

正交信号校正正在人体血糖近红外光谱分析中的研究

罗运平 张军^① 郑仁富^a 曹燕^a 温旺荣^a 陈星旦

(暨南大学光电工程系 广州市黄埔大道西 601 号 510632)

^a(暨南大学附属第一医院临床学检验中心 广州市 510632)

摘 要 本文以血清的近红外光谱为研究对象, 首先运用载荷向量分析确定最佳定标谱区为 4192—4943 cm^{-1} 与 5294—7200 cm^{-1} 两个波段, 然后分别用平滑、一阶导数、正交信号校正(OSC) 预处理方法, 结合偏最小二乘回归方法建立了血清中血糖的模型。用平滑、一阶导数所建立模型的预测标准偏差(RMSEP) 分别是 0.545、0.568, 用 OSC 校正后所建立模型的 RMSEP 是 0.390, 结果表明使用 OSC 校正对光谱进行合理的校正, 能够滤除光谱矩阵与浓度矩阵无关的信号, 降低模型的因子数, 从而降低模型的复杂性, 提高模型的稳健性。

关键词 正交信号校正, 偏最小二乘回归, 血糖, 近红外光谱。

中图分类号: O657.33

文献标识码: A

文章编号: 1004-8138(2007) 05-0773-05

1 引言

在血液的临床分析中, 无创血糖光谱分析的简单性, 少试剂, 逐渐体现了其在检测许多生物体组织成分的优越性^[1-3]。但是由于生物体组织的复杂性, 使其分析起来仍有些难度, 原因在于血糖浓度相对血液中的其他成分含量较低, 并且光谱很大程度上受到血液中水的吸收以及检测条件的影响^[4]。因此, 在使用多元校正方法进行建立模型时, 我们通常先使用各种预处理方法降低噪声的影响, 更好地提取光谱中的有效信息。在本文中, 作者先用载荷向量分析法寻找血糖的最佳定标模型光谱区, 然后使用平滑、一阶导数和 OSC 三种预处理方法对光谱矩阵进行光谱处理, 使用偏最小二乘回归方法^[5]对光谱矩阵和浓度矩阵进行建模, 对结果进行比较, 最后得出血糖的最佳定标模型。

2 实验部分

2.1 理论

S. wold 等人在 1998 年提出了正交信号校正(Orthogonal Signal Correction, OSC)方法^[6], 其目的是去除原始光谱矩阵中包含的与浓度矩阵不相关的变异信息, 减少模型中的主因子数, 提高模型的预测能力和稳健性。目前正交信号校正存在多种算法, 本文采用由 Fearn 提出的算法^[7], 其具体算法如下:

- (1) 将原始校正集光谱矩阵 X 和浓度矩阵 Y 进行标准化处理
- (2) 计算 $M = (I - X \cdot Y(Y'X X'Y)^{-1}Y'X)$, I 为单位矩阵
- (3) 计算 $Z = XM$

① 联系人, 电话: (020) 85220484; E-mail: ccdbyss@163.com

作者简介: 罗运平(1982—), 男, 广东省梅州市人, 硕士研究生, 主要从事近红外光谱技术方面的研究。

收稿日期: 2007-06-04; 接受日期: 2007-06-19

- (4) 对 ZZ' 进行奇异值分解, 提取前 p 个需要正交处理的特征值 g_i 及对应的特征向量 C_i
- (5) 计算权重向量 $w_i = MX' C_i g_i$
- (6) 计算得分向量: $t_i = C_i g_i$
- (7) 计算载荷向量: $p_i = X' t_i (t_i' t_i)$
- (8) 在 X 中将正交信号减去: $X_{osc} = X - \sum t_i p_i'$
- (9) 对预测向量 X_{new} , 由权重 w 和载荷 p 求出校正后的光谱: $t = X_{new}' w, X_{new} = X_{osc} - t' p$

