

文章编号: 1674-2869(2016)04-0333-04

# 基于LS-SVM苯乳酸发酵过程的建模

张正风

徐州生物工程职业技术学院信息中心, 江苏 徐州 221006

**摘要:**为了解决苯乳酸发酵过程中关键生物参数难以直接在线检测的问题,提出了基于最小二乘支持向量机(LS-SVM)的软测量建模方法.通过使用径向基核函数来对菌体浓度、苯乳酸浓度建立模型,对建模的理论进行了分析和并进行了仿真研究,同时还采用支持向量机对过程进行了建模,对两种方法的优缺点进行了比较.结果表明,基于LS-SVM的建模方法预测精度高、跟踪性能好,能提高在线预估的效率,非常适合于苯乳酸发酵过程的在线预估.

**关键词:**发酵;建模;径向基核函数;支持向量机;最小二乘支持向量机

**中图分类号:**TQ92      **文献标识码:**A      **doi:**10.3969/j.issn.1674-2869.2016.04.005

## Modeling of Phenylactic Acid Fermentation Process Based on Least Square Support Vector Machine

ZHANG Zhengfeng

Information Center of Xuzhou Vocational College of Bioengineering, Xuzhou 221006, China

**Abstract:** To solve the difficulties of online measurement for crucial biological variables in the phenylactic acid fermentation process, a soft sensor modeling method was proposed based on the least squares support vector machine (LS-SVM), and the model for concentration of mycelium and phenylactic acid was built by kernel of Radial Basis Function. Theoretical analysis and simulation study of the modeling was investigated, and a second modeling process was constructed by the support vector machine. Finally, the effects of the two methods modeling were compared. The results show that the modeling method based on the LS-SVM has the advantages of accuracy prediction, good tracking performance, improving efficiency of on-line prediction, which is very suitable for the on-line estimation of the phenylactic acid fermentation process.

**Keywords:** fermentation; modeling; kernel of Radial Basis Function; SVM; LS-SVM

## 1 引言

苯乳酸是近几年才开始研究的一种新型高效生物防腐剂,它的发酵过程与其他生化反应一样,具有非线性和不确定性的特点.要想掌握苯乳酸的发酵过程,并且对其进行进一步的优化和控制,就要能够获得充足的发酵环节数据.在线测定发酵过程的重要状态信息,在目前常用的测量条件和仪器下,难度较大,甚至几乎是不可能的.因此,离线化验一直是目前测定菌体浓度、基质浓度和

产物浓度等的最常用方法.然而离线测定有很多缺点,比如测量结果滞后,限制了过程优化控制方案的实施.

国内对苯乳酸的研究刚起步,目前关于苯乳酸的研究成果大多集中于苯乳酸的乳杆菌株的筛选<sup>[1]</sup>,有关其发酵工艺的研究很少.苯乳酸发酵生产过程可持续20~100 h不等<sup>[2]</sup>,在整个过程中,菌体和产物浓度不断变化,互相影响和抑制.这导致人们对其发酵过程认识不足,不能精确地进行过程控制,因此,要在估算方法和估算精度上下功

收稿日期:2016-04-11

基金项目:徐州生物工程职业技术学院科研课题项目(2014B04);第二期江苏省职业教育教学研究课题项目(ZY94)

作者简介:张正风,硕士,讲师. E-mail: 4508454@qq.com

夫, 本文采用基于LS-SVM的软测量技术来改善估算方法, 对苯乳酸发酵的生产过程进行建模研究。

最小二乘支持向量机(LS-SVM)是在标准支持向量机(SVM)的基础上提出的一种扩展和改进算法<sup>[3]</sup>。标准SVM算法用求解二次规划的方法, 力争在现有的有限采样数据的条件下推出最优的结果, 而且SVM拥有很多优良的特性, 如不存在局部最小问题, 以及强大的小样本学习和泛化能力。但是随着样本数据规模的增大, 相应的二次规划问题会变得非常复杂, 一个直接的问题就是导致了计算速度会越来越慢。LS-SVM通过一个线性方程组来进行相应的训练, 有严格的数学理论支撑, 优势相当明显, 算法精度高, 速度快, 且易于实现, 因而在模式识别和非线性建模领域的应用度较高<sup>[4]</sup>。

