

基于 SVR 的傅里叶变换型近红外光谱仪间数学模型传递的研究

赵龙莲^{1,2}, 李军会¹, 张文娟¹, 王建才¹, 张录达^{3*}

1. 中国农业大学信息与电气工程学院, 北京 100094
2. 清华大学医学院生物医学工程系, 北京 100084
3. 中国农业大学理学院, 北京 100094

摘要 近红外分析的一个重要基础是数学模型。不同的近红外光谱仪间由于对同一个样品响应的差异, 导致一台仪器上建立的数学模型不能直接用于另一台仪器上样品的分析, 需要进行模型传递。文章以两台傅里叶变换近红外光谱仪为实验研究对象, 以玉米粉末样品为实验材料, 采用移动窗口支持向量回归机(SVR)方法, 把一台仪器上建立的近红外定量分析数学模型传递到另一台仪器上: 当 SVR 回归的窗口大小为 31 个波长点, 传递样品个数为 15 个时, 模型传递效果较好, 以“主机”所建蛋白含量的数学模型分析“从机”上修正后的光谱, 化学测定值和近红外预测值间的相关系数提高到 0.943 4, 相对标准差为 4.23%。表明采用移动窗口 SVR 法进行傅里叶变换近红外光谱仪间数学模型的传递是可行的。

关键词 模型传递; 近红外; 支持向量回归机

中图分类号: O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2008)10-2299-05

引言

近红外分析的一个重要基础是数学模型。不同的近红外光谱仪间由于光路设计、元器件选用和装配误差, 以及外部环境不同等原因, 使得对相同样品测得的光谱数据间存在一定的差异, 导致一台仪器上建立的数学模型不能直接用于另一台仪器上样品的分析。模型传递(Calibration transfer)是解决这一问题的基本途径, 它通过修正“从机”(Slave instrument, 对未知样品扫描光谱的仪器)上的光谱, 使得这些光谱等效于是在“主机”(Master instrument, 建立模型的仪器)上扫描的, 应用“主机”上已建立的模型预测这些修正后的光谱, 即可得到未知样品所分析成分的含量^[1]。模型传递实际是通过一定数量的传递样品, 以数学方法在两台不同仪器之间寻求一种变换关系, 来增强不同仪器间光谱数据的通用性、可比性^[2]。目前, 模型传递问题已经成为近红外技术进一步推广的主要技术问题之一, 是化学计量学研究的热点课题。

国外从 20 世纪 80 年代起, 就开始有关于近红外数学模型传递的研究报道。主要方法有: 直接校正法(direct standardization, DS)^[3,4]、分段直接校正法(piecewise direct standardization, PDS)^[5]、Shenk 专利算法^[6]、FIR 法(finite impulse response, FIR)^[7]和神经网络方法等^[8]。由于不同

仪器间光谱数据的变异具有非线性特征, 加上样品集的有限性, 使得解决小样本条件下非线性关系的模型传递问题显得尤为重要^[9]。支持向量机(support vector machine, SVM)是 Vapnik 等人 20 世纪 90 年代提出的一种全新的机器学习方法, 能较好地解决小样本、非线性、高维数等实际问题^[10,11], 已在模式识别(SVC)和回归分析(SVR)方面得到了越来越多的应用^[12-17]。本文以傅里叶变换近红外光谱仪为实验研究对象, 采用移动窗口 SVR 模型传递方法, 建立算法流程, 实现了两台仪器间近红外光谱定量分析数学模型的传递。

1 实验材料与方法

1.1 实验材料与仪器

玉米粉末(过 40 目筛)样品共 90 个, 由中国农科院品种资源所提供。该样品于 1998 年 11 月在 Bruker Vector 22/N 傅里叶变换近红外光谱仪上采集漫反射光谱并建立模型, 光谱范围为 4 000~12 000 cm^{-1} , 分辨率为 8 cm^{-1} , 该仪器称为“主机”。

样品置于冰柜中冷藏保存, 剩余其中的 52 个, 2007 年 4 月在 Bruker MPA 傅里叶变换近红外光谱仪上扫描漫反射光谱, 光谱范围也为 4 000~12 000 cm^{-1} , 分辨率为 8 cm^{-1} , 该仪器称为“从机”。

