

非度量多维测度及其在群落 分类中的应用*

余世孝

(中山大学生物学系, 广州 510275)

摘要 排序与聚类分析法是近代最为常用的植被数量分析方法。原理上, 排序较之于聚类法, 一般具有较为严格的数学基础。在植被研究的应用方面, 排序比聚类法能更好反映植被的连续性。同时, 在反映植被分类结果上, 排序图不仅象聚类树形图一样, 使全部实体之间的划分关系得以反映, 而且两两实体间的关系可以通过排序图上彼此间的距离比较而得以较好的反映。然而常规的排序方法主要适用于具有线性结构的数据分析, 在 2—3 维排序图上常难以充分反映这些实体的生态学关系, 造成大量生态学数据信息的损失。非度量多维测度(Non-metric multidimensional scaling)是近期发展起来的适用于非线性数据结构分析的一种复杂的迭代排序方法。它的基本思想是通过排序, n 个实体在尽可能低维($t < n$)排序空间上之间的距离与实体之间的实际相异性一致。本文在较为详尽介绍这一多元分析方法的同时, 以海南岛霸王岭热带山地植被的群落分类为例, 说明非度量多维测度技术在植被生态学中的应用。结果显示, 其排序结果明显优于其它分析方法。同时, 非度量多维测度具有能处理大量的数据且常能以低维排序图反映结果的优点, 是一种具有广泛应用前景的排序技术。

关键词 非度量多维测度; 排序; 群落分类; 热带森林; 海南岛

排序(Ordination), 旨在较低维度的几何空间上以点表示各分析实体之间的关系, 它与聚类分析法(Cluster analysis)一起, 是近代最为常用的植被数量分析方法。原理上, 排序较之于聚类法, 一般具有较为严格的数学基础。在植被研究应用方面, 排序比聚类法能更好反映植被的连续性。同时, 在反映植被分类结果上, 排序图不仅象聚类树形图一样, 使全部实体之间的划分关系得以反映, 而且两两实体间的关系可以通过排序图上彼此间的距离测度而得以较好的反映, 如图 1(a) 中反映实体 1, 2, 3 的聚类树形图, 根据树形图, 我们可知实体 1 和 2 之间的相似程度高于实体 1 和 3 或 2 和 3。但如何比较实体 1 和 3 以及实体 2 和 3 间的相似性? 通过聚类树形图的观察, 我们无法得出结论。相反地, 这种相似性可以通过排序图上各点间的欧氏距离的比较而得出结论(图 1(b))。

常规的排序方法, 如主分量分析(PCA), 主坐标分析(PRA), 相互平均分析(RA), 应用于生态学数据的分析, 具有一定的实际意义(王伯荪等, 1987; 1988; 余世孝等, 1986; 张宏达等, 1989), 但应用于复杂的生态学数据, 也必然存在着一定的不足之处, 主要地表现在这些排序方法, 适用于具有线性结构的数据分析, 而复杂的生态学数据集, 更常呈现出

本文于 1993 年 5 月收到, 1993 年 9 月收到修改稿。

* 国家教委基金资助研究项目。程序设计与数据处理为作者于 1991—1992 年在加拿大访问期间完成, 承蒙 L. Orlóci 教授热情帮助, 谨致谢意。

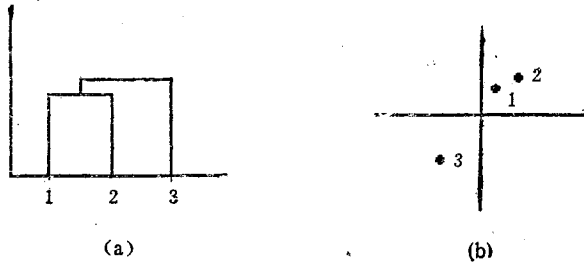


图 1 反映三个实体间相似关系的聚类树形图(a)与排序图(b)

Fig. 1 Similarity among three entities represented by cluster analysis (a) and ordination(b)

