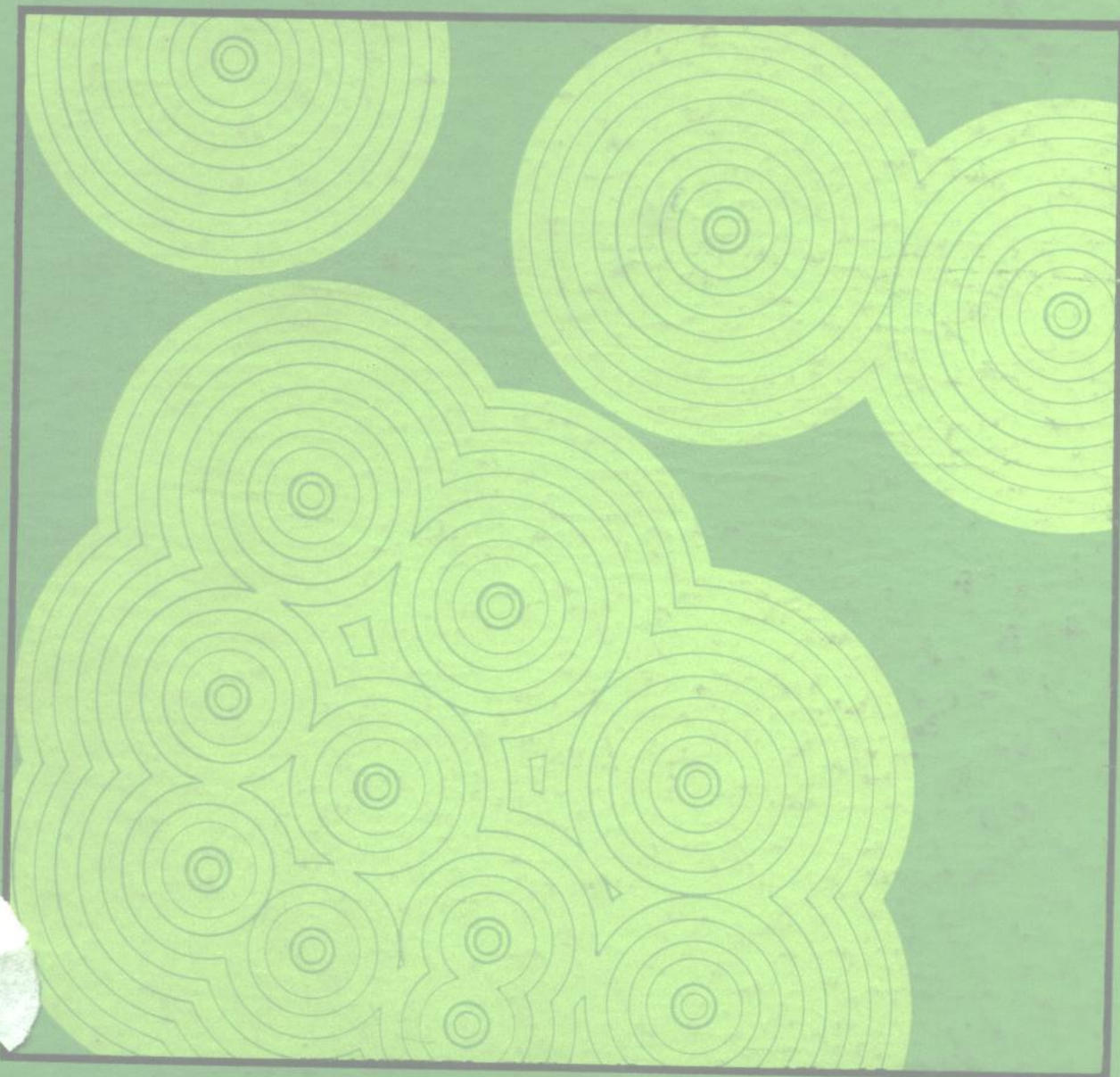


[英] P. 史尼斯 著
[美] R. 索卡尔

数值分类学

—数值分类的原理和应用



数值分类学

数值分类的原理和应用

[英] P. 史尼斯 著
[美] R. 索卡尔

赵铁桥 译

汪振儒 校



科学出版社

1984

8510181

DS84/55

内 容 简 介

生物分类理论方面近年来进展十分迅速,如进化系统学、数值分类学、分支系统学等学派热烈争鸣,本书是数值分类学方面的一部重要理论著作。全书共分十二章,前四章阐述了数值分类学的目的、原理和各项基础理论,中间的五章详尽叙述了该学科的内容、方法及相互关系。最后的三章对数值分类学的地位、应用及其今后发展作了估计。本书所述理论系统,推理严密,引用文献资料也很丰富,使读者对这一崭新分支学科的发展、理论和今后需要进一步探讨的问题,有比较全面系统的了解。

