

构建数据矩阵的方法之三项情形分析初探

袁文伟^{1,2}

(1. 中科院南京地质古生物研究所,江苏 南京 210008; 2. 现代古生物学和地层学国家重点实验室,江苏 南京 210008)

摘要:三项情形分析是分支分析中一种构建数据矩阵的方法。介绍了这种数据矩阵的构建、执行及其与传统的标准数据矩阵的区别,讨论了三项情形分析的优点和存在的争议。

关键词:三项情形分析;分支分析;数据矩阵;构建

中图分类号:Q 91-3 **文献标识码:**A **文章编号:**1674-3636(2009)02-0184-03

0 引言

由德国昆虫学家 Willi Hennig 明确提出的分支分析方法在古生物学研究中的应用已经越来越广泛。毫无疑问,构建数据矩阵是分支分析中最重要的一环之一,但其方法很少在科学文献中提及。构建数据矩阵的标准方法(standard approach)是以特征为行,类元为列构建矩阵,其中的每个单元为某特征在某类元中所表现的状态。但是, Nelson and Platnick 提出了另一种构建数据矩阵的方法,称为三项情形分析(three-item statement analysis)。三项情形分析是基于“类元”和“同源”代表相同关系的另一种数据编码方法。它表达的是各个特征所反映的分类关系中的最小可能组合,也就是三类元关系。通过简约分析,三项情形矩阵通常可以得到比标准矩阵解析度更高的分支图。笔者简要介绍三项情形方法中数据矩阵的构建、执行,和标准数据矩阵的区别,以及该方法的优点和存在的争议。

1 三项表达数据矩阵的构建

引用 Nelson 和 Platnick(1991)的例子解释三项表达数据矩阵的构建。考虑一个有 4 个类元(A-D),6 个二态特征(1-6)的标准数据矩阵 1(图 1)。

矩阵 1

类元	特征					
	1	2	3	4	5	6
A	1	0	1	0	0	0
B	1	0	0	0	1	1
C	0	1	1	1	0	1
D	0	1	0	1	1	1

图 1 一个具有 4 个类元、6 个二态特征的标准数据矩阵

每一个特征都可以用三项表达的组合作表示。例如,假定状态(0)为祖态,特征 1 包含了 2 个三项表达 C(AB)和 D(AB),每一表达都可视为 A 和 B 具有一个相对于 C 或 D 的近裔共性,另一类元可以用缺失值(?)编码,即图 2。

类元	三项表达	
	1a	1b
A	1	1
B	1	1
C	0	?
D	?	0

图 2 图 1 中特征 1 的三项表达编码

特征 2-5 和 1 类似,都只在 2 个类元存在裔态。特征 6 的裔态存在于 3 个类元,因此可以用 3 个三项表达的组合作表示:A(BC)、A(BD)和 A(CD)。这样我们就得到了对所有 6 个特征以三项表达编码的矩阵 2(图 3)。

收稿日期:2008-10-13;编辑:齐运铎

基金项目:国家自然科学基金(编号 40772007);科学院方向性项目(编号 KZCX3-SW-149);现代古生物学和地层学国家重点实验室(中科院南京地质古生物研究所)项目(编号 033107)资助

作者简介:袁文伟(1969—),男,江苏南京人,博士,副研究员,主要从事三叶虫和早古生代地层研究。

执行 Hennig 86 分析标准数据矩阵 1 得到 2 个同等最简约分支图 $XA(B(CD))$ 和 $X(AB)(CD)$, 这 2 个图的区别在于类元 A 和 B 的位置差异。而分析三项情形表达矩阵 2 只得到单一的最简约分支图 $XA(B(CD))$, 也就是将 $(B(CD))$ 视为一个单系组分。这是因为类元 B 有 3 个裔征, 其中特征 1 支持组分 (AB) , 而特征 5 和 6 支持组分 (BCD) 。

自从三项表达方法提出以后, 也有一些对此的争论, 有些学者就认为这种方法存在事先假定特征演化不可逆的缺陷。但笔者认为, 假定特征演化不可逆不应看作三项分析的缺陷, 这只是符合了简约法则的要求, 简约法则将逆转一律视为同塑。因为从数学定义上看, 等级可以理解为相当于一种嵌套关系。而承认可逆性的存在相当于允许出现(中国(南京(中国)))这样的嵌套关系, 这显然在数学上就已经是悖论。

5 致 谢

特别感谢法国巴黎第 6 大学的 René Zaragüeta i Bagils 博士和笔者就文章内容进行的讨论。感谢国家自然科学基金(40772007)、科学院方向性项目(KZCX3-SW-149)、现代古生物学和地层学国家重点实验室(中科院南京地质古生物研究所)项目

(033107)资助。

参考文献:

- [1] HENNING W. Grundzüge einer Theorie der phylogenetischen Systematik[M]. Berlin;Deutsche Zentralverlag,1950.
- [2] NELSON G J, PLATNICK N I. Three - taxon statements: a more precise use of parsimony[J]. Cladistics,1991,7: 351 - 366.
- [3] NELSON G J. Homology and systematic[C]//HALL B K. Homology:the hierarchical basis of comparative biology. San Diego;Academic Press,1994.
- [4] NELSON G J, LADIGES PY. Information content and fractional weight of three-taxon statements[J]. Systematic Biology,1992,41:490 - 494.
- [5] KITCHING I J,FOREY P L,HIMPHRIES C J,et al. Cladistics: The theory and practice of parsimony analysis [M]. Oxford;Oxford University Press,1998.
- [6] PLATNICK N I, GRISWOLD C E, CODDINGTON J A. On missing entries in cladistic analysis [J]. Cladistics, 1991,7:337 - 343.
- [7] NELSON G, WILLIAMS D M, EBACH M C. A question of conflict: three-item and standard parsimony compared [J]. Systematics and Biodiversity,2003,1:145 - 149.
- [8] LAET J D, SMETS E. On the three - taxon approach to parsimony analysis[J]. Cladistics,1998,14:363 - 381.

On three-item statement analysis of constructing data matrix

YUAN Wen-wei^{1,2}

(1. Nanjing Institute of Geology and Paleontology, Chinese Academy of Sciences, Nanjing 210008, China; 2. State Key Laboratory of Paleobiology and Stratigraphy, Nanjing 210008, China)

Abstract: Three item statements analysis codes data was based on the idea that ‘taxon’ and ‘homology’ representing the same relationship. The author presented the coding, implementation of the three item statements analysis, and the differences between the new data matrix and the standard data matrix. The debates on the three item statements analysis were discussed as well in the text.

Keywords: Three-item statements analysis; Cladistics; Data matrix; Construction