

· 综述 ·

中药指纹图谱相似性评价研究进展

梁宗锁^{1,2}, 杨东风^{1,2}

(1, 西北农林科技大学 生命科学学院, 陕西 杨陵 712100;

2, 陕西省中药指纹图谱与天然产物库研究中心 陕西 杨陵 712100)

关键词: 中药; 指纹图谱; 相似性; 评价指标; 评价方法

中图分类号: R284

文献标识码: A

文章编号: 1004-2199(2006)05-0055-05

中药指纹图谱是一种旨在从整体上对中药进行评价和鉴定的分析技术, 为中药的质量控制提供了一种很好的模式。因此, 其相关研究论文也逐渐增多, 但多集中在方法的建立上, 而针对数据信息处理和图谱相似性评价的研究较少。

1 评价原则 中药指纹图谱的评价应考虑图谱的整体性和模糊性两个方面。整体性是要尽可能包括所有具有指纹意义的特征峰, 同时也要排除溶剂峰和杂质峰。模糊性是着眼于宏观规律的特征分析, 重在辨认图谱的整体面貌, 而不追求细枝末节^[1]。

2 评价指标

相似度的评价方法主要有峰重叠率法(Nei 系数法)、距离系数法、相关系数法、夹角余弦法和峰重叠率与共有峰强度结合法(改进的 Nei 系数法)等方法。聂磊等人把它们分为三种类型: 一是利用指纹图谱间的差异性进行评价, 如距离系数法; 二是利用指纹图谱间的相似性进行评价如夹角余弦法, 相关系数法和 Nei 系数法; 三是在相似性的基础上引入差异性来评价, 如改进的 Nei 系数法^[2]。

2.1 距离系数法

距离系数常用于描述样品间的亲疏程度, 系数越大, 二者的差异也就越大。

2.1.1 欧氏距离 欧氏距离是距离测度中使用最广泛的一种, 公式为:

$$d_{ir} = \left[\sum_{k=1}^m (x_{ik} - x_{rk})^2 \right]^{1/2}$$

其中, x_{ik} 代表第 i 个样品第 k 个特征峰值($k=1, 2, \dots, m$)。 x_{rk} 代表共有模式均值向量第 k 个特征峰值($k=1, 2, \dots, m$), r 代表共有模式均值向量。

欧氏距离在量上反映指纹图谱间的距离和化学成分含量的差异, 侧重于特征变量值的大小亲疏程度的相似性。中药各成分常常作为协同作用的整体达到治疗疾病的目的, 但该系数并没有考虑各指纹峰的相关性和方差的差异性, 有一定的局限性。

2.1.2 斜交空间距离 它采用正交空间距离来计算, 克服了各成分之间相关性的影响。

$$d_{ij} = \left[\frac{1}{m^2} \sum_{k=1}^m \sum_{l=1}^m (x_{ik} - x_{jk})(x_{ik} - x_{jk})r_{kl} \right]^{1/2}$$

其中 m 代表共有峰数, x_{ik} 、 x_{jk} 代表第 i 个样品和第 j 个样品的第 k 个特征峰值, r_{kl} 代表峰 k 与 l 峰之间的相关系数。

2.2 相似系数法

2.2.1 夹角余弦系数 夹角余弦系数是指纹图谱特征变量在变化模式上的相似性测度, 可以提供样品鉴别真伪相似性的信息。其计算公式为:

$$\cos\theta = \frac{\sum_{k=1}^m x_{ik} x_{jk}}{\sqrt{\sum_{k=1}^m x_{ik}^2 \sum_{k=1}^m x_{jk}^2}}$$

其中 m 代表共有峰数, x_{ik} 、 x_{jk} 代表第 i 个样品和第 j 个样品的第 k 个特征峰值

夹角余弦系数从质上判断样品与标准图谱之间的相似性, 识别化学组成在含量比例上的相似性, 值越小, 相似性越大。王龙星等利用向量夹角法, 对 11 个不同产地及炮制方法的吴茱萸样品的液相色谱

收稿日期: 2006-04-04

基金项目: 陕西省中药产业发展基金项目资助(2005-2)

