DSWT-FTIR-RBFNN 鉴别罂粟和虞美人

张长江, 程存归2

1. 浙江师范大学数理与信息工程学院, 浙江 金华 321004

2. 浙江师范大学化学与生命科学学院, 浙江 金华 321004

摘 要 采用水平衰减全反射傅里叶红外光谱法(HATR-FTR)测定罂粟和虞美人的 FTR,由于两者为同科 同属中药材,所含化学成分较为相近,为了更好地突出罂粟和虞美人在 FTR 上的差异,并据此进行正确分 类识别,利用离散平稳小波变换(DSWT)分别对罂粟和虞美人的种皮和种仁的 FTB 进行若于尺度的变换。 从中选择 2个最具代表性的尺度作为特征提取的尺度空间。根据罂粟和虞美人的 FTR分布情况. 确定将 DSWT域内 2个尺度的 FTR 分别划分为 2个特征区域并以每个区域内的光谱能量作为特征参数。从而构造 一个包含 8个特征参数的特征向量,将这个特征向量输入到径向基函数神经网络(RBFNN)进行训练,从而 达到正确识别罂粟和虞美人的目的。实验中共取罂粟和虞美人的 FTR 数据 128对, 其中训练样本 78对, 测 试样本 50对。实验结果表明利用文章的方法对罂粟和虞美人的正确识别率分别为 99.8% 和 99.9%,从而验 证了方法的有效性。

关键词 傅里 叶变换红外光谱;离散平稳小波变换;径向基函数神经网络;罂粟;虞美人 中图分类号: 0657.3 文献标识码: A **DOI** 10. 3964/j issn 1000 0593(2009) 05-1255 05

引 言

鸦片为严重的成瘾性毒品,由罂粟科植物罂粟 Papaversom niferum L. 果实中的液汁凝固而成。而虞美人为优良观赏 花品种、与罂粟为同科同属植物、花、果实等植物形态非常 相似,有人曾将虞美人误作罂粟而报案,引起误会。由于植 物果实含有较高的脂肪油,采用传统的药典鉴别方法如性状 鉴别、显微鉴别和理化鉴别、以及现代仪器薄层色谱等均得 不到理想的效果。而傅里叶变换红外光谱 (FTR)由于具有 指纹特征分析、谱图整体分析、宏观推断分析等特点,适合 干分析复杂化学物质组成的稳定性, 故成为当今植物质量评 价研究领域的前沿课题,引起光谱学界的高度重视[1]。近年 来,小波变换已经逐步应用于化学及其相关领域[2-10]。本文 基于我们先前的研究工作^[11-13],利用傅里叶变换红外光谱 仪,借助水平衰减全反射(HATR)直接、快速、准确地测定 了罂粟和虞美人种子的不同部位的红外光谱,采用 Daubechies小波做小波母函数, 对罂粟和虞美人种仁及外表皮的 红外光谱进行若干尺度的一维离散平稳小波变换,在各个尺 度下观察罂粟和虞美人种仁及外表皮的红外光谱的差异程 度,从中选择 2个差异程度最为明显的尺度来提取罂粟和虞

美人的特征,然后输入到径向基函数神经网络进行训练和识 别。

1 理论部分

1.1 离散平稳小波变换

与经典的离散正交小波变换相比,离散平稳小波变换的 主要特点是冗余性和平移不变性, 能对连续小波变换给出一 个更为近似的估计。从矩阵的观点来看,离散平稳小波变换 可视为矩阵乘法。离散平稳小波变换由于其具有冗余性,因 而其逆变换有多种形式,本文采用 Guo等给出的逆变换形 式^[14]。

1.2 RBF人工神经网络

对于本文的二分类识别问题,把提取的 8个特征参数作 为径向基函数 (radial basis function RBF) 神经网络的输入矢 量[15],因而网络的输入层需 8个神经元;要求网络能识别出 2类不同的目标: 罂粟和虞美人, 因此建立的 RBF网络具有 8维输入和 2个输出, RBF网络结构如图 1所示。隐层结点 选取基函数作为转移函数,一般采用高斯函数

$$\Phi(x, \ \ \rho) = \exp[-(s - C_{\rm h}) 2/\rho^2]$$
(1)

