

三维坐标异常数据判定方法的模拟与实验研究

王 林, 马雪洁, 孟丹蕊, 刘 蓉*, 徐可欣

天津大学精密测试技术及仪器国家重点实验室, 天津 300072

摘 要 近红外漫反射光谱具有无创伤、连续、无感染、速度快等诸多优势,在人体成分无创伤检测方面有很好的应用前景。但是在测量过程中,随机噪声、干扰组分以及检测条件的改变等容易导致异常光谱。判定并剔除异常光谱对于提高近红外无创血液成分检测的可靠性具有重要意义。首先分析了近红外漫反射光谱无创血糖检测中可能出现的异常数据类型,提出了一种综合利用马氏距离、光谱残差和化学值残差三个指标构造三维空间对样本集进行检验的三维坐标异常数据判定方法。其次,针对三层皮肤组织模型,在参数中设置人为失误、极端成分含量以及异常温度变化的样本,通过蒙特卡罗(MC)模拟程序得到一组正常模拟数据以及一组包含化学值异常和光谱异常的模拟数据,并利用三维坐标法进行异常数据的判定。结果显示,该方法能识别出全部异常样本,剔除这些异常样本后,偏最小二乘(PLS)校正模型的交互验证均方根误差(RMSECV)由 $21.2 \text{ mmol} \cdot \text{L}^{-1}$ 降低到 $1.1 \text{ mmol} \cdot \text{L}^{-1}$,初步验证了该方法的可行性。进一步,对三位受试者开展了口服葡萄糖耐量试验(OGTT),通过在测量受试者血糖参考值的同时同步采集其手指部位的漫反射光谱,获得了三组在体实验数据。并利用三维坐标法和蒙特卡罗交互验证法进行异常数据的判定和剔除,最后建立 PLS 模型比较两种异常数据判别方法的效果:剔除三维坐标法识别出的异常数据后,三组样本建立的校正模型的决定系数显著提升, RMSECV 平均值由 $2.1 \text{ mmol} \cdot \text{L}^{-1}$ 降低至 $0.8 \text{ mmol} \cdot \text{L}^{-1}$,效果优于蒙特卡罗交互验证法的结果。这些结果表明,基于马氏距离、光谱残差和化学值残差的三维坐标异常数据判定方法能有效识别近红外无创血糖测量中的异常数据,在在体成分检测应用中有显著优势。

关键词 近红外光谱; 无创检测; 异常光谱; 三维坐标法; 蒙特卡罗模拟

中图分类号: O433.4 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)09-2774-06

引 言

近红外漫反射光谱在人体成分无创检测方面具有诸多优势,目前已成功应用于血氧饱和度和血红蛋白的检测。该技术在人体血糖的无创检测方面也很有前景^[1-4],但由于血液中其他成分、测量条件改变、仪器系统误差等因素的影响,容易出现异常数据。

异常数据一般可分为光谱异常和化学值异常两种。光谱异常主要包括两类,一类是由样品中极端组成导致的高杠杆样本;另一类是由外界测量环境及样品来源变化导致的光谱特征峰异常。由于近红外光谱分析是一种间接分析方法,需要借助化学计量学手段建立糖浓度与光谱数据之间的校正模型,在测量糖浓度参考值的过程中,样品来源的变化以及操作人员的失误等可能会造成血糖浓度的化学值异常。这些异

常数据会严重影响近红外光谱分析的精度,因此,对异常数据的判定和剔除成为建立血糖浓度预测模型的重要环节之一。

目前已有多种异常数据判定方法。一类判定方法,考虑了校正集样本中待测物质的参考值信息,例如常用的留一交互验证法^[5]。但该方法容易出现过拟合现象,为了克服过拟合问题,逐渐发展出了蒙特卡罗交互验证法(Monte Carlo cross validation, MCCV)^[6],主要利用预测误差对异常样本的敏感特性,来检验光谱矩阵和性质矩阵的奇异点,具有较高的识别异常数据的能力^[7]。刘翠玲^[8]等在利用近红外光谱分析小麦粉时,使用 MCCV 方法对 150 个样本数据进行异常数据的判定,剔除异常数据后模型的交互验证均方根误差明显下降。何佳艳^[9]等在采用近红外光谱测定奶粉的脂肪含量时,使用 MCCV 方法两次剔除异常数据,提高了模型的泛化性和预测能力。还有一类判定方法不考虑校正集样本中待测物质含量的参考值信息,以欧氏距离、Cook 距离、马氏距

