文章编号:1671-4598(2020)09-0131-06 DOI:10.16526/j.cnki.11—4762/tp.2020.09.026 中图分类号:TP274 文献标识码:A

基于变分贝叶斯算法的青霉素发酵过程建模

蔡子君，谢 莉，杨慧中
（江南大学 轻工过程先进控制教育部重点实验室，江苏 无锡 214122）

摘要：青霉素发酵过程具有明显的阶段特征，同时由于操作条件多变、生产环境复杂等原因导致其存在极大的不确定性，故本文在变分贝叶斯框架下建立了青霉素浓度预测的FIR融合模型；首先选取调度变量对发酵阶段进行划分，然后基于变分贝叶斯算法辨识得到各FIR子模型的参数，最后根据阶段特征计算样本隶属于各子模型的概率并融合子模型的输出得到青霉素浓度的预测值；文中利用工业规模青霉素发酵罐的实际数据进行仿真实验，模型预测青霉素浓度的相关误差为0.24%，表明提出模型具有较高的拟合度，能够更为精准的预测青霉素浓度并适应实际的复杂工业环境。

关键词：青霉素发酵过程；变分贝叶斯算法；融合模型

Modeling for Penicillin Fermentation Processes Based on Variational Bayesian Algorithm

Cai Zijun, Xie Li, Yang Huizhong
(Ministerial Key Laboratory of Advanced Process Control for Light Industry, Jiangnan University, Wuxi 214122, China)

Abstract: Penicillin fermentation processes have obvious stage characteristics, meanwhile which have great uncertainties due to some reasons of variable operation conditions and complex production environments, this paper aims to establish a finite impulse response (FIR) fusion model under the variational Bayesian (VB) framework for online prediction of penicillin concentration. First, the scheduling variable is selected to divide fermentation stages, then the parameters of each FIR sub-model are identified based on the VB algorithm. Finally, the probability of the sample belonging to each sub-model is calculated according to the stage characteristics, and further applied to fuse sub-model outputs for obtaining the penicillin concentration predictions. The paper uses the actual industrial scale penicillin fermentation data to carry out simulation experiments. The correlation error of the model predicting penicillin concentration is 0.24% which shows that the model has a high degree of fitting, which can provide more accurate prediction of the penicillin concentration and adapt to the actual complex industrial environments.

Keywords: penicillin fermentation process; variational Bayesian algorithm; fusion model

0 引言

抗生素是一类生物产生的具有抑制某些细胞生长的次级代谢产物[1]，其对于某些病原微生物的抑制和灭杀作用使其成为一种重要的药物。作为一种典型的抗生素，青霉素发酵过程具有以下特点：

1）时变性。发酵过程中青霉素的浓度取决于生成青霉素的菌丝浓度，而影响菌丝生长的因素众多，例如发酵罐中糖浓度、溶解氧浓度、pH值、温度等，并且这些变量都会随着时间改变，导致青霉素的发酵过程呈现很强的时变性。

2）非线性。青霉素是一种次级代谢产物，生成青霉素所需的众多基质、前体、以及副产物之间会发生复杂的反应，形成了一个多输入多输出的非线性系统，加大了青霉素的浓度预测难度。

3）阶段性。青霉素发酵通常经历4个阶段，分别为准备期、对数生长期、平稳期和消亡期[2]，如图1所示，准备期中菌丝慢慢生成，青霉素主要在对数生长期生成，经过平稳期后菌丝逐渐水解，进入消亡期。

4）测量难度大。许多关键变量无法在线测量，需要进行离线分析，例如青霉素浓度、氮浓度、浓稠度等，导致关键变量的数据稀缺。

上述特点使得对青霉素发酵过程的预测和控制较为困难。过去提出的各种青霉素发酵过程的预测模型大致可以分为机理模型和数据驱动模型两类。最著名的机理模型是由Birol于2002年改进的非结构模型[3]。由于青霉素化学结构复杂，发酵过程涉及的反应众多，大部分机理模型不对内部结构讨论[4]，而是对青霉素与菌丝的不同部分之间的动态关

收稿日期:2020-01-14； **修回日期**:2020-03-20。
基金项目:国家自然科学基金资助项目(61403166,61773181)；江苏省自然科学基金资助项目(BK20140164)；中央高校基本科研业务费专项资金资助项目(JUSRP51733B)。
作者简介:蔡子君(1995-)，男，江苏无锡人，硕士研究生，主要从事青霉素发酵过程建模方向的研究。
谢 莉(1985-)，女，重庆长寿人，博士，副教授，主要从事软测量建模与多率系统辨识方向的研究。
杨慧中(1955-)，女，江苏无锡人，博士，教授，主要从事复杂过程建模和优化控制方向的研究。

