

雷达图表示与相似度计算相结合的麻黄品种鉴别方法^①

崔建新^{②*} 崔建凤^{***} 洪文学^{*} 高海波^{③*} ****

(* 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)

(*** 秦皇岛职业技术学院旅游系 秦皇岛 066004)

(**** 中国中医科学院针灸研究所 北京 100000)

摘 要 将指纹图谱技术用于中药麻黄的鉴别,提出了指纹图谱雷达图表示与相似度计算相结合的麻黄品种鉴别分析方法。该方法首先对麻黄样品的 HPLC 指纹图谱的峰面积分别进行相似度计算和雷达图表示,然后对雷达图进行重心图特征提取,最终得到用于分类比较的雷达图特征,最后结合雷达图特征与相似度特征对麻黄品种进行鉴别。实验结果显示该方法鉴别准确率明显高于只采用相似度模式特征的鉴别,且综合特征的最高分类准确率达 98.25%,表明雷达图表示与相似度特征相结合的方法是一种可行的麻黄品种鉴别方法。

关键词 麻黄, HPLC 指纹图谱, 雷达图, 相似度, 图特征

0 引言

中药麻黄是一种重要的药用植物,在我国有非常悠久的药用历史,古时别名称作龙沙、卑相,指的是麻黄科的三种植物草麻黄(*Ephedra sinica*)、中麻黄(*Ephedra intermedia*)以及木贼麻黄(*Ephedra equisetina*)的干燥草质茎。临床表明,麻黄在中药治疗传染病中起着非常关键的作用。在治疗 SARS 时,麻黄是一味重要的中药,取得了非常好的效果^[1]。麻黄在对病毒性黄疸型肝炎的治疗中取得了满意的疗效^[2]。另外,麻黄植物的挥发油能有效地抑制甲型流感病毒^[3]。麻黄中含有多种生物碱,麻黄碱是其中最有效成分,伪麻黄碱次之,此外,还含有其他一些微量的生物碱,包括甲基麻黄碱和甲基伪麻黄碱以及去甲基麻黄碱和去甲基伪麻黄碱等。目前商品以草麻黄为主流商品,但从麻黄碱含量衡量,则为木贼麻黄为佳,草麻黄次之,中麻黄最次。麻黄的品种不同,它们所含有的有效成分也有很大差别,所以麻黄品种的鉴别非常重要。

随着现代分析技术水平的提高,指纹图谱技术已成为国际公认的控制中药质量的方法。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^[4]。数学分析方法在中药指纹图谱的研究中也得到了广泛应用^[5-7]。但是,这些方法在分析过程中打破了图谱数据的整体性。而指纹图谱特征参数的雷达图表示可以将抽象的数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各种关系^[8]。而且形象的图形特征可能会有利于视觉上的分类或聚类。为得到更好的中药鉴别效果,本文提出了计算指纹图谱相似度和用雷达图表示指纹图谱特征参数相结合的麻黄中药指纹图谱鉴别方法。

1 计算方法

1.1 相似度计算方法

指纹图谱的相似度表示的是指纹图谱之间的相

① 中国博士后科学基金(2012M510722),中国中医科学研究院项目(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:2013-12-27)

似程度。常用的相似度计算方法有夹角余弦法和相关系数法。夹角余弦法通过计算两个指纹图谱数据的夹角余弦计算其相似程度,相关系数法是在指纹图谱数据作标准化处理后再计算其夹角余弦。相似度的值在 0 和 1 之间,越接近 1 表示相似程度越高^[4]。下面给出这两种方法的数学表示。

(1) 夹角余弦法。指纹图谱数据 x_s, x_t 的夹角余弦 $\cos\alpha_{st}$ 表示为

$$\cos\alpha_{st} = \frac{\sum_{i=1}^l x_{is}x_{it}}{\sqrt{\sum_{i=1}^l x_{is}^2 \cdot \sum_{i=1}^l x_{it}^2}} \quad (1)$$

(2) 相关系数法。指纹图谱数据 x_s, x_t 的相关系数常用 r_{st} 来表示,它是数据作标准化处理之后的夹角余弦,表示为

$$r_{st} = \frac{\sum_{i=1}^l (x_{is} - \bar{x}_s)(x_{it} - \bar{x}_t)}{\sqrt{\sum_{i=1}^l (x_{is} - \bar{x}_s)^2 \cdot \sum_{i=1}^l (x_{it} - \bar{x}_t)^2}} \quad (2)$$

