

# 基于雷达图特征提取和近红外光谱技术的葛根粉鉴别方法研究<sup>①</sup>

崔建新<sup>②\*\*\*</sup> 高海波<sup>③\* \*\*\*</sup> 洪文学\*

(\* 燕山大学电气工程学院 秦皇岛 066004)

(\*\* 河北省测试计量技术及仪器重点实验室 秦皇岛 066004)

(\*\*\* 中国中医科学院针灸研究所 北京 100000)

**摘要** 基于雷达图表示原理,对葛根粉的近红外光谱进行了鉴别研究。首先,利用近红外光谱技术获得了中药葛根粉在  $12000\text{cm}^{-1}$  ~  $4000\text{cm}^{-1}$  范围内的近红外漫反射光谱作为指纹图谱,然后采用多层雷达图对指纹图谱逐步进行特征提取,最后用优选图特征进行分类研究。该方法在保证数据整体性的前提下,有很好的直观可视性,最小识别误差率达 1.04%。研究结果表明该方法既是一种可行的葛根粉鉴别新方法,同时又可作为可视化中药指纹图谱模式分析新方法。

**关键词** 葛根, 近红外光谱, 雷达图, 中药, 指纹图谱, 特征提取

## 0 引言

葛根为豆科植物野葛或甘葛藤的干燥根<sup>[1]</sup>。据《本草纲目》记载,葛根性凉、味甘、气平,具清热、降火、排毒诸功效。葛根化学成分复杂,含有丰富的淀粉、葛根素、异黄酮苷、大豆黄酮和木黄酮等物质<sup>[2]</sup>,另外,还含有多种人体所必需的微量元素如钙、硒、铁、铜、磷、钾等。现代医学研究表明葛根对多种疾病都有一定疗效,因而有广阔的市场前景。葛根以粉状销售,各国专家将葛根粉(也称葛粉)誉为“长寿粉”。葛根粉的生产成本要高出其他淀粉 2~6 倍,有些生产厂家为了谋取暴利,在葛根粉中混入了其他淀粉。葛根粉、马铃薯粉、藕粉、玉米粉等外观上非常相似,很难区别,混在一起就更难鉴别。所以采用合适的方法对葛根粉的真假优劣进行有效鉴别是一项非常重要的研究课题。

近年来,近红外光谱分析技术以高效、无损、结

果稳定等优点在食品工业、环境工程、石油化工、纺织工业以及制药工业等领域得到广泛的应用<sup>[3-7]</sup>。近红外光谱的测量不需要样本的化学预处理,以非破坏的方式从样本直接获取信息,有效地避免了预处理对样品所造成的微量的组分损失或者组分形态的变化,最大限度地保留了不同中药材样本之间的微小差异。而且,谱图中蕴含了极其丰富的被测物信息,有利于反映被测物的整体性征,这与中药的“君臣佐使,协同作用”的理念是完全一致的。同时,近红外光谱技术可以分析药物的固体、浸膏等多种不同状态,也适合用于中药领域的分析。另外,近红外光谱技术可以多个成分和多个参数同时分析,这有利于缩短分析时间,提高分析效率。所以,近红外光谱技术在中药的鉴别和评价中具有很大的发展空间。目前,许多中药是通过计算指纹图谱相似度来进行鉴别与评价的,通过相似度计算。我们可以得到指纹图谱数据间的相关系数或夹角余弦,这是数据间的抽象关系。而雷达图表示可以将抽象的数

① 中国博士后科学基金(2012M510722),中国中医科学院(ZZ12001),燕山大学博士基金(B692)和秦皇岛市科学技术研究与发展计划(201001A119)资助项目。

② 女,1977 年生,博士;研究方向:模式识别,中药指纹图谱;E-mail: ydcuijianxin@yahoo.com.cn

③ 通讯作者,E-mail: hhghb@ysu.edu.cn

(收稿日期:2014-12-31)

据以形象的雷达图的形式表现出来,提取的图形特征同时刻画数据样本的内部数据结构以及样本间的关系。而且形象的图形特征可能会有利于视觉上的分类或聚类。因此本文采用多层次雷达图特征提取方法对葛根粉的近红外光谱进行分析。