式中: C₁表示基函数的中心, P表示宽度。结点计算输入向量

作者简介:张长江, 1974年生,浙江师范大学数理与信息工程学院副教授 email zcj74922@ zjnu.cn © 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

收稿日期: 2007-12-23 修订日期: 2008-03-29

基金项目:国家自然科学基金项目(60774011)和浙江省自然科学基金项目(Y506203)资助

辦率 4 m⁻¹, 扫描累加次数 32次。

2.2 光谱测定和数据分析

与中心的欧式距离,然后通过转移函数进行变换。第三层为 输出层, 第 *;*个输出结点的输出为

$$y_{j} = \sum W_{j} \phi_{i}(\| s - C_{h} \|, \rho)$$
(2)
式中 h 1, 2, ..., H; ||・||表示欧式范数。



仪器与样品 2

2.1 实验仪器和实验样品

0.06

0

-0.02

0.06

0.02

-0.02

0

Absorbance 0.04 0.02

Absorbance 0.04

实验仪器:美国 Nicolet公司生产的 NEXUS 670型傅里 叶变换红外光谱仪, DTGS 检测器, OMN IC E.S.P. 5.1智 能操作软件. HATR 附件. 光谱范围 4 000~650 m⁻¹. 分



图 2为罂粟与虞美人的种皮和种仁的 FTR 图谱。从图 2 中可以看出, 它们的图形主要差别在 2 900~ 3 000 m⁻¹之间 的范围内。

实验样品: 罂粟种子为罂粟科植物罂粟 Papaver son niferum L. 的干燥成熟果实. 虞美人种子为罂粟科植物虞美人 Pap aver rhoeas的干燥成熟果实。所有材料均由浙江省金华市

光谱测定: 在采集数据前, 根据仪器测试要求把 HATR 水平放置在傅里叶变换红外光谱仪的样品仓中,采用单面刀

数据分析:通过测定,得到样品种仁及外表皮的 FT R。

根据吸收峰的吸光度值特点,利用自编程序读取不同波数段 上的吸光度;采用 matlab软件进行小波变换分析,提取平稳

小波域罂粟和虞美人的种仁及外表皮的红外光谱的特征。

分别切取样品种仁及外表皮部置于 HATR 金刚石与压力校正 装置之间,按照所给定的测试条件直接测定样品的 FT R。为

了降低测定误差,图谱采用自动校正法进行基线校正。

药检所中药科提供,并经浙江师范大学植物学教研室鉴定。



Fig 2 FTIR spectra of coat and kernel of Papaver som niferum L. (top) and Papaver rhoeas (bottom)

(a): Coat of Papaver som niferum L; (b): Coat Papaver rhoeas (c): Kernel of Papaver som niferum L; (d): Kernel Papaver rhoeas

罂粟在 2 928 cm⁻¹处有一个吸收峰,而虞美人在此处是 个双重峰. 分别为 2 971 和 2 922 cm⁻¹。由于罂粟和虞美人 为同科同属植物,所含化学成分尽管有差异,但主要化学成 分基本相同. 故在 FTR 上吸收峰比较相近. 区别不是很明 显。但是仅仅依靠图 2直接判断二者的区别并不明显,因而 有必要寻求一种更为有效直观的技术以鉴别罂粟和虞美人。 3.2 主成分分析

对 FTR 谱图中的数据进行主成分分析, 通过主成分分 数矩阵 C、特征值 λ 和方差矩阵 S, 可得到主成分荷重矩阵 CL, 主成分荷重反映了主成分和原始 FTR 变量之间的相关





性。选取分析结果中的 PCA case scores数据作二维图,结果 如图 3所示。从主成分分析图也可以看出信息负荷量的分 布,信息负荷量比较大的区域出现在 2 000~ 800 m⁻¹,由于 此区域包括指纹区范围,所表现的分子结构信息较为丰富, 而在高波数吸收区域主要是羟基及氨基的伸缩振动吸收,特 征性不明显。故本文在 2 000~ 800 m⁻¹区域进行小波特征 提取并进行人工神经网络分析。