通讯作者: 梁宗锁, liangzs@ms.iswc.ac.cn

谱指纹图谱进行相似度计算,结果此法能较好地评价指纹图谱间的相似性,并清楚地区别了汤洗 7 遍这种炮制方法对指纹图谱的影响^[3]。

2.2.2 相关系数 相关系数对数据座标准化处理后的夹角余弦系数,公式为:

$$r_{ir} = \frac{\sum_{k=1}^m (x_{ik} - \bar{X}_i)(x_{rk} - \bar{X}_r)}{\sqrt{\sum_{k=1}^m (x_{ik} - \bar{X}_i)^2 \sum_{k=1}^m (x_{rk} - \bar{X}_r)^2}}$$

$$\bar{X}_i = \frac{1}{m} \sum_{k=1}^m x_{ik}, \bar{X}_r = \frac{1}{m} \sum_{k=1}^m x_{rk}$$

其中, x_{ik} 代表第 i 个样品第 k 个特征峰值; \bar{X}_i 代表第 i 个样品所有变量的均值; x_{rk} 代表共有模式第 k 个特征变量的值; \bar{X}_r 代表共有模式所有变量的均值。

相关系数多用于评价两个母体如两个不同产地药材之间的相似性,也是从质上判断,识别化学组成在含量比例上的差异,值越小,相似性越大,而对各特征变量值上的变化不敏感。由于它测度的是样品间在特征变量的变化模式上相似形状的相似性,因此又称为形状测度。

2.2.3 指数相似系数

$$r_{ir} = \frac{1}{m} \sum_{k=1}^m \exp \left\{ -\frac{3}{4} \frac{(x_{ik} - x_{rk})^2}{S_k^2} \right\}$$

$$S_k^2 = \frac{1}{n-1} \sum_{i=1}^n (x_{ik} - \bar{X}_k)^2; \bar{X}_k = \frac{1}{n} \sum_{i=1}^n x_{ik}$$

其中 n 表示样品的批数, X_i 代表第 i 批样品。指数相似系数同样是判断样品成分在含量比例上的差异,但它多用于比较同一产地或厂家药材之间的相似性。

2.3 相对关联度

所谓关联度,是指两个系统或两个因素间关联性大小的量度,它描述了系统发展过程中因素间相对变化的情况。相对关联度为: $r = \frac{r_1}{r_1 + r_2}$ 其中 r_1 为相对于最优参考序列的关联度, r_2 为相对于最差参考序列的关联度。设有若干个中药样本,每个样本有若干项指纹特征量化指标,这样组成了评价单元序列。 r 越大,评价单元越佳,可根据各评价单元相对关联度的大小,给出各评价单元的优劣顺序,这样可最终得出中药质量优劣得评价结果。吴忠等人以定义的相对关联度为测度,构建了新的评价中药质量的模式识别模型,适合多组分多指标的中药质量进行综合评价并将其用于连翘质量的评价模型^[4]。

2.4 改进的 Nei 系数法

Nei 系数法只能从质的角度考查指纹图谱的相似性,不能反映指纹图谱峰强度的变化对相似性的影响。基于此,孟庆华等提出峰重叠率和共有峰强度结合算法(改进的 Nei 系数法):

$$NIr = \frac{2N_{xy}}{N_x + N_y} - \frac{2}{N_x + N_y} \sum \left| \frac{h_{1xy} - h_{2xy}}{h_{1xy} + h_{2xy}} \right|$$

其中 NIr 为改进的 Nei 系数值, N_x, N_y 分别为两待比较色谱指纹图谱中实际色谱峰数, N_{xy} 表示两待比较指纹图谱的共有峰数, h_{1xy}, h_{2xy} 表示共有峰强度, $\frac{2N_{xy}}{N_x + N_y}$ 表示两待比较色谱指纹图谱共有峰的重叠率, $\frac{2}{N_x + N_y} \sum \left| \frac{h_{1xy} - h_{2xy}}{h_{1xy} + h_{2xy}} \right|$ 表示共有峰强度差异对相似度降低的程度。

