

应用模糊聚类分析方法划分贵州 含笑属植物区系地理成分

左家哺

(贵州省林业学校, 修文县)

摘要 本文通过贵州含笑属 *Michelia* 植物介绍了植物区系地理成分划分的模糊聚类分析方法和步骤。这种方法较常规分析更能反映其本质特征, 结果比较理想; 此外, 还能准确地反映各类地理成分之间的关系。

关键词 模糊聚类分析; 植物区系地理成分; 含笑属; 贵州

模糊数学自1965年创立以来^[1], 发展迅猛, 各个领域应用广泛, 植物学科亦不例外, 在分类、生态、育种等方面我国学者都已作了研究^[2—5], 但是在植物区系地理成分划分工作中还未曾有报道。为此, 本文以贵州含笑属 *Michelia* 植物为例, 就植物区系地理成分划分的模糊聚类分析进行初步尝试, 就教同行。

一、植物区系地理成分划分的模糊聚类分析方法及步骤

1. 收集植物地理分布资料, 建立其模糊矩阵

对研究范围内所有植物的地理分布资料以某一种类在某一分布区(以行政界线为依据)存在记为1, 相反记为0, 列出植物地理分布数据表(表1, 分布资料引自参考文献^[6])。

设所有植物种类

$u_1, u_2, u_3, \dots, u_i, \dots, u_m$

为论域 $U = \{u_1, u_2, u_3, \dots, u_i, \dots, u_m\}$ (其中: $i = 1, 2, 3, \dots, m$)。

再设m个种类的分布区

$v_1, v_2, v_3, \dots, v_k, \dots, v_n$

为论域 $V = \{v_1, v_2, v_3, \dots, v_k, \dots, v_n\}$ (其中: $k = 1, 2, 3, \dots, n$)。

若 x_{ik} 表示任何一种类 u_i 在分布区 v_k 的分布情况, 由此构成的植物地理分布数据表(表1)就是一个模糊矩阵, 它表示从论域 U 到 V 的一种模糊关系。

2. 建立模糊相似关系

设论域 $U = \{u_1, u_2, u_3, \dots, u_i, \dots, u_m\}$ 为待分类的全体, 且每二个被分类的对象 u_i, u_j 是由二组数据

$x_{i_1}, x_{i_2}, x_{i_3}, \dots, x_{i_k}, \dots, x_{i_n}$

$x_{j_1}, x_{j_2}, x_{j_3}, \dots, x_{j_k}, \dots, x_{j_n}$

(其中: $i, j = 1, 2, 3, \dots, m$) 所表征的。

对于建立 u_i 、 u_j 之间的模糊相似关系的方法常有十三种^[7], 其中大多数不适宜于讨论植物地理分布的二元状态数据, 本文采用绝对值减数法^[7] (下文称方法 1) :

$$r_{ij} = \begin{cases} 1; & \text{当 } i = j \text{ 时} \\ 1 - c \sum_{k=1}^n |x_{ik} - x_{jk}|; & \text{当 } i \neq j \text{ 时} \end{cases} \quad (1)$$