可供专门从事生物分类研究的科技工作者参考,也可以作为有关大专院校师生的参考书。本书对其它学科直至社会科学,亦有价值。

P. Sneath, R. Sokal

NUMERICAL TAXONOMY

The Principles and Practice of Numerical Classification

W. H. Freeman and Company

数 值 分 类 学

数值分类的原理和应用

〔英〕P. 史尼斯 著
〔美〕R. 索卡尔

赵铁桥 译
汪振儒 校

责任编辑 张志强 关 国

科学出版社出版

北京朝阳门内大街137号

中国科学院印刷厂印刷

新华书店北京发行所发行 各地新华书店经售

*

1984年10月第一版 开本:787×1092 1/16

1984年10月第一次印刷 印张:26

插页:精3 平4

印数:精1—3,400 字数:603,000

平1—2,500

统一书号:13031·2680

本社书号:3684·13—6

定价:布脊精装 5.10 元
平 装 4.10 元

序 言

从林奈时代起到我们自己的现代,生物科学的一个弱点就是我们的分类术语没有任何定量的含义。“纲”是什么? A纲和B纲的差异与C纲和D纲的差异是否相同?对目、科、属和种等其它分类梯级也可以提出这个问题。对此,没有一个答案是圆满的,而在大多数情况下则完全不能回答……除非到能对类似的问题给以适当答复以后,我们的分类方案才能令人满意,即“自然的”。现有的方案不比记忆口诀(只挂有一些毫不相关的片断知识的空格架)好多少。曾支持上一世纪所有分类系统的进化理论,对这些难题也没有提供实质性答案。在地质学上虽有一些零散的片断,但是,对不同生物组群发生的方式加以概括和对那些组群作出定量的评价,是非常不同的两件事。

辛格(C. Singer):《生物学史》(1959),
第200页。

数值分类学迅速的、几乎爆炸性的发展和对这个领域日益广泛的兴趣,使我们明白了我们首次在1963年出版的《数值分类学原理》一书需要新编。我们很快就体会到,我们的任务不能单纯限于修订内容。不仅有许多发表了的新材料需予编入,而且学科的重点和理论体制上发生了变化,需以适当篇幅表述。鉴此,写下了一本全新的书。

以往十年,足证生物系统学和居群生物学领域看法和方法论的显著变化。许多新途径是定量的,使用种种数学分支学科。分类学上,数值方法有了可观发展,有许多由计算机来完成。这些方法也影响了其它分支学科,又转过来为系统学提供了新的概念和技术。

数值分类学——根据其性状状态按分类单位的数值方法归类——在它目前的迅速发展中得助于计算机技术的同时发展。数值分类学旨在发展分类关系的估价和类元的建立上都是客观的、明确的和可重复的方法。加之数值方法在准确测度进化速率和在种系发生分析上,都揭开了一个广阔的领域。过去直觉途径的成功,在于迅速地、虽然不精确地识别形态学细节上的全面相似性的记忆能力。对于分类学上现在就像微生物学、化学或生理学的性状那样,往往具有表格形式的、日益增加的数据基础,这样的识别并不容易;对这些性状使用数值方法成为必要。

本书目的是为数值分类学提供时新的理论基础,使读者熟悉它的程式,说明它胜过常规分类学的优点,并报道该领域的现状。

我们无法论述已经在分类上使用的数值分析的所有形式,这可以轻而易举地参考关于在生物学上使用统计学和数学方法的大量教科书。我们只把本书范围限制于表明分类

关系和产生分类学学派的方法, 尽管为了完备, 我们也简括地论述一些其它技术。然而, 我们试图尽可能广泛地论述我们的论题, 使本书对动物学者、植物学者、微生物学者和古生物学者有用, 也对有关领域的学者有用。

熟悉《数值分类学原理》的读者会注意到, 章节和论题的顺序已经改编, 以形成我们认为较好的一个完整统一体。原有若干节已经删除或更简略了——另一些则大为增加。例如, 对分类学上常规实践作详细批判和争论, 现在看来没有什么必要了。观点已经确立并被广泛接受, 而目前需要的是见解较新的理论和方法论的阐明和推敲。反之, 种系发生分析的数值方法——在原书写成之时是不发达的, 在本教程中需要相当篇幅均衡地论述。我们尝试把系统学和分类学理论基础的讨论统成浑然一体, 不再把我们对常规系统学的批判性评论同我们提倡的观点分开。

在关于数值分类学理论和假定的各节中, 会发现大有变化。最近几年的工作使我们领悟到: 我们早先对迅速而利索地解决分类学问题的有些希望和期待, 时机还不成熟。于是, 举例来说, 我们还是不知道用于最优分类学的方法(甚或方法有无), 因而不可能提倡一种。不过, 除了这个领域里最保守的工作者外, 全都清楚: 未来的分类学若不是完全由计算机进行也会大得其助, 虽然要在数值分类学的发展中为一切问题提供“锦囊妙计”(“cookbook” recipes), 还是太早, 但是在自己的工作中忽视数值分类方法的系统学者则对此自负其责。

重写关于方法的各节最难应付。完备说明搞过或提过的一切, 结果将是一部体积大此二倍的书, 妨碍按时出版, 也使初学者为难。我们不得不在选取要讲述的方法上折衷, 在此报道使用最频繁的途径, 并为另一些较不常用的或我们觉得普遍意义不大的途径提供引证。虽然, 正如我们将经常指出的, 数值分类学的许多方面统计学的和概念的根据是不足的, 但在此我们甚至介绍那些我们也知道不够严谨的方面, 因为我们想报道学术的现状, 也想鼓励这些方面的深入工作。