## 1 材料与方法

### 1.1 实验仪器

德国布鲁克光学仪器公司(Bruker Optics Inc.)的 MPA 型傅立叶近红外光谱仪,光谱扫描范围为  $12000\text{cm}^{-1} \sim 4000\text{cm}^{-1}$ ,扫描次数为 64 次,光谱分辨率为  $8\text{cm}^{-1}$ ; TE-InGaAs 检测器,分析软件为 OPUS 6.5。

### 1.2 样品制备与光谱测量

为了保证纯葛根粉样本品质,纯葛根粉样本采用由医药公司采购的葛根片粉碎加工获得,假葛根粉从网上选择售价低于市场平均价格 40% 以上的 3 种葛根粉产品。葛根片选取比较有代表性的江西、福建、湖南、河南以及泰国 5 个不同产地的葛根样品。每种葛根样品制备 10 份,共计 50 个样本,依次编号,置于实验室室内 2h 后,进行光谱采集,每个样本采集 3 次,这样我们就得到 150 个葛根光谱,其中一组葛根光谱如图 1 所示。

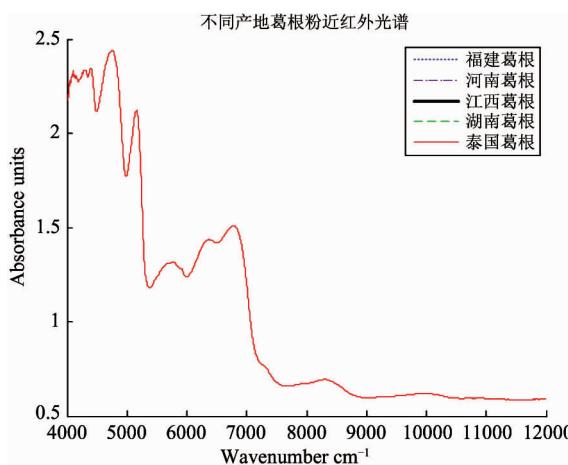


图 1 不同产地的葛根近红外光谱曲线

结果表明不同产地葛根的近红外光谱特性具有同一性。即江西、福建、湖南、河南以及泰国 5 个不同产地的葛根样品在本项目中所使用的近红外光谱

仪扫描范围内的近红外光谱曲线完全重合,即葛根粉自身差异在近红外光谱特性上不显著。

假葛根粉每种样品 10 份,共计 40 个样本,从 4 类样本获得的光谱数据中各随机选择一个得到 4 种葛根粉的近红外光谱曲线如图 2 所示。从图 2 中可以看出,劣质的市售葛根粉与纯葛根粉之间存在明显的差异性,且市售葛根粉的特性比较接近,这是由于市售劣质葛根粉中掺加的多为各种淀粉。

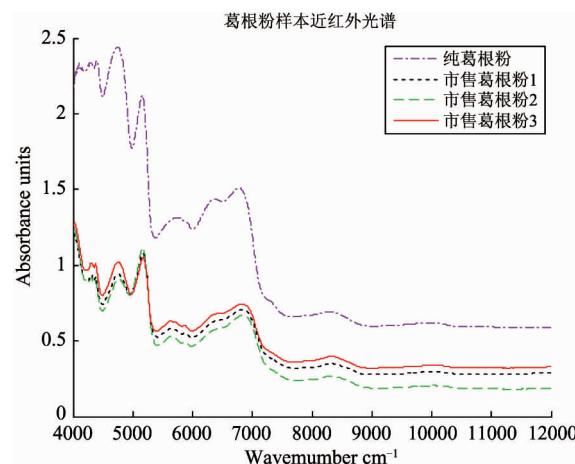


图 2 葛根粉样本的近红外光谱曲线

同样我们得到葛根粉和红薯淀粉、玉米淀粉、土豆淀粉的近红外光谱如图 3 所示。可以看出,葛根粉和红薯淀粉、玉米淀粉、土豆淀粉的近红外光谱具有明显的差异性,且纯葛根粉与淀粉的差异性以及不同淀粉之间的相似性与图 2 相似。这说明利用近红外光谱既可以鉴别葛根的真伪,并且可以进一步确定伪劣葛根产品其中的假冒成分。

