

模糊数学排序及其应用

(5)

张金屯

(山西大学生物系, 太原, 030006)

Q 948.15

Q 14

摘 要

本文应用模糊数学排序法, 对英国威尔士北部山地草甸进行了研究, 结果证明模糊数学排序是一个有效的植被分析方法。

关键词: 模糊子集, 植被分析, 数量方法。

草甸

随着电子计算机的出现和广泛使用, 排序已成为现代植被分析的重要手段之一。在过去30多年中, 排序技术得到了很大发展。由于缺乏统一的检验标准, 至今对各种排序方法尚难作出合理的评价。一般生态学家认为, 只要某种方法在实际应用中所得到的结果符合生态学意义, 就是可取的。

排序是研究植物群落之间及植物与环境间关系的数量分析方法, 其结果是用植物群落(样地)在低维空间的分布图来表示, 从中可以推断出植被类型间的关系及环境因子在这种关系中的作用, 从而可以作出生态学意义的解释。

自从1965年模糊数学诞生以来^[1], 在很多领域中得到了广泛应用^[2,3]。它把经典集合中元素与集合的隶属关系只取{0,1}的二值逻辑, 推广到模糊集合可取(0,1)上任何实数的连续逻辑, 特别是它所定义的隶属函数是用数学方法来刻划和描述定性信息的一种度量, 因而它适用于以研究植被连续性为特点的排序分析。模糊数学排序法是以植物群落种类数据和环境因子数据为基础, 通过模糊集的运算, 得出排序坐标值, 从而实现排序。本文主要介绍模糊数学排序原理和运算过程, 并用英国威尔士北部山地草甸数据对其进行效果检验。

一、方 法

1. 模糊集合及其运算法则

(1) 集合的表示法 模糊数学是以模糊集合论为基础。凡是具有某种共同性质的事物总体{x}称为集合X, 也称为论域, 表示讨论范围。其中每个事物 x_i 称为集合的元素。把论域中一部分元素构成的集合A称为X的一个子集合。比如, 选取10个样地的调查数据为样品, 组成论域:

$$X = \{x_1, x_2, \dots, x_{10}\} \quad (1)$$

本文于1991年3月12日收到。修改稿于1991年10月6日收到。

每个样地调查了3个因子,这样每个样地被视作3维空间上的一个点,即

$$x_i = (x_{i1}, x_{i2}, x_{i3}) \quad (2)$$

组成了样本容量为10,因子数为3的样品集(表1):

$$X = \{x_{ik} | i=1, 2, \dots, 10; k=1, 2, 3\} \quad (3)$$

表1 模拟样地记录数据

Table 1 Artificial site data for ten quadrats

样地 Sample sites	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
海拔高度 Altitude	1000	1400	1100	1150	1450	1250	1300	1500	1450	1150
坡向 Slope direction	N	NW	NE	N	S	SE	S	E	SE	NE
坡度 Slope	10	20	22	24	30	25	21	24	22	15

若只考虑海拔1300m以上的样地,其集合为:

$$A = \{x_{ik} | x_{ik} \geq 1300, i=2, 5, 7, 8, 9; k=1, 2, 3\} \quad (4)$$

称A为X的一个子集合。

在经典集合论中,元素与集合之间的关系,要么 $x \in A$,要么 $x \notin A$,二者必居其一。这种元素与集合之间的绝对隶属关系,称为特征函数,记作 χ ,通常用1表示真,用0表示假,记作 $\{0, 1\}$,称为二值逻辑,通过各元素的特征函数与集合中元素的对应关系来刻画一个集合。如 x_2 样地是在1300m以上,即 $x_2 \in A$ 取真值,即 $\chi_A(x_2) = 1$,据此得:

$$A = \left\{ \frac{0}{x_1} + \frac{1}{x_2} + \frac{0}{x_3} + \frac{0}{x_4} + \frac{1}{x_5} + \frac{0}{x_6} + \frac{1}{x_7} + \frac{1}{x_8} + \frac{1}{x_9} + \frac{0}{x_{10}} \right\} \quad (5)$$

可以定义一个线性隶属函数来描述每个样地的海拔高度对于高海拔集合B隶属程度上的差异,这个程度可用 $[0, 1]$ 的任何实数来表示。

设以海拔高度为论域,取 $X = [1000, 1500]$,则所谓高海拔是X上的一个模糊子集B,其隶属函数定义为:

$$\mu_B(x) = \begin{cases} 1, & x > 1500; \\ 1 - \frac{1}{500}(1500 - x), & 1000 \leq x \leq 1500; \\ 0, & x < 1000. \end{cases} \quad (6)$$