(其中: c 取适当的常数使 $0 \leq r_{ij} \leq 1$) 来作为模糊相似关系的隶属度 (或称隶属函数)。

表 1 贵州含笑属植物地理分布数据表

Table 1. Table of distributional data of *Michelia* of Guizhou

Ui	V _k	X _{ik}	地区 Areas															
			Guizhou	Yunnan	Sichuan	Guangxi	Guangdong	Hunan	Hubei	Zhejiang	Fujian	Henan	Hainan	Jiangxi	India	Vietnam	Burma	Tibet
1	<i>M. champaca</i>		1	1	0	0	0	0	0	0	0	0	0	0	1	1	1	1
2	<i>M. szechuanica</i>		1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
3	<i>M. foveolata</i>		1	1	0	1	1	1	0	0	0	0	1	1	0	0	0	0
4	<i>M. macaciurei</i>		1	0	0	1	1	0	0	0	0	0	1	0	0	0	0	0
5	<i>M. fulgens</i>		1	1	0	1	0	0	0	0	0	0	1	0	0	1	0	0
6	<i>M. cavaleriei</i>		1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
7	<i>M. platyptala</i>		1	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0
8	<i>M. maudiae</i>		1	0	0	1	0	1	0	1	1	0	0	0	0	0	0	0
9	<i>M. martinii</i>		1	1	1	0	0	1	1	0	0	1	0	0	0	0	0	0
10	<i>M. chapensis</i>		1	0	0	1	1	1	0	0	0	0	0	1	0	0	0	0
11	<i>M. angustioblonga</i> ¹⁾		1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	<i>M. skinneriana</i>		1	0	0	1	1	1	0	1	1	0	0	1	0	0	0	0
13	<i>M. crassipes</i>		1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0
14	<i>M. figo</i>		1	0	0	1	1	1	0	1	1	0	0	1	0	0	0	0
15	<i>M. yunnanensis</i>		1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	<i>M. balansae</i> var. <i>appressi pubescens</i> ²⁾		1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
17	<i>M. brevipes</i> ³⁾		1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

1) 见“植物研究”6卷2期, 97—98页; 2) 见“中国树木志”, 第一卷, 490—491页; 3) 见“植物分类学报”, 25卷, 5期, 408—409页;

1) See “Bull. Bot. Res.”, 6(2):97—98; 2) See “Sylva Sin.”, 1:490—491; 3). See “Act. Phytotaxon. Sin.”, 25(5):408—409.

这里笔者按信息论熵表达式^[8]

$$I = -N \sum_{k=1}^p q_k \log_a q_k \quad (2)$$

导出的样方A、B间种的信息系数公式^[9, 10]

$$\Delta I_{AB} = t_{AB} 2 \ln 2 \quad (3)$$

(其中： t_{AB} 为样方A、B中仅有一个样方出现的种数)

再度引伸到建立 u_i 、 u_j 的模糊相似关系之中。

假令 $t_{AB} = t_{ij}$ 表示 u_i 、 u_j 中仅有一个对象存在于分布区 v_k 数之和。那么， u_i 、 u_j 间的信息相异系数为

$$\Delta I_{ij} = t_{ij} 2 \ln 2 \quad (4)$$

由式(4)可算出：当 $t_{ij} = 0, 1, 2, \dots$ 时，有 $\Delta I_{ij} = 0, 1.386, 2.773, \dots$ ；加之， ΔI_{ij} 值为相异性指标^[9]，故式(4)不符合模糊相似关系的定义^[7]，故修改如下：

$$r_{ij} = 1 - M t_{ij} \quad (5)$$

(其中： $M = c \cdot 2 \ln 2$ ， c 取适当的常数使得 $0 \leq r_{ij} \leq 1$)

由式(2)极易导出^[9, 10]，当 $i = j$ 时，式(5)为

$$r_{ii} = 1 \quad (6)$$

综合式(5)、(6)得到 u_i 、 u_j 间模糊相似关系的隶属度：

$$r_{ij} = \begin{cases} 1; & \text{当 } i = j \text{ 时} \\ 1 - M t_{ij}; & \text{当 } i \neq j \text{ 时} \end{cases} \quad (7)$$

不难看出，式(1)和式(7)对二元状态数据而言，其 $\sum_{k=1}^n |x_{ik} - x_{jk}| = t_{ij}$ 。这

样似乎导出的式(7)显得多余，其实不然：1)、式(1)需要将分布资料转化为数值，式(7)则可不必；2)、式(1)是以对象 u_i 、 u_j 按 x_{ik} 、 x_{jk} 在二维空间的绝对距离的补来描述 r_{ij} 的^[9, 10]，而式(7)则是用对象 u_i 、 u_j 仅有一对象存在于分布区 v_k 的概率来描述^[11, 12]。

以式(7)来建立 u_i 、 u_j 间模糊相似关系的方法称为信息系数减数法 (Method of Information coefficients Subtraction) (下文称方法2)。

