

## DS 算法在近红外光谱多元校正模型传递中的应用

李庆波<sup>1</sup>, 张广军<sup>1\*</sup>, 徐可欣<sup>2</sup>, 汪 曩<sup>2</sup>

1. 北京航空航天大学仪器科学与光电工程学院, 北京 100083

2. 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** 模型传递问题是近红外光谱分析技术中一个关键的共性基础技术问题。模型传递通过在同一型号的两台仪器之间寻求一种变换关系,使得在一台仪器上建立起的模型能够应用于另外一台仪器样品光谱响应的预测。文章将直接校正(direct standardization, DS)算法应用于化学计量学多元校正中的模型传递,并研究了模型转换集样品的选择方法。在两台 AOTF 近红外光谱仪上进行了模型传递实验,首先采用 Kennard/Stone 算法选择转换集样品,然后采用 DS 方法进行模型传递。实验结果表明采用 DS 算法取得了较好的模型传递效果。DS 算法不仅可以应用于不同仪器之间的光谱分析模型传递,而且该方法对于同一仪器的长时间漂移或者由于部件的更换、测量环境的改变等引入的光谱差异校正也是适用的。

**关键词** DS 算法; 模型传递; 多元校正; 化学计量学; 光谱分析

**中图分类号:** O657.3 **文献标识码:** A **文章编号:** 1000-0593(2007)05-0873-04

### 引言

光谱仪器与化学计量学理论相结合形成的近红外光谱分析技术具有快速、高效、非破坏、样品无需预处理、无污染、无浪费等优点。近年来,该技术已广泛应用于食品工业、农业、石油化工、医药、生命科学等许多领域<sup>[1]</sup>,实现样品多种品质参数的同时测量。近红外光谱分析方法多采用多元校正方法,通过对样品光谱和其浓度等品质参数进行关联,建立校正模型,然后通过校正模型和未知样品的光谱信息来预测样品的品质参数。

在近红外光谱分析实际应用中经常遇到这样的情况,在某一仪器或者某一种测量条件下建立的多元校正模型,在另一台相同型号的仪器或者其他的测量条件下预测结果会产生较大的偏差,甚至校正模型根本无法使用。这是由于测量光谱中不仅包含了样品组分引起的光谱响应,而且还包含了仪器、测量条件的特性。当仪器的响应函数发生变化的时候,所建立的模型不再适用。这种响应函数的变化主要是由于仪器的老化、部件的更换、新仪器的使用以及测量环境的变化,比如测量地点、环境温度的变化等因素引起的。这些都是因为光谱响应中混入了没有关联进入校正模型的信息。一种解决方法是重新进行建模。这往往需要花费大量的人力和物力,所以该方法并不经济实用。另一种解决方法就是进行

模型传递。该方法是一种数学解决方法,更具有实际应用价值。模型传递问题是近红外光谱分析技术中一个关键的共性基础技术问题,该课题的研究不仅具有理论研究意义,而且具有应用价值。因为建立满足实际应用要求的多元校正模型往往花费大量人力和物力,如果所建立的模型不能在仪器之间进行有效的传递,或者模型受到严格测量条件的限制,将直接影响着这项技术的应用<sup>[2]</sup>。

目前模型传递算法可以分为两大类<sup>[3]</sup>,一类是有标样方法。这类算法必须选择一定数量的样品组成标样集(也称转换集 transfer set),并在主仪器和从仪器上分别测得其样品光谱,从而找出两台仪器之间的数学转换关系。如基于预测结果校正的斜率/偏差算法(slope/bias,简称 S/B)<sup>[4]</sup>,基于仪器所测光谱信号校正的直接校正算法(direct standardization,简称 DS)<sup>[5]</sup>、分段直接校正算法(piecewise direct standardization,简称 PDS)<sup>[6]</sup>、普鲁克算法(procrustes analysis,简称 PA)<sup>[3]</sup>等;另一类是无标样方法,使用这类算法不需要任何标准样品,主要以有限脉冲响应算法(finite impulse response,简称 FIR)<sup>[2]</sup>为代表。本文主要针对有标样方法中的 DS 算法进行研究。该方法适用于相对复杂情况下的模型传递。对于不同仪器之间的差别非常大或者测量条件不同时,相同样品测量得到的光谱之间的差异会很大而且比较复杂。这种情况下,DS 算法能够取得较好的模型传递效果。

