

模糊聚类在荆条灌丛(*Scrub. Vitex negundo* var. *heterophylla*)分类中的应用*

张金屯

(山西大学生物系)

摘 要

本文应用模糊聚类分析对荆条灌丛分类进行了研究。聚类过程可分为三步:

1. 计算相似矩阵 \tilde{R} : 这一步与其它聚类方法相同,相似系数可有各种选择。
2. 寻找模糊等价关系: 取 \tilde{R} 的乘幂 $\tilde{R}^2, \tilde{R}^4, \tilde{R}^8, \dots$,

若在某一步,有 $\tilde{R}^K = \tilde{R}^{2^K} \triangleq \tilde{R}^*$

\tilde{R}^* 便是一个模糊等价关系。

3. 聚类, 选取适当的置信水平 λ 进行聚类。

本文相似系数采用 $r_{jk} = 1 - \frac{d_{jk}}{M}$ ($j, k = 1, 2, 3, \dots, 41$)

式中 r_{jk} 代表二样方 j 和 k 的相似系数; M 为一适当的常数, 以使 $0 < r_{jk} \leq 1$, 本文 $M = \max D_{jk} + 1$, d_{jk} 表示二样方 j, k 间的距离系数, 用下式计算:

$$d_{jk} = \sqrt{\sum_{i=1}^{30} (X_{ij} - X_{ik})^2}$$

X_{ij} 和 X_{ik} 分别代表原始数据矩阵中第 i 行第 j 列和第 i 行第 k 列元素。

数据为记名样方中30个种的多度数据。结果将41个样方分为11组。根据各样方组各层片的优势植物, 我们将11个样方组定为11个群丛, 其中3个群丛具有明显的过渡性质。这一结果同用组平均聚类法、图论聚类法及以群落结构为主要依据的定性分类方法的结果是吻合的。说明模糊聚类是一个有效的植物群落数量分类方法。

一、引 言

植物群落的分类是植被分析的一个重要方面。从50年代开始, 在植物群落分类中引入数量方法。早期它包括了Curtis & McIntosh (1951), Brown & Curtis, Ramensky (1952), Goodall (1953, 1954), Webb (1954), Poore (1956)等人的工作。以后数量分类的方法和技术不断发展和完善。其中, Williams和Lambert (1959—1961)作了大量工作。1957年Greig-Smith出版了专著《数量植物生态学》(1964年增订重版), 对植物群落数量分类方法的原理和技术作了详尽的论述。进入60年代, 由于数字计

* 本文在傅子祯教授指导下完成。周光裕、陈灵芝等同志提出了宝贵意见, 特致谢意。

计算机的高速发展,大大地促进了植物群落数量分类的发展,近20年来,发表了大量研究论文,新方法和新技术不断涌现。

1965年,美国控制论专家查德(L. A. Zadeh)教授首次提出“模糊集合”的概念,标志着模糊数学的诞生。模糊数学扬弃了排中律,把数学从二值逻辑的基础上转移到连续逻辑,把绝对的“是”“非”变为更加灵活,在适当的界限上去相对地划分“是”与“非”。这在理论上完全适合于植物群落的聚类分析。

本文应用把模糊聚类方法对暖温带荆条灌丛的分类进行了研究。

二、群落与取样

本文所研究的荆条灌丛取自山西南部垣曲县和夏县境内的低山丘陵区,属中条山系。该区位于东径 $111^{\circ}15'$ — $112^{\circ}00'$,北纬 $35^{\circ}00'$ — $35^{\circ}25'$,处在中条山主峰以西。海拔500—1000m,平均气温约 13.3°C ,7月份均温 26.1°C ,1月份均温 -0.8°C , $\geq 10^{\circ}\text{C}$ 的积温 4250.9°C ,无霜期220天左右,全年降水量 667.6mm ,相对湿度60%,属于暖温带季风气候类型¹⁾。

中条山位于山西最南部。此地温暖湿润,雨量充沛,自然地理条件较为优越。因此,植物种类丰富,比之山西其他地区植物群落类型较多,且结构复杂。植被区划仍属于暖温带落叶阔叶林地带。荆条灌丛

