最大适应度值和平均适应度值的差小于 100, 这满足了收敛条件。横坐标代表 RGA 的进化代数。由于在第一代就找到了全局最大值, 所以最大值没有变化。随着种群的进行, RGA 的平均适应度值递增。由于只进化了两代, 两点成线, 所以看起来平均值是线性的。如果按照最优控制模式控制整个发酵过程, 生产结果如图 9 所示, 最终效价比没有采用优化控制策略的生产结果提高了 22.88%。

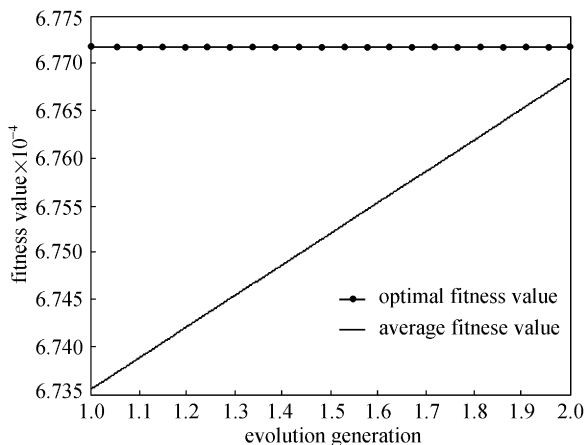


图 8 RGA 进化关系曲线

Fig. 8 Evolution trace of RGA

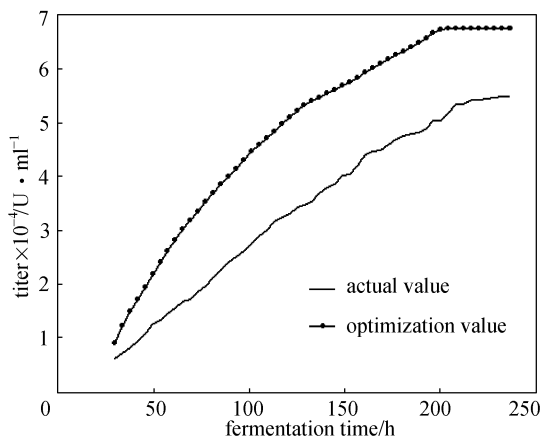


图 9 优化控制前后青霉素效价比较

Fig. 9 Penicillin titer compared between fermentation optimized and not optimized

5 结 论

为了提高微生物发酵单位, 本文提出了 SVM 与 RGA 相互耦联的优化控制策略。同时, 引入了模式的概念, 避免了发酵参数之间的耦合。SVM 建立微生物发酵过程的预估模型, RGA 基于此模

型计算最优控制模式。将此策略应用于青霉素发酵过程, 实验结果表明, SVM 模型具有很高的拟合精度和很强的泛化能力; 基于此模型 RGA 寻优得到了最优控制模式, 控制后的效价比没有进行优化控制的生产结果提高了 22.88%, 充分证明了此策略的有效性。

符 号 说 明

- F_s —— 补料速率, $L \cdot h^{-1}$
- $P(t)$ —— t 时刻的产物浓度, $U \cdot ml^{-1}$
- $P(t+\Delta t)$ —— $t+\Delta t$ 时刻的产物浓度, $U \cdot ml^{-1}$
- $S(t)$ —— t 时刻的基质浓度, $g \cdot L^{-1}$
- $S(t+\Delta t)$ —— $t+\Delta t$ 时刻的基质浓度, $g \cdot L^{-1}$
- T —— 发酵温度, $^{\circ}C$
- Δt —— 时间间隔, h
- $X(t+\Delta t)$ —— $t+\Delta t$ 时刻的菌体浓度, $g \cdot L^{-1}$

References

- [1] Qian Mingyong (钱铭镛). Optimization Control of Fermentation Engineering (发酵工程最优化控制). Nanjing: Jiangsu Science and Technology Publishing House, 1998: 1-10
- [2] Qi Xiangjun (齐香君). Modern Biopharmaceutical Technology (现代生物制药工艺学). Beijing: Chemical Industry Press, 2003
- [3] Holland J H. Adaptation in Natural and Artificial Systems. Michigan: The University of Michigan Press Ann Arbor, 1975
- [4] Tu Zhenguo, Lu Yong. A robust stochastic genetic algorithm (StGA) for global numerical optimization. *IEEE Trans. on Evolutionary Computation*, 2004, **8** (5): 456-470
- [5] Chiu C C, Hsu P L. A constraint-based genetic algorithm approach for mining classification rules. *IEEE Trans. on SMC*, 2005, **35** (2): 205-220
- [6] Zuo K, Wu W T. Semi-realtime optimization and control of a fed-batch fermentation system. *Computers and Chemical Engineering*, 2000, **24** (2): 1105-1109
- [7] Becker T, Enders T, Delgado A. Dynamic neural networks as a tool for the online optimization of industrial fermentation. *Bioprocess Biosyst. Eng.*, 2002, **24** (2): 347-354
- [8] Fang Lide (方立德), Zhang Tao (张涛), Luo Yi (罗翼). Wet gas correlation for horizontal mounted Venturi based on neural network. *Journal of Chemical Industry and Engineering (China)* (化工学报), 2007, **58** (4): 957-962

- [9] Gao Xuejin (高学金), Wang Pu (王普), Sun Chongzheng (孙崇正), Yi Jianqiang (易建强), Zhang Yating (张亚庭), Zhang Huiqing (张会清). Research and application of method on modeling by dividing phase for fermentation process. *Journal of System Simulation* (系统仿真学报), 2007, **19** (11): 2574-2577, 2598
- [10] Vapnik V N. The Nature of Statistical Learning Theory. New York: Springer Verlag, 1995
- [11] Vapnik V N. An overview of statistical learning theory. *IEEE Trans. on Neural Network*, 1999, **10** (5): 988-999
- [12] Cao L J, Tay F E H. Support vector machine with adaptive parameters in financial time forecasting. *IEEE Trans. on Neural Networks*, 2003, **14** (6): 1506-1518
- [13] Deng Naiyang (邓乃扬), Tian Yingjie (田英杰). A New Method of Data Mining: Support Vector Machine (数据挖掘中的新方法——支持向量机). Beijing: Science Press, 2004
- [14] Michalewicz Z. Genetic Algorithms + Data Structure = Evolution Programs. New York: Springer-Verlag, 1996
- [15] Zou Lin, Xia Juchen, Hu Guo'an. Real coding based multi-population parallel genetic algorithm. *Mini-micro Systems*, 2004, **25** (6): 982-986
- [16] Wang Shuqing (王树青), Yuan Yingjin (元英进). Automation Technology of Biochemistry Process (生化过程自动化技术). Beijing: Chemical Industry Press, 1999: 223-224
- [17] Gunn S R. MATLAB Support Vector Machine Toolbox [DB/OL]. [1997-10-11]. <http://www.isis.ecs.soton.ac.uk/resources/svminfo.html>