3.3 FTIR的离散平稳小波特征提取

在小波多尺度分解过程中,根据 FTIR信号的特性并比

较不同尺度下信号分解的效果来确定合适的小波基函数及小 波尺度。其标准是突出原始光谱中的若干个特征峰,并选取 平滑性好的小波基。这里我们选择 Daubech ies小波作为"分 析小波"。在本文中对罂粟和虞美人的傅里叶红外光谱分别 进行 DSWT,分解的层数为 5,我们选择其中二层(4和 5)提 取其特征。图 4和图 5中的第一行表示 2 000~ 800 m⁻¹区域 内的罂粟(左)和虞美人(右)的种皮和种仁的傅里叶红外光 谱曲线,第 2~ 6行分别表示平稳小波域的罂粟和虞美人的 种皮和种仁在 5个尺度下的小波系数。



Fig. 4 FTIR and its DSWT coefficients of coat of Papaver somniferum L. (left) and Papaver rhoeas (right)





(a): Papaver somni ferum L.; (b): Papaver rhoeas

由图 5可见,当 DSWT的分解尺度比较小的时候,细节 信号对光谱变化比较敏感,对原始光谱中各个特征峰反应过 于强烈,反而不利于特征提取。为此取第 4和 5这 2个尺度 的 DSWT细节信号作为特征变量提取空间。特征变量定义为 DSWT内种仁和种皮的第 4和 5等 4个尺度下光谱的能量。 图 6表示特征区间的划分示意图,由图 6可知,我们将 2 000 ~ 800 m⁻¹区域划分二个特征区域: 1 800~1 300 m⁻¹和 1 200~900 m⁻¹。这样特征向量由小波域的种皮和种仁的第 4 和 5等 4个尺度下的 8个特征区间构成,特征值分别为这 8 个特征区间的光谱能量。



Fig. 6 D ivision of feature region in the CWT domain

3.4 分类结果

为验证本文方法的有效性,在试验中取 128对 罂粟和虞 美人的傅里叶红外光 谱进行实验,其中包括 罂粟和虞美人的 种皮和种仁样本各 64对。利用 78对数据 (罂粟和虞美人的 种皮和种仁样本各 39对)利用神经网络进行训练,剩余其余 50对数据(罂粟和虞美人的种皮和种仁样本各 25对)用于测 试。结果表明:利用 RBF神经网络对罂粟和虞美人的训练样 本的正确识别率均为 100%,而对测试样本的正确识别率分 别为 99.8%和 99.9%。可见利用 RBF神经网络结合离散平 稳小波特征可以对罂粟和虞美人进行高精度的正确分类识 别。

4 结 论

应用离散平稳小波对所得样品 FT R 进行特征提取后, 采用 R BF 神经网络进行罂粟和虞美人的分类具有较高的识 别率。与传统的 PCA方法相比,利用 DSWT提取罂粟和虞美 人的特征有如下优点:利用 PCA提取傅里叶红外光谱特征 时必须要求样本数大于信号的维数,因而为了满足这个条 件,需要再采样已达到降维的目的。而利用 DSWT提取傅里 叶红外光谱特征时对样本数没有限制。利用 PCA 提取特征 时是通过所有样本的统计特性来实现的,这就要求新样本应 该与原来的样本集应该有类似的性质。而利用 DSWT提取特 征时是直接在小波域内划分特征区间,所有的特征向量都直 接被提取出来。这样利用 DSWT提取特征就对新样本具有很 强的自适应性。对于同样的测试数据,分别利用 PCA 和 DSWT法抽取信号特征时,二者有相似的收敛性,但是利用 DSWT方法时的正确识别率明显优于 PCA法。