非线性关系。通常,排序结果常通过 2—3 维排序图来表示实体间的关系,在如此低维的空间上常难以充分反映这些实体的生态学关系,造成大量生态学数据信息的损失。

在适用非线性结构数据的排序方法中,非度量多维测度被认为是一种较有效的可取方法。这一方法,最初由 Shepard, R. N. (1962) 提出,常用的则是经 Kruskal, J. B. (1964) 修改的方法。但是涉及这一领域的有关文章,主要仍在方法介绍与理论探讨上,它的实际应用,尤其在生态学领域的应用,有关报道不多,特别在我国,有关这一方法在理论上的介绍或其实际应用,迄今都未有详细的报道。本文的目的,在于介绍这一多元分析方法,并以实例说明其在群落分类上的应用。

1 非度量多维测度法

根据 Shepard 和 Kruskal 等人的观点,这一方法的基本思想可以这样归纳:假设有 n 个实体 $1, 2, \dots, n$, 实体 i 和 j 之间的相异(似)性为 δ_{ij} , 将其列为半矩阵, 实体自身的相异(似)性略去, 则共有 $m = n(n-1)/2$ 个相异(似)值。假设这 m 个值按序值从小到大排列:

$$\delta_{i_1 j_1} \leq \delta_{i_2 j_2} \leq \delta_{i_3 j_3} \leq \dots \leq \delta_{i_m j_m}$$

若为相似值则按序值从大到小排列。非度量多维测度的目的就是用 t 维($t \leq n$) 欧氏空间上的 n 个点 x_1, x_2, \dots, x_n , 来表示这 n 个实体, 而且点间欧氏距离大小:

$$d_{ij} = \sqrt{\sum_{s=0}^t (X_{is} - X_{js})^2}$$

顺序与相异值序相一致。

非度量多维测度的计算过程可以归纳如下:

1. 基于原始数据, 计算 n 个实体两两之间的相异(似)值 δ_{ij} 。
2. 随机或人为地确定 t 维欧氏空间上的 n 个初始排序点 x_1, x_2, \dots, x_n , 将数据加以标准化, 并计算 n 个点间的欧氏距离 d_{ij} 。
3. 以欧氏距离 d_{ij} 为 x 轴, 相异值 δ_{ij} 为 y 轴, 作出 m 个对应值的二维图形(图 2)。
4. 确定出 m 个新的排序点 $(\hat{d}_{ij}, \delta_{ij})$, 它们满足:

$$\hat{d}_{i_1 j_1} \leq \hat{d}_{i_2 j_2} \leq \hat{d}_{i_3 j_3} \leq \dots \leq \hat{d}_{i_m j_m}$$

的条件。 \hat{d}_{ij} 值的确定是依据相异值 δ_{ij} 以及点间距离 d_{ij} , 通过一较为复杂逐步回归过程而得出(Kruskal, J. B., 1964)。

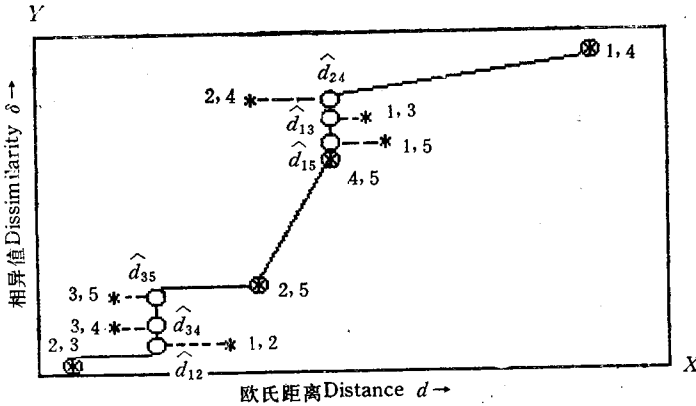


图 2. m个对应值的二维排序图

Fig. 2 Relationship between m points of original configuration and m new points to be fitted