则有:

$$B = \{0.0, 0.8, 0.2, 0.3, 0.9, 0.5, 0.6, 1.0, 0.9, 0.3\} \quad (7)$$

同法,可以定义其它因子的模糊子集及其隶属函数。

(2) 模糊集的运算 集合运算是由给定的一个或两个集合,通过并、交、差、余运算得出新集合的一种方法。设有A、B两个模糊子集,其运算公式^[1,4]为:

$$\text{并集} \quad \mu_{A \cup B}(x) = \max(\mu_A(x), \mu_B(x)) \quad (8)$$

$$\text{交集} \quad \mu_{A \cap B}(x) = \min(\mu_A(x), \mu_B(x)) \quad (9)$$

$$\text{余集} \quad \mu_{\bar{A}}(x) = 1 - \mu_A(x) \quad (10)$$

$$\text{差集} \quad \mu_{A-B}(x) = \min(\mu_A(x), \mu_{\bar{B}}(x)) \quad (11)$$

对于模糊数学排序来说,还需计算一特殊子集,叫反换差集(Anticommutative Difference)^[4],其公式定义为:

$$\mu_{A\bar{B}}(x) = [1 + (\mu_{\bar{B}}(x))^2 - (\mu_{\bar{A}}(x))^2] / 2 \quad (12)$$

反换差集可以允许定义更复杂的模糊子集, 这些子集与选定的环境梯度有相似的特点。

(3) 模糊关系 模糊关系是指同一论域内两个元素间的相互关系。在植被生态学中, 一个非常重要的关系是群落(样地)间的相似关系。它是一特殊的模糊关系^[2,4], 描述两个植物群落间的相似程度, 其值一般变化于[0, 1]闭区间。模糊数学排序需要计算群落相似关系, 本文用夹角余弦法求得:

$$R_{ij} = \frac{\sum_{k=1}^m X_{ik} \cdot X_{jk}}{\sqrt{\left(\sum_{k=1}^m X_{ik}^2\right)\left(\sum_{k=1}^m X_{jk}^2\right)}} \quad \begin{array}{l} i, j=1, 2, \dots, n(\text{样地数}) \\ k=1, 2, \dots, m(\text{种群数}) \end{array} \quad (13)$$

这里 R_{ij} 是样地 i 和 j 间的相似系数, X_{ik} 和 X_{jk} 表示第 i 和第 j 个样地 k 个植物种的观测值。

2. 模糊数学排序

根据以上原理, 可以定义一系列模糊子集, 选取适当的模糊子集作为排序轴, 即可得到排序图。首先计算样地间模糊相似系数矩阵 R , 然后选一环境梯度, 比如海拔高度。接着定义以下5个模糊子集: (1) 高海拔样地模糊子集 A , 它的隶属函数用前面提到的线性方法加以定义。(2) 低海拔样地模糊子集 B , B 是 A 的余集。(3) 相似于高海拔样地模糊子集 C , 其隶属函数为:

$$\mu_C(x) = \frac{\sum_{y \sim x} (R_{xy}(\mu_A(y)))}{\sum_{y \sim x} (\mu_A(y))} \quad (14)$$

其中 $\mu_C(x)$ 是样地 x 在模糊子集 C 中的隶属度, $\mu_A(y)$ 是样地 y 在 A 中的隶属度, R_{xy} 是样地 x 和 y 的相似系数。

(4) 同法定义相似于低海拔样地模糊子集的新子集 D :

$$\mu_D(x) = \frac{\sum_{y \sim x} (R_{xy}(\mu_B(y)))}{\sum_{y \sim x} (\mu_B(y))} \quad (15)$$

这里 $\mu_D(x)$ 是样地 x 在模糊子集 D 中的隶属度, $\mu_B(y)$ 是样地 y 在 B 中的隶属度。

(5) 计算模糊子集 C 和 D 的反换差集(12式), 记作 E 。 E 是一个相似于高海拔样地模糊子集而不相似于低海拔样地模糊子集的新模糊子集。

用模糊子集 E 和 A 分别作为 x 轴和 y 轴, 可以得到一个二维排序图。这里 y 轴是海拔高度由低向高的梯度序列; x 轴是由相似于低海拔到相似于高海拔的梯度序列, 二轴有着较高的相关性。

