

基于 FT-NIR 光谱技术在线监测乙醇发酵过程的标定建模

王旭东, 刘 涛, 孙旭东

大连理工大学控制科学与工程学院, 辽宁 大连 116024

基金项目: 国家自然科学基金重点项目(61633006); 大连理工大学重点培育基金项目(DUT18ZD201)

通信作者: 刘涛, tliu@dlut.edu.cn 收稿/录用/修回: 2018-10-26/2019-03-14/2019-04-08

摘要

针对在线检测乙醇发酵过程中葡萄糖浓度、乙醇浓度和生物量问题, 提出了一种基于 FT-NIR 光谱技术在线检测这些参数的光谱标定建模方法. 采用偏稳健 M 回归 (PRM) 的方法消除了采集光谱异常值对于标定建模的影响, 给出了一种网格搜索寻优方法确定最优因子数和权重系数, 并从准确性、稳定性和分辨率三方面给出评价模型指标. 结果表明, PRM 方法建立的模型具有较好的预测效果. 最后, 通过对一个乙醇发酵过程的在线监测实验, 验证了所提标定建模和在线监测方法的有效性.

关键词

乙醇发酵过程
近红外光谱分析技术
偏稳健 M 回归
偏最小二乘法
模型评价
中图分类号: TP29
文献标识码: A

Calibration Modeling for On-line Monitoring of Ethanol Fermentation Processes Based on FT-NIR Spectra Technology

WANG Xudong, LIU Tao, SUN Xudong

School of Control Science and Engineering (SCSE), Dalian University of Technology, Dalian 116024, China

Abstract

Aiming at the on-line detection of glucose concentration, ethanol concentration, and biomass in ethanol fermentation process, we propose a spectral calibration model building method for measuring these parameters on-line by the Fourier transform-near infrared (FT-NIR) spectroscopy technology. We adopt the partial robust M-regression (PRM) method to eliminate the influence of collected spectral outliers to model calibration and present a grid-searching method to determine the optimal component number as well as the weighting coefficients for modeling. Moreover, we evaluate the model validity in terms of accuracy, stability, and resolution. The results show that the model established by the PRM method has better prediction effect. Finally, an on-line monitoring experiment for the ethanol fermentation process is performed to verify the effectiveness of the proposed calibration model and on-line monitoring method.

Keywords

ethanol fermentation process;
near-infrared spectra analysis
technology;
partial robust M-regression
(PRM);
partial least-squares (PLS)
method;
model evaluation

0 引言

生物发酵过程普遍具有多组份体系混合、反应机理复杂等特点, 实时在线监测发酵过程, 如各组分浓度和生物细胞量, 对提高生产质量和效益具有重要的指导作用^[1]. 传统的测量方法是利用酶标仪等测量吸光光度值 (OD) 代表生物量, 使用气相色谱仪和液相色谱仪测量底物和产物浓度, 这些方法需要对样本预处理, 测量耗时较长, 不能对发酵过程进行实时监测, 亦不能用于发酵过程的调控优化.

基于傅里叶变换的近红外光谱分析技术 (Fourier transform near-infrared spectra, FT-NIR) 是一种新发展的间接测

量技术, 近红外光能够深入样本内部, 并且不同谱区进入的深度不同, 取得不同层次物质的信息, 适应于对复杂样本的快速无损分析. 近红外光谱已逐渐应用于石化、制药、烟草、食品和化工等领域^[2-4]. 现已有少量文献介绍了利用近红外光谱技术检测生物发酵过程中重要的物质浓度. 例如, 文[5]基于近红外光谱技术离线检测乳酸链球菌发酵过程的还原糖浓度、乳酸链球菌效价、细胞浓度和 PH. 文[6]采用近红外光谱对 1, 3-丙二醇发酵过程的生物量进行了在线测量.

目前, 偏最小二乘法 (PLS) 因简便和有效性被较多地应用于分析光谱数据和标定建模^[7]. 但 PLS 对于过程数据的异常特征非常敏感, 异常点可能严重影响 PLS 建模的有

效性^[8]. 为了克服经典 PLS 的这一缺点, 一些稳健的 PLS 建模方法被提出^[9]. Cummins 等提出了迭代加权偏最小二乘(IRPLS)^[10], 该方法只对垂直异常点较敏感; Hubert 等提出了稳健简单偏最小二乘(RSIMPLS)的方法^[11], 对样本点采用了固定阈值加权; Serneels 等提出了连续加权的稳健偏 M 回归(PRM)^[12].

本文针对乙醇发酵过程研究葡萄糖浓度、乙醇浓度及生物量的在线监测方法. 采用 PRM 算法建立定量校正模型, 其计算速度相类似于较快的 IRPLS 且准确性优于 RSIMPLS 和 IRPLS, 迭代过程中采用了连续加权, 通过降低异常点的权重可以减小其影响. 为了保证模型预测精度和实现全局最优, 本文通过网络搜索寻优和交叉验证^[13]的策略来实现对 PRM 算法中因子数和权重函数中关键参数的寻优.