*: 初始排序点 Original configuration (d_{ij}, δ_{ij}) ○: 迭代排序点 To be fitted ($\hat{d}_{ij}, \delta_{ij}$)

5. 如何反映由(2)确定的 n 个初始排序点在多大程度上表示了 n 个实体的实际差异, Kruskal 提出了一胁迫(stress)系数:

$$s = \sqrt{\frac{\sum_{i < j} (d_{ij} - \hat{d}_{ij})^2}{\sum_{i < j} d_{ij}^2}}$$

当 $s > 20\%$, 那么这种吻合较差; 如果为 $10-20\%$, 吻合一般; 如果为 $5-10\%$, 吻合较好; 若 $s < 5\%$, 吻合极好。

在确定了排序维度 (t 值) 以后, 非度量多维测度的目的应使得胁迫系数 s 最小, 也即通过逐步逼近改变初始排序点, 重复步骤 2—5, 使 s 尽可能的小。大致上讲, 如果 $\hat{d}_{ij} < d_{ij}$, 则使得点 x_i 和 x_j 靠近, 相反则移开, 从而使得 \hat{d}_{ij} 接近 d_{ij} 。

归纳起来, 非度量多维测度技术的步骤可以用图 3 来表示。

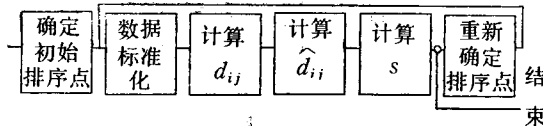


图 3 非度量多维测度技术的计算步骤

Fig. 3 Illustration of algorithm of non-metric multidimensional scaling

上述计算过程由作者以 C++ 语言编成计算机程序 NMMDS, 并收入作者所编制的“植被结构分析”计算软件 GINGKO U1.0¹⁾ 中“聚类与排序”模块。排序图可由 GINGKO U1.0 中程序 SCATTER (在“作图”模块) 打印。

1) Yu Shixiao, Orlóci L., Structural analysis of uegetation, SPB Academic Publishing bu, The Hague (in press).

2 实例

本文将以海南岛霸王岭热带山地植被的群落分类为例, 说明非度量多维测度技术在植被生态学研究中的应用。

热带森林植被的分类, 尤其在低级植被分类上的划分, 由于群落中物种组成的丰富性和常有不明显的优势种或建群种, 以及植被分布的连续性, 群落之间的分界线常不明显, 因此比其它地区的植被分类远为困难。

海南岛霸王岭长臂猿自然保护区, 位于海南岛昌江县境内与白沙县交界处, 整个林区约当北纬 $N18^{\circ}50' - 19^{\circ}05'$, 东经 $E109^{\circ}05' - 109^{\circ}25'$, 面积约 72000ha, 其中于 1980 年建立的长臂猿 (*Hylobates concolor*) 保护区, 覆盖有约 2500ha 的以原始潮湿热带的常绿而带有季节性的森林。

有关霸王岭山地植被的概况, 已有过报道 (余世孝, 1989; 余世孝等, 1993; 陆阳等, 1986)。用于本文分析的取样地, 包括了在蚂蟥冈的样地区 SP2B 的 12 个样地, 保护站北侧