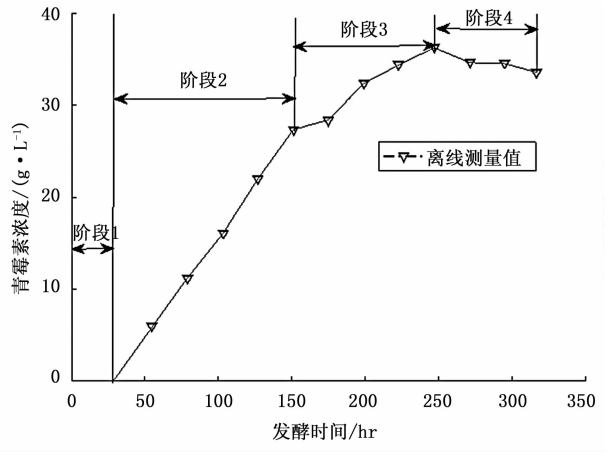

图 1 青霉素浓度曲线阶段图

系进行描述。这种非结构模型能够一定程度上描述反应过程并在特定参数环境下模拟发酵过程，但是在面对实际发酵过程多变的环境时就显露出适应性差的缺点。近年来，许多基于数据驱动的青霉素发酵模型被提出，利用诸如神经网络、支持向量机等方法模拟青霉素发酵过程[5]。相对机理模型，数据驱动模型具有以下两点优势：1）不需要复杂的发酵机理知识，降低了建模难度；2）在发酵环境改变时不需要大量实验重新确定模型参数，降低了模型的维护成本。但现有的数据驱动模型大多利用 Pensim 仿真平台产生的数据，只能描述出在实验室规模下青霉素发酵的大致过程。

本文针对实际工业中的青霉素发酵过程，基于变分贝斯算法建立 FIR 融合模型。在模型选择方面，由于青霉素发酵过程具有明显的阶段性[6]，本文采用多模型融合的思路，选取能够反应阶段特征的关键过程变量作为调度变量进行阶段划分，确定几个典型工况点以计算各局部模型的权重。工业青霉素发酵过程中生产环境复杂性使得过程存在着不确定性，这对发酵过程的辨识造成了很大的困难[7]。不确定性在工业过程辨识中的处理方法可分为两种[8]，一种是通过不同模型的变化体现系统的不确定性；而本文采用的是第二种，即在模型结构已知的条件下，将系统不确定性通过模型参数的不确定性来体现。本文采用变分贝叶斯算法作为辨识算法，它是期望最大化（EM）算法在贝叶斯方法上的一种推广，相较于 EM 算法的点估计，变分贝叶斯算法可以估计参数和隐变量的整个后验概率分布，能够更好地描述青霉素发酵过程的不确定性。

1 青霉素发酵阶段划分

调度变量是指可以反应系统工作状态与阶段特征并能够人为调控的过程变量[9]。在 Pensim 仿真平台的环境下，冷水流加速率能比较好的反映发酵的阶段特征，所以在以往的青霉素发酵模型中经常被作为模型的调度变量。但是，在实际工业过程中，操作人员会频繁地改变冷水流加速率以保持稳定的发酵罐温度，导致实际过程中冷水流加速率（图 2）不宜作为过程的调度变量。

考虑实际工业过程的操作环境，本文将使用葡萄糖流

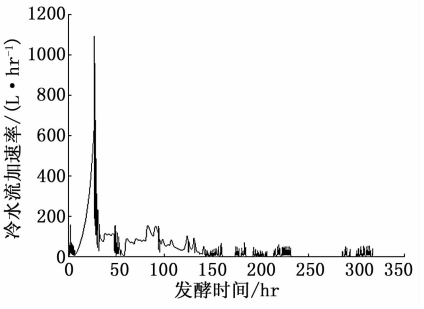

图 2 工业环境下冷水流加速率图

加速率作为调度变量。葡萄糖是青霉素发酵中底物的一种，其流加速率虽然在发酵过程中经过人为调整，但调整频率相对较低并且整体曲线能够体现发酵过程的阶段特性[10]。

为划分发酵过程的各个阶段，需要进一步对调度变量进行聚类。本文采用模糊 C 均值（Fuzzy C－Means，FCM）聚类方法[11]对葡萄糖流加速率进行聚类，并将各聚类中心作为发酵过程的典型工作点。FCM 是一种基于目标函数的聚类算法，首先对聚类中心设置初值，然后通过最小化数据点与各聚类中心的距离和模糊隶属度的加权和为目标，不断修正聚类中心和分类矩阵直到符合终止准则，将具有类似特征的数据聚为一类。

在青霉素发酵过程的准备期中，菌体快速生长消耗氧气，当溶解氧水平下降到一定程度时，菌体会产生中间代谢物并开始生成青霉素[12]，所以本文假定在准备期中青霉素浓度为零，而对数生长期、平稳期和消亡期这 3 个阶段分别对应 3 个不同的工作点。因此，在采用 FCM 算法对实际工业青霉素发酵过程中的葡萄糖流加速率进行聚类时，将聚类中心个数设置为 3，相应的聚类结果如图 3 所示，其中三角形表示计算得到的聚类中心。