## 2 最小二乘支持向量机

设有如下训练样本集, 其中为 $n$ 维的样本输入向量, 为相应的样本输出,  $n$ 为样本数目。构造如式(1)所形式的函数来拟合样本<sup>[5]</sup>:

$$f(x) = \omega^t \varphi(x) + b \quad (1)$$

其中,  $\omega \in R^n$ , 是原始空间的权向量,  $b$ 是偏差量, 并且使用非线性映射  $\varphi(\cdot): R^n \rightarrow H$  将原来的输入空间  $R^n$  映射到高维 Hilbert 空间  $\varphi(x)$ , 特征空间的维数不固定, 有可能为无穷维。此函数的目的是求解满足条件的  $\omega^t$  和  $b$ 。利用结构风险最小化原则, 降低学习机器的 VC 维, LS-SVM 回归预测可以表示为如式(2)所示的约束优化问题<sup>[6]</sup>:

$$\min_{\omega, \xi} J(\omega, \xi) = \frac{1}{2} \omega^t \omega + c \sum_{i=1}^k \xi_i^2 \quad (2)$$

约束条件:  $y_i = \varphi(x_i) \cdot \omega + b + \xi_i, i = 1, \dots, k$ 。

其中  $\omega^t$ 、 $\omega$  用来控制模型的复杂度,  $c(c>0)$  表示函数的平滑过渡和允许误差大于  $x$  的数值之间的折衷值<sup>[6]</sup>。采用线性回归的方法, 相应的损失函数就是平方损失函数, 也就是计算误差  $\xi$  的平方和来作为度量和调整因选择不同参数导致的损失。

求解这个优化的问题可以用拉格朗日法:

$$L(\omega, b, \xi, a, \gamma) = \frac{1}{2} \omega^t \omega + c \sum_{i=1}^k \xi_i^2 - \sum_{i=1}^k a_i (\omega \cdot \varphi(x_i) + b + \xi_i - y_i) \quad (3)$$

对式(3)的变量  $(\omega, b)$  求偏导<sup>[7]</sup>, 进一步得到优化条件:

$$\frac{\partial L}{\partial \omega} = 0, \frac{\partial L}{\partial b} = 0, \frac{\partial L}{\partial \xi} = 0, \frac{\partial L}{\partial a} = 0 \quad (4)$$

根据式(4)可以得到:

$$\begin{aligned} \omega &= \sum_{i=1}^k a_i \varphi(x_i), \sum_{i=1}^k a_i = 0, a_i = c \xi_i \\ \omega \cdot \varphi(x_i) + b + \xi_i - y_i &= 0 \end{aligned} \quad (5)$$

定义满足 Mercer 条件的线性核函数  $K(x_i, y_j) = \varphi(x_i) \cdot \varphi(x_j)$ , 该核函数的作用是用来替代内积, 再对其进行进一步进行隐式非线性变换, 从而顺利地解决了维数灾难问题<sup>[8]</sup>。根据式(5), 优化问题最终演变为非线性模型:

$$f(x) = \sum_{i=1}^k a_i K(x, x_i) + b \quad (6)$$

通过推导可知, 把待求解的优化问题转化为线性方程的求解问题, 因此, 优化问题的算法复杂度就大大的降低。根据推导过程还能看出, LS-SVM 需要预先确定的参数为正则化参数  $C$  和核函数参数  $\delta$ , 而对于标准的 SVM 算法, 以 RBF 核函数为例, 标准 SVM 的参数为  $\varepsilon$ 、 $C$ 、 $\delta$  三个, 因此 LS-SVM 算法需要预先确定的参数少, 这也体现了 LS-SVM 算法的可操作性更强。