现对表1以式(1)、(7)即 r_{ij} 为元素构成模糊相似矩阵R

从式(1)(7)即 r_{ij} 的定义和式(8)、(9)可知，R满足自反性和对称性，因此，R表示论域u上的一种模糊相似关系。

通过计算R。 $R \neq R'$ ，即R未满足传递性，非模糊等价矩阵，不能以R对论域u进行分类^[7]。

3. 建立模糊等价矩阵

基于R非模糊等价矩阵而不能对论域u直接进行分类。因此，采用求包含R的最小传递矩阵即传递闭包 $t^+(R)$ 的方法来构造一个模糊等价矩阵。为了节省计算量，采用

$$r_{ij}' = \bigvee_{i=1}^m (r_{ij} \wedge r_{ii}) \quad (10)$$

求R的幂的形式

	1	0.70	0.55	0.60	0.75	0.70	0.55	0.55	0.60	0.55	0.75	0.45	0.60	0.45	0.80	0.70	0.75
	1	0.65	0.70	0.75	0.90	0.75	0.65	0.90	0.65	0.85	0.55	0.70	0.55	0.90	0.75	0.85	
	1	0.85	0.80	0.75	0.80	0.70	0.65	0.90	0.70	0.80	0.85	0.80	0.75	0.80	0.70		
	1	0.85	0.80	0.85	0.75	0.60	0.85	0.85	0.75	0.90	0.75	0.80	0.95	0.85			
	1	0.85	0.70	0.70	0.65	0.70	0.80	0.60	0.75	0.60	0.85	0.80	0.80	0.85			
	1	0.75	0.75	0.80	0.75	0.85	0.65	0.80	0.65	0.90	0.85	0.85					
	1	0.80	0.75	0.90	0.80	0.80	0.95	0.80	0.75	0.90	0.80						
R =		1	0.65	0.80	0.80	0.90	0.85	0.90	0.75	0.80	0.80						
(方法 1)			1	0.65	0.75	0.55	0.70	0.55	0.80	0.65	0.75	(8)*					
			1	0.80	0.90	0.95	0.90	0.75	0.90	0.80							
			1	0.70	0.85	0.70	0.95	0.90	1.00								
			1	0.85	1.00	0.65	0.80	0.70									
			1	0.85	0.80	0.95	0.85										
			1	0.65	0.80	0.70											
			1	0.85	0.95												
			1	0.90													
			1														
	1	0.67	0.50	0.56	0.72	0.67	0.50	0.50	0.56	0.50	0.72	0.39	0.56	0.39	0.78	0.67	0.72
	1	0.61	0.67	0.72	0.89	0.72	0.61	0.89	0.61	0.83	0.50	0.67	0.50	0.89	0.72	0.83	
	1	0.83	0.78	0.72	0.78	0.67	0.61	0.89	0.67	0.78	0.83	0.78	0.72	0.78	0.67		
	1	0.83	0.78	0.83	0.72	0.56	0.83	0.83	0.72	0.89	0.72	0.78	0.95	0.83			
	1	0.83	0.67	0.67	0.61	0.67	0.78	0.56	0.72	0.56	0.83	0.78	0.78				
	1	0.72	0.72	0.78	0.72	0.83	0.61	0.78	0.61	0.89	0.83	0.83	0.83				
	1	0.78	0.72	0.89	0.78	0.78	0.95	0.78	0.72	0.83	0.78						
R =		1	0.61	0.78	0.78	0.89	0.83	0.89	0.72	0.78	0.78						
(方法 2)			1	0.61	0.72	0.50	0.67	0.50	0.78	0.61	0.72	(9)					
			1	0.78	0.89	0.95	0.89	0.72	0.89	0.78							
			1	0.67	0.83	0.67	0.95	0.89	1.00								
			1	0.83	1.00	0.61	0.78	0.67									
			1	0.83	0.78	0.95	0.83										
			1	0.61	0.78	0.67											
			1	0.83	0.95												
			1	0.89													
			1														

