

文章编号: 1000-5013(2013)01-0022-04

# 基于相关向量机的赖氨酸反应过程参数软测量

嵇小辅, 张孟尧, 王博, 黄丽

(江苏大学 电气信息工程学院, 江苏 镇江 212013)

**摘要:** 针对支持向量机(SVM)计算复杂度高、参数不容易确定等局限性,提出一种基于相关向量机(RVM)的赖氨酸反应过程关键参量的软测量方法.根据过程经验,确定发酵液的溶解氧浓度、pH 值、二氧化碳释放率、氧吸收率和葡萄糖流加速率为辅助变量,利用相关支持向量机的拟合与泛化能力,建立了赖氨酸反应过程基质浓度、菌体浓度、产物浓度等不可直接测量参量的软测量模型.基于 *L*-赖氨酸反应过程开展的试验研究表明:所建立的相关向量机软测量模型拟合精度高、泛化能力强,较好地满足了赖氨酸反应过程的控制要求.

**关键词:** *L*-赖氨酸; 相关向量机; 软测量; 贝叶斯方法

**中图分类号:** TQ 922.3; TQ 018

**文献标志码:** A

赖氨酸发酵过程是一个多变量、非线性、强耦合的复杂动态过程.目前使用的测量方法主要是在线取样,离线分析的方法,存在较大的测量延时,难以实施在线实时控制,同时在线取样容易引入人为污染,降低发酵过程品质.在现有技术条件下,一些直接反映发酵品质的基质浓度、菌体浓度、产物浓度等关键生物量参数目前还缺乏在线实时测量的仪器与手段,成为制约在线控制的技术瓶颈.软测量是利用在线可测的辅助变量对难以在线测量的主导变量进行在线估计的技术,是解决生物反应过程测量瓶颈的有效途径.生物反应过程的软测量建模有很多方法,大体可分为机理建模、黑箱建模和混合建模三大类.黑箱建模是基于过程数据,采用数据拟合回归方法建立软测量模型.文献[1-4]分别提出了采用标准支持向量机(SVM)建立了赖氨酸反应过程的软测量模型,以及基于模糊最小二乘支持向量机、聚类动态最小二乘支持向量机、粒子群最小二乘支持向量机等软测量方法.为了解决支持向量机实际应用时存在的难题<sup>[5]</sup>,Michael 提出一种与支持向量机相似的稀疏概率模型,即相关向量机(RVM),其训练在贝叶斯框架下进行,克服了支持向量机的上述所有缺点,已被证实在线回归精度、泛化能力等方面优于前者.本文将 RVM 引入赖氨酸反应过程的软测量,建立基质浓度、菌体浓度、产物浓度等不可直接测量参量的软测量模型,实现了赖氨酸反应过程不可直接测量的关键参量的在线实时测量.

## 1 相关向量机原理

给定训练数据样本集  $\{\mathbf{x}_i, t_i\}_{i=1}^n$ , 其中  $\mathbf{x}_i \in \mathbf{R}^n$  是输入变量,  $t_i \in \mathbf{R}$  是与  $\mathbf{x}_i$  对应的输出变量,是回归过程中的真实值.假设输出变量全部由带有白噪声的回归模型产生,即有

$$t_i = y(\mathbf{x}_i; \boldsymbol{\omega}) + \epsilon_i. \quad (1)$$

式(1)中:样本噪声  $\epsilon_i$  满足均值为 0,方差为  $\sigma^2$  的高斯分布,即  $\epsilon_i \sim N(\epsilon|0, \sigma^2)$ .

与 SVM 类似,RVM 将回归函数表示成基函数  $K(\mathbf{x}, \mathbf{x}_i)$  的线性组合形式,即

$$y_i = \sum_{i=1}^N \omega_j K(\mathbf{x}, \mathbf{x}_i) + \omega_0. \quad (2)$$

**收稿日期:** 2012-06-15

**通信作者:** 嵇小辅(1979-),男,副教授,主要从事生物反应过程软测量与优化控制、鲁棒控制等方面的研究. E-mail: xfji@msn.com.

**基金项目:** 江苏省自然科学基金资助项目(BK2011465);轻工过程先进控制教育部重点实验室开放课题资助项目(APCLI1009);江苏省高校优势学科建设工程科研基金资助项目(苏政发(2011)6号)

式(2)中:  $\omega = [\omega_0, \omega_1, \dots, \omega_M]^T$  是可调的权重向量;  $\omega_0$  是回归函数中的偏置. 由于噪声  $\epsilon_i$  满足高斯分布, 由式(1)可知:  $t_i$  也满足于均值  $y(x_i; \omega)$ , 方差  $\sigma^2$  的高斯分布, 即

$$p(t_i | x_i) = N(t_i | t(x_i; \omega), \sigma^2). \quad (3)$$

为了描述方便起见, 引入超参数  $\beta = 1/\sigma^2$ , 则整个训练数据集的似然函数为

$$p(t | \omega, \beta) = \left(\frac{\beta}{2\pi}\right)^{N/2} \exp\left(\frac{\beta \|t - \Phi\omega\|^2}{2}\right). \quad (4)$$