改进的 Nei 系数法考虑了色谱指纹图谱的整体性,把色谱指纹图谱的重叠率和共有峰的峰强度有机结合起来,无须对数据进行标准化处理,计算简单,判断准确,不需要复杂的模式识别方法,且灵敏度高于距离系数和相似系数。若两指纹图谱完全没有共有峰,则 NIr 为 0,若两指纹图谱完全一致,则 NIr 等于 1; NIr 数值一般介于 0~1 之间; NIr 数值的大小能够完全、定量地反映出两个指纹图谱的相似程度。孟庆华等人对文献中的数据预处理,并用距离系数和相似系数进行相似性判断,结果与实际并不相符,而改进的 Nei 系数却能很好的反映实际情况^[5]。聂磊等人以痛必定粉针高效液相色谱指纹图谱为例,通过理论分析,数字模拟以及统计方法从峰强度波动、峰缺失及数据标准化处理等几方面对各种相似度评价方法进行了综合比较。结果显示距离系数和改进的 Nei 系数对峰强度的变化较敏感,Nei 系数和改进的 Nei 系数的对小峰的缺失比较敏感,相似系数和夹角余弦对大峰的缺失较为灵敏,而距离系数在一定范围内无论是大峰还是小峰都表现出较高的敏感性。数据标准化处理对距离系数、相似系数和夹角余弦相似度评价的影响较小,它们比较适合于指纹图谱的模式识别研究^[2]。

峰数弹性、峰比例同态性及峰面积同态性是评价色谱指纹图谱相似度算法的三个指标,峰数弹性是用于评价某相似性测度检测指纹图谱间对应谱峰个数波动能力的指标,它可定义为相似性测度值变化的相对量与指纹图谱间谱峰个数差异的相对量之比。峰数弹性的取值总是取决于小峰增减所得到的

计算值,峰数弹性值越大,表明该测度检测指纹图谱间对应谱峰数差异和小峰的能力越强。峰比例同态性是一个反映相似性测度对指纹图谱间谱峰比例关系变动情况检测能力的评价指标,峰比例同态性数值在 0~1 之间,其数值越大,表示相似度的算法能更好地反映一张指纹图谱上各谱峰比例关系与另一张指纹图谱上各谱峰比例关系间的差异。峰面积同态性数值也在 0~1 之间,数值越大表示相似度算法对指纹图谱中所有谱峰的总面积波动情况检测能力越强。程翼宇等通过数值仿真研究,认为对指纹图谱间小峰数和谱峰比例关系差异检测能力最强的测度是分别是峰匹配度和夹角余弦或相似系数,对各指纹图谱间谱峰总面积波动情况的检测能力较强的是大类成分相似度、指标成分相似度以及欧氏距离等,而欧氏距离测度在三方面都具有较好的评价能力^[6]。孟庆华等人提出利用综合信息指数(改进的 Nei 系数)判断指纹图谱相似性,也用峰数弹性、峰比例同态性及峰面积同态性三个量化评价指标评价综合信息指数,结果综合信息指数对指纹图谱间小峰差异、色谱峰比值的差异、总峰面积的差异都有较强的检测能力,而夹角余弦法和相似系数法只对色谱峰比值的差异具有很强的检测能力。而且无论是以中位值共有模式,还是以均值共有模式为标准,综合信息指数法计算的相似度均能反映出样品的差异^[7]。

3 评价方法

3.1 特征信息的分析与抽提

特征信息分析与提取的主要方法有基于统计的多元分析方法,如聚类、偏最小二乘回归分析、主成分分析等;基于人工智能的信息处理方法,如人工神经网络法;基于数理统计、人工智能和数据仓库技术结合的数据库知识发现方法,如关联规则、遗传算法等。

3.1.1 主成分分析法(PCA) 主成分分析是在损失样本特征值的数量,对多变量进行降维处理的一种数据线性组合方法,它在尽可能保留原有信息的基础上,对数据进行空间转换,将高维空间中的样本转换到较低维的主成分空间中,找出能反映原始数据特征的主成分,通过对各个变量在主成分上的载荷因子大小找出影响结果的重要变量,进一步选择主要表征变量。