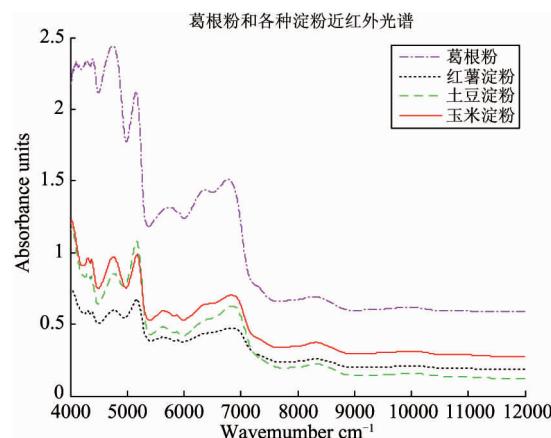


图 3 葛根粉和淀粉的近红外光谱曲线

### 1.3 雷达图特征提取

通常情况下,中药指纹图谱数据的样本包含以下参数值:相对保留值(也就是相对保留时间)、n 强峰、峰值、峰面积值,以及特征峰数、共有峰数和重叠率等<sup>[7]</sup>。这些参数被映射为雷达图,并作为中药指纹图谱数据在特征空间中的描述,通常可以得到 3 个以上的雷达图,且维数都比较高,这不利于分类过程的直观理解及分类计算。

基于上述问题,本研究将中药指纹图谱数据进行非等间距(或等间距)雷达图表示,然后将表示出的多个非等间距(或等间距)雷达图进行图形特征提取,多元图的图形特征可以为多元图的整体特征以及局部特征,而整体特征包含有面积、重心矢量、方向和位置等特征,局部特征则包含有相邻幅值比、分区面积比以及对称性等特征。然后把得到的图形特征作为特征参数绘制下一级雷达图,这样就实现了降维,如果得到的雷达图仍然不利于分类及可视化,那么再进一步进行雷达图的图形特征提取和绘制雷达图,如此类推,直到最后变量的数目满足分类要求,可以采用一个雷达图清楚描述整个指纹图谱特征为止。多层次雷达图的特征融合原理如图 4 所示。

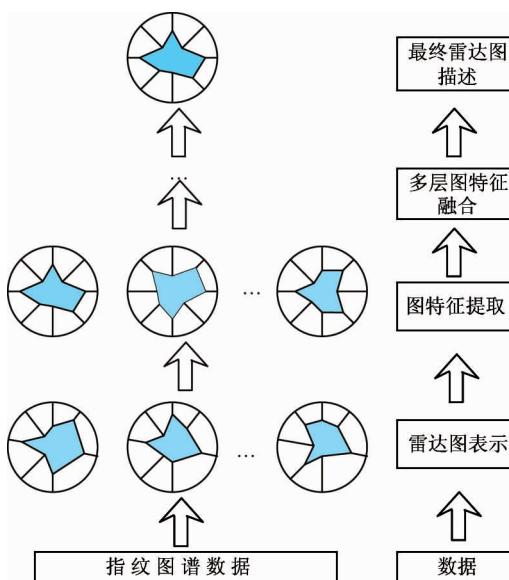


图 4 雷达图的特征融合过程

### 1.4 雷达图的重心图形特征

Tang 等<sup>[8]</sup> 和 Liu 等<sup>[9]</sup> 分别在《Science》和《Na-

ture》杂志上发表研究论文指出,果蝇能够通过物体外形的重心高度、朝向等参数实现对不同物体的分辨识别并形成记忆,这表明图形的重心可能是一种有利于图形辨识的重要特征。

