文章编号：1672-8785(2013)01-0037-05

# 近红外光谱技术在快速鉴别南蛇藤品种中的应用

罗威强[1]　杨海清[1]　李　云[1]　徐　宁[2]

(1. 浙江工业大学信息工程学院，浙江 杭州 310032；

2. 浙江工业大学药学院，浙江 杭州 310032)

**摘　要**：提出了一种利用近红外光谱技术对南蛇藤品种进行快速无损鉴别的方法。收集了 6 种南蛇藤样本，并用光谱仪获得了它们在 12493~4000 $cm^{-1}$ 范围的光谱曲线。通过用主成分分析法对预处理后的光谱数据进行聚类分析，获得了 10 个主成分。再结合不同的化学计量分析方法建立了品种鉴别模型。由于主成分 1 和 2 上的得分分布对不同样本的聚类效果明显，可根据得分分布定性地区分南蛇藤品种。从 220 个样本中随机抽取 165 个样本作为建模集，并将其分别用于建立线性鉴别分析、人工神经网络和支持向量机模型。剩下的 55 个样本用于预测验证。经过主成分数的优化，鉴别精度均达到了 100%。结果表明，本文提出的方法对南蛇藤的品种具有很好的分类和鉴别作用。

**关键词**：近红外光谱；南蛇藤；主成分分析

**中图分类号**：O433　　**文献标识码**：A　　**DOI**：10.3969/j.issn.1672-8785.2013.01.07

# Application of Near Infrared Spectroscopy in Rapid Identification of Celastrus Varieties

LUO Wei-qiang[1], YANG Hai-qing[1], LI Yun[1], XU Ning[2]

*(1. College of Information Engineering, Zhejiang University of Technology, Hangzhou 310032, China;*

*2. College of Pharmaceutical Science, Zhejiang University of Technology, Hangzhou 310032, China)*

**Abstract:** A method for rapid non-destructive identification of Celastrus varieties by near-infrared spectroscopy is proposed. Six kinds of Celastrus are collected. Their spectra in the region from 12493 $cm^{-1}$ to 4000 $cm^{-1}$ are obtained with a spectrometer. Through the clustering analysis of pre-processed spectral data by principal component analysis (PCA), ten principal components are obtained. Then, several variety identification models are established by combining the PCA with different stoichiometries. Since the score disribution in PC1 and PC2 has a remarkable clustering effect for different samples, it can be used to discriminate Celastrus varieties qualitatively. 165 samples randomly selected from 220 samples are used as modeling sets so as to establish linear discrimination analysis (LDA), back-propagation artificial neural network (BPANN) and support vector machine (SVM) models. The remaining 55 samples are used to validate the prediction. After optimization of principal components, all models have an identification precision up to 100%. The result shows that the proposed method is effective in the classification and identification of Celastrus varieties.

**Key words:** near infrared spectroscopy; Celastrus; principal components analysis

**收稿日期**：2012–11–19

**基金项目**：浙江省自然科学基金 (Y1090885); 浙江省 2012 年度留学人员科技活动择优资助项目

**作者简介**：罗威强 (1989-)，男，江西宜春人，硕士研究生，主要从事光传感和检测技术研究。

E-mail: luowq89@hotmail.com

**通讯作者**：杨海清 (1971-)，男，浙江温岭人，博士，副教授，硕士生导师，主要从事先进传感技术和传感器研究。 E-mail: yanghq@zjut.edu.cn

## 0 引言

全世界约有 50 种卫矛科南蛇藤属植物。这些植物主要分布在亚洲，我国有 30 种，集中在西南、华南和东北等地区。该属植物富含生物碱、萜类、强心苷及黄酮类化合物，具有消炎、杀菌、止痛、防病毒和肿瘤等生物活性 [1]。南蛇藤成为许多药物制剂的替代品，例如大芽南蛇藤、独子藤、粉背南蛇藤、灯油藤等，它们对人体的伤筋折骨、软组织受损、关节疼痛、痈肿疮疡等病症具有明显的治愈效果 [2]。由于气候、环境和人为因素的影响，不同地区的南蛇藤的化学成分和药用价值不同，不同品种差异较大，因此研究一种简单、快速、无损的南蛇藤品种鉴别方法，对于南蛇藤产业的健康持续发展具有重要的意义。

