

连续幂系数回归在人体血糖无创检测中的应用研究

李庆波, 刘杰强, 张广军*

北京航空航天大学仪器科学与光电工程学院, 精密光机电一体化技术教育部重点实验室, 北京 100191

摘要 血糖浓度的准确检测对于糖尿病的治疗具有重要的意义。本文采用连续幂系数回归方法有效地提高了近红外无创检测人体血糖浓度的预测精度。该方法是传统偏最小二乘法(PLS)的扩展, 实现简易, 且当幂系数选取恰当时, 能够明显地提高预测精度。应用该方法分别建立了四成分葡萄糖实验和人体口服葡萄糖实验的定量分析模型, 并利用该模型对预测集样本进行预测。实验结果表明, 与 PLS 相比, 该方法建立的定量分析模型不仅可以提高预测精度, 而且可以针对不同的测量对象设定不同的幂系数以达到最佳的建模效果。根据不同个体灵活地选取幂系数, 对于人体血糖浓度近红外无创检测研究具有很大的应用价值。

关键词 连续幂系数回归; 近红外光谱; 血糖浓度; 偏最小二乘法; 无创检测

中图分类号: O657.3 文献标识码: A DOI: 10.3964/j.issn.1000-0593(2011)06-1481-05

引言

糖尿病是一种由于血糖失控高出正常水平所造成的全身性进行性疾病, 并发症多, 且严重。据统计, 每 10 秒钟世界上就有 1 名糖尿病患者死亡; 到 2050 年, 全球糖尿病患者人数将猛增至 3.8 亿。截止 2007 年, 全球糖尿病患者人数已达 2.46 亿; 其中中国人占 4000 万(仅次于美国)。目前, 对糖尿病有效的治疗方法是要求患者经常对自身血糖浓度进行检测, 以便将血糖浓度控制在安全范围内。

血糖检测方法可以分为有创检测和无创检测。血糖有创检测需要对患者抽血, 会给患者带来不便、疼痛, 并且存在安全隐患、容易交叉感染、检测成本高、无法做到连续的实时检测。为了克服血糖有创检测的以上缺点, 血糖无创检测技术已经成为国内外相关领域的研究热点。相对于近年来出现的电化学方法、光学方法、阻抗法等无创检测方法, 近红外光谱分析技术具有高效、快速、易操作等优点^[1], 能够在几秒至几分钟内同时测定样本中多种成分的浓度或者性质参数。

近红外光谱分析技术已经在农业、食品、化工等领域得到了应用^[2], 将其用于血糖无创检测具备充足的理论依据。在化学计量学和计算机技术的推动下, 近红外光谱定量分析的稳定性、快速性和准确性不断地得到了提高, 成为了目前血糖无创检测研究的首选方法。然而, 由于受到个体差异、

信号微弱、测量条件等因素的制约, 现阶段近红外光谱分析技术在血糖无创检测中的应用还处于研究阶段^[3-7]。目前, 不同测量对象吸收光谱之间的差异使血糖定量分析模型只能以个体为单位。传统的光谱分析主要采用偏最小二乘法^[8]建立分析模型, 但是个体之间的差异在一定程度上制约着 PLS 预测精度的进一步提高。为了减小个体差异对定量分析模型的影响, 可以设想对不同的个体和环境设定不同的系数以提高预测精度。而连续幂系数回归(continuum power regression, CPR)就能够很好地解决这一问题。

本文研究了连续幂系数回归在人体血糖浓度近红外无创检测中的应用。CPR 是由 Wise 和 Ricker 提出的一种可改变幂系数的回归方法^[9], 传统的最小二乘法(OLS)、主成分回归(PCR)和偏最小二乘法都只是该方法某个幂系数下的特例。文中给出了 CPR 的实现算法, 并利用该方法建立了离体实验(四成分葡萄糖水溶液实验)数据和人体无创实验(人体口服葡萄糖耐量实验)数据各种幂系数下的定量分析模型, 探讨了 CPR 在人体血糖浓度近红外无创检测的可行性和优越性。

1 CPR 算法介绍

PLS 作为当前化学计量学中定量分析的传统方法, 已经在很多领域得到了应用。但是, 应用 PLS 建立的定量分析模型容易受到样本中可能存在奇异点的影响, 导致其预测精度

