

文章编号: 1001-0920(2011)11-1753-04

## 基于连续隐 Markov 模型的发酵过程软测量方法

刘国海, 江兴科, 梅从立

(江苏大学 电气信息工程学院, 江苏 镇江 212013)

**摘要:** 针对生物发酵过程中一些生物参量难以用仪表进行在线检测的问题, 提出一种基于连续隐 Markov 模型(CHMM)的发酵过程软测量建模方法. 为减少建模过程的计算量, 提出了改进最小分类误差准则, 用于 CHMM 软测量模型参数估计. 为避免软测量结果在发酵过程监测与控制实际应用中存在的盲目性, 提出了在线评价软测量结果可靠性的可信度评价指标. 实验结果表明了所提出方法的有效性以及可信度评价指标的实际意义.

**关键词:** 软测量; 连续隐 Markov 模型; 最小分类误差; 发酵过程

**中图分类号:** TP273

**文献标识码:** A

### Research on soft sensing method based on continuous hidden Markov model in fermentation process

LIU Guo-hai, JIANG Xing-ke, MEI Cong-li

(School of Electrical and Information Engineering, Jiangsu University, Zhenjiang 212013, China. Correspondent: MEI Cong-li, E-mail: clmei@ujs.edu.cn)

**Abstract:** A soft sensing modeling method based on continuous hidden Markov model(CHMM) is developed to deal with the problem that some biologic variables cannot be measured directly online in fermentation process. In order to reduce the computation quantity of modeling process, improved minimum classification error criteria is used to train the CHMM-based soft sensor. Meanwhile, a soft sensing credibility evaluation index is proposed to avoid blindness problem during the practical application of soft sensing result to monitoring in fermentation process. The testing result shows the effectiveness of the proposed method and the practical significance of the credibility evaluation index.

**Key words:** soft sensing; continuous hidden Markov model; minimum classification error; fermentation process

## 1 引言

生物发酵过程是一个具有高度非线性、不确定性, 同时兼有生物、化学、物理和热力变化的复杂生化反应过程, 生物参量(如菌体浓度、代谢产物浓度和菌体比生长速率等)在线测量困难, 只能采用人工采样离线分析的方法. 该方法采样间隔时间长、数据滞后大, 难以满足发酵过程监测和控制要求. 为此, 人们提出了软测量方法并应用于生物发酵过程<sup>[1]</sup>.

目前, 软测量建模主要采用神经网络(NN)和支持向量机(SVM)方法. Gonzaga 等人<sup>[2]</sup>将 NN 应用于聚对苯二甲酸乙二醇酯(PET)生产过程关键变量的软测量建模, Rivera 等人<sup>[3]</sup>将 NN 应用于连续急骤发酵过程乙醇浓度软测量建模, 杨强大等人<sup>[4]</sup>提出一种改进 BP 神经网络, 用以建立发酵过程菌体浓度软测

量模型. 上述应用结果均表明了 NN 软测量方法的有效性, 但神经网络存在收敛速度慢、过拟合和难以确定隐层神经元个数的缺点. Desai 等人<sup>[5]</sup>将 SVM 应用于分批补料生物反应过程关键变量的软测量建模, Liu 等人<sup>[6]</sup>研究了基于 SVM 的软测量建模方法及其在红霉素发酵过程中的应用. 然而, SVM 良好的学习能力和泛化能力很大程度上取决于支持向量机参数的选择, 因此该方法存在参数寻优问题, 而且算法计算量较大.

连续隐 Markov 模型(CHMM)是一种可用于动态过程时间序列建模的概率分类统计模型. 相对于 NN 和 SVM 模型, CHMM 模型结构简单, 易于实现, 现已在模式识别、机械故障诊断等领域得到成功应用<sup>[7-10]</sup>. 就理论而言, CHMM 也应该可以作为生物发酵过程

收稿日期: 2010-07-20; 修回日期: 2010-09-17.

基金项目: 国家863计划项目(2007AA04Z179); 江苏高校优势学科建设工程项目.

作者简介: 刘国海(1964-), 男, 教授, 博士生导师, 从事生物发酵控制、非线性智能控制等研究; 江兴科(1985-), 男, 硕士生, 从事生物发酵软测量的研究.