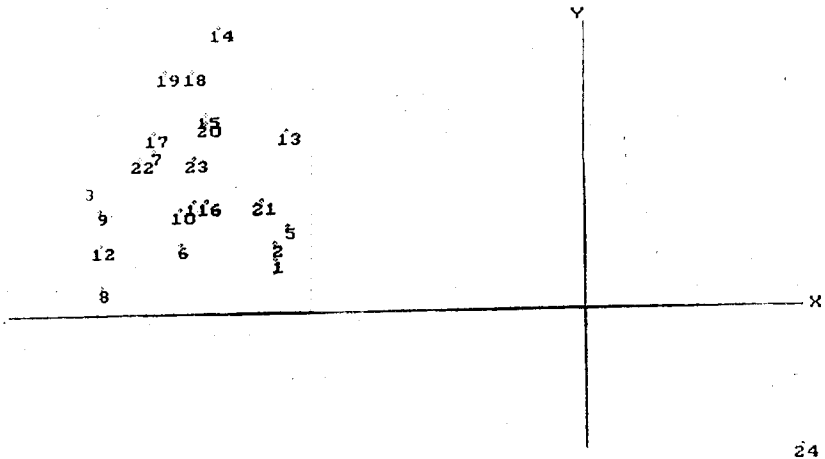


图 4 24 个样地的非度量多维测度二维排序图

Fig. 4 Scatter of 24 sampling units. Two axes were extracted in non-metric multidimensional scaling.

的样地区 SP2A 的 11 个样地, 每一样地面积为 50m × 50m, 并进一步划分为 25 个 10m × 10m 样方, 在海拔约 1340m 的第三斧头岭样地区 SP3, 设置了 1 个 50m × 50m 样地。

各样地统计了树干胸径大于 10 cm 的植株, 共 277 种, 计算了各物种在各样地的重要值, 表 1 按顺序列出了各样地前 5 个物种的重要值。

在此基础上, 将物种的重要值列成 277 × 24 维矩阵, 作为进行非度量多维测度分析的基本数据 (排序维数 t 值取 3)。同时, 用于比较的目的, 进行了主分量分析和聚类分析 (离差平方和

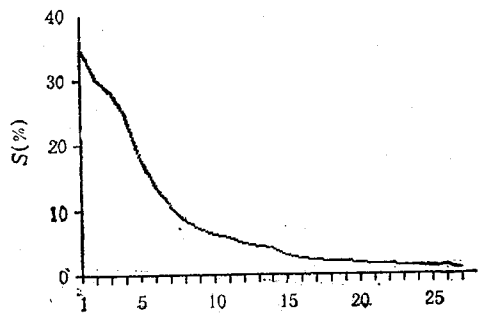


图 5 胁强系数 S 值随着迭代计算过程的递减状况

Fig. 5 Declining stress value during the computations

表1 各样地前5个物种的重要值(每一样地面积为50m×50m)

Table 1 The importance values of the first five species in each sampling unit at Bawangling Nature Reserve. The sample unit size is 50m×50m

