

文章编号: 1005-5630(2013)03-0020-05

血糖浓度的短波近红外定量分析*

王 伟^{1,2}, 李振庆^{1,2}, 倪 一^{1,2}, 刘晨晨^{1,2}, 赵志强³

(1. 上海理工大学 光电信息与计算机工程学院, 上海 200093;

2. 上海理工大学 上海市现代光学系统重点实验室, 上海 200093;

3. 河南省电子产品质量监督检验所, 河南 郑州 450003)

摘要: 采用间隔偏最小二乘法(IPLS)和移动窗口偏最小二乘方法(MWPLS), 在 640~1 100 nm 范围内建立血糖短波近红外的优化模型。使用马氏距离对人血清样品中的奇异样品进行筛选, 将检测光谱分别等分为 2~15 份进行 IPLS 分析, 对比建立预测模型。设窗宽为 151 nm, 成分数范围(1~20), 全谱进行 MWPLS, 对预测模型进行优化。结果显示, 依据马氏距离采用最小半球体积法能有效筛选所采集光谱中的奇异光谱, IPLS 可以有效地找到葡萄糖分子官能团对应的近红外特征谱段, MWPLS 能够找到适合建模的精确起止波长点, 通过偏最小二乘法建立血糖浓度的预测模型, 相关系数 $R=0.9822$, 预测均方差 $RMSEP=0.1635$ mmol/L, 偏差 $Bias=-0.0873$ mmol/L。

关键词: 短波近红外光谱; 血糖; 间隔偏最小二乘法; 移动窗口偏最小二乘法

中图分类号: O 657.33 **文献标识码:** A **doi:** 10.3969/j.issn.1005-5630.2013.03.005

The quantitative analysis of glucose in human serum by short-wave NIR spectra

WANG Wei^{1,2}, LI Zhenqing^{1,2}, NI Yi^{1,2}, LIU Chenchen^{1,2}, ZHAO Zhiqiang³

(1. School of Optical-Electrical and Computer Engineering, University of Shanghai for
Science and Technology, Shanghai 200093, China;

2. Shanghai Key Laboratory of Modern Optical System, University of Shanghai for
Science and Technology, Shanghai 200093, China;

3. Henan Electronic Products Quality Supervision and Inspection Institute, Zhengzhou 450003, China)

Abstract: The optimal model for the short wave near-infrared (640~1 100 nm) spectroscopic analysis of serum glucose was established by interval partial least square (IPLS) and moving window partial least square (MWPLS) methods. The outlier samples were eliminated based on mahalanobis distance, and then the spectra was equally divided into 2~15 parts for IPLS model. The spectra window for MWPLS was 151 nm, and the principal component was between 1 and 20. Results demonstrated that mahalanobis distance method can effectively eliminate the outlier samples. The optimal spectra domain by IPLS corresponded well with the functional groups of the glucose molecules. By MWPLS, the relative coefficient R , root meant square error for prediction $RMSEP$, and $Bias$ error reached 0.9822, 0.1635 mmol/L, and -0.0873

* 收稿日期: 2012-12-10

基金项目: 国家自然科学基金资助项目(21205078); 上海市科技发展基金资助项目(10540500700); 上海市重点学科项目第三期资助项目(S30502)

作者简介: 王 伟(1987-), 男, 上海市人, 硕士研究生, 主要从事光谱检测技术方面的研究。

mmol/L, respectively.

Key words: short-wave near-infrared spectrum; blood glucose; interval partial least square method; moving window partial least square method

引言

糖尿病是一种常见的慢性终身性疾病。近年来,随着人民生活水平的提高,糖尿病发病率不断攀升。最新调查数据显示^[1],1994年我国糖尿病患病率仅为2.5%,到2008年,糖尿病患病率已高达9.7%,其中前期患病率为15.5%。血糖是诊断糖尿病的关键指标。目前,小型便携式的电子血糖仪得到了广泛的应用。