收稿日期: 2018-08-01, 修订日期: 2018-12-06

基金项目: 国家自然科学基金项目(81471698, 81401454)资助

作者简介: 王 林, 1992 年生, 天津大学精密测试技术及仪器国家重点实验室硕士研究生 e-mail: linwang2016@tju.edu.cn

* 通讯联系人 e-mail: rongliu@tju.edu.cn

离、杠杆值或光谱残差为基础来识别异常数据^[10]。杨峰^[11]等针对高光谱数据异常值影响叶绿素密度反演精度的问题,基于马氏距离判定并剔除异常样本,提高了校正模型的精度。

以上方法大多考虑单一类型指标来进行异常数据判定,不能全面地识别异常数据,还可能会造成模型错误。针对此问题,提出了一种基于马氏距离、光谱残差和化学值残差构造三维空间的三维坐标异常数据判定方法,通过对蒙特卡罗(Monte Carlo, MC)模拟光谱中异常数据的判定,初步验证了方法的有效性;进一步,开展了人体口服葡萄糖耐量实验,利用三维坐标法和 MCCV 对在体近红外实验进行了异常数据的判定,验证了三维坐标异常数据判定方法在人体近红外无创血糖检测应用中的有效性和优越性。

1 三维坐标异常数据判别原理

样本的马氏距离是指样本光谱与标准光谱集的平均光谱之间的距离,可以衡量一个样本对整个校正集的影响,从而识别出样本中的高杠杆点异常样本。其计算公式为

$$MD_i = (t_i - \bar{t})^T Cov^{-1}(t)(t_i - \bar{t}) \quad (1)$$

其中, MD_i 为校正集中样本 i 的马氏距离; t 为原光谱数据的得分矩阵; t_i 为样本 i 的得分向量; \bar{t} 为所有样本的平均得分向量; Cov 为得分矩阵的协方差矩阵。马氏距离的阈值通常设定为各样本马氏距离平均值的 2~3 倍,马氏距离大于该阈值的样本被认定为异常光谱。

光谱残差谱是指原始光谱与重构光谱之间的残差谱,反映了被忽略部分的信息。正常情况下,每个样本的光谱残差应处于同一水平,样本光谱残差

$$SR_i = (x_i - \hat{x}_i)^T (x_i - \hat{x}_i) \quad (2)$$

式(2)中, SR_i 是校正集中样本的光谱残差, x_i 是样本 i 的原始光谱数据, \hat{x}_i 为样本 i 的光谱经主成分分析重构的光谱数据。

求出所有样本光谱残差的平均值和标准偏差,并构造统计量 T_i 进行 T 检验。

$$T_i = \frac{SR_i - \overline{SR}}{S} \quad (3)$$

给定自由度 n 和显著水平 α (置信度 $1-\alpha$),查表可得到统计量 T_i 的阈值 $T_{(\alpha, n)}$,若 $T_i > T_{(\alpha, n)}$,则判定 SR_i 对应的原始样本 i 为异常样本,应予以剔除。

化学值残差是指样本的化学值即参考值与模型预测值之间的差。样本 i 的化学值残差 Res_i

$$Res_i = y_i - \hat{y}_i \quad (4)$$

其中, y_i 为样本的化学值; \hat{y}_i 为模型的预测值。进一步构造统计量 F_i 进行 F 检验

$$F_i = \frac{Res^2(i)}{Res^2} = \frac{Res^2(i)}{\sum_{j=1}^{n-1} Res^2(j)/(n-1)} \quad (5)$$

给定自由度 n 和显著水平 α (置信度 $1-\alpha$),查表可得到统计量 F_i 的阈值 $F_{\alpha}(1, n-1)$,大于该阈值的样本判断为异常样本。