收稿日期: 2005-12-08, 修订日期: 2006-03-01

基金项目: 航天科技创新基金项目和“863-703”专题项目资助

作者简介: 李庆波, 1975 年生, 北京航空航天大学仪器科学与光电工程学院副教授

\* 通讯联系人 e-mail: gjzhang@buaa.edu.cn

## 1 原理与算法

### 1.1 DS 模型传递方法原理

模型传递通过在两台同一型号不同仪器之间寻求一种变换关系,使得在一台仪器上建立的模型能够应用于另外一台仪器样品光谱响应的预测。在模型传递中,建立模型的仪器称为主仪器,提供需要进行传递和预测的新光谱的仪器,称之为从仪器。寻求两台仪器间的变换关系的目的是希望将从仪器上测得一个样品的光谱  $\mathbf{X}'$ ,通过一定的数学模型,同相应的主仪器上的光谱联系起来,得到转换光谱  $\mathbf{X}$ ,

$$\mathbf{X} = g(\mathbf{X}') \quad (1)$$

进行完转换后,原始的校准模型就可以用来间接预测新仪器上测得的光谱,

$$\hat{y} = f(g(\mathbf{X}')) \quad (2)$$

除了对于不同仪器之间的测量光谱进行转换以外,这种转换思想对于同一仪器的长时间漂移,或者由于部件的更换,测量环境的改变等引入的光谱差异纠正也是适用的。

DS 模型传递方法的基本思想为:通过数学方法建立主仪器和从仪器所测光谱信号之间的函数关系,由确定的函数关系变换从仪器所测光谱信号,尽可能从仪器上测定的谱图,与主仪器测定的同一样品的谱图相同,从而实现模型传递。

DS 模型传递算法<sup>[5]</sup>的流程如下:设主仪器光谱矩阵为  $\mathbf{S}_1$ ,从仪器光谱矩阵为  $\mathbf{S}_2$ ,其维数为  $m \times p$ , $m$  为样品个数, $p$  为波长个数,为了进行转换,进行矩阵运算

$$\mathbf{S}_1 = \mathbf{S}_2 \mathbf{F} + \mathbf{1} \mathbf{b}_s^T \quad (3)$$

其中  $\mathbf{F}$  ( $p \times p$ ) 为转换矩阵, $\mathbf{b}_s$  ( $p \times 1$ ) 为背景矫正矩阵。引入一个中心化矩阵  $\mathbf{C}_m$  ( $m \times m$ ),

$$\mathbf{C}_m = \mathbf{I}_m - (1/m) \mathbf{1} \mathbf{1}^T \quad (4)$$

其中, $\mathbf{I}_m$  为 ( $m \times m$ ) 阶单位矩阵, $\mathbf{1}$  为各元素均为 1 的 ( $m \times 1$ ) 列向量。在公式(3)两边同时乘以这个矩阵,得到主、从仪器上光谱矩阵各自的中性化矩阵,由于  $\mathbf{1} \mathbf{b}_s^T$  每行向量相同,所以  $\mathbf{C}_m \mathbf{1} \mathbf{b}_s^T = 0$ ,因此:

$$\bar{\mathbf{S}}_1 = \bar{\mathbf{S}}_2 \mathbf{F} \quad (5)$$

所以得到

$$\mathbf{F} = \bar{\mathbf{S}}_2^+ \bar{\mathbf{S}}_1 \quad (6)$$

其中,“+”代表广义逆矩阵,可以通过奇异值(SVD)分解计算得到, $\bar{\mathbf{S}}_1$  和  $\bar{\mathbf{S}}_2$  为中心化的光谱矩阵。另外,将公式(3)两边同时乘以  $(1/m) \mathbf{1}^T$ ,则背景矫正矩阵  $\mathbf{b}_s$  可由公式(7)计算得到

$$\mathbf{b}_s = \mathbf{s}_{1m}^T - \mathbf{F}^T \mathbf{s}_{2m}^T \quad (7)$$

式中, $\mathbf{s}_{1m}$  和  $\mathbf{s}_{2m}$  分别为由矩阵  $\mathbf{S}_1$  和  $\mathbf{S}_2$  每列元素的平均值组成的行向量。