1 乙醇发酵在线监测系统与数据采集

1.1 实验平台和检测仪器

葡萄糖经酿酒酵母的糖酵解过程产生乙醇和二氧化碳, 释放少量能量. 在此反应过程中, 反应底物葡萄糖的浓度、反应产物乙醇的浓度及生物量是非常重要的参数.

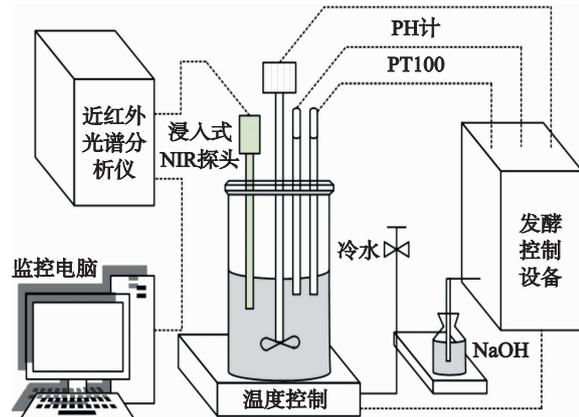


(a) 实验装置

本文的研究工作基于一个 NIR 技术的实时检测系统, 如图 1 所示. 该实验发酵反应釜容量为 2.5 L, 使用 PH 计采集酸碱度, 用 NaOH 控制反应罐中 PH 值, 反应釜温度由热电偶 PT100 采集, 利用加热装置和冷水将温度控制到指定温度, 采用电动搅拌桨对发酵液进行均匀的搅拌. 近红外光谱分析仪的结构主要分为 3 个部分: 光源模块、进样与光谱监测模块及分光、检测器与电子系统模块^[14], 通过以太网接口与监控计算机连接, 在发酵过程中近红外光谱分析仪实时采集发酵液在不同波长下的吸光度值(即近红外光谱), 将其传输到监控计算机中, 利用预先标定的葡萄糖浓度、乙醇浓度和生物量模型做实时计算和显示结果.

1.2 实验材料与操作条件

本实验使用的菌种为酿酒酵母(*saccharomyces cerevisiae* 4126). 发酵实验前需要对种子进行培养. 种子活化培养阶段中, 接种 100 μ L 菌种到活化培养基中, 接着将培养基置入在 30 $^{\circ}$ C、150 r/min 的摇床中进行 24 h 的活化. 将活化后的菌种接入种子培养基, 在 30 $^{\circ}$ C、150 r/min 的摇床中进行 12 h 的摇床培养, 种子培养基和活化培养基的成分组成如表 1 所示.



(b) 发酵过程检测示意图

图 1 乙醇发酵过程在线监测实验平台

Fig.1 Online monitoring platform for ethanol fermentation process

表 1 种子培养基与活化培养基成分

Tab.1 Components of seed medium and active medium

物质	浓度/(g/L)
葡萄糖	20
蛋白胨	20
Yeast extract powder	10

将经过培养活化的种子接种到发酵培养基中, 培养基均经过高温灭菌处理, 其成分组成如表 2 所示. 发酵过程中使用控制面板设定发酵罐 PH 值(5 ± 0.5)、温度(30 ± 0.5 $^{\circ}$ C)和搅拌桨转速(150 r/min). 发酵初始时期使用空气泵对发酵罐中进行通气, 直到尾气分析仪中显示发酵罐中气体组成成分和空气一致.

1.3 光谱和参考数据采集

本文实验采用 ABB 公司的 TALYS-ASP531 傅里叶近

表 2 发酵培养基成分

Tab.2 Components of fermentation medium

物质	浓度/(g/L)
葡萄糖	60
蛋白胨	3
Yeast extract powder	5

红外光谱分析仪以及配套的浸入式漫反射探头采集近红外光谱数据, 由 Horizon MB 软件进行光谱参数设置, 连续采集 3 次数据作为样品该时刻的近红外光谱数据. 光谱采集参数: 测量波数范围为 $4\ 790\ \text{cm}^{-1} \sim 12\ 000\ \text{cm}^{-1}$, 仪器分辨率为 $16\ \text{cm}^{-1}$, 光谱扫描次数为 64 次, 检测器增益为 237.84. 每隔 0.5 h 对发酵液进行取样一次, 对采集样本液进行离心, 稀释等处理后, 采用液相色谱仪(Waters)测量发酵液中葡萄糖浓度, 采用气相色谱仪(Agilent 6890 Se-

ries GC System) 测量发酵液中乙醇浓度, 采用 Multiskan Ascent 酶标仪测量发酵液中生物量. 这些离线测量数据用于标定建模对应检测的近红外光谱数据.

