

ISSN 1000-0933
CN 11-2031/Q

生态学报

Acta Ecologica Sinica



第33卷 第11期 Vol.33 No.11 **2013**

中国生态学学会
中国科学院生态环境研究中心
科学出版社

主办
出版



中国科学院科学出版基金资助出版

生态学报

(SHENGTAI XUEBAO)

第 33 卷 第 11 期 2013 年 6 月 (半月刊)

目 次

前沿理论与学科综述

- 新一代 Landsat 系列卫星: Landsat 8 遥感影像新增特征及其生态环境意义 徐涵秋, 唐 菲 (3249)
- 两种自然保护区设计方法——数学建模和计算机模拟 王宜成 (3258)
- 家域研究进展 张晋东, Vanessa HULL, 欧阳志云 (3269)
- 浅水湖泊生态系统稳态转换的阈值判定方法 李玉照, 刘 永, 赵 磊, 等 (3280)
- 辐射传输模型多尺度反演植被理化参数研究进展 肖艳芳, 周德民, 赵文吉 (3291)
- 微囊藻毒素对陆生植物的污染途径及累积研究进展 靳红梅, 常志州 (3298)

个体与基础生态

- 年龄、性别及季节因素对千岛湖岛屿社鼠最大活动距离的影响 叶 彬, 沈良良, 鲍毅新, 等 (3311)
- 寄主大小及寄生顺序对蝇蛹佣小蜂寄生策略的影响 詹月平, 周 敏, 贺 张, 等 (3318)
- 两种苹果砧木根系水力结构及其 PV 曲线水分参数对干旱胁迫的响应
..... 张林森, 张海亭, 胡景江, 等 (3324)
- 三种根系分泌脂肪酸对花生生长和土壤酶活性的影响 刘 苹, 赵海军, 仲子文, 等 (3332)

种群、群落和生态系统

- 象山港春季网采浮游植物的分布特征及其影响因素 江志兵, 朱旭宇, 高 瑜, 等 (3340)
- 洞头海域网采浮游植物的月际变化 朱旭宇, 黄 伟, 曾江宁, 等 (3351)
- 狗牙根与牛鞭草在三峡库区消落带水淹结束后的抗氧化酶活力 李兆佳, 熊高明, 邓龙强, 等 (3362)
- 三亚岩相潮间带底栖海藻群落结构及其季节变化 陈自强, 寿 鹿, 廖一波, 等 (3370)
- 长期围封对不同放牧强度下草地植物和 AM 真菌群落恢复的影响 周文萍, 向 丹, 胡亚军, 等 (3383)
- 北京松山自然保护区森林群落物种多样性及其神经网络预测 苏日古嘎, 张金屯, 王永霞 (3394)
- 藏北高寒草地生态补偿机制与方案 刘兴元, 龙瑞军 (3404)
- 辽东山区次生林生态系统不同林型树干茎流的理化性质 徐天乐, 朱教君, 于立忠, 等 (3415)
- 施氮对亚热带樟树林土壤呼吸的影响 郑 威, 闫文德, 王光军, 等 (3425)
- 人工高效经营雷竹林 CO₂ 通量估算及季节变化特征 陈云飞, 江 洪, 周国模, 等 (3434)
- 新疆典型荒漠区单食性天花吉丁虫磷元素含量对环境的响应 王 晶, 吕昭智, 宋 菁 (3445)
- 双斑长跗萤叶甲越冬卵在玉米田的空间分布型 张 聪, 葛 星, 赵 磊, 等 (3452)
- 舟山群岛四个养殖獐种群遗传多样性和遗传结构 林杰君, 鲍毅新, 刘 军, 等 (3460)

景观、区域和全球生态

- 乡镇尺度金塔绿洲时空格局变化 巩 杰, 谢余初, 孙 朋, 等 (3470)
- 合并与不合并: 两个相似性聚类分析方法比较 刘新涛, 刘晓光, 申 琪, 等 (3480)

资源与产业生态

基于投入产出表的中国水足迹走势分析 王艳阳,王会肖,张 昕 (3488)

基于 MRICES 模型的气候融资模拟分析 朱潜挺,吴 静,王 铮 (3499)

黄东海陆架区沉积物中磷的形态分布及生物可利用性..... 张小勇,杨 茜,孙 耀,等 (3509)

鄱阳湖采砂南移扩大影响范围——多源遥感的证据 崔丽娟,翟彦放,邬国锋 (3520)

温度、盐度及其互作效应对吉富罗非鱼血清 IGF-I 与生长的影响 强 俊,杨 弘,王 辉,等 (3526)

城乡与社会生态

福建省城镇-交通系统的景观分隔效应 张天海,罗 涛,邱全毅,等 (3536)

研究简报

青藏高原高寒草原区工程迹地面积对其恢复植物群落特征的影响 毛 亮,周 杰,郭正刚 (3547)

黄土山地苹果树树体不同方位液流速率分析..... 孟秦倩,王 健,张青峰,等 (3555)

期刊基本参数:CN 11-2031/Q * 1981 * m * 16 * 314 * zh * P * ¥90.00 * 1510 * 33 * 2013-06



封面图说: 清晨的天山马鹿群——家域是动物行为学和保护生物学的重要概念之一,它在动物对资源环境的适应与选择,种群密度及社会关系等生态学过程研究中有着重要的作用。马鹿属于北方森林草原型动物,在选择生境的各种要素中,隐蔽条件、水源和食物的丰富度是最重要的指标。野生天山马鹿是中国的特产亚种,主要分布在北天山深山海拔1500—3800m 地带的森林草原中,在高山至谷地之间不同高度的坡面上,马鹿按季节、昼夜变化的不同进行采食。

彩图及图说提供: 陈建伟教授 北京林业大学 E-mail: cites.chenjw@163.com

DOI: 10.5846/stxb201203090319

刘新涛, 刘晓光, 申琪, 张书杰, 杨党伟, 任应党. 合并与不合并: 两个相似性聚类分析方法比较. 生态学报, 2013, 33(11): 3480-3487.