现代光谱分析技术是实现无损血糖检测的一个重要的发展方向。以往的血糖检测研究多集中于长波近红外(1 100~2 500 nm)^[2-7],其主要原因是这些区域,含有葡萄糖的甲基 C-H 对称伸缩振动和 O-H 伸缩振动的一级、二级倍频,具有较强的葡萄糖特征吸收信号。然而这些区域内水分子也会产生较强的吸收信号,在此干扰条件下使得在长波近红外内实现血糖的无损检测几乎是不可行的。本文则关注短波近红外(700~1 100 nm)的检测分析。在短波近红外处,人体骨骼、肌肉、脂肪、皮肤等的吸收系数很小,该波段被誉为人体“光学治疗窗”^[8],处于该窗口的光学波段可以更有效地携带人体内部的信息。而且短波近红外包含了葡萄糖的甲基 C-H 对称伸缩振动的三倍频和 O-H 伸缩振动的三倍频^[9],具有检测的可能性。长波近红外区域制作成快速并行检测所需的阵列检测器,价格较为昂贵,且检测时间较长。相比较,短波近红外光谱区域可以使用成本较低的 Si 基光敏元件,硅阵列检测器 CCD 的技术也已成熟,可以大幅度提高检测速度,有利于实现检测设备的小型化和低成本。

本文对短波近红外波段用于血液样本中葡萄糖浓度的定量检测的可行性进行探索。与长波近红外相比,短波近红外是更高倍频(或合频)的振动吸收,谱峰更宽,光谱重叠的情况会更严重。为了充分提取光谱信息,需要研究有效可靠的定量校正方法。本文在 640~1 100 nm 范围内测量血清短波近红外光谱,采用间隔偏最小二乘法(interval partial least squares, IPLS)和移动窗口偏最小二乘法(moving window partial least squares, MWPLS)分析了吸光度与血清中葡萄糖浓度之间的量化关系。

1 原理及方法

偏最小二乘法(partial least squares, PLS)是一种较为常用的近红外光谱数据处理方法^[10],主要用于线型模型的建立。为了更有效利用光谱信息,选择血糖光谱信息最为丰富的波段进行建模,文章采用 IPLS 和 MWPLS 技术,其主要原理和步骤如下:

IPLS:(1)将整个光谱分割成若干个波段,对每个波段进行偏最小二乘分析,根据交叉验证均方差(root mean square error of cross validation, RMSECV),从中找出最佳主成份数;(2)改变光谱分割的波段数目,按照(1)中的过程分析;(3)对比(2)中不同波段数时所得到的结果,最终找出最适宜建模的光谱区间。

MWPLS:选取一个宽度为 w 的光谱窗口,从整个光谱(假设有 n 个波长点)的第一个波长点开始依次向右移动一个波长点直至最后,设置一最大主成分数,对每个子波长区分别建立偏最小二乘法 PLS 模型,分别得到不同主成分数里对应 PLS 模型的预测误差(root mean square error of prediction, RMSEP),从而找出含有有用信息的一个或几个波长区。然后利用这些波长区建立 PLS 模型。

文章采用 IPLS 和 MWPLS 对血清样品的短波近红外光谱进行数据分析和优化时,模型最终的预测能力通过相关系数 R ,预测均方差 $RMSEP$,偏差 $Bias$ 的值来检验,其中 $RMSEP$ 和 $Bias$ 的定义如下:

$$RMSEP = \sqrt{\frac{1}{I_p} \sum_{i=1}^{I_p} (\hat{y}_i - y_i)^2} \quad (1)$$

$$Bias = \frac{1}{I_p} \sum_{i=1}^{I_p} (\hat{y}_i - y_i) \tag{2}$$

其中 \hat{y}_i 表示样本的预测值, y_i 表示样本的实测值, I_p 表示检验的样本数。

2 数据测量及模型建立

2.1 数据测量

对 13 名志愿者在午饭后 3h 各抽取静脉血约 5 mL,为了扩大样本血糖范围,每名志愿者第一次抽血后口服葡萄糖 50 g,30 min 后再次抽血。每个样品取少量全血用血糖仪测量其血糖值作为参考值,血糖浓度范围 4.9~8.3 mmol/L,剩余血样经分离得到血清约 2.5 mL。使用 UV1900 紫外可见分光光度计室温下测定样品的吸收光谱,带宽 1 nm,测量波长范围 640~1 100 nm,采用 10 mm 光程的石英比色皿,以空气作参比,测量其吸光度。样品中,血清的提取及血清中葡萄糖浓度参考值由上海交通大学医学院采用血糖仪测量后提供。

