

近红外无创生化检测中不同光程的光谱校正模型研究

高洪智, 卢启鹏*, 丁海泉

中国科学院长春光学精密机械与物理研究所应用光学国家重点实验室, 吉林 长春 130033

摘要 近红外无创生化检测中采用不同血流容积光谱相减的方法理论上能够消除人体组织产生的强背景干扰, 但人体光谱相减后得到的血液光谱的光程是未知的, 这给校正模型的建立带来了困难。通过设计模拟实验并采用正交信号校正的方法对光程信息进行校正, 提高不同光程的光谱建立校正模型的预测精度。分别建立了光程校正前后的模型, 模型的交叉验证标准偏差从 $90.17 \text{ mg} \cdot \text{dL}^{-1}$ 下降到 $31.62 \text{ mg} \cdot \text{dL}^{-1}$, 相关系数从 0.978 7 提高到 0.996 8。实验结果表明, 采用正交信号校正的方法能够有效抑制光程信息的干扰, 提高模型的预测精度, 这为不同血流容积光谱相减法的实际应用提供了基础。

关键词 近红外光谱; 无创生化检测; 光程差异; 正交信号校正

中图分类号: O433.1; O657.3 文献标识码: A DOI: 10.3964/j.issn.1000-0593(2012)05-1423-04

引言

血液的生化检测是健康诊断的重要手段。利用近红外光谱分析技术实现无创生化检测具有无痛苦、无感染、无需试剂、便于实现生化指标的实时监测等优点^[1]。采用不同血流容积光谱相减的方法理论上能够消除人体组织产生的强背景干扰, 得到纯净的血液光谱, 解决目前近红外无创生化检测所面临的个体差异、背景干扰等问题。

不同血流容积光谱相减法的目的是要对通过光谱相减得到的不含人体背景干扰的纯净血液光谱与其各生化成分的程度信息建立多元校正模型, 对未知血液生化指标的光谱进行预测, 最终实现无创生化检测。然而由于自然或人为作用, 在血流容积发生变化时测得的人体光谱中所反映的血液光程是变化的, 两幅人体光谱相减后得到的血液光谱的光程将是未知的, 如果将各种未知光程的血液光谱直接建立多元校正模型, 将会使模型的性能降低, 这给血流容积光谱相减法带来了困难。

本工作通过不同光程不同葡萄糖浓度的脂肪乳溶液实验, 对不同血流容积光谱相减法中不同光程的近红外光谱与其浓度信息的多元校正模型建立方法进行了研究, 采用正交信号校正的方法以抑制光程差异, 从而提高模型的性能, 解决不同血流容积光谱相减法在建立校正模型中遇到的问题。

1 实验部分

1.1 实验仪器

光谱采集采用 Nicolet 6700 傅里叶变换红外光谱仪。光源为卤钨灯, CaF_2 为分束器, 探测器为 TE 制冷的 InGaAs 探测器, 利用光谱仪配套的 OMNIC 软件采集样品近红外光谱。实验中 Nicolet 6700 傅里叶变换红外光谱仪的光谱采集参数设置如下: 分辨率为 8 cm^{-1} ; 扫描次数为 32 次; 光谱采集范围为 $1000 \sim 2500 \text{ nm}$; 每隔 30 min 测量一次背景(本实验中以空气为背景)以防止仪器带来的基线漂移。

采用可调光程样品池(如图 1)来改变含有葡萄糖的脂肪乳溶液的光程。样品池的最大量程为 5 mm, 分辨率为 0.01 mm。



Fig 1 Variable light path length cell

收稿日期: 2011-11-23, 修订日期: 2012-02-16

基金项目: 国家自然科学基金项目(60878052, 60938002, 61078038), 中国科学院知识创新工程领域前沿项目(Y00232Q100), 应用光学国家重点实验室开放课题(Y1Q03FQ113)和吉林省科技发展计划项目(20100338)资助