是落叶阔叶林被砍伐后的次生植被类型。除荆条外还有一些混生的灌木种类。如黄刺梅(*Rosa xanthina*),小叶鼠李(*Rhamnus parvifolia*),酸枣(*Zizyphus jujuba*),黄栌(*Cotinus coggygria* var. *cinerea*)等。群落垂直结构大致可分为三层。灌木层,荆条占优势;半灌木层,以蒿属(*Artemisia*)植物占优势;草本层,主要种类有白羊草(*Bothriochloa ischaemum*),羊胡子草(*Carex lanceolata*),翻白草(*Potentilla discolor*),黄背草(*Themeda triandra* var. *japonica*),绵枣(*Scilla sinensis*)等。各群丛的层次结构不尽相同。

本文作者于1983年6—9月份采用样方取样方法在野外获得数据。首先确定该群落最小样方面积为 15m^2 ,所以选用 $4 \times 5\text{m}^2$ 的长方形样方,并使长边与海拔梯度方向保持一致。在样方内记录种类成分及灌木半灌木的数量性状。然后再在样方内选一有代表性的小

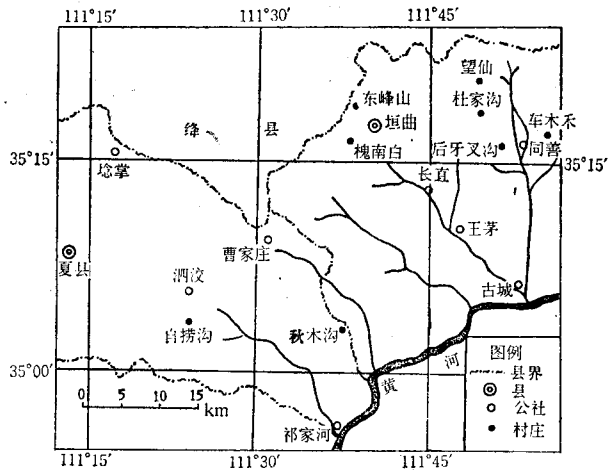


图1 调查地的地理位置

1) 气象资料根据垣曲县气象站 1957—1970 年的统计平均值。本站位于东径 $111^{\circ}12'$, 北纬 $35^{\circ}10'$, 海拔 484.1m 。该站与调查地最为接近。

样方(1×1m²),用以记录草本植物的数量性状。

本研究共取41个样方,分别分布在傅家岭(1—5号样方),后牙叉沟(6—7号),车木禾(8—10号),望仙(11—15号),杜家沟(16—17号),槐南白(18—25号),秋木沟(26—31号),自捞沟(32—33号),东峰山(34—41号)(图1)。

调查所得资料首先用以群落结构为主要依据的传统定性分类方法进行分类,结果得到11个群丛(群丛名称见后),其中3个群丛(Ⅷ、Ⅹ、Ⅺ)具有明显的过渡性质。

三、计算方法

模糊聚类(原理参考汪培庄,1980)的过程一般为以下三步:

1. 计算相似矩阵: 设 U 是需要分类的对象的全体,即为论域,建立 U 上的相似关系 R , $R(u, v)$ 表示 u 与 v 之间的“相似”程度。当 U 为有限集时, R 是一矩阵,叫相似矩阵。这一步与其他聚类方法相同。相似系数的计算,原则上说,可以照搬普通聚类分析的相似系数的确定方法,可有各种选择。

2. 寻找模糊等价关系: 相似关系 R 一般只满足自反性和对称性,不满足传递性,不是模糊等价关系。为此,取 R 的乘幂 R^2, R^4, R^8, \dots ,

若在某一步,有, $R^k = R^{2^k} \xrightarrow{\Delta} R^*$ 则 R^* 便是一个模糊等价关系。

3. 聚类: 选取任一置信水平 λ 进行聚类,得一动态聚类图。根据实际需要,在图上选一适当水平,便得到所需的分类。

由于本文所研究的荆条灌丛,数据庞大,计算复杂,不便在此列出。所以在此用一虚拟数据说明计算过程。假设我们要用4个种对5个样方进行聚类。聚类的第一步是计算相似矩阵,这一步与其他聚类方法相同,所以在此省去原始数据及相似系数计算,直接列出相似矩阵:

$$\tilde{R} = \begin{pmatrix} 1 & 0.1 & 0.8 & 0.5 & 0.3 \\ 0.1 & 1 & 0.1 & 0.2 & 0.4 \\ 0.8 & 0.1 & 1 & 0.3 & 0.1 \\ 0.5 & 0.2 & 0.3 & 1 & 0.6 \\ 0.3 & 0.4 & 0.1 & 0.6 & 1 \end{pmatrix}$$