2.2 数据预处理

为了避免除血糖浓度以外的因素对血清样本光谱的干扰,根据马氏距离采用最小半球体积法 (smallest half-volume, SHV)^[11]对样本中的奇异样品进行筛选,结果如图 1 所示。由图 1 可以发现标记 1 的样品其马氏距离明显高于其它样品,经检查发现是由于该血清较其它样品混浊,呈乳浊液状,而乳浊液在光谱检测时,由于溶液中颗粒的存在容易导致光线通过样品池时产生散射,因此建立模型时将样品 1 剔除较好。将奇异光谱剔除后,为了去除高频噪声,基线漂移等影响,对剩余的样品在 640~1 100 nm 波段内作基线校正,然后采用窗口宽度为 10 nm,多项式次数为 2 的 Savitzky-Golay 平滑法对光谱作预处理^[12-13],处理后的血清光谱如图 2 所示。

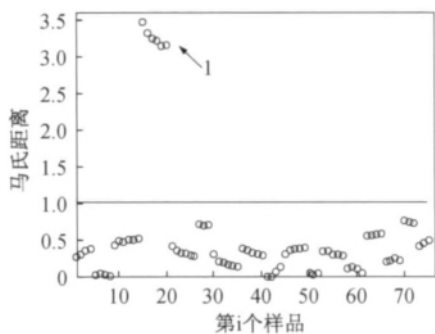


图 1 奇异样品检测

Fig. 1 Detection of the outlier sample by SHV method

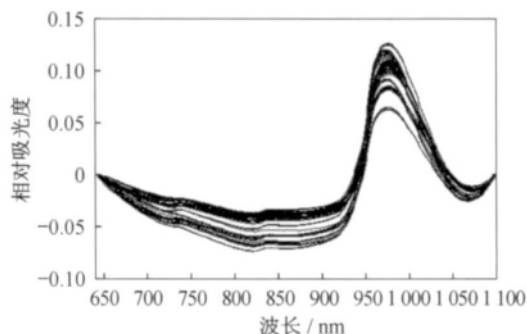


图 2 预处理后的血清近红外光谱

Fig. 2 NIR spectra of serum after pretreatment

在图 2 中对应于 980 nm 处的吸收峰主要来源于水分子中 O-H 键的二级倍频吸收,由于在 640~1 100 nm 波段内,葡萄糖的光谱信息被水分子对光的吸收掩盖,因此在这一波段内并不能明显观察到葡萄糖的吸收峰。为了从中提取葡萄糖分子的光谱信息,这里采取化学计量学方法中的 IPLS 和 MWPLS 方法建立定量预测模型。

2.3 IPLS 光谱分析

将整个光谱(640~1 100 nm)平均分成 10 段,设置最大主成分数为 20,对每个区间光谱采用 PLS 建立模型,利用 RMSECV 判断每个区间 PLS 建模时的最佳主成分数,结果如图 3 所示,其中图中灰色柱体中数字表示该波谱区间 PLS 建模时最佳主成分数,柱高表示该波谱区间采用最佳主成分数建模时所对应的 RMSECV。例如,640~687 nm 柱高中 2 表示此间区 PLS 建模时的最佳主成分数,灰色柱高表示此区间采用 PLS 建模时对应 RMSECV 大小,图中纵坐标位于 0.3 附近的虚线表示整段光谱(640~1 100 nm)进行 PLS 建模时最佳主成分数 9 时所对应的 RMSECV。图 3 显示,当主成分数为 8 时在光谱区间

917~963 nm 建立模型,所对应的 $RMSECV$ 最小,而此区间正好对应 C-H 的三级倍频,表明 IPLS 所提取的光谱波段与葡萄糖分子官能团对应的近红外特征波段基本吻合。

2.4 利用 IPLS 对葡萄糖光谱进行数据分析

为了进一步优化最适宜血糖建模的光谱区间,将 640~1 100 nm 光谱分别等分为 2~15 份分别建立 IPLS 模型,之后对比每次选出的最佳光谱区间建模后所对应预测集的预测均方差 $RMSEP$ 、相关系数 R 及预测偏差 $Bias$ 。为了验证马氏距离在奇异样品检测中的作用,将含奇异光谱和最小半球体积法排除奇异样品后的结果作了对比分析,建模结果分别如表 1 和表 2 所示。

表 1 采用 IPLS 建立模型时得到的预测结果

Tab.1 Prediction results by IPLS method

| 波长范围/ nm | R | $RMSEP/(mmol \cdot L^{-1})$ | $Bias/(mmol \cdot L^{-1})$ |
|-------------|---------|-----------------------------|----------------------------|
| 870~1 100 | 0.793 9 | 0.488 7 | 0.031 2 |
| 870~985 | 0.575 3 | 0.852 5 | 0.031 5 |
| 916~1 008 | 0.731 4 | 0.563 3 | 0.029 3 |
| 903~968 | 0.572 2 | 0.818 2 | 0.077 6 |
| 927~985 | 0.708 5 | 0.625 3 | 0.011 3 |