软测量建模的一种方法,而且CHMM用概率统计学描述生物发酵这一随机过程的概率统计特性,显得更为合理,所建立的软测量模型更具有可解释性:CHMM是一个双重随机过程,在软测量模型建模中,内在可以相互转移的有限隐状态Markov链对应于生物发酵过程内在不可直接观测的生化反应链;Markov链中每个状态相关联的观测矢量对应于生物发酵过程可见(可测)连续状态变量。

综上所述,本文将CHMM引入生物发酵过程软测量建模,提出一种新的软测量方法,并针对现有软测量模型预测结果在发酵过程监测与控制实际应用中存在一定的盲目性问题,提出软测量结果可信度评价指标。最后将所提出的软测量方法应用于红霉素发酵过程菌体浓度软测量建模,实验结果表明了所建立的软测量模型具有良好的预估性能,软测量结果可信度评价指标具有一定的实际意义。

## 2 基于CHMM的发酵过程软测量建模

CHMM是一种用于描述随机过程概率统计特性的概率分类统计模型。CHMM建模方法通过样本在模型内的输出概率来衡量样本到模型的相似度,实现样本的分类与识别。实际建模应用时,该输出概率由Viterbi算法实现。CHMM基本定义及Viterbi算法的详细内容可参见文献[10-11]。

### 2.1 CHMM软测量建模基本思想

本文将CHMM引入生物发酵过程菌体浓度软测量建模,提出一种新的软测量建模方法,其模型结构如图1所示。其中: $O = [o_1, o_2, \dots, o_T]$ 为软测量模型输入, $y$ 为模型输出, $\{w_1, w_2, \dots, w_K\}$ 为根据历史发酵数据集确定的一个菌体浓度离线分析值数据库。

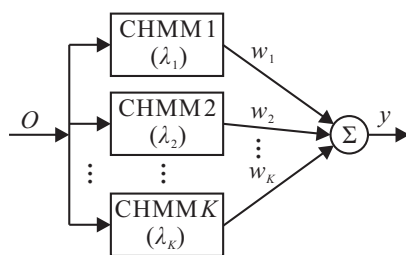


图1 基于CHMM的软测量模型结构

在软测量建模阶段,针对数据库中每一种浓度状态 $w_k$ 建立一个相应的CHMM子模型,记为 $\lambda_k$ 。所有 $K$ 个子模型构成CHMM模型库(即CHMM软测量模型)。在预测阶段,将待测样本输入CHMM模型库,模型库中每个CHMM子模型都获得一个输出概率 $P(O|\lambda_k)$ 。该输出概率表征了待测样本与子模型的相似度,将其作为子模型 $\lambda_k$ 对应浓度 $w_k$ 的权重,并通过如下加权平均运算得出软测量结果:

$$y = \frac{\sum_{k=1}^K w_k \times P(O|\lambda_k)}{\sum_{k=1}^K P(O|\lambda_k)} \quad (1)$$

### 2.2 改进最小分类误差准则

最小分类误差(MCE)训练作为一种有判别力的训练方法,已广泛应用于分类器的设计,文献[12-13]证明了这种学习方法与传统最大似然估计(ML)法相比,能够较好地改善识别性能。但是,MCE法花费的训练时间较长,因为对于一个有 $K$ 个子模型的系统,每一类别的分类误差都需要计算 $K-1$ 次的判别函数,随着 $K$ 的增加,使得计算量大大增加。为此,本文提出一种改进MCE方法,并应用于CHMM软测量模型参数训练。

传统MCE法的误差函数<sup>[12-13]</sup>为

$$d(O_{k,n}, \Lambda) = -g(O_{k,n}, \lambda_k) + \left[ \frac{1}{K-1} \sum_{p \neq k} g(O_{k,n}, \lambda_p) \right]^{1/\eta} \quad (2)$$

式中: $\Lambda$ 表示系统所有模型参数集合, $g(O_{k,n}, \lambda_k)$ 表示学习样本 $O_{k,n}$ 属于第 $k$ 类的程度(判别函数), $\eta$ 为竞争模型判别函数的加权系数。