收稿日期: 2007-10-08, 修订日期: 2008-01-08

基金项目: 国家自然科学基金项目(30370915, 60571027)资助

作者简介: 赵龙莲, 女, 1972 年生, 中国农业大学信息与电气工程学院讲师

e-mail: caunir@cau.edu.cn

* 通讯联系人

“主机”和“从机”属于同一公司两个型号的傅里叶变换红外光谱仪，两台仪器的结构设计、元器件选用等都有所不同。

1.2 模型传递方法原理

1.2.1 模型传递中样品集的构成

数学模型传递过程中需要有 3 组样品集：校正样品集 (Calibration sample set)，传递样品集 (Transfer sample set) 和预测样品集 (Prediction sample set)。校正样品集用于在“主机”上建立模型；传递样品集需分别在“主机”、“从机”上扫描其光谱，以计算模型传递参数；预测样品集用于检验模型传递的效果。3 个样品集的分布示意图 1， N 个校正样品用来建立模型， L 个传递样品和 P 个预测样品是“主机”和“从机”两台仪器上共有的。

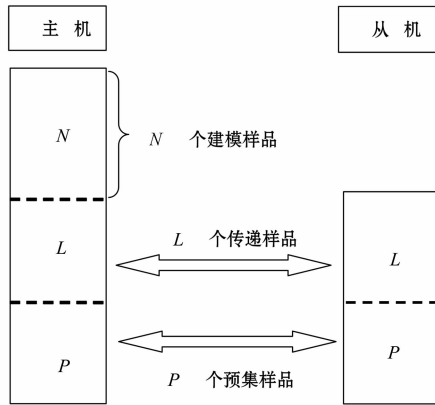


Fig. 1 Sample sets in calibration transfer

1.2.2 基于 SVR 的模型传递方法

对于实际的光谱数据，波长点的漂移通常只局限于一个小的区域。因此，“主机”上的每个波长点只是与“从机”上相应波长点附近的波长关联大，而并非与全谱区的光谱点都关联。为此，模型传递时我们不是采用全谱区的 SVM 回归，而是对一小段窗口 (window) 内的数据进行 SVM 回归以修正相应光谱点的吸光度，然后移动窗口求出全光谱的修正参数，该方法我们称之为移动窗口式 SVR 模型传递方法^[18]。

设 L 为传递样品的个数， M 为光谱的点数， $\mathbf{A}(L \times M)$ 表示“主机”上传递样品集的吸光度矩阵， $\mathbf{B}(L \times M)$ 表示“从机”上传递样品集的吸光度矩阵。

(1) 对光谱点 $i(i = 1, 2, \dots, M)$ ，令 \mathbf{X}_i 表示“从机”上从 $i-k$ 到 $i+k$ 共 $2k+1$ 个点这一窗口 (Window = $2k+1$) 内的吸光度矩阵

$$\mathbf{X}_i = [\mathbf{B}_{i-k}, \mathbf{B}_{i-k+1}, \dots, \mathbf{B}_{i+k-1}, \mathbf{B}_{i+k}]$$

对“主机”上的列向量 $\mathbf{Y}_i = [\mathbf{A}_i]$ ，利用 SVM 法建立 \mathbf{X}_i 与 \mathbf{Y}_i 间的回归模型，求出波长点 i 处“主机”与“从机”间光谱的修正参数 $\beta_i, bias_i$ 。

(2) 移动 i 和“从机”上相应窗口，求出每一个波长点的传递参数，构成传递参数矩阵。

具体算法过程示意如图 2。

传递参数矩阵计算得到后，就可以将“从机”上未知样品的光谱转化成与“主机”光谱相匹配的光谱了，从而可以用

“主机”上的模型进行预测分析。

1.3 模型传递效果评价指标

模型传递的效果可以用预测集的近红外预测值与化学测定值之间的相关系数 r 以及预测的相对标准差 RSD% 两个指标共同评价^[1]。

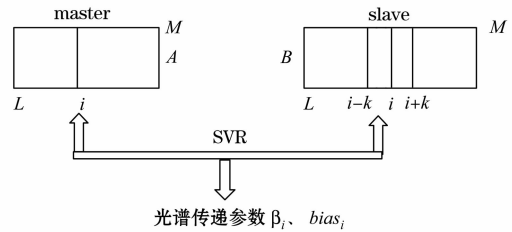


Fig. 2 Computation of the transfer parameters in SVR procedure

2 结果与分析

2.1 模型结果

本文在研究模型传递时玉米粉末样品总计 90 个，考虑到尽量增加建模样品个数，90 个样品除 25 个用做预测外，剩余 65 个全部用于在“主机”上建模，用于研究模型传递方法的 27 个传递样品亦取自于建模样品之中，即建模样品的个数为 65 个，其中 27 个用于传递样品。