1.2 重心图特征提取

中药指纹图谱的雷达图绘制非常简单,对于一个有 p 个指纹特征的图谱,先作一个圆,然后把圆周进行 p 等分;连接圆心和各分点,则得到 p 条半径,分别为各指标的坐标轴,标以适当的刻度,圆心到圆周方向为正向,变量归一化到 $[0,1]$;在 p 条半径上按特征值取值,依次联结得到一个 p 边形。这样我们就得到了一个样本的雷达图。根据实际需要,也可以采用加权雷达图来进行表示。

传统的特征指纹均是在选择的内参照峰基础上建立的。内参照峰是指,在图谱上的出峰时间比较居中,且各个样品中均存在的组分。一般情况下采用叠加的方法来标定,可是标定的标准却不尽相同,而且不能较完全地反映出全部作用峰的共同作用。所以本文基于雷达图的图形特征,提出了重心参照峰这一参数指标。

重心是雷达图的一个重要图形特征。2001 年 Science 杂志^[9]及 2006 年 Nature 杂志^[10]分别有研究论文发表,并指出对于不同的物体,果蝇可通过物体外形重心高度及朝向等进行分辨识别而且能够形成记忆,由此可以说明图形的重心特征可能是一种适于图形识别的重要特征。

重心的表示与求法如下:一个多元信息 R , 包含 n 个变量,从 n 个变量中任意选择 m ($2 \leq m \leq n$) 个变量为 $(r_1, r_2, \dots, r_m)$, 并且它们依次相邻,从而

构成 $m+1$ 边形 $Or_1r_2 \dots r_m$, 如图 1 所示。

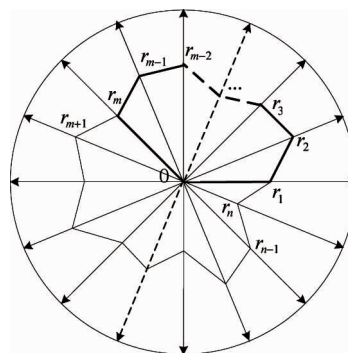


图 1 雷达图的相邻 m 变量重心图特征

根据质量集中在顶点(顶点即为质点)的任意多边形的重心计算公式我们很容易推导出由 $(r_1, r_2, \dots, r_m)$ 相邻 m 变量所构成的多边形的雷达图重心特征的幅值公式为

$$abs_m = f(r_1, r_2, \dots, r_m) = \sqrt{\left(\sum_{i=1}^m p_i r_i \cos \frac{2(i-1)\pi}{n}\right)^2 + \left(\sum_{i=1}^m p_i r_i \sin \frac{2(i-1)\pi}{n}\right)^2} \quad (3)$$

式中 p_i 为相邻 m 变量各个变量信息质点的权值,由各个变量的信息熵函数确定,实际应用中为了方便通常取为 $1/m$ 。重心参数峰融合了全部峰值,它是全部峰值共同作用的结果,更具有可解释性,且其定量表示更加规范。以 t_{Gs} 表示重心参照峰的出峰时间,可以得到相对重心参照峰保留值

$$\alpha' = t_{Ri}/t_{Gs} \quad (4)$$

式中 t_{Ri} 为各组分的出峰时间, t_{Gs} 为重心参照峰的出峰时间。

2 实验结果及分析

2.1 实验数据

实验数据来自文献^[11]表 7-64、表 7-65、表 7-66 的高效液相色谱(HPLC)麻黄指纹图谱数据。

2.2 实验过程

整个研究过程如图 2 所示。

首先对数据进行归一化处理。图 3 所示为 28 个样品归一化后的峰面积雷达图,第 1-18 为草麻黄,19-22 为木贼麻黄,23-18 为中麻黄。由图 3 可见,三种麻黄的雷达图形状各异,直观上很难区分,所以要进行图特征提取。