同样可以选取另一环境梯度, 计算出相应的5个模糊子集, 得到另一排序图。也可以把两个或多个环境因子结合起来, 完成排序, 这样排序结果能更好地反映植被与环境的关系, 因为它包括更多的信息。

二、植被数据

本文所用数据取自英国威尔士北部Snowdonia山地草甸。这里条件优越，草甸植被发育良好，除山顶部和一些裸岩较多的地方外，群落覆盖度多在95%以上，许多地方达100%。

在该区沿三条位于不同山坡上的样带取74个 $5 \times 5\text{m}^2$ 的样地，同一样带上的两个相邻样地垂直间隔20m。在样地中记录植被和环境数据，并取土样，以备化学分析。本文仅用盖度(5级制：1<5%，2=6%—20%，3=21%—50%，4=51%—75%，5=76%—100%)、海拔高度和土壤有机质含量数据。

74个样地先用双向指示种分析法(TWINSPAN, Two-way Indicator Species Analysis)进行分类。结果74个样地被分成6个植被类型：

- I. *Eriophorum vaginatum* + *Juncus squarrosus*;
- II. *Nardus stricta* + *Festuca ovina* + *Juncus squarrosus*;
- III. *Festuca ovina* + *Nardus stricta*;
- IV. *Festuca ovina* + *Galium saxatile*;
- V. *Agrostis capillaris*;
- VI. *Calluna vulgaris* + *Vaccinium myrtillus*.

三、结果分析与讨论

首先取海拔高度为环境梯度，依前面所述方法计算出5个模糊子集A、B、C、D和E。以模糊子集A和E分别为x和y轴，得一排序图(略)。这里x坐标值是样地的相对海拔高度，y坐标值反映群落组成上与海拔高度的关系。这一排序图虽然也反映了各样地之间的关系，但仅依据海拔高度这一环境因子，其含信息量小，分辨率较低，使一些植被类型混合在一起，尤其是两个相似程度较高的植被类型II和III在图上交错分布。所以，这一结果不十分满意。

为了得到更好的结果，再选一新的环境梯度——土壤有机质含量。同法求得5个新模糊子集，记作 A_1 、 B_1 、 C_1 、 D_1 、 E_1 。为了把海拔高度和土壤有机质二因子结合起来，现以模糊子集 E_1 和E分别为x和y轴组成排序(图1)。这一结果包含信息量大，较好反映了植被类型间的关系，是比较满意的。植被类型II和III在种类组成和环境特点上都比较相似，二者分布在图的中间，非常接近。但二者又被明显区分开来，说明这一排序分辨率较高。由图1可知，植被分布与环境因素之间密切相关。沿x轴土壤有机质由低到高，沿y轴，海拔高度逐渐上升。植被分布与这两个因子之间的关系是不言而喻的。另外，这一排序图还明显反映出土壤水分的分布规律。从图的左下角向右上角，群落土壤水分逐渐增高。出现在图左下角的植被类型VI由*Calluna vulgaris*占优势，它分布于海拔低，土层薄，裸岩多的地带，土壤保水力差，较干燥。类型V分布海拔低的山麓，虽然土层较厚，但排水性能好，相对也较干燥。而位于图右上方的类型I多出现在地势平缓、低洼积水之地，土壤水分丰富，有机质含量很高。类型II和III土壤水分状况介于中间水平，位于图的中间。