2.2 仪器与软件

本实验仪器 5700 傅里叶变换型近红外光谱仪(美国 Necolet 公司), 探测器为铟镓砷(InGaAs)。偏最小二乘回归方法(PLS), OSC, Savitzky-Golay 窗口移动多项式最小二乘求导^[8]用 MATLAB7.0 编写。

2.3 实验样品

87 份血清样品由华侨医院提供, 血糖含量由全自动生化分析仪测定, 样品含量范围 3.5—7.04mmol/L。血清样品随机分成两部分, 57 个样品用于建立模型, 30 个样品用于模型预测, 测量光谱时, 将血清样品置于 2mm 厚石英比色皿中, 利用 OMNIC 软件采集光谱数据, 每个样品扫描 64 次然后取平均, 扫描谱区为 4000—10000cm⁻¹, 考虑到水在 5200cm⁻¹左右有强烈的吸收峰^[9], 本文并没有给出 4943—5294cm⁻¹谱区, 如图 1。

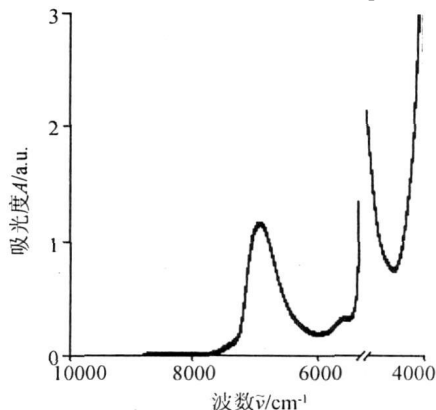


图 1 人血清原始近红外光谱图

3 建模与讨论

建立模型时, 选择适合的谱区能提高模型的预测结果, 增强模型的预测能力, 因此有必要对定标谱区进行选择, 一种有效的方法是载荷向量分析法^[10, 11]。首先对 57 个样品的光谱矩阵标准化处理, 用偏最小二乘回归方法建立模型, 利用 leave-one-out 内部交叉验证法确定模型的最佳因子数, 然后对模型的各个因子的载荷向量进行分析, 为了避免水分子强烈吸收的干扰, 选取 4192—4943cm⁻¹和 5294—10000cm⁻¹谱区。

对各个载荷向量进行分析得出, 第一、二因子的载荷向量对模型的贡献比较大, 如图 2(a) 和图 2(b) 所示, 而第三个因子的载荷向量明显的特点是 7200—10000cm⁻¹的载荷系数都接近于零, 如图 2(c) 所示。对其他因子进行分析得出同样的结果, 这说明这些谱区对回归结果并没有明显影响。因此, 建模时最终选取的谱区为 4192—4943cm⁻¹与 5294—7200cm⁻¹。

近红外光谱仪所采集的光谱除样品自身信息外, 还包含了其他无关信息和噪音。因此, 在用多元校正方法建立模型前, 应该采取有效的光谱预处理。本文分别采用了平滑、一阶导数和 OSC 校正三种预处理方法。图 3 为血清样品导数光谱, 这里采用 Savitzky-Golay 窗口移动多项式最小二乘 23 点 2 次平滑求导, 图 4 为光谱矩阵经过 OSC 校正后的光谱图, 此光谱和原始光谱相比, 两者的光谱图形状基本相同, 但经过 OSC 后的光谱图更加平整。

本文使用偏最小二乘回归方法对光谱矩阵和浓度矩阵进行回归分析, 采用内部交叉验证法(cross-validation) 确定最佳因子数, 结果见表 1。从表中可以看出, 经过预处理后相关系数都有所提高, 用正交信号校正后的模型的分析误差小于用原始光谱, 平滑、一阶导数光谱建立的模型。值得注意的是, 用平滑和一阶导数光谱建立的模型与用原始光谱建立的模型相比, 预测标准差(RMSEP)要大, 其 RMSEP 分别是 0.545 和 0.568, 这是因为对血糖光谱进行求导或平滑时, 在消