Liu X T, Liu X G, Shen Q, Zhang S J, Yang D W, Ren Y D. Comparison of merged and non-merged similarity clustering analysis methods. Acta Ecologica Sinica, 2013, 33(11): 3480-3487.

合并与不合并: 两个相似性聚类分析方法比较

刘新涛¹, 刘晓光¹, 申琪², 张书杰³, 杨党伟¹, 任应党^{1,*}

(1. 河南省农业科学院植物保护研究所, 河南省农作物病虫害防治重点实验室,
农业部华北南部作物有害生物综合治理重点实验室, 郑州 450002;
2. 河南中医学院, 郑州 450008; 3. 郑州大学生物工程系, 郑州 450001)

摘要:以山西省 4638 种昆虫在 7 个地理小区的分布、内蒙古 7766 种昆虫在 14 个地理小区的分布和中国 16804 属昆虫在 67 个生态区域的分布 3 组数据为样本, 用传统的层层合并的相似性聚类分析法 (SCA) 和新的不需合并的多元相似性聚类分析法 (MSCA) 进行运算分析, 对比结果表明, 不合并法都能得到既符合统计学逻辑, 又符合地理学、生物学逻辑的结果; 合并法在参与小区较少时, 还能够得到与不合并法类似的结果, 随着参与小区的增多, 聚类结构发生变化, 以致聚类功能彻底丧失。无论两种聚类结果差异大小, 其性质都迥然不同: 不合并法的相似性系数是固有的、互相独立的、同时存在的, 聚类结果是所有小区之间关系亲疏、距离远近的状态; 合并法的每个相似性系数都是合并的依据或结果, 前一个系数是后一个系数产生的条件, 后一个系数是前一个系数消亡的结果, 严格按照顺序, 当最后一个系数产生时, 前面所有系数和所有小区都已不复存在, 聚类结果只是记录不断合并、不断消亡的过程。因此在肯定合并法历史价值的同时, 认为申效诚等创建的多元相似性系数公式及多元相似性聚类分析法摒弃合并降阶这一产生偏差和错误的根源, 能够得出相对客观的聚类结果, 是生物地理学研究领域有效的聚类分析工具, 必将推动生物地理学定量研究迈入一个新阶段。

关键词:多元相似性聚类分析; 多元相似性系数; 生物地理学

Comparison of merged and non-merged similarity clustering analysis methods

LIU Xintao¹, LIU Xiaoguang¹, SHEN Qi², ZHANG Shujie³, YANG Dangwei¹, REN Yingdang^{1,*}

1 Institute of Plant Protection, Henan Academy of Agricultural Sciences; The Key Laboratory of Crops Pests and Diseases Control of Henan Province; The Key Laboratory of Integrated Pest Management on Crops in the Southern Region of North China; the Agriculture Ministry of China, Zhengzhou 450002, China
2 Henan University of Traditional Chinese Medicine, Zhengzhou, 450008, China
3 Bioengineering Department of Zhengzhou University, Zhengzhou 450001, China

Abstract: Distribution data of 4638 species in seven geographic regions of Shanxi Province were examined as a small sample, of 7766 species in 14 geographic regions of Inner Mongolia as a medium sample, and of 16804 genera in 67 ecological regions of China as a large sample. Statistical analyses of the three data groups were conducted separately, using a traditional merged method (similarity clustering analysis, SCA) and a new non-merged method (multivariate similarity clustering analysis (MSCA)). A critical comparison of the two methods demonstrates that the non-merged method can attain a result suitable for both logistics of biological statistics and geography, regardless of the scale of the data. The merged method (SCA) may achieve a result closely resembling that of the non-merged method when dealing with a fewer number of geographic regions. However, with an increased number of geographic regions, the clustering structure with the merged method may create a change at a different level — so much as to cause a complete loss of functionality. Regardless

基金项目:河南省基础与前沿技术研究计划项目(河南省昆虫区系研究(082300430370)); 河南省重点实验室建设专项(河南昆虫地理分布及区划研究(112300413221))

收稿日期: 2012-03-09; 修订日期: 2012-12-14

* 通讯作者 Corresponding author. E-mail: renyd@126.com

of the magnitude of difference between results of the two kinds of clustering, their nature will be totally different. The non-merged method similarity coefficients are inherent, independent of each other and exist simultaneously, the clustering result reflects the relationship and distance of all involved geographic regions, and all the coefficients are easily calculated with no strict orders. In the merged method, however, every coefficient was considered to be founded upon or be the result of clustering. The non-merged coefficient is the basis for the merged coefficient's emergence, which is a result of the non-merged coefficient's disappearance after merging. All of the calculations depend on input data and the deduced result is strictly in alphabetical order. It should be noted that the newest or final coefficients were worked out or generated, whereas the non-merged coefficients as well as the involved geographic regions had to be eliminated or discarded. The newest clustering coefficients were constantly generated, subsequently disappearing with the circulation. MSCA, in agreement with the value and huge contribution by SCA methods, can correct errors or inaccuracy that caused by merging or descending order during clustering by SCA method. It especially avoids some lost branches in the clustering result that are very important to the relationship, and cannot find any similarity level that requires indication in some detail. In summary, the MSCA method can solve many of the problems of the SCA method. The clustering achieves greater accuracy, which makes the results fit ecological reality. Also, our modified MSCA method can easily perform macroscopic clustering analysis of ecosystem data, which has never been completely accomplished before.