从式(2)可以看出,误差函数为正则是错误分类,为负则是正确分类。但是,式(2)第2部分计算了所有 $K-1$ 个竞争模型的判别函数,过于复杂。本文将式(2)简化为

$$d(O_{k,n}, \Lambda) = -g(O_{k,n}, \lambda_k) + g(O_{k,n}, \lambda_c) \quad (3)$$

式中 $\lambda_c = \arg \max_{p \neq k} g(O_{k,n}, \lambda_p)$ ,即最优竞争模型。

式(3)只考虑了2个类别的误差函数。同样,误差函数为正则是错误分类,为负则是正确分类。

将上式嵌入一个值域在 $0 \sim 1$ 之间的连续函数(如Sigmoid函数),便可得到所需要的损失函数

$$l(O_{k,n}, \Lambda) = \frac{1}{1 + \exp(-d(O_{k,n}, \Lambda))} \quad (4)$$

对于具有 $K$ 个子模型,每类子模型训练样本数为 $N$ 的系统,总损失函数可表示成如下形式:

$$L(O, \Lambda) = \sum_{k=1}^K \sum_{n=1}^N l(O_{k,n}, \Lambda) \quad (5)$$

于是,以式(5)作为目标函数,利用最佳梯度下降法使其最小化,即可实现基于改进MCE准则的CHMM软测量模型参数估计。

### 2.3 软测量结果可信度评价指标

生物发酵软测量研究的实际意义在于:通过软测量模型在线实时预测发酵过程中难以用仪表进行在线检测的生物参量,再将预测结果应用于发酵过程的实时监测和控制,可解决由传统人工采样离线分析方

法所产生的测量滞后所带来的监测和控制不及时的问题。但是, 生物发酵过程是一个极其复杂的不确定性过程, 在缺少可信度指标的前提下, 将软测量结果直接应用于发酵过程的监测和控制具有一定的盲目性。

为解决上述问题, 本文在 CHMM 软测量建模方法的基础上, 提出了软测量结果可信度评价指标, 用于在线评价软测量模型预测结果的可靠性。该可信度指标计算公式为

$$V = \frac{\max_k(P(O|\lambda_k))}{\sum_{k=1}^K P(O|\lambda_k)} \times 100\% \quad (6)$$

从式 (6) 可以看出:  $V$  越大, 表示待测样本与某个 CHMM 子模型相似度越大, 所得软测量结果越可靠; 反之,  $V$  很小, 表示待测样本与模型库中每个 CHMM 子模型相似度很小, 从而所得软测量结果不可靠。

## 2.4 基于改进 MCE 准则的 CHMM 菌体浓度软测量建模步骤

基于改进 MCE 准则的 CHMM 发酵过程菌体浓度软测量建模步骤如下:

Step 1: 数据采集及预处理。

Step 2: 菌体浓度值数据库确定。根据 Step 1 获取的历史发酵数据集, 确定一个菌体浓度值数据库  $\{w_1, w_2, \dots, w_K\}$ 。

Step 3: 辅助变量选择及特征提取。选择发酵过程在线可测状态变量按序排列组成 CHMM 观测值序列, 并从训练样本集中提取出对应于菌体浓度值数据库中每个浓度状态的多组观测值序列, 将其作为训练样本序列用于 CHMM 模型训练。

Step 4: CHMM 模型结构选取。包括模型状态数和状态转移方式的选择, 本文选取四状态各态历经型 CHMM。

Step 5: CHMM 软测量模型训练。通过 Step 3 提取的训练样本序列, 在改进 MCE 准则下进行 CHMM 软测量模型参数估计。

Step 5.1: 初始化。设定算法最大迭代步数和收敛误差, 在满足 CHMM 参数归一化约束条件<sup>[10]</sup>下随机初始化模型库参数集。

Step 5.2: 计算总损失函数对各参数的偏导数, 并根据下式调整模型参数:

$$\theta(n) = \theta(n-1) - \varepsilon \frac{\partial L}{\partial \theta(n-1)} \quad (7)$$

