文章编号:1001-5078(2008)04-0342-03

·红外技术·

近红外光谱技术鉴别连翘产地

张晓慧,刘建学
(河南科技大学食品与生物工程学院,河南 洛阳 471003)

摘 要:采用近红外光谱结合SIMCA模式识别方法鉴别五个不同产地的连翘。研究结果表明,在6700~5300cm^{-1}波数范围内的光谱,通过SNV预处理方法后,利用SIMCA的模式识别方法分别为卢氏、栾川、洛宁、新安和山西安泽等五个产地连翘建立了模型。交叉验证的最佳主成分数分别为3,2,2,3,4。在$\alpha=5\%$的显著水平下,预测集的10个样品中只有1个被错判,表明该方法具有良好的鉴别分类功能。

关键词:连翘;近红外光谱;SIMCA;模式识别
中图分类号:O657.33 **文献标识码**:A

Identification of Forsythia Suspense from Different Habitats by NIR Spectra

ZHANG Xiao-hui, LIU Jian-xue
(Food & Bioengineering Department, Henan University of Science & Technology, Luoyang 471003, China)

Abstract: Identification method of forsythia suspense from 5 different habitats by near infrared spectroscopy coupled with pattern recognition based on SIMCA was proposed in this paper. In the spectra region between 6700cm^{-1} and 5300cm^{-1}, 5 predictive models of Lushi, Luanchuan, Luoning, Xin'an in Henan province, and Anzhe in Shanxi province were built separately by the standard normal variate (SNV) preprocessing method with SIMCA pattern recognition method, optimal principal components were 3,2,2,3,4. Under the $\alpha=5\%$ significance level, 1 sample wasn't identified correctly in 10 prediction sample, it prove that method was established can identify exactly forsythia suspense from different habitats

Key words: forsythia suspense; near infrared spectroscopy; SIMCA; pattern recognition

1 前 言

连翘为木犀科植物连翘 forsythia suspense (Thunb1) Vahl 的干燥果实,味苦,性微寒,归肺、心、小肠经,具有清热解毒、消肿散结的功效。连翘苷是连翘的主要有效成分,具有抗菌、抗病毒及强心、抑制毛细血管通透性、抗肝损伤等作用[1]。传统的连翘鉴别法如性状鉴别法、紫外光谱法、薄层色谱法、离子色谱法等均需复杂的提取、分离、富集等前处理,操作繁琐、费时、不易在线测定。近红外光谱分析则具有快速省时、操作简单、无损伤测定和不受样品状态影响的特点,十分符合药物分析的要求[2]。因此,在制药业中原料药的分析、药物制剂中的水分、有效成分分析、药物生产品质的过程控制等方面,近红外光谱技术均得到了广泛的应用。

本文介绍在开发中药材光谱法在线检测系统研究的基础上,将中药材复杂的化学组分作为整体,采用近红外漫反射光谱构建连翘的图谱库,同时结合模式识别的SIMCA法对不同产地的连翘样品进行了模式识别方法学研究,取得了满意的分类结果,为中药材道地性的质量鉴别研究提供了新的研究思路,对中药现代化研究具有积极意义。

2 实验仪器与方法

2.1 仪器和样品

近红外光谱仪(matrix-2,德国 Bruker 公司),

作者简介:张晓慧(1979-),女,在读研究生,主要从事近红外光谱的研究。E-mail:zhang_xiao.hui@163.com

收稿日期:2007-10-15

漫反射积分球附件,OPUS 软件。连翘药材购买于河南卢氏、栾川、洛宁、新安和山西安泽等5个产地,各10个样本。

2.2 光谱采集

实验时将连翘药材粉碎后80目药材标准筛,将样本倒入样品杯中。采集条件:分辨率4cm^{-1},扫描次数64次,扫描范围4000~12000cm^{-1},数据格式为Log(1/R),每个样品3张光谱,计算平均光谱以建立模型。

2.2.1 粒度条件 分别取过40,60,80,100和120目的连翘药材采集光谱,结果显示大于60目筛的样品测定光谱的重现性好。本实验取样品过80目筛。

2.2.2 扫描次数的选择 扫描次数分别设置为16,32,64和128次时,取过80目筛的连翘样品进行测定,结果表明,扫描次数越多,噪声影响越小,扫描时间越长。本实验设扫描次数为64次。

2.2.3 分辨率 分辨率分别设置为4,8,16,32和64cm^{-1}时,取过80目筛的连翘样品进行测定,结果以分辨率4cm^{-1}为佳。当分辨率大于4cm^{-1}时,部分样品信息损失。本实验取分辨率为4cm^{-1}。

2.3 分析方法

模式识别一般是根据物以类聚的原则进行样本的分类,目前所采用的方法主要有聚类分析、判别分析、KNN法和SIMCA等方法,由于连翘近红外光谱特征变量数多,本研究最终选用SIMCA模式识别方法。