Key Words: multivariate similarity clustering analysis; multivariate similarity coefficients; biogeography

1901 年 Jaccard 提出的用于生物区系比较的相似性系数计算公式^[1], 由于简明、准确, 迅速得到人们普遍认可, 在生物学等自然科学以及社会科学的众多领域被广泛应用^[2], 以相似性作为尺度的聚类分析技术 (SCA) 也日渐普及。由于 Jaccard 的公式只能计算 2 个地区间的相似性系数, 于是“合并降阶”便成为相似性聚类分析方法中的核心技术环节, 并被奉为经典。人们在大中型相似性聚类分析的运算中得不到既符合统计学逻辑, 又具有地理学、生物学意义的结果时, 往往只怀疑自己的基础数据欠缺, 而不去质疑“合并”的合理性, 只能将研究和数据束之高阁。申效诚等从创立多元相似性系数计算公式入手^[3-4], 彻底摒弃层层合并的环节, 创建了新的多元相似性聚类分析方法 (MSCA)^[5-6], 经过多类群、多地理区域的运算实验^[7-11], 不仅简便省时, 而且聚类能力强大合理。

为了更直接鲜明地对比 SCA 和 MSCA 由于合并与不合并所产生的差异, 选用小、中、大型 3 组数据, 分别用两种方法运算, 比较聚类结果。以便为 MSCA 的广泛应用、为生物地理学的发展提供科学依据。

1 材料与方法

1.1 材料

所用材料均来自我们建造的中国昆虫分布数据库: 山西省 4638 种昆虫在 7 个小区的分布; 内蒙古自治区 7766 种昆虫在 14 个小区的分布; 中国 16804 属昆虫在 67 个生态区域的分布。

1.2 方法

用 Jaccard 的二元相似性系数计算公式将上述材料分别制出 7×7 、 14×14 、 67×67 的相似性系数三角矩阵备用。

1.2.1 传统的聚类分析法 (SCA)

选择相似性系数最大的两个小区首先聚类, 将这两个小区的分布资料合并为一个新的小区, 使参与聚类的小区降为 $n-1$ 个, 再用 Jaccard 的公式计算 $n-1$ 个小区的两两相似性系数, 排成 $(n-1) \times (n-1)$ 矩阵, 重新选择相似性系数最大的两个小区予以合并, 使参与聚类的小区数降为 $n-2$ 个。如此反复, 直至全部小区聚类完成。最后, 根据聚类与合并的顺序作出支序图。

Jaccard 公式:

$$SI = C / (A + B - C)$$

式中, SI 是两个小区间的相似性系数, A 、 B 分别是两个小区的种类数, C 是两个小区的共有种类数。

1.2.2 多元相似性聚类分析法(MSCA)

选择相似性系数最大的两个小区首先聚类,但不将这两个小区的分布资料合并,而是将其视为一个“单元群”,与其它没有聚类的 $n-2$ 个小区一同进入下一轮聚类分析。每轮均挑选相似性系数最大者聚类,扩大原单元群或形成新的单元群,如此反复,直到聚类完成。相似性系数采用申效诚等创立的多元相似性系数公式计算。最后作出支序图。

申效诚等的公式:

$$SI_n = \sum H_i / nS = (\sum S_i - \sum T_i) / nS$$

式中, SI_n 是要比较的 n 个小区间的相似性系数; S_i 、 H_i 、 T_i 分别是 i 小区的种类数、共有种类数、独有种类数,且满足 $S_i - T_i = H_i$; S 为 n 个小区的总种类数。这些数据都可以从数据库的查询表上直接获得。

两个公式原理完全相同,前者是后者在 n 为 2 时的一个特例。后者是前者在 n 大于 2 时的通式。也即本文要比较的两个方法的区别在于合并与不合并所引起的差异。

例如表 1 的山西省 7 个小区中,5、6 小区的相似性系数最大(0.412),合并法是将其合并成一个有 1413 种(974+1021-582)昆虫的新小区,然后全省降为 6 个小区,再重新寻找相似性最大的两个小区予以合并,直到最后。

不合并法是不将 5、6 小区合并,而是将其视为新的聚类单元进入下一轮比较,当计算新聚类单元(包含 5、6 小区)与其它任一小区(1、2、3、4、7 小区)间的相似性系数时,参与计算的是 3 个小区即 $n=3$,在这一轮中,(5、6)和 3 小区间的相似性系数最大,因此,((5、6)、3)聚在一起形成了新的聚类单元;在下一轮比较时,将计算((5、6)、3)分别与 1、2、4、7 小区之间的相似性系数(共 4 个),以及(1 和 2)、(1 和 4)、(1 和 7)、(2 和 4)、(2 和 7)、(4 和 7)小区间的相似性系数(共 6 个),挑选其中相似性系数最大者(1 和 4 小区)聚成一类;再下一轮比较,需要比较的聚类单元有:2 小区、(1、4 小区)、((5、6)、3 小区)和 7 小区,计算相似性系数时, n 值是实际参与的小区数,如计算(1、4 小区)与((5、6)、3 小区)之间的相似性系数时,实际参与的小区数是 5 个小区即 $n=5$ 。依次类推,直至全部小区聚类完成。

2 结果与分析

2.1 山西省昆虫分布的聚类分析结果比较

山西省共记录 4638 种昆虫,其中有省下分布记录的有 2619 种,分布在 7 个小区内的种类数、共有种类数及其相似性系数如表 1。

表 1 山西省各地理小区的昆虫种类(对角线)、共有种类数(上三角)和相似性系数(下三角)