收稿日期: 2010-09-15, 修订日期: 2010-11-10

基金项目: 国家自然科学基金项目(60708026), 北京航空航天大学蓝天新星项目和长江学者和创新团队发展计划资助

作者简介: 李庆波, 女, 1975 年生, 北京航空航天大学仪器科学与光电工程学院副教授 e-mail: qbleebuaa@buaa.edu.cn

* 通讯联系人 e-mail: gjzhang@buaa.edu.cn

© 1994-2012 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

下降。作为一种全面的线性回归方法, CPR 不仅包含了 PLS 的优点, 而且可以通过设定合适的幂系数减弱奇异点对分析模型的影响, 进而提高其预测精度。CPR 算法^[10]的具体实现过程如下:

设光谱矩阵为 $X(m \times n)$, 浓度矩阵为 $Y(m \times 1)$ 。

(1) 通过交互验证选取合适的主成分个数 pcs。

(2) 计算矩阵 $\tilde{K}^Y$ 。

①设定幂系数 $\alpha(0 \sim 1)$, 由幂系数 α 得出相应的 γ 值

$$\gamma = \frac{\alpha}{1 - \alpha} \quad (1)$$

②通过奇异值分解可得

$$X = UL^{1/2}V^T \quad (2)$$

其中 U 为 $m \times q$, V 为 $n \times q$, 且满足 $U^T U = V^T V = E_q$ (q 为 X 矩阵的秩); $L^{1/2}$ 为非零特征值 $\lambda_i^{1/2}$ ($i = 1, \dots, q$) 构成的对角矩阵。

③令

$$\tilde{K}^Y = [pE^Y, pE^{2Y}, \dots, pE^{pcs \cdot Y}] \quad (3)$$

其中 $p = U^T Y / |Y|$ 。

(3) 求出得分矩阵 T 。

$$T = U\tilde{T} \quad (4)$$

其中 $\tilde{T} = GS(\tilde{K}^Y)$ (GS 为格莱姆-施密特正交化)。

(4) 计算线性回归模型的回归系数 β_{CPR} 。

由线性回归模型

$$Y = X\beta + \xi \quad (5)$$

可以计算 β 的估计值 $\hat{\beta}$ 为

$$\hat{\beta} = \beta_{CPR} = |Y| V^{1/2} \tilde{T}^T p \quad (6)$$

其中 $\beta(n \times 1)$ 为回归系数, ξ 为测量误差。

2 材料与方法

2.1 牛血红蛋白和白蛋白的葡萄糖水溶液实验

牛血红蛋白和白蛋白的葡萄糖水溶液实验(又称四成分实验)光谱采集范围为 1 000~2 500 nm, 数据获取 1 501 个波长点。

(1) 实验仪器: 美国 Pekin-Elmer 公司 Spectrum GX 傅里叶变换红外光谱仪, 精密电子天平, 容量瓶, 液态氮冷却 InSb 检测器, 1 mm 石英样本池, 蠕动泵自动进样系统, 日本 TBA30 全自动生化分析仪。

(2) 实验试剂来源: 牛血红蛋白由上海丽珠东风生物技术有限公司提供, 牛血白蛋白由 TBD 生物科技发展有限公司提供, 葡萄糖由天津化学试剂三厂提供。

(3) 四成分实验样本配制: 称取牛血红蛋白、白蛋白和葡萄糖放入容量瓶, 再加入生理盐水至 100 mL。配置 44 个样本, 牛血红蛋白浓度范围为 0~1 000 mg · dL⁻¹, 牛血白蛋白浓度范围为 0~400 mg · dL⁻¹, 两者随机配比; 葡萄糖浓度范围为 100~5 000 mg · dL⁻¹, 均匀分布。