用 PLS 法对“主机”上 65 个建模样品建立模型，模型近红外预测值和化学值之间的相关系数以及相对标准差见表 1。用该模型分别预测“主机”和“从机”上未经传递的预测样品集 (25 个)，并计算相关系数和相对标准差，结果见表 1。

Table 1 Results of cross-validation for master and prediction for master and slave before transfer

	模型结果	主机上预测集的结果	模型传递前从机预测集的结果
r	0.964 6	0.968 7	0.881 7
RSD/%	3.61	2.89	6.81

由表 1 可见，利用“主机”上建立的模型分析“主机”上的预测样品集效果很好，相关系数为 0.968 7，相对标准差为 2.89%，而用这一模型直接分析“从机”上的预测样品集效果则较差，相关系数变小为 0.881 7，相对标准差变大为 6.81%，可见“主机”上的近红外数学模型不再适合于直接预测“从机”上的样品光谱，需要进行模型传递。

2.2 SVR 回归窗口大小的影响

选择 15 个传递样品，对两台仪器上传递样品集的光谱进行 SVR 回归 (poly 核函数)，算法流程见图 2，求出“主机”和“从机”光谱间的修正参数。根据得到的参数再修正“从机”上预测样品集的光谱，然后用“主机”上的模型预测修正后的光谱，SVR 回归的窗口分别选择 11, 21, 31, 41, 51 个波长点时预测结果见表 2。

由表 2 可见，采用移动窗口 SVR 法进行模型传递 (传递样品为 15 个) 时，回归窗口的大小对模型传递效果有影响，

窗口太小(11 和 21 个波长点)和太大(41 和 51 个波长点)时传递效果较差。当回归窗口取 31 个波长点时效果较好, 相关系数达 0.943 4, 相对标准差为 4.23%。

Table 2 Transfer results of different window size with 15 transfer samples

窗口小大(波长点个数)	r	RSD/%
11	0.898 7	14.28
21	0.903 6	8.39
31	0.943 4	4.23
41	0.849 0	8.22
51	0.823 8	7.87

2.3 传递样品个数对模型传递效果的影响

传递样品个数对模型传递的效果也有影响。SVR 回归(窗口为 31 个波长点)时, 传递样品的个数分别为 5, 10, 15, 20, 25 和 27 个时模型传递后预测检验样品集的结果见表 3。

Table 3 Transfer results of different transfer samples with 31 window size

传递样品个数	r	RSD/%
5	0.864 5	19.28
10	0.908 0	12.81
15	0.943 4	4.23
20	0.960 4	3.99
25	0.956 0	3.69
27	0.953 7	3.82

由表 3 可见, 模型传递时传递样品的个数由 5 个逐渐增加到 15 个, 再到 20 个时, 传递效果由差变好, 考虑到尽量减少传递样品的个数, 本文选择 15 个。由此可见, 只有当传递样品的个数达到一定的数目时, 才能充分修正“主机”和“从机”间的光谱差异, 样品太少不足以修正这些差别。

由以上结果可知, 采用移动窗口 SVR 法在两台傅里叶变换近红外光谱仪上进行模型传递, 当窗口大小为 31 个波长点, 传递样品的个数达 15 个时, 模型传递效果较好, 相关系数为 0.943 4, 相对标准差为 4.23%。其中一个样品在“主机”上的光谱和“从机”修正前后的光谱见图 3, 由图可见“从机”上的光谱经修正后与“主机”上的光谱基本重叠; 图 4 是“从机”上预测样品集在模型传递前后的散点图, 相关系数由传递前的 0.881 7 变为 0.943 4, 相对标准差由传递前的 6.81% 变为 4.23%, 也表明模型传递效果较好。因此采用移动窗口 SVR 法进行傅里叶变换近红外光谱仪间数学模型的传递是可行的。

