

偏最小二乘法方法在光谱定性分析中的应用研究

褚小立 许育鹏 陆婉珍

(石油化工科学研究院 北京 100083)

摘要 传统的光谱多元定性方法大都基于主成分分析(PCA),即首先将光谱进行PCA降维并选取特征变量后,再进行聚类或判别分析。近年来,用于光谱多元定量校正的偏最小二乘法(PLS)也被越来越多地用于定性分析中,并得到更优的结果。本文以实例介绍PLS方法在光谱模式识别以及建立定量模型适用性判据等方面的应用,其有望成为一种常用的光谱定性方法。

关键词 偏最小二乘法 化学计量学 主成分分析 模式识别 光谱定性分析

偏最小二乘法(PLS)是光谱多元定量校正最常用的一种方法,已被广泛应用于近红外、红外、拉曼、核磁和质谱等波谱定量模型的建立,几乎成为光谱分析中建立线性定量校正模型的通用方法^[1,2]。

近年来,随着PLS方法在光谱分析尤其是分子光谱如近红外、红外和拉曼中应用的深入开展,PLS方法还被用来解决模式识别、定量校正模型适用性判断以及异常样本检测等定性分析问题。由于PLS方法同时从光谱阵和浓度阵中提取载荷和得分,克服主成分分析(PCA)方法没有利用浓度阵的缺点,可有效降维,并消除光谱间可能存在的复共线关系,因此取得令人非常满意的定性分析结果^[3~5]。本文主要介绍PLS方法在光谱定性分析方面的原理及应用实例。

1 PLS方法在光谱模式识别中的应用

基于光谱的模式识别用途很广泛,主要有两种情况,一是无监督的聚类分析,常用的方法是将一组待分析样本的光谱进行PCA降维,选取特征主成分得分作为输入变量,通过聚类方法,或者直接作图样本在主成分空间的分布图来确定样本之间的亲疏关系^[6],如中草药的道地性分析等;二是有监督的模式识别,即已知样本的类别,如不同牌号的成品汽油,通过光谱和模式识别方法实时判别远程输油管道中的成品汽油类型。这类方法大都先对样本的光谱阵进行PCA降维,选取特征变量后再用线性判别分析(LDA)或SIMCA等方法来识别未知样本^[7]。这些基于PCA的模式识别方法通常可以解决大多数的光谱定性分析问题。但是对于类别间光谱差异非常小的样本,这些方法则往往得不到满意的结果。

PLS方法本质上是一种基于特征变量的回归方

法,但若将已知类别样本的浓度阵分别设为0、1(对两类用PLS1方法而言),或-1、0、+1(对三类用PLS2方法而言),或0 1、1 0(对两类用PLS2方法而言),或0 0 1、0 1 0、1 0 0(对三类用PLS1方法而言),多类问题依此类推,则PLS方法可用于有监督的模式识别光谱定性分析,通常称为伪PLS回归(Dummy partial least-squares regression, DPLS)^[8]。与定量校正类似,由于PLS方法同时对光谱阵和类别阵进行分解,加强类别信息在光谱分解时的作用,以提取出与样本类别最相关的光谱信息,即最大化提取不同类别光谱之间的差异,因此PLS方法通常可以得到比PCA方法更优的分类和判别结果。目前将PLS用于模式识别受到越来越多的关注和应用,许多研究也证明其判别结果要优于基于PCA的模式识别方法^[9,10]。下面举一个利用PLS方法结合红外光谱识别渣油类型的实例进行说明^[11]。

根据原油加工工艺不同,有常压渣油、减压渣油和加氢渣油等多种类型,由于不同类型渣油的组成有较大差异,很难通过线性定量校正算法得到光谱与性质之间精确的数学关系,须对同类样本单独建立校正模型。因此有必要建立识别不同类型渣油的定性模型,对未知样本首先判断其类别,然后调用相应的定量校正模型进行预测分析。这也是模式识别在光谱结合化学计量学定量分析中的一个重要应用方面^[12]。

