

# 科技期刊选订的系统聚类分析与决策

黄 敏 都平平 王 群(中国矿业大学图书馆)

**【摘要】** 论文选择期刊类型、质量、信息量和订购价格四类指标作为聚类要素,对某一学科领域内的 23 种中文科技期刊进行了系统聚类分析。当聚类距离  $\lambda=3.29$  时,可将这 23 种科技期刊划分为 7 类。在分析各类期刊优缺点的基础上,合理确定其订购原则,旨在对高校图书馆科技期刊选订工作产生一定的指导作用。

**【关键词】** 科技期刊;选订;系统聚类分析

**【Abstract】** Choosing periodical type, quality, information content and subscription price as cluster elements, we carried out a systematic cluster analysis with 23 kinds of Chinese sci-tech periodicals in an academic subject. When the cluster distance is 3.29, these 23 kinds of sci-tech periodicals can be divided into 7 categories. On the basis of advantages and disadvantages analysis of each category, the subscription principle can be determined, which aims to have some important guiding significance on selected subscribing of sci-tech periodicals for the university library.

**【Keywords】** Sci-tech periodical; Selected subscribing; Systematic cluster analysis

科技期刊选订是一个多目标决策问题。首先需要根据期刊特征进行系统分类,然后根据分类结果进行合理选择,因此聚类分析成为科技期刊选订优化决策的重要前提。本文以 23 种某一学科领域内的中文科技期刊为例,选择期刊类型、质量、信息量、订购价格四类指标作为聚类要素进行了系统聚类分析,并在探讨各类期刊优缺点的基础上确定了订购原则。

## 1 聚类要素选择与数据处理

在聚类分析中,聚类要素的选择非常重要,它直接影响到分类结果的准确性与可靠性。科技期刊选订决策分类不同于简单的期刊质量评价,需要考虑期刊类型、内容丰富度、订购价格等多种因素,因此需要选择多种指标作为聚类要素。本文选择了期刊类型、质量、信息量和订购价格四类指标作为聚类要素。

### 1.1 期刊类型

期刊类型是高校图书馆选订期刊时需要考虑的一个重要方面。科技期刊根据其发文内容可划分为三类:综合研究型、专业研究型和政策宣传型。由于期刊类型无法用具体特征值表示,因此需要人为定义聚类要素的数值。笔者根据各类型期刊亲疏关系所定义的特征值如图 1 所示。

### 1.2 期刊质量

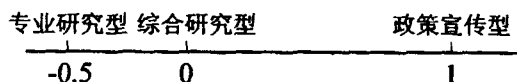


图 1 期刊类型亲疏关系图

期刊质量评价指标与体系有多种。本文选择影响因子、被引次数、引用发文比三个典型评价指标来表示期刊质量。在聚类分析过程中,期刊质量只能作为一个聚类要素,因此在聚类分析前需要将数据在标准化处理的基础上进行合并,取算术平均值作为聚类要素值。数据标准化处理及合并方法为:

$$x'_{ij} = \frac{1}{m} \sum_{i=1}^m \frac{x_{ij} - \bar{x}_j}{S_j}$$

其中:

$$\bar{x}_j = \sum_{i=1}^m x_{ij} / n \quad S_j = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}$$

$$(i = 1, 2, \dots, m; j = 1, 2, \dots, n)$$

### 1.3 期刊信息量

科技期刊内容丰富程度是影响高校图书馆期刊选订的一个重要因素。当前,期刊信息量评价主要有两个指标:期发文量与年发文量。在进行聚类分析前,同样需要将表示期刊信息量的数据在标准化处理的基础上进行合并,取两者算术平均值作为聚类要素值。

### 1.4 订购价格

在图书馆采购经费有限的情况下，订购价格成为影响期刊选订决策的关键因素。期订购价与年订购价是评价期刊订购成本的两项基本指标。期刊订购价格聚类要素值的确定同期刊质量、信息量数据处理及合并方法。