作者简介: 高洪智, 1984 年生, 中国科学院长春光学精密机械与物理研究所助理研究员 e-mail: gaohz@ciomp.ac.cn

* 通讯联系人 e-mail: luqipeng@126.com

1.2 溶液配制与光谱采集

脂肪乳溶液在组织光学中的应用比较广泛, 由于其光学参数(如散射特性、吸收特性及各向异性因子)与生物组织的光学参数较为接近, 被许多学者作为生物组织的模拟溶液^[2]。本实验选用浓度为 20% 的 250 mL 脂肪乳注射液, 其中大豆油含量为 50 g, 注射用卵磷脂含量为 3 g, 注射用甘油 5.5 g。

用电子天平称取不同重量的分析纯葡萄糖(纯度为 99.99%) 配制成各种不同浓度的葡萄糖溶液, 在每种葡萄糖溶液中加入一定量的 20% 脂肪乳, 最终配制成的溶液中葡萄糖的浓度范围为 100~1 500 mg·dL⁻¹, 脂肪乳的浓度为 2%, 共配制了 35 种不同葡萄糖浓度的脂肪乳溶液。可调光程样品池的光程变化范围为 0.05~0.6 mm, 间隔为 0.05 mm。图 2 为各种浓度各种光程的葡萄糖脂肪乳溶液的近红外光谱。

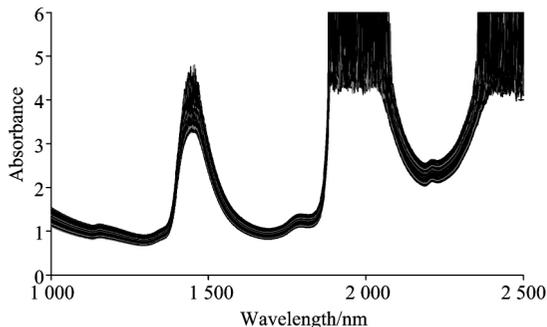


Fig 2 NIR spectra of Intralipid mixed with different glucose concentration at different pathlength

图 2 中可以看到, 水在 1 450 和 1 940 nm 的两个近红外特征吸收峰将整个光谱区域分成三部分: 合频区(2 200~2 400 nm)、第一倍频区(1 600~1 800 nm)和短波区(1 000~1 400 nm)。为了避免葡萄糖脂肪乳溶液中水产生的干扰, 有必要选取葡萄糖的特征波段, 在特征波段内建立多元校正模型将会得到更好的模型预测效果。第一倍频波段的吸光度值在 1 附近, 具有较好的信噪比, 且葡萄糖在第一倍频具有较多的特征吸收^[3]。因此, 本工作不同光程光谱的模型都是在第一倍频区域建立的。

1.3 正交信号校正算法

在近红外光谱分析中, 光谱矩阵 R 不可避免地存在着一些系统误差和随机误差^[4], 即与浓度矩阵 Y 不相关的部分 R_1 。因此希望能够将光谱矩阵 R 中与浓度矩阵 Y 正交不相关的部分 R_1 去掉, 即 $R_{new} = R - R_1$, 然后利用这个新的光谱矩阵 R_{new} 来建立多元校正模型。

正交信号校正(OSC)方法能够去除光谱矩阵 R 中与分析成分矩阵 Y 相正交的部分, 并认为正交的这部分信息是不包含待测成分信息的无关量^[4]。本工作中不同血流容积光谱相减后得到的血液光谱中光程差异信息是对建立校正模型最大的干扰, 同时由于光谱相减的两幅光谱选取的随机性, 所得到的血液光谱的光程干扰信息与血液的生化成分含量是不相关的。采用正交信号校正有望降低光程差异对校正模型的影

响, 提高模型的精度。

正交信号校正的算法原理如下^[5,6]:

- (1) 将原始光谱矩阵 R 和浓度矩阵 Y 进行标准化处理;
- (2) 采用奇异值分解(SVD)方法计算光谱矩阵 R 的主成分 t ;
- (3) 将 t 与浓度矩阵 Y 作正交处理, 得到 $t_{new} = (I - Y(Y'Y)^{-1}Y')t$, 并计算特征向量 $\omega = X - t_{new}$;
- (4) 根据新计算得到的 X, ω 计算正交主成分 $t, t = X\omega$;
- (5) 检查收敛性。如果 $\|t - t_{old}\| / \|t\| < 10^{-6}$, 则认为已经收敛, 转入下一步骤, 否则返回步骤(3)继续处理;
- (6) 计算载荷向量 $p' = t'X / (t't)$;
- (7) 去除光谱矩阵 R 与浓度矩阵 Y 正交的信号, 得到 $R = R - tp'$;
- (8) 回到步骤(2), 计算新的 R 与 Y 正交的主成分, 直到完成所需要的次数;
- (9) 计算得到去除与 Y 正交部分的新光谱矩阵 R_{osc} , $R_{osc} = R - TP'$, 将此矩阵作为后续建模分析的光谱矩阵。

2 结果分析

2.1 直接偏最小二乘法建立的校正模型

偏最小二乘法(PLS)是近红外光谱分析中应用最广的经典建模方法^[7], 它能够建立具有一定精度和稳定性的校正模型。首先, 对 0.2 mm 光程的不同浓度的葡萄糖脂肪乳溶液在 1 600~1 800 nm 区域采用 PLS 建立葡萄糖的多元校正模型, 以便与不同光程的葡萄糖多元校正模型进行对比。所建 PLS 模型的预测效果如图 3 所示, 最佳主成分数为 4, 模型的相关系数为 0.990 5, 校正均方根误差 RMSEC 为 51.13 mg·dL⁻¹, 交叉验证均方根误差 RMSECV 为 82.73 mg·dL⁻¹。

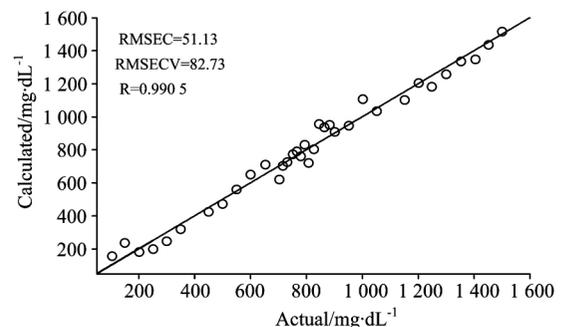


Fig 3 The PLS model of Intralipid mixed with glucose at 0.2 mm pathlength

从采集到的各种浓度葡萄糖脂肪乳溶液光谱中随机选取某一光程的光谱, 共得到 35 种不同光程不同浓度分布的葡萄糖脂肪乳溶液光谱。同样采用偏最小二乘法对不同光程的葡萄糖模拟溶液建立多元校正模型, 模型的预测效果如图 4 所示, 最佳主成分数为 4, 相关系数 R 为 0.978 7, 校正均方根误差 RMSEC 为 76.31 mg·dL⁻¹, 交叉验证均方根误差 RMSECV 为 90.16 mg·dL⁻¹。很明显, 不同光程的葡萄糖

脂肪乳溶液光谱所建立的多元校正模型性能要比同光程时所建的模型差。

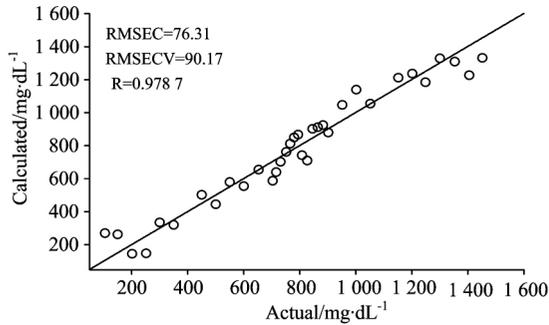


Fig 4 The PLS model of Intralipid mixed with glucose at different pathlength

对不同光程的葡萄糖脂肪乳溶液进行主成分分解, 得到由第一主成分(PC1)和第二主成分(PC2)所组成的主成分空间(如图5), 图中同一光程的样品采用相同的符号表示, 可以看到, 随着光程的增加, PC1 逐渐增大。由于建立校正模型的过程中没有消除光程的影响, 在主成分空间中不同光程的样品的第一主成分即光程的干扰是导致模型性能下降的重要原因。

