

滤光片型近红外仪器模型传递的研究

陈嘉威¹, 周昌乐^{1*}, 张晔晖², 徐晓洁³, 林坤辉⁴, 叶楠²

1. 厦门大学信息科学与技术学院, 福建 厦门 361005
2. 中国农业大学信息与电气工程学院, 北京 100094
3. 北京农学院, 北京 102206
4. 厦门大学软件学院, 福建 厦门 361005

摘要 模型传递问题是近红外光谱分析技术中解决数据通用性的关键问题。文章以玉米籽粒近红外光谱图(检测其中水分含量)为例,考察了模型传递的问题。使用斜率截距算法,直接校正法和目标因子分析等算法,在5台滤光片型近红外仪器上实现了模型传递,并比较了各种方法的模型传递效果。研究表明,直接校正法的模型传递效果最好,4台从仪器的平均传递差异度为7.01%。文章还研究了标准样品数量对模型传递效果的影响。作为转换集的标准样品数目越多,模型传递效果越好,一般有20个标准样品就能达到稳定的效果。当转换集小于20时,直接校正法的传递效果急剧下降,而标准样品数量对斜率截距法和目标因子分析法的影响不明显。

关键词 光谱分析;模型传递;近红外光谱;多元校正

中图分类号: TP39 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2008)10-2459-04

引言

近红外光谱分析技术由于样品处理简单,分析速度快,不用试剂和无损检测等特点,具有广阔的应用前景。其中滤光片型近红外仪器由于价格低廉、性能稳定,将对近红外光谱分析技术的普及与推广起着至关重要的作用。

近红外光谱分析的前提是数学模型,然而建立一个稳定、可靠、准确的数学模型是相当复杂的过程,需要耗费大量的时间、人力、物力和财力。由于仪器之间的差异,一台仪器建立的模型在其他仪器上往往不能使用,同一台仪器也会因为仪器元件的老化和更换使原来建立的模型不再适用。模型传递技术通过寻求两台仪器之间的一种变换关系来避免重复建模,从而大大降低近红外光谱分析的成本,提高分析的效率。模型传递的成功与否直接影响近红外光谱分析技术的推广应用。

模型传递问题主要通过有标样法来解决,即选择一定数量的样品作为标准样品转换集,并在主仪器和从仪器上分别测得其信号,从而找出该函数关系。有标样法包括基于预测结果的校正方法和基于光谱信号的校正方法。基于预测结果的校正以斜率截距算法(Slope/Bias)为代表^[1]。基于光谱信

号的校正主要分为两类:一是全光谱校正法,如校正吸光度矩阵的直接校正法(DS)^[2],对波长分段进行校正的PDS(Piecewise Direct Standardization)算法^[3],以及分别对波长和吸光度进行校正的Shenk算法等^[4];二是以提取光谱数据中的有效信息为目的,对光谱数据进行筛选过滤后进行校正,如普鲁克分析法^[5]、目标因子分析法(TFA)^[6]、正交信号校正法^[7]、基于支持向量机的模型传递法^[8]等。这些算法主要应用于连续扫描型近红外仪器(主要是傅里叶型和光栅型)的模型传递之中。

由于滤光片型近红外仪器的波长点较少,存在数据处理量小等特点,适用于其他类型仪器的PDS算法、Shenk算法等,对滤光片型近红外仪器并不适用。本文选择最有代表性的斜率截距算法、直接校正法、目标因子分析法对滤光片型近红外仪器的模型传递进行了研究,为滤光片型近红外仪器的推广应用提供依据。