## 3 结果与讨论

使用 LS-SVM 来进行苯乳酸发酵过程的建模, 可测变量  $X$  通常为溶氧量、pH 值、温度、搅拌功率、发酵液体积等; 输入变量  $u$  为通气量、热水流速、冷水流速等; 可测输出变量为  $CO_2$  浓度、反应热量等; 不可测量变量主要包括: 菌体浓度、苯乳酸浓度、基质浓度<sup>[9]</sup>。

输入变量的选择是直接影响软测量结果的重要因素, 因此要在各个输入变量中找出与模型关系非常直接和重要的一组输入变量作为 LS-SVM 测量模型的输入参数, 不可测变量作为软测量模型的输出, 使用不同的参数来实现输入输出的非线性函数关系。

在用神经网络和向量机建立菌体浓度估计模型时, 刘毅<sup>[9]</sup>、张蓓<sup>[10]</sup>等人的研究表明, 由于微生物生长速率和氧气消耗量有很大关系, 所以最重要的辅助变量是溶氧利用率  $DO(t)$ 。同时发酵底物和发酵产物还将产生比较明显的抑制作用, 这里选择在  $t$  时刻实际物理意义较大的苯乳酸浓度  $P(t)$ 、基质浓度  $K(t)$  和菌体浓度  $W(t)$  纳入输入变量中。另外, 为了减小建模的复杂度, 不采用变化量较小的温度以及 pH 值这两个变量。

除了输入参数的选择外, 在支持向量机研究领域, 另一个要事先确定的重要问题是模型选择。本文使用 RBF 核函数来进行 LS-SVM 的建模, RBF 核函数输入参数有正则化参数  $C$  和核函数参数  $\delta$

两个参数,为了得到最佳模型选择,要力争寻找到最佳的参数搭配.使用最小二乘支持向量机的一个非常明显的优势就是求解速度快,所以可以很快的对数据进行训练.首先确定正则化参数集 $SC=\{C1,C2,\cdots,Cn\}$ 和核参数 $S\delta=\{\delta1,\delta2,\cdots,\delta n\}$ ,再对这些核函数输入参数集分别进行组合,使用组合参数对LS-SVM进行反复训练,最后用确认集对参数的选择进行筛选和优化,从而选出最佳参数组合作为使用LS-SVM进行苯乳酸发酵建模模型的最终参数.

苯乳酸发酵过程在5L生物发酵罐中进行,建模过程中所使用的数据都是来自发酵过程中的实际采样数据,由于其生产规模小,采样间隔时间较长,所以发酵实验获取到的原始数据量比较少,为了弥补这个缺陷,根据生物发酵过程两次采样之间的数据变化不大的特点,使用3次平滑样条插值拟合<sup>[11]</sup>的方法来适当填充丢失的实验数据.通过以上的数据测量和平滑补差,得到了用于发酵实

验建模的数据集合.

因为待用的建模数据尺度不统一,不同的数据可能有数量级上的差距,如果直接计算,大数的变化会掩盖小数的变化,使其不能反映小的测量值的变化,同时,收敛速度也比较慢.所以在得到实验数据之后,还要对数据进行归一化处理.数据归一化主要有两种方法<sup>[12]</sup>,一是把数据变为(0,1)之间的小数,另一个方法是把有纲量表达式变为无纲量表达式.经过归一化处理后,数据就变得更加容易训练和学习.

本次实验一共采集10个批次的发酵数据,每个完整的发酵批次时间跨度30 h,任意选择8个批次用于软测量模型误差的最小训练,其余2批作为测试集对模型进行检验,对菌体浓度、苯乳酸浓度分别使用SVM和LS-SVM两种方法进行建模预测.在不同批次样本数据下,LS-SVM和SVM建模性能的比较见表1.