为了更有效地识别出所有可能的异常样本,综合利用马氏距离、光谱残差和化学值残差三个指标构造三维空间对样

本集进行检验,称为三维坐标异常数据判别方法。只要样本有一项指标是异常的,就判定其为异常样本。

2 三维坐标异常数据判别方法的模拟验证

2.1 MC 模拟样本设置

MC 模拟是研究光子在生物组织中传输问题的理想模型^[12]。针对人体三层皮肤组织模型,利用 MC 模拟获取血糖近红外漫反射光谱。以无限细点光源入射,波长范围为 1 000~1 700 nm(间隔 20nm),光源-探测器距离设定为 1.04 mm。为了更容易观察到光谱随浓度变化的规律,设置样品葡萄糖浓度范围为 0~150 mmol·L⁻¹(间隔 10 mmol·L⁻¹),共 16 个样本,第一个样本为参考样本,其他样本编号为 1—15 号,温度设定为 37 °C,入射光子数设定为 10⁸,皮肤光学参数来自参考文献^[13,15],采用 MCX 程序^[14]模拟光子在组织中的传输过程,获取不同糖浓度下的漫反射光强,得到正常组数据。并通过人为参数设置的方式来获得异常组数据:将 4 号样品糖浓度设置为 49 mmol·L⁻¹,模拟得到化学值异常样本;将 7 号样品的温度设为 38 °C,模拟谱峰异常的光谱异常样本;将 10 号样品设置为含极端组成的样品,模拟得到高杠杆点的光谱异常样本,其他设置与正常组相同。

2.2 模拟结果与分析

以无糖样本为参考样品,分别计算正常组和异常组中的

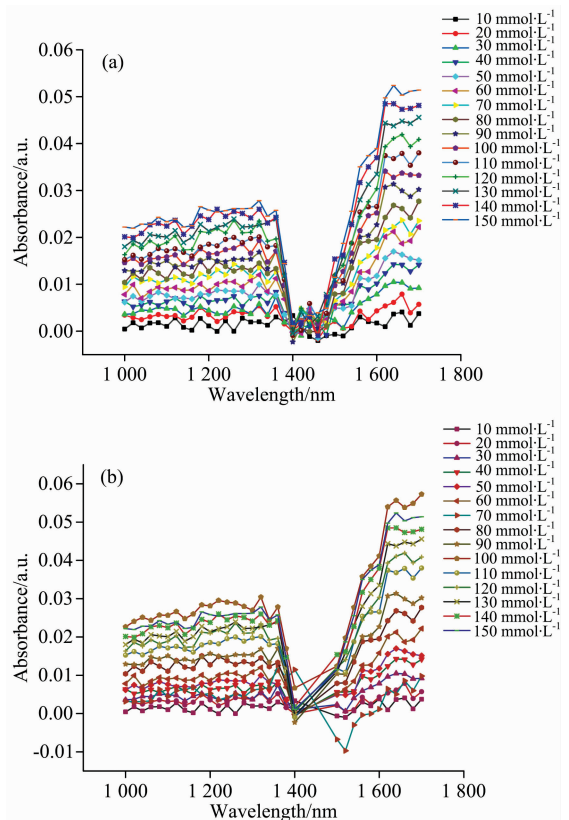


图 1 模拟样品吸光度变化量曲线

(a): 正常组; (b): 异常组

Fig. 1 Absorbance variation of the simulated samples

(a): Normal group; (b): Abnormal group

测量样本相对于无糖参考样本的吸光度变化量,结果分别如图 1(a)和(b)所示。

从图 1(a)正常组模拟样本吸光度变化量曲线可以看出,由不同糖浓度变化引起的吸光度变化量曲线很容易区分开,且随着浓度变化呈现出阶梯型变化,吸光度变化量随糖浓度增加而增大。从图 1(b)异常组模拟样品吸光度变化量曲线可以看出,受温度影响的 7 号样本的吸光度变化量曲线形状发生了较明显的改变,且在 1 400 nm 处出现了正的吸收峰,在 1 500 nm 处出现了负的吸收峰。而其余两个异常样本的吸光度变化量曲线与正常样品的曲线形状基本相同,较难直接从光谱分辨出来。