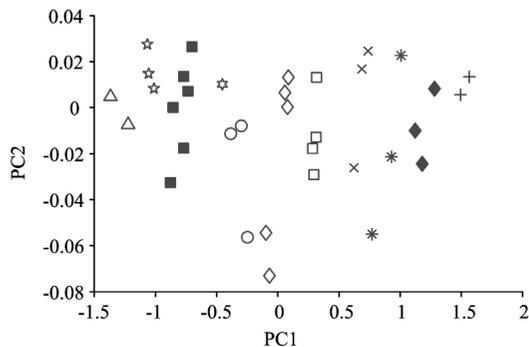


Fig 5 Principal component space of Intralipid mixed with glucose at different pathlength

2.1 基于 OSC-PLS 方法的校正模型

不同光程的葡萄糖脂肪乳溶液中含有与葡萄糖浓度信息无关的光程信息的干扰, 所建校正模型的性能与同光程的模型相比较差。由于光程信息是与葡萄糖浓度信息是无关的, 从 OSC 的算法原理上可以看到, 所有与浓度信息无关的变量都能通过校正得到消除, 那么 OSC 校正后的葡萄糖脂肪乳溶液光谱中将不再包含光程的信息。

采用介绍的 OSC 算法对不同光程的葡萄糖脂肪乳溶液光谱进行校正, 此时可以看到校正后的主成分空间中第一主成分不再随着样品光程的改变而改变, 也就是光程信息得到了消除(如图6)。

对 OSC 校正后的光谱采用 PLS 建立校正模型, 如图7所示, 模型的主成分数为 1, 模型相关系数 r 为 0.9968, 校

正均方根误差 RMSEC 为 $29.57 \text{ mg} \cdot \text{dL}^{-1}$, 交叉验证均方根误差 RMSECV 为 $31.62 \text{ mg} \cdot \text{dL}^{-1}$ 。与直接建立的不同光程 PLS 校正模型相比, r 和 RMSEC 都得到了提高, 特别是 RMSECV 得到了很大的提升, 模型的稳定性得到了增强。可以看出, 经过 OSC 校正后的 PLS 模型性能优于直接 PLS 所建立的模型。