不难看出, R 满足自反性和对称性,但不满足传递性。所以要对其进行改造。取 R 的乘幂, R^2, R^4, R^8, \dots ,

经计算得 $R^4 = R^8 \xrightarrow{\Delta} R^*$

$$\tilde{R}^* = \begin{pmatrix} 1 & 0.4 & 0.8 & 0.5 & 0.5 \\ 0.4 & 1 & 0.4 & 0.4 & 0.4 \\ 0.8 & 0.4 & 1 & 0.5 & 0.5 \\ 0.5 & 0.4 & 0.5 & 1 & 0.6 \\ 0.5 & 0.4 & 0.5 & 0.6 & 1 \end{pmatrix}$$

取 $\lambda=0.5$, 则

$$R_{0.5} = \left\{ \begin{array}{ccccc} 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 \end{array} \right\}$$

因此, 5 个样方分为两类 $\{x_1, x_3, x_4, x_5, \}$ 和 $\{x_2\}$ 。

取不同的 λ 值, 如此重复, 便可得到一动态聚类图。再根据实际需要, 选取适当的 λ 水平, 便可得到我们所需要的分类。

按照上述计算程序, 我们对荆条灌丛进行模糊聚类分析。本文中每一个实体(样方)的属性, 我们用多度数据刻划。

调查所得数据根据下面原则进行简缩: (1) 淘汰偶见种, 90% 样方中不出现的种给以淘汰。例如糙叶败酱 (*Patrinia scabra*) 仅在一个样方中出现, 给以淘汰; 石防风 (*Peucedanum terebinthaceum*) 在 4 个样方中出现, 但多度很小, 各样方都仅有一个个体, 也给以淘汰。(2) 淘汰建群种, 荆条在 41 个样方中均起建群作用, 尽管在各样方中的多度有所差异, 但这种差异反映在群落关系上不十分明显, 对分类作用不大, 给以淘汰。(3) 淘汰其他灌木种, 从野外实地考察来看, 同一个灌木种的不同个体, 大小悬殊, 盖度差别颇大。盖度小的个体, 在群落中的作用甚微, 可以略去; 盖度较大的个体, 由于多度小, 在整个样方中盖度也不过 5%, 这对群落是有一定作用的, 但这种作用可以反映在对草本层的影响上, 所以我们采用下层植物作为分类依据, 也就反映了它的作用。简缩后得到 30 个植物的多度数据, 故原始数据矩阵为 30×41 维矩阵(见表)。

相似系数(周新远, 1982)用

$$r_{jk} = 1 - \frac{d_{jk}}{M} \quad (j, k = 1, 2, 3, \dots, 41)$$

式中 r_{jk} 代表二样方 j 和 k 的相似系数; M 为一适当的常数, 以使 $0 < r_{jk} \leq 1$, 本文 $M = \max D_{jk} + 1$; d_{jk} 表示二样方 j, k 间的距离系数, 用下式确定

$$d_{jk} = \sqrt{\sum_{i=1}^{30} (X_{ij} - X_{ik})^2} \quad (i = 1, 2, 3, \dots, 30; j, k = 1, 2, \dots, 41)$$

X_{ij} 和 X_{ik} 分别代表原始数据矩阵中第 i 行第 j 列和第 i 行第 k 列元素。

本文计算是在国产 TQ-16 计算机上完成。

四、结果分析

经计算得到模糊等价关系 R^* 。分别取 $1 > \lambda \geq 0.9$, $0.9 > \lambda \geq 0.85$, $0.85 > \lambda \geq 0.8$, $0.8 > \lambda \geq 0.75$, $0.75 > \lambda \geq 0.7$, $0.7 > \lambda \geq 0.65$, $0.65 > \lambda \geq 0.6$, $0.6 > \lambda \geq 0.55$, $0.55 > \lambda \geq 0.49$ 得到动态聚类图(图 2)。