2 近红外光谱数据的标定建模

2.1 光谱和参考数据分析

近红外光谱为分子振动光谱, 其波段范围为 780 nm ~ 2 500 nm, 该段谱区主要包含的信息为分子中含氢基团 (如 C-H, N-H, O-H) 振动的合频与倍频特征信息, 这些信息包含了有机分子含氢基团的主要结构信息, 有机分子一般都包含这些含氢基团, 如蛋白质、糖和油脂等^[15]. 图 2 显示出了一个典型批次乙醇发酵过程的近红外光谱. 可见发酵液在合频区 (4 790 cm^{-1} ~ 5 000 cm^{-1}) 和 1 倍频区 (6 300 cm^{-1} ~ 7 500 cm^{-1}) 的平均吸收相对较强, 在高倍频区 (8 000 cm^{-1} ~ 12 000 cm^{-1}) 平均吸收相对较弱. 说明乙醇发酵过程对近红外光谱的吸收特征比较明显, 因此近红外光谱技术适合分析乙醇发酵过程中与含氢基团有直接或间接关系的成分, 如葡萄糖 ($\text{C}_6\text{H}_{12}\text{O}_6$)、乙醇 ($\text{C}_2\text{H}_6\text{O}$) 和生物量.

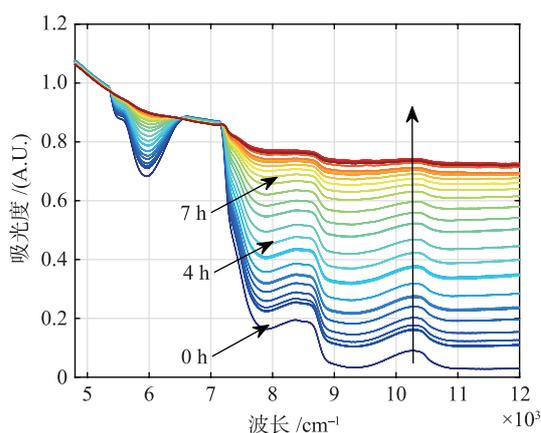


图 2 乙醇发酵过程光谱图

Fig.2 Spectra of ethanol fermentation process

由近红外光谱分析仪采集的近红外光谱数据为表观光谱, 样品的表观光谱中包含真实光谱信息和噪声. 为了消除工况环境给测量光谱带来的噪声, 采用了背景差减的方式去噪^[16], 即选用光学性质稳定的空气作为参考背景, 每次采样检测的近红外光谱数据由仪器测量的表观光谱与参考背景的差谱得到, 由此消除噪声引起的测量误差.

在上述实验平台进行 5 次发酵实验, 共得 369 个样本的数据样本点. 葡萄糖浓度、乙醇浓度及 OD 值的数据统计如表 3 所示. 其中对于葡萄糖组分明显偏离正常趋势的异常样本点按照采样时间顺序标记为 19, 20, 21, 82, 83, 84, 175, 176, 177, 190, 191, 192. 将全部数据点进行数值排序, 按 3:1 的比例从中挑选校正集和验证集, 最大和最小的浓度数据点归入校正集, 以保证校正集中的待测量物质浓度范围大于待预测的未知样本的浓度范围, 从而确保对未知样本预测是通过模型内插实现.

表 3 校正集与验证集数据统计

Tab.3 Data statistics of calibration set and validation set

样品集	样本个数	组分	最小值	最大值	平均值	标准偏差
校正集	279	葡萄糖	0	66.000 0	30.720 4	19.480 1
		乙醇	0	28.760 0	11.216 6	8.714 6
		OD	0.051 0	4.806 7	2.015 7	1.515 0
验证集	90	葡萄糖	0	60.000 0	31.866 7	19.414 0
		乙醇	0	26.380 0	11.746 1	8.684 6
		OD	0.055 0	4.556 7	2.005 7	1.489 3

2.2 建模方法

2.2.1 偏最小二乘法

PLS 在分解解释变量的同时引入预测变量, 消除解释变量间的线性相关性, 并使预测隐变量和解释隐变量间有最大的线性相关性^[17].

解释变量 X 包括 n 条光谱, p 个变量; 预测变量 Y 包括 n 组参考浓度, q 个变量:

$$X = \begin{bmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_n^T \end{bmatrix} \in \mathbb{R}^{n \times p}, Y = \begin{bmatrix} \mathbf{y}_1^T \\ \vdots \\ \mathbf{y}_n^T \end{bmatrix} \in \mathbb{R}^{n \times q} \quad (1)$$

$$\mathbf{x}_i \in \mathbb{R}^p, \mathbf{y}_i \in \mathbb{R}^q, i = 1, \dots, n$$

$$T = [t_1, \dots, t_h] \in \mathbb{R}^{n \times h} \quad (2)$$

投影 X 和 Y 到潜变量空间可得解释潜变量 T . 其中, h 为模型的潜变量数, PLS 模型可以表示为

$$X = TP^T + E \quad (3)$$

$$Y = TQ + R = XB + R \quad (4)$$

其中, P 为解释变量的载荷阵; E 代表观测的 X 与预测的 X 之间的差异; Y 为预测变量阵; Q 为解释隐变量阵与预测变量阵的回归系数; B 为模型回归系数:

$$B = W(P^T W)^{-1} Q \quad (5)$$

其中, W 为自变量权重向量, $P = X^T X W$ ^[18-19].