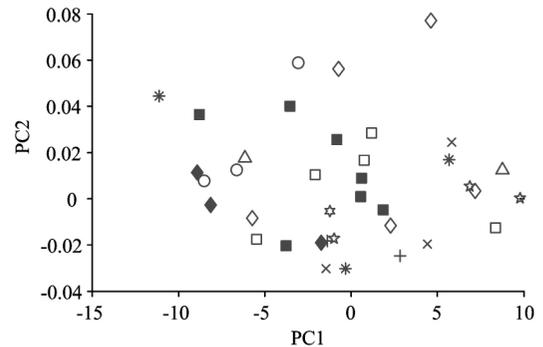


Fig 6 The OSC corrected principal component space

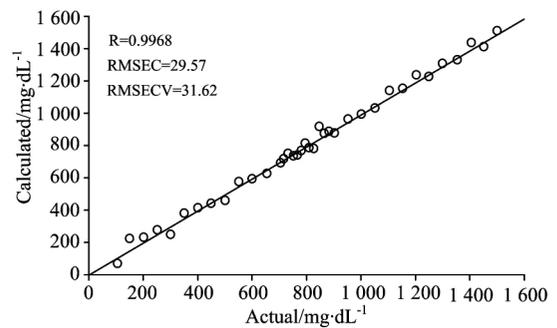


Fig 7 The OSC-PLS model of Intralipid mixed with glucose at different pathlength

3 结论

不同血流容积光谱相减法为血液无创生化检测提供了一种新思路, 但在实际应用中还有一些关键问题需要解决。通常, 使用偏最小二乘法(PLS)对样品的光谱和化学值建立多元校正模型, 在用于建模的各个样品处于同一光程时, 采用这一方法能够得到精度较高的校正模型。但使用不同血流容积光谱相减法得到的血液光谱的光程是各异的, 这给校正模型的建立带来了困难。本工作以不同光程不同葡萄糖浓度的脂肪乳溶液模拟人体组织, 采用 OSC 校正结合 PLS 建模的方法建立的不同光程葡萄糖脂肪乳溶液光谱的校正模型, 与直接 PLS 建立的模型相比, RMSEC 从 76.31 降低到 $29.57 \text{ mg} \cdot \text{dL}^{-1}$, RMSECV 从 $90.17 \text{ mg} \cdot \text{dL}^{-1}$ 下降到 $31.62 \text{ mg} \cdot \text{dL}^{-1}$, R 从 0.9787 提高到 0.9968 。

研究结果表明, 采用正交信号校正的方法能够有效抑制光程信息的干扰, 提高模型的预测精度, 为不同血流容积光谱相减法的实际应用提供基础。

References

- [1] CHEN Xing-dan(陈星旦). Optics and Precision Engineering(光学精密工程), 2008, 16(5): 759.
- [2] ZHANG Lin, ZHANG Chun-ping(张琳, 张春平). Chinese Journal of Luminescence(发光学报), 2009, 30(6): 858.
- [3] Amerov A K, Chen J, Arnold M A. Applied Spectroscopy, 2004, 58(10): 1195.
- [4] REN Qian, XIE Guo-ling, DONG Shou-long, et al(任芊, 解国玲, 董守龙, 等). Transactions of Beijing Institute of Technology(北京理工大学学报), 2005, 25(3): 272.
- [5] Wold S, Antti H, Lindgren F, et al. Chemometrics and Intelligent Laboratory Systems, 1998, 44: 175.
- [6] Westerhuis J A, Jong S, Smilde A K. Chemometrics and Intelligent Laboratory Systems, 2001, 56: 13.
- [7] 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(北京: 中国石化出版社), 2006.

Study of the Calibration Model of Different Pathlength Spectra in Near Infrared Noninvasive Biochemical Sensing

GAO Hong-zhi, LU Qi-peng*, DING Hai-quan

Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China

Abstract The spectral subtraction approach with different flow blood volume in near infrared noninvasive biochemical sensing can eliminate human tissue background interference, but at the same time the pathlength of subtracted spectrum is unknown. Thus, the model will be insufficiently effective if established directly with pathlength-unknown spectra, which is a big problem for subtracted blood volume spectrometry. In the present paper, a simulated experiment was designed to simulate this issue. The orthogonal signal correction method was proposed to eliminate the influence brought by pathlength. Compared with the PLS model before and after orthogonal signal correction, the root mean square error of cross validation (RMSECV) was from 90.17 $\text{mg} \cdot \text{dL}^{-1}$ down to 31.62 $\text{mg} \cdot \text{dL}^{-1}$, and the correlation coefficient was improved from 0.9787 to 0.9968. The experimental results show that using the orthogonal signal correction method can effectively restrain the interference information of pathlength, and improve the prediction precision of calibration model. The conclusion in this paper sets the stage for the practical application of the spectral subtraction approach with different flow blood volume.

Keywords Near-infrared spectroscopy; Noninvasive biochemical sensing; Different pathlength; Orthogonal signal correction

(Received Nov. 23, 2011; accepted Feb. 16, 2012)

* Corresponding author