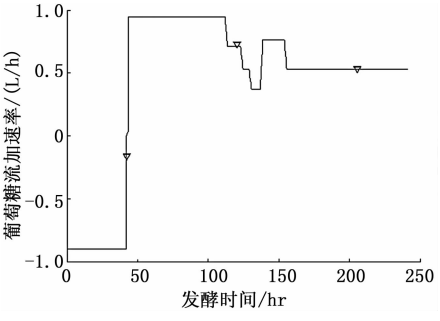

图 3 葡萄糖流加速率 FCM 聚类结果

2 模型结构：FIR 融合模型

考虑到在工业现场青霉素浓度为慢速采样[13]，无法用于模型输入，所以本文采用 FIR 模型作为局部模型，然后通过权重函数将 3 个局部模型融合为全局模型来描述青霉素发酵的动态特性。局部 FIR 模型结构如下：

$$y_k = x_k^T \theta_{l_k} + e_k \tag{1}$$

$$x_k = [u_{l_{k-1}} u_{l_{k-2}} \cdots u_{l_{k-n_a}} \cdots u_{m_{k-1}} u_{m_{k-2}} \cdots u_{m_{k-n_b}}]^T \tag{2}$$

其中：$k = 1,2\cdots N$ 表示发酵过程的各时刻，输出变量

y_k 为 k 时刻发酵罐中青霉素浓度，$I_k=1,2,3$ 表示 k 时刻变量隶属的局部模型，模型选择的输入变量个数为 m，输入阶数设为 n_a，e_k 是均值为 0 方差为 σ_i^{-1}（未知的待辨识参数）的高斯分布噪声，下标 $i=I_k$ 表示各个局部模型。

采用高斯分布作为权重函数[14]：

$$\lambda_{ki}=\exp\left[\frac{-(H_k-H_i^0)^2}{2O_i^2}\right] \tag{3}$$

其中：H_k 表示 k 时刻的调度变量即葡萄糖流加速率，H_i^0 和 O_i 分别表示第 i 个局部模型的工作点和有效宽度。将 λ_{ki} 归一化后得到的权值函数代表在 k 时刻第 i 个局部模型对全局模型影响大小：

$$\alpha_{ki}=\frac{\lambda_{ki}}{\sum_{i=1}^{3}\lambda_{ki}} \tag{4}$$

若各个局部模型的有效宽度 O_i 已知，根据式（3）～（4）计算出各子模型的权重后，容易得到系统全局模型的预测输出：

$$\hat{y}_k=\sum_{i=1}^{3}\alpha_{ki}x_k^T\theta_i \tag{5}$$

然而，局部模型的有效宽度 O_i 未知，需要与各子模型的参数 θ_i 以及噪声方差 σ_i^{-1} 同时辨识，增加了系统辨识的难度。此外，考虑到实际过程的不确定性，本文通过引入参数不确定性，在 VB 算法框架下推导相应的辨识算法。

3　基于 VB 算法的模型参数辨识

3.1　VB 算法回顾

令模型的参数集为 Θ，隐变量集为 C_{mis}，观测数据集为 C_{obs}。对于存在未知参数的模型，边缘似然函数可以由下式计算：

$$p(C_{obs})=\int p(C_{obs},C_{mis},\Theta)\,\mathrm{d}C_{mis}\mathrm{d}\Theta \tag{6}$$

而式（6）中含有难以计算的高维积分，VB 算法通过构造联合分布 $q(C_{mis},\Theta)$ 来近似计算后验分布 $p(C_{mis},\Theta)$，运用 Jensen 不等式[15]得到：

$$\ln p(C_{obs})=\ln\int q(C_{mis},\Theta)\frac{p(C_{obs},C_{mis},\Theta)}{q(C_{mis},\Theta)}\,\mathrm{d}C_{mis}\mathrm{d}\Theta\geqslant\int q(C_{mis},\Theta)\ln\frac{p(C_{obs},C_{mis},\Theta)}{q(C_{mis},\Theta)}\mathrm{d}C_{mis}\mathrm{d}\Theta \tag{7}$$

假定联合分布 $q(C_{mis},\Theta)$ 是可分解的[16]，得到对数边缘函数的下界函数：

$$F[q(C_{mis})q(\Theta)]=\int q(C_{mis})q(\Theta)\ln\frac{p(C_{obs},C_{mis},\Theta)}{q(C_{mis})q(\Theta)}\mathrm{d}C_{mis}\mathrm{d}\Theta \tag{8}$$

将对数边缘函数与下界函数做差，可得：

$$\ln p(C_{obs})-F[q(C_{mis})q(\Theta)]=\int q(C_{mis})q(\Theta)\ln\frac{q(C_{mis})q(\Theta)}{p(C_{obs},C_{mis},\Theta)}\,\mathrm{d}C_{mis}\mathrm{d}\Theta=KL(q\,|\,p) \tag{9}$$

其中：$KL(q\|p)$ 表示 Kullback－Leible 散度，即联合分布 $q(C_{mis},\Theta)$ 与实际分布 $p_(C_{mis},\Theta|C_{obs})$ 之间的差异，所以由式（9）可知，最大化 $F[q(C_{mis}),q(\Theta)]$ 等价于最小化 $KL(q\|p)$[17]。