Table 1 The insect species number (on diagonal line), shared species number (above diagonal) and similarity coefficient (below diagonal) in every regions in Shanxi Province

地区 Region	1	2	3	4	5	6	7
五台山	855	318	335	361	415	367	437
太行山	0.288	566	392	277	392	433	332
中条山	0.218	0.328	1020	294	433	544	393
恒山	0.347	0.331	0.231	547	392	337	373
晋中平原	0.294	0.342	0.277	0.347	974	582	497
晋南丘陵	0.243	0.375	0.363	0.274	0.412	1021	457
吕梁山	0.314	0.275	0.246	0.326	0.343	0.298	972

使用合并法和不合并法分别得到两个聚类图(图 1,图 2)。

比较图 1 和图 2,图 2 中 7 个小区在相似性系数为 0.30 时聚为两群。1、4、7 小区聚为一群,以中低山地为主,居该省北、西方;其余 4 小区为一群,以平原、丘陵、低山为主,居该省中、东、南部,7 个小区的总相似性系数为 0.248。图 1 中 7 个小区起初并为 3 个新小区,(1、4、7),(2、3),(5、6)各为一新小区。2、3 为低山,5、6 则为平原丘陵,生态学意义更为突出,但它们难以以更低的相似性系数合并在一起,其生态学意义在高一级的聚类中丧失。7 个小区最后的相似性系数为 0.308,最多可在 0.32 处区分成两个新小区,同样找不到辨别 3

个新小区的相似性水平。两种聚类方法的结果在地理学、生物学上不存在差异,聚类结构基本没有变化。在统计上的差异:第一,相似性系数的含义不同,合并法最后的相似性系数 0.308 是最终合并成的山地区与平原区之间的相似性系数,必须层层合并到最后才能完成,完成时,7 个小区已不复存在,支序图只是合并过程图;不合并法的相似性系数 0.248 确实是 7 个小区的总相似性系数,它不受聚类过程的影响,也不因聚类结构变动而变化,甚至可以最先计算出来;第二,合并法在 2、3 合并区和 5、6 合并区之间的相似性系数 0.382 比 2、3 合并时的系数 0.328 还高,这种“倒挂”现象是由合并引起的后果,致使支序图出现“凹陷”,不再是典型的梯形结构。

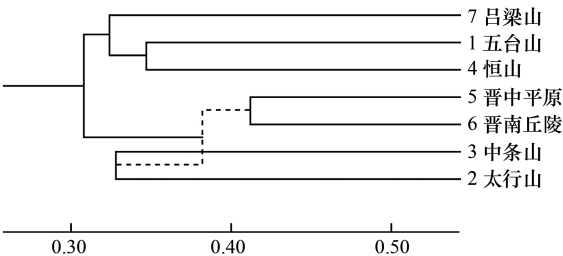


图 1 山西省昆虫分布合并法聚类图

Fig.1 The clustering graph of insect fauna of Shanxi Province by merge method

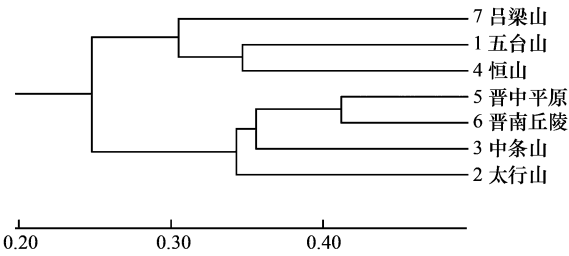


图 2 山西省昆虫分布不合并法聚类支序图

Fig.2 The clustering graph of insect fauna of Shanxi Province by non-merged method

2.2 内蒙古自治区昆虫分布的聚类结果比较

内蒙古自治区有昆虫 7766 种,有区下分布记录的共 5543 种。分布在 14 个小区的种类数、共有种类数和相似性系数如表 2,两种聚类法得到两个支序图(图 3,图 4)。

表 2 内蒙古各地理小区的昆虫种类(对角线)、共有种类数(上三角)和相似性系数(下三角)

Table 2 The insect species number (on diagonal line), shared species number (above diagonal) and similarity coefficient (below diagonal) in every regions in Inner Mongolia

地区 Region	01	02	03	04	05	06	07	08	09	10	11	12	13	14
01	1466	473	405	342	797	542	391	483	281	502	569	438	433	329
02	0.222	1134	496	371	380	430	390	561	290	278	453	398	480	411
03	0.199	0.308	972	431	338	386	344	504	289	243	392	358	555	420
04	0.157	0.204	0.269	1060	300	301	295	372	266	235	331	277	433	382
05	0.389	0.178	0.168	0.140	1680	489	395	387	243	558	508	400	392	293
06	0.259	0.230	0.220	0.156	0.237	1169	391	472	277	327	605	479	427	314
07	0.195	0.232	0.220	0.174	0.206	0.228	935	418	234	276	436	432	436	342
08	0.221	0.315	0.301	0.196	0.176	0.248	0.243	1208	298	280	481	446	578	413
09	0.145	0.187	0.201	0.171	0.128	0.168	0.160	0.179	758	219	288	245	291	258
10	0.309	0.183	0.175	0.158	0.376	0.218	0.209	0.176	0.183	661	395	311	270	240
11	0.261	0.231	0.211	0.169	0.236	0.328	0.246	0.240	0.165	0.256	1280	484	451	331
12	0.232	0.249	0.242	0.183	0.217	0.308	0.316	0.275	0.146	0.256	0.291	863	422	336
13	0.182	0.241	0.316	0.220	0.169	0.205	0.237	0.294	0.161	0.156	0.208	0.237	1341	452
14	0.170	0.271	0.312	0.260	0.156	0.190	0.247	0.260	0.199	0.198	0.190	0.254	0.269	794