表 2 剔除奇异样本后采用 IPLS 建立模型时得到的预测结果

Tab.2 Prediction results by IPLS method without outlier sample

| 波长范围/ nm | R | $RMSEP/(mmol \cdot L^{-1})$ | $Bias/(mmol \cdot L^{-1})$ |
|-------------|---------|-----------------------------|----------------------------|
| 870~1 100 | 0.974 3 | 0.188 2 | -0.089 4 |
| 870~985 | 0.966 7 | 0.235 3 | -0.054 6 |
| 916~1 008 | 0.963 4 | 0.232 2 | -0.122 5 |
| 903~968 | 0.954 4 | 0.235 0 | -0.083 6 |
| 927~985 | 0.968 8 | 0.209 1 | -0.101 9 |

结果表明:(1)根据马氏距离采用最小半球体积法来剔除奇异样品可以有效改良预测模型,可以提高预测集的相关系数 R 以及降低 $Bias$ 即预测值与实际值的偏差;(2)采用 IPLS 建立光谱与血糖浓度模型,可以实现血糖光谱信息区间的优化;(3)870~1 100 nm 范围内可以用于建立血糖浓度预测模型,其光谱信息主要源于葡萄糖甲基 C-H 键对称振动三级倍频与 O-H 键伸缩振动的二级倍频。对应于 870~1 100 nm 光谱,剔除奇异样品后的预测结果如图 4 所示。

2.5 利用 MWPLS 对葡萄糖光谱进行数据优化

虽然 IPLS 方法指出选择波段建模的效果优于全谱建模,并给出全谱中残差变小波段的大致位置,但仍无法精确波段起止点。因此采用 MWPLS 方法进一步决定所需波段位置并建立模型。设定窗口宽度,并对全谱移动窗口。在每一个窗口位置,计算每个成分数在此窗口位置时的 $RMSECV$,选取以某一给定的成分数达到较小的误差水平的波段,以此区间及其对应窗宽、窗口位置建模。设窗宽为 151 nm,成分数范围(1~20),对全谱作 MWPLS 处理,结果如图 5 所示。

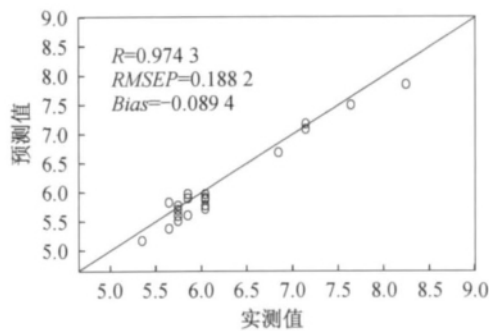


图 4 利用 10 个主成分数时 IPLS 预测结果
Fig.4 IPLS prediction results with 10 principal components

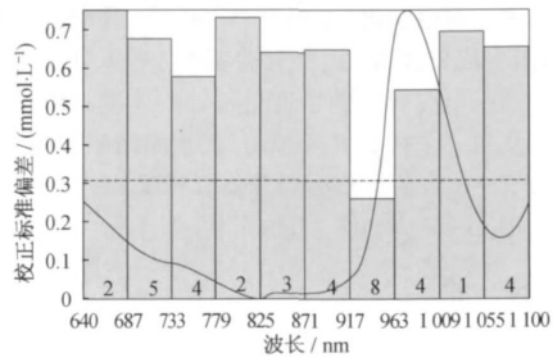


图 3 血糖近红外光谱 IPLS 建模分析
Fig.3 Analysis of blood glucose spectra by IPLS method

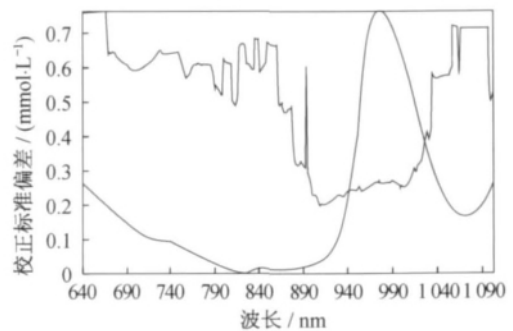


图 5 血糖近红外光谱 MWPLS 建模分析
Fig.5 Analysis of blood glucose spectra by MWPLS method

图 5 可以看出,在 915~1 065 nm 这个波段建立 PLS 模型,当主成分数为 9 时交叉验证均方差 $RMSECV$ 最小,因此选用这一谱段对血糖浓度进行回归建模。用建立的模型对验证集每一个样品进行预测,其预测结果如图 6 所示,相关系数 $R=0.982 2$,预测均方差 $RMSEP=0.163 5$ mmol/L,偏差 $Bias=-0.087 3$ mmol/L。