传统的中药鉴别方法主要包括眼看、手摸、鼻闻、口尝四个方面。这些方法可鉴定差异很大的中草药 [3]，但要鉴别品性近似的南蛇藤比较困难。许多现代仪器分析法如色谱法、质谱法和生物 DNA 等 [4]，需对药材进行分离提取，既耗时，又成本高，不能完全满足南蛇藤快速、无损鉴定的需要。

现代近红外光谱分析技术，具有快速省时、操作简单、无损伤测定、不受样品状态影响等特点 [5]，可充分利用全谱段或多波长光谱数据进行定性或定量分析，已被越来越多地应用于中草药材的质量监测和分析 [6]。例如，汪劲和程存归 [7] 利用傅里叶变换红外光谱快速鉴定传统中药玄参，测试精度高达 96%~100%；王丽等 [8] 应用光纤近红外光谱技术在线快速检测中药甘草中有效成分甘草酸含量；丁念亚和黎微等 [9] 应用近红外漫反射光谱分析技术对白芷、葛根、当归、白术等几种常见中药建立了一种简单、快速、有效的分类和真伪鉴别方法；陈雪英等 [10] 建立了一种用近红外透射光谱法快速测定赤芍水提过程有效成分含量的新方法；张爱军等 [11] 以丹参提取过程为研究对象，利用近红外在线检测技术建立了一种中药提取过程在线控制方法。但对南蛇藤进行品种快速鉴别方法的研究尚未见文献报道。

本研究利用近红外光谱主成分分析技术建立南蛇藤品种鉴别方法，通过提取其中主要的主成分信息，对南蛇藤样本进行聚类分析，再分别结合线性鉴别分析 (Linear Discriminant Analysis，LDA)、人工神经网络 (Backpropagation Artificial Neural Networks，BPANN) 和支持向量机 (Support Vector Machine，SVM) 方法建立三种鉴别模型，再用独立样本集进行模型性能验证，以确定最优模型。

表 1 实验用南蛇藤样本统计表

| 样本品种 | 1<br>大芽南蛇藤 | 2<br>独子藤 | 3<br>粉背南蛇藤 | 4<br>灯油藤 | 5<br>灰叶南蛇藤 | 6<br>青江藤 | 总计 |
|---|---|---|---|---|---|---|---|
| 建模集 | 17 | 33 | 27 | 31 | 24 | 33 | 165 |
| 预测集 | 5 | 11 | 9 | 10 | 9 | 11 | 55 |
| 样本总数 | 22 | 44 | 36 | 41 | 33 | 44 | 220 |

## 1 材料和方法

### 1.1 光谱的获取及预处理

实验所用仪器为 MPA 傅里叶变换光谱仪 (BRUKER，德国)，其吸收光谱采样间隔为 3.8 nm，测定范围为 12493~4000 $cm^{-1}$，分辨率为 2 $cm^{-1}$。从制药厂收集大芽南蛇藤、独子藤、粉背南蛇藤、灯油藤、灰叶南蛇藤、青江藤共六种南蛇藤，置于恒温箱干燥后，研磨成细粉，然后用其制备 220 个样本 (见表 1)。光谱探头置于样本上方，距离样本表面 15 cm，视场角为 25°，每一个样本扫描 30 次。为消除高频随机噪音、基

线漂移、样本不均匀、光散射等影响，采用光谱处理软件 Unscrambler X10.1 的 Savitzky-Golay 进行平滑处理，平滑点数为 11 个。将南蛇藤样本随机分成建模集 (165 个) 和预测集 (55 个)，见表 1 。

### 1.2　线性判别分析

LDA 是模式识别领域中常用的一种快速、高效定性分析方法，其基本思想是寻找一个向量使 Fisher 准则函数达到极值，然后让样本在此向量方向上投影，以达到类间离散度最大而类内离散度最小的目的 [12] 。在小样本高维数情况下抽取 Fisher 最优鉴别特征较为困难。因此，本文采用主成分分析 (Principal Component Analysis ，PCA) 与 LDA 相结合的解决方案，即将 PCA 作为预处理步骤，以消除噪声及变量间的复共线性，保证投影后样本的类内离散度矩阵是非奇异的，得到若干个主成分 [13] ，再将它们组合成新的分类特征矢量作为 LDA 的输入，建立 PCA-LDA 模型。