01 呼伦贝尔高原;02 锡林格勒高原;03 乌兰察布高原;04 阿拉善沙漠;05 大兴安岭北段;06 大兴安岭南段;07 赤峰山地;08 察哈尔;09 贺兰山;10 大兴安岭北段山前平原;11 大兴安岭南段山前平原;12 辽河上游平原;13 河套地区;14 鄂尔多斯

图 4 中,在相似性系数 0.20 的水平上,14 个小区聚为两类,一类内蒙古的东北部,以大兴安岭等山地为主要地理特征,另一类在内蒙古西南部,以高原沙漠为主要地理特征,14 个小区的总相似性系数为 0.159。图 3 中,起初 12 个小区分别合并为 6 个新小区,在以后的 7 次系数计算中,有 3 次出现了“倒挂”,而且由于合

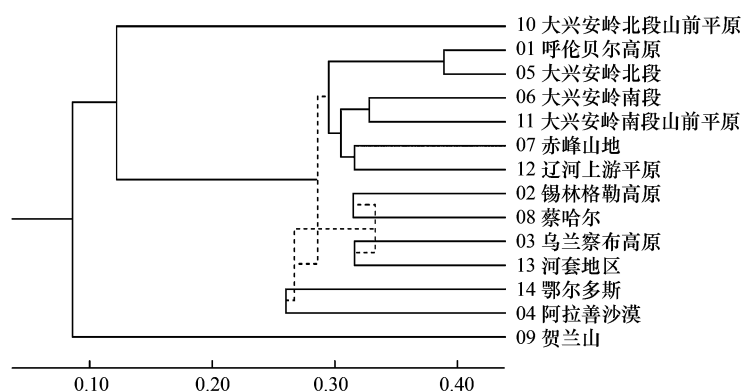


图3 内蒙古昆虫分布合并法聚类图

Fig. 3 The clustering graph of insect fauna of Inner Mongolia by merge method

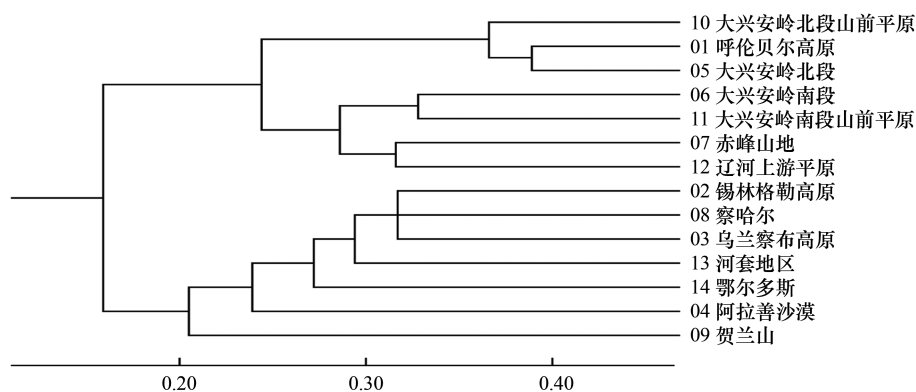


图4 内蒙古昆虫分布不合并法聚类支序图

Fig. 4 The clustering graph of insect fauna of Inner Mongolia by non-merged method

并,第9小区贺兰山和第10小区大兴安岭北段山前平原面积最小、昆虫种类最少,被排斥在外,直到最后是贺兰山和全内蒙古的比较,相似性系数为0.086,聚类结构产生较大变化,找不到一个合适的相似性水平把14个小区划分成几个有统计学和生态学意义的“类”来。“并而不类”,常常是合并法的最终结果。

图4中,也出现一次“倒挂”,2、8小区之间相似性系数为0.315,3、13小区之间为0.316,但2、8、3三者的相似性系数为0.317,3小区只能放弃13小区,和2、8小区聚在一起,由于2、8没有合并,可以将3个小区并列。

2.3 中国昆虫属级分布的聚类结果比较

数据库记录到的中国昆虫共91179种,隶属于16804属,按生态条件将全国分成67个基础地理单元,对于16904属在67个单元中的分布,用两个聚类方法得到两个支序图(图5,图6)。

图6中,67个基础地理单元在相似性系数为0.25时,聚合为9群,每群所辖单元在地理上都相邻相连,在昆虫区系性质上都具有相同或相似的成分构成,可以不加任何修饰地作为我国昆虫的9个分布区。图5中,67个单元最后合并成两区,一个是由5个单元合并,包括东北的小兴安岭、三江平原,西北的阿尔泰山,和新疆南部的帕米尔高原、昆仑山,违背地理学逻辑;另一个由其余62个单元合并而成,没有生态学和生物地理学价值。66个相似性系数中,除去23个有意义的最低层次系数外,其余43个中有21个系数是倒挂的。整个过程,除是一场数字游戏外,没有出现任何有积极意义的结果。