图2是6个主要优势种在排序图上的分布，图中数值为种的盖度值。由图2可知，模糊

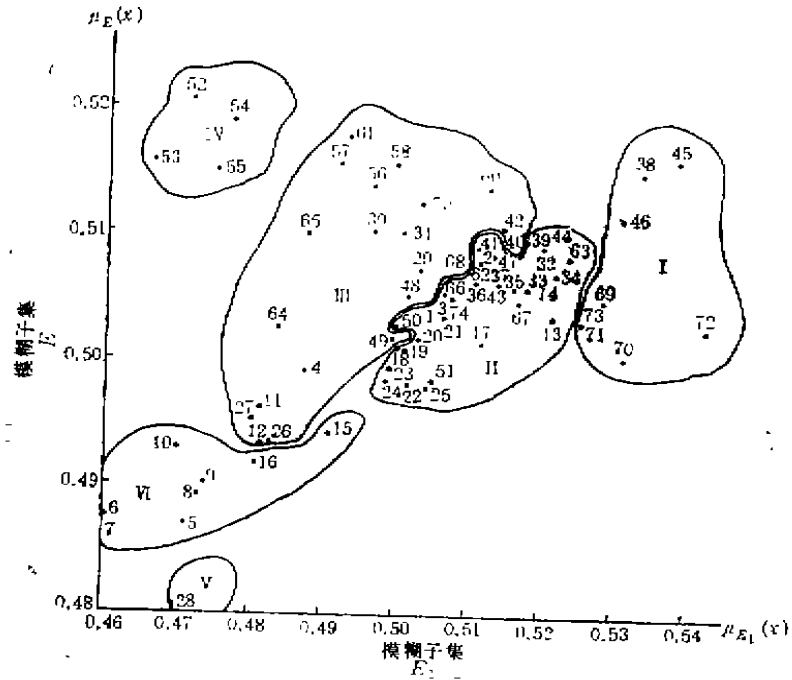


图 1 模糊数学排序图(模糊子集 E_1 为 x 轴, E_2 为 y 轴)。I, II, ..., 代表植被类型。
Fig.1 Fuzzy set ordination diagram, x axis = $\mu_{E_1}(x)$, y axis = $\mu_{E_2}(x)$.

数学排序很好地反映了植物种类的分布规律及种类分布与环境间的关系。每个种在图上都有其特定的分布中心和范围, 它的分布和由它占优势所形成的植被类型的分布相吻合, 但不完全相同。图 2 也较好地反映了植物种间的关系。比如 *Festuca* 在本区分布最广, 几乎见于所有样地中, 说明它能与大多数种类共同生存, 适应不同的环境条件; 而种类 *Juncus* 的出现则与 *Nardus* 有较大的正相关。这一点也说明了排序的合理性。

从本文结果看, 模糊数学排序较好地描述了植物群落之间及植被与环境间的关系, 是一个有效的排序方法。其分辨率较高, 可以较好地与分类方法结合使用, 其本身也可以作为一个植被分类方法^[5]。

与其它排序方法相比, 模糊数学排序有以下几个特点: (1) 它是一个直接梯度分析方法, 而其它大多数排序方法为间接梯度分析^[6-8]。也就是模糊数学排序把环境数据和植被组成数据直接结合起来。这样模糊数学排序实际上是一个分析生态关系的过程, 一旦排序完成, 各种生态关系都十分清楚。而其它排序过程大多数只是数学计算过程, 当排序完成后, 再用生态学知识试图解释排序轴所包含的生态意义^[4]。(2) 模糊数学排序是一个综合信息的过程。当完成一个模糊排序后, 一些重要的生态关系可能会更明显, 从而提示研究者进行新的排序分析。在环境因子选择上, 当一个环境因子不能给出满意结果时, 可以选择第二个, 第三个……, 也可以把由单独环境因子所确立的模糊子集进行综合而形成更复杂、包含更多信息的新模糊子集。所以, 模糊数学排序可以综合大量信息。这一特点, 使它的低维空间(二维或三维)排序能够包含大量生态信息, 这正是所有排序方法试图做到的。(3) 在数学