2.3 模拟样本中异常数据的判别

采用三维坐标异常数据判别方法检验模拟得到的异常组样本数据。基于马氏距离、光谱残差和化学值残差三个指标构造三维空间,分别求出各个样本的指标值进而确定其三维坐标,并以各指标的阈值确定正常数据空间范围(马氏距离阈值为其平均值的 2.5 倍,光谱残差统计量和化学值残差统计量的阈值为 95%置信水平下的统计量阈值),进行异常数据的判定,判定结果如图 2 所示。

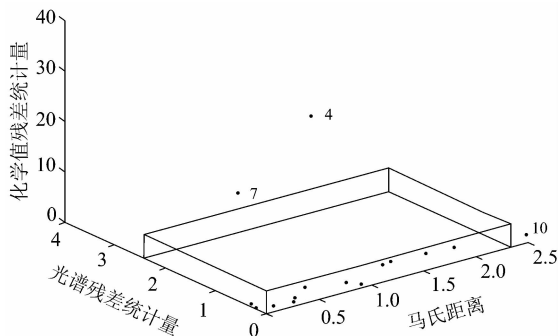


图 2 异常组模拟样本的判定结果

Fig. 2 Determination of outliers in the simulated samples

从图 2 可以看出,异常的 4, 7 和 10 号样本在正常数据空间范围外,三维坐标法准确地将其判定为异常数据。进一步,基于偏最小二乘(partial least square, PLS)建立光谱数据与葡萄糖浓度之间的校正模型,并通过留一交互验证方法对建模结果进行评价。与全部数据的建模结果相比,剔除异常样本后建立的校正模型的决定系数 R^2 由 0.84 提高到了 0.99,交互验证的均方根误差(root-mean-square error of cross-validation, RMSECV)由 $21.2 \text{ mmol} \cdot \text{L}^{-1}$ 下降到 $1.1 \text{ mmol} \cdot \text{L}^{-1}$ 。由此可见,基于三维坐标法剔除异常数据后,校正模型的精度得到提升,初步验证了该方法的效果。

3 实验部分

3.1 试验系统

实验采用了自行搭建的光谱采集系统,系统示意图如图 3 所示。

试验系统中,光源采用 3 dB 带宽为 40~60 nm 的 6 个 SLD 光源,中心波长分别为 1 050, 1 219, 1 314, 1 409,

1 550 和 1 609 nm。通过计算机控制光开关,实现 6 个光源的分时测量,其中光开关采用低耗高耦合效率六路 1×1 多通道单模光开关,其工作波长范围为 1 000~1 650 nm。光纤采用芯径为 0.2 mm(含包层和涂覆层 0.28 mm)的石英光纤,光纤探头中心入射,由光源-探测器距离 1.04 mm 处的环状接收光纤束接收漫反射光信号。

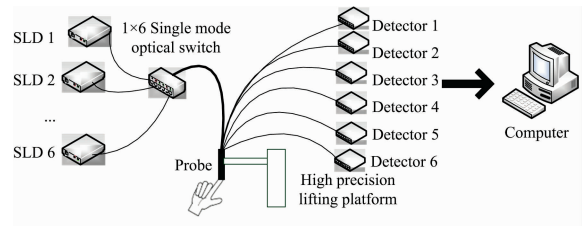


图 3 试验系统示意图

Fig. 3 Schematic diagram of the experimental system

3.2 方法

针对健康志愿者,选取手指作为测量部位,开展口服葡萄糖耐量试验(oral glucose tolerance test, OGTT)^[12]。实验系统开机 2 h 后,光源和检测器基本达到稳定,让受试者调整好坐姿,将右手放置于人机接口的平台上,手指与光纤探头接触,标记测量部位。手指与光纤探头接触 10 min,使接触温度稳定,开始实验。实验过程中为保证血糖测量准确性,通过针刺取血后采用两台便携式血糖仪(GT-1820, Arkray, Japan)同时测量血糖浓度值,并取其均值作为血糖浓度参考值。