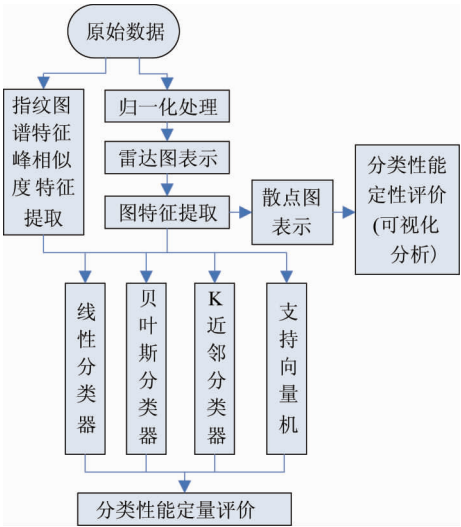


图2 研究整体流程框图

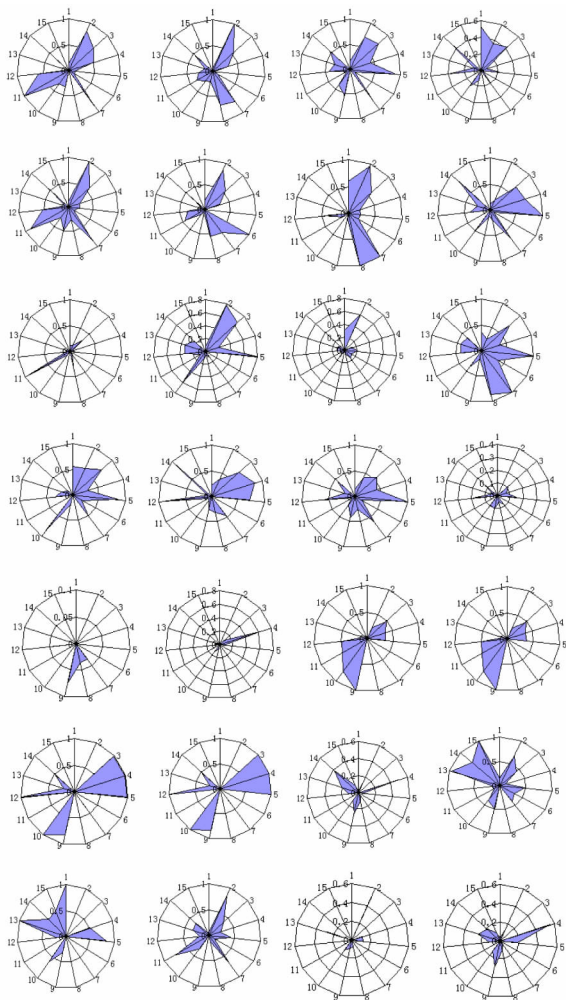


图3 麻黄样本的峰面积雷达图

对28个样品峰面积雷达图进行重心坐标值计算,将计算所得结果用二维散点图来表示,如图4所示。图中“+”为草麻黄,“*”为木贼麻黄,“o”为中

麻黄。从图中可以看出,中麻黄与草麻黄两者分布有部分重合,聚类效果较差,很难区分,而木贼麻黄与草麻黄则相对独立,较易分开。

文献[5]中表7-64、表7-65、表7-66以同一基准计算相关系数及夹角余弦,结果木贼麻黄与草麻黄的相似度测度数据相近,而中麻黄差别比较显著。

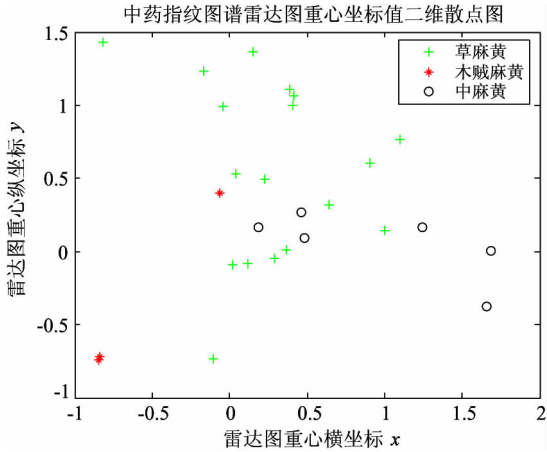


图4 28个麻黄样品峰面积雷达图重心坐标值二维散点图

由以上分析,基于指纹图谱特征峰相似度和雷达图重心特征提取两种方法相结合,采用基于指纹图谱特征峰相似度方法对木贼麻黄与另两种麻黄进行分离,再对另两种麻黄采用雷达图重心特征提取进行分类,应能大大提高麻黄的分类鉴别效果。

分类器选用线性(Ldc)分类器、贝叶斯(Bayes)分类器、k近邻(kNN)分类器、支持向量机(SVM)分类器。用于分类的特征包括相似度特征、重心图特征、融合特征。结果采用10%样本测试100次交叉检验误差估计作为对比评价检验指标。