与期望最大化算法类似，变分贝叶斯算法不断迭代地更新隐变量和模型参数的后验分布，直至算法收敛，得到真实后验分布 $p(C_{mis},\Theta|C_{obs})$ 的近似分布 $q(C_{mis},\Theta)$[18]。

3.2　基于 VB 的辨识算法推导

3.2.1　模型参数的先验分布

在 FIR 融合模型中，观测到的数据集 $C_{obs}=\{y_{1:N},u_{1:N},H_{1:N},H_{1:3}^0\}$，缺失数据集 $C_{mis}=\{I_{1:N}\}$，待估计的参数集 $\Theta=\{\theta_{1:3},\sigma_{1:3},\eta_{1:3},O_{1:3}\}$，其中 η_i 表示局部模型参数的精度[19-20]。初始化待估计参数 θ_i，η_i，σ_i 的先验分布：

$$p(\theta_i\,|\,\eta_i)=N(0,\eta_i^{-1}\boldsymbol{D}_{n_a\times n_a})$$

$$p(\eta_i\,|\,a_0,b_0)=g(a_0,b_0,X)=\frac{{a_0}^{b_0}}{\Gamma(a_0)}X^{a_0-1}e^{-b_0X}p(\sigma_i\,|\,c_0,d_0)=g(c_0,d_0,X)=\frac{{c_0}^{d_0}}{\Gamma(c_0)}X^{c_0-1}e^{-d_0X}$$

其中：$\boldsymbol{D}$ 表示单位矩阵，g 表示伽马分布，a_0,b_0,c_0,d_0 为伽马分布的超参数，Γ 表示伽马函数，其表达式为：

$$\Gamma(x)=\int_0^{\infty}t^{x-1}e^{-t}\mathrm{d}t$$

将上述先验分布用一个联合先验分布表示：

$$p(\Theta)=\prod_{i=1}^{3}p(\Theta_i)=\prod_{i=1}^{3}p(\theta_i\,|\,\eta_i)p(\eta_i)p(\sigma_i) \tag{10}$$

3.2.2　VB 算法：E 步

在 E 步骤中，固定参数集，关于隐变量对下界函数求极值，得到隐变量的更新后验分布 $q(I)$。下界函数 $F[q(I),q(\Theta)]$ 可表示为：

$$F[q(I),q(\Theta)]+\int q(\Theta)\ln\frac{p(\Theta|O)}{q(\Theta)}\mathrm{d}\Theta \tag{11}$$

求解如下优化问题：

$$\max_I F[q(C_{mis}),q(\Theta)]$$
$$s.t.\ \sum_I q(I)=1$$

计算关于 $q(I)$ 的拉格朗日函数的导数：

$$\frac{\partial\left\{F[q(C_{mis}),q(\Theta)]+\lambda\left[\sum_I q(I)-1\right]\right\}}{\partial\,q(I)}=0$$

逐项求导后可以得到：

$$q(I)=\exp\left\{\lambda-1+<\ln p(y_{1,N},u_{1,N},H_{1,N},H_{1,3}^0,I_{1,N}\,|\,\Theta,O)>_{q(\Theta)}\right\} \tag{12}$$

将 $\sum_I q(I)=1$ 代入，得到：

$$e^{1-\lambda}=\sum_I\exp\{<\ln p(y_{1,N},u_{1,N},H_{1,N},H_{1,3}^0,I_{1,N}\,|\,\Theta,O)>_{q(\Theta)}\}$$

令：

$$Z_I=\sum_I\exp\{<\ln p(y_{1,N},u_{1,N},H_{1,N},H_{1,3}^0,I_{1,N}\,|\,\Theta,O)>_{q(\Theta)}\}$$

得到：

$$q(I)=\frac{1}{Z_I}\exp\{<\ln p(y_{1,N},u_{1,N},H_{1,N},H_{1,3}^0,I_{1,N}\,|\,\Theta,O)>_{q(\Theta)}\} \tag{13}$$

其中：$<\cdot>_{q(\Theta)}$ 代表对 $q(\Theta)$ 求期望。

将 $\ln p(y_{1:N},u_{1:N},H_{1:N},H_{1:3}^0,I_{1:N}\mid\Theta,O)$ 表示为：

$$\ln p(y_{1:N},u_{1:N},H_{1:N},H_{1:3}^0,I_{1:N}\mid\Theta,O) = \sum_{k=1}^{N}\ln\begin{Bmatrix} p(y_k\mid u_k,I_k,H_{1:3}^0,H_k,\Theta,O)\times \\ p(I_k\mid u_k,H_{1:3}^0,H_k,\Theta,O)\times \\ p(u_k,H_{1:3}^0,H_k\mid\Theta,O)\end{Bmatrix} \tag{14}$$