受试者在空腹状态下(血糖值大约 $3.8 \sim 5.5 \text{ mmol} \cdot \text{L}^{-1}$)口服溶解了 75 g 葡萄糖的糖水溶液 250 mL(健康志愿者口服葡萄糖后的血糖峰值为 $8.9 \sim 12.6 \text{ mmol} \cdot \text{L}^{-1}$)。每隔 5~15 min 用便携血糖仪测定血糖浓度值,同时采集手指部位的漫反射光谱。由于口服葡萄糖后血糖浓度上升较快,因此在血糖上升阶段采血相对频繁,下降阶段采血间隔时间较长。当受试者的血糖值恢复到空腹状态水平时,停止在体测量。实验采集 3 名年龄 25~30 岁健康受试者的数据,每位受试者的测量时间 2 h 左右,每名志愿者共测量 10 次血糖参考值及同步的漫反射光谱数据。

4 结果与讨论

对 3 名受试者的实验数据进行处理,分别以空腹状态下的光谱作为参考,计算其余九个时刻下的吸光度变化量。3 名受试者的数据分别记为数据组 1~3,并对每组 9 个样本数据按测量时刻编号,记为 1~9。以数据组 2 为例,吸光度变化曲线如图 4 所示。

从图 4 可以看出,由不同糖浓度变化引起的吸光度变化量曲线很容易区分开,但某些样本的吸光度变化量与样本浓度的相关性较弱。例如,9 号样品的吸光度变化量曲线发生了较为明显的改变,其样本数据可能受到了外界干扰。

由于在体实验中存在诸多的不确定性,不清楚哪些样本是异常的,这完全不同于模拟数据中已知异常样本的情况,

因此,选择三维坐标法和 MCCV 方法分别对异常数据进行判别及剔除。对 OGTT 实验得到的三组样本,以马氏距离、光谱残差和化学值残差三个指标构造三维空间,分别求出各个样本的指标值进而确定其三维坐标,并以各指标的阈值确定正常数据空间范围(马氏距离阈值为其平均值的 2.5 倍,光谱残差统计量和化学值残差统计量阈值为 95%置信水平下的统计量阈值),据此进行异常数据的判定,判定结果如图 5 所示。

从图 5 中可以看出,数据组 1 中 1, 4 和 6 号样本数据点在正常数据空间范围外,被判定为异常数据[图 5(a)];数据组 2 中的 4 和 9 号样本数据点在正常数据空间范围外,被判定为异常数据[图 5(b)];数据组 3 中 3, 5 和 9 号样本数据点在正常数据空间范围外,被判定为异常数据[图 5(c)]。

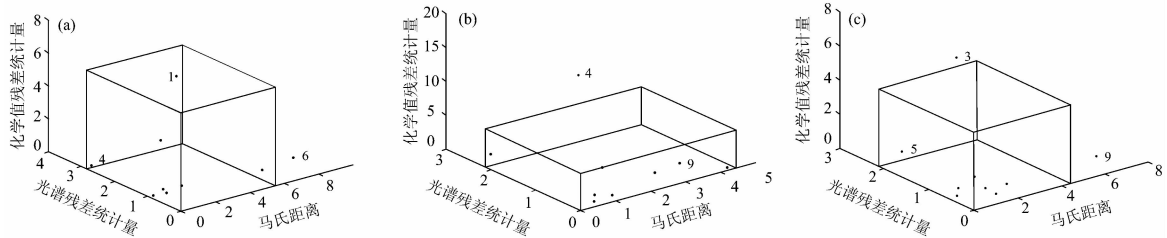


图 5 在体实验异常数据判定结果(三维坐标法)

(a): 数据组 1; (b): 数据组 2; (c): 数据组 3

Fig. 5 Determination of outliers in the *in vivo* experimental samples (Three-dimensional coordinate method)