No.	1	2	3	4	5
1	鸡毛松 25.434 <i>Podocarpus imbricatus</i>	线枝蒲桃 18.216 <i>Syzygium araiocladum</i>	鸭脚木 18.098 <i>Schefflera octophylla</i>	白颜树 14.464 <i>Gironniera subaequalis</i>	饭甑青冈 10.477 <i>Quercus fleuryi</i>
2	鸡毛松 24.766 <i>Podocarpus imbricatus</i>	线枝蒲桃 16.664 <i>Syzygium araiocladum</i>	白颜树 15.985 <i>Gironniera subaequalis</i>	海南水团花 12.680 <i>Adina hainanensis</i>	卵叶樟 11.896 <i>Cinnamomum ovatum</i>
3	线枝蒲桃 37.261 <i>Syzygium araiocladum</i>	黄叶树 22.267 <i>Xanthophyllum hainanense</i>	油丹 20.809 <i>Alseodaphne hainanensis</i>	米椎 19.825 <i>Castanopsis carlesii</i>	陆匀松 18.527 <i>Dacrydium pierrei</i>
4	线枝蒲桃 54.602 <i>Syzygium araiocladum</i>	五列木 31.252 <i>Pentaphylax euryoides</i>	黄叶树 24.986 <i>Xanthophyllum hainanense</i>	陆匀松 20.026 <i>Dacrydium pierrei</i>	泡花树 12.698 <i>Meliosma squamulata</i>
5	海南水团花 19.191 <i>Adina hainanensis</i>	岭南青冈 15.861 <i>Quercus championii</i>	线枝蒲桃 14.567 <i>Syzygium araiocladum</i>	油丹 13.026 <i>Alseodaphne hainanensis</i>	药用狗牙花 11.385 <i>Ervatamia officinalis</i>
6	线枝蒲桃 40.804 <i>Syzygium araiocladum</i>	白颜树 27.781 <i>Gironniera subaequalis</i>	厚壳桂 16.622 <i>Cryptocarya chinensis</i>	白花含笑 14.276 <i>Michelia mediocris</i>	黄叶树 11.287 <i>Xanthophyllum hainanense</i>
7	线枝蒲桃 29.625 <i>Syzygium araiocladum</i>	陆匀松 28.861 <i>Dacrydium pierrei</i>	黄叶树 23.103 <i>Xanthophyllum hainanense</i>	油丹 18.618 <i>Alseodaphne hainanensis</i>	五列木 17.456 <i>Pentaphylax euryoides</i>
8	厚壳桂 31.392 <i>Cryptocarya chinensis</i>	岭南青冈 20,558 <i>Quercus championii</i>	线枝蒲桃 20,245 <i>Syzygium araiocladum</i>	白颜树 17.758 <i>Gironniera subaequalis</i>	海南韶子 13.299 <i>Nephelium lappaceum</i> var. <i>topengii</i>
9	线枝蒲桃 33.446 <i>Syzygium araiocladum</i>	黄叶树 24.036 <i>Xanthophyllum hainanense</i>	油丹 23.256 <i>Alseodaphne hainanensis</i>	米椎 7.097 <i>Castanopsis carlesii</i>	鸭脚木 13.959 <i>Schefflera octophylla</i>
10	线枝蒲桃 29.355 <i>Syzygium araiocladum</i>	黄叶树 25.970 <i>Xanthophyllum hainanense</i>	短穗棕柯 18.142 <i>Lithocarpus fenzelianus</i>	白花含笑 14.952 <i>Michelia mediocris</i>	五列木 12.956 <i>Pentaphylax euryoides</i>
11	线枝蒲桃 24.606 <i>Syzygium araiocladum</i>	白花含笑 13.633 <i>Michelia mediocris</i>	黄叶树 13.158 <i>Xanthophyllum hainanense</i>	厚壳桂 11.183 <i>Cryptocarya chinensis</i>	隐脉红淡比 10.763 <i>Cleyera obscurinervia</i>
12	白颜树 27.273 <i>Gironniera subaequalis</i>	厚壳桂 22.207 <i>Cryptocarya chinensis</i>	线枝蒲桃 18.783 <i>Syzygium araiocladum</i>	黄叶树 18.079 <i>Xanthophyllum hainanense</i>	荔枝红豆 16.186 <i>Ormosia semicastrata</i>
13	线枝蒲桃 25.040 <i>Syzygium araiocladum</i>	陆匀松 24.969 <i>Dacrydium pierrei</i>	米椎 16.218 <i>Castanopsis carlesii</i>	凸脉冬青 13.706 <i>Ilex kobuskiana</i>	黄叶树 13.336 <i>Xanthophyllum hainanense</i>