### 1.3　人工神经网络

人工神经网络主要用于处理非线性问题，其基本原理是模拟人脑细胞的工作来建立分类和预测模型。目前常用的是多层结构反向传播 BPANN ，其主要特点是信号前向传递，误差反向传播 [14] 。为了提高建模运算速度，减少运算量，本文把提取 PCA 降维后的特征数据用来建立 BPANN 判别模型。模型采用 Sigmoid 传递函数，网络隐含层神经元数为 14 ，输出神经元数为 1 ，学习速率为 0.05 ，训练迭代次数为 10000 次。

### 1.4　支持向量机

SVM 是一种新的 NIR 模式识别技术，它基于统计学习理论和结构风险最小原理，通过非线性映射将样本空间映射到一个高维的特征空间 [15] 。本研究用 PCA 降低全谱段数据维数，并将提取的特征信息作为 SVM 的输入。尽管 SVM 又提升了数据维数，看似加大了计算难度，但它将原本非线性区分的问题通过在高维空间中的一个线性超平面实现线性划分，在一定程度上避免了“维数灾难”，使得我们可以利用 PCA-SVM 算法建立南蛇藤样本与光谱信息之间的关系模型。

## 2　结果与讨论

### 2.1　南蛇藤的近红外吸收光谱

经过光滑预处理后， 6 种南蛇藤的平均吸收光谱曲线如图 1 所示。可以看出，在范围 12493~9000 $cm^{-1}$ 内，不同南蛇藤的光谱曲线有明显区分。从 9000 到 4000 $cm^{-1}$ ，品种 2 和 4 的光谱曲线始终重叠在一起，品种 1 、 5 和 6 大部分重叠，只有品种 3 单独分离。这些曲线都具有较为明显的指纹特征，例如 6900 $cm^{-1}$ 、5700 $cm^{-1}$ 、 5100 $cm^{-1}$ 和 4700~4300 $cm^{-1}$ 处的波峰，这些光谱差异为南蛇藤的品种鉴别提供了机理依据。

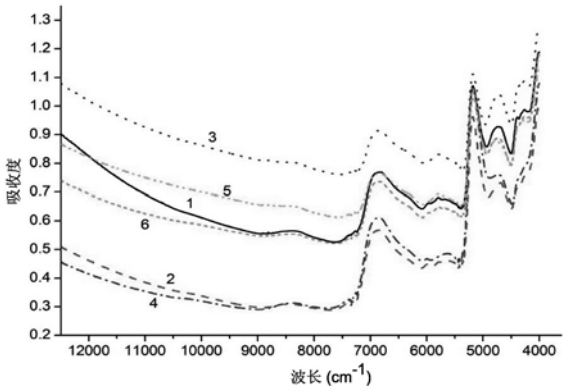


图 1　南蛇藤的平均近红外光谱曲线图

### 2.2　PCA

PCA 法可保留低阶主成分，忽略高阶主成分，以减少光谱数据集的维数，同时可保留数据集对方差贡献最大的特征。本研究对 6 种南蛇藤共 220 个样本应用了 PCA 法，得到的前 2 个主成分对方差贡献率达到 99.55% ，适合对不同品种的南蛇藤进行聚类分析。图 2 表示 220 个样本在主成分 1 和 2 上的得分图，图中横坐标显示每个样本的第一主成分得分值，纵坐标显示每个样本的第二主成分得分值。图 2 中品种 2 、 4 和 5 明显分成 3 堆，分别紧密地分布在图 2 的第三、二和四象限。但是品种 1 、 3 和 6 却有部分

重叠。这说明主成分 1 和 2 对南蛇藤有很好的聚类作用，尤其是对品种 2 、 4 和 5 的分类效果明显，品种 1 和 6 的聚合度也很好，大部分样本位于第一象限，也有少部分落在第四象限和坐标轴上。品种 3 的聚合度一般，样本分两堆聚集。鉴于仅用前 2 个主成分不能完全区分出品种 1 、 3 和 6 ，还必须通过融合其它主成分来鉴别南蛇藤品种光谱鉴别精度。