(a): Group 1; (b): Group 2; (c): Group 3

利用 MCCV 方法对实验数据进行判定时,首先利用蒙特卡罗随机取样将三组样本数据分为校正集(80%样本数)和验证集(20%样本数),利用校正集数据建立 PLS 模型预测验证集数据。循环此过程 25 次以保证每个样本均被预测到。计算每个样本预测误差的均值和方差,根据样本的均值-方差

分布图进行异常数据的判定,判定结果如图 6 所示。数据组 1 中 4, 6 号样本[图 6(a)]、数据组 2 中的 1 和 9 号样本[图 6(b)]、数据组 3 中 3 和 9 号样本[图 6(c)]在正常数据空间范围外,被 MCCV 方法判定为异常数据。

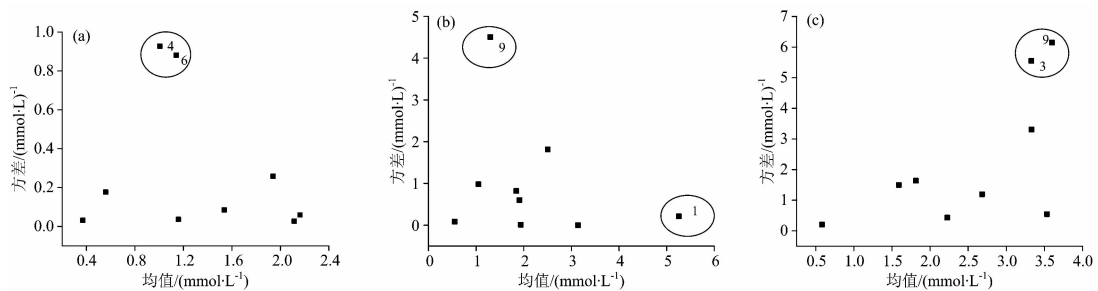


图 6 在体实验异常数据判定结果(MCCV 方法)

(a): 数据组 1; (b): 数据组 2; (c): 数据组 3

Fig. 6 Determination of outliers in the *in vivo* experimental samples (MCCV method)

(a): Group 1; (b): Group 2; (c): Group 3

进一步,分别剔除三维坐标法和 MCCV 方法识别出的异常样本后重新建立 PLS 校正模型,剔除异常样本前后校正模型的决定系数和交互验证预测均方根误差如图 7 所示。

从图 7(a)可以看到,未剔除异常样本的校正模型的决定系数最低, R^2 平均值为 0.57,对于受试者 1 和受试者 3 的校

正模型,利用 MCCV 和三维坐标法剔除异常数据后,模型的决定系数 R^2 显著增加;对于受试者 2, MCCV 方法剔除异常数据后,模型 R^2 变化不大,但三维坐标剔除异常数据后, R^2 提高到 0.98。从图 7(b)可以看到,未剔除异常样本的校正模型的 RMSECV 较大,平均值为 $2.1 \text{ mmol} \cdot \text{L}^{-1}$;利用

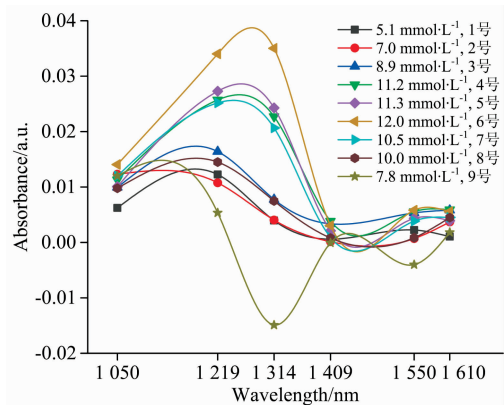


图 4 2 号受试者漫反射光谱的吸光度变化量

Fig. 4 Absorbance change of the diffuse reflectance (2th subject)