2.2 人体无创实验

人体口服葡萄糖耐量实验(又称人体无创实验)光谱采集范围为 1 000~2 500 nm, 数据获取 3 112 个波长点。

(1) 实验仪器: Nicolet FT-IR6700 光谱仪, InGaAs 2.6

nm 探测器, CaF₂ 光学分束器和白光光源, 光谱扫描范围为 1 000~2 500 nm, 分辨率设置为 4; 强生稳豪 II 代血糖仪测量血糖参考值。

(2) 实验试剂来源: 北京燕京药业生产的口服葡萄糖。

(3) 人体无创实验步骤: 选一名健康志愿者为受试者, 在实验前禁食 8 h, 实验时 5 min 内饮入 100 mL 含 75 g 葡萄糖的水溶液, 每 6 min 采集手指指腹的近红外光谱, 每次实验时尽量保持测量的外界条件(测量部位、压力等)相同。采集光谱的同时, 采用强生便携式血糖仪测定血糖浓度作为建模参考值。每测定一次血糖参考值时采集一次近红外光谱, 共采集 28 个样本, 每个样本包含 3 112 个数据点, 样本的血糖参考值范围为 75.6~243.0 mg · dL⁻¹。

3 结果与分析

3.1 实验原始光谱及建模波段的选取

图 1 和图 2 中的光谱保留了丰富的物质吸收信息, 可以用于建立相关物质的定量分析模型。

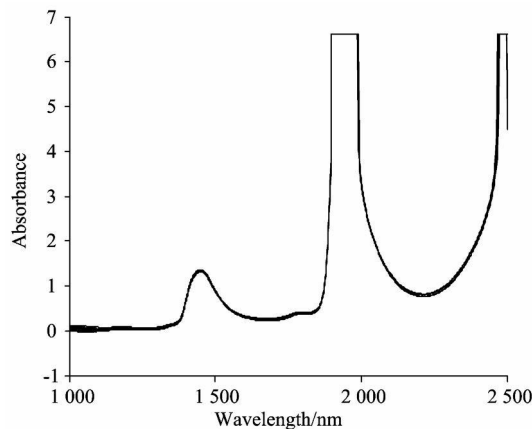


Fig 1 Original spectra of four ingredient experiment

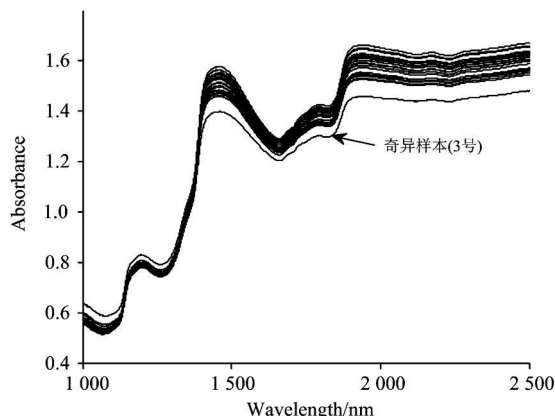


Fig 2 Original spectra of noninvasive experiment of body

但是, 由四成分实验的原始光谱可知, 吸光度在 1 950 nm 左右的波段出现了饱和现象。由于葡萄糖在 2 273 和 2 326 nm 附近有两个吸收峰; 在 1 923 和 2 632 nm 附近为水的吸收峰, 2 222 nm 附近水吸收较弱, 且随温度漂移较小,

由此可知波段 2 200~ 2 400 nm 的信噪比较高^[11]。综合考虑吸光度饱和现象和信噪比高低对葡萄糖定量分析模型的影响,四成分实验可以采用 2 200~ 2 400 nm 的波段建立模型。图 3 给出了四成分实验建模的光谱图。

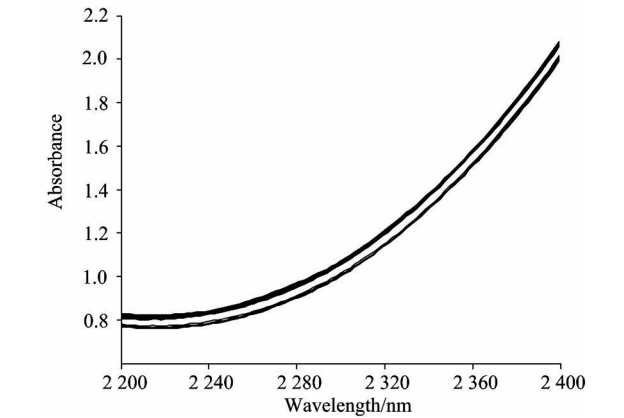


Fig 3 Spectra of model of four ingredient experiment

在人体无创实验中 3 号、4 号、13 号、17 号和 19 号样本均为坏样本,会产生奇异光谱(在人体无创实验光谱图中只标出了 3 号奇异样本产生的奇异光谱)。为了验证 CPR 可以在一定程度上减弱奇异值对建模分析的影响,对奇异样本不予剔除。在人体无创实验的原始光谱中并没有出现吸光度的饱和现象,并且在全谱中包含多个吸收峰和吸收谷。因此,人体无创实验可以采用全谱进行建模。