PLS 对所有样本都有一样的权重, 因此没有削弱异常样本点的能力.

2.2.2 稳健 M 回归 (PRM)

PRM 是能提高稳健性的偏最小二乘法, 其目的是构建一个能趋向于大多数数据的定量校正模型. 通过在数据集引入权重来减小对回归系数可能存在的异常影响. 坏的杠杆点和垂直异常点对校正模型均有影响, 为了削弱它们对模型预测准确性和稳健性的影响, PRM 引入了两种权重系数. 其中, 第 i 个样本的杠杆权重 w_i^* 定义为

$$w_i^* = \Psi \left(\frac{\|t_i - S(T)\|}{\text{med}_i \|t_i - S(T)\|}, c \right) \quad (6)$$

其中, $\|\cdot\|$ 为欧氏范数, $\Psi(z, c) = \frac{1}{(1 + |z/c|^2)}$ 为公平函数, c 是可调节常数, S 是对得分矩阵 T 取 L_1 中位数.

第 i 个样本的残差权重 w_i' 定义为

$$w_i' = \Psi \left(\frac{r_i}{\sigma}, c \right) \quad (7)$$

其中, r_i 为第 i 个样本点预测变量的观测值与预测值的残差; σ 为稳健的残余尺度估计, 由绝对中位偏差计算得到:

$$\sigma = \text{med}_i |r_i - \text{med}_j(r_j)| \quad (8)$$

则全局权重 w_i 可定义为

$$w_i = w_i^s w_i^r \quad (9)$$

所有样本的全局权重计算后经过归一化处理,通过迭代加权偏最小二乘实现 PRM 算法,计算中的权重会被一直计算直至收敛. PRM 算法的简要步骤为:

1) 利用式(6)、式(7)、式(9)计算稳健的初始权重 w_i , 其中初始残差 $r_i = y_i - \text{med}_j y_j$.

2) 使用重新加权的数据进行 PLS 回归分析,即将原始数据 \mathbf{X} 和 \mathbf{y} 中每行乘以权重 $\sqrt{w_i}$, PLS 的具体算法通过简单偏最小二乘算法实现 (SIMPLS). 将 PLS 运行的结果更新到回归系数 \mathbf{q} 和得分矩阵 \mathbf{T} 中,然后将得分阵的每行除以相应的 $\sqrt{w_i}$.

3) 重新计算残差 $r_i = y_i - \mathbf{t}_i \mathbf{q}$ 且使用式(6)、式(7)、式(9)更新权重.

4) 如果前后两次得到的回归系数的绝对残差大于预设阈值(如 10^{-2}),则跳转至第2)步. 如果收敛,则继续进行第5)步.

5) 最终的回归系数 $\hat{\beta}$ 由最后一次加权 PLS 得到.

2.2.3 网格搜索寻优权重调节值及最佳因子数

上述算法权重函数中 c 为可调节参数,增大 c 的值会减小排除异常值的能力,反之可以会消除较多的异常值. 其值通常由经验决定,没有结合实际模型实现全局最优. 本文通过网格搜索寻优交叉验证的方法获得最佳因子数 k 和可调节参数 c . 采用留一交叉验证 (leave one out cross validation, LOOCV) 获得可靠的模型参数. LOOCV 在训练样本集中每次留下一个样本点,使用训练集中其它样本点建立校正模型,然后对剩下的样本点进行测试,直至训练集中每一个样本点都被留下一次,把全部测试所得的误差的平均值作为参数指标. 网格搜索寻优即让因子数 k 和 c 遍历在一定范围内划分的等步长网格,其中 c 和因子数的划分范围分别参照文[11]和文[12]给出, c 的取值范围为 $2^{-8} \sim 2^8$,为了提升计算效率,选用指数步长^[20],最大因子数 $k_{\max} = \min\{p, 10\}$,网格中每一组因子数和 c 都是一组潜在解,遍历网格所有的参数进行交叉验证,计算交叉验证均方根误差 (RMSECV):

$$\text{RMSECV} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_{-i})^2} \quad (10)$$

其中, y_i 是预测变量中第 i 个样本的测量值; \hat{y}_{-i} 是每次留下第 i 个样本,用其它数据点建模并预测它得到的预测值,取较小交叉验证均方根误差对应的参数为模型参数.

2.3 模型评价指标

近红外光谱分析模型的性能指标包括模型的准确性、稳定性和分辨率三个方面^[21-23].