3 结论与讨论

3.1 两种聚类方法的差异显而易见,随着比较单元的增多愈加剧烈

使用同一组数据,两种聚类分析方法得到不同的结果,而且随着参与比较的地理单元的增多,差异愈加剧

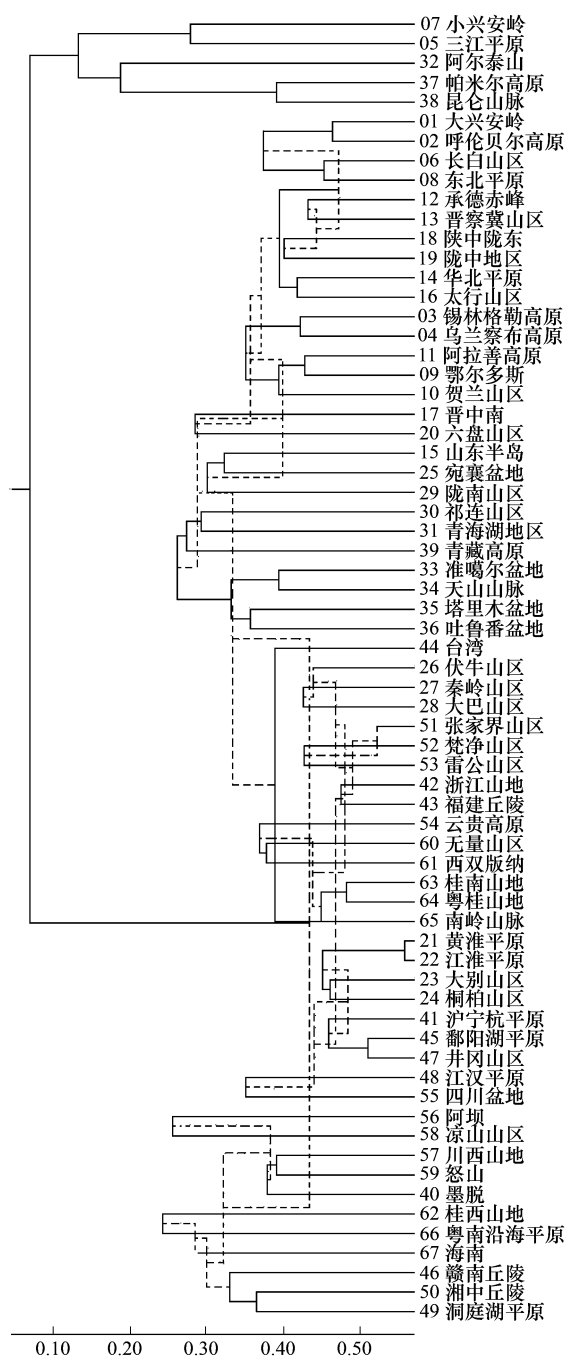


图5 中国昆虫属级分布合并法支序图

Fig. 5 The clustering graph of generic fauna from China by merge method

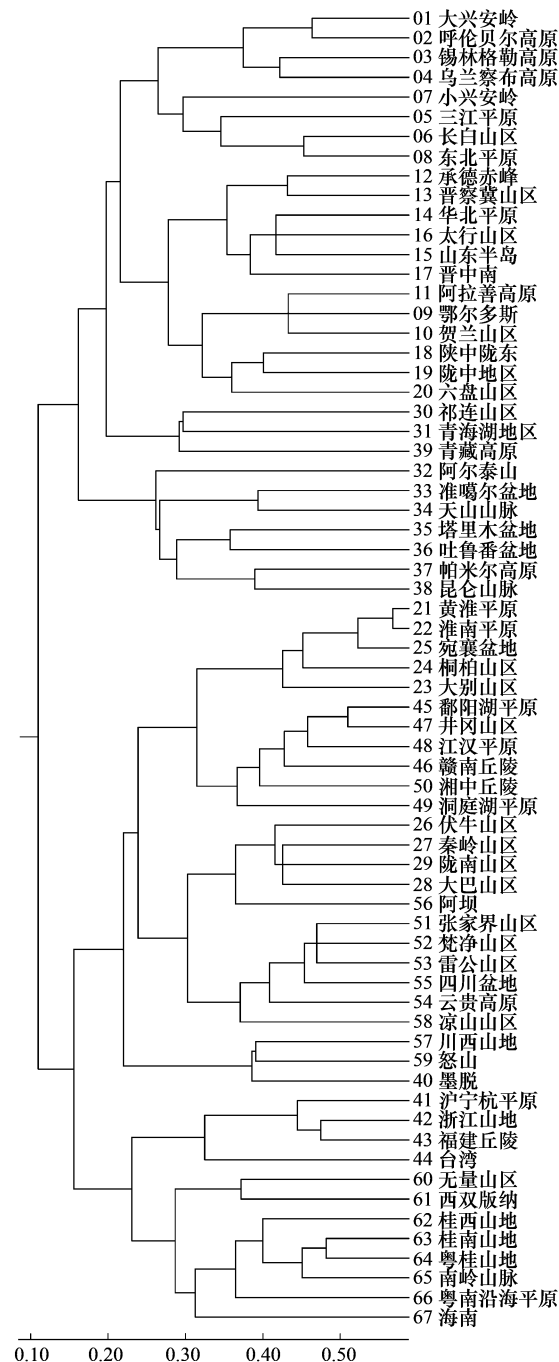


图6 中国昆虫属级分布不合并法支序图

Fig. 6 The clustering graph of generic fauna from China by non-merged method

烈,从相似性系数大小,到聚类结构变化,再到聚类功能丧失与否。这不是使用计算公式的错误,而是由于合并改变了原参与小区资料的性质所引发的变化。在参与比较的地理单元较少时(例如7个以下),聚类结构还不至于发生不合理变动,聚类结果还有一些应用价值。参与小区达到10个以上,聚类结果则难堪相信。所以目前聚类分析的报道多是较少地理单元的应用,多地理区域、多单元参与的报告寥若晨星。这也是人们已经看到合并法的应用局限性的结果。