收集472个渣油样本,其中常压渣油105个样本(校正集70个、验证集35个)、减压渣油98个样本(校正集68个、验证集30个)、加氢渣油269个样本(校正集164个、验证集105个)。通过ATR方式在相同的测量条件下(波数范围400~4000 cm^{-1} 、分辨率4 cm^{-1} 、扫描次数64次、样本恒定温度60 $^{\circ}\text{C}$)采

集所有样本的红外光谱。在进行定性分析前,选取信息较强的 $1850 \sim 780\text{cm}^{-1}$ 光谱范围,并经过二阶微分处理。

首先对校正集 302 个渣油光谱进行主成分分析,图 1 给出前 3 个主成分得分的三维图(解释 84.2% 光谱变量),可以看出,由于常压渣油和减压渣油在组成上极为相近,这两类样本存在显著的重叠。尝试从前 8 个主成分中选取不同得分变量结合 SIMCA 方法均未得到满意结果,说明 PCA 方法不能将类别间存在的微弱信息差异提取出来。

为此采用 PLS 方法对三类渣油的红外光谱进行分类和判别分析,将参与伪回归的 302 个常压渣油、减压渣油和加氢渣油样本的类别值分别为 0、-1 和 1,采用留一交互验证方法确定最佳主因子数为 6。图 2 给出交互验证的结果,其中预测值在 $-1.5 \sim -0.5$ 范围的为常压渣油,预测值在 $-0.5 \sim 0.5$ 范围的为减压渣油,预测值在 $0.5 \sim 1.5$ 范围的为加氢渣油。图 3 给出验证集 170 个样本的预测结果。从图 2 和图 3 可以看出,通过 PLS 方法三种渣油得到 100% 成功分类和判别。图 4 给出 PCA 和 PLS 对校正集光谱分解得到的第一主成分载荷图,对比可以看出,两者之间存在显著差异,PLS 的载荷信息更为丰富,说明由于 PLS 方法在对光谱阵进行分解时考虑类别因素,最大化提取出因类别引起的光谱间的差异,所以得到优于 PCA 方法的分类和判别结果。