### 2.3 PCA-LDA, PCA-BPANN 和 PCA-SVM 模型

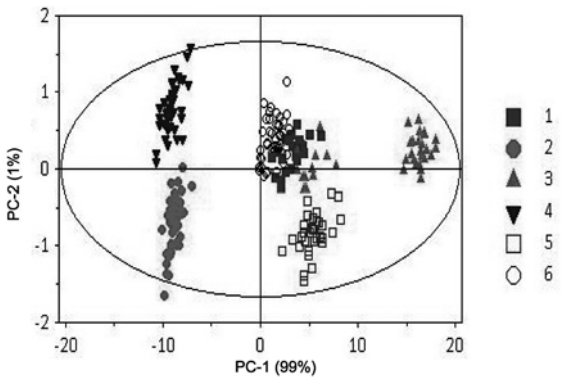


图 2 南蛇藤样本在主成分 1 和 2 上的得分图

表 2 PCA-LDA, PCA-BPANN 和 PCA-SVM 模型预测集确定系数 ($R^2$)

| 鉴别模型 | 主成分数 | 南蛇藤品种 | | | | | | 平均 $R^2$ |
|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | |
| PCA-LDA | 1 | 0.2 | 0.36 | 0.56 | 0.7 | 1 | 0.91 | 0.65 |
| | 2 | 0.8 | 1 | 0.56 | 1 | 1 | 0.82 | 0.87 |
| | 3 | 1 | 1 | 0.56 | 1 | 1 | 1 | 0.93 |
| | 4 | 1 | 1 | 0.56 | 1 | 1 | 1 | 0.93 |
| | 5 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| PCA-BPANN | 1 | 0 | 0.09 | 0.56 | 0.3 | 0.11 | 0.73 | 0.33 |
| | 2 | 0.4 | 0.91 | 0.56 | 0.7 1 | 0.27 | 0.65 | |
| | 3 | 1 | 1 | 0.78 | 1 | 1 | 0.91 | 0.95 |
| | 4 | 1 | 1 | 0.78 | 1 | 1 | 1 | 0.96 |
| | 5 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| PCA-SVM | 1 | 0.2 | 0.64 | 0.56 | 0.5 | 1 | 1 | 0.69 |
| | 2 | 0.6 | 1 | 0.78 | 1 | 1 | 0.91 | 0.91 |
| | 3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

将建模集的 165×n ( n 为主成分数 ) 样本数据作为 LDA 、 BPANN 和 SVM 的输入变量，品种 1 、 2 、 3 、 4 、 5 和 6 作为输出变量，建立南蛇藤品种鉴别模型。然后利用这些回归模型对预测集的 55×n 样本数据进行预测验证，预测结果见表 2 。结果显示 PCA-LDA 和 PCA-BPANN 模型在 5 个主成分鉴别时精度都达到了 100% ，而 PCA-SVM 只需要 3 个主成分就能 100% 识别全部样本。区别在于 SVM 可根据统计学理论中的结构风险最小化原则，从有限的数据集判别函数也能得到独立预测集较小的误差。而 LDA 的投影向量则是通过最大化数据集的类间散度同时最小化类内散度来获得的，这就需要大样本数据来支持投影方向的准确性。 BPANN 采用牛顿梯度迭代法进行回归，因此建模时间较长且易受样本维数影响。 SVM 通过求解二次规划问题代替迭代，其建模所需时间与样本维数关系不大，耗时明显小于 BPANN ，所以 PCA-SVM 的测量精度和泛化能力优于 PCA-BPANN ，更适用于南蛇藤近红外光谱的品种鉴别。从表 2 还可以看出，对于这三种模型，品种 2 、 4 和 5 的鉴别精度在主成分数较少的情况下就能达到 100% ，例如 PCA-LDA 的 2 个， PCA-BPANN 的 3 个和 PCA-SVM 的 2 个。这说明品种 2 、 4 和 5 南蛇藤相对于其他品种更容易被正确识别，这也从另一方面验证了图 2 中的聚类情况。

## 3 结论

本研究提出了一种基于近红外光谱的南蛇藤品种快速、无损鉴别方法。采用主成分分析对预处理过的全光谱信息进行数据压缩，在保留大多数信息的前提下降低了光谱数据维数。通过 LDA 、 BPANN 和 SVM 建模对未知样本进行了预测。结果显示，这三种回归模型的鉴别精度均能达到 100% 。经过主成分的优化， PCA-SVM 模型的预测性能比 PCA-LDA 和 PCA-BPANN 的更佳。结果表明，用近红外光谱技术结合模式识别的方法鉴别南蛇藤品种是可行的 。

## 参考文献


[1] 张舰，刘廷庆．南蛇藤属植物的化学成分与药理作用 [J]. **国外医药．植物药分册**, 2005, **20(5)**: 197-200.