2 聚类要素数据统计及处理结果

根据上述方法对这 23 种中文科技期刊的聚类要素数据进行统计和处理,结果如表 1 所示。

表 1 聚类要素数据统计及处理结果

期刊 编号	期刊 类型	期刊质量				期刊信息量			订购价格		
		要素 数值	影响 因子	被引 次数	引次 发文	要素 数值	期发 文数	年发 文数	要素 数值	期订 购价	年订 购价
1	0	1.93	0.46	977	5.37	-0.78	15.2	182	3.20	26	312
2	0	2.23	0.68	867	5.22	-0.51	27.7	166	0.23	15	90
3	0	-0.14	0.23	433	1.20	-0.08	30.2	362	0.10	11	132
4	0	1.33	0.55	815	2.70	-0.31	25.2	302	-0.10	10	120
5	0	1.37	0.52	611	4.33	-0.65	23.5	141	-0.89	8	48
6	-0.5	-0.81	0.10	211	0.62	-0.17	28.3	339	0.52	13	156
7	-0.5	-0.61	0.16	200	1.06	0.01	47.3	189	-0.37	12.5	50
8	-0.5	-0.56	0.08	468	0.45	2.52	86.3	1035	0.93	15	180
9	-0.5	-0.97	0.08	141	0.54	0.06	43.8	263	-0.09	13	78
10	1	-0.55	0.15	308	0.85	-0.08	30.3	364	0.31	12	144
11	1	0.42	0.34	582	1.89	-0.29	25.7	308	-0.10	10	120
12	0	0.20	0.32	537	1.34	0.87	66.8	401	0.23	15	90
13	0	-0.99	0.09	154	0.26	0.82	49.6	595	0.10	11	132
14	-0.5	0.72	0.44	476	2.92	-0.52	27.2	163	-0.57	10	60
15	-0.5	0.34	0.30	412	2.84	-0.63	24.2	145	-0.25	12	72
16	-0.5	-0.32	0.15	542	0.48	2.92	94.9	1139	0.93	15	180
17	0	-0.29	0.30	250	0.92	0.11	45.2	271	-1.05	7	42
18	-0.5	-0.58	0.17	191	1.18	-0.53	27	162	-0.57	10	60
19	-0.5	-0.85	0.17	99	0.46	-0.22	35.7	214	-0.89	8	48
20	0	-0.72	0.25	65	0.46	-0.37	35.3	141	-0.73	10	40
21	0	-0.49	0.27	109	1.15	-0.73	23.8	95	-1.02	8	32
22	-0.5	-0.02	0.32	328	1.53	-0.66	17.8	214	-0.93	6	72
23	-0.5	-0.63	0.21	122	1.03	-0.78	19.8	119	1.03	20	120

注:影响因子、被引次数、引次/发文、年发文量数据来源于维普咨询网期刊评价 2006 年统计数据。

3 系统聚类分析与结果

在进行系统聚类过程中，首先需将聚类要素数据进行二次标准化转换,公式为：

$$x'_{ij} = \frac{x_{ij} - x_j}{S_j}$$

$(i = 1,2,\cdots,m; j = 1,2,\cdots,n)$

然后采用欧氏聚类距离、离差平方和法进行系统聚类分析。欧氏距离计算公式为：

$$d_{ij} = \sqrt{\sum_{k=1}^n (x_{ik} - x_{jk})^2}$$

$(i,j = 1,2,\cdots,m;k = 1,2,\cdots,n)$

采用上述方法得到的系统聚类分析结果如图 2 所示。

4 聚类分析对期刊选订的指导作用

经过系统聚类分析后,各类期刊在类型、质量、内容丰富性和订购价格方面都具有其鲜明的特点。对期刊选订工作者来说,通过分析各类期刊的优缺点来确定其订购原则,是实现高校图书馆科学合理选订期刊的前提。根据图 2 系统聚类分析结果,当聚类距离  $\lambda=3.29$  时,可将这 23 种中文科技期刊划分为 7 类。分析各类期刊的优缺点及订购原则结果如表 2 所示。