该法具有变差最优性、熵达到极小值、相关最优

性、回归最优性等特点。

3.1.2 非线性映照法(NLM) NLM 是在维持原有数据结构的情况下,将多维空间数据点非线性映照在二维平面上的模式分类方法。该法的优点是可以直观地对不同类别的点分类,缺点是其投影图坐标没有明确的物理意义和函数表达式,难以看出每个特征对分类的贡献,且维数太高或样本数太大,收敛困难,迭代容易失败。

3.1.3 小波变换 利用小波变换可以滤除色谱信号中的干扰信号、低频信号和低频信号,以实现背景信号的扣除、基线校正和减小信号的波动,从而把基线信号和色谱峰信号分离,而得到基线平直的色谱信号,同时还可对重叠信号进行解析。袁海龙等使用采用小波变换对茵陈注射液的高效液相色谱基线进行校正,使漂移的基线变得平直,提高了 HPLC 指纹图谱的质量^[8]。

3.2 评价模式

化学识别模式是中药指纹图谱评价的主要模式,它是根据物质所含化学成分用计算机对其进行分类或描述。过程是先建立标准样本模式的色谱指纹图谱,然后对待鉴定样品色谱指纹图谱进行计算机解析,依据它与标准样本模式的隶属度,判别未知模式的真伪及优劣^[9]。

聚类分析识别模式和神经网络法识别模式是较为常用的两种化学识别模式。

3.2.1 聚类分析识别模式 聚类分析是用“相似度”来衡量样品之间的亲疏程度,并以此来实现分类。通常先将待聚类样本集的 N 个样本各自看成一类,然后定义样本间的距离或相似性量度,将相似度大的样本归为一类,相似度小的样本归为不同类,然后开始进行聚类分析。对于不同批次的中药样品,其色谱指纹图经计算机快速辨识处理,可依据样品批与批之间的相似度,确定中药样品批间的稳定性。

色谱指纹图谱常用模糊聚类分析,它是依据样品的特征、亲疏关系程度和相似性,通过建立模糊相似关系对样品进行分类的方法,能反映样品整体的、主要的特性,具有很强的结构性知识的表达能力,可在相当程度上抗干扰与畸变,但准确合理的隶属度函数往往难以建立,一般不具备学习能力。周敬泉等人采用模糊聚类分析方法,提取郁金裂解气相色谱中反映不同样本在化学成分和含量上有差异的信

息特征(保留时间、相对浓度和峰面积等),并对图谱进行综合评价,得到了一系列与相似系数对应的分类,认为此法能比较全面、综合地反映不同产地郁金裂解气相色谱之间的相似关系,为分析和鉴定提供了一种简单快捷的方法^[10]。陈彦等采用高分辨气相色谱—一阶程序升温的方法测定了术类 12 种药材的 Kovats 保留指数及相应组分的相对百分含量,按照模糊聚类分析的数学方法,正确划分了术类药材的种类^[11]。张雪辉等利用高效液相色谱法,结合模式识别技术,对 8 种 1 变种植物不同产地、不同采收日期共计 10 个品种 20 个样品进行系统聚类分析,认为此法方法作为一种辅助手段用于风轮菜屑药材的分类及鉴定是可行的^[12]。

3.2.2 人工神经网络(ANN) 人工神经网络是以数学网络拓扑结构为理论基础,仿照神经网络结构的非线性预测模型,可用于模式识别、数据处理、预测、分析组效关系等,具有超强的处理复杂信息的能力。ANN 法可以获取反映中药内在质量的宏观、综合和隐含的信息,并以模糊非线性观点进行比较,评价中药内在质量的真实性、一致性和稳定性,它在中药质量评价、产地、种属、真伪的识别应用广泛。