3 结 论

本文主要研究了使用短波近红外进行血糖浓度检测的可行性。通过 IPLS 与 MWPLS 对 640~1 100 nm 血清光谱的数据建模,结果表明:(1)根据马氏距离对样品的光谱进行奇异样品筛选,可以提高建模的准确性;(2)采用 IPLS 方法对全谱进行分析,能够提取葡萄糖分子官能团对应的近红外特征谱段,从而有效提取光谱中有效信息建立较好的预测模型;(3)采用 MWPLS 的方法,能够找出包含光谱有效信息的精确波长起止点,进一步优化其预测模型。下一步的工作是将样本浓度范围扩大,从而使建立的模型适用范围更广泛,并提高血糖定量分析的准确性。本文的研究结果将为进一步发展短波近红外的无损血糖检测技术提供基础。

参考文献:

- [1] YANG W, LU J, WENG J, *et al.* Prevalence of diabetes among men and women in China[J]. *The New England Journal of Medicine*, 2010; 362(12): 1090-1101.
- [2] 肖 君, 王 龙, 骆清铭, 等. 基于近红外三波长的血糖检测系统研制[J]. *光电子·激光*, 2007, 18(9): 1135-1138.
- [3] 高丽丽, 陶 卫, 赵 辉, 等. 用于无创血糖检测的光声池设计与研究[J]. *光学仪器*, 2012, 34(2): 55-60.
- [4] 刘 娟, 王尚奇, 简水生. 血糖浓度检测技术的最新进展[J]. *激光生物学报*, 2005, 14(5): 393-396.
- [5] 李 刚, 刘玉良, 林 凌, 等. 采用多光程长建模方法检测血液成分含量[J]. *分析化学*, 2007, 35(10): 1495-1498.
- [6] 张 兰, 张清悦, 南 群, 等. 中红外光谱法在无创血糖检测技术中的应用[J]. *光学与光电技术*, 2008, 6(2): 81-84.
- [7] 王 云, 徐可欣, 常 敏. 近红外光谱技术检测牛奶中脂肪及蛋白质含量校正模型的建立[J]. *光学仪器*, 2006, 28(3): 3-7.
- [8] 陈文亮, 徐可欣, 杜振辉, 等. 人体无创血糖检测技术[J]. *仪器仪表学报*, 2003, 24(4): 258-264.
- [9] 陆婉珍, 袁洪福, 徐广通, 等. 现代近红外光谱分析技术[M]. 北京: 中国石化出版社, 2001: 18-19, 148-149.
- [10] 李振庆, 黄梅珍, 倪 一, 等. 改进偏最小二乘法在近红外牛奶成分测量中的应用[J]. *光学技术*, 2009, 35(1): 70-73.
- [11] 刘 蓉, 陈文亮, 徐可欣, 等. 奇异点快速检测在牛奶成分近红外光谱测量中的应用[J]. *光谱学与光谱分析*, 2005, 25(2): 207-210.
- [12] SAVITZKY A, GOLAY M J E. Smoothing and differentiation of data by simplified least squares procedures [J]. *Analytical Chemistry*, 1964, 36(8): 1627-1639.
- [13] 梁逸曾, 俞汝勤. 分析化学手册(十)化学计量学[M]. 2 版. 北京: 化学工业出版社, 2000: 161-162.

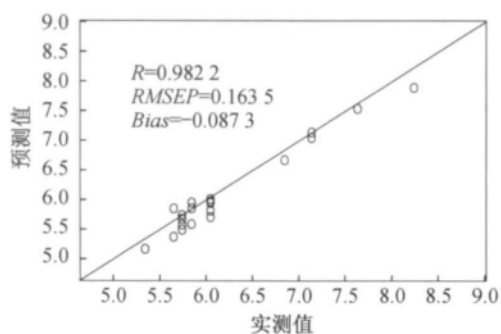


图 6 利用 9 个主成分数时 PLS 预测结果
Fig. 6 PLS prediction results with
9 principal components