对于样本数据的雷达图表示,从几何角度看,一个样本的雷达图中相邻  $m$  个变量构成一个  $m + 1$  边形(当  $m = n$  时构成  $n$  边形)都产生一个重心,这构成多元数据雷达图重心特征。容易知,一个  $n$  元变量样本的雷达图的  $m$  变量重心特征的维数为  $C_n^m$ ,  $n$  元变量样本的雷达图的重心特征总维数为  $C_n^2 + C_n^3 + \dots + C_n^n = (2^n - n - 1)$  个。其中比较特殊的有两种情况:一是相邻 2 变量构成三角形,从几何上讲三角形三条中线相交于一点,这个交点叫做三角形的重心,而且三角形的重心与顶点的距离等于它与对边中点的距离的 2 倍,这种情况所具有的优秀几何特性是实际应用中受到青睐的基础;另一种情况是所有变量构成  $n$  边形的重心,显然全变量雷达图重心特征的维数为 1,它包含了样本的全局信息。雷达图重心图特征包括重心矢量模长(原点到重心的幅值)和方向,下面以雷达图 2 变量重心矢量模长图特征为例予以介绍,其它情况类似。如图 5 所示,8 维数据  $(x_1 \ x_2 \ \dots \ x_8)$  雷达图表示如图,相邻变量  $x_1$  和  $x_2$  构成三角形  $AOB$ ,  $G$  表示重心,  $OG$  表示原点  $O$  到重心  $G$  的距离。

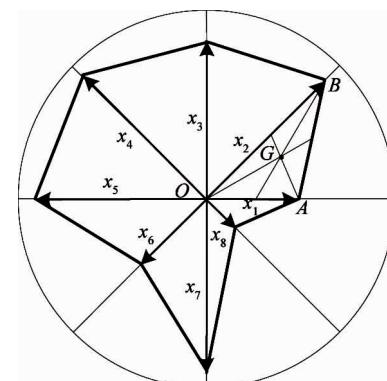


图 5 雷达图二变量重心图特征提取

那么,雷达图中由相邻二变量构成的局部重心矢量模长图特征提取公式为

$$\begin{cases} abs_i = \sqrt{\left(\frac{r_i + r_{i+1} \cos \omega_i}{3}\right)^2 + \left(\frac{r_{i+1} \sin \omega_i}{3}\right)^2}, \\ angle_i = \arctg\left(\frac{r_{i+1} \sin \omega_i}{r_i + r_{i+1} \cos \omega_i}\right) \end{cases},$$

$$i = 1, \dots, d \quad (1)$$

这里  $abs_i$ ,  $angle_i$  分别表示第  $i$  维变量和第  $i + 1$  维变量组成的三角形的重心的幅值和真实的角度;  $r_i$  和  $r_{i+1}$  表示归一化特征样本的第  $i$  维变量和第  $i + 1$  维变量的值,一般由预处理计算得到;弧度  $w_i = 2\pi/d$  为第  $i$  维变量和第  $i + 1$  维变量间的夹角弧度,这里认为圆周角被样本维数  $d$  平分。