[2] 郭远强，李铣．南蛇藤属植物化学成分研究进展 [J].**沈阳药科大学学报**, 2003, **20(3)**: 226-229.

[3] 沈昌明，王琳．传统中药鉴定方法的应用 [J].**医药论坛杂志**, 2010, **31**(20): 205-207.

[4] 陈士林，郭宝林，张贵君．中药鉴定学新技术新方法研究进展 [J].**中国中药杂志**, 2012, **37**(8): 1043-1055.

[5] Osborne B G. Principles and practice of near infrared (NIR) reflectance analysis [J].*International Journal of Food Science and Technology*, 1981, **16**(1): 13-19.

[6] 胡咏川，田晓鑫，刘蕾．近红外光谱技术鉴定中药的进展 [J]. **中国中药杂志**, 2012,**37**(8): 1066-1071.

[7] 汪劲，程存归．傅立叶变换红外光谱的 SVM 快速中药鉴别 [J].**仪器仪表学报**, 2005, **26**(8): 710-715.

[8] 王丽，何鹰，邱招钗．光纤近红外光谱法在中草药分析中的应用 - 甘草中甘草酸含量的测定 [J].**光谱学与光谱分析**, 2005, **25**(9): 1379-1399.

[9] 丁念亚，黎薇．近红外漫反射光谱在中药分类及真伪鉴别中的应用 [J].**计算机与应用化学**, 2008,**25**(4): 499-502.

[10] 陈雪英，李页瑞，陈勇．近红外光谱分析技术在赤芍提取过程质量监控中的应用研究 [J].**中国中药杂志**, 2009, **34**(11): 1355-1358.

[11] 张爱军，戴宁，赵国磊．丹参产业化提取中近红外在线检测技术的研究 [J].**中草药**, 2010, **41**(2): 238-240.

[12] James G M, Hastie T J. Functional Linear Discriminant Analysis for Irregularly Sampled Curves [J]. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2001, **63**(3): 533-550.

[13] Giordani P, Kiers H A L. Principal Component Analysis with Boundary Constraints [J]. *Journal of Chemometrics*, 2007, 21(12): 547-556.

[14] Ciampi A, Zhang F. A New Approach to Training Back-propagation Artificial Neural Networks: Empirical Evaluation on Ten Data Sets from Clinical Studies [J]. *Statistics in Medicine*, 2002, **21**(9): 1309-1330.

[15] Lins L D, Moura M C, Zio E, et al. A Particle Swarm-optimized Support Vector Machine for Reliability Prediction [J]. *Quality and Reliability Engineering International*, 2012,**28**(2): 141-158.


（上接第 29 页）

## 参考文献


[1] 葛小青．红外与可见光图像融合的研究 [D]. 重庆：重庆大学， 2010.

[2] da Cunha A L, Zhou J P, Do M N. The Nonsubsampled Contourlet Transform:Theory, Design and Applications [J]. *IEEE Transactions on Image Processing*, 2006, **15**(10): 3089-3101.

[3] Zhou Jian ping, Cunha A L, Do M N. Nonsubsampled Contourlet Transform: Construction and Applicationin Enhancement [C]. *International Conference on Image Processing*, 2005:469-472.

[4] 唐国维．嵌入式小波图像编码算法及应用研究 [D]. 哈尔滨：哈尔滨工程大学， 2010.

[5] 王跃华，陶忠祥．基于 NSCT 的红外与可见光图像融合算法 [J]. **四川兵工学报**, 2012, **33**(7) :117-119.

[6] Eckhorn R, Reitboeck H J. Feature Linking Via Synchronization Among Distributed Assemblies ： Simulations of Results from Cat Visual Cortex [J]. *Neural Computation (S0899-7667)*, 1990, **2**(3): 293-307.

[7] 张强，郭宝龙．一种基于非采样 Contourlet 变换红外图像与可见光图像融合算法 [J]. **红外与毫米波学报**, 2007, **26**(6): 476-480.

[8] 温黎茗，彭力，徐红．基于 NSCT 和 PCNN 的遥感图像融合算法 [J]. **计算机工程**, 2012,**38**(11):196-198.