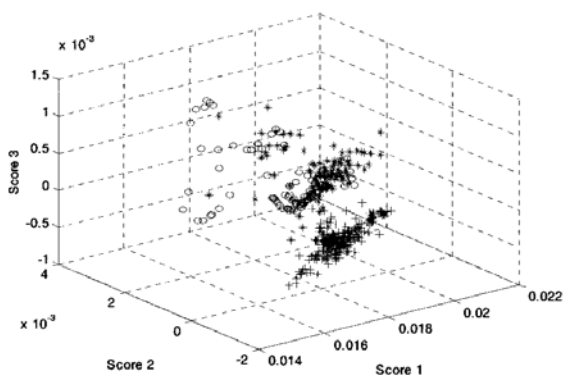


图 1 302 个渣油校正样本光谱进行主成分分析得到的前三个主成分得分图
0 常压渣油; * 减压渣油; + 加氢渣油

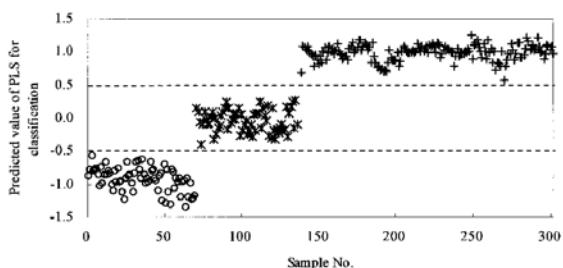


图 2 302 个渣油校正样本 PLS1 交互验证得到的分类结果
0 常压渣油; * 减压渣油; + 加氢渣油

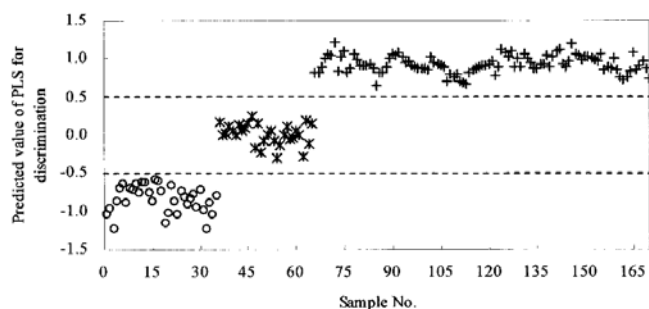


图 3 170 个渣油样本 PLS1 的判别分析结果

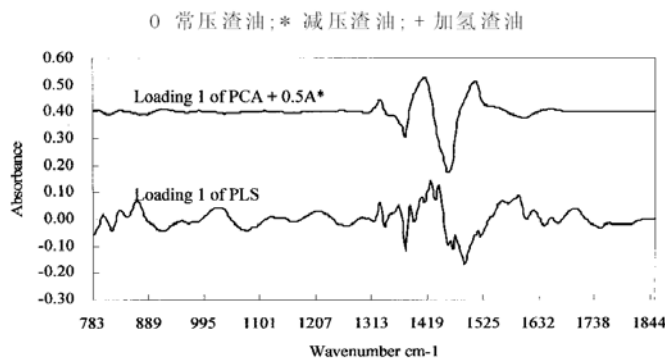


图 4 PLS 和 PCA 所得第一主成分载荷的对比图

0 常压渣油; * 减压渣油; + 加氢渣油

2 PLS 方法用于光谱定量模型适用性判据的建立

对于光谱结合多元校正的定量分析方法而言,由于不可能建立一个覆盖所有未知样品的校正模型,尤其对于天然产品如农产品和石油等,建立模型的适用性判据尤为重要和必要,也称为校正模型界外样本检测。在对未知样本进行预测分析时,只有待测样品在模型覆盖的范围之内,才能保证分析结果的有效性和准确性^[13]。

通常通过三个判据来保证定量校正模型的适用性^[14],一是马氏距离,如果待测样品的马氏距离大于校正集样品的最大马氏距离,则说明待测样品中的一些组分浓度超出校正集样品组分浓度的范围;二是光谱残差,如果待测样品的光谱残差大于规定的阈值,则说明待测样品中的含有校正集样品中所没有的组分;三是最邻近距离,如果待测样品与所有校正集样本之间距离的最小值(最邻近距离)大于规定的阈值,则说明待测样品落入校正集分布比较稀疏的地方,预测结果的准确性将受到置疑。

目前以上这三个判据大都基于 PCA 方法得到^[15],即首先将校正集的光谱进行主成分分析,选择合适的主因子数后,通过每个校正样本的得分向

量分别计算马氏距离和最邻近距离阈值,通过校正样本光谱PCA分解后的光谱残差计算出光谱残差阈值,显然对于一个校正集只得到一组校正模型适用性判别阈值。这种计算方法存在的问题是,对于一个包含多种性质的样品集,如果某些校正样本会缺少某一种或多种性质的基础数据,这时基于整个校正集样本光谱和PCA得到的阈值便不适合这些缺少基础数据的性质,显然会存在扩大模型适应范围的风险。但如果使用PLS1方法对每种性质校正回归时得到的得分向量和光谱残差,分别计算每一种性质的三个判据阈值,上述问题就迎刃而解,下面举例进行说明。

对于一个由350个汽油样本的近红外光谱和6种性质数据(研究法辛烷值、马达法辛烷值、烯烃、芳烃、苯和氧含量)组成的校正集,其中有研究法辛烷值数据的样本为335个,有马达法辛烷值数据的样本仅为85个,有氧含量数据为250个,有苯含量数据也仅为120个,有烯烃和芳烃数据的样本为315个。对于这个校正集,如果用所有350个校正样本基于PCA方法建立模型适用性判据阈值,显然对所有性质的定量校正模型适应范围都存在风险,尤其是马达法辛烷值和苯含量。因为实际参与PLS1回归的样本远小于350,而且有些样本的基础数据在回归过程中因偏差较大还有可能被剔除。这时可以采用每一性质进行PLS1回归时得到的主成分得分和载荷来分别计算对应性质的模型适用性判据阈值,对于未知样本的模型界外判别也是针对每一性质逐个进行的,尽管这种方法要比PCA方法略显繁琐,但得到的判据更为合理和可靠。这种方式有望成为商品化学计量学光谱分析软件中建立定量模型适用性判据的一种可选方法。