1 原理与算法

1.1 斜率截距校正法

斜率截距校正法(Slope/Bias)是建立主仪器与从仪器所得预测结果间函数关系的方法。通过校正预测结果,实现模

收稿日期:2008-02-08,修订日期:2008-04-20

基金项目:国家“863”计划项目(2006AA01Z129)资助

作者简介:陈嘉威,1982年生,厦门大学信息科学与技术学院博士研究生 e-mail: jw_chen@126.com *通讯联系人

型传递。

在主仪器上建立校正模型 b 。选择 N_s 个标准样品作为转换集, 用主仪器与从仪器测得的标准样品光谱阵 T_m , T_s 分别计算标样集预测矩阵 y^m , y^s , 如式(1)所示。

$$y^m = T_m \times b, \quad y^s = T_s \times b \quad (1)$$

假定 y^m 和 y^s 存在关系 $y^m = bias + slope \times y^s$, 用最小二乘法求出 $slope$ 和 $bias$, 即为这一线性模型的斜率和截距。过程如下:

$$S_{xx} = (y_j^s - \bar{y}^s)^2, \quad S_{yy} = (y_j^m - \bar{y}^m)^2, \\ S_{xy} = (y_j^s - \bar{y}^s)(y_j^m - \bar{y}^m) \quad (2)$$

$$slope = S_{xy} / S_{xx}, \quad bias = \bar{y}^m - slope \times \bar{y}^s \quad (3)$$

从仪器上测得未知样品光谱阵 X_s , 通过斜率截距校正因子, 就可利用主仪器建立的模型 b 预测未知样品的化学值, 如式(4)所示。

$$y_{im} = bias + slope \times (X_s \times b) \quad (4)$$

1.2 直接校正法

直接校正法(DS)是一种多元全光谱模型传递方法。通过比较主从仪器的光谱数据的差异, 实现对光谱数据吸光度的校正。它的基本思想是利用转换矩阵 F 将从仪器上测得的未知样品光谱 X_s 转换为 X_{st} , 理论上 X_{st} 与主仪器上测得的光谱是一致的。

主仪器上得到标准样品光谱矩阵 T_m , 从仪器上得到标准样品光谱矩阵 T_s , 二者因为在不同条件下测定, 所以存在着差异, 通过传递矩阵 F 将 T_s 转换成 T_m , 使二者相匹配。其中矩阵 F 描述了主、从两种测定条件下光谱的差异。

由 $T_m = T_s \times F + E$, E 为残差矩阵, 则通过最小二乘法计算可得到转换矩阵 $F = T_s^{-1} \times T_m$, 其中 T_s^{-1} 为 T_s 的广义逆, F 为 $m \times m$ 矩阵(m 为波长点数)。

对在从仪器上测得的未知样品光谱矩阵 X_{su} , 利用 F 将 X_{su} 转化成与主仪器匹配的光谱 X_{st} 。

$$X_{st} = F \times X_{su} \quad (5)$$

再由主仪器上建立的校正模型 b 计算最终结果。

$$y_{im} = X_{st} \times b \quad (6)$$

1.3 目标因子分析法

目标因子分析法(TFA)是一种多元统计分析方法, 目的是将原变量表达为新变量的线性组合。在分析过程中根据有关吸光度矩阵的基本情况, 找到最佳因子数目, 使多余信息减到最小。目标因子分析是在不损失原有信息的前提下, 从光谱信号中获取有用信息的一种有效途径。

基于目标因子分析的模型传递算法计算过程如下:

(1) 在主仪器测得的标准样品光谱 T_m , 当各组分均符合比耳定律, 其吸收曲线线性无关且满足吸光度加和性原理时, 经因子分析, 可将 T_m 分解为两个矩阵相乘, 得到载荷矩阵和因子矩阵:

$$T_m = T_{ms} \times T_{mc} \quad (7)$$

(2) 对从仪器测得的标准样品光谱进行主成分分析, 得到载荷矩阵和因子矩阵:

$$T_s = T_{ss} \times T_{sc} \quad (8)$$

(3) 通过目标变换使二者的因子矩阵相等, 即

$$T_{mc} = F_s \times T_{sc} \quad (9)$$

(4) 目标变换后从仪器的载荷矩阵为 $T_{ss}^{new} = T_{ss} \times F_s$, 由于主仪器的因子矩阵可以通过 $T_{mc} = T_{ms}^{-1} \times T_m$ 计算, 同时由于变换后从仪器的因子矩阵等于主仪器的因子矩阵 $T_{sc}^{new} = T_{mc}$, 所以从仪器的光谱可以表达为:

$$T_s = T_{ss}^{new} \times T_{sc}^{new} = T_{ss} \times F_s \times T_{ms}^{-1} \times T_m \quad (10)$$