1) 准确性:模型的准确性通过模型的预测均方根误差 RMSEP 来评价,它的值越小说明模型预测准确性越好. 评价指标为

$$\text{RMSEP} = \sqrt{\frac{1}{n_p} \sum_{i=1}^n (y_{pi} - \hat{y}_{pi})^2} \quad (11)$$

其中, y_{pi} 为验证集的第 i 个样本参考值, \hat{y}_{pi} 为验证集的第 i 个样本预测值.

2) 稳定性:模型的稳定性由预测均方根误差和校正均方根误差的比值 (RMSEP/RMSEC) 来评价. RMSEC 的形式为

$$\text{RMSEC} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (12)$$

如果 RMSEP/RMSEC < 1.2, 则说明模型的稳定性较好,不会产生过拟合的现象.

3) 分辨率:模型的分辨率由性能对标准差之比 (RPD) 指标判别,指标形式为

$$\text{RPD} = \frac{\text{SD}}{\text{RMSEP}} = \frac{\sum_{i=1}^n (y_i - y_m)^2}{\sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (13)$$

其中, SD 为校正集样本分布标准差,可以代表样本分布范围; y_m 是待预测样本均值.

满足 RPD > 5 的定量分析模型可应用于可以接受的质量预测; 满足 RPD > 10, 则可以应用于实时过程控制和优化.

3 结果与讨论

图 3 给出了 PLS、RSIMPLS、IRPLS 与 PRM 方法对葡萄糖含量的拟合关系图. 图 3(a) 是离线检测参考值同 PLS 方法的预测值得到的葡萄糖含量的拟合图,可见观测点比较均匀分布在拟合线的两侧. 图 3(b) ~ 图 3(d) 分别是离线检测参考值同 RSIMPLS、IRPLS 和 PRM 方法的预测值得到的葡萄糖含量的拟合图,相对于 PLS 方法,更多的点贴近拟合线. 这说明 PLS 方法更易受到异常点的影响. 相比之下,3 种稳健的算法可以在一定程度上克服异常点的影响,PRM 方法能取得较贴近的拟合效果.

图 4 示出了应用 PRM 方法建立葡萄糖浓度模型的样本权重系数. 可以看到该葡萄糖浓度模型中偏离正常趋势的样本点对应较小的权重系数,由此说明 PRM 算法能有效减小异常样本点对建模的影响.

表 4 为 PLS、RSIMPLS、IRPLS 和 PRM 算法建立的葡萄糖浓度模型、乙醇浓度模型和生物量的预测性能指标,通过网格搜索寻优得到 PRM 方法建立的葡萄糖浓度模型、乙醇浓度模型和生物量模型的可调节常数 c 分别为 2^4 、 2^5 、 2^8 ,最佳主成分数为 9、10、10,其中 3 个组分的 RMSEP 均为 PRM 模型最小,因此本文提出方法建立的模型有更好的准确性. 相应的 PRM 建立的 3 个模型都有 RMSEP/RMSEC < 1.2,说明建立的模型都具有较好的稳定性. 从分辨率来看,PRM 建立模型的 RPD 均较大,尤其指出,对于生物量 (OD) 预测模型的 RPD 为 13.422 1,说明该模型可用于过程实时控制和优化,葡萄糖浓度预测模型和乙醇浓度预测模型的 RPD 亦都大于 5,说明由此建立的定量分析模型能用于可以接受的质量预测.