### 1.2 模型转换集样品的选择方法

在模型传递的具体步骤上,首先需要选取转换样品集(transfer set),将转换集内的样品在主、从仪器上分别进行光谱测量,从而计算转换矩阵。转换样品集的选取应该能够同建模的样品以及需要预测的样品的覆盖范围尽可能相同,而且选择代表性强、覆盖范围广的样品。转换集样品选择的

不同对模型传递效果存在影响,研究中发现转换集在主仪器和从仪器所测信号中必须包含足够的信息来校正仪器或测量条件变化造成的偏差,其中在主仪器校正集样品中选择转换集效果最好。Kennard/Stone 算法<sup>[7]</sup>是一种有效、应用广泛的转换集选取方法,在本文中使用方法来选择转换集样品,并讨论了该转换集样品的选择方法的有效性。

Kennard/Stone 算法是通过计算样本之间的欧氏距离(Euclidean distance)来选择转换集样品,选择样本间距离大的样品,这样所选的转换集样本张成的空间较大,覆盖范围也较大。样本  $i$  和  $j$  之间的欧氏距离  $d_{ij}$  通过样品的原始光谱根据下式进行计算, $p$  为样品光谱的波长个数:

$$d_{ij} = \sqrt{\sum_{k=1}^p (r_{ik} - r_{jk})^2} \quad (8)$$

采用 Kennard/Stone 算法选择样品的步骤为:

(1) 首先选择两两样本之间距离最大的两个样本作为第一个和第二个转换集样品。

(2) 然后分别计算剩余的样本与已选样本之间的距离。

(3) 对于每个剩余样本而言,其与已选样品之间的最短距离被选择,然后选择这些最短距离中相对最长的距离所对应的样本,作为第三个转换集样品。

(4) 重复步骤(3),直至所选的转换集样品的个数等于事先确定的数目为止。

在上述的 Kennard/Stone 算法样品选择过程中,如果待选的校正样品集中存在奇异样品(Outlier),那么它必将会影响转换集样品的选择。因此在采用 Kennard/Stone 算法进行样品选择之前,应该首先对待选的样品集进行奇异点检测,剔除样品奇异点之后才可以采用 Kennard/Stone 等算法进行转换集样品选择。样品奇异点可以采用 RHM(resampling half-mean)方法<sup>[8]</sup>判断。

## 2 实验部分

### 2.1 样品准备

实验中采用葡萄糖水溶液进行模型传递验证,这主要是因为葡萄糖水溶液成份简单,而且葡萄糖在近红外的吸收区处于  $6200 \sim 6400 \text{ cm}^{-1}$  范围内,而这个区域正好在仪器的测量范围内。同时为了使光谱差异明显,配制的葡萄糖水溶液的浓度差异较大,浓度变化  $2.96 \sim 32.75 \text{ g} \cdot \text{dL}^{-1}$ ,共 54 个样品,样品葡萄糖浓度的标准偏差为  $9.13 \text{ g} \cdot \text{dL}^{-1}$ 。

样品的配制采用稀释的方法,首先配制母液,采用精密电子天平称出 329.4 g 溶质。放入容量瓶中,加去离子水至 1000 mL。然后使用移液管提取每个样品所需的母液体积,转移到容量瓶中,并使用去离子水稀释到 100 mL。

### 2.2 实验过程

为了消除其他的测量条件带来的影响,对样品池的摆放位置进行了固定,同时对光纤使用支架进行了位置的固定。由于光纤探头的位置影响较大,采用同一光纤探头进行测量。测量时,测试分两天在两台不同的仪器进行,每天首先进行仪器的重复性测试,测量时保持室内测量条件,样品池更换时,光纤探头和样品池进行擦拭和清洗,并使用少量待