因而传递矩阵 F 可由式(11)得到

$$FF = T_{ss} \times F_s \times T_{ms}^{-1}, \quad F = T_s^{-1} \times FF \times T_s \quad (11)$$

传递矩阵 F 可以用于传递两台仪器的光谱:

$$X_{st} = X_s \times F \quad (12)$$

由传递矩阵 F 和主仪器上建立的校正模型 b 可以得到模型传递后的预测结果。

$$y_{im} = X_{st} \times b \quad (13)$$

2 实验部分

2.1 仪器及试剂

105 种玉米籽粒样品, 用国标法测得水分值, 范围为 10.4% ~ 18.8%, 平均水分含量为 13.6%。5 台滤光片型近红外玉米品质分析仪, 由中国农业大学研制; 工作谱区为 880 ~ 1100 nm; 吸光度范围 0 ~ 5 AU; 滤光片带宽 10 ~ 15 nm; 共 14 个波长点。

2.2 光谱测定

在 5 台滤光片型近红外玉米品质分析仪上用漫透射方式分别测定 105 个玉米样品结果图略。

3 结果与讨论

3.1 各台仪器自建模型预测的结果

将每台仪器测定的 105 个样品, 分 80 个作为校正集, 25 个作为检验集, 各台仪器的校正集和检验集的样品相应一致; 用偏最小二乘法(PLS)建立模型, 预测结果表 1 所示。

Table 1 Result of PLS model

| Instrument | R^2 | SEP | RSD/ % |
|------------|---------|------|--------|
| S1 | 0.970 7 | 0.36 | 2.65 |
| S2 | 0.971 1 | 0.35 | 2.57 |
| S3 | 0.975 6 | 0.33 | 2.41 |
| S4 | 0.956 6 | 0.43 | 3.23 |
| S5 | 0.977 9 | 0.31 | 2.30 |

表中 r^2 是决定系数, 接近 100% 表示预测浓度值接近真值。SEP 为预测标准差, 是衡量模型好坏的一个重要指标。RSD 是相对标准差, 由预测标准差除以检验集的样品均值得到, 反映的是模型对某一组分的总体预测效果。

由表 1 可知, 5 台仪器的决定系数都大于 0.95, 相对标准差也很小, 说明这几台滤光片型近红外仪器的性能能达到分析要求。但 4 号仪器和 5 号仪器相比预测标准差相差近 1/4, 相对标准差也将近差 1%, 说明这类仪器的一致性不好, 可能是滤光片的波长准确性和带宽一致性差引起的。

3.2 转换集样品数和不同模型传递算法对模型传递的影响

分别使用 10, 15, 20, 30, 40, 50, 60, 80 数目样品作为转换集, 把 1 号机作为主仪器, 2, 3, 4 和 5 号机作为从仪器, 用偏最小二乘法在主仪器上建立模型, 分别用斜率截距法、直接校正法和目标因子分析法三种模型传递算法进行校正, 结果见图 1。

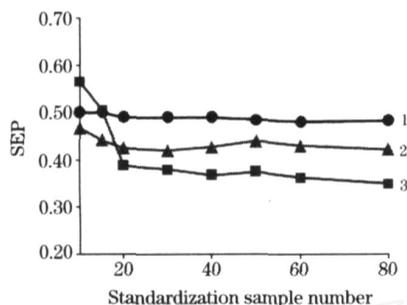


Fig 1 Average calibration transfer result at 4 instruments

1: S/B; 2: TFA; 3: DS

从图 1 可以发现, 随着转换集样品数目的增加, 三种模型传递的预测标准差呈下降趋势, 模型传递效果越来越好。当转换集样品数目达到 20 后预测标准差降低不再明显, 模型传递效果达到最佳。