这里取  $\varepsilon = 1/n$ , 以保证算法收敛。

Step 5.3: 计算总损失  $L(O, \Lambda(n))$ 。

Step 5.4: 终止条件判断。若不满足终止条件, 则转到 Step 5.2; 若满足终止条件, 则训练结束。

Step 6: 软测量实现。将待测样本输入训练好的 CHMM 软测量模型, 用 Viterbi 算法计算每个 CHMM 子模型的输出概率, 再用式 (1) 和 (6) 分别得到软测量结果及其可信度。

## 3 实验与分析

通过对镇江某制药公司红霉素发酵过程菌体浓度进行软测量建模, 并与 NN 软测量方法进行对比研究来验证本文 CHMM 软测量方法的有效性。

红霉素发酵周期约为 168 h, 其中 50~160 h 为发酵过程的主要阶段, 取 6 h 为采样周期, 共采集 301# 发酵罐中 10 批 (200 个样本点) 红霉素发酵过程数据, 其中每个样本点数据包含 10 个输入变量 (分别为溶解氧、pH 值、糊精体积、豆油体积、丙醇体积、水体积、温度、相对气压、转速和空气流量) 和一个输出变量 (通过离线化验分析获得的菌体浓度离线分析值), 每批数据均是一个完整的红霉素发酵过程的记录。

为更直观地描述一个完整红霉素发酵过程菌体浓度的连续变化过程, 采用最小二乘法将同一批发酵过程离散的菌体浓度离线分析值拟合成连续光滑曲线, 如图 2 所示。

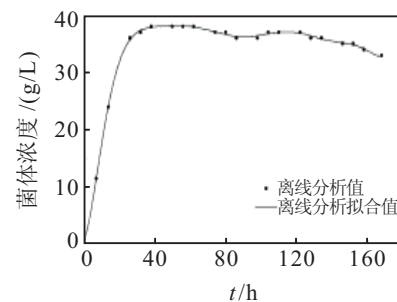


图 2 菌体浓度离线分析值最小二乘拟合结果

实验分训练与测试 2 个模块, 任取 7 批发酵数据作为训练数据集, 剩下的 3 批作为测试数据集。通过对训练数据集中菌体浓度值的分析, 选取 6 个浓度值 (33 g/L, 34 g/L, 35 g/L, 36 g/L, 37 g/L, 38 g/L) 组成菌体浓度值数据库。为了提高准确性, 在输入变量中加入上一时刻的菌体浓度, 将这 11 个辅助变量按序排列组成 CHMM 观测值序列, 并将训练数据集处理成统一标准的观测值序列集合, 作为总学习样本集用于 CHMM 模型库训练。建模时, 改进 MCE 训练算法最大迭代步数为 100, 收敛误差为  $1 \times 10^{-4}$ 。为简化 MCE 训练过程, 采用简单易实现的 ML 法初始化模型库参数集<sup>[13]</sup>。改进 MCE 准则训练过程的迭代次数和总损失变化曲线如图 3 所示。由图 3 可知, 随着迭代次数的增加, 模型总损失在 50 步的迭代次数内达到收敛。

为验证本文提出的 CHMM 软测量建模方法的有效性, 利用前面训练好的 CHMM 软测量模型对 3 批

测试数据样本点进行菌体浓度值预测,并结合NN软测量模型的预测效果进行对比分析.图4和图5分别是CHMM和NN采用相同的训练数据集对同一批发酵过程样本点菌体浓度进行预测的结果.