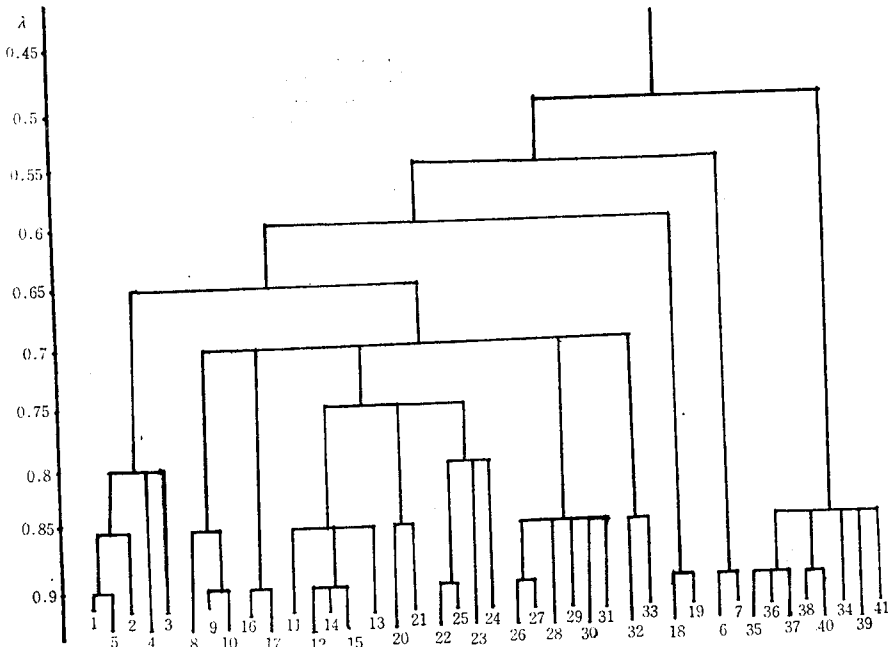


图 2 荆条灌丛动态聚类图

在实际应用中,我们必须把连续值逻辑离散化,按需要分为若干个等级。取定 $0.8 \geq \lambda > 0.75$, 将41个样方分为11个样方组:

{1—5}, {6, 7}, {8—10}, {11—15}, {18, 19}, {22—25}, {26—31}, {34—41}, {16, 17}, {20, 21}, {32, 33}。

按照各样方组各层片的优势种类,我们将11个样方组定为11个群丛,其中最后3个群丛(Ⅸ, X, XI)过渡性质较明显。这与传统的定性分类方法的结果是一致的

群丛 I (1—5), 荆条-野艾蒿+山蒿-圆枝卷柏 (*Selaginella sanguinolenta*)。分布高度700—800m。

群丛 II (6—7), 荆条-白羊草。分布高度550—600m。

群丛 III (8—10), 荆条-山蒿-隐子草+白羊草。分布高度800—850m。

群丛 IV (11—15), 荆条-山蒿+变蒿-黄背草+茅叶荩草。分布高度600—900m。

群丛 V (18—19), 荆条-野艾蒿-白羊草。分布高度600—680m。

群丛 VI (22—25), 荆条-羊胡子草+圆枝卷柏。分布高度750—800m。

群丛 VII (26—31), 荆条-圆枝卷柏。分布高度580—650m。

群丛 VIII (34—41), 荆条-绵枣+羊胡子草。分布高度800—900m。

群丛 IX (16—17), 荆条-变蒿-荻 (*Miscanthus sacchariflorus*)。分布高度 550—600m。

群丛 X (20—21), 荆条-变蒿+山蒿-羊胡子草。分布高度680—730m。

群丛 XI (32—33), 荆条-变蒿+山蒿-翻白草。分布高度730—800m。

聚类效果是满意的,所得结果可以用生态意义加以解释:

1. 水分是制约群丛分布的重要因素。本区位于中条山主峰以西,来自东面的湿润气流受到主峰的阻隔,雨量要少于主峰以东的地区。因此,群落多属旱生类型。例如,群丛Ⅱ、Ⅲ、Ⅴ是典型的旱生类型,群落下层以喜温耐旱的白羊草占优势。群丛Ⅰ、Ⅳ、Ⅹ、Ⅺ也都是偏旱生类型,群落中耐旱的中生植物较多,如蒿属植物。

2. 海拔高度对群丛分布有重要影响。总的来讲,在海拔较高的地方,由于水分条件较为优越,植物群落生长繁茂,覆盖度较大,具有较好的景观,草本层以喜湿的羊胡子草,圆枝卷柏等为主,如群丛Ⅵ、Ⅷ。在海拔较低的地方,多分布着旱生类型,如群丛Ⅱ、Ⅴ。在同一地方,海拔高度对群丛分布的影响更为明显,如在槐南白海拔较低处分布着群丛Ⅴ,为旱生类型;在海拔较高处分布着群丛Ⅵ,为湿中生类型;在二者之间分布着具有过渡性质的群丛Ⅹ,为中间类型。这也说明植被随着环境梯度的变化而呈连续分布。