式(4)中:  $t = [t_1, t_2, \dots, t_N]^T$ ;  $\Phi \in \mathbf{R}^{N \times (N+1)}$  是设计矩阵, 定义为  $\Phi = [\Phi_1(x_1), \Phi_2(x_2), \dots, \Phi(x_N)]^T$ ; 基函数向量  $\Phi(x_i) = [1, K(x_i, x_1), \dots, K(x_i, x_N)]^T$ ,  $i = 1, 2, \dots, N$ , 即

$$\Phi = \begin{bmatrix} 1 & K(x_1, x_1) & K(x_1, x_2) & \cdots & K(x_1, x_N) \\ 1 & K(x_2, x_1) & K(x_2, x_2) & \cdots & K(x_2, x_N) \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & K(x_N, x_1) & K(x_N, x_2) & \cdots & K(x_N, x_N) \end{bmatrix}.$$

RVM 训练过程的目标是求取权值向量  $\omega$  的后验分布. 为了保证模型稀疏性, 需要定义权值  $\omega_i$  的先验分布. 假设  $\omega_i$  满足均值为 0, 方差为  $\alpha_i^{-1}$  的高斯分布, 则  $\omega$  的先验可以表示为

$$p(\omega | \alpha) = \prod_{j=0}^N N(\omega_j | 0, \alpha_j^{-1}). \quad (5)$$

式(5)中:  $\alpha = [\alpha_0, \alpha_1, \dots, \alpha_N]^T$  每个独立的超参数  $\alpha_j$  只与其对应的权值  $\omega_j$  相关. 根据贝叶斯公式, 利用样本似然函数(4)和  $\omega$  先验分布(5)可得  $\omega$  的后验分布为

$$p(\omega | t, \alpha, \beta) = \frac{p(\omega | \alpha) p(t | \omega, \beta)}{p(t | \alpha, \beta)}. \quad (6)$$

由于  $p(\omega | \alpha)$  和  $p(t | \omega, \beta)$  均满足高斯分布, 因此其乘积  $p(\omega | \alpha) p(t | \omega, \beta)$  也满足高斯分布, 而  $p(t | \omega, \beta)$  不含有参数  $\omega$ , 可以看作是归一化系数, 所以  $\omega$  的后验能进一步表示为

$$p(t | \omega, \beta) = N(\omega | \mu, \Sigma). \quad (7)$$

其协方差均值为

$$\Sigma = (\beta \Phi^T \Phi + \mathbf{A})^{-1}, \quad (8)$$

而其均值为

$$\mu = \beta \Sigma \Phi^T t. \quad (9)$$

式(9)中:  $\mathbf{A} = \text{diag}\{\alpha_0, \alpha_1, \dots, \alpha_N\}$ .

可以看出, 超参数  $\beta$  和  $\alpha_j$  直接影响  $\omega$  的后验分布, 需要对其进行优化, 从而获得  $\omega$  的最大后验分布. 优化过程可以依据贝叶斯证据框架, 通过最大化边缘似然  $p(\omega | \alpha, \beta)$  取负对数得到目标函数, 然后令目标函数分别对超参数  $\alpha_j$  和  $\beta$  求偏导并令偏导数为 0 而得到. 即

$$\alpha_j = \frac{1}{\mu_j^2 + \Sigma_{j,j}} = \frac{\gamma_j}{\mu_j^2}, \quad j = 0, 1, 2, \dots, N, \quad (10)$$

$$\beta = \frac{N - \sum_{j=0}^N \gamma_j}{\|t - \Phi \mu\|^2}. \quad (11)$$

式(10), (11)中:  $\mu_j$  为权值后验均值向量  $\mu$  的第  $j$  个元素;  $\Sigma_{j,j}$  为协方差矩阵  $\Sigma$  的第  $j$  个对角元素;  $\gamma_j = 1 - \alpha_j \Sigma_{j,j}$ . 在 RVM 模型训练过程中, 式(8)~(11)要依次迭代计算, 直到所有参数都收敛或者达到最大收敛次数为止, 这时可认为 RVM 建模过程完成.