3 结束语

由于PLS算法的特点,将其用于光谱定性分析受到越来越多的关注,并逐步得到实际应用,尤其是在有监督的判别分析以及建立定量模型适用性判据等方面表现出优于PCA方法的性能。此外,利用PLS回归得到的得分向量,结合马氏距离还可用于校正集奇异样本的筛选等。可以相信,PLS方法有望成为解决光谱(如红外、近红外和拉曼)多元定性分析中一种常用的可选方法。

参考文献

1 Hrdko E, Frank, John H, Kalivas, Bruce R, Kowalski Partial least squares solutions for multicomponent analysis (J)

- Anal Chem, 1983, 55 (11): 1800 ~ 1804
- 2 Richard G Breton Chemometrics: data analysis for the laboratory and chemical plant (M) Wiley, 2003
- 3 Ortiz C, Zhang D, Xie Y, Davison V J, Amotz DB Identification of insulin variants using Raman spectroscopy (J) Anal Biochem, 2004, 332 (2): 245 ~ 252
- 4 Cozzolino D, Chiree A, Scaife J R, Murray I Usefulness of Near-Infrared Reflectance (NIR) Spectroscopy and Chemometrics To Discriminate Fishmeal Batches Made with Different Fish Species (J) J Agric Food Chem, 2005, 53 (11): 4459 ~ 4463
- 5 Lutz U, Lutz R W, Lutz W K Metabolic Profiling of Glucuronides in Human Urine by LC-MS/MS and Partial Least-Squares Discriminant Analysis for Classification and Prediction of Gender (J) Anal Chem, 2006, 78 (13): 4564 ~ 4571
- 6 褚小立, 袁洪福, 陆婉珍 光谱结合主成分分析和模糊聚类方法的样品聚类与识别 (J) 分析化学, 2000, 28 (4): 421 ~ 427
- 7 Young-ah Woo, Chang-hee Cho, Hyo-jin Kim, Jong-soo Yang, Ki-yong Seong Classification of cultivation area of ginseng by near infrared spectroscopy and ICP-AES (J) Microchemical Journal, 2002, 73 (3): 299 ~ 306
- 8 Cozzolino D, Murray I Identification of animal meat muscles by visible and near infrared reflectance spectroscopy (J) Lebensmittel-Wissenschaft und -Technologie, 2004, 37 (4): 447 ~ 452
- 9 Oliver P, Wehlan, Mark E, Earll, Erik Johansson, Marianne Toft, Lennart Eriksson Detection of ovarian cancer using chemometric analysis of proteomic profiles (J) Chemometrics and Intelligent Laboratory Systems, 2006, 84 (1 ~ 2): 82 ~ 87
- 10 陆婉珍, 袁洪福, 褚小立编 当代中国近红外光谱技术 (M), 北京: 中国石化出版社, 2006
- 11 Yuan Hongfu, Chu Xiaoli, Li Haoran, Xu Yupeng Determination of multi-properties of residual oils using mid-infrared attenuated total reflection spectroscopy (J) Fuel, 2006, 80 (12 ~ 13): 1720 ~ 1728
- 12 Andrés J M, Bona M T ASTM clustering for improving coal analysis by near-infrared spectroscopy (J) Talanta, 2006, 70 (4): 711 ~ 719
- 13 D Jouan-Rimbaud, E Bouveresse, D L Massart and O E de Noord Detection of prediction outliers and inliers in multivariate calibration (J) Analytica Chimica Acta, 1999, 388 (3): 283 ~ 301
- 14 ASTM E1655 Standard Practices for Infrared Multivariate Quantitative Analysis (S)
- 15 陆婉珍主编 现代近红外光谱分析技术(第二版) (M), 北京: 中国石化出版社, 2006