测溶液浸泡探头, 保证探头没有气泡粘附, 然后进行测量。

## 3 结果与讨论

### 3.1 转换集样品的选择

首先采用 Kennard/Stone 算法选择转换集样品, 并根据所选样品来计算转换矩阵。转换集样品个数对模型传递效果存在影响。对于有标算法, 若转换集中标样数目太少, 会导致所求的转换矩阵包含的转换信息不充分; 若标样数太多, 实际应用时不方便。研究中可以通过转换集样品个数与验证集的预测标准偏差 (RMSEP) 作图来确定合适的转换集样品个数。在本例中转换集样品个数确定为 5 个, 这是由于 RMSEP 随转换集样品个数的增多而逐渐下降, 当转换集样品个数增至 5 个时, RMSEP 变化幅度趋于平稳, 所以转换集样品个数确定为 5 个。通过 K/S 算法计算, 最终选择了 11 号、12 号、45 号、47 号和 54 号样品。图 1 为 5 个转换集样品在主仪器样品集光谱得分图中的分布情况, 可见这些样品具有一定的代表性, 而且覆盖范围较广。

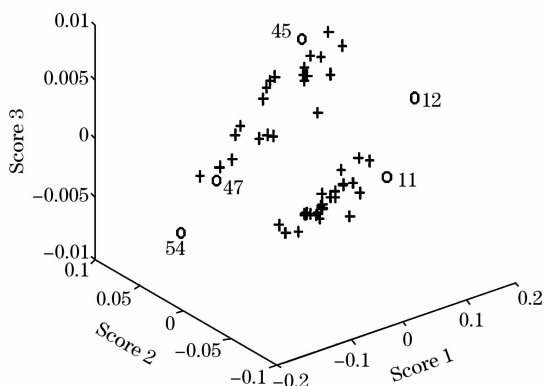


Fig. 1 Distribution of transfer set samples (marked as "O") in score plot of training set samples for master instrument

### 3.2 模型传递结果

首先观察一下在主、从仪器上测量得到的相同样品的出射光能量谱图 (见图 2), 可以看到两台仪器测量的光谱之间的差异是非常明显的。

采用主仪器在全波段上的样品光谱和葡萄糖浓度参考值建立 PLS 校正模型, 最佳主成分数为 5 个。从仪器的测量样品作为验证集进行预测。经过计算得到从仪器测量样品预测标准偏差 RMSEP 为  $515.11 \text{ g} \cdot \text{dL}^{-1}$ , 相关系数为  $-0.04$ 。可见, 主仪器上建立的校正模型根本无法适用于从仪器测量样品的预测。

下面采用 DS 方法进行模型传递。使用 DS 算法, 将两台

仪器上测得的转换集样品光谱进行转换处理, 得到了转换矩阵  $F$  和背景修正向量  $b_s$ 。根据公式  $X_{2\text{std}} = X_2 F + b_s$ , 将从仪器所测的样品光谱  $X_2$  转换为光谱矩阵  $X_{2\text{std}}$ 。该转换光谱与主仪器测定的样品光谱比较图如图 3 所示。可见, 经过 DS 传递后主从仪器的样品光谱已无明显的差异。

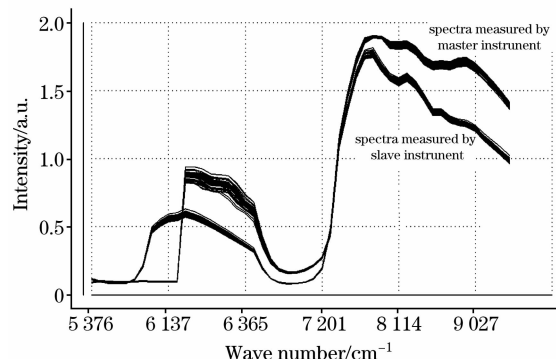


Fig. 2 Sample spectra measured by master instrument and slave instrument

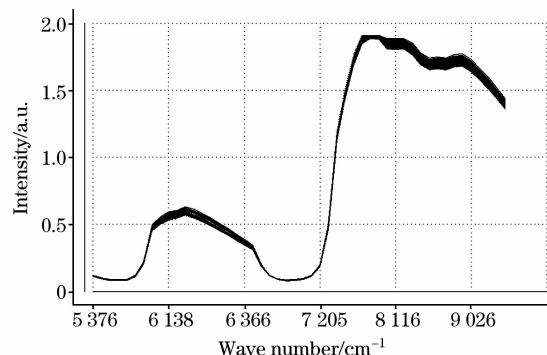


Fig. 3 Comparison of sample spectra obtained from master instrument and those from slave instrument transferred by DS algorithm