3.2 建立模型

- (1) 数据预处理。本文对光谱数据主要采用简单的中心化预处理。
- (2) 分配校正集和预测集。在四成分实验样本中选取 30 个样本作为校正集,剩余 14 个样本作为预测集;在人体无创实验样本中选取 19 个样本作为校正集,剩余 9 个样本作为预测集。
- (3) 选取主成分个数。采用“留一法”进行交互验证,根据交互验证的均方根误差(RMSECV)选取合适的主成分个数。
- (4) 选取幂系数 α 值。根据不同的 α 值建立相应的 CPR 定量分析模型,并对模型的预测能力进行比较与评估。

3.3 各种模型间的数据对比

表 1 和表 2 分别给出了不同 α 值下四成分实验和人体无

Table 2 Comparison of data of different α in the model of noninvasive experiment of body

α	pcs	RMSEP/(mg · dL ⁻¹)	r
0.2	7	10.41	0.958 2
0.3	7	9.94	0.961 8
0.4	8	10.87	0.959 4
0.5(PLS)	10	10.83	0.955 6
0.6	12	10.69	0.957 4
0.7	13	10.57	0.958 1
0.8	12	17.56	0.865 6

创实验定量分析模型的建模数据和评估数据;图 4 和图 5 分别给出了四成分实验和人体无创实验最佳定量分析模型(包括预测模型和校正模型)的回归直线。

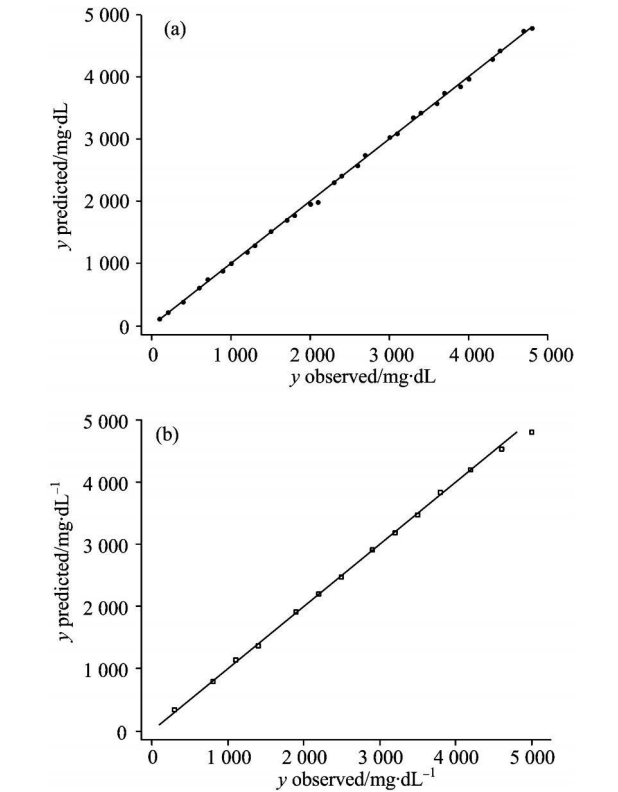


Fig 4 The four ingredient experiment's correlation ($\alpha = 0.4$) of the observed Y and the predicted Y with the calibration models (a) and prediction models (b)

传统的光谱分析方法 PLS 只是幂系数 $\alpha = 0.5$ 时 CPR 的一个特例。通过比较 PLS 和 CPR 两种方法建立的定量分析模型,可以对 CPR 在人体血糖浓度无创检测中应用的可行性和优越性进行评估。