- mass spectrometry (J) *Journal of Chromatography A*, 2001, 939: 49 ~58
- 5 David N Heller, Maureen A Ngohl, Dan Donoghue, et al Identification of incurred sulfonamide residues in eggs: methods for confirmation by liquid chromatography tandem mass spectrometry and quantitation by liquid chromatography with ultraviolet detection (J) *Journal of Chromatography B*, 2002, 774: 39 ~52
- 6 Roman Hirsch, Thomas A Temes, Klaus Haberera, et al Determination of antibiotics in different water compartments via liquid chromatography electrospray tandem mass spectrometry (J) *Journal of Chromatography A*, 1998, 815: 213 ~223
- 7 Bing Shao a, Dan Donga, Yongning Wu, et al Simultaneous determination of 17 sulfonamide residues in porcine meat, kidney and liver by solid-phase extraction and liquid chromatography - tandem mass spectrometry (J) *Analytica Chimica Acta*, 2005, 546: 174 ~181
- 8 Jungju Seo a, Hye-Young Kim a, Bong Chul Chungb, et al Simultaneous determination of anabolic steroids and synthetic hormones in meat by freezing-lipid filtration, solid-phase extraction and gas chromatography - mass spectrometry (J) *Journal of Chromatography A*, 2005, 1067: 303 ~309
- 9 George Stubbings, Jonathan Tarbin, Andrew Cooper, et al A multi-residue cation-exchange clean up procedure for basic drugs in produce of animal origin (J) *Analytica Chimica Acta*, 2005, 547: 262 ~268
- 10 Anne Marie Jacobsen, Bent Halling-Sørensen, Flemming Ingerslev, et al Simultaneous extraction of tetracycline, macrolide and sulfonamide antibiotics from agricultural soils using pressurised liquid extraction, followed by solid-phase extraction and liquid chromatography - tandem mass spectrometry (J) *Journal of Chromatography A*, 2004, 1038: 157 ~170
- 11 Shinwoo Yang, Jongmun Cha, Kenneth Carlson Simultaneous extraction and analysis of 11 tetracycline and sulfonamide antibiotics in influent and effluent domestic wastewater by solid-phase extraction and liquid chromatography electrospray ionization tandem mass spectrometry (J) *Journal of Chromatography A*, 2005, 1097: 40 ~53
- 12 E Benito-Pena, A I Parta-Rodera, M E Leon-Gonzalez, et al Evaluation of mixed mode solid phase extraction cartridges for the preconcentration of beta-lactam antibiotics in wastewater using liquid chromatography with UV-DAD detection (J) *Analytica Chimica Acta*, 2006, 556: 415 ~422
- 13 Jose Blanca, Patricia Muñoz, Miguel Morgado, et al Determination of clenbuterol, ractopamine and zilpaterol in liver and urine by liquid chromatography tandem mass spectrometry (J) *Analytica Chimica Acta*, 2005, 529: 199 ~205

The application of solid-phase extraction in analysis of veterinary drugs residue

Zhu Weixia¹ Yang Jizhou² Zhang Shusheng¹ Liang Wei²

(1 Chemistry Department of Zhengzhou University Zhengzhou 450001)

(2 Henan Entry - Exit Inspection & Quarantine Bureau, Zhengzhou 450003)

Abstract Solid-phase extraction (SPE) has extensively applied in determination of veterinary residue according to different separation mechanism, the paper described the usage of SPE in sample pretreatment. At the same time, it generalized how to select the SPE cartridges for compounds of different chemical property.

key words Solid-phase extraction Residue Separation mode

(下接第 15 页)

The study of use of partial least squares in spectroscopy qualitative analysis

Chu Xiaoli Xu Yupeng Lu Wanzhen

(Research Institute of Petroleum Processing, Beijing 100083, China)

Abstract In this paper, the applications of partial least squares method in spectroscopy multivariate qualitative analysis, such as discriminant analysis and outliers detection of quantitative calibration models, were introduced based on several examples. Compared with the traditional principal component analysis method, partial least squares discriminant analysis method usually gives better results.

Key words Partial least squares Chemometrics Principal component analysis Pattern recognition Spectroscopy analysis