蔡煜东等运用人工神经网络法,对中药厚朴气相色谱分析得到的各组分相对含量进行分析评价,得到了相应的评价模式^[13]。苏薇薇等采用反向传播人工神经网络模式识别技术,对 78 个苦丁条样品的高效液相色谱数据进行处理,实现了苦丁样品的计算机快速分类鉴定,结果与生药学鉴定完全一致^[14]。李一波等人采用一种自组织模糊神经网络方法,提出一种基于高效液相色谱的中药材识别模式,整个网络既有神经网络的学习能力和自动聚类能力,还有特征维数不等的模糊识别模式的适应能力。并经大量中药材样本和实际测试表明,此法的网络抗平移、形变和适应新产地药材的能力都很强^[15]。汤丹等运用 B-P 网络法对广藿香 GC-MS 指纹图谱进行解析,建立了广藿香 GC-MS 指纹图谱人工神经网络模型,并通过训练优化的 BP 网络对不同产地广藿香的识别,证明其有较好的识别功能^[16]。

但 ANN 为一个黑箱系统,无法观察中间的学习过程,无法利用已有的专业知识设定初始连接权值,网络参数和输出结果无明确物理意义^[17]。

4 评价研究的不足

中药指纹图谱研究尚处于起步阶段,从最初指纹图谱的建立到多维多息指智能指纹图谱库的应用都存在严重的不足,特别是指纹图谱评价的研究还很混乱,缺乏可靠、规范、统一的评价方法,已成为中药指纹图谱研推广和应用重要障碍。

首先,直接获得的图谱信息数据量大,包含有杂质、溶剂以及许多与指纹研究无关的信息,怎样对这些数据进行预处理,滤除掉无用的信息,最大程度地保留原始特征信息,实现对特征信息的抽提和转换是数据处理方法研究的重点和难点。目前许多处理方法的适用范围不明确,针对不同研究方法获得的指纹图谱,不同处理方法的适用性和灵敏度研究还不够。特别是在特征信息抽提方面,很难判断有用信息和无用信息,主要峰、N 强峰是样品内含量较高的化学成分,应该严格的控制,而对次要峰和无意义峰则允许有较大变化或直接滤去,这对许多有效成分不明确的中药并不一定适用。

其次,指纹图谱评价是对整个图谱面貌和特征指纹峰的综合评价,要兼顾整体性和模糊性两个方面。相似性系数、距离系数和角度余弦系数是当前较为通用的相似性评价指标,它们分别在峰比例、峰强度、峰数等不同的方面对相似性进行评价,但还缺少一个综合的评价指标,尽可能把图谱的所有特征信息有机的统一起来,实现对指纹图谱更准确的判断。改进的 Nei 系数法虽然把图谱的重叠率和共有峰的峰强度有机结合起来,但其有效性尚待进一步确定,而且只统一这两方面的信息还是远远不够的。

参考文献

- [1] 周玉新. 中药指纹图谱研究技术[M]. 北京: 化学工业出版社, 2002, 109.
- [2] 聂磊, 曹进, 罗国安, 等. 中药指纹图谱相似度评价方法的比较[J]. 中成药, 2005, 27(3): 249-252.
- [3] 王龙星, 肖红斌, 梁鑫淼, 等. 一种评价中药色谱指纹谱相似性的新方法: 向量夹角法[J]. 药学学报, 2002, 37(9): 713-717.
- [4] 吴忠, 苏薇薇, 何新新. 中药连翘质量的灰色模式识别研[J]. 中药材, 2000, 23(9): 536-538.
- [5] 孟庆华, 刘永锁, 王健松, 等. 色谱指纹图谱相似度的新算法及其应用[J]. 中成药, 2003, 25(1): 4-8.
- [6] 程翼宇, 陈闽军, 吴永江. 化学指纹图谱的相似性测度及其评价方法[J]. 化学学报, 2002, 60(11): 2017-2021.
- [7] 孟庆华, 刘永锁, 蒋淑敏, 等. 色谱指纹图谱综合信息指数在中药质量控制中的应用研究[J]. 中国天然药物, 2004, 2(6): 359-