两种方法的计算,简便程度也差别颇大。以手工计算为例,从制成二元相似性系数表开始,到绘出支序图

为止,合并法和不合并法的 3 个对比分别为 130min 和 50min,4.5h 和 1.8h,7d 和 2d。合并法所浪费的时间主要在合并数据的环节。

3.2 两种聚类方法的性质迥然不同,认识须逐步到位

无论两种方法的结果差异大小,即使完全相同的情况下,其性质也决然不同。不合并法的每一个相似性系数都是所辖小区的共同的相似性关系,不受所辖小区之间的聚类顺序变动的影响;每一个系数都是独立的,它的产生没有顺序,既可从下到上,也可从上到下,又可从中间任何层次算起;所有系数都是同时存在的。所以,不合并法的支序图是一个“状态”,一个所参与地理单元在共同存在的情况下表明彼此关系亲疏、距离大小的状态。

合并法的每一个相似性系数都是有关小区经过多次合并而成的两个新小区的相似性关系,受有关小区之间的合并顺序变动的影响;每一个系数都不是独立的,它的产生遵循从下到上的顺序,前一个系数是后一个系数产生的条件,后一个系数是前一个系数消亡的结果;所有系数都不可能同时存在。所以,合并法的支序图是一个“过程”,一个所参与地理单元不断消亡新单元不断产生的过程,一个不断肯定又不断否定的过程。

相似性系数越聚越高的“倒挂”是两个方法都遇到的现象,但其性质也不相同。不合并法的倒挂是由于涉及到的 3 个或 4 个小区互相都有较高的相似性,聚类后的共同相似性系数更高的罕见现象,只出现在聚类过程中的初级层次,极少出现在较高层次,出现频次不高,出现时可以用并列法表示;合并法中的倒挂是由于合并后的两个新小区之间的较高的相似性,它主要出现在合并过程的较高层次,而且频次很高,几占较高层次的 1/2。由于涉及到的小区已经合并,没有办法再把已经合并消失掉的它们并列,只能使支序图出现凹陷,失去正常的梯形结构。

3.3 合并法的历史作用值得肯定,终结其历史阶段的条件已经具备

1848 年,植物学领域首先提出相似性的概念,1901 年,Jaccard 提出了计算两个地区间生物区系的相似性系数公式,由于其简便性、科学性,迅速得到科学界认可。此后,人们又相继提出 40 余个相似性公式,但都未动摇 Jaccard 公式的经典地位,成为多学科、多领域中相似性计算的最基础、最常用、最直观方法。由于 Jaccard 公式不能计算多地区的相似性系数,在相似性聚类分析中采用“合并降阶”的办法,能够在较少小区比较时得到相对满意的结果,实现了人们多区比较的愿望,使生物地理由定性研究向定量研究发展迈出了第一步,其历史性价值不容低估。随着其局限性的逐渐显现,人们曾试图对合并后的二元系数进行修饰改良^[12],但由于未脱离合并的窠臼,也难以达到预期的效果。因此在经历了短期的热情之后,众多领域的中大型聚类需求得不到满足,其积极作用便逐渐消失,以致成为制约生物地理发展的瓶颈。申效诚等人创建的多元相似性系数公式及 MSCA 法,彻底摒弃合并降阶这一产生偏差和错误的根源,能够得出相对客观的聚类结果,是生物地理学研究领域有效的聚类分析工具,必将使生物地理学的定量研究迈入一个新阶段。

References:

- [1] Jaccard P. Distribution de la flore alpine dans le Bassin des Dranses et dans quelque region voisines. Bulletin de la Societe vaudoise des Sciences naturelles. Lausanne. 1901,37: 241-272.
- [2] Zhan Y L. Coefficient of Similarity — An Important Parameter in Floristic Geography, Geographical Research, 1998, 17(4): 429-434
- [3] Shen X C, Sun H, Zhao H D. A discussion about the method for multivariate similarity analysis of fauna. Acta Ecologica Sinica, 2008, 28(2): 849-854.
- [4] Shen X C, Wang A P. A Simple Formula for Multivariate Similarity Coefficient and Its Contribution Rate in Analysis of Insect Fauna. Journal of Henan Agricultural Sciences, 2008, (7): 67-69.
- [5] Shen X C, Wang A P. Zhang S J. Studies on the Fauna of Noctuidae II. Distribution and Similarity of Noctuidae in China. Acta Agriculturae Boreali-Sinica, 2008, 23(5): 151-156.
- [6] Shen X C, Zhang S J, Ren Y D. The elements of insect fauna in China and distribution characteristics. Journal of Life Science, 2009, 3(7): 19-25.
- [7] Zhao H D, Shen X C. A study on the Biogeography of Family Arctiidae in China//Shen X C, Zhang R Z, Ren Y D. Classification and Distribution

of Insects in China, Beijing: China Agricultural Science and Technology Press, 2008, 381-388.

- [8] Sheng M L, Shen X C. Distribution and Multivariate Similarity Clustering Analysis of Ichneumonidae in Every Provinces, China//Shen X C, Zhang R Z, Ren Y D. Classification and Distribution of Insects in China, Beijing: China Agricultural Science and Technology Press, 2008, 389-393.
- [9] Shen X C, Ren Y D, Wang A P. Zhang S J. A multivariate similarity clustering analysis for geographical distribution of insects, spiders and mites in Henan Province. Acta Ecologica Sinica, 2010, 30(16): 4416-4426.
- [10] Shen X C, Sun H, Ma X J. The multivariate similarity clustering analysis for 40,000 species of insect and spider fauna in China. Journal of Life Science, 2010, 4(2): 35-40.
- [11] Ren Y D, Shen X C, Sun H, Ma X J. The Fauna Element and Geographical Distribution of Insect, Spider and Mite in Henan, China. Acta Agriculturae Boreali-Sinica, 2011, 26(1): 204-209
- [12] Ward J H. Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association. 1963, 58: 236-244.