* R 为对称矩阵故简写成这样，下文相同。

$$R \rightarrow R^2 \rightarrow R^4 \rightarrow \dots R^{2^h} \rightarrow \dots \quad (11)$$

来计算 $t(R)$ 。经过 $h+1$ 次运算后，一定有

$$R^{2^h} = R^{2^{h+1}}$$

此时， $t(R) = R^{2^h}$ 为传递矩阵，且 R^{2^h} 就是模糊等价矩阵 R^* 。换句话说， R 的 $t(R)$ 就是 R^* 。 R^* 满足传递性，表示一种模糊等价关系，即可用 R^* 对论域 u 实现分类。

对式 (8)、(9) 以式 (10) 即 r_{ii}' 为元素经过 4 次运算后构成的模糊等价矩阵 $R^*[R^* = t(R) = R^8 = R^{16}]$

4. 归类处理即 λ —截矩阵 R_λ 分析

虽然能以 R^* 对论域 u 实现判类但是比较麻烦^[7]，若以

$$r_{ij}(\lambda) = \begin{cases} 1, & \text{当 } r_{ij}' \geq \lambda \text{ 时} \\ 0, & \text{当 } r_{ij}' < \lambda \text{ 时} \end{cases} \quad (15)$$

对 R^* 中的元素进行变换。那么，由 $r_{ij}^{(\lambda)}$ 为元素构成的矩阵称为 λ —截矩阵 R_λ （ λ 为阈值或置信水平），它表示一种普通等价关系；加之， $0 \leq \lambda < r_{ij}' \leq 1$ ，故按 R^* 将论域 u 分成的每一类必定是按 R_λ 将论域 u 分成的类的子类^[7]。因此，不仅可用 R_λ 代替 R^* 对论域 u 的分类，而且，用 R_λ 对论域 u 的判类比用 R^* 更简单明了。

λ 每取一次值，论域 u 中某些对象分为一类或几类，直至最小 λ 值。同时， λ 由 1 降到 0 时，分类越来越粗，逐步形成一类。这一过程中形成了一个动态树状图（或聚类图）（图 1）。至此，模糊聚类分析完毕。

对式(13)(14)按式(15)作 λ -截矩阵分析,其结果如下(第一个数为方法1的取值,括弧内为方法2的取值):

λ 值 [方法 1(2)]

归并类型

1.00(1.00)	$\{u_1\}$, $\{u_2\}$, $\{u_3\}$, $\{u_4\}$, $\{u_5\}$, $\{u_6\}$, $\{u_7\}$, $\{u_8\}$, $\{u_9\}$, $\{u_{10}\}$, $\{u_{11}, u_{17}\}$, $\{u_{12}, u_{14}\}$, $\{u_{13}\}$, $\{u_{15}\}$, $\{u_{16}\}$
0.95(0.95)	$\{u_1\}$, $\{u_2\}$, $\{u_3\}$, $\{u_4, u_7, u_{10}, u_{13}, u_{16}\}$, $\{u_5\}$, $\{u_6\}$, $\{u_8\}$, $\{u_9\}$, $\{u_{11}, u_{15}, u_{17}\}$, $\{u_{12}, u_{14}\}$
0.90(0.89)	$\{u_1\}$, $\{u_2, u_3, u_4, u_5, u_6, u_7, u_8, u_9, u_{10}, u_{11}, u_{12}, u_{13},$ $u_{14}, u_{15}, u_{16}, u_{17}\}$, $\{u_5\}$
0.85	$\{u_1, u_2, u_3, u_4, u_5, u_6, u_7, u_8, u_9, u_{10}, u_{11}, u_{12}, u_{13}, u_{14},$ $u_{15}, u_{16}, u_{17}\}$
(0.83)	$\{u_1\}$, $\{u_2, u_3, u_4, u_5, u_6, u_7, u_8, u_9, u_{10}, u_{11}, u_{12}, u_{13},$ $u_{14}, u_{15}, u_{16}, u_{17}\}$
(0.78)	$\{u_1, u_2, u_3, u_4, u_5, u_6, u_7, u_8, u_9, u_{10}, u_{11}, u_{12}, u_{13}, u_{14},$ $u_{15}, u_{16}, u_{17}\}$