SIMCA(soft independent modeling of class analogy)方法实际上是相似分析方法,该方法在光谱、色谱的定性分析中得到了广泛的应用。在本研究中,SIMCA模式识别方法首先针对每一类样品的光谱数据矩阵进行主成分分析,建立主成分回归类模型,然后依据该模型对未知样品进行分类,即分别试探将该未知样本与各样本的类模型进行拟合以确定未知样本类别。具体的分析原理和步骤如下:

(1)建立类的主成分回归模型对于第q类样本中的第k个样本矢量$\boldsymbol{x}_{ik}^{(q)}$可用如公式(1)的主成分的回归模型表示:

$$x_{ik}^{(q)} = a_i^{(q)} + \sum_{a=1}^{A_q} \beta_{ia}^{(q)} \theta_{ak}^{(q)} + \varepsilon_{ik}^{(q)} \quad (1)$$

式中,$a_i^{(q)}$为变量的均值;A_q为主成分数;$\beta_{ia}^{(q)}$为变量i在主成分a上的载荷;$\theta_{ak}^{(q)}$为样本k关于主成份a的得分;$\varepsilon_{ik}^{(q)}$为偏差。

(2)用所建的q类模型拟合未知样本p,用拟合残差$S_p^{(q)^2}$表示样本p与q类模型的相似性,计算q类模型的总体偏差$S_0^{(q)^2}$和拟合残差$S_p^{(q)^2}$,分别如公式(2)和公式(3)所示:

$$S_0^{(q)^2} = \sum_{k=1}^{n_q}\sum_{i=1}^{m} (\varepsilon_{ik}^{(q)})^2 / [(n_q - A_q - 1)(m - A_q)] \quad (2)$$

$$S_p^{(q)^2} = \sum_{i=1}^{m} (\varepsilon_{ip})^2 / (m - A_q) \quad (3)$$

式中,n_q为第q类模型的样本数目;m为变量数;ε_{ip}为偏差。

(3)由公式(4)和(5)计算值与临界值F_0,通过F显著性检验判断未知样本p是否属于该类模型。如果$F < F_0$则样本p属于该类模型;否则样本p不属于该类模型。

$$F = S_p^{(q)^2} / S_0^{(q)^2} \quad (4)$$

$$F_0 = F_\alpha((m - A_q), (n_q - A_q - 1)(m - A_q)) \quad (5)$$

式中,α为显著水平;$(m - A_q)$,$(n_q - A_q - 1)(m - A_q)$为F自由分布的自由度。

3 结果与分析

3.1 波长范围的选择

图1是连翘的原始光谱图(a)和一阶导数光谱图(b)。从图1(a)中可以看出原始光谱在波数为5155cm^{-1}(波长1940nm)的附近有一个明显的吸收峰,从图1(b)中可以看一阶导数光谱在波数为5155cm^{-1}(波长1940nm)和6944cm^{-1}(波长1440nm)的附近有明显的波动。因为纯水中的O-H伸缩振动的一级基频区在1440nm附近,它的一个合频区在1940nm附近,在这两个波长附近是水分吸收的敏感区,从图1中可以看出在这两个区域,水分对连翘的近红外图谱的影响还是很大的。为了减少水分的影响,选择光谱波长范围尽量避开水分吸收峰的特征波长区。本研究有比较地选用了各段的波长进行了分析,结果显示选用6700~5300cm^{-1}范围内的光谱数据既避开了水分的影响,且取得了较好的实验结果。

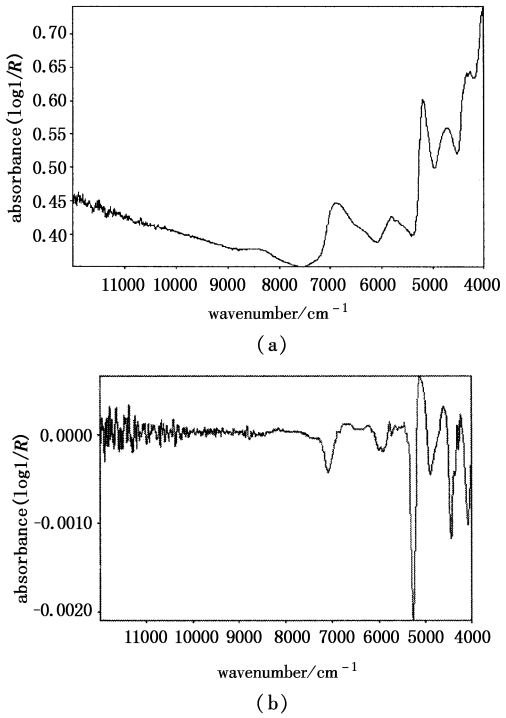

图1 连翘的原始光谱图(a)和一阶导数光谱图(b)

Fig.1 the original NIR spectra (a) and first derivative NIR spectra (b) of forsythia suspensa from five different habitats