由式（1）的局部 FIR 模型可知，当前时刻的输出 y_k 与历史时刻输入 x_k 、参数 θ_{I_k} 以及隐变量 I_k 有关，并且由于 I_k 代表 k 时刻变量隶属的局部模型，所以 $p(I_k)$ 代表了权值函数。而由式（3）可知权值函数与调度变量、系统工作点和局部模型有效宽度有关，又由于 $p(u_k,H_k,H_{1:3}^0\mid\Theta,O)$ 是常数，将其表示为 C，可以将式（14）简化为：

$$\ln p(y_{1:N},u_{1:N},H_{1:N},H_{1:3}^0,I_{1:N}\mid\Theta,O) = \sum_{k=1}^{N}\ln\{p(y_k\mid x_k,I_k,\Theta)p(I_k\mid H_{1:3}^0,H_k,O)C\}$$

由此可以将 Z_I 表达为：

$$Z_I = \sum_I\prod_{k=1}^{N}\exp\{<\ln p(y_{1:N},u_{1:N},H_{1:N},H_{1:3}^0,I_{1:N}\mid\Theta,O)>_{q(\Theta)}\}$$

继而定义：

$$Z_{I_s} = \sum_{I_s}\exp\{<\ln p(y_{1:N},u_{1:N},H_{1:N},H_{1:3}^0,I_{1:N}\mid\Theta,O)>_{q(\Theta)}\}$$

则 $Z_I=\prod_{k=1}^{N}Z_{I_s}$ ，结合式（13），可得：

$$q(I)=\prod_{i=1}^{3}q(I_k=i)$$

其中：$q(I_k=i)$ 可以表示为：

$$\frac{1}{Z_{I_s}}\exp\{<\ln p(y_{1:N},u_{1:N},H_{1:N},H_{1:3}^0,I_{1:N}\mid\Theta,O)>_{q(\Theta)}\} \tag{15}$$

为简化表达，令：

$$A_{ki}=\exp\{<\ln p(y_{1:N},u_{1:N},H_{1:N},H_{1:3}^0,I_{1:N}\mid\Theta,O)>_{q(\Theta)}\}$$

式（15）可简化为：

$$q(I_k=i)=\frac{A_{ki}}{\sum_{i=1}^{3}A_{ki}} \tag{16}$$

接下来要求出 A_{ki} 的表达式，由 y_k 服从均值为 $x_k^T\theta_{I_s}$ ，方差为 σ_i^{-1} 的高斯分布，可以得到：

$$p(y_k\mid x_k,I_k,\Theta)=\frac{\sqrt{\sigma_i}}{\sqrt{2\pi}}\exp\left\{\frac{\sigma_i}{-2}(y_k-x_k^T\theta_{I_s})^2\right\}$$

由于 $p(I_k\mid H_{1:3}^0,H_k,O)$ 代表了权值函数 α_{ki}，可得：

$$\ln A_{ki}=[\ln\{p(y_k\mid x_k,I_k,\Theta)p(I_k\mid H_{1:3}^0,H_k,O)C_a\}]_{q(\Theta)} = \frac{1}{2}(-\ln 2\pi+<\ln\sigma_i>_{q(\sigma)}-<\sigma_i>_{q(\sigma)}<(y_k-x_k^T\theta_i)^2>_{q(\theta)})+\ln\alpha_{ki}+\ln C_a \tag{17}$$

其中：

$$<(y_k-x_k^T\theta_i)^2>_{q(\theta)}=y_k^2-2y_kx_k^T<\theta_i>_{q(\theta)}+x_k^T<\theta_i\theta_i^T>_{q(\theta)}x_k$$

将 A_{ki} 代入式（16）即可得到 $q(I_k=i)$ 。

3.2.3 VB 算法：M 步

在 M 步骤中，固定隐变量，关于参数集对下界函数（11）求极值，得到参数集的更新式 $q(\Theta)$ 。下界函数可以表示为：

$$F[q(I),q(\Theta)]=\int q(\theta)\,q(\sigma)<\ln p(y_{1:N},u_{1:N},H_{1:N},H_{1:3}^0,I_{1:N}\mid\Theta,O)>_{q(I)}\mathrm{d}\theta\mathrm{d}\sigma-<\ln q(I)>_{q(I)}+\int q(\theta)q(\eta)q(\sigma)\ln\frac{p(\theta\mid\eta)p(\eta)p(\sigma)}{q(\theta)q(\eta)q(\sigma)}\mathrm{d}\theta\mathrm{d}\eta\mathrm{d}\sigma \tag{18}$$

其中：

$$<\ln p(y_{1:N},u_{1:N},H_{1:N},H_{1:3}^0,I_{1:N}\mid\Theta,O)>_{q(I)}=\sum_{k=1}^{N}\sum_{i=1}^{M}q(I_k=i)\begin{Bmatrix}-\frac{1}{2}\ln 2\pi+\frac{1}{2}\ln\sigma_i+\ln\alpha_{ki}-\\ \frac{1}{2}\sigma_i(y_k-x_k^T\theta_i)^2+\ln C_a\end{Bmatrix}$$

下面利用拉格朗日乘子法，依次关于 $q(\theta_i)$,$q(\eta_i)$ 和 $q(\sigma_i)$ 最大化下界函数。

1) $q(\theta_i)$ 部分：计算关于 $q(\theta_i)$ 的拉格朗日函数的一阶导数

$$\frac{\partial\{F[q(I),q(\Theta)]+\lambda(q(\theta_i)-1)\}}{\partial q(\theta_i)}=0$$