除样品光谱干扰信号的同时,也不可避免地去除了与浓度矩阵相关的信号,从而降低了信噪比。而对光谱矩阵使用 OSC 校正后其 RMSEP 降低到 0.390,这说明 OSC 能够滤除光谱矩阵与浓度矩阵无关的光谱信号,保持最大的有效光谱信息,并且其因子数也降低到 5,从而简化了光谱模型的复杂性。

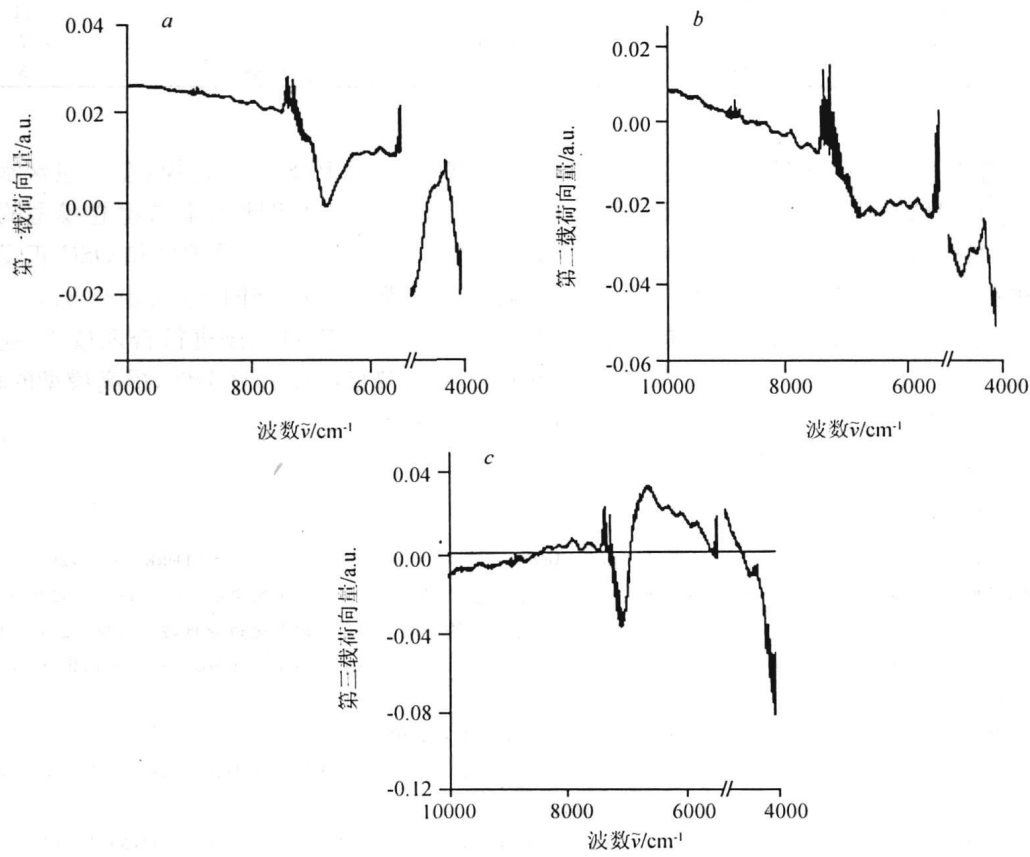


图 2 基于 PLS 模型的前三个载荷向量
a——第一载荷向量; b——第二载荷向量; c——第三载荷向量。

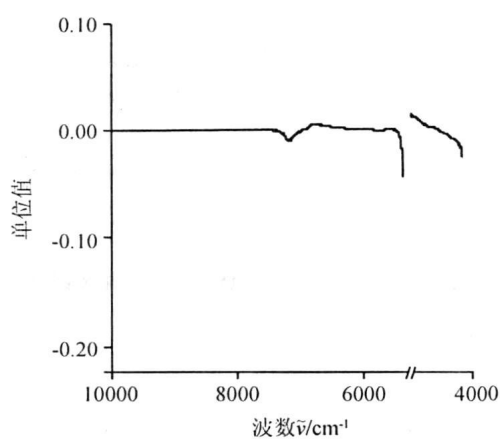


图 3 人血清一阶导数光谱

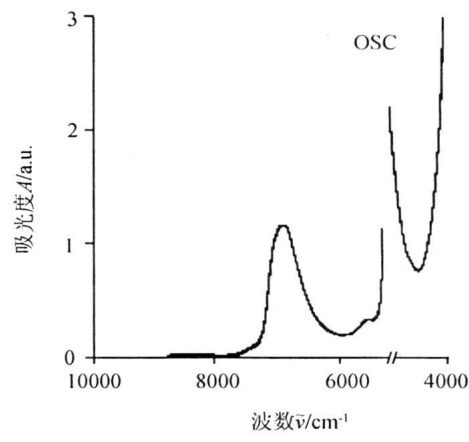


图 4 人血清经过 OSC 校正后光谱

表 1 最佳因子数

光谱区(cm ⁻¹)	预处理	R	预测标准差	因子数
4192—4943	——	0. 935	0. 489	8
5294—10000				
	——	0. 943	0. 471	8
4192—4943	平滑	0. 901	0. 545	11
5294—7200	一阶导数	0. 939	0. 568	7
	OSC	0. 930	0. 390	5

4 结论

本文首先利用载荷向量分析法选择了最佳的光谱区,运用 PLS 进行建模,在不包含水强烈吸收峰 5200cm⁻¹情况下,使用载荷向量分析得出的谱区建立模型结果比用全谱建立模型的结果更好。然后对选择的光谱用平滑, Savetaky-Golay 窗口移动多项式最小二乘求导和 OSC 正交信号预处理方法进行预处理,并作出比较,用平滑、一阶导数所建立模型的 RMSEP 分别是 0. 545、0. 568,用 OSC 校正后所建立模型的 RMSEP 是 0. 390,表明运用 OSC 校正对光谱进行合理校正,能够滤除光谱矩阵与浓度矩阵无关的信号,降低模型的因子数,从而降低模型的复杂性,提高模型的稳健性,这对人体血糖浓度在无创伤检测中的进一步研究具有指导意义。

参考文献

[1] Burns D A, Ciurczak E W. *Handbook of Near-Infrared Analysis*[M]. 2nd ed. , New york: Marcel Dekker Inc. , 2001. 633—647.