## 2 结果与分析

本文中的实验均在 Matlab7 环境下编写。

### 2.1 葛根粉光谱曲线的峰值检测和特征表示

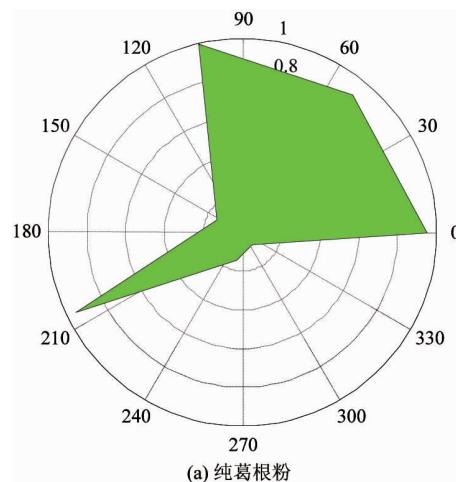
表1为从各类实验样本数据中任选一个样本在

选择不同波身宽度设定参数下检出的波峰数目。

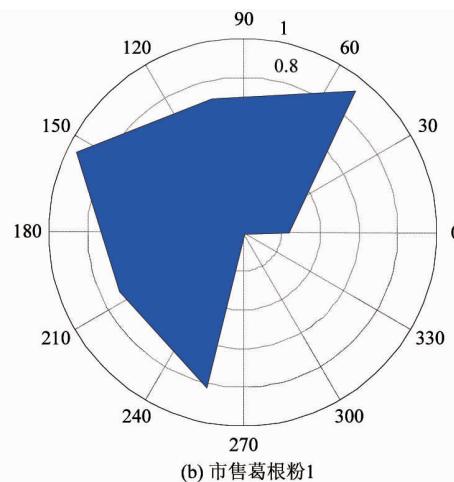
表1 葛粉光谱检出波峰数目

数据样本	波峰数(波数)						
	4	8	12	16	20	24	28
纯葛根粉	97	78	35	11	8	7	7
市售葛根粉 1	93	75	37	15	10	8	8
市售葛根粉 2	92	74	40	13	8	8	8
市售葛根粉 3	88	66	35	17	7	7	7

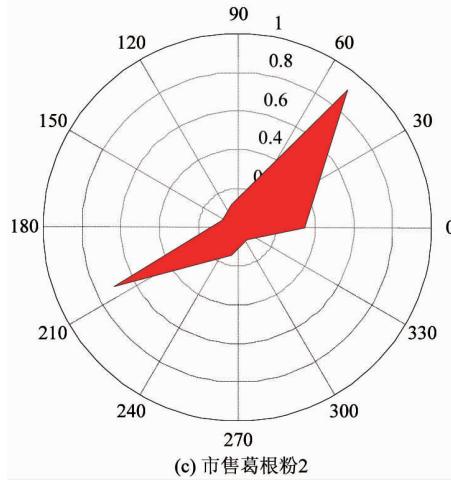
应用峰值检测程序对所有样本数据进行峰值检测(整个波段),选择波峰起伏宽度设定参数为 24 个波数检出的波峰特征值最少为 7,对于所有样本数据经波峰检测检出特征大于 7 的一律选取前 7 个波峰特征,用各类样本数据的特征均值作为该类特征模式,绘制各类样本的雷达图如图6所示。图6



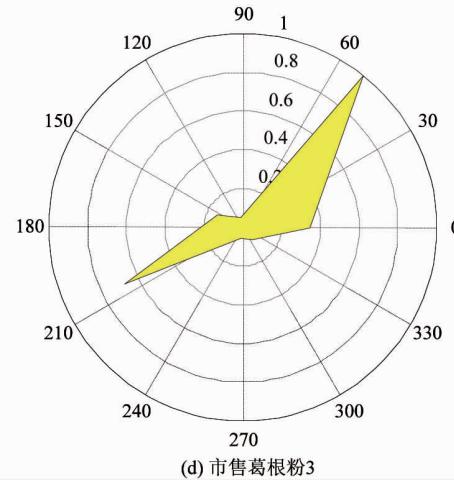
(a) 纯葛根粉



(b) 市售葛根粉1



(c) 市售葛根粉2



(d) 市售葛根粉3

图6 葛根粉样本波峰特征雷达图

中,(a)为纯葛根粉样本,(b)(c)(d)分别为市售葛根粉1,2,3样本。为了能够更充分地利用光谱信息来进行研究对象的鉴别区分,下面的特征融合研究中选择波峰起伏宽度设定参数为4个波数的检出波峰特征组成初始数据集,即利用光谱曲线中的所有波峰特征。

## 2.2 葛根粉光谱的多层雷达图特征融合

由前面的光谱曲线峰值检测得到了4种样本的波峰特征,即波数递增序列下对应的波峰峰值。由于当维数较高时(一般不大于30维),多元数据雷达图表示将很难对多元图进行观察。所以,采用多层次雷达图特征融合方法(即前面提到的多层次雷达图特征提取过程)将高维数据用多个雷达图进行表示,特征融合过程中多层次参数设定为3层。

## 2.3 结果及讨论

这里的数据实验选择标准化的原始特征表示、原始特征的主成分特征提取、选择波峰起伏宽度设定参数为4个波数的检出波峰特征表示和波峰特征的主成分特征提取的分类性能作为对照。分类器分别选择的是线性分类器(ldc)和以标准概率密度为基准的贝叶斯分类器(nbayesc)以及k近邻分类器(knnc)。从数据中随机选取80%作为训练集,20%作为测试集,测试集的10次交叉检验误差率作为分类结果评价指标。表2所示为实验结果。