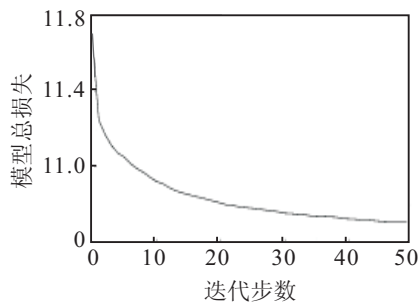


图3 改进MCE准则训练过程模型总损失变化曲线

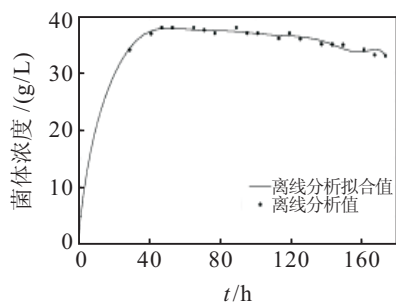


图4 基于CHMM软测量模型对菌体浓度的预测仿真

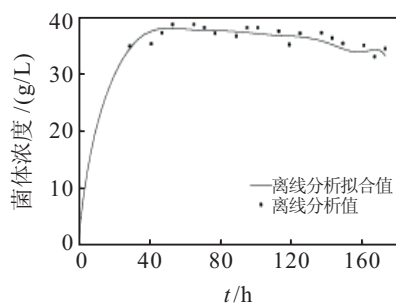


图5 基于NN软测量模型对菌体浓度的预测仿真

由图4和图5的仿真结果对比发现,在利用相同训练样本的情况下,CHMM的预测效果较NN优越.其中CHMM软测量模型测试样本的均方差(MSE)为0.178,NN软测量模型测试样本的均方差为0.904.

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2. \quad (8)$$

式中: $N$ 为预测样本点的个数, $y_i$ 为样本离线分析值, $\hat{y}_i$ 为相应软测量模型预测值.

为更直观地说明本文所建立的CHMM软测量模型的优良性能以及可信度指标的实际意义,基于前面所得实验结果,表1列出了采用相同训练数据集和测试数据集条件下,CHMM和NN软测量模型对3批测试数据中的部分测试样本(10个样本点)进行预测的结果.

从表1可以看出,本文提出的CHMM软测量模型不但预测精度比NN软测量模型有显著提高,而且每个预测结果都附有可信度这个评价指标.其中:

表1 部分测试样本及预测结果

序号	离线分析值/(g/L)	预测值/(g/L)		误差/(g/L)		可信度/%	
		CHMM	NN	CHMM	NN	CHMM	NN
1-52	37	36.99	37.96	0.01	0.96	99	-
1-70	38	38	37.04	0	0.96	100	-
1-124	36	36.04	37.20	0.04	1.20	95	-
1-142	34	34.50	35.17	0.50	1.17	50	-
2-52	35	34.92	36.35	0.08	1.35	91	-
2-82	37	36.22	36.41	0.78	0.59	32	-
2-100	38	37.99	36.37	0.01	1.62	99	-
3-58	36	35.99	36.76	0.01	0.76	99	-
3-76	37	36.98	36.43	0.02	0.57	99	-
3-148	33	33.04	34.06	0.04	1.06	98	-

注:序号“\*-\*”表示第\*批测试数据\*\*时刻采样的样本点.

序号为1-142和2-82的样本点软测量模型预测结果的可信度较低,分别为50%和32%,其测试误差也较大,分别为0.5 g/L和0.78 g/L,而高可信度预测结果普遍具有较好的预测精度,这说明了本文所提出的可信度指标的有效性.因此,在软测量模型实际应用时,可根据可信度这一在线评价指标有选择地选取高可信度预测结果用于发酵过程的实时监测和控制,以避免盲目控制.由此可见,本文提出的CHMM软测量模型具有优良性能,软测量结果可信度评价指标具有一定的实际意义.

## 4 结论

本文将CHMM应用于生物发酵过程软测量建模,提出了一种新的软测量建模方法.采用改进MCE准则训练CHMM软测量模型,减少了建模过程计算量.为避免软测量结果在发酵过程监测和控制实际应用中存在的盲目性,提出了软测量结果可信度评价指标.将所提出的方法应用于红霉素发酵过程进行软测量建模,并与NN软测量建模方法进行了对比研究,实验结果表明了新方法的可行性和可信度评价指标的实际意义,为软测量建模提供了一个新的研究思路.

## 参考文献(References)

- [1] Bogaerts P H, Wouwer A V. Software sensors for bioprocesses[J]. The Instrumentation, Systems, and Automation Society(ISA Transactions), 2003, 42(4): 547-558.
- [2] Gonzaga J C B, Meleiro L A C, Kiang C, et al. ANN-based soft-sensor for real-time process monitoring and control of an industrial polymerization process[J]. Computers and Chemical Engineering, 2009, 33(1): 43-49.
- [3] Rivera E C, Atala D L P, Costa A C D, et al. Soft-sensor for real-time estimation of ethanol concentration in continuous flash fermentation[J]. Computer Aided Chemical Engineering, 2009, 27: 1653-1658.

(下转第1760页)