将上述模型用于在线检测时,对于可能出现的光谱数据异常值,可以通过光谱信号预处理方法进行排除,并且可以利用发酵过程历史数据和变化趋势来判断是否当前光

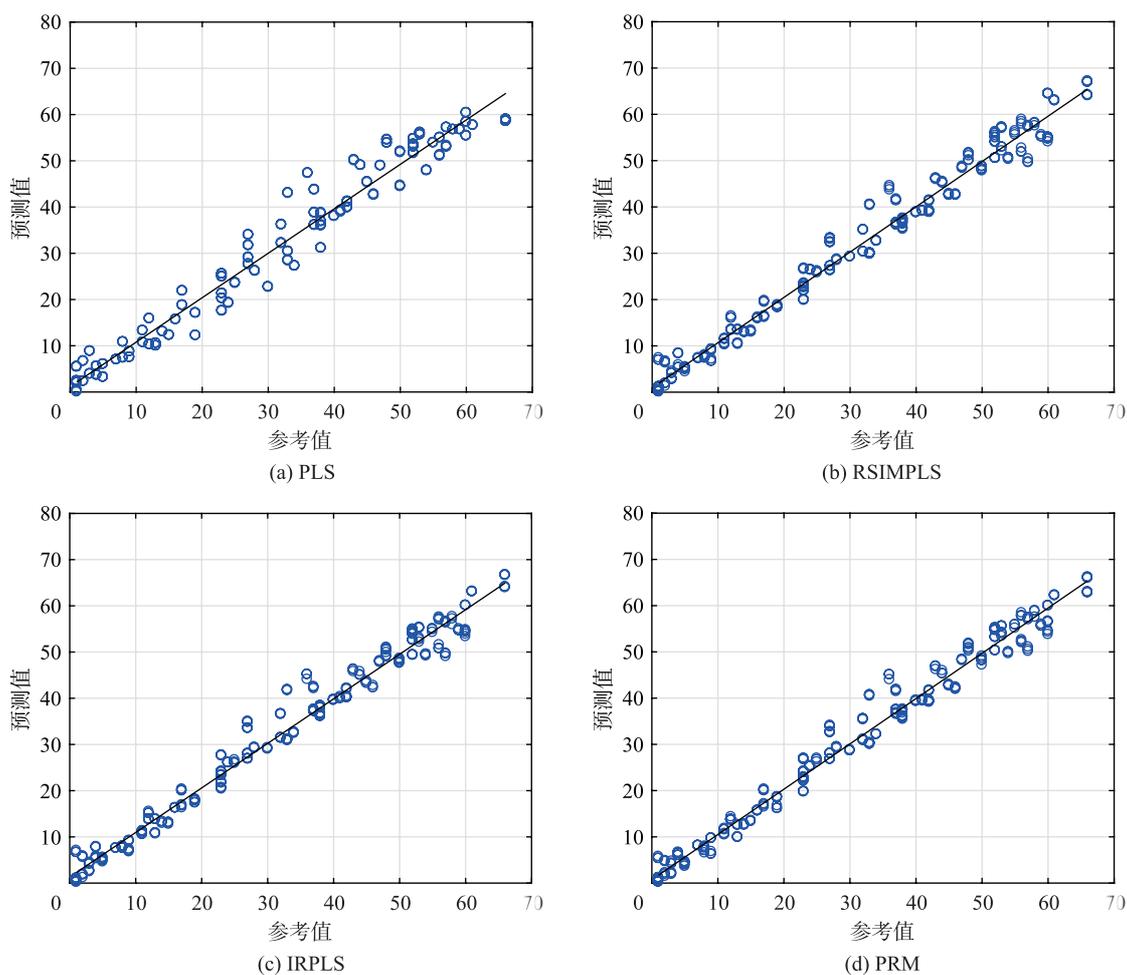


图3 葡萄糖浓度测定的拟合结果

Fig.3 Fitting results of glucose concentration measurement

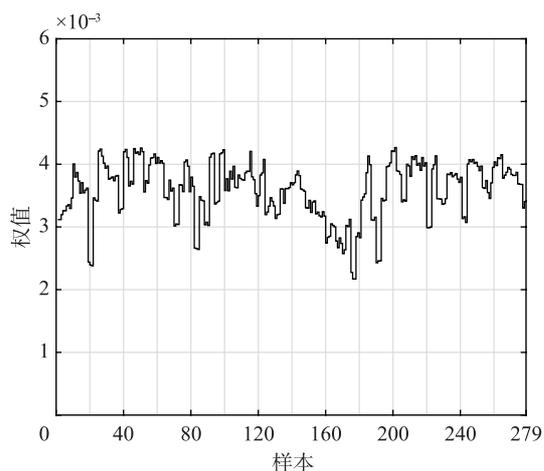


图4 PRM模型的权重系数

Fig.4 Weight coefficients of the PRM model

谱数据是否出现异常. 可供参考的一个策略是: 若一次出现异常, 则维持该预测值为上一时刻预测值, 若连续多次的预测值出现异常情况, 应需要对发酵液进行取样做离线检测, 以判断异常来自光谱数据采集还是发酵过程状况.

表4 不同建模方法的预测性能指标对比

Tab.4 Comparison between predictive performance indices of different modeling methods

方法	物质	RMSEP	RMSEP/RMSEC	RPD
PLS	葡萄糖	4.421 0	1.157 7	4.406 3
	乙醇	1.625 6	0.913 0	5.360 9
	OD	0.185 1	0.937 0	8.186 4
RSIMPLS	葡萄糖	3.813 3	1.366 2	5.108 4
	乙醇	1.550 7	0.749 5	5.619 7
	OD	0.142 0	0.867 6	10.668 2
IRPLS	葡萄糖	3.658 5	1.287 5	5.324 7
	乙醇	1.311 6	0.818 7	6.644 2
	OD	0.120 8	0.979 5	12.541 4
PRM	葡萄糖	3.530 6	1.196 3	5.517 5
	乙醇	1.246 3	0.875 5	6.992 4
	OD	0.112 9	1.188 8	13.422 1

4 实验验证

通过一个批次发酵实验对已建立的乙醇发酵在线监测模型进行验证. 图5为上述建立的模型对发酵过程葡萄糖

浓度、乙醇浓度和生物量的在线监测结果. 由图 5 可知, 上述方法建立的模型可以有效的对乙醇发酵过程进行在线监测, 葡萄糖浓度、乙醇浓度、OD 检测值同传统方法的测

量参考值趋势一致. 此外, 葡萄糖浓度、乙醇浓度和生物量都能在线监测 (1 min 检测一次), 比传统的检测方式 (约 30 min 得到一个检测结果) 显著提高效率.

