

对引文分析专题高被引论文的同被引聚类分析

赵悦阳

(中国医科大学附属盛京医院图书馆 沈阳 110004)

[摘要] 针对引文分析这一专题,通过同被引聚类分析和战略坐标图,确定该领域若干个研究分支的发展情况,进而对该领域的研究状况有一个系统的认识。指出同被引聚类分析具有避免主观性、简单易行的优势,战略坐标图的引入能使结果更加直观。

[关键词] 引文分析;聚类;战略坐标图

Co-citation Clustering Analysis of Highly Cited Papers in Citation Analysis Subject ZHAO Yue-yang, Library of Shengjing Hospital of China Medical University, Shenyang 110004, China

[Abstract] Aiming the topics of citation analysis, the paper defines the development status of several study branches in this field through co-citation clustering analysis and strategy coordinate diagram, thus systematically understanding about current research situation in this area. It points out co-citation clustering analysis has advantages of avoiding subjective judgment and easy to use, the introduction of strategy coordinate diagram could make the displaying of the results much more directly.

[Keywords] Citation analysis; Clustering; Strategic coordinate diagram

同被引聚类分析是文献计量学的主要方法之一,可以用来表示某一学科或专题的研究结构和状况^[1]。对某学科和专题的高被引论文的同被引聚类分析可以动态地反映该学科和专题的变化情况^[2]。本研究选择引文分析这一专题,力图通过该专题高被引论文的同被引聚类分析,进一步画出战略坐标图,确定该领域若干个研究分支的发展情况,进而对该领域的研究状况有一个系统的认识。

1 资料与方法

[修回日期] 2010-12-16

[作者简介] 赵悦阳,硕士研究生,助理研究员,发表论文 5 篇。

1.1 检索文献

利用 ISI Web of Knowledge,以“Citation”为检索词,检索 2000-2009 年与引文分析有关的文献,并套录每篇文献的引文(包括引文的作者、年代、发表刊名缩写等内容)。

1.2 记录处理

采用书目信息共现挖掘系统(Bibliographic Item Co-occurrence Mining System, BICOMS)^[3]处理记录,截取被引次数超过某一阈值的论文,处理得到这些论文的同被引相关矩阵。对于这些高被引论文,查找其研究主题。对所得的相关矩阵,导入到 SPSS 进行聚类分析,形成聚类分析的树状图。然后,根据所得的树状图,选择恰当的距离,分出数

目较多、较清晰的若干类。根据题名、主题词和副主题词概括每一类内容，命名各类。

1.3 制作战略坐标图

战略坐标 (Strategic Diagram) 是 Law^[4] 等人于 1988 年提出的, 用来描述研究领域内部受关注和新颖情况。在战略坐标图中, X 轴表示某个领域出现的时期, 也就是新颖度, 通过每一类平均发表时间的离均差来计算; Y 轴表示某个领域受关注的强度, 也就是关注度, 用每一类平均被引次数的离均差来计算。

2 结果与分析

2.1 高频引文

将检索得到的 175 194 篇引文, 由书目信息共现挖掘系统统计出被引频次超过 50 次的 31 篇高频引文及同被引相关矩阵。

2.2 聚类分析

图 1 为运用 SPSS 所做出的聚类树图。选择 7.5 (距离) 为分类标准, 根据文献的题目、主题词和副主题词, 将这 31 篇高频引文归纳成 13 类: 第 1 类是 meta 分析研究; 第 2 类是应用文献计量学方法进行随机对照实验的质量研究; 第 3 类是影响因子

可靠性研究; 第 4 类是正确认识影响因子的研究; 第 5 类是同被引分析; 第 6 类是 meta 分析在应用中的改进研究; 第 7 类是引文分析应用: 分析国家、杂志等研究; 第 8 类是引文分析中存在的问题研究; 第 9 类是论文之间网络结构研究; 第 10 类是引文分析总论; 第 11 类是引文索引总论; 第 12 类是 h 指数研究; 第 13 类是影响因子不适用于评价研究。