可得：

$$-\ln q(\theta_i)+\ln p(\theta_i\mid<\eta_i>_{q(\eta)})+C_\theta-\frac{1}{2}\sum_{k=1}^{N}q(I_k=i)\left\{<\sigma_i>_{q(\sigma)}\begin{pmatrix}\theta_i^Tx_kx_k^T\theta_i-\\ 2\theta_i^Tx_ky_k\end{pmatrix}\right\}=0$$

即：

$$q(\theta_i)\propto p(\theta_i\mid<\eta_i>_{q(\eta)})\times\exp\left\{-\frac{1}{2}\sum_{k=1}^{N}q(I_k)<\sigma_i>_{q(\sigma)}\begin{pmatrix}\theta_i^Tx_kx_k^T\theta_i-\\ 2\theta_i^Tx_ky_k\end{pmatrix}\right\}=\exp\left\{-\frac{1}{2}\begin{bmatrix}\theta_i^T<\eta_i>_{q(\eta)}\theta_i+\\ \sum_{k=1}^{N}q(I_k=i)<\sigma_i>_{q(\sigma)}\begin{pmatrix}\theta_i^Tx_kx_k^T\theta_i-\\ 2\theta_i^Tx_ky_k\end{pmatrix}\end{bmatrix}\right\}$$

因此，$q(\theta_i)$ 服从高斯分布，即：

$$q(\theta_i)\propto\exp\left\{-\frac{1}{2}\frac{(\theta_i-E(\theta_i))^2}{Var(\theta_i)}\right\}$$

其中：

$$Var(\theta_i)=(<\eta_i>_{q(\eta)}D+\sum_{k=1}^{N}q(I_k=i)x_kx_k^T)^{-1} \tag{19}$$

2) $q(\eta_i)$ 部分：类似地关于 $q(\eta_i)$ 对式（18）的求导：

$$\frac{\partial\{F[q(I),q(\Theta)]+\lambda(q(\eta_i)-1)\}}{\partial q(\eta_i)}=0$$

可得：

$$-\ln q(\eta_i)+(a_0+1)\ln p(\eta_i)-(b_0+\frac{1}{2}<\theta_i^T\theta_i>_{q(\theta_i)})\eta_i+C_\eta=0$$

其中：

$$p(\theta_i \mid \eta_i) = \frac{e^{-\frac{1}{2}\theta_i^T(\eta_i D)\theta_i}}{\sqrt{(2\pi)^{n_a} \mid \eta_i^{-1} D \mid}}$$

$$p(\eta_i) = \frac{b_0^{a_0}}{\Gamma(a_0)} \eta_i^{a_0-1} e^{-b_0\eta_i} \mathrm{d}f$$

整理后得到：

$$\ln q(\eta_i) = (a_0+1)\ln\eta_i - \left(b_0 + \frac{1}{2} < \theta_i^T\theta_i >_{q(\theta_i)}\right)\eta_i + C_\eta \tag{20}$$

其中：C_η 是与 η_i 无关的常数。由式（20）可知 $q(\eta_i)$ 服从伽马分布，即 $q(\eta_i) = g(a_i, b_i)$，且：

$$a_i = a_0 + \frac{n_a}{2}$$

$$b_i = b_0 + \frac{1}{2} < \theta_i\theta_i^T >_{q(\theta_i)}$$

$$< \theta_i\theta_i^T >_{q(\theta_i)} = Var(\theta_i) + E(\theta_i)E(\theta_i)^T$$

3）$q(\sigma_i)$ 部分：关于 $q(\sigma_i)$ 对式（18）的三部分求导：

$$\frac{\partial\{F[q(I), q(\Theta)] + \lambda(q(\sigma_i)-1)\}}{\partial q(\sigma_i)} = 0$$

可得：

$$\ln q(\sigma_i) = \left(c_0 - 1 + \frac{1}{2}\sum_{k=1}^{N} q(I_k = i)\right)\ln\sigma_i + C_\sigma - \left(d_0 + \frac{1}{2}\sum_{k=1}^{N} q(I_k = i)\begin{pmatrix} y_k^2 - 2 < \theta_i^T >_{q(\theta_i)} y_k x_k + \\ x_k^T < \theta_i\theta_i^T >_{q(\theta_i)} x_k \end{pmatrix}\right)\sigma_i \tag{21}$$

其中：C_σ 是与 σ_i 无关的常数。根据式（21）可知 $q(\sigma_i)$ 服从伽马分布，即 $q(\sigma_i) = g(c_i, d_i)$，且：

$$c_i = c_0 + \frac{1}{2}\sum_{k=1}^{N} q(I_k = i)$$

$$d_i = d_0 + \frac{1}{2}\sum_{k=1}^{N} q(I_k = i)\begin{pmatrix} y_k^2 - 2 < \theta_i^T >_{q(\theta_i)} y_k x_k + \\ x_k^T < \theta_i\theta_i^T >_{q(\theta_i)} x_k \end{pmatrix}$$