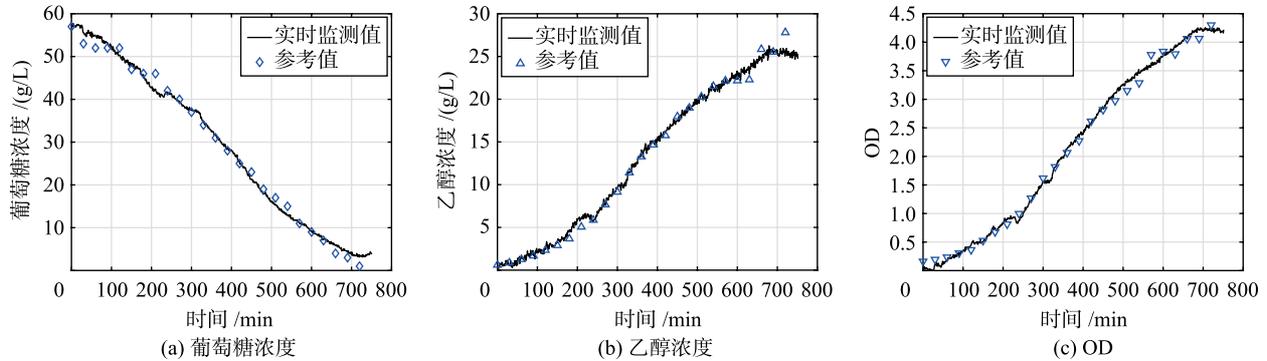


图 5 乙醇发酵过程的近红外光谱在线监测结果

Fig.5 On-line monitoring result of ethonal fermentation process using NIR spectra

5 结论

本文提出了一种针对乙醇发酵过程葡萄糖浓度、乙醇浓度和生物量的近红外光谱标定建模和在线实时监测方法, 相较于传统方法有更好的在线检测准确性和实时性. 为了克服近红外光谱采集数据中异常值的不利影响, 给出了一种基于网格搜索参数寻优的 PRM 建模方法, 并从准确性、稳定性和分辨率系统评价了乙醇发酵过程各组分浓度 NIR 模型, 结果显示 PRM 方法建立的 3 个模型均有最

小的 RMSEP, 3 个模型的 RPD 分别为 5.517 5、6.992 4、13.422 1. 本文建立的生物量模型可用于实时过程控制, 葡萄糖浓度和乙醇浓度模型也可以用于质量预测. 通过对乙醇发酵过程的在线监测实验, 有效地验证了本文给出的方法对葡萄糖浓度、乙醇浓度和生物量在线检测的有效性. 针对更复杂的发酵过程, 如何对发酵过程相近成分含量进行有效监测、发酵过程中的故障诊断及质量控制有待于进一步的研究.

参考文献

- [1] 张嗣良. 发酵工程原理[M]. 北京: 高等教育出版社, 2013: 1-9.
Zhang S L. Principles of fermentation engineering[M]. Beijing: Higher Education Press, 2013: 1-9.
- [2] 褚小立, 陆婉珍. 近五年我国近红外光谱分析技术研究与应用进展[J]. 光谱学与光谱分析, 2014, 34(10): 2595-2605.
Chu X L, Lu W Z. Research and application progress of near infrared spectroscopy analytical technology in china in the past five years[J]. Spectroscopy and Spectral Analysis, 2014, 34(10): 2595-2605.
- [3] 冯红年, 陈英斌, 黎世文, 等. 在线近红外光谱分析技术及应用[J]. 控制工程, 2010(s1): 96-99.
Feng H N, Chen Y B, Li S W, et al. On-line near infrared spectroscopy analytical technique and its applications[J]. Control Engineering of China, 2010(s1): 96-99.
- [4] Lourenço N D, Lopes J A, Almeida C F, et al. Bioreactor monitoring with spectroscopy and chemometrics: A review[J]. Analytical & Bioanalytical Chemistry, 2012, 404(4): 1211-1237.
- [5] Fan G F, Yan J, Li G M, et al. The application of near infrared spectroscopy in process monitoring of solid-state fermentation of sweet sorghum stalks[J]. Journal of Near Infrared Spectroscopy, 2015, 23(5): 293.
- [6] 王路, 刘涛, 陈洋, 等. 基于傅里叶变换近红外光谱实时分析 1, 3-丙二醇发酵过程生物量的在线监测方法[J]. 生物工程学报, 2017, 33(1): 68-78.
Wang L, Liu T, Chen Y, et al. On-line monitoring method for 1, 3-propanediol fermentation biomass based on FT-NIR real-time analysis[J]. Chinese Journal of Biotechnology, 2017, 33(1): 68-78.
- [7] Giovenzana V, Beghi R, Guidetti R. Rapid evaluation of craft beer quality during fermentation process by vis/NIR spectroscopy[J]. Journal of Food Engineering, 2014, 142(142): 80-86.
- [8] Rousseeuw P J, Hubert M. Robust statistics for outlier detection[J]. Wiley Interdisciplinary Reviews Data Mining & Knowledge Discovery, 2011, 1(1): 73-79.
- [9] Yin S, Ding S X, Zhang P, et al. Study on modifications of PLS approach for process monitoring[J]. IFAC Proceedings Volumes, 2011, 44(1): 12389-12394.