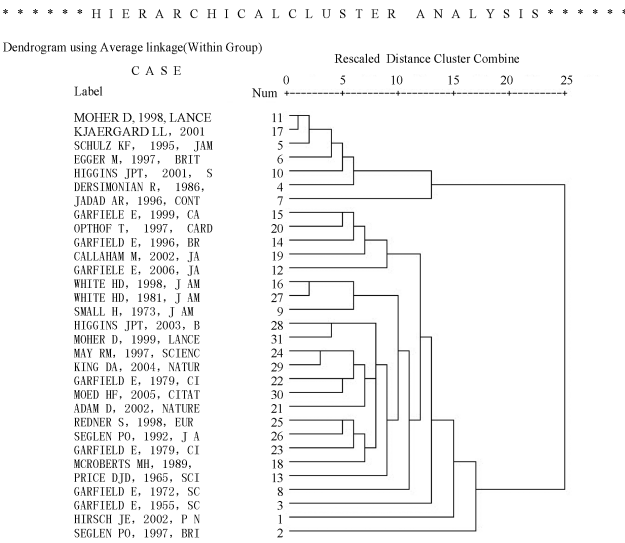


图 1 聚类树图

2.3 战略坐标图

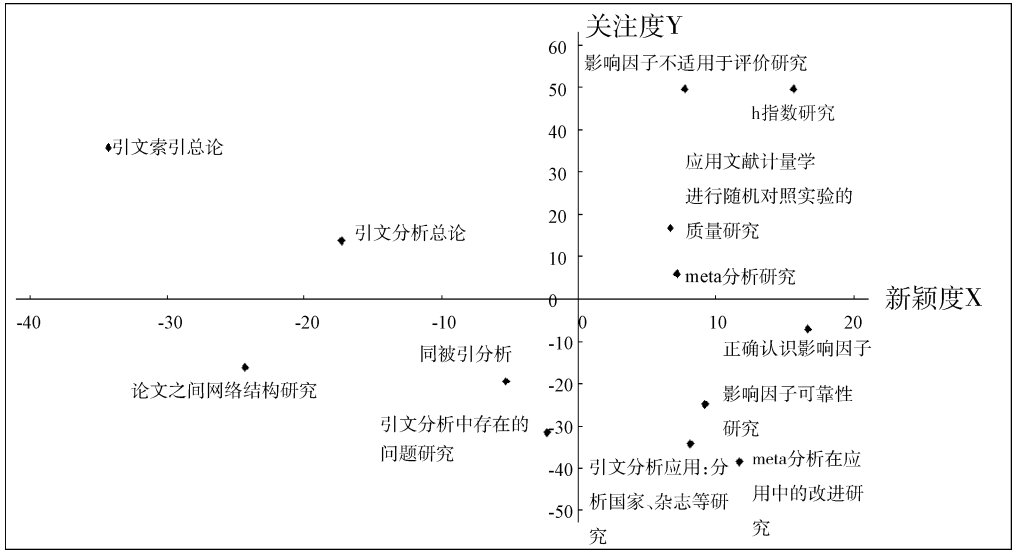


图 2 战略坐标图

图2为战略坐标图。以原点为中心, X, Y轴为分界限, 右上的是第1象限, 在这一象限中的内容代表新颖度和关注度都很高; 左上的是第2象限, 代表新颖性较差却比较受关注; 左下的是第3象限, 代表比较过时而且不再受到关注; 右下的是第4象限, 代表比较新颖却还没受到关注和重视。第1象限: h指数研究的平均发表时间是2005年, 平均被引次数是139次; 影响因子不适用于评价研究的平均值分别是1997年和139次; 应用文献计量学方法进行随机对照实验的质量研究的平均值分别是1996年和106次; meta分析研究这一类的平均值分别是1997年和95.167次。这几类都是平均发表时间较近而且被引次数都很多的, 说明这些研究比较前沿、受关注较多。第2象限: 引文索引研究平均发表时间是1955年, 平均被引次数是125次; 引文分析总论研究平均值分别是1972年和103次。这两类虽然发表时间较久远但都属于最基础的研究, 所以还是比较受关注的, 因此都位于第2象限。第3象限: 同被引分析平均发表时间是1984年, 平均被引次数是69.667次; 论文之间网络结构研究平均值分别是1965年和73次; 引文分析中存在的问题研究平均值分别是1987年和57.5次。这3类研究比较过时, 不再受到太多关注。第4象限: 正确认识影响因子研究平均发表时间是2006年, 平均被引次数是82次; 影响因子可靠性研究平均值分别是1999年和64.25次; 引文分析的应用研究平均值分别是1997年和55次; meta分析在应用中的改进研究平均值分别是2001年和50.5次。这些研究也都是属于比较前沿的方面, 但是却未受到太多关注的。

3 结论

本次研究通过聚类树图和战略坐标图比较好地反映了引文分析领域各分支学科的发展及受关注情况, 反映了一些比较陈旧而且受关注程度差的分支, 同时也提示一些目前比较新颖的有待研究解决的分支学科。由于选取的年代距离短, 文献的数量有限, 会导致分析不够全面。另外高频频次和距离的选择也会影响结果, 这也是需要进一步研究和解决的问题。

同被引聚类分析避免了依靠主观判断的传统方法所带来的主观性, 同时也为数据的积累和分析提供了简便的路径, 比起通过阅读剖析两篇文献的相关性要简便易行。战略坐标图的加入使结果更直观, 对各分支学科的了解更具体。

参考文献

- 1 Small H G, et al. The Structure of Scientific Literature I: identifying and graphing specialties [J]. Science Studies, 1974, (4): 17-40.
- 2 崔雷. 专题文献高被引论文的连续同被引聚类分析 [J]. 情报理论与实践, 1996, 19 (1): 46-48.
- 3 崔雷, 刘伟, 闫雷, 等. 文献数据库中书目信息共现挖掘系统的开发 [J]. 现代图书情报技术, 2008, (8): 70-75.
- 4 Law J, Bauin S, Courtial J-P, et al. Policy and the Mapping of Scientific Change: a co-word analysis of research into environmental acidification [J]. Scientometrics, 1988, 14 (3-4): 251-264.

敬告作者

《医学信息学杂志》网站现已开通, 投稿作者请登录期刊网站: <http://www.yxxxx.ac.cn>, 在线注册并投稿。

《医学信息学杂志》编辑部