必然地, 自从《数值分类学原理》发表, 掌握这门学问所需的数学知识水平有所提高。幸好, 学问日趋复杂同近年来系统学和其它生物学领域里年轻生物学者们的数学造诣日臻纯熟相称。这样的时代若非业已来临, 也正在迅速逼近, 那时透彻了解生物度量分析和颇为熟悉计算机处理, 将是抱负不凡的分类学者的一项不可缺少的先决条件。我们料想在书读者具有一些统计学知识、初等的集和图论的知识, 也具有一些矩阵代数的知识。现本市面上有许多作这样介绍的书; 选列在附录 B 里。

在《数值分类学原理》中, 我们力图尽我们所知评述数值分类学在生物学领域的一切应用。以后这方面的文献如此之多, 以致我们窘于估计文献总数而没有在本书里尝试提供一份详尽无遗的书目。乘引用某些关键性评论文章之便, 我们试图给读者指出可以找到数值分类学特定分支上的文献综述的来源。然而, 我们在附录 A 里提出了一份我们所知的全部新近论文的清单, 这些论文已经把数值分类学应用于各种各样机体的分类, 按分类学大类开列。我们希望这份清单将对愿意在他们的专业方面概观这一领域的分类学者具有价值。

在说明数值分类学在生物系统学之外领域的应用上, 我们不得不较为拘谨。应用在数量和多样性上是令人惊奇的。此处提及的方法, 相似的方法, 已被应用于范围从考古学到政治科学、从材料分类到语言学、从电视节目编排到生物地理学的领域, 并且虽然我们

力图对数值分类学在这些学科上的应用作象样报道,但篇幅的限制和我们自己对这些领域的知识局限要求从简。

主要兴趣不在生物系统学而是想把本书用作一般的类聚和数值分类学原理入门的读者,可以用连续的方法只读如下章节:第 1.1—5、1.9, 2.1—2、2.8, 3.1—3、3.8—9, 4.1—8、4.10、4.12—13, 5.1—2、5.4—10, 6.4 节,第八章和第十一章,漏失不多。

我们不想依靠简单的例子,象我们在《数值分类学原理》的附录中做的那样,去说明技术。要举例说明分类学上现有的更多得多的方法又保持本书的大小和价格仍然恰当,将是不可能的。再者,有许多方法的复杂性实在需要计算机处理数据。为要有用,方法论细节的讨论将要求对计算机处理详加指点(也讲到一些性状编码和数据的初步处理)。这在弗利曼公司(W. H. Freeman and Company)将出版的、罗尔夫(F. J. Rohlf)和尼利(P. M. Neely)的一本现成的书里有。

在计算方面我们限于简短讨论,我们在附录 B 中为那些需要更深入这个领域的工作者列出了一些资料来源,同样也提供了一般意见和情报。在那儿我们还列出了一些关于数值分类学的种种其它方面的书刊和评论。

本书的准备工作开始于 1967 年秋 9 月,那时史尼斯访问教授应聘在堪萨斯大学讲学。有几章的初稿在大学的生物系统学和数值分类学午餐会团体(Biosystematics and the Numerical Taxonomy Luncheon Groups)中宣读过,并受益于其成员们的建设性批评。当英国医学研究会(The Medical Research Council of the United Kingdom)让史尼思访问斯托尼布鲁克纽约州大学时,继续了共同工作。在古尔本吉基金会(Gulbenkian Foundation)主办的葡萄牙威拉什高等研究院(The Institute of Advanced Studies in Oeiras, Portugal)的夏季课程中,作者们幸运地能够完成最后校订。我们感激范尤登博士(Dr. N. Van Udan)听任我们利用一切方便。

我们乐于感谢三位同行,他们读过全部文稿,并以渊博的建设性批评帮助我们。堪培拉联邦科学和工业研究组织昆虫学部科利斯博士(Dr. D. H. Colless)、哈彭登罗桑斯塔德实验站高尔(J. C. Gower)和斯托尼布鲁克纽约州大学罗尔夫对改进我们的阐述贡献巨大。法里士博士(Dr. J. R. Farris)阅读了种系发生一章,基尔希博士(Dr. J. A. W. Kirsch)阅读了有关性状种类的各节。施特劳斯博士(Dr. J. H. Strauss)阅读并改进了精神病分类的一节。克罗维洛博士(Dr. T. J. Crovello)对贯穿正文的许多讨论作出了贡献。这些同行在编排他们校阅过的材料方面帮助很大。在斯托尼布鲁克纽约州大学和威拉什高等研究院听夏季课程的许多学生对内容的改进提出了建议。我们感谢来自朋友、同行和学生们的盛情帮助,而如果我们未能一一听从他们的忠告,则请他们海量。

我们的成果得到萨瓦雷斯女士(Mrs. Ethel Savarese)细心熟练的秘书工作的巨大帮助,她看了贯穿许多个准备阶段的手稿,使我们受惠甚多。也要感谢琼斯女士(Mrs. B. Jones)为附录和参考书目打字。斯托尼布鲁克纽约州大学数值分类学课程的学生们在约定时间内为本书清稿。应该指出斯通(R. Stone)和米克维支(M. Mickvich),他们校对了大部分打字稿,还有柯恩菲尔德(I. Kornfield),他负责汇总。托雷卡(B. Torraca)和翟纽保(Che Nu Paul)在校对索引上帮助最大。