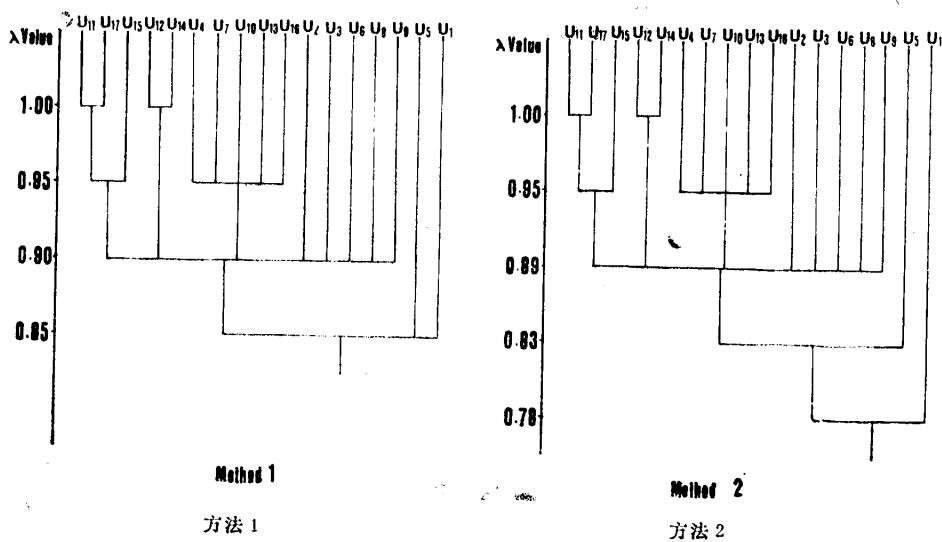


图 1 贵州含笑属植物的模糊聚类分析树状图

Fig. 1 The dendrogram of fuzzy clustering analysis on the species of *Michelia* of Guizhou

二、结果的初步讨论

依贵州含笑属植物区系地理成分的模糊聚类分析结果 [式 (13)、(14) 和图 1] 试讨论如下:

1. 当 $0.85(0.78) \leq \lambda < 0.90(0.89)$ (括弧外的数值为方法 1 的取值, 括弧内数值为方法 2 的取值) 时, 贵州含笑属植物中黄兰 (*M. champaca*) 和亮叶含笑 (*M. fulgens*) 被归为一类, 产我国西南 (华南) 至中南半岛, 为泛北极植物区 (东亚地区) 与古热带植物区 (马来西亚地区) 的联结分布成分^[13, 14, 15]。若按方法 2 的分析, 可分为二

类：前者产东亚的喜马拉雅至马来西亚，后者产东亚的华南（稍延至西南）至马来西亚东部。

2.当 $\lambda = 0.90(0.89)$ 时，贵州含笑属植物中川含笑(*M. szechuanica*)、金叶含笑(*M. foveolata*)、平伐含笑(*M. cavaleriei*)、深山含笑(*M. maudiae*)和黄心夜合(*M. martinii*)被归为一类，它们主产西南、华中（至华南），为西南（云南高原）—华中分布成分^[16, 17]，个别类群延伸到海南（古热带植物区）。

3.当 $\lambda = 0.95(0.95)$ 时，贵州含笑属植物中醉香含笑(*M. macclurei*)、阔瓣含笑(*M. platypetala*)。乐昌含笑(*M. chapensis*)、紫花含笑(*M. crassipes*)和细毛含笑(*M. oalansae* var. *appressipubescens*)被归为一类，它们主产华南至西南、华中（或海南），为华南（包括滇黔桂）至西南（云南高原）成分^[16, 17]，个别成分或向华中或向海南伸延。另外，云南含笑(*M. yunnanensis*)被归为另一类，产贵州（西部威宁和盘县）和云南，为典型的西南（云南高原）成分^[16, 17]，系云贵两省植物区系的重要链锁植物之一。