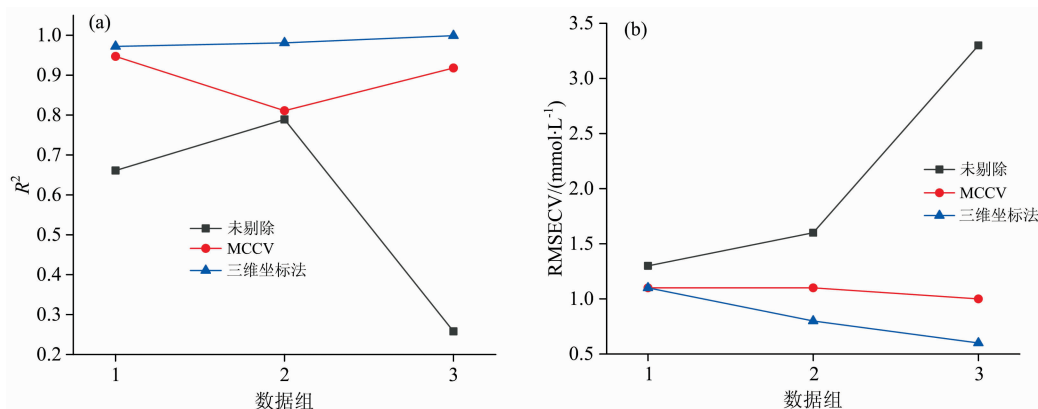


图 7 在体实验异常样本剔除前后模型 R^2 和 RMSECV

(a): 决定系数; (b): 交互验证均方根误差

Fig. 7 R^2 and RMSECV before and after elimination of abnormal samples in *in vivo* experiments

(a): Coefficient of determination; (b): RMSECV

MCCV 方法和三维坐标法剔除异常样本后, 受试者 2 和受试者 3 的预测误差明显降低, 三组数据建立的校正模型的 RMSECV 平均值分别降低到 1.0 和 0.8 $\text{mmol} \cdot \text{L}^{-1}$ 。

综上所述, 三维坐标法和 MCCV 方法都能识别异常数据, 且剔除识别出的异常数据后, 模型的精度都有所提升。从剔除异常样本后模型的结果看, MCCV 方法的效果要逊色于三维坐标法。因此, 在近红外无创血糖测量的建模过程中, 可优先考虑利用三维坐标法对异常数据进行判定并剔除, 优化校正模型, 提高近红外无创血糖测量的精度。

5 结 论

针对常用判定方法判定异常样本不全面的问题, 提出了

一种基于化学值残差、马氏距离和光谱残差的三维坐标异常数据判定方法, 并通过对 MC 获得的模拟数据中的异常样本的判别, 初步验证了方法的可行性; 最后开展了人体近红外漫反射光谱实验, 利用三维坐标法和 MCCV 方法分别在体实验中的异常光谱进行判别, 建立的校正模型的预测结果表明, 剔除异常样本后模型的预测性能得到显著提高, 且三维坐标法的效果要明显优于 MCCV 方法。结果表明, 三维坐标法能更有效地识别在体实验中的异常样本, 更适用于近红外光谱无创血糖测量中的异常数据判定。在判定过程中, 相比于单一指标的判定, 这种方法会存在判定时间复杂度增加的问题。