图 1 还说明用不同模型传递算法, 模型传递效果有很大的差异。斜率截距法最差, 因为它是一种简单的对预测值的一元校正法, 对模型的修正是有限的, 像滤光片型仪器, 主从仪器之间差异很大, 只是简单的拟合直线来预测没有标准化的光谱, 很难获得满意的传递结果; 目标因子分析法由于去除了光谱矩阵中的多余信息, 效果稍好; 而直接校正法效果最好, 它是一种多元全光谱标准化方法, 它通过比较主从仪器的光谱数据的差异, 实现光谱数据的校正, 从而有效地提高了预测结果。

3.3 转换集样品的选择方式对模型传递的影响

把 80 个玉米样品按水分含量排序, 然后分别在 1 到 80 号、11 到 70 号、20 到 60 号中尽可能等间隔选 10 个标准样品作为转换集, 计算这 3 组转换集在三种不同模型传递算法下 4 台从仪器的平均预测标准差; 同样选 3 组 20 个标准样品和 3 组 30 个标准样品计算。结果表明, 模型传递效果与转换集样品的选择方式相关, 在 1 到 80 号中选择的标准样品好于在 20 到 60 号中选择的标准样品。为了使模型传递达到最佳效果, 应该选择各类具有代表性的样品作为转换集。

3.4 不同建模方法对模型传递的影响

用 20 个样品作为转换集, 分别用偏最小二乘法 (PLS)、多元线性回归 (MLR) 和主成分回归 (PCR) 方法在主仪器上建立模型, 再分别用斜率截距法、直接校正法和目标因子分析法进行校正, 求得 4 台从仪器平均预测标准差, 结果如表 2 所示。

表中模型传递差异度 (calibration transfer difference, CTD) 用来衡量模型传递效果。模型传递差异度反映的是用模型传递方法得到的预测标准差与用建模方法得到的预测标准差的差异程度, 由下式表示

$$CTD = (SEP_{ST} - SEP_S) / SEP_S \times 100\%$$

其中 SEP_S 为从仪器自己建立模型的预测标准差, SEP_{ST} 为模型传递后的预测标准差。模型传递差异度越接近 0%, 则模型传递的效果越好。

从表 2 明显可见, 偏最小二乘法最佳, 多元线性回归次之, 但与偏最小二乘法比较接近, 主成分回归方法很差。

Table 2 Average transfer difference under different models

| CTD | SLOPE/BIAS/ % | DS/ % | TFA/ % |
|-----|---------------|-------|--------|
| PLS | 24.89 | 7.01 | 15.3 |
| MLR | 36.64 | 8.29 | 18.58 |
| PCR | 46.72 | 24.74 | 22.91 |

3.5 不同主仪器对模型传递的影响

分别把自建模型预测效果最差的 4 号机和预测效果最好的 5 号机作为主仪器, 1, 2 和 3 号机作为从仪器, 先用偏最小二乘法在主仪器上建立模型, 再分别用斜率截距法、直接校正法和目标因子分析法进行校正, 结果如表 3 所示。

Table 3 Average transfer result when using No. 4 and No. 5 as the master

| Master | Slaves | SEP | SLOPE/BIAS | DS | TFA |
|----------|--------|------|------------|------|------|
| Master 4 | 1 | 0.35 | 0.68 | 0.4 | 0.41 |
| | 2 | 0.33 | 0.91 | 0.54 | 0.46 |
| | 3 | 0.36 | 0.42 | 0.42 | 0.36 |
| Master 5 | 1 | 0.35 | 0.51 | 0.3 | 0.46 |
| | 2 | 0.33 | 0.44 | 0.31 | 0.41 |
| | 3 | 0.36 | 0.55 | 0.4 | 0.55 |