3 结论与讨论

本文以两台傅里叶变换近红外光谱仪为实验研究对象, 采用移动窗口支持向量回归机(SVR)方法, 把一台仪器上的

玉米粉末近红外数学模型传递到另一台仪器上, 取得了较满意的结果, 由此表明采用移动窗口 SVR 法进行傅里叶变换近红外光谱仪间数学模型的传递是可行的。

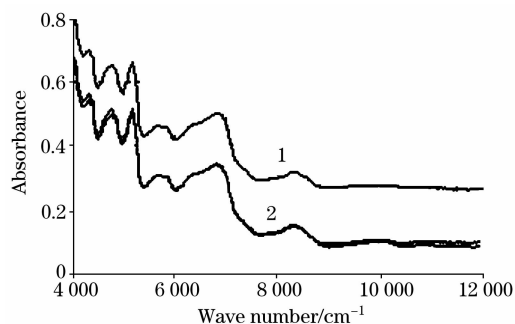


Fig. 3 Spectra of the same sample on the master and slave before and after transfer

1: Spectrum of slave before transfer;

2: Spectrum of master and slave after transfer

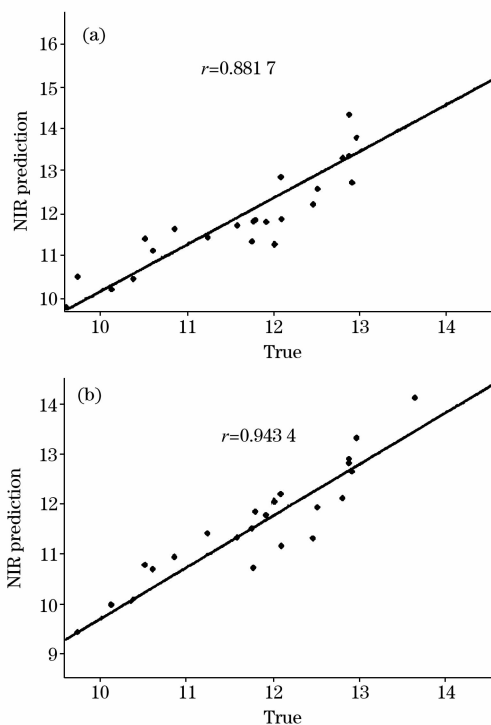


Fig. 4 Plot of NIR prediction vs. true of prediction sample set before(a) and after(b) transfer

模型传递中, 传递样品集用来计算传递参数, 需要分别在“主、从”两台仪器上扫描光谱。如果传递样品集的样品来源于建模样品集, 就需要保存样品, 而样品长时间保存是非常困难的, 在保存的过程中很可能发生变质。本文在研究模型传递时考虑到建模样品个数较少, 又要保证传递样品的个数, 因此部分传递样品取自于建模样品集。为了应用的方便, 最大程度不保存样品, 可以使传递样品和建模样品集不重合, 只要将传递样品同时在“主机”和“从机”上扫描光谱计算传递参数进行模型传递即可。

参 考 文 献

- [1] YAN Yan-lu, ZHAO Long-lian, HAN Dong-hai, et al(严衍禄, 赵龙莲, 韩东海, 等). Foundation of NIR Spectral Analysis and Its Application(近红外光谱分析基础与应用). Beijing: China Light Industry Press(北京: 中国轻工业出版社), 2005.
- [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] de Noord O E. Chemometrics and Intelligent Laboratory Systems, 1994, 25: 85.
- [4] LI Qing-bo, ZHANG Guang-jun, XU Ke-xin, et al(李庆波, 张广军, 徐可欣, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2007, 27(5): 873.
- [5] Wang Yongdong, Velamp David J, Kowalski Bruce R. Anal. Chem., 1991, 63: 2750.
- [6] Shenk J S, Westerhaus M O. U. S. Patent.: 4866644, Sept. 12, 1989.
- [7] Blank Thomas B, Brown Steven D. Anal. Chem., 1996, 68(17): 2987.
- [8] Goodacre Royston, Timmins Eadaoin M, Jones Alun, et al. Analytica Chimica Acta, 1997, 348(1-3): 511.
- [9] 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.
- [10] Vapnik V N. The Nature of Statistical Learning Theory. New York: Springer, 1995.
- [11] BIAN Zhao-qi, ZHANG Xue-gong(边肇祺, 张学工). Pattern Recognition(模式识别)(2nd ed.). Beijing: Tsinghua University Press(北京: 清华大学出版社), 2000.
- [12] CHEN Quan-sheng, ZHAO Jie-wen, ZHANG Hai-dong, et al(陈全胜, 赵杰文, 张海东, 等). Acta Optica Sinica(光学学报), 2006, 26(6): 933.
- [13] ZHANG Lu-da, JIN Ze-chen, SHEN Xiao-nan, et al(张录达, 金泽宸, 沈晓南, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2005, 25(9): 1400.
- [14] LIU Shu-hua, ZHANG Xue-gong, ZHOU Qun, et al(刘沐华, 张学工, 周群, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2006, 26(4): 629.
- [15] BAI Peng, XIE Wen-jun, LIU Jun-hua(白鹏, 谢文俊, 刘君华). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2007, 27(7): 1323.
- [16] WU Jing-zhu, WANG Yi-ming, ZHANG Xiao-chao, et al(吴静珠, 王一鸣, 张小超, 等). Journal of Agricultural Mechanization Research(农机化研究), 2007, (1): 155.
- [17] LIN Ji-peng, LIU Jun-hua(林继鹏, 刘君华). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2006, 26(12): 2232.
- [18] Alamar M Carmen, Bobelyn Els, Lammertyn Jeroen, et al. Postharvest Biology and Technology, 2007, 45: 38.