References

- [1] Liu J, Liu R, Xu K. *Applied Spectroscopy*, 2015, 69(11): 1313.
- [2] Goodarzi M, Sharma S, Ramon H, et al. *Trends in Analytical Chemistry*, 2015, 67: 147.
- [3] Yadav J, Rani A, Singh V, et al. *Biomedical Signal Processing and Control*, 2015, 18: 214.
- [4] Zhang W, Liu R, Zhang W, et al. *Biomedical Optics Express*, 2013, 4(6): 789.
- [5] LIU Xue-yi, LI Ping, GAO Chuan-hou(刘学艺, 李平, 郜传厚). *Journal of Shanghai Jiaotong University(上海交通大学学报)*, 2011, (8): 1140.
- [6] Liu Z C, Cai W S, Shao X G. *SCIENCE CHINA Chemistry*, 2008, 51(8): 751.
- [7] LI Shui-fang, SHAN Yang, FAN Wei, et al(李水芳, 单杨, 范伟, 等). *Food Science(食品科学)*, 2011, 32(8): 182.
- [8] LIU Cui-ling, SUN Xiao-rong, WU Jing-zhu, et al(刘翠玲, 孙晓荣, 吴静珠, 等). *Journal of Agricultural Mechanization Research(农机化研究)*, 2014, (4): 46.
- [9] HE Jia-yan, LI Ting, GUO Chang-kai, et al(何佳艳, 李亭, 郭长凯, 等). *Food and Fermentation Industries(食品与发酵工业)*, 2017, 43(10): 228.
- [10] HE Yong, LIU Fei, LI Xiao-li, et al(何勇, 刘飞, 李晓丽, 等). *Spectroscopy and Imaging Technology in Agriculture(光谱及成像技术在农业中的应用)*. Beijing: Science Press(北京: 科学出版社), 2016.
- [11] YANG Feng, ZHANG Yong, CHEN Jun-xu, et al(杨峰, 张勇, 谌俊旭, 等). *Remote Sensing Information(遥感信息)*, 2017, (4): 64.
- [12] Jacques S L. *Physics in Medicine & Biology*, 2013, 58(11): 37.
- [13] Bashkatov A N, Genina E A, Tuchin V V. *Journal of Innovation in Optical Health Science*, 2011, 4(1): 9.
- [14] Fang Q. *Biomedical Optics Express*, 2010, 1(1): 165.

[15] Seike M, Saitou T, Kouchi Y, et al. *Journal of Physiological Sciences*, 2011, 61(4): 321.

Simulation and Experiment Study on Three-Dimensional Coordinate Outlier Detection Method

WANG Lin, MA Xue-jie, MENG Dan-ruì, LIU Rong*, XU Ke-xin

State Key Laboratory of Precision Measuring Technology and Instruments, Tianjin University, Tianjin 300072, China

Abstract Near-infrared diffuse reflectance spectroscopy has many advantages, such as being non-invasive, continuous, non-infectious, fast, in the non-invasive detection of body components. It has a great prospect in the application of blood glucose measurement in vivo. However, outliers often occur in the process of measurement due to the random noise, the change in interference components or the measurement conditions. Therefore, it is of great significance to eliminate the outliers in the near-infrared spectroscopy and thus improve the reliability of non-invasive blood components measurement. In this paper, the types of outliers that may occur in the blood glucose sensing by near-infrared diffuse reflectance were analyzed, and a three-dimensional coordinate outlier determination method based on the three-dimensional space constructed by the residual of chemical value, the Mahalanobis distance and the spectral residuals was proposed firstly. Then, it was used to discriminate the outliers in the simulated spectra of three-layer skin model by Monte Carlo program, where the abnormal data was obtained by adding the artificial errors, abnormal chemical values and abnormal temperature changes in the parameters setting in Monte Carlo simulation. All the outliers could be found successfully by the three-dimensional coordinate outlier determination method, and the root-mean-square error of cross-validation (RMSECV) of the Partial Least Square (PLS) model was reduced from 21.2 to 1.1 mmol · L⁻¹ after the removal of outliers. Further, the oral glucose tolerance tests (OGTTs) of three volunteers were carried out, where three groups of experimental data were obtained by measuring the reference blood glucose concentrations and collecting the diffuse reflectance of finger synchronously, and Monte Carlo Cross-Validation outlier detection method and three-dimensional coordinate method were used to detect the outliers, respectively. Results showed that, after the removal of outlier by the three-dimensional coordinate method, the coefficient of determination of calibration model increased significantly, and the average RMSECV value of calibration model for three sets of samples was reduced from 2.1 to 0.8 mmol · L⁻¹, which was better than that of MCCV method. All these results indicated that, three-dimensional coordinate method can effectively determine the outlier in the near-infrared diffuse reflectance and it's more suitable for the non-invasive blood glucose measurement in vivo by near-infrared diffuse reflectance spectroscopy.

Keywords Near-infrared spectroscopy; Non-invasive detection; Outliers; Three-dimensional coordinate method; Monte Carlo simulation

(Received Aug. 1, 2018; accepted Dec. 6, 2018)

* Corresponding author