根据伽马分布的相关知识，可得：

$$E(\eta_i) = \frac{a_i}{b_i} \tag{22}$$

$$E(\sigma_i) = \frac{c_i}{d_i} \tag{23}$$

最后通过非线性数值优化的方法求得局部模型宽度 O_i 的点估计，优化目标函数如下：

$$\max_{O_i, i=1,2,3} \sum_{k=1}^{N}\sum_{i=1}^{3} q(I_k = i)\ln\alpha_{ki}$$
$$s.t. \quad O_{i,\min} < O_i < O_{i,\max} \tag{24}$$

其中：$q(I_k = i)$ 由式（16）计算得到：

$$\alpha_{ki} = \frac{\lambda_{ki}}{\sum_{i=1}^{3}\lambda_{ki}}, \lambda_{ki} = \exp\left[\frac{-(H_k - H_i^0)^2}{2O_i^2}\right]$$

3.2.4　VB 辨识算法计算步骤

基于 VB 的参数辨识算法计算步骤总结如下。

1）根据式（10）给模型的未知参数 Θ 设置合适的先验分布 $p(\Theta)$，初始化未知参数 Θ、O 以及先验分布中的超参数 a_0, b_0, c_0, d_0。

2）E 步：关于隐变量最大化下界函数，根据式（17）计算 A_{ki}，并由式（16）得到隐变量 I_k 后验分布的更新式 $q(I_k)$。

3）M 步：固定隐变量，分别关于 $q(\theta_i)$、$q(\eta_i)$、$q(\sigma_i)$ 最大化下界函数，根据式（19）、式（22）以及式（23）得到各参数的期望 $E(\theta)$、$E(\eta)$ 以及 $E(\sigma)$，并根据式（24）计算局部模型有效宽度 O_i 的点估计。

4）根据以上两步得到的模型参数以及隐变量计算下界函数 F：

$$F = \sum_{k=1}^{N}\sum_{i=1}^{3} A_{ki} - KL[q(\Theta) \mid p(\Theta \mid O)]$$

5）将获得的估计值代入 2），不断迭代计算 2）～4），直至下界函数收敛。

由上述步骤得到模型的参数后，根据式（5）$\hat{y}_k = \sum_{i=1}^{3}\alpha_{ki}x_k^T\theta_i$ 计算得到青霉素浓度预测结果。

4　仿真验证

本文利用来自文献［13］的 10 个 100 000 L 的青霉素流加发酵罐产生的数据训练并测试模型以验证模型对实际工业环境的适应性。实际工业过程中采集到的数据夹杂着大量噪声，以发酵罐排放气体中的 CO_2 浓度（图 4）为例。

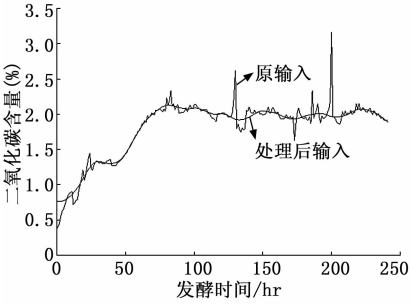

图 4　预处理前后排放气体中 CO_2 浓度图

为减少数据中的噪声对模型稳定性的影响，首先对输入数据中的异常值进行处理，由于发酵时间较长导致数据前后浮动较大，故本文首先根据发酵过程的阶段特性分阶段利用 3σ 准则剔除了输入数据中的异常值，然后对输入作滤波处理。

输入变量选择方面，在充分考虑反应机理并计算各输入变量与青霉素浓度相关系数后，选取 5 个过程变量作为输入变量，分别为：排放气体中 CO_2 百分比（%），排放气体中 O_2 百分比（%），pH 值，C 的生成率（g/min），发酵罐中物质总质量（kg）。利用本文第 1 节对葡萄糖流加速率进行聚类得到的典型工作点和经过预处理的输入数据，应用本文所述方法辨识得到子模型的参数，融合后得到青霉素浓度拟合曲线，如图 5 所示，其中三角形表示实际测量得到的数据。模型的质量通过相关误差进行测量，其计算公式如下：

$$Err = \frac{\mathrm{var}(y - \hat{y})}{\mathrm{var}(y)} \times 100\%$$

其中：var 为信号方差，y 为真实输出，$\hat{y}$ 为预测输出。

经过计算得到相关误差为 0.24%。

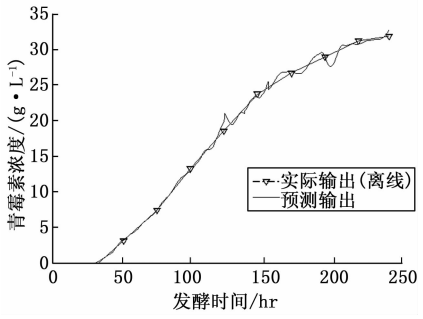

图 5 青霉素浓度拟合曲线

选取正常发酵情况下的另一批次数据对得到的模型进行测试，并与文献 [9] 中基于 EM 算法的青霉素发酵建模方法进行对比，预测结果如图 6 所示，计算得到本文模型与文献 [9] 的相关误差分别为 0.23%和 0.75%。