采用已建好的主仪器 PLS 校正模型来预测经 DS 算法传递后的从仪器样品光谱, 经过计算可知 RMSEP 为  $9.54 \text{ g} \cdot \text{dL}^{-1}$ , 相关系数为 0.68。可见, DS 算法对于修正仪器之间的光谱差异, 进行模型传递的效果是比较有效的。

## 4 结论

在两台 AOTF 近红外光谱仪上进行了模型传递实验, 实验结果表明采用 DS 算法取得了较好的模型传递效果。DS 算法不仅可以应用于不同仪器之间的光谱分析模型传递, 而且该方法对于同一仪器的长时间漂移, 或者由于部件的更换、测量环境的改变等引入的光谱差异校正也是适用的。

## 参 考 文 献

- [1] XU Guang-tong, YUAN Hong-fu, LU Wan-zhen(徐广通, 袁洪福, 陆婉珍). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2001, 21(4): 459.
- [2] CHU Xiao-li, YUAN Hong-fu, LU Wan-zhen(褚小立, 袁洪福, 陆婉珍). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2001, 21(6): 881.
- [3] CHU Xiao-li, YUAN Hong-fu, LU Wan-zhen(褚小立, 袁洪福, 陆婉珍). Chinese Journal of Analytical Chemistry(分析化学), 2002, 30(1): 114.
- [4] Bouveresse E, Casolino C, de la Pezuela C. Journal of Pharmaceutical and Biomedical Analysis, 1998, 18: 35.
- [5] Annette Behrens. Spectrochimica Acta, Part B, 1997, 52: 445.
- [6] Frédéric Despagne, D Luc Massart, Martin Jansen, et al. Analytica Chimica Acta, 2000, 406: 233.
- [7] Sales F, Callao M P, Rius F X. Chemometrics and Intelligent Laboratory Systems, 1997, 38: 63.
- [8] Leo H Chiang, Randy J Pell, Mary Beth Seasholtz. Journal of Process Control, 2003, 13: 437.

## Application of DS Algorithm to the Calibration Transfer in Near-Infrared Spectroscopy

LI Qing-bo<sup>1</sup>, ZHANG Guang-jun<sup>1\*</sup>, XU Ke-xin<sup>2</sup>, WANG Yan<sup>2</sup>

1. College of Instrument Science and Opto-Electronics Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100083, China
2. State Key Laboratory of Precision Measuring Technology and Instruments, Tianjin University, Tianjin 300072, China

**Abstract** Calibration transfer in multivariate calibration is one of the most important and key issues in near-infrared spectral analysis technology. The model was transferred by means of finding the transformation relation between two instruments of the same type, so that the model established on one instrument could be used on the other to predict the spectral response. In this paper, direct standardization (DS) algorithm was investigated for resolving calibration transfer in multivariate calibration of chemometrics. In addition, the selection method for the transfer set samples was studied in this paper. The experiment of model transfer was carried out between two AOTF near-infrared spectrometers. Firstly, the Kennard/Stone algorithm was adopted to select the transfer set samples, and then DS algorithm was applied to the calibration transfer. The experimental results showed that DS algorithm was effective for model transfer. DS algorithm is applicable to not only instrument standardization, but also the correction of the spectral differences induced by baseline drift for a long time or the replacement of measurement modules for the same instrument.

**Keywords** Direct standardization algorithm; Calibration transfer; Multivariate calibration; Chemometrics; Spectral analysis

(Received Dec. 8, 2005; accepted Mar. 1, 2006)

\* Corresponding author