3.4 实验数据分析与讨论

由表 1 和表 2 可知,在四成分实验和人体无创实验定量分析模型中,预测集样本的预测值与参考值之间的相关系数都很高,根据相关系数无法分辨出不同幂系数下模型的优劣。但是,从预测集的预测均方根误差可以很容易地判别出各种模型的差异。且当 $\alpha = 0.4$ 时,四成分实验的预测均方根

Table 1 Comparison of data of different α in the model of four ingredient experiment			
α	pcs	RMSEP/(mg · dL ⁻¹)	r
0.2	6	72.71	0.998 7
0.3	5	62.41	0.999 2
0.4	5	60.74	0.999 4
0.5(PLS)	7	67.53	0.999 3
0.6	6	65.98	0.999 4
0.7	10	66.76	0.999 4
0.8	8	66.91	0.959 4

误差达到最小($60.74 \text{ mg} \cdot \text{dL}^{-1}$); 当 $\alpha = 0.3$ 时, 人体无创实验的预测均方根误差达到最小($9.94 \text{ mg} \cdot \text{dL}^{-1}$)。因此, 可以分别将 $\alpha = 0.4$ 和 $\alpha = 0.3$ 作为四成分实验与人体无创实验的最佳幂系数, 从而可以在该幂系数下建立最佳定量分析模型。

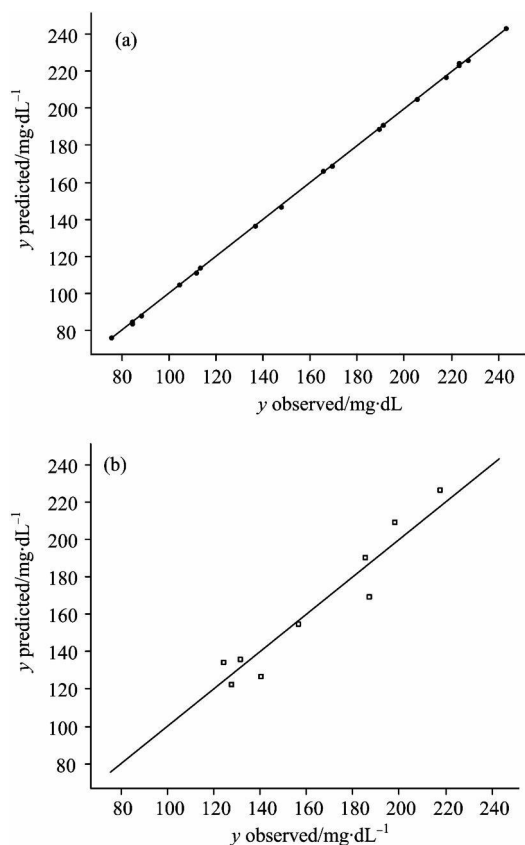


Fig 5 The noninvasive experiment of body's correlation ($\alpha = 0.3$) of the observed Y and the predicted Y with the calibration models (a) and prediction models (b)

从校正模型的回归直线[图 4(a)和图 5(a)]中可以很明显地看出, 四成分实验和人体无创实验校正集的预测值和参考值之间具有很高的相关性, 可以很好地满足实验建模的要求。从预测模型的回归直线[图 4(b)和图 5(b)]中可知, 四成分实验和人体无创实验的预测结果都比较理想, 并且比较人体无创实验的预测值和参考值可得人体无创实验的相对误差为 $\pm 9.66\%$, 能够达到预期的效果。

表 1 和表 2 的数据对比以及图 4 和图 5 的相关性分析, 说明 CPR 在人体血糖浓度无创检测中的应用是可行的。比较不同幂系数 α 值下的预测均方根误差可以很清楚地辨别出 CPR 相对于 PLS 的优越性。通过选取合适的幂系数 α , 不仅可以提高预测模型的预测精度, 而且能够减弱奇异点对建模的影响。与传统的光谱分析方法 PLS 相比, 四成分实验的预测精度提高了 10.05% , 人体无创实验的预测精度提高了 8.22% 。

采用 PLS 建模的人体血糖浓度预测精度容易受到测试者心理、外界温度、测试压力等不确定因素的影响。而采用 CPR 建模可以灵活地设定特定时段和环境下的幂系数, 从而根据最佳幂系数建立特定的最佳定量分析模型, 进而最大程度地提高预测精度。可见, CPR 建模的灵活性是 PLS 建模无法比拟的。