Calibration Transfer between Two FTNIR Spectrophotometers Using SVR

ZHAO Long-lian^{1, 2}, LI Jun-hui¹, ZHANG Wen-juan¹, WANG Jian-cai¹, ZHANG Lu-da^{3*}

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

2. Department of Biomedical Engineering, Medical School of Tsinghua University, Beijing 100084, China

3. College of Science, China Agricultural University, Beijing 100094, China

Abstract In the present research, a set of maize powder samples was used to study the calibration transfer between two fourier transform near-infrared (FTNIR) spectrophotometers, and a method of moving window support vector regression machines (SVR) was used to correct the differences between the two instruments. Bruker Vector 22/N was referred to as “master” on which the maize protein calibration model was built. Bruker MPA was referred to as “slave” instrument. A transformation matrix was constructed based on the spectra of a sample set (for calibration transfer) measured on both instruments. After transfer, NIR spectra acquired on “slave” will appear as if they were measured on master instrument. The calibration model available for the master can then be used to predict the transformed spectra measured on the slave. The transfer parameters were computed as follows. For wavelength i , the absorbance vector obtained on the master instrument was regressed against the corresponding absorbance matrix of a spectral window obtained on the slave instrument. Method of SVR was used for regression. Moving the wavelength i and corresponding window, the transfer parameter for each wavelength can be obtained. For the two FTNIR spectrophotometers, a window size of 31 wavelengths and a subset of 15 transfer samples were chosen to establish the SVR regression model between “master” and “slave”. Applying the calibration model to the prediction samples after being corrected by

the transfer parameters, a good transfer performance can be achieved. The correlation coefficient (r) is 0.943 4, while the relative standard deviation(RSD) is 4.23%. These results suggest that the SVR method can be used to successfully transfer the calibration model for protein of maize developed on a FTNIR spectrophotometer to another.

Keywords Calibration transfer; Near-infrared (NIR); Support vector regression machines(SVR)

(Received Oct. 8, 2007; accepted Jan. 8, 2008)

* Corresponding author

2008 第 3 届亚太地区冬季等离子体光谱化学会议 2008 Third Asia-Pacific Winter Conference on Plasma Spectrochemistry (2008 APWC)

由日本化学会和日本分析化学学会主办的第 3 届亚太地区冬季等离子体光谱会议(2008 APWC)将于 2008 年 11 月 16~21 日在日本筑波国际会议中心举行。

征文范围:

1. ICP-MS 和 ICP-OES 的仪器与原理; 2. 样品的制备和导入; 3. ICP-MS 的碰撞反应池技术与应用; 4. 新仪器与等离子体源; 5. 辉光放电光源; 6. 环境、金属组学和生命科学中的元素形态; 7. ICP-MS 和 ICP-OES 的应用, 包括: 冶金、RoHS、高纯物质及半导体材料、核工业材料、环境、地质、海洋科学、食品、农业、生物学、临床和药物、质量控制和标准化等; 8. 同位素比和同位素稀释; 9. 激光烧蚀进样。

会议地点:

筑波国际会议中心(Tsukuba International Congress Center) (日本茨城县筑波市(Tsukuba, Ibaraki, Japan))

联系人: Prof. Naoki Furuta

Chuo University, Faculty of Science and Engineering, Department of Applied Chemistry, 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan

Tel: +81-3-3817-1906, Fax: +81-3-3817-1699

e-mail: nfuruta@chem.chuo-u.ac.jp

会议使用英语交流

会议详细信息请登录网站查询 <http://envsun.chem.chuo-u.ac.jp/plasma/2008apwc>