4.当 $\lambda = 1.00(1.00)$ 时，贵州含笑植物中狭叶含笑(*M. angustioblonga*)和短梗含笑(*M. brevipes*)被归为一类，而野含笑(*M. skinneriana*)和含笑(*M. figo*)被归为另一类。前一类系贵州特有成分^[13, 14, 15]，为新近发现的新种（见“植物研究”6卷2期97~98页；“植物分类学报”25卷5期408—409页）；后一类产贵州至浙江一线（约北纬27°）以南但不到古热带植物区^[13, 14, 15, 18]，西止于广西中部，为滇黔桂至华南的区系成分^[16, 17]。

由上述讨论看出：以模糊等价关系衡量划分地理成分类型的指标比常规分析^[13, 14, 15]更能反映其本质特征，结果比较理想。另外，这种分析还能准确地反映各类地理成分之间的关系，而常规分析是不能做到的。

最后必须指出：掌握了每一个类群的分布资料的情况下，对它们的分布区进行模糊聚类分析（即把它们的分布区视为论域u），即可划分出某一地区各类群的分布区类型^[17]。

参 考 文 献

- 1 Zadeh L A. *Information and Control* 1965; 8:338—353
- 2 徐克学. 植物分类学报 1982; 20:502—509
- 3 赵松岭等. 植物学报 1982; 24:366—373
- 4 杨远攸等. 生态学杂志 1987; 6:20—26
- 5 赵德玉. 模糊数学 1987; (1):39—42
- 6 左家哺. 黔南林业科技 1988; (1): 1—10
- 7 葛苏林. 模糊子集. 模糊关系. 模糊映射. 北京: 北京师范大学出版社, 1985; 31—38
- 8 程明, 朱筱. 信息论讲座. 北京: 科学普及出版社, 1984:11—25
- 9 阳含熙, 卢泽愚. 植物生态学的数量分类方法. 北京: 科学出版社, 1981:38—46; 49—58
- 10 史尼斯 P, 索卡尔 R. 数值分类学（赵铁桥译）. 北京: 科学出版社, 1984:80—84; 92—95
- 11 Goodall D W. *Nature* 1964; 203:1098
- 12 Goodall D W. *Biometrics* 1966; 22:882—907
- 13 武吉华, 张绅. 植物地理学（第二版）. 北京: 高等教育出版社, 1983:85—86; 90—93

- 14 阿略兴 B B (П. В. 库德里亚绍夫改编). 植物地理学 (傅子桢、王燕译). 北京: 财政经济出版社, 1955:31—34; 378—379
- 15 沙菲尔 W. 普通植物地理学原理 (傅子桢译). 北京: 高等教育出版社, 1958:339; 360—362
- 16 吴征镒. 云南植物研究, 1979; 1:1—22
- 17 吴征镒, 王荷生. 中国自然地理·植物地理 (上). 北京: 科学出版社, 1983:29—103
- 18 H. 沃尔特. 世界植被 (李世英等译). 北京: 科学出版社, 1984:80—84; 92—95

致谢 本文承《云南植物研究》编委提出许多有益的建议及宝贵意见供作者修改。

DIVISION OF THE FLORISTIC ELEMENTS OF MICHELIA OF GUIZHOU BY METHOD OF FUZZY CLUSTERING ANALYSIS

Zuo Jiafu

(Guizhou Forestry School, Xiuwen County)

Abstract This paper introduces the method and step of fuzzy clustering analysis that has been used for dividing the floristic elements of *Michelia* of Guizhou, and the floristic elements are discussed in detail. The result shows: 1. The method of fuzzy clustering analysis can reflect more the intrinsic character of the floristic elements than that of the conventional analysis. 2. The method of fuzzy clustering analysis can reflect accurately the relationships among each floristic elements.

Key words Fuzzy clustering analysis; Floristic element; *Michelia*; Guizhou