我们感谢出版者艾贝拉德-舒曼公司(Abelard-Schuman Ltd.)允许引用辛格《生物学史》的一段。全部资料来源的作者已在正文中示明并列入参考书目之中。

我们想在此感谢英国医学研究会(给予史尼斯)和美国国家科学基金会(给予索卡尔)在数值分类学研究上的支持。这些组织的帮助在本门学问早先发展时期是决定性的,而它们的持续支持深受感激。

彼得·H·A·史尼斯

(Peter H. A. Sneath)

罗伯特·R·索卡尔

(Robert R. Sokal)

1973年2月

目 录

献辞	iv
序言	v
第一章 数值分类学的目标和原理	1
1.1 分类学上一些术语的定义	1
1.2 数值分类学的定义	2
1.3 数值分类学的基本见解	3
1.4 类似性的估计	4
1.5 类元的构成	4
1.6 种系关系的识别	5
1.7 表相分类学和种系发生分类学	6
1.8 数值分类学的优点	7
1.9 标本的鉴定	8
1.10 命名法	8
1.11 分类学上数值方法的发展	8
第二章 分类学原理	12
2.1 经验途径和运筹途径	12
2.2 自然系统	14
2.3 分类学关系	20
2.4 估计表相关关系问题	22
2.5 估计分支关系问题	28
2.6 选取分类基础	37
2.7 分类(学)秩级	42
2.8 分类系统的理想特性	43
第三章 分类学证据	47
3.1 运筹的分类单位	47
3.2 分类学性状的定义	48
3.3 单位性状	49
3.4 同源	51
3.5 性状的种类	61
3.6 性状的选取	65
3.7 不可采纳的性状	69
3.8 性状的必需数目	70
3.9 性状加权问题	72
第四章 分类学类似性的估计	76
4.1 数据矩阵	76
4.2 相似性系数引论	77
4.3 距离系数	80

4.4	结合系数	85
4.5	相关系数	90
4.6	概率相似性系数	92
4.7	系数的比较	95
4.8	对性状编码和标度	96
4.9	生长和形态学	103
4.10	相似性系数的统计显著性	106
4.11	表相类似性的组分	109
4.12	不恰当的比较	116
4.13	OTU 内性状变异	119
第五章	分类学结构	123
5.1	类似性矩阵	124
5.2	型式和类聚	125
5.3	生物学上的分类目标	130
5.4	成聚方法的分类学	131
5.5	序贯、凝结、等级、非迭代的成聚方法	139
5.6	标位方法	161
5.7	图形和树	167
5.8	数值分类学中 Q 技术和 R 技术的关系	170
5.9	分类学结构的表示法	171
5.10	最优性标准和分类的比较	182
5.11	秩级标准	193
5.12	分类借生化方法确证	199
5.13	关于进行和发表数值分类学研究的建议的概述	202
5.14	OTU 和类元在表相空间的分布	204
第六章	种系发生的研究	207
6.1	表相学和时维	207
6.2	进化速率	209
6.3	分支学分析	213
6.4	分支学分析的数值途径	216
6.5	古生物学中的数值分类学	240
第七章	居群表相学	244
7.1	定义问题	244
7.2	居群水平的数值分类学	247
7.3	表相型式和进化结构	249
7.4	表相学和环境	251
7.5	地理变异分析	253
第八章	鉴定和判别	256
8.1	鉴定矩阵	256
8.2	一般考虑	257
8.3	序贯检索表	260
8.4	同时检索表	265
8.5	判别式分析	270

第九章 牵涉命名法的问题.....	276
9.1 一些一般考虑	276
9.2 数值分类学和命名法问题	279
第十章 数值分类学的批评考查.....	281
10.1 对数值分类学的批评	281
10.2 数值分类学的缺点	287
10.3 数值分类学的启示性方面	289
第十一章 生物系统学以外领域的数值分类学.....	293
11.1 生态学和生物地理学	293
11.2 医学	296
11.3 社会科学	298
11.4 地球科学	300
11.5 其他科学和工艺学	302
11.6 艺术和人文学	303
第十二章 系统学的未来.....	304
附录 A 数值分类学在生物系统学上的应用.....	306
附录 B 关于技术、来源和引证的一些提示	319
参考文献.....	324
译校后记.....	379
索引.....	382

第一章 数值分类学的目标和原理

本书内容分为三大部分。第一部分两章提供一般分类学理论、特别是数值分类学的入门。本章的目的是以摘要形式概述数值分类学的目标和原理。在第二章,我们相当详细地讨论分类的概念基础,同近年越来越多的分类学者所袒护的因袭观点相对照。

本书中心部分是按计划布置的,计划密切反映分类学者在进行分类处理时所遵循的相继步骤。在第三章,我们讨论分类学证据,包括用于研究的机体选择、分类性状的选取和定义、与同源的标准。机体间的分类学相似性的估计接在第四章里。第五章考虑在这些类似性的基础上将机体归作类元。

最后部分论述数值分类学对系统学研究的关系。种系发生研究的数值途径在第六章论述。下一章,第七章探讨居群的数值分类学,导向表相型式和进化结构的讨论。数值分类学对检索和鉴定的牵连,在第八章讨论;对命名法的牵连在第九章讨论。数值分类学的批判性审查——其优点及其缺点——在第十章专述。下一章,第十一章为数值分类学在生物系统学之外领域应用的必要简短的概观。正文在第十二章以关于计算机时代里系统学的未来的简短展望告终。