参考文献

- CHENG Currgui, RUAN Yongming LIBing lan(程存归, 阮永明, 李冰岚). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2004 24(11): 1355.
- [2] Ehren treich F. Analytical and Bioanalytical Chemistry, 2002, 372 (1): 115.
- [3] Ek strom P A, H ales J M. M onth ly W eather Review, 2000, 128(9): 3169.
- [4] Singh S, Rastigejev Y, Paolucci S, et al Combustion, Theory and Modeling 2001, 5(2): 163
- [5] Liu Y, Cameron IT, Bhatia S K. Computers & Chemical Engineering 2001, 25 (11-12): 1611.
- [6] Shao Lin in, Lin Xiangqin, Shao Xueguang Applied Spectroscopy Reviews 2002, 37(4): 429.
- [7] Jyengar S S, Frisch M J, Journal of Chemical Physics, 2004, 121(11): 5061.
- [8] Ying Y B, Liu Yande, Fu Xiaping et al. Proceedings of the SPIE-the International Society for Optical Engineering 2004, 5587 (1): 29
- [9] Wang D, Pan J Computational Materials Science, 2004, 29(2): 221.
- [10] Chen Jun, Wang Xue Z. Chen ical Engineering Communications, 2005, 192(4-6): 499
- [11] ZHANG Chang-jiang CHENG Currgui(张长江,程存归). RareMetalMaterials and Engineering(稀有金属材料与工程), 2006, 35 614
- [12] ZHANG Chang-jiang LID arr ting LIANG Jiu-zhen, et al(张长江,李丹婷,梁久祯,等). Spectroscopy and Spectral Analysis(光谱学与光谱 分析), 2007, 27(1): 50
- [13] ZHANG Chang-jiang CHENG Curr gu (张长江,程存归). Spectroscopy and Spectra l Analysis(光谱学与光谱分析), 2006 (Supp): 63
- [14] LangM, Guo H, Odegend JE, et al In SPIE Conference on Wavelet Applications vol 2491, Orkindo, E, April 1995, 2491
- [15] Sing Jamuna Kanta, Basu Dipak Kumar, Nasipuri Mita et al Applied Soft Computing 2007, 7(1): 58.

Iden tification of *Papaver Sonniferum* L. and *Papaver Rhoeas* Using DSWT-FTIR-RBFNN

ZHANG Chang-jiang¹, CHENG Curr gu f

 $1. \ College of M \ a them \ atics, \ Physics \ and \ Inform \ ation \ Engineering \ Zhejiang \ Norm \ a University, \ Jinhua \ 321004, \ China \ Model \ She \$

2 College of Chemistry and Life Science, Zhejiang Normal University, Jinhua 321004, China

Abstract Infrared spectra of Papaver som niferum L and Papaver rhoeas were obtained directly, quickly and accurately by Fourier transform infrared spectroscopy (FTR) with OMNI sampler D iscrete stationary wavelet transform was used to extude local region of ir frared spectra of Papaver som niferum L and Papaver thoeas. The difference of infrared spectra between Papaver som niferum L and Papaver thoeas was extuded Accurate identification rate is improved greatly. One dimensional discrete stationary wavelet transform was in plen ented to the infrared spectra of Papaver som niferum L and Papaver thoeas. The difference between Papaver som niferum L and Papaver thoeas was observed at all scales in wavelet domain. Two scales, at which the difference between Papaver som niferum L and Papaver thoeas is them ost obvious, were selected to extract the features of Papaver som niferum L and Papaver rhoeas. A feature vector including eight feature parameters as constructed. The feature vectors as input to RBFNN for training in order to accurately identify Papaver som niferum L and Papaver thoeas. In experiment, the authors used one hundred and wenty eight couples of data of Papaver som niferum L and Papaver som niferum L and Papaver thoeas. The experiment L and Papaver rhoeas (including seventy eight couples of training samples and fifty couples of testing samples). The experimental results show that it is effective to apply discrete stationary wavelet transform on the basis of FTR to identify the Papaver som niferum L and Papaver rhoeas. The accurate identification rate of Papaver som niferum L and Papaver rhoeas is 99.8% and 99.9% respectively.

Keywords Fourier transform infrared spectroscopy, Stationary wavelet transform; Radius basis function neural network, Papaver son niferum L; Papaver rhoeas

(Received Dec 23, 2007; accepted Mar 29, 2008)