[2] 陈华才, 杨仲国, 李惠英, 陈星旦. 人血清中血糖的近红外光谱快速检测[J]. 中国计量学院学报, 2004, **15**(3): 0235—0237.

[3] 吕丽娜, 刘蓉. O-PLS 在近红外无创人体血糖浓度测量基础研究中的应用[J]. 光谱学与光谱分析, 2005, **25**(12): 1950—1954.

[4] Arnold M A, Small G W. Determination of Physiological Levels of Glucose in an Aqueous Matrix with Digitally Filtered Fourier Transform Near-Infrared Spectra[J]. *Analytical Chemistry*, 1990, **62**(14): 1457—1464.

[5] 王惠文. 偏最小二乘加回归方法及其应用[M]. 北京: 国防工业出版社, 1999. 201—234.

[6] Svante Wold· Henrik Antti, Fredrik Lindgren *et al.* Orthogonal Signal Correction of Near-Infrared Spectra[J]. *Chemometrics and Intelligent Laboratory Systems*, 1998, **44**(1— 2): 175—185.

[7] Tom Fearn. On Orthogonal Signal Correction[J]. *Chemometrics and Intelligent Laboratory Systems*, 2000, **50**(1): 47—52.

[8] Savitzky A·Golay M J E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures[J]. *Anal·Chem·* , 1964, **36**(8): 1627—1639.

[9] Hale G M, Querry M R. Optical Constants of Water in the 200nm to 200μm Wavelength Region[J]. *Applied Optics*, 1973, **12**(3): 555—563.