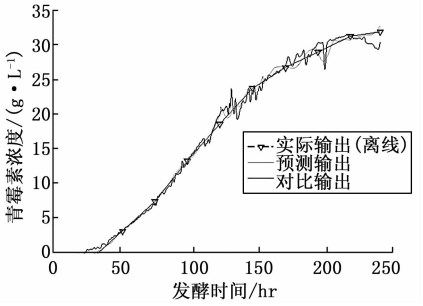

图 6 对比实验结果图

可以看出，通过对青霉素发酵准确的阶段划分并充分考虑工业环境中的变化因素，本文所建立的模型能够更好地预测实际工业环境中青霉素发酵过程的产物浓度，对青霉素生产过程的控制与优化具有一定的指示作用。

5 结束语

本文采用变分贝叶斯算法建立了基于不同工作点的青霉素发酵过程各阶段子模型，采用由调度变量计算得到的标准化权值将子模型进行融合，变分贝叶斯算法通过估计参数的后验分布，能够将发酵中过程变量的不确定利用均值、方差等统计特性解析地表达出来，在环境复杂的工业青霉素发酵过程中表现出优异的性能。仿真实验表明本文方法能够精确建立青霉素发酵过程产物浓度模型，在复杂的工业环境中准确地预测青霉素发酵过程。

参考文献：

[1] 张 粤. 青霉素发酵罐温度模糊控制 [J]. 计算机自动测量与控制，2002 (2)：115－117.

[2] Gao Y，Zhao Z，Liu F. DMFA－based operation model for fermentation processes [J]. Computers & Chemical Engineering，2018，109：138－150.

[3] Birol G，ündey C，Cinar A. A modular simulation package for fed－batch fermentation：penicillin production [J]. Computers & Chemical Engineering，2002，26 (11)：1553－1565.

[4] Bajpai R K，Reuss M. A mechanistic model for penicillin production [J]. Journal of Chemical Technology and Biotechnology，1980，30 (1)：332－344.

[5] Yao Y，Gao F. A survey on multistage/multiphase statistical modeling methods for batch processes [J]. Annual Reviews in Control，2009，33 (2)：172－183.

[6] Gao Y，Zhao Z，Liu F. DMFA－based operation model for fermentation processes [J]. Computers & Chemical Engineering，2018，109：138－150.

[7] Syddall M T. Improving the identification of a penicillin fermentation model [D]. University of Birmingham，1999.

[8] Guo F，Kodamana H，Zhao YJ，et al. Robust identification of nonlinear Errors－in－Variables systems with parameter uncertainties using variational Bayesian approach [J]. IEEE Transactions on Industrial Informatics，2017，13 (6)：3047－3057.

[9] 熊伟丽，姚 乐，徐保国. 基于 EM 算法的青霉素发酵过程多阶段融合建模 [J]. 化工学报，2014，65 (12)：4935－4041.

[10] Patnaik P R. Penicillin fermentation：mechanisms and models for industrial－scale bioreactors [J]. Critical reviews in microbiology，2001，27 (1)：25－39.

[11] 张晓磊，潘卫军，陈佳炀，等. 基于均值漂移与空间信息的导向模糊 C 均值遥感图像分割算法 [J]. 计算机测量与控制，2019，27 (11)：243－248.

[12] Mears L，Stocks S M，Albaek M O，et al. Mechanistic fermentation models for process design，monitoring，and control [J]. Trends in biotechnology，2017，35 (10)：914－924.

[13] Stephen G，Andrei S，David L，et al. The development of an industrial－scale fed－batch fermentation simulation [J]. Journal of Biotechnology，2015，193：70－82.

[14] 熊伟丽，姚 乐，徐保国. 混沌最小二乘支持向量机及其在发酵过程建模中的应用 [J]. 化工学报，2013，64 (12)：4585－4591.

[15] 李寒霜，赵忠盖，刘 飞. 基于变分贝叶斯算法的线性变参数系统辨识 [J]. 化工学报，2018，69 (7)：3125－3134.

[16] Chen J，Huang B，Ding F，et al. Variational Bayesian approach for ARX systems with missing observations and varying time－delays [J]. Automatica，2018，94：194－204.

[17] Yang X，Yin S. Variational Bayesian inference for FIR models with randomly missing measurements [J]. IEEE Transactions on Industrial Electronics，2016，64 (5)：4217－4225.

[18] Zhao Y，Fatehi A，Huang B. Robust estimation of ARX models with time varying time delays using variational Bayesian approach [J]. IEEE Transactions on Cybernetics，2018，48 (2)：532－542.

[19] Nasios N，Bors A G. Variational learning for Gaussian mixture models [J]. IEEE Transactions on Systems，Man，and Cybernetics，Part B (Cybernetics)，2006，36 (4)：849－862.

[20] Lu Y，Huang B，Khatibisepehr S. A variational Bayesian approach to robust identification of switched ARX models [J]. IEEE Transactions on Cybernetics，2016，46 (12)：3195－3208.