五、讨 论

植物群落的分类发展到今天,已产生了大量的数量方法,其各有所长,不尽统一。总括起来,有三种实现途径:聚合,分划和排序。模糊聚类属于前者。为了检验模糊聚类效果,本文选用组平均聚类法,图论聚类法进行结果比较。

组平均法(Sokal & Michener, 1958; 阳含熙等, 1981)是常用的一种聚类方法。它既是空间保持,又是单调的,比较起来是较理想的聚合方法。Prichard & Anderson (1971)在全面比较了各种聚合方法后,认为该方法是最有用的方法之一。所以笔者选用这一方法(图3)。取距离系数为22,将41个样方分为10组: {1—5}, {6—7}, {8—10}, {11—15}, {18—19}, {20—21}, {22—25}, {16—33}, {26—31}, {34—41},这与模糊聚类结果基本一致。所不同的仅在于Ⅹ、Ⅺ两个过渡性质明显的群丛被合为一组,这就更加肯定了它们的过渡性质。

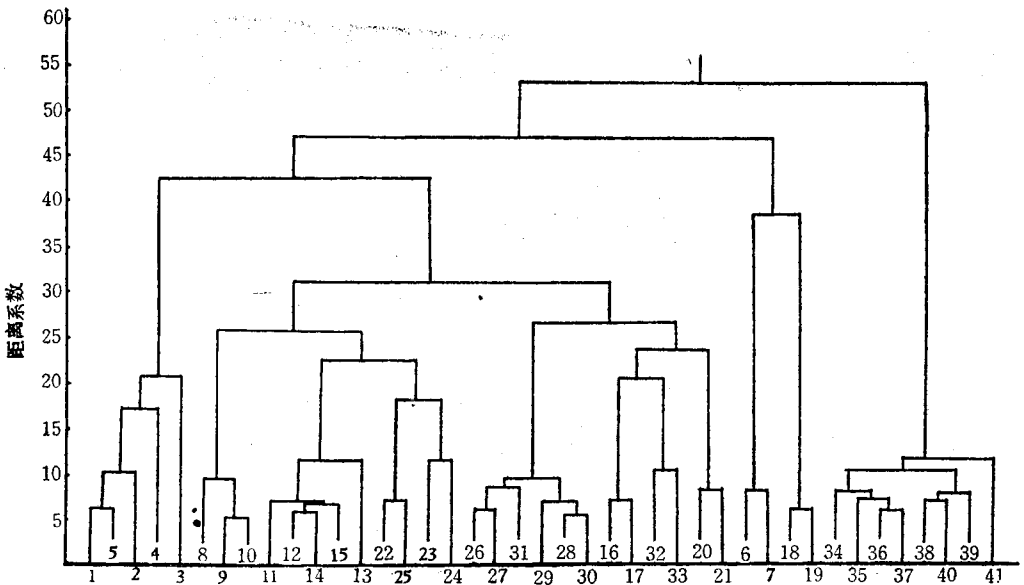


图3 组平均聚类图

图论聚类法 (Cliford et al., 1975; Sneath, Sokal, 1973; 阳含熙等, 1981) 是一种较简单的聚类方法。它是以图论为基础的, 以几何图形表示各样方间的相似关系的。取相似系数为 0.75, 41 个样方也被分为 10 组 (图 4), 与模糊聚类结果基本吻合, 不同的是具有过渡性质的群丛 X 被归入群丛 VII。

由此可以认为, 模糊聚类是一种有效的植被分类方法, 它适合于植被低级单元的聚类分析。模糊聚类分辨能力较强, 它把三个过渡性质明显的群丛都区分得很清楚。但模糊聚类中门坎 λ 值的选择, 需要使用者根据野外

调查的植被资料和经验加以确定。选取阈值越高, 分划出的类群越多, 反之则少。这正说明, 同其它数量分类方法一样, 模糊聚类的应用也是以生态学专业知识和野外工作经验为基础的, 如果忽略了这一点, 就难以得到满意的结果。