续表 1

No.	1	2	3	4	5
14	陆匀松 55.566 <i>Dacrydium pierrei</i>	线枝蒲桃 23.593 <i>Syzygium araiocladum</i>	黄叶树 21.023 <i>Xanthophyllum hainanense</i>	厚皮香八角 18.092 <i>Illicium ternstroemioides</i>	异株木樨榄 11.925 <i>Olea divoca</i>
15	线枝蒲桃 35.352 <i>Syzygium araiocladum</i>	陆匀松 30.852 <i>Dacrydium pierrei</i>	厚壳桂 18.042 <i>Cryptocarya chinensis</i>	饭甌青冈 15.048 <i>Quercus fleuryi</i>	腺叶山矾 13.983 <i>Symplocos adenophylla</i>
16	黄叶树 21.924 <i>Xanthophyllum hainanense</i>	油丹 19.424 <i>Alseodaphne hainanensis</i>	线枝蒲桃 17.341 <i>Syzygium araiocladum</i>	白颜树 12.320 <i>Gironniera subaequalis</i>	密脉蒲桃 9.810 <i>Syzygium chunianum</i>
17	线枝蒲桃 38.796 <i>Syzygium araiocladum</i>	密花树 27.424 <i>Rapanea nerifolia</i>	黄叶树 18.817 <i>Xanthophyllum hainanense</i>	密脉蒲桃 16.314 <i>Syzygium chunianum</i>	厚壳桂 13.581 <i>Cryptocarya chinensis</i>
18	陆匀松 43.045 <i>Dacrydium pierrei</i>	黄叶树 25.356 <i>Xanthophyllum hainanense</i>	线枝蒲桃 24.949 <i>Syzygium araiocladum</i>	油丹 14.458 <i>Alseodaphne hainanensis</i>	隐脉红淡比 12.112 <i>Cleyera obscurinervia</i>
19	陆匀松 52.082 <i>Dacrydium pierrei</i>	线枝蒲桃 36.320 <i>Syzygium araiocladum</i>	黄叶树 20.366 <i>Xanthophyllum hainanense</i>	厚壳桂 12.860 <i>Cryptocarya chinensis</i>	隐脉红淡比 10.666 <i>Cleyera obscurinervia</i>
20	陆匀松 34.514 <i>Dacrydium pierrei</i>	线枝蒲桃 33.420 <i>Syzygium araiocladum</i>	黄叶树 14.916 <i>Xanthophyllum hainanense</i>	海南肖榄 13.727 <i>Platea hainanensis</i>	厚皮香八角 12.096 <i>Illicium ternstroemioides</i>
21	线枝蒲桃 36.829 <i>Syzygium araiocladum</i>	鸡毛松 22.155 <i>Podocarpus imbricatus</i>	斜脉暗罗 15.000 <i>Polyalthia plagioneura</i>	油丹 12.433 <i>Alseodaphne hainanensis</i>	子凌蒲桃 9.599 <i>Syzygium championii</i>
22	线枝蒲桃 26.463 <i>Syzygium araiocladum</i>	黄叶树 16.715 <i>Xanthophyllum hainanense</i>	厚壳桂 15.302 <i>Cryptocarya chinensis</i>	短穗棕柯 14.328 <i>Lithocarpus fenzelianus</i>	凸脉冬青 13.387 <i>Ilex kobuskiana</i>
23	黄叶树 24.469 <i>Xanthophyllum hainanense</i>	线枝蒲桃 23.081 <i>Syzygium araiocladum</i>	陆匀松 19.289 <i>Dacrydium pierrei</i>	白花含笑 11.366 <i>Michelia mediocris</i>	腺叶山矾 10.096 <i>Symplocos adenophylla</i>
24	钝叶水丝梨 111.970 <i>Sycopsis tutcheri</i>	碎叶蒲桃 23.861 <i>Syzygium buxifolium</i>	线枝蒲桃 21.330 <i>Syzygium araiocladum</i>	黄杞 14.155 <i>Engelhardtia roxburghii</i>	隐脉红淡比 11.842 <i>Cleyera obscurinervia</i>

法)。在进行 PCA 分析的情况下,由于计算限制,删除了频率低于10%的物种,计138种。计算程序及排序图、聚类图的打印依据 GINGKO U1.0。

非度量多维测度分析的结果以二维排序图表示(图4)。迭代计算过程中肋强系数 s 值的递减状况如图5所示。依据 s 值,可见排序结果是令人满意的。据此,将霸王岭山地植被中的二个植被亚型划分为五个群落类型:

I. 热带山地雨林

1. 陆均松 (*Dacrydium pierrei*) + 黄叶树 (*Xanthophyllum hainanense*) - 线枝蒲桃 (*Syzygium araiocladum*) 群落: 包括样地 7, 13-15, 17-23, 为霸王岭山地植被占优势的代表群落类型。
2. 油丹 (*Alseodaphne hainanensis*) - 线枝蒲桃 (*Syzygium araiocladum*) 群落: 包括样地 3, 4, 8, 9, 12。
3. 厚壳桂 (*Cryptocarya chinensis*) + 线枝蒲桃 (*Syzygium araiocladum*) - 光叶白颜 (*Gironniera subaqualis*) 群落: 包括样地 6, 10, 11, 16。
4. 鸡毛松 (*Podocarpus imbricatus*) - 线枝蒲桃 (*Syzygium araiocladum*) 群落: 包括样地 1, 2, 5。

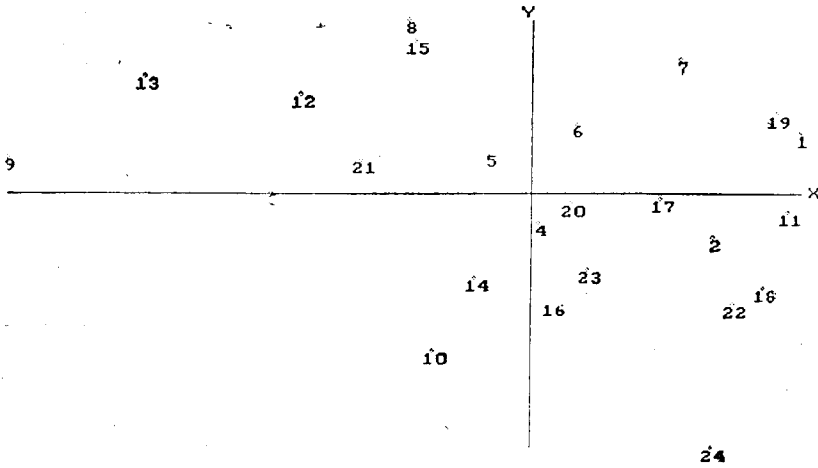


图 6 24个样地的主分量分析二维排序图

Fig. 6 Scatter of 24 sampling units. Two axes were extracted in principal components analysis, they accounted for 40.326% of total variance

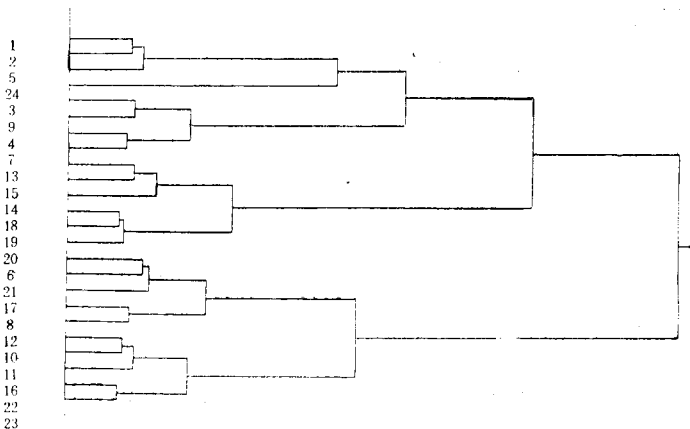


图 7 24个样地的聚类树形图

Fig. 7 Dendrogram of sampling units

II. 热带山地常绿林

5. 钝叶水丝梨(*Sycopsis tutcheri*) + 碎叶蒲桃(*Syzygium buxifolium*) + 线枝蒲桃(*Syzygium araiocladum*)群落: 代表样地 24。

主分量分析的二维排序图和聚类树形图也分别用图 6 和图 7 表示。在主分量分析的情况下, 二维的排序图所反映的信息较低 (40.372%), 因此无论是在二个植被亚型, 即热带山地雨林(陆均松 + 线枝蒲桃群系)和热带山地常绿林之间的区别上, 还是在热带山地雨林中的进一步划分, 都明显比不上非度量多维测度分析结果。聚类分析法则没有区分出霸王岭山地植被的二个植被亚型, 而这一点应该是整个山地植被划分上的最重要方面。