2.3 实验结果及分析

实验结果如表1所示。采用基于指纹图谱特征峰相似度和雷达图重心特征提取的综合特征进行分类的准确率都明显高于只采用相似度模式特征进行分类的准确率,且综合特征的最高分类准确率达98.25%,这说明了该方法的优越性。

表1 麻黄样本数据分类实验结果

特征 \ 分类器	Ldc	Bayes	3NN	SVM
相似度特征表示	0.6667	0.7933	0.9023	0.9144
雷达图特征表示	0.8741	0.9102	0.9155	0.9186
融合特征表示(相似度特征、雷达图特征)	0.9113	0.9218	0.9825	0.9851

3 结 论

传统的中药麻黄鉴别方法依赖于专业人士的专业知识和经验,通过观察药物的外观来区分不同的品种。这对工作人员的要求较高,依赖于个人的专业水平高低,比较容易出现问题。所以,建立一种准确性高、容易使用的麻黄品种判别方法具有重要的意义,这不仅有利于麻黄药物资源的科学合理利用,同时还能对其他中草药的研究与利用提供有价值的参考。本文采用相似度计算及雷达图表示相结合的方法,研究了不同种类麻黄指纹图谱的差异,建立了一种基于 HPLC 指纹图谱技术的麻黄品种鉴别新方法。主要研究了不同种类麻黄的指纹图谱特征峰,对峰面积分别进行相似度计算和雷达图表示,并对雷达图进行图特征提取最终得到用于比较的雷达图特征。数据实验结果显示,该方法是一种可行的麻黄品种鉴别方法。

参考文献

- [1] 赵春惠. 77 例 SARS 患者中西医结合治疗随机对照研究. 中国中西医结合急救杂志, 2003, 10(4): 197-200
- [2] 赵会忠, 付广臣, 杨福庆等. 从肺论治病毒性黄疸型肝炎 34 例. 中西医结合肝病杂志, 2001, 11(3): 171-172
- [3] 马振亚, 居民建. 麻杏石甘汤对甲型流感病毒等病原微生物的影响. 陕西中医学院学报, 1998, 4: 37-39
- [4] 邹纯才, 鄢海燕. 中药指纹图谱及其数字化. 安徽: 科学技术出版社, 2008
- [5] 赵惠茹, 王锐平, 王燕. 聚类分析法在金银花及其伪品鉴别分类中的应用. 陕西中医, 2005, 26(1): 72-73
- [6] 周璇, 冯毅凡, 郭晓玲. 主成分分析法用于人参皂苷薄层色谱分离的研究. 广东药学院学报, 2003, 19(2): 101-102
- [7] 赵宇, 谢培山, 梁逸曾等. 中药枳壳 HPLC 指纹图谱分析及化学模式识别分类研究. 中国药理学杂志, 2005, 40(11): 55-58
- [8] 洪文学. 基于多元统计图表示原理的信息融合和模式识别技术. 北京: 国防工业出版社, 2008
- [9] Tang S, Guo A. Choice behavior of drosophila facing contradictory visual cues. Science, 2001, 294: 1543-1547
- [10] 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
- [11] 谢培山. 中药质量控制模式的发展趋势. 中药新药与临床药理, 2001, 12(3): 188-191

Identification of ephedra species by radar graph representation and similarity calculation

Cui Jianxin^{**}, Cui Jianfeng^{***}, Hong Wenxue^{*}, Gao Haibo^{****}

(^{*} School of Electric and Electronic Engineering, Yanshan University, Qinhuangdao 066004)

(^{**} Measurement Technology and Instrumentation Key Lab of Hebei Province, Qinhuangdao 066004)

(^{***} Qinhuangdao Institute of Technology, Qinhuangdao 066400)

(^{****} Institute of Acupuncture and Moxibustion China Academy of Chinese Medical Sciences, Beijing 100000)

Abstract

The fingerprint technique was applied to identification of ephedra species, and a novel ephedra species identification method based on fingerprints' radar graph representation and similarity calculation was advanced. The method is characterized by the following. Firstly, the peak areas of ephedra HPLC fingerprints are analyzed by similarity calculation and radar graph. Secondly, the radar graph features are obtained for classification and comparison. Finally, the radar graph features and similarity features are combined to identify the ephedra species. The accuracy of this combining method is better than the identifications using similarity features, and its highest accurate classification rate is 98.25%. Thus, the method combining radar graph representation with similarity features is a feasible ephedra species identification method.

Key words: ephedra, HPLC fingerprint, radar graph, similarity, graph feature