## 2 赖氨酸反应过程软测量模型构建

赖氨酸软测量模型的精度和泛化能力与软测量模型输入数据的预处理有很大关系. 为了消除测量噪声, 首先对输入数据采用巴特沃思滤波器进行滤波处理, 产生一个平滑的样本数据. 同时, 由于各变量的变化范围差别很大, 因此采用最大最小值归一化方法对数据进行归一化处理, 将原始数据映射到区间  $[0, 1]$  内, 即

$$x_{s,i} = \frac{x_i - \min x_i}{\max x_i - \min x_i}.$$

(12)

式(12)中: $x_{s,i}$ 为标准化数据, $x_i$ 为原始数据; $\max x_i$ 和 $\min x_i$ 分别为原始数据的最大值和最小值.

基于滤波和归一化处理后的数据,对 RVM 模型进行学习训练,由此可以看出,RVM 学习过程是基于贝叶斯框架对权重的后验分布进行推理的过程. 具体算法有如下 6 个主要步骤:1) 初始化超级参数  $\alpha_i, i=1,2,\cdots,M$ ;2) 计算权重后验分布的均值  $\boldsymbol{\mu}$  和方差  $\boldsymbol{\Sigma}$ ;3) 计算  $\gamma_i, i=1,2,\cdots,M$ ,再估计  $\alpha_i, i=1,2,\cdots,M$ ;4) 重复步骤 2),3),直到所有的超级参数都收敛;5) 由于  $\alpha_i=\infty$ 对应的权重均值为  $\mu_i=0$ ,删除其对应的权重;6) 对于新的输入数据  $\mathbf{x}$ ,通过收敛的  $\alpha_{\text{MP}}$ 和  $\sigma_{\text{MP}}^2$ 对目标数据进行预测,所采用的预测分布函数为

$$p(t_m | \mathbf{t}) = \int p(t_m | \boldsymbol{\omega}, \alpha_{\text{MP}}) p(\boldsymbol{\omega} | \mathbf{t}, \alpha_{\text{MP}}, \sigma_{\text{MP}}^2) d\boldsymbol{\omega}.$$

3 仿真结果与分析

针对实验室的 WKT 型生物反应器,以 L-赖氨酸流加发酵过程为对象进行实验研究. 为使实验过程能客观反映实际生产过程,对实验和建模过程进行设计.

1) 根据前期对赖氨酸反应过程的机理和数据分析,赖氨酸反应过程的基质浓度、菌体浓度、产物浓度的软测量模型与溶解氧值、发酵液 pH 值、二氧化碳释放率、氧吸收率和葡萄糖流加速率等参量紧密相关. 因此,选择上述变量为 RVM 软测量模型的输入变量.

2) 每批次发酵周期 72 h,采样周期为 15 min,通过测试仪器对溶解氧值、发酵液 pH 值、二氧化碳释放率、氧吸收率、葡萄糖流加速率等参量进行实时采集,每 2 min 取样并离线化验得到基质质量浓度  $\rho$  (基质)、菌体质量浓度  $\rho$  (菌体)、产物质量浓度  $\rho$  (产物)的值. 其中:菌体质量浓度采用细胞干重法计算得到,即取 10 mL 发酵液于离心管中,在  $3\,000\text{ r}\cdot\text{min}^{-1}$ 下离心 5 min,弃上清,蒸馏水洗涤 2 次,在  $105\text{ }^{\circ}\text{C}$ 干燥至恒质量后称量;基质质量浓度采用 SBA-40C 型多功能生物传感器测定;产物质量浓度采用改进的茚三酮比色法进行测定,即取上清液 2 mL 加茚三酮试剂 4 mL 混合,沸水浴加热 20 min,冷却后通过分光光度计测定 475 nm 处的吸光度值,通过查标准 L-赖氨酸曲线得到.

3) 考虑 5 个批次培养以检验赖氨酸反应过程的 RVM 软测量模型,每批次间的初始条件设为不同,补料策略亦有相应变化,以增强各批次间的差异性. 发酵罐罐压控制在  $0\sim0.25\text{ MPa}$ ,发酵罐罐温控制在  $((0\sim50)\pm0.5)\text{ }^{\circ}\text{C}$ ,发酵前搅拌电机转速为  $400\text{ r}\cdot\text{min}^{-1}$ 时标定溶解氧电极的基准读数,其中前 4 批次发酵数据作为训练样本集,离线训练获得 RVM 软测量模型,第 5 个批次用于检验 RVM 软测量模型的泛化能力.