- 364.
- [8] 袁海龙,雷长海,肖小河,等.小波变换校正茵陈注射液 HPLC 指纹图谱基线的研究[J].中国新医药,2003,2(9):13-14.
- [9] 吴忠.质量计量学—中药色谱指纹图谱的解析与特征表达[J].中药材,2003,26(8):598-600.
- [10] 周敬泉,颜春兰,袁鹏,等.用模糊聚类法研究中药成分特征谱[J].控制理论与应用,2004,21(4):570-574.
- [11] 陈彦,叶崇义,刘舞霞,等.高分辨气相色谱—模糊聚类分析法在术类药材分类中的应用[J].药学学报,1996,30(3):230-234.
- [12] 张雪辉,陈建民.高效液相色谱—系统聚类分析方法在风轮菜属药材分类中的应用[J].中国中药杂志,2003,28(9):812-816.
- [13] 蔡煜东,官家文,程兆年.中药质量的人工神经网络评价方法[J].中草药,1994,25(4):187-189.
- [14] 苏薇薇,吴忠,何新新,等.中药苦丁茶的化学模式识别研究(II)[J].中药材,1998,21(4):170-173.
- [15] 李一波,黄小原.基于高效液相色谱的中药材模式识别新方法[J].电子与信息学报,2004,26(3):382-388.
- [16] 汤丹,李薇,许毅,等.广藿香指纹图谱解析的人工神经网络方法研究[J].中药材,27(7):534-536.
- [17] 石志红,何建涛,常文保.中药指纹图谱技术[J].大学化学,2004,19(1):33-39.

半夏有性繁殖研究

李西文,张晓柠,刘海涛,马小军*,陈震,赵鑫

(中国医学科学院中国协和医科大学,药用植物研究所,北京,100094)

摘要:目的 研究半夏 *Pinellia ternate* (Thunb) Breit 的有性繁殖方式,为半夏新品种繁育提供依据。方法 采用不同的去雄方式、不同的套袋材料、不同的授粉方式研究半夏的授粉方式。结果 自花授粉和异花授粉实验都有种子发育,但百分比很低。结论 初步认为半夏为常异花授粉,但以自花为主,此为杂交育种的可行性提供了依据。

关键词:半夏;佛焰苞;有性繁殖;授粉

中图分类号:S567

文献标识码:A

文章编号:1004-2199(2006)05-0059-03

半夏 *Pinellia ternate* (Thunb) Breit. 别名又叫麻芋子,三叶半夏,老鸦芋头,地巴豆等。为天南星科多年生草本植物。以块茎入药,生产上常用其珠芽、块茎繁殖,用种子繁殖较少。半夏在自然进化中形成了以株芽、块茎部位无性繁殖为主的生殖特点,由于繁殖系数低而限制了大田生产。在单纯的野生变家栽过程中,尚未形成栽培品种,所以产量一直提不上去。目前,市场对半夏的需求有增无减,野生资源越来越趋于濒危,因此,探讨半夏的有性繁殖,对良种培育具有重要意义。

关于半夏的有性繁殖文献报道较少,据《中国药用植物栽培》^[1]一书描述,半夏为自交不亲和,但从前期实验观察发现自交仍有个体结实现象,为摸清半夏有性繁殖方式,我们进行了初步试验,现将具体情况介绍如下。

1 材料与方 法

1.1 实验材料 半夏块茎来源于四川南充,选用多年生 1.0~1.5 cm 块茎作种。

实验用具和药品:塑料袋,牛皮纸,沙网;显微镜;蔗糖;载玻片;盖玻片;硬纸牌;细线;烧杯 I—KI 溶液;剪刀;细毛笔;磁盘;放大镜;滴管;吸纸;蒸馏水;容量瓶;小量筒;天平;镊子。

1.2 种植方法 2004 年 4 月 2 日平地,浇水,5 日整地,施肥,17 日把种栽播于北京中国医学科学院药用植物研究所实验田内,播种 10 畦,每畦 10 m²,按 0.2 kg/m² 撒播,常规方法进行苗期管理,5 月中旬盛花期时每种处理随机取样 30 株进行有性繁殖

收稿日期:2006-04-3

基金项目:中央级科研院所科技基础性工作专项资金(ZYS-365-01)和四川省重点科技攻关项目(2002-2004)

作者简介:李西文,在读硕士研究生。E-mail:xiweijia2004@yahoo.com.cn

通讯作者:马小军,男,博士,研究员,博士生导师。Tel:(010)62819410, E-mail: xjma@public.bta.net.cn