在对菌体浓度的预测中,LS-SVM方法选择的

表1 LS-SVM与SVM预测性能比较

Tab. 1 Comparison of prediction performance of LS-SVM and SVM

| measurement object              | sample size | training time/t |        | MSE       |           |
|---------------------------------|-------------|-----------------|--------|-----------|-----------|
|                                 |             | SVM             | LS-SVM | SVM       | LS-SVM    |
| cell concentration              | 90          | 21.43           | 0.353  | 0.003 043 | 0.003 165 |
|                                 | 130         | 56.91           | 0.476  | 0.002 067 | 0.002 149 |
| phenyllactic acid concentration | 90          | 19.62           | 0.317  | 0.001 257 | 0.001 317 |
|                                 | 130         | 47.59           | 0.405  | 0.001 094 | 0.001 160 |

参数为: $C90=15\ 500$ , $\delta90=1.50$ ;  $C130=15\ 000$ , $\delta130=1.70$ . SVM方法参数选择均为: $C=1\ 000$ , $\varepsilon=0.000\ 1$ , $\sigma=0.5$ .在对苯乳酸浓度的预测中,LS-SVM的模型参数为: $C100=16\ 500$ , $\delta100=1.34$ ;  $C150=16\ 000$ , $\delta150=1.55$ ,SVM模型的参数均选择为: $C=2\ 000$ , $\varepsilon=0.000\ 1$ , $\sigma=0.5$ .图1和图2分别是部分数据的训练结果和测试结果.

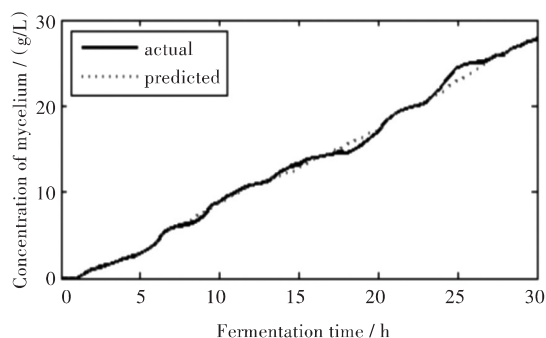


图1 LS-SVM模型的训练曲线  
Fig. 1 Training curve of LS-SVM model

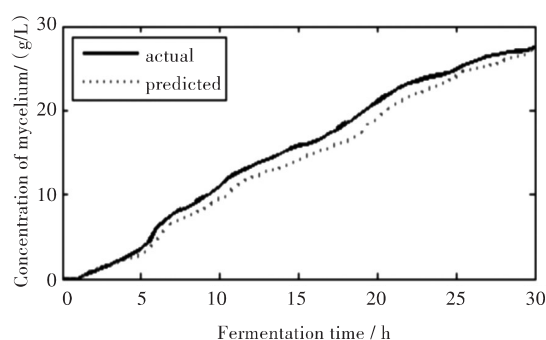


图2 LS-SVM模型的测试曲线  
Fig. 2 Test curve of LS-SVM model

### 3.2 建模结果分析

表1是在同样的训练集以及训练条件下,采用LS-SVM与SVM算法的预测性能比较,文中采用了训练时间和均方差作为衡量建模效果优劣的两个指标.图1和图2分别是对菌体浓度和苯乳酸浓度预测的训练曲线和预测曲线.从图1和图2可以

看出,通过定义合理的参数,LS-SVM能够把误差控制在一个合理的范围内,也使LS-SVM算法具有了相当高的拟合精度和泛化能力.继续观察和分析表1的数据可以看出,同样的样本数下,LS-SVM模型训练需要的时间更少,效率更高;通过均方差比较可以看出,两种算法的均方差都控制在非常低的水平,说明实验的数据精度都较高.