采用训练好的 RVM 模型对第 5 批数据进行泛化能力检验. 基质质量浓度  $\rho$  (基质)、菌体质量浓度  $\rho$  (菌体)、产物质量浓度  $\rho$  (产物)的预测曲线,如图 1~3 所示. 从图 1~3 可以看出:所建立的 RVM 模型具有拟合精度高、泛化能力强的优点.

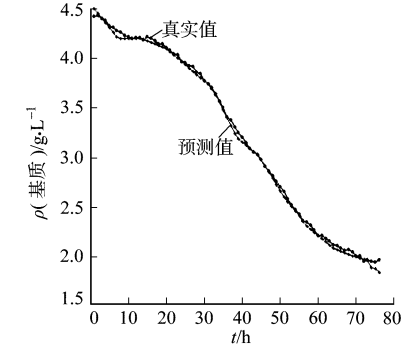


图 1 基质浓度预测曲线图  
Fig.1 Predicted value of  
glucose concentration

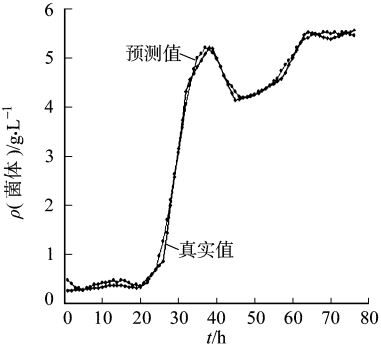


图 2 菌体浓度预测曲线  
Fig.2 Predicted value of  
cell concentration

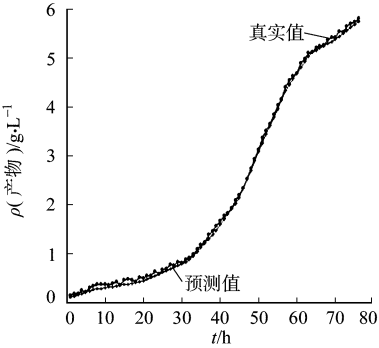


图 3 产物浓度预测曲线  
Fig.3 Predicted value of  
product concentration

## 4 结论

在分析赖氨酸反应过程特性的基础上,提出一种基于相关向量机的赖氨酸反应过程基质浓度、菌体浓度和产物浓度的软测量模型. 由于采用贝叶斯方法进行概率预测,相对于支持向量机稀疏性更好,所需要的核函数向量个数更少,因此测试时间更短,可以实现赖氨酸反应过程的实时、在线测量.

### 参考文献:

[1] 孙玉坤,陈明忠,嵇小辅,等. 基于支持向量机的赖氨酸发酵生物参数软测量[J]. 仪器仪表学报,2008,29(10):2067-2071.

[2] 王博,嵇小辅,孙玉坤. 基于自适应模糊支持向量机的 *L*-赖氨酸发酵过程建模研究[J]. 仪器仪表学报,2010,31(8):467-481.

[3] 王博,孙玉坤,嵇小辅,等. 基于 PSO-SVM 逆的 *L*-赖氨酸发酵过程软测量方法[J]. 化工学报,2012,26(3):224-227.

[4] 孙玉坤,王博,黄永红,等. 基于聚类动态 LS-SVM 的 *L*-赖氨酸发酵过程软测量方法[J]. 仪器仪表学报,2010,24(2):1024-1028.

[5] TIPPING M E. Sparse Bayesian learning and the relevance vector machine[J]. Journal of Machine Learning Research,2001,1(3): 211-244.

## Soft Sensor of Lysien Fermentation Biological Parameters Based on Relevance Vector Machine

JI Xiao-fu, ZHANG Meng-yao, WANG Bo, HUANG Li

(School of Electrical and Information Engineering, Jiangsu University, Zhenjiang 202013, China)

**Abstract:** To overcome the high computational complexity and difficulty in design kernel parameters of support vector machine, the soft sensor model of lysien fermentation biological parameters is proposed based on relevance vector machine. According to procedure experience, dissolved oxygen, the parameters of pH value, discharge rate of CO<sub>2</sub>, absorption rate of O<sub>2</sub>, and flow acceleration rate of glucose are chosen as auxiliary variables, and the concentration of matrix, thallus and product are chosen as primary variables. The soft sensor of lysien fermentation procedure is built using the fitting and generalization capacity. The experiments results show that the obtained soft sensor is with high fitting precision and generalization capacity, which well satisfies the control requirement of lysien fermentation procedure.

**Keywords:** *L*-lysien; relevance vector machine; soft sensor; Bayesian method

(责任编辑: 黄晓楠 英文审校: 吴逢铁)