[10] Kim Y J, Yoon G. Multicomponent Assay for Human Serum using Mid-Infrared Transmission Spectroscopy Based on Component-Optimized Spectral Region Selected by A First Loading Vector Analysis in Partial Least-Squares Regression[J]. *Applied Spectroscopy*, 2002, **56**(5): 625—632.

[11] Kim Y J, Yoon G. Prediction of Glucose in Whole Blood by Near-Infrared Spectroscopy: Influence of Wavelength Region, Preprocessing, and Hemoglobin Concentration[J]. *Journal of Biomedical Optics*, 2006, **11**(4): 041128.

Study of Orthogonal Signal Correction Analyze in Human Serum Glucose with Near Infrared Spectroscopy

Abstract Two optimal spectral region $4192\text{--}4943\text{cm}^{-1}$ and $5294\text{--}7200\text{cm}^{-1}$ are selected based on human serum near-infrared spectroscopy for human glucose scale by loading vector analysis method. The spectra are preprocessed by smooth, first derivative and orthogonal signal correction (OSC) data preprocessing methods. The calibration models of serum glucose are established with partial least squares regression (PLSR). The root mean squared errors of prediction (RMSEP) are 0.545 and 0.568 for spectra pretreated by smooth and first derivative preprocessing methods. But the RMSEP is 0.390 for OSC spectra preprocessing methods. The OSC preprocessing method can correct the spectral matrix, and remove the information that is orthogonal to the concentration matrix and can reduce the factors number of calibration. Thereby the complexity is diminished and the stability is improved to the serum glucose model by OSC method.

Key words Orthogonal Signal Correction, PLS, Glucose, Near-Infrared Spectroscopy.

封四: “保质、高效 —— 《光谱实验室》主要特色”的附件 1

主编不编与主编不主

闲下翻阅地方杂志,有标明主编也有不标的。这倒各随其规而悉听其便。但也有令人疑惑的事,比如记忆中某人在某部门任行政要职,突然间成为一个地方杂志的主编(并非顾问或名誉主编)。是同名还是改了行?无意中渐渐知道,有一些确系既未改行也未重名的,是在“遥控”机制中兼了职。

兼职这事不好妄论,但主编要编,却可以论定。因为,抛开真正的编辑者或者为了某人之名而拉大旗,或者为了某部门捐赠拨款之利而钓大鱼,暂且不去论它。那些遥兼主编的同志真的能够切切实实地履行主编的职责吗?如果不能切切实实地履行,还是以不挂虚名为好,免得闹出盗名欺世的笑话。

说白一些,“空头主编”并没有看到主编也是一种重要的专职业务岗位,来不得任何一点“名存实亡,失其所业”。这正像企业家不敢贸然兼之,科学家和学者不敢贸然兼之一样,编辑尤其是总其成的主编,同样不好贸然兼之,因为,这是有责、权、利的问题,有术业专长的的问题,也还有“法人”而不是声名徒自远扬的问题。

由此还想到“期刊衙门”与“编辑官”。由于体制方面的原因,编辑部门机关化的倾向颇严重,而编辑头们把自己首先当成“官员”而后才是编辑的意识,也很根深蒂固。人们觉得“处级和尚”可笑,局级企业也不妥,殊不知局级处级报刊杂志大约也不那么顺理成章啊!

大概是那个所谓的“官本位”或曰“行政级本位”作祟,刊物升级之风曾经有些洋洋乎盈耳。也是这样一个原因,不仅出现生拉硬扯“空头主编”的事,也还出现了“生编不主”的现象:当主编而不主编务,干吗非要挂这个衔呢?

主编,“是名也,止于是实也”。随着行政机关同企事业单位的逐渐分开,编辑终究会成为编辑。此前,主编不编与主编不主,首先应当纳入革除之列。一些人一定还要去当“空头主编”而不干实事,不妨赠以孟子的一句话:“先生之志则大矣,先生之号则不可”。

其实,主编挂名,这种杂志原本也不该核准的。从法律角度上思量,不是这样么?

(原载 1988 年 1 月 6 日《新闻出版报》,作者:冯并)