3.2 数据的预处理

实验中,药材由于产地、气候、生长期、收获季节及储藏条件等的不同,会带来很多变化,这些都对光的漫反射有一定影响;同时样本的密实度也影响了光在样品中的传播。通常,NIR 光谱信号含有随机噪音、基线漂移、信号本底、样品不均、光散射等干扰,运用合理的光谱预处理方法可提取 NIR 光谱的特征信息,消除各种噪声和干扰,降低样品表面不均匀和色差等因素影响,提高模型的预测精度和稳定性[3]。实验中运用了多元散射校正(MSC)、标准归一化(SNV)、一阶导数和二阶导数等四种预处理方法,通过对比发现用 SNV 或 MSC 的方法明显优于一阶导数和二阶导数预处理方法,本实验最终采用了 SNV 的预处理方法。

3.3 模型的建立与主成分数确定

在 6700 ~ 5300cm^{-1}波数范围内分别截取这 50 个连翘样本的近红外光谱,在卢氏、栾川、洛宁、新安和山西安泽等 5 个产地随机挑选 9,8,7,7 和 9 个样本做为校正集,剩下的 10 个样本作为预测集,来检验模型的可靠性。按照公式(1)对已知样本进行主成分分解,通过交互验证来确定上述 5 个产地连翘模型的最佳主成分数,结果卢氏、栾川、洛宁、新安和山西安泽等五个产地的最佳主成分数分别为 3,2,2,3,4。

3.4 结果分析

图 2 是校正集光谱矩阵第一主成分和第二主成分得分图,表明校正集中的样本点在该二维平面上的投影。校正集光谱矩阵的第一主成分和第二主成分的方差贡献率边分别为 89% 和 8%,累计方差贡献率达到了 97%,所以样本在该二维平面上的投影分布可以充分表征样本在多维空间中的分布特征。由图 2 可以看出,5 个产地连翘的基本可以分开来。新安与山西安泽的相距较近,它们与卢氏的距离相对较远,而与栾川和洛宁的距离较远。这主要是其地理位置、气候、土壤等外在条件的不同而引起连翘内部品质的差异,同时不同的采摘时间也对连翘内部品质有较大的影响。

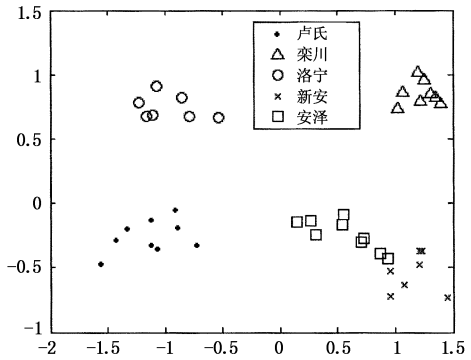

图 2 校正集光谱矩阵第一主成分和第二主成分得分图

Fig. 4 score cluster plot using first and second principal component(PC) for samples in calibration set

在显著水平 $\alpha = 5\%$ 条件下来检验模型的可靠性。在预测集的 10 个样品中只有 1 个被错判(新安的被判为山西安泽的),产生这种结果的原因从图 2 也可以看出,新安与山西安泽样本距离较近,投影点之间有相互交叉的区域,这样也就造成了新安的被错判为山西安泽的结果。

4 结果与展望

本实验利用近红外光谱方法识别了 5 个产地连翘,在主成分分析的基础上利用 SIMCA 模式识别原理对 5 个产地连翘分别建立了类模型,模型基本能正确识别这五个产地连翘,结果充分表明了近红外光谱结合 SIMCA 方法在连翘产地识别中的可行性。近红外光谱鉴别连翘产地方法操作简便、快速、结果准确,且无污染,是药物分析的新发展方向,扩展了近红外光谱在制药也在线检测中的应用,确保了原料质量,进一步提高了药物生产流程的效率,促进了制药工业的现代化进程。

中药品种繁多,应用历史悠久,产区广泛。由于地理和历史的原因,药材品种混乱,中药的同名异物或同物异名现象普遍存在,加之伪品、混淆品和误用品等因素,对中药的化学成分和药理作用的研究、制剂生产、临床疗效及推广使用等都有直接影响,因此,中药鉴定即真伪和道地性的鉴定是保证中药的真实性、确切疗效和用药安全的重要环节。随着先进的科学仪器与计算机技术相结合,仪器分析技术同化学计量学等方法联合建立起来的计算机信息处理技术(模式识别、人工神经网络、褶合变换等)的应用,模型的准确性、稳定性、适用性及传递性得到了增强。根据所建立的校正模型得到药材的物种、产地等重要信息,并由此得到一个定性的药材质量指数,这个指数相当于多位药材专家人工判别结果的总和,同时避免了个人主观因素对中药材质量定性判别的影响。因此,近红外技术在保证中药材质量的一致性,促进中药现代化、产业化和国际化方面具有广阔的发展前景。

参考文献:

[1] Zheng H Z, Dong Z H, She J, et al. Modern study application of traditinla Chinese medicine[M]. Beijing: Xueyuan Press, 1997. (in Chinese)

[2] 陆婉珍,袁洪福,徐广通,等. 现代近红外光谱分析[M]. 北京:中国石化出版社,2000.

[3] 刘荔荔,李力,邢旺兴. 不同种丹参药材的近红外漫反射光谱模式识别法鉴别[J]. 药物服务与研究,2002,2(1):23.