与传统方法相比, 模糊聚类有它的优点。它可以处理大量数据, 在差别不大明显的数据中, 经过多次运算, 从中找到亲疏关系, 而得到较客观的分类结果。

使用不同的相似系数, 动态聚类图的结构会有所变化, 也就是聚类结果会有所差异。理论上讲, 各种相似系数均可选择。究竟何种系数较为理想, 尚需进一步探讨。

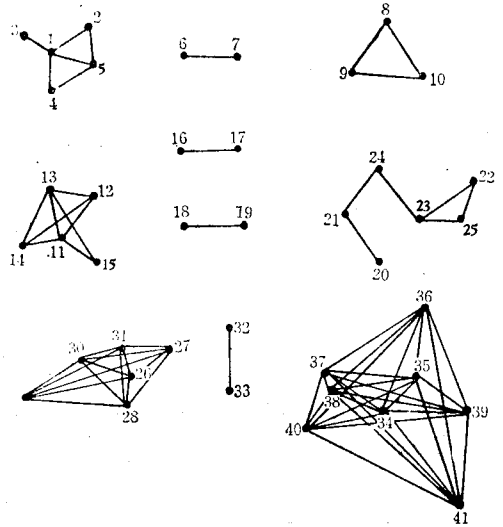


图 4 图论聚类结果图

参 考 文 献

- (1) 阳含熙等, 1981: 植物生态学的数量分类方法。科学出版社。
- (2) 汪培庄, 1980: 模糊数学简介 (I)。数学的实践与认识, (2): 45—59 页。
- (3) 周新远, 1982: 模糊数学在生态学中的应用研究: I、桔全爪螨种群动态的模糊聚类分析。生态学报, 2: 159—170。
- (4) Bray, J. R. & J. T. Curtis, 1957: An Ordination of the upland forest communities of southern Wisconsin. Ecol. Monogr. 27: 325—349。
- (5) Clifford, H. T. & W. Stephenson, 1975: An introduction to numerical classification. Academic Press. New York.
- (6) Greig-Smith, P. 1957: Quantitative plant Ecology. Butterworths, London. 2nd ed. 1964.
- (7) Pritchard, N. M. & A. J. B. Anderson, 1971: Observations on the use of cluster analysis, in botany with an ecological example. J. Ecol. 59: 727—747.
- (8) Sneath, P. H. A. & R. R. Sokal, 1973: Numerical Taxonomy. Freeman San Francisco & London.
- (9) Sokal, R. R. & C. D. Michener, 1958: A statistical method for evaluating systematic relationship. Univ. Kansas Sci. Bull. 38: 1409—1438.

**APPLICATION OF FUZZY MATHEMATICS TO VEGETATION
CLASSIFICATION OF SCRUB. *VITEX NEGUNDO*
VAR. *HETEROPHYLLA***

Zhang Jintun

(Department of Biology, Shanxi University)

Abstract

In this paper, Fuzzy clustering analysis is used to study the scrub of *Vitex negundo* var. *heterophylla*. The procedure of this method is as follows:

1. The calculation of similarity matrix \tilde{R} : This is the same as other approaches. Various similarity coefficients may be used here.

2. The transformation of R to Fuzzy equivalent relation \tilde{R}^* :
Make $\tilde{R}^2, \tilde{R}^4, \tilde{R}^8, \dots$, if $\tilde{R}^k = \tilde{R}^{2k} \triangleq \tilde{R}^*$
then, \tilde{R}^* is a fuzzy equivalent matrix.

3. Clustering: select suitable values of λ and cluster accordingly.

Establishing the similarity matrix \tilde{R} : let $r_{jk} = 1 - \frac{d_{jk}}{M}$ ($j, k = 1, 2, 3, \dots, 41$)

where r_{jk} is the similarity coefficient of both quadrants j and k ; M , the proper constant chosen at will to make $0 < r_{jk} \leq 1$, here $M = \max D_{jk} + 1$; d_{jk} is the distance between j and k , determined by the formula,

$$d_{jk} = \sqrt{\sum_{i=1}^{30} (X_{ij} - X_{ik})^2}$$

where X_{ij} and X_{ik} are the values of the i th row the j th column, and the i th row the k th column respectively in the data matrix.

The data matrix is made up of abundances of 30 species and 41 quadrants. In the result, 41 sampling units are clustered into 11 groups, 11 associations. The result, being identical with the results of group average clustering, graphs clustering and our traditional method, proves that Fuzzy clustering is a useful method for numerical classification of plant communities.