- [10] Cummins D J, Andrews C W. Iteratively reweighted partial least squares: A performance analysis by monte carlo simulation[J]. *Journal of Chemometrics*, 1995, 9(6): 489 – 507.
- [11] Hubert M, Branden K V. Robust methods for partial least squares regression[J]. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 2003, 17(10): 537 – 549.
- [12] Sermeels S, Croux C, Filzmoser P, et al. Partial robust M-regression[J]. *Chemometrics & Intelligent Laboratory Systems*, 2005, 79(1/2): 55 – 64.
- [13] 於帮兵, 王华忠, 颜秉勇. 基于长短时记忆网络的工业控制系统入侵检测[J]. *信息与控制*, 2018, 47(1): 54 – 59.
Yu B B, Wang H Z, Yan B Y. Intrusion detection of industrial control system based on long short term memory[J]. *Information and Control*, 2018, 47(1): 54 – 59.
- [14] 陆婉珍, 袁洪福, 褚小立. 近红外光谱仪器[M]. 北京: 化学工业出版社, 2010: 62 – 69.
Lu W Z, Yuan H F, Chu X L. Near infrared spectroscopy instrument[M]. Beijing: Chemical Industry Press, 2010: 62 – 69.
- [15] 杰尔·沃克曼, 洛伊斯·文依. 近红外光谱解析实用指南[M]. 北京: 化学工业出版社, 2009: 1 – 18.
Workman J, Weyer L. Practical guide to interpretive near-infrared spectroscopy[M]. Beijing: Chemical Industry Press, 2009: 1 – 18.
- [16] Amerov A K, Chen J, Arnold M A. Molar absorptivities of glucose and other biological molecules in aqueous solutions over the first overtone and combination regions of the near-infrared spectrum[J]. *Applied Spectroscopy*, 2004, 58(10): 1195.
- [17] 罗晓, 陈耀, 孙优贤. 基于统计回归的质量推断方法[J]. *信息与控制*, 2001, 30(5): 422 – 426.
Luo X, Chen Y, Sun Y X. Statistical-regression-based quality inference methods[J]. *Information and Control*, 2001, 30(5): 422 – 426.
- [18] 汤健, 贾美英, 刘卓, 等. 基于偏最小二乘算法的高维谱数据特征选择[J]. *控制工程*, 2015, 22(6): 1127 – 1130.
Tang J, Jia M Y, Liu Z, et al. Feature selection approach of high dimension spectral data base-on partial least squares algorithm[J]. *Control Engineering of China*, 2015, 22(6): 1127 – 1130.
- [19] 王桂增, 叶昊. 主元分析与偏最小二乘法[M]. 北京: 清华大学出版社, 2012: 200 – 233.
Wang G Z, Ye H. Principal component analysis and partial least squares[M]. Beijing: Tsinghua University Press, 2012: 200 – 233.
- [20] Chang C C, Lin C J. LIBSVM: A library for support vector machines[J]. *ACM transactions on intelligent systems and technology (TIST)*, 2011, 2(3): 27.
- [21] 严衍禄, 陈斌, 朱大洲, 等. 近红外光谱分析的原理、技术与应用[M]. 北京: 中国轻工业出版社, 2013: 162 – 175.
Yan Y L, Chen B, Zhu D Z, et al. Near infrared spectroscopy-principles, technologies and applications[M]. Beijing: China Light Industry Press, 2013: 162 – 175.
- [22] Blanco M, Peguero A. Influence of physical factors on the accuracy of calibration models for NIR spectroscopy[J]. *Journal of Pharmaceutical & Biomedical Analysis*, 2010, 52(1): 59 – 65.
- [23] Vohland M, Michel K, Ludwig B. Use of near-infrared spectroscopy to distinguish carbon and nitrogen originating from char and forest-floor material in soils; usefulness of a genetic algorithm[J]. *Journal of Plant Nutrition & Soil Science*, 2011, 174(5): 695 – 701.

作者简介

王旭东(1994 –), 男, 硕士生. 研究领域为工业过程检测技术与微生物发酵过程质量监测.

刘涛(1974 –), 男, 博士, 教授. 研究领域为工业过程辨识建模、在线监测、鲁棒控制与运行优化.

孙旭东(1962 –), 男, 学士, 副教授. 研究领域为工业过程检测技术与先进控制方法.