参考文献:

- [2] 张德铨. 植物区系地理研究中的重要参数——相似性系数. 地理研究, 1998, 17(4): 429-434.
- [3] 申效诚, 孙浩, 赵华东. 昆虫区系多元相似性分析方法. 生态学报, 2008, 28(2): 849-854.
- [4] 申效诚, 王爱萍. 昆虫区系多元相似性的简便计算方法及其贡献率. 河南农业科学, 2008, (7): 67-69.
- [5] 申效诚, 王爱萍, 张书杰. 夜蛾科昆虫区系研究 II. 中国各省区夜蛾的分布及相似性分析. 华北农学报, 2008, 23(5): 151-156.
- [6] 申效诚, 张书杰, 任应党. 中国昆虫区系成分构成及其分布特点. 生命科学, 2009, 3(7): 19-25.
- [7] 赵华东, 申效诚. 中国灯蛾科昆虫的生物地理学研究//申效诚, 张润志, 任应党. 昆虫分布与分类. 北京: 中国农业科学技术出版社, 2008, 381-388.
- [8] 盛茂领, 申效诚. 中国各省区姬蜂科昆虫的分布及多元相似性聚类分析//申效诚, 张润志, 任应党. 昆虫分布与分类. 北京: 中国农业科学技术出版社, 2008, 389-393.
- [9] 申效诚, 任应党, 王爱萍, 张书杰. 河南昆虫、蜘蛛、蜚蠊地理分布的多元相似性聚类分析. 生态学报, 2010, 30(16): 4416-4426.
- [10] 申效诚, 孙浩, 马晓静. 中国 40000 种昆虫蜘蛛区系的多元相似性聚类分析. 生命科学, 2010, 4(2): 35-40.
- [11] 任应党, 申效诚, 孙浩, 马晓静. 河南昆虫、蜘蛛、蜚蠊的区系成分和分布地理研究. 华北农学报, 2011, 26(1): 204-209.

《生态学报》2013 年征订启事

《生态学报》是由中国科学技术协会主管,中国生态学学会、中国科学院生态环境研究中心主办的生态学高级专业学术期刊,创刊于 1981 年,报道生态学领域前沿理论和原始创新性研究成果。坚持“百花齐放,百家争鸣”的方针,依靠和团结广大生态学科工作者,探索自然奥秘,为生态学基础理论研究搭建交流平台,促进生态学研究深入发展,为我国培养和造就生态学科人才和知识创新服务、为国民经济建设和发展服务。

《生态学报》主要报道生态学及各分支学科的重要基础理论和应用研究的原始创新性科研成果。特别欢迎能反映现代生态学发展方向的优秀综述性文章;研究简报;生态学新理论、新方法、新技术介绍;新书评价和学术、科研动态及开放实验室介绍等。

《生态学报》为半月刊,大 16 开本,300 页,国内定价 90 元/册,全年定价 2160 元。

国内邮发代号:82-7,国外邮发代号:M670

标准刊号:ISSN 1000-0933 CN 11-2031/Q

全国各地邮局均可订阅,也可直接与编辑部联系购买。欢迎广大科技工作者、科研单位、高等院校、图书馆等订阅。

通讯地址:100085 北京海淀区双清路 18 号 电 话:(010)62941099; 62843362

E-mail: shengtaixuebao@rcees.ac.cn 网 址: www.ecologica.cn

本期责任副主编 朱永官

编辑部主任 孔红梅

执行编辑 刘天星 段 靖

生 态 学 报

(SHENGTAI XUEBAO)

(半月刊 1981 年 3 月创刊)

第 33 卷 第 11 期 (2013 年 6 月)

ACTA ECOLOGICA SINICA

(Semimonthly, Started in 1981)

Vol. 33 No. 11 (June, 2013)

编 辑	《生态学报》编辑部 地址:北京海淀区双清路 18 号 邮政编码:100085 电话:(010)62941099 www.ecologica.cn shengtaixuebao@rcees.ac.cn	Edited by	Editorial board of ACTA ECOLOGICA SINICA Add:18, Shuangqing Street, Haidian, Beijing 100085, China Tel:(010)62941099 www.ecologica.cn shengtaixuebao@rcees.ac.cn
主 编	王如松	Editor-in-chief	WANG Rusong
主 管	中国科学技术协会	Supervised by	China Association for Science and Technology
主 办	中国生态学学会 中国科学院生态环境研究中心 地址:北京海淀区双清路 18 号 邮政编码:100085	Sponsored by	Ecological Society of China Research Center for Eco-environmental Sciences, CAS Add:18, Shuangqing Street, Haidian, Beijing 100085, China
出 版	科 学 出 版 社 地址:北京东黄城根北街 16 号 邮政编码:100717	Published by	Science Press Add:16 Donghuangchenggen North Street, Beijing 100717, China
印 刷	北京北林印刷厂	Printed by	Beijing Bei Lin Printing House, Beijing 100083, China
发 行	科 学 出 版 社 地址:东黄城根北街 16 号 邮政编码:100717 电话:(010)64034563 E-mail:journal@cspg.net	Distributed by	Science Press Add:16 Donghuangchenggen North Street, Beijing 100717, China Tel:(010)64034563 E-mail:journal@cspg.net
订 购	全国各地邮局	Domestic	All Local Post Offices in China
国外发行	中国国际图书贸易总公司 地址:北京 399 信箱 邮政编码:100044	Foreign	China International Book Trading Corporation Add:P. O. Box 399 Beijing 100044, China
广告经营 许 可 证	京海工商广字第 8013 号		



ISSN 1000-0933
CN 11-2031/Q

国内外公开发行

国内邮发代号 82-7

国外发行代号 M670

定价 90.00 元