3 结论

与其它常规的排序方法比较, 非度量多维测度由于不涉及矩阵中特征根的求解等过程, 因此相对能处理较大量的数据。

与同样可基于相异矩阵进行排序的主坐标分析比较, 非度量多维测度由于利用的是秩信息, 因此无论相异矩阵的来源如何, 是定量的或定性的, 只要能将各实体之间的关系按相异(似)值的顺序列出, 就可以进行排序分析。

要求事先确定排序维数(r 值)从而达到良好的排序效果, 即胁强系数 s 值很小, 也许是非度量多维测度的一个弱点。因为同时获得较低的胁强系数与较低的排序维数两者常互相矛盾。所以, 一般是在计算量允许的条件下, 通过不同排序维数情况下的胁强系数比较来加以判断。通常是在符合要求的胁强系数情况下取尽可能少的排序轴。

参 考 文 献

- 王伯荪等, 1988: 香港岛森林群落的聚类与排序, 植物生态学与地植物学学报, 12: 1—11。
 王伯荪等, 1987: 香港岛森林群落的数量分析——模糊聚类, 相互平均排序与 R 型主分量分析, 中山大学学报, (4)66—73。
 余世孝等, 1986: 千里香在中国的地理分布及其限制因素, 生态科学, (1)20—26。
 余世孝, 1989: 海南霸王岭热带山地雨林群落种群分布, 生态科学, (2)118—123。
 余世孝等, 1993: 海南霸王岭热带山地植被研究 I: 永久样地设置与群落类型, 生态科学, (2)。
 余世孝等, 1993: 海南霸王岭热带山地植被研究 II: 代表群落的结构分析, 生态科学, (2)。
 陆阳等, 1986: 海南岛霸王岭长臂猿自然保护区植被, 植物生态学与地植物学学报, 10: 106—114。
 张宏达等, 1989: 香港植被, 中山大学学报论丛, 8(2)1—171。
 Kruskal, J. B., 1964: Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 29: 1—27。
 Kruskal, J. B., 1964: Nonmetric multidimensional scaling: A numerical method. *Psychometrika* 29: 115—129。
 Shepard, R. N., 1962: The analysis of proximities: Multidimensional scaling with an unknown distance function. *Psychometrika* 27: 125—139, 219—246。

THE APPLICATION OF NON-METRIC MULTIDIMENSIONAL SCALING IN COMMUNITY CLASSIFICATION

Yu Shi-xiao

(Department of Biology, Sunyatsen University, Guangzhou 510275)

Abstract

Ordination is one of the most popular multivariate analysis methods used in vegetation research. However, most of the commonly used techniques involved in this aspect are concerned with linear data set that is difficult to find in the ecological world. In this paper, non-metric multidimensional scaling, a non-linear ordination technique put forward in Shepard and Kruskal, is introduced and its algorithm is programmed in the C++ computer language as program NMMDS, and it is included in the computer software Ginkgo VI.O used for structural analysis of vegetation. An example with the classification of tropical montane forest in Bawangling Nature Reserve, Hainan island, is presented. Also other multivariate analysis methods, such as principal components analysis, cluster analysis, are also used for comparison. The result shows that non-metric multidimensional scaling is suitable for higher species richness community classification, and much better than principal components analysis. With this, two sub-vegetation types in Bawangling montane forest have been classified. The tropical montane rain forest is represented by *Dacrydium pierrei* + *Syzygium araiocladum* community, while the tropical montane evergreen forest is represented by *Sycopsis tutcheri* + *Syzygium araiocladum* community. Also five associations among them are recognized.

Key words Non-metric multidimensional scaling, Ordination, Community classification, Tropical forest, Hainan Island