备有两个附录。一份数值分类学在不同组群有机体的应用的清单,见于附录 A;关于数值分类研究上数据处理的意见,见于附录 B。

我们作了审慎的努力,使代表性状、运筹分类单位(OTU)和类元(OTU 组群)的符号通卷保持统一。这就意味着在许多地方,我们不得不使用不同于原出版物所用符号来报道其它作者的工作。我们觉得,本书有统一的符号,足以消弥给那些想查询原文的读者所造成的、混乱的危险而有余,我们希望,本书的出版会在阻止数值分类学上不同符号的发展方面,稍微起些作用。虽有极好愿望,还不能百分之百地使符号完全统一,部分因为字母表上适用的字母数目所限。我们尽量对所有这样的场合作出记号,使读者进而减少混淆的机会。

1.1 分类学上一些术语的定义

分类学术语的确切定义本身就几乎需要一本书。许多术语在如此大不相同的含义上使用,以致我们在以一致和精确的方式使用它们时有了困难,并且无疑仍有许多我们忽视了的模棱两可之处。本书所用符号和术语的意思,可以通过索引查对。然而,有几个用得如此频繁,我们认为以此处一开头就予介绍最好。

分类、系统学和分类学往往相互交替使用。近年,尤其在美国,有一种将这些术语的各别意义分定的倾向。在这个意义上,它们已由 Simpson (1961) 恰当予以定义,而我们按他的用法。

系统学 (Systematics)。这被 Simpson (1961, 第 7 页) 定义为:“机体种类和多样性的科学研究与它们之间任何的或全部的关系的科学研究。”这一定义在系统学的广义上使用

它,就其本身而论不仅包含机体成类元排列和为之命名,而且包含这些排列的原因和起源。

分类 (Classification)。Simpson 的定义本身限定动物分类,我们将它变成更为比较普遍的用法。“分类是在机体的关系基础上将它们整理成组群(或集)”。关于术语“关系”(relationship),可能有混淆,它可能暗示(在第 2.3 节分析的)种系发生的关系,或者它可能单纯指示依机体性状时的类似性或全面相似性,至于它们的祖先关系就毫不相干了。如 Cain 和 Harrison (1960b) 使用的那样,采用 H. K. Pusey 的方便术语,将基于全面相似性的关系称为表相(的)关系 (phenetic relationship) 来指依机体的表(现)型而不依其种系发生判断的关系,可以将它与种系发生关系相区别。我们将分类的这一定义限定于机体,因为本书原来预定是给生物学者的。然而,有许多分类方法,包括数值分类学,是同等适用于机体以外的概念和实体的。如上述定义的分类是一个过程的名称;然而它也可以为这一过程的终结产物所用。于是,分类的结果是一项分类。主要是生物学以外的领域,也在与将实体安排在一连续的谱系、倾差或另一不表示明显划分的排列(标位)相对立而将它们放入明显的类别的严格意义上,使用分类这一术语。我们没有按这种方式限定分类。术语分类也应该同“鉴定”相区别。

鉴定 (Identification)。这是一项分类一经确立就将新加的未鉴定对象分配或指定到正确类别。于是,使用检索表认识黄石国家公园 (Yellowstone National Park) 的野花的人将一既定标本“鉴定”为一种黄花(菊科植物)。有些数学家和哲学家将这一过程也叫做分类,但是我们依从生物学者的常规用法,严格地区分二者。

分类学 (Taxonomy)。Simpson 使用这个术语意为“分类的理论研究,包括其基础、原理、程式和规则”(Simpson, 1961, 第 11 页)。按此定义,我们的这本书的主体内容都涉及分类学。我们想在分类学之下也包括鉴定的理论方面。分类学正如分类,也用以指分类过程的终结产物。因为这是一个被普遍接受的用法,我们将偶而在这个意义上采用它。分类学经常成为分类的同义词,我们不试图严格区分对理论而言的“数值分类学”(numerical taxonomy) 和对实践而言的“数值分类”(numerical classification) 这两术语。Blackwelder (1967a) 以为分类学指“……处理机体种类的日常业务。这包括标本的处理和鉴定,资料的发表,文献的研究和标本所示变异的分析。”这些活动多数是程式性的,与分类学有关,但毫无理论意义。对于它们,我们将不多说。

我们将术语类元 (taxon, 复数 taxa)* 用作任何性质或秩级的分类群 (taxonomic group) 的略语,正如 H. J. Lam (在 Lanjouw, 1950 中) 和 Rickett (1958) 所提议的。

1.2 数值分类学的定义

入题前,需要清楚界定我们的术语“数值分类学”的用法。我们用它来指:借数值方法根据其性状状态将分类单位归类成类元。这一术语包括尽可能地用统计学或其它数学方法从数据引出种系发生的推论。这些方法要求将有关分类实体的信息转换成数值量。我们宁可用“数值分类学”而不是“定量分类学”(quantitative taxonomy),因为后者包括其它