## 4 结 语

苯乳酸发酵过程各个变量之间相互制约,是一个高度非线性的过程.本文使用LS-SVM算法,对苯乳酸发酵过程进行仿真建模,并且和SVM算法进行比较.通过数据和图表等实验结果表明,LS-SVM是一种十分有效的仿真算法,该算法预测精度高、跟踪性能好,并且能大幅提高在线预估的效率,适合于对苯乳酸发酵过程的建模,并且对于其他类似的需要在线预估的场合,该方法也有一定的参考价值.

### 参考文献:

- [1] 高学金,王普,孙崇正,等.基于动态SVM的发酵过程建模[J].仪器仪表学报,2006,27(11):1497-1500.  
GAO X J, WANG P, SUN C Z, et al. Modeling of fermentation process based on dynamic SVM[J]. Chinese journal of scientific instrument, 2006, 27(11): 1497-1500.
- [2] 李兴峰,江波.乳酸菌来源的苯乳酸及其对食源性致病菌的抑菌活性[J].中国食品学报,2014,35(3):250-253.  
LI X F, JIANG B. Lactic acid bacteria from lactic acid bacteria and their antimicrobial activity against food borne pathogens[J]. Journal of Chinese institute of food science and technology, 2014, 35(3): 250-253.
- [3] 阎威武,朱宏栋,邵惠鹤.基于最小二乘支持向量机的软测量建模[J].系统仿真学报,2003(10):189-191.  
YAN W W, ZHU H D, SHAO H H. Soft sensor modeling based on least square support vector machine[J]. Journal of system simulation, 2003(10): 189-191.
- [4] 黄丽,孙玉坤,嵇小辅,等.基于tPSO-BPNN的赖氨酸发酵软测量[J].仪器仪表学报,2010,31(14):2317-2320.  
HANG L, SUN Y K, JI X F, et al. Soft sensing of lysine fermentation based on tPSO-BPNN[J]. Chinese journal of scientific instrument, 2010, 31(14): 2317-2320.
- [5] SUN J, FENG B, XU W B. Particles swarm optimization with particles having quantum behavior[C]. USA: Proceedings of IEEE conference on evolutionary computation, 2004(1): 325-331.
- [6] 田雨波.混合神经网络技术[M].北京:科学出版社,2009.
- [7] 陈凯,张静怡.基于RBF网络的传动滚筒测温系统[J].仪表技术与传感器,2011(9):56-58.  
CHEN K, ZHANG J Y. Temperature measurement system of transmission roller based on RBF network[J]. Instrument technique and sensor, 2011(9): 56-58.
- [8] 袁晓芳,王耀南.基于混沌优化算法的支持向量机参数选取方法[J].控制与决策,2006,21(1):111-117.  
YUAN X F, WANG Y N. Parameter selection of support vector machine based on chaos optimization algorithm[J]. Control and decision, 2006, 21(1): 111-117.
- [9] 刘毅,王海清.采用最小二乘支持向量机的青霉素发酵过程建模研究[J].生物工程学报,2006,22(5):144-149.  
LIU Y, WANG H Q. Modeling of penicillin fermentation process using least square support vector machine[J]. Chinese journal of biotechnology, 2006, 22(5): 144-149.
- [10] 张蓓,熊明勇,张克旭.人工神经网络在发酵工业中的应用[J].生物技术通讯,2003,14(1):74-76.  
ZHANG B, XIONG M Y, ZHANG K X. Application of artificial neural network in fermentation industry[J]. Letters in biotechnology, 2003, 14(1): 74-76.
- [11] AN S, LIU W, VENKATESH S. Fast cross-validation algorithms for least squares support vector machine and kernel ridge re-gression[J]. Pattern recognition, 2007, 40(8): 2154-2162.
- [12] 周兆勇,汪西莉,曹燕龙.基于GA优选参数的SVM水质评价方法研究[J].计算机工程,2008,44(4):190-193.  
ZHOU Z Y, WANG X L, CAO Y L. Study on SVM water quality evaluation method based on GA optimization parameter[J]. Computer engineering, 2008, 44(4): 190-193.

本文编辑:张 瑞