表2 各类期刊的优缺点及订购原则

类号	期刊编号	期刊类型	优点	缺点	订购原则
I	1	综合研究型	质量高、影响范围较大	发文量少,订购价格高	经费充足情况下优先订购
II	2,4,5	综合研究型	质量高,影响范围大	发文量较少	经费有限情况下优先订购
III	10,11	政策宣传型	宣传行业最新政策	刊发研究型论文量少	在调查读者需求的基础上选择性订购
IV	3,12,13,17,20,21	综合研究型	发文量较大,订购价格较低	质量一般,影响范围不大	经费有限情况下选择性订购
V	6,7,9,18,19,23	专业研究型	刊发研究型论文专业性强	质量一般,影响范围小	结合高校专业特色及经费情况选择性订购
VI	14,15,22	专业研究型	质量较高,订购价格低	发文量少	结合高校专业特色优先订购
VII	8,16	专业研究型	内容丰富,发文量大	订购价格高,影响范围不大	结合高校专业特色选择性订购

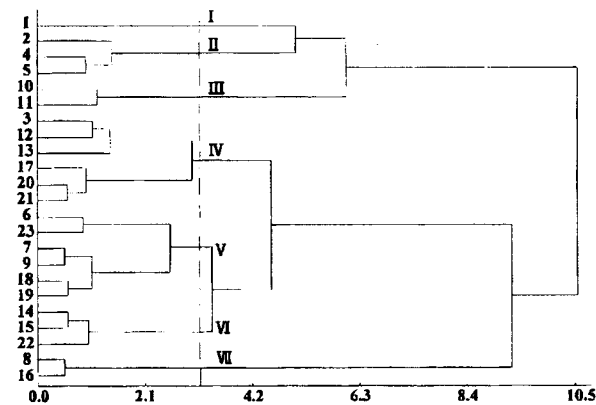


图2 系统聚类分析结果

5 结论

选择期刊类型、质量、信息量和订购价格四类指标作为聚类要素，以系统聚类分析的方法对某一学科领域内的23种中文科技期刊进行了科学分类,并通过对各类期刊优缺点的分析合理确定了其订购原则。研究表明,系统聚类分析方法对高校图书馆的科技期刊选订工作具有重要的指导意义。

参考文献

1 刘振西.期刊变化对管理利用的影响及对策.图书馆杂志, 2001(5):23~26

2 王晓慧. 科技期刊选订的模糊聚类决策. 情报探索,2008 (5):32~33

3 王筠.科技期刊选订的灰色聚类决策.情报学报,2001,20 (1):112~115

4 张红芹,黄水清.期刊质量评价指标研究综述.图书馆理论与实践,2008(3):20~23

(作者信息:黄敏,中国矿业大学图书馆助理馆员;都平平,中

国矿业大学图书馆研究馆员;王群,中国矿业大学图书馆助理馆员,邮编 221116。收稿日期:2009-02-06。) 编校:刘勇定

无锡市图书馆荣膺  
“全国古籍重点保护单位”称号

继去年首批48部珍贵善本入选《国家珍贵古籍名录》后,无锡市图书馆日前又有22部古籍入选,无锡市图书馆也被评为“全国古籍重点保护单位”。目前,锡图入选“古籍国榜”的善本数量在江苏省排名第二,在全国地市级图书馆中处于领先地位。

无锡市图书馆藏有古籍30多万册,其中善本古籍2.15万册、无锡地方文献1万余册、旧家谱0.8万册以及荣德生、荣毅仁父子捐赠的图书11万余册,均为特色馆藏。首批48部古籍入选《国家珍贵古籍名录》更是位于全国同级图书馆之冠,引起了业内的高度关注。

在此次入选的22部古籍善本中,《先德集一卷》明弘光刻本和《日本国考略一卷补遗一卷》明嘉靖九年王文光刻本是仅存于世的“孤本”。《先德集一卷》是明代秦鏞(寄畅园秦氏后人)为其父亲所辑的人物传记,其开篇即高攀龙所写的《秦文学彦熙墓志铭》。而介绍日本地理的《日本国考略一卷补遗一卷》,虽已出现微小的损坏现象,但是因其“独一无二”的价值,毫无争议地被列入了名录。由明代无锡人顾经起注的《类笺唐王右丞诗集》也特别引人关注。这套古籍据称是现存最早的王维诗注本,代表了明代嘉靖刻本的最高水平。