* 尚有分类单元、分类单位、分类群等多种译法。本书统一译作“类元”。——译者。

方法,诸如血清学或纸层析。

鉴于《数值分类学原理》出版后接踵而来的、用以估计分支(的)(cladistic)关系的数值方法的发展,(尤其 W. H. Wagner, 私人通讯)提议应该将整个领域重新命名为“数值系统学”(numerical systematics)。由于术语系统学的含义(见前节)被普遍接受的广度和系统学的无数方面不适合于本书所讨论的途径,我们宁愿保持这确立了的定义。Jardine 和 Sibson (1971) 提出名称“数学分类学”(mathematical taxonomy), 而 Blackith 和 Reyment (1971) 采用“多变量形态度量学”(multivariate morphometrics)。提出了一个同义语“分类度量学”(taxometris) (Mayr, 1966, 代替 Rogers, 1963 所提的 taximetrics), 而我们将偶而使用形容词“分类度量的”(taxometrics) 代替比较累赘的“数值分类的”(numerical taxonomic)。然而,数值分类学这个术语看来已牢固确立,我们保留于此。

我们不想把数值分类学的概念扩大到包括系统学研究上统计的或其他的数值方法的每种应用。我们的途径有多种的数值技术;但当这样的技术不用于分类的问题时,它们就不包括在数值分类学里。

数值分类学的实践,拥有许多基本假定和对分类学工作的基本观点,我们将在下面各节详加讨论和辩解。观点和假定无一是新颖的。它们,以及分类学关系的数值处理上的个别尝试,早在二百多年前就有过。然而,我们毋宁将数值分类学这一名义局限于近年的完整的途径,它引起了分类学理论和实践的修改。

1.3 数值分类学的基本见解

数值分类学的基本见解可以概括于下列原理(改自 Sneath, 1958),体现了可以追溯到一位法国植物学者 Michel Adanson (1727—1806) 的概念,他的观点在第 2.2 节相当详细地加以讨论。由这些原理代表的因此往往被称为新阿当松学派 (Neo-Adansonian) 观点。

1. 一项分类的类元里信息量越大,且分类所根据的性状越多,则既定的分类会越好。
2. 先验地 (a priori), 建立自然的类元时每一性状是等权的。
3. 任何二实体的全面相似性,是它们被比较着的许多性状的每项性状相似性的函数。
4. 能够由性状相关研究的机体组群上存在差别,识别有区别的类元。
5. 给予有关进化途径和机制的某些假定,可以从组群的分类学结构和从性状相关作出种系发生的推论。
6. 分类学是作为一门经验科学来看待和实践的。
7. 分类是以表相相似性为基础的。

原理 1—3 在第 1.4 节《类似性的估计》里阐述;原理 4 在第 1.5 节《类元的构成》中论述;原理 5 在第 1.6 节《种系关系的识别》里讨论;原理 6 和 7 在第 1.7 节《表相分类学和种系发生分类学》中研究。

实践上,数值分类学的操作按如下顺序进行:选取并记录机体和性状;计算机体间的类似性;类元建立在这些类似性上;而最后对类元作概述(诸如有关它们的种系发生的推论,判别性状的选取等等)。

应该明白:识别类元之前不能对类元作出概述;了解机体间的类似性之前不能识别类元;而在检查了机体及其性状之前不能估计这些类似性。因此,虽然这些步骤有的实际

上可以结合成某类计算机方法,或者由于某种特殊缘故整个程式可以第二次重复,但程式内步骤的秩序不能加以改变,而不然就破坏了分类过程的理论基础。

1.4 类似性的估计

类似性的估计是数值分类学上最重要、最根本的一步。它从收集有关要研究的分类群中的性状的信息开始。这种信息可能业已现成而仅需从文献引出,或许可能得全部或部分地从新发现。大多数情况下,这两种程式将都要应用。为了方法可靠,需要许多性状。所有种类的性状,形态的、生理的、行为生物学的、而时常连分布的性状都同样可用。必须警惕选取性状时带着先入的偏见和避免选取无意义的性状(第3.7节)。

照我们每一性状都具有相等分类价值的主张来看(我们对性状的定义见第3.2和3.3节),在用它来评定分类学关系时,这只是走向给每一性状分配等权的实践的小小一步。这是同传统分类学实践直接相悖的一点,对此争论很多。这一论题,我们在第3.9节讨论得相当详细。不可能为分类学性状的先验加权建立标准,这已被十分广泛地承认了。

类似性的测度的实际计算,能够以许多办法完成(第四章)。类似性是用相似性的系数(值域通常在一和零之间,一代表完全相符,零代表毫不相符)或用相异性(距离)的系数(值域通常在零和某未定值之间,前者代表等同,后者代表极大距离或悬殊)来表达的。计算当然是令人厌烦的,而除了很简单的研究,都需要电算机处理。

类似性系数是表列为矩阵形式的,每一分类实体的对偶有一项系数。若为对称(镜像)矩阵,表列 i 个实体,则结果是 $i \times i$ 矩阵。三个以上分类实体间的相似性,可以表示为几何空间中的点。为在此空间里正确表示 i 个点(分类实体),需要极大值为 $i-1$ 的维(数)。点间距离可以视为分类学距离。

1.5 类元的构成

数值分类学上的分类,一般以类似性矩阵为基础,类元是通过为了揭示和概括矩阵结构而设计的技术在矩阵中构成的。按某些方法,分类可以在原始数据矩阵上进上。如果事先对元素作了粗略归类而使假定相似的类型互相接近,用不同深浅的阴影表示矩阵的各种各样元素,就能画出粗略的矩阵结构的图解(见图5-16)。纵令方法上可取,如果实体不按预定次序排在矩阵里,不作调整就不易看出归类。多种用于成聚的计算机方法,将同等有效地处理数据,不论数据是否有序。因为计算方法同时提供一些分类学关系的数值评价,所以它们被采用了。