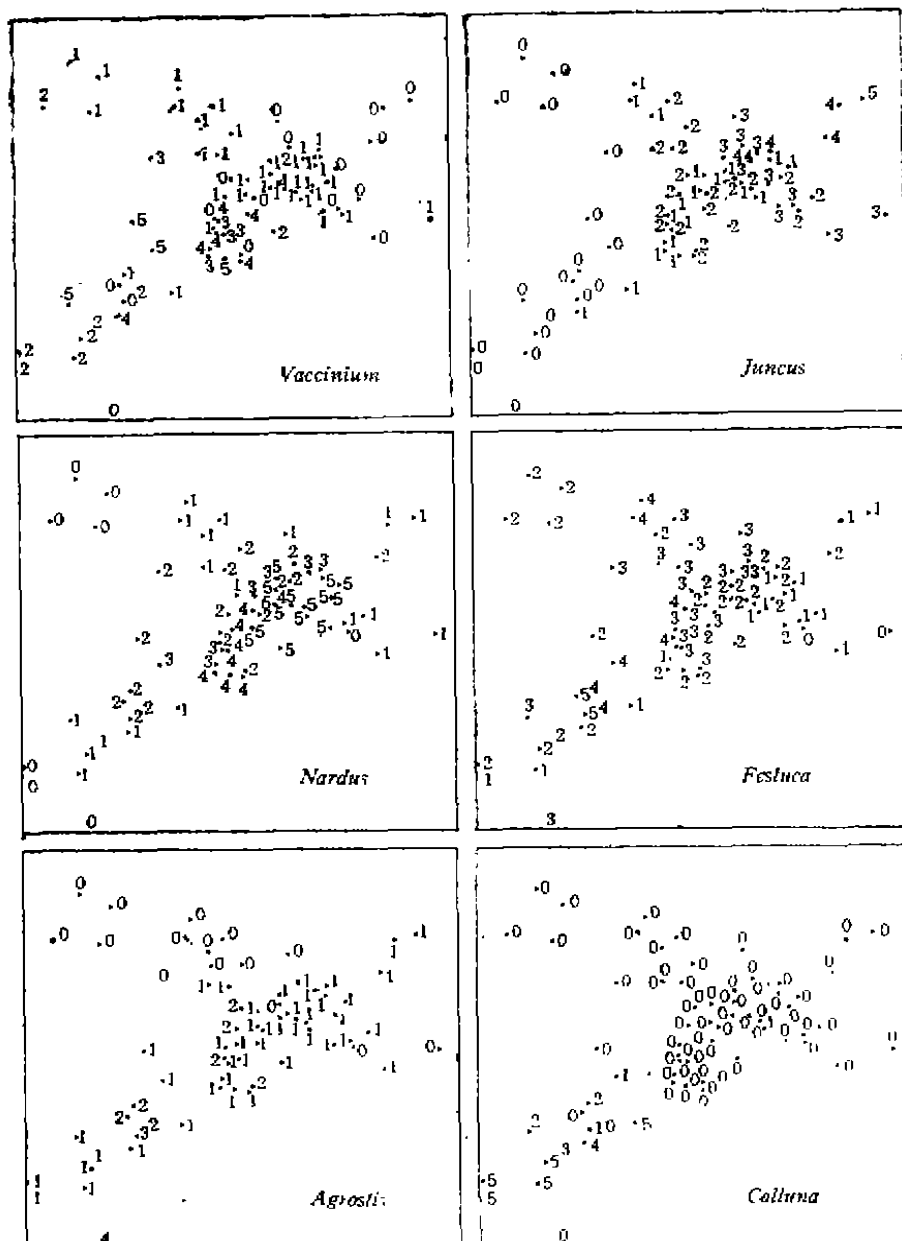


图 2 6个主要优势植物种在排序图上的分布
Fig.2 Distribution of six dominant species on fuzzy set ordination diagram

原理上, 模糊数学排序是以集合理论为基础的, 而大多数其它排序则是依据于几何理论^[4]。

另外, 模糊数学排序要求主观选择生态因子, 研究者需要根据所研究的植被类型、研究地的生态特点、环境差异性等等来综合考虑。当环境梯度比较明显时, 可能一个或两个因子就可以得到较满意的排序。象本文用两个因子, 结果已相当满意。如果环境因子差异不十分明显, 可以考虑多个因子。这一过程需要生态学专业知识和野外工作经验。

参 考 文 献

- [1] Zadeh, L. A. *Fuzzy Sets. Inf. Control.* 1965, 8:338—353.
- [2] 贺仲雄, 《模糊数学及其应用》, 天津科学技术出版社, 1983
- [3] 张金屯, 模糊聚类在荆条灌丛分类中的应用, *植物生态学与地植物学丛刊*, 1986, 9(4):306—314.
- [4] Roberts, D. W. Ordination on the basis of fuzzy set theory. *Vegetatio*, 1986, 66:123—131.
- [5] 阳含熙、卢泽愚, 《植物生态学的数量分类方法》, 科学出版社, 北京, 1981.
- [6] Gauch, H. G. *Multivariate Analysis in Community Ecology*, Cambridge University Press, New York, 1982.
- [7] Greig-Smith, P. *Quantitative Plant Ecology*. 3rd Eds. Blackwell Scientific Publications, London 1983.
- [8] Whittaker, R. H. *Ordination and Classification of Communities*. The Hague. Junk, 1973

FUZZY SET ORDINATION AND ITS APPLICATION

Zhang Jin-Tun

(Department of Biology, Shanxi University, Taiyuan, 030006)

Fuzzy set ordination technique is introduced and applied to a mountain grassland study in Snowdonia, North Wales, in this paper. The results suggest that fuzzy set ordination is an effective and time-saving method.

Key words: fuzzy set, vegetation analysis, numerical methodology.