从表 3 可以看到, 模型传递的效果受主仪器模型的影响。用预测效果最差的 4 号机作为主仪器, 则从仪器原来自建模型预测效果好的传递后变差了, 效果差的传递后反而变好了; 用预测效果最好的 5 号机作为主仪器, 则从仪器原来效果好的传递后还是好的, 效果差的传递后还是差的; 也就是原来预测效果与主仪器越相近, 则模型传递差异度越小, 模型传递效果越好。总体说来, 主仪器自建模型预测效果越好, 则模型传递效果越好, 所以应该选效果最好的仪器作为主仪器。

4 结论

滤光片型近红外仪器的波长准确性和带宽一致性不高, 因此通过斜率截距法对预测值进行一元校正对模型的修正是有限的。由于滤光片型近红外仪器的波长点较少, 目标因子分析法在提取光谱数据的主成分降低数据维数的同时减少了数据中的有效信息, 因此目标因为分析法的模型传递效果不如直接校正法稳定。采用直接校正法, 在只选用 20 个标准样品作为转换集的情况下, 平均模型传递差异度为 7.01%, 完全能够满足滤光片型近红外仪器测量的要求。

参 考 文 献

- [1] Bouveresse E, Casolino C, de la Pezuela C. *Journal of Pharmaceutical and Biomedical Analysis*, 1998, 18: 35.
- [2] Annette Behrens. *Spectrochimica Acta*, Part B, 1997, 52: 445.
- [3] Frédéric Despagne, D Luc Massart, Martin Jansen, et al. *Analytica Chimica Acta*, 2000, 406: 233.
- [4] Shen J S, Westerhaus M O. U. S. Patent, 1989, No. 4. 866. 644.
- [5] CHU Xiao-li, YUAN Hong-fu, LU Wan-zhen(褚小立, 袁洪福, 陆婉珍). *Chinese J. Anal. Chem. (分析化学)*, 2002, 30(1): 114.
- [6] WANG Yan-bin, YUAN Hong-fu, LU Wan-zhen(王艳斌, 袁洪福, 陆婉珍). *Spectroscopy and Spectral Analysis (光谱学与光谱分析)*, 2005, 25(3): 398.
- [7] ZHANG Lin, ZHANG Li-ming, LI Yan, et al(张琳, 张黎明, 李燕, 等). *Chinese Journal of Analytical Chemistry (分析化学)*, 2005, 33(12): 1709.
- [8] XIONG Yu-hong, WEN Zhi-yu, LIANG Yu-qian, et al(熊宇虹, 温志渝, 梁玉前, 等). *Spectroscopy and Spectral Analysis (光谱学与光谱分析)*, 2007, 27(1): 147.

Research on Calibration Transfer of NIR Filter Spectrophotometer

CHEN Jia-wei¹, ZHOU Chang-le^{1*}, ZHANG Ye-hui², XU Xiao-jie³, LIN Kun-hui⁴, YE Nan²

1. School of Information Science and Technology, Xiamen University, Xiamen 361005, China

2. College of Information and Electrical Engineering, China Agricultural University, Beijing 100094, China

3. Beijing Agricultural College, Beijing 102206, China

4. School of Software, Xiamen University, Xiamen 361005, China

Abstract Calibration transfer is an important issue to building up universal and comparable performance of spectrometer data in near infrared spectral analysis technology. Methods of slope/ bias correction, direct standardization (DS), and target factor analysis (TFA) were used for the calibration transfer among five NIR filter spectrophotometers using maize as the samples. The effects of three calibration transfer methods were compared. The DS method has the best performance. The average calibration transfer difference of DS is 7.01%. This study also relates to the dependence of calibration transfer on the number of standardization samples. It was proven by experiment that the results of calibration transfer will be better as the number of samples is increased and will be generally stable when there are twenty standardization samples. However, the effect of calibration transfer attained by DS is degraded sharply when the number of standardization samples decreases to be below twenty. Moreover, slope/ bias and TFA are not sensitive to the number of standardization samples.

Keywords Spectral analysis; Calibration transfer; Near infrared spectroscopy; Multivariate calibration

(Received Feb. 8, 2008; accepted Apr. 28, 2008)

* Corresponding author