这些数值方法总称类聚分析(见第5.4和5.5节)。它们是从 $i \times i$ 类似性矩阵来建立和界定彼此相似实体的类聚的方法。这些类聚可以看作地形图上的山和峰,而建立类聚用的标准和地形图上的等高线相类。严格的标准相当于孤立高峰——例如种间类似性矩阵中的种群(species groups)周围的高海拔线。标准放宽时,类聚扩大而相接,有如孤峰随着从高海拔过渡到低海拔等高线而底部变宽广并连成峰群直到形成山脉。类聚一般仅以表相类似性为基础而没有必然的种系内涵。成聚方法的差别主要涉及在分类学(性状)空间里形成类聚和划分机体的规则。

所有这些方法的主要共同点是：给定了关系系数矩阵就能客观地为分类群划界。用于分类群的界线可以形象化为业已讨论过的等高线，或者它们可以用水平断面同数值分类学上常用的关系树状图的分枝相交来表示。对特定研究内的所有分类群能画出可比较的界线。类似性水平逐渐降低的界线或断面会产生分类秩级越来越高的类元。

界线或断面的数目和位置应按某种预先排列的系统而定。显然它们将取决于相似性矩阵的大小：界线太多给出的分类太细；太少则遗留下许多结构没有揭示。研究者的目标、特定组群中的惯例、方便问题和审美，都会影响界线的设置。我们认为，要确定的类元数目在任何秩级都是相对次要而任意的细节。但界线一经划定，类元内部结构（而因此其中所含低秩类元的数目）就完全取决于实体的类似性值，并且是不受研究者支配的。换言之，界线或断面的位置和数目是任意的，但是它们在考虑之中的分类学空间的全部区域里，都必须根据可比较的标准。

我们对单型类元或很大的类元一视同仁。它们的出现并不使我们分别加以归并或分裂。是因为使类元具有一些客观的和可定义的群内凝聚的标准的努力，促使了我们持这样一种立场。这大多将取决于类聚形成的方法（第 5.3 和 5.4 节）。

应用分类学研究结果的生物学者更关心的是分类的稳定性。根据数值分类学的方案的稳定性，可能受到两个方面的影响。

1. 可能积累更多的（成新性状形式的）信息。如果原先的类似性评价已经根据大的性状样品，则我们的看法是：增加更多性状，相对相似性也变化不大（第 3.8 节）。

2. 在后继的研究中可能包含新的分类实体。应用原先的水平和断面数目的标准，可以得出新的不同类元。关于所建立的系统是否应该作适合新结果的调整或新数据应否按业已确立的关系标准来判断——即，界线应否继续划在首次研究所画的水平上，将由应用数值分类学的人们来取得一致（见第 5.10 节）。

在等级制的何种水平上我们称亚属、属和科？这些术语除作为界线的相对水平的指征外，它们本身还有什么意义？属（和其它类目）在诸如昆虫、鸟类和有花植物那样歧异的组群中并不代表等价关系的类元，这在常规系统学上是普遍接受的，尽管类目会根据完全不同的性状群（在第 5.11 节讨论），数值分类学能建立这样的等价类目否？看来，更为可取的是数值分类学所确立的等级系统采用新的一系列术语，术语里包括组群成员间的相似性的定量估计。表元（phenon）* 这一术语（见第 5.11 节）是这种新系统的一次尝试，别的更高级的方法有待发展。

构成自然类元的一个重要方面是它们的表达法。这可以借指明表相关系的树状图、网络（graph）或表相超空间中的类元三维模型和此类模型的二维射影或立体像来实现（见第 5.9 节）。

二维或三维空间中的分类学关系表达法（标位；见第 5.6 节），导向分类学结构型式的调查研究，有着很大的生态和进化意义（见第 5.14 和 7.3 节）。

1.6 种系关系的识别

尽管数值分类学的主要重点，至今放在稳定的表相群的建立上，但方法论也为种系传

* 或译“分浓”——译者。

代线和进化的研究提供了建设性途径(第六章)。不同地质时期的机体间的表相相似性的测度,将提供关于进化的速率和方向的客观信息,或者可能有助于一些地层学问题的解决。借表相差异与遗传或地理数据的比较,也能为研究种的形成获得许多有用信息。

近年见到了重新强调用运筹的途径研究种系或种系发生。正如我们在第 2.3 节所解释的,这两个词有着有时交迭的含意。这样的分析,旨在估计一项研究中类元所表达的进化路线的分枝顺序。基本数据包括常规形态性状以及蛋白质的氨基酸顺序。迄今所用的方法,力图找出确定类元表相型式所需的过吝的(极小长度)种系发生树。此类证据已被种系发生学者采用多年。新的东西是程式的系统化和同时考虑许多性状,分给它们先验的等权,但有时在根据一些运筹标准成聚时给它们加权。用这些方法和其它发展中的方法,进化传代线是以电算机来计算的。