表2 鉴别实验误差率

特征表示方法	Ldc	Bayesc	Knnc
标准化原始特征表示(data)	0.3710	0.4148	0.3725
Data + PCA	0.3623	0.3739	0.3652
4个波数的检出的波峰特征表示(fengdata4)	0.0734	0.1052	0.1052
fengdata4 + PCA	0.1014	0.0953	0.0961
基于多层次雷达图重心图特征融合表示	0.0225	0.0104	0.0236

## 3 结论

本研究首先对葛根粉的近红外光谱进行峰值检测和特征表示,然后进行多层次雷达图特征融合提取葛根光谱峰值特征的雷达图重心特征,最后进行葛根近红外光谱的鉴别分析。实验结果表明利用近红外光谱鉴别葛根真假优劣是可行的,同时雷达图特征提取方法在保证数据整体性的前提下,能有很好的直观可视性,可以作为一种新的中药指纹图谱特征提取方法。

### 参考文献

- [1] 中国药典[S].一部. 2005
- [2] Cai R L, Li M, Xie S H, et al. Antihypertensive effect of total flavone extracts from Puerariae Radix. *Journal of Ethnopharmacology*, 2011, 133(1):177-183
- [3] 严衍禄,赵龙莲,韩东海等.近红外光谱分析基础与应用.北京:中国轻工业出版社,2005
- [4] 陆婉珍,袁洪福,徐广通等.现代近红外光谱分析技术,第2版.北京:中国石化出版社,2007
- [5] 王丽,卓林,何鹰等.近红外光谱技术鉴别海面溢油.光谱学与光谱分析,2004, 24 (12): 1537-1541
- [6] 孙通,徐惠荣,应义斌.近红外光谱分析技术在农产品/食品品质在线无损检测中的应用研究进展.光谱学与光谱分析,2009, 29(1): 122-126
- [7] 梁欣健,刘家水,张丹雁.浅谈近红外光谱技术在中药鉴定中的应用.广州轻工业,2011, 2:4-7
- [8] Tang S, Guo A. Choice behavior of Drosophila facing contradictory visual cues. *Science*, 2001, 294: 1543-1547
- [9] Liu G, Seiler H, Wen A, et al. Distinct memory traces for two visual features in the Drosophila brain. *Nature*, 2006, 439(7076): 551-556

# Research on identification of radix pueraria power based on radar graph feature extraction and near infrared spectra

Cui Jianxin<sup>\* \*\*\*</sup>, Gao Haibo<sup>\* \*\*\*</sup>, Hong Wenxue<sup>\*</sup>

(<sup>\*</sup>School of Electric and Electronic Engineering, Yanshan University, Qinhuangdao 066004)

(<sup>\*\*</sup>Measurement Technology and Instrumentation Key Lab of Hebei Province, Qinhuangdao 066004)

(<sup>\*\*\*</sup>Institute of Acupuncture and Moxibustion China Acadimy of Chinese Medical Sciences, Beijing 100000)

## Abstract

According to the principle of radar graph representation, the near infrared spectra of Radix Pueraria power were identified. Firstly, by using the near infrared spectroscopy, the near infrared diffuse reflection spectra of Radix Pueraria Power in the scope of  $12000^{-1} \sim 4000\text{cm}^{-1}$  were obtained to be as fingerprints. Secondly, the feature extraction for the fingerprints was performed gradually by using the multilayer radar graph. Finally, the classification was studied by using the optimized graph features. On the premise of guaranteeing the data integrity, this method has a good intuitive visibility, and its minimal recognition error rate can reach 1.04%. The experimental results showed that the study provided not only is a feasible new method for Radix Pueraria power identification, but also a new method for visualized pattern analysis of the fingerprints of traditional Chinese medicine.

**Key words:** radix pueraria, near infrared spectra, radar graph, traditional Chinese medicine, fingerprint, feature extraction