4 结 论

本文提出了一种应用连续幂系数回归对人体血糖浓度近红外无创检测的研究方法。通过建立不同幂系数 α 下的定量分析模型发现, CPR 可以根据不同的实验模型灵活地选取幂系数 α , 从而在最大程度上提高预测精度和减弱奇异点对建模的影响, 进而满足实际人体血糖浓度测量应用中的要求。本文研究结果表明, 应用连续幂系数回归对人体血糖浓度近红外无创检测是可行的, 可以作为人体血糖浓度测量的理论依据, 也为今后开发相应的无创血糖检测仪奠定了基础。

References

- [1] YAN Yan-lu, ZHAO Long-lian, HAN Dong-hai, et al(严衍禄, 赵龙莲, 韩东海, 等). The Foundation and Application of the Near Infrared Spectroscopy Analysis(近红外光谱分析基础与应用). Beijing: China Light Industry Press(北京: 中国轻工业出版社), 2005.
- [2] LI Gang, WANG Yan, LI Qiu-xia, et al(李刚, 王焱, 李秋霞, 等). J. Infrared Millim. Waves(红外与毫米波), 2006, 25(5): 345.
- [3] LI Qing-bo, YAN Hou-lai, LI Li-na, et al(李庆波, 阎侯赖, 李丽娜, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2010, 30(8): 2115.
- [4] LI Qingbo, LI Lina, ZHANG Guang-jun. Infrared Physics & Technology, 2010.
- [5] ZHANG Guang-jun, LI Li-na, LI Qing-bo, et al(张广军, 李丽娜, 李庆波, 等). J. Infrared Millim. Waves(红外与毫米波), 2009, 28: 107.
- [6] LUO Yun-han, XU Ke-xin(罗云瀚, 徐可欣). Acta Laser Biology Sinica(激光生物学报), 2008, 17(1): 24.
- [7] XIAO Jun, WANG Long, LUO Qing-ming, et al(肖君, 王龙, 骆清铭, 等). Journal of Optoelectronics Laser(光电子激光), 2007, 18(9): 1135.
- [8] LU Wan-zhen(陆婉珍). Modern Near Infrared Spectroscopy Analytical Technology(现代近红外光谱分析技术). Beijing: China Petrochemical Press(中国石化出版社), 2007. 44.
- [9] Wise B M, Ricker N L. J. Chemometrics, 1993, 7: 1.
- [10] Sijmen de Jong, Barry M Wise, N Lawrence Ricker. J. Chemometrics, 2001, 15: 86.
- [11] Kevin H Hazen, Mark A Arnold, Gary W Small. 1994, 48: 4.

Research on Continuum Power Regression in Noninvasive Measurement of Human Blood Glucose

LI Qing-bo, LIU Jie-qiang, ZHANG Guang-jun*

Key Laboratory of Precision Opto-mechatronics Technology, Ministry of Education, School of Instrument Science and Opto-Electronics Engineering, Beihang University, Beijing 100191, China

Abstract Accurate measurement of human blood glucose concentration is very significant for the treatment of diabetes. In the present paper, the method of continuum power regression can improve the predictive accuracy of noninvasive measurement of human blood glucose concentration with near infrared spectroscopy. This method is the expansion of the traditional method of partial least squares (PLS). It can achieve simpleness, and can significantly improve predictive accuracy when the power coefficient is fit. Using the method, quantitative analysis models of four ingredient experiment and noninvasive experiment of body were established, and these models can be used to predict the predictive samples. Experimental results show that compared with the PLS, the quantitative analysis models of this method not only can improve predictive accuracy, but also can set different power coefficient for different individuals to achieve the best results of models. According to different individuals, the power coefficient can be selected flexibly, which is of great value to the research on noninvasive measurement of human blood glucose concentration with near infrared spectroscopy.

Keywords Continuum power regression; Near infrared spectroscopy; Blood glucose concentration; Partial least squares; Noninvasive measurement

(Received Sep. 15, 2010; accepted Nov. 10, 2010)

* Corresponding author