1.7 表相分类学和种系发生分类学

Sokal 和 Camin (1965) 区分了分类学上运筹的 (operational)、经验的和数值的途径,认为它们的重要性也顺此次序排列。在运筹分类学 (operational taxonomy) 上,关于自然界的陈述和假说必须服从于能由观测或实验来检验的、有意义的问题(见第 2.1 节)。经验分类学 (empirical taxonomy) 是以许多观测到和记录了的性状为基础的,类元就按大量共有性状归类 (Simpson, 1961, Sokal, 1962b)。用以决定共有性状数目的方法,因经验学派而异。它可以是定量的,可以不是。一般实践的数值分类学,既是运筹的又是经验的。

本章前几节中概述的分类学程式,是严格经验性质的。就此而论,它们与以往一些模式论者 (typologist) 的程式有关。然而,正如 Simpson (1961) 暗示地说到了,这不是自动取消我们的观点[关于“统计类型论”(statistical typology)的辩护,见 Sokal, 1962b]。分类学上对经验论有效性的根本检验,必定是它能否用作排列有机界的、必从而一致的方法。我们认为能这样用并且它是唯一合理的途径。

数值分类学者的基本看法是将种系发生的臆测同分类学程式严加区分。分类学关系是纯粹根据现在手上材料的类似性来评价的。这些表相关系并不考虑所找到的类似性的起源,也不考虑类似性在以往可能增减了的速率。

持此看法,正如在第 2.6 节详细讨论的,是因为此刻我们并不拥有一项图解式的或别样的分类方案,能够同时由之得出类似性的程度、血缘和进化发展速率的信息。任何试图兼有这些途径的方案都会过分错综纷繁。全面相似性[表相学 (phenetics)] 和进化分枝顺序[分支学 (cladistics)] 的分离,是分类学思想的重要进步(详细讨论见第 2.3 节)。鉴于现在一代多数从业分类学者由初期训练学得的思想方法确立已久,作这种分离是困难的。不过,刚从研究院出来的最新一代分类学者对这一概念困难不大,并且承认关于分类学关系的各别方法的价值。

我们不但坚持在分类程式上将表相的考虑理由与种系发生的分开,并且我们觉得只有表相学证据才能用来建立满意的分类。我们持此信念有几项理由。

1. 可以得到的化石记录如此零碎,以致绝大多数类元是未知的。进化分枝顺序必须从现存机体中的表相关系大体推知。

2. 表相分类可用于所有组群。反之,以分枝顺序为基础的分支分类,需要关于机体组群进化方向的历史推论。

3. 即便可以得到化石证据,首先也必须以严格的表相方式来解释证据本身——附上能限定若干种系发生解释的时间标度者除外——因为在种系发生上,选取祖型的标准是表相的,并且取决于推定祖先和后裔间的表相关系。

4. 一般从生物学观点来看,描述机体的全面相似性或许比描述其分枝顺序更有意义。若要分类有预测性价值,很显然,基于全面相似性的分类将是最富预测性的。

对许多当代生物学者说来,仅据表相学标准的分类学关系是难以接受的。分类学的和种系发生的论证间的紧密概念结合,几乎已有一个世纪了,所以诸如“特化的”(specialized)、“原始的”(primitive)、“同功的”(analogous)以及许多其它术语,已经具有双重含义,其差异很少分清。不过,必须作出努力,澄清其含义并在分类原理的理解上取得进展。

在本节扼要讲述的论题,将在第 2.2—2.5 节更详细讨论。

1.8 数值分类学的优点

本书所讨论的方法有许多优点,将编列于下。然而,首先我们应该提及它们的主要目标是可重复性(repeatability)和客观性(objectivity)。如果在接受的误差范围内可以重复观测,如果分类程式明确规定,数值方法就有希望引导采用相同数据基础而独立工作的不同科学家,去取得可比较的、任何组群机体间的类似性估计。

客观性的提法是与可重复性密切联系着的。《兰德豪斯英语辞典》(The Random House Dictionary of the English Language)将“客观的”定义为“与个人感情和成见无关的;基于事实的;不偏不倚的。”这是一个相对的概念,难得圆满实现。通过将许多未经预先任意选择或剔除的性状包括在内,通过提供处理数据和估计结果的明确方法,我们减少了主观偏见而增加了客观性。

关于数值分类学的种种优点,我们可以简略提及如下,我们将在正文适当章节更详细地加以讨论。

1. 数值分类学有综合多种来源数据的能力,诸如形态学、生理学、化学、DNA 线间亲合性、蛋白质的氨基酸顺序等等。这是用常规分类学很难办到的。

2. 经过大部分分类过程的自动化,效能大为提高(Sokal 和 Sneath, 1966)。于是,许多分类学工作可以由不太熟练的工作者或自动装置来做。

3. 以数值形式编码的数据,在分类学研究机关能用现成的电子数据处理系统来综合,并能用于编制记述、检索、目录、地图和其它文件。

4. 由于方法是定量的,就循着分类差分谱提供了更大示差判别且在划分类元上更为灵敏。于是,它们应该给出比由常规方法所能得到的更好的分类和检索。

5. 用于数值分类学的明晰的数据表的制定,已迫使这个领域的工作者使用更多的、描述得更好的性状。这也必将改善常规分类学的质量。

6. 数值分类学的基本优点是分类学原理和分类目的的重新审查。这一般是对分类学有利的,并且已经导致一些根本问题的提出,这些问题在本书后面的许多章节里、在近年系统学文献中的许多出版物里作了讨论。