武汉大学学术丛书

Wuhan University Academic Library

学术期刊主题可视化研究

安璐 著

武汉大学出版社
WUHAN UNIVERSITY PRESS

武汉大学人文社会科学学术丛书
湖北省社会科学基金项目『十一五』规划资助课题

学术期刊主题可视化研究

武汉大学学术丛书

安璐 著

武汉大学出版社
WUHAN UNIVERSITY PRESS

图书在版编目(CIP)数据

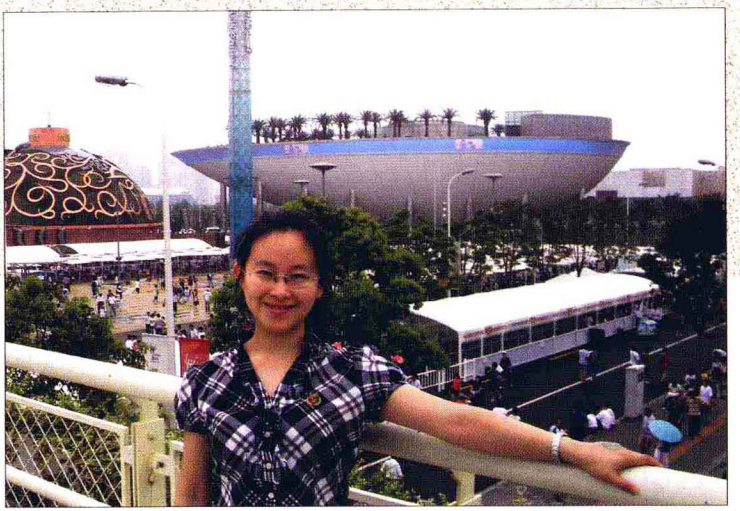
学术期刊主题可视化研究/安璐著.—武汉：武汉大学出版社 2011.10
武汉大学学术丛书
ISBN 978-7-307-08915-0

Ⅰ.学… Ⅱ.安… Ⅲ.可视化软件—应用—学术期刊—主题法—研究 Ⅳ.①G237.5 ②G254.2

中国版本图书馆 CIP 数据核字(2011)第 131108 号

责任编辑:李 琼 责任校对:刘 欣 版式设计:支 笛

出版发行:武汉大学出版社 (430072 武昌 珞珈山)
(电子邮件:cbs22@whu.edu.cn 网址:www.wdp.com.cn)
印刷:武汉中远印务有限公司
开本:720×1000 1/16 印张:14.75 字数:208 千字 插页:3
版次:2011 年 10 月第 1 版 2011 年 10 月第 1 次印刷
ISBN 978-7-307-08915-0/G·2068 定价:32.00 元



安　璐　1979年生，情报学博士，讲师。1997年考入武汉大学信息管理学院信息管理与信息系统专业，2001年获得信息管理与信息系统学士学位，并被推荐免试攻读情报学硕士研究生，2004年获情报学硕士学位。2007年9月至2008年8月受国家建设高水平大学公派研究生项目资助在美国威斯康辛密尔沃基大学信息研究学院做联合培养博士生。2009年获得情报学专业博士学位，随后留校工作。2009年7月至2011年7月在管理科学与工程流动站做博士后。目前已发表学术论文20余篇，其中被SSCI收录的论文4篇，被EI收录的论文3篇，在国内情报学权威期刊《情报学报》上发表论文5篇，参编专著2部，在核心期刊上发表论文十余篇，主持中国博士后科学基金面上资助项目、湖北省社会科学基金项目、武汉大学自主科研项目（人文社会科学）和国家社会科学基金青年项目各一项，参加国家自然科学基金青年科学基金项目、教育部人文社会科学重点研究基地重大项目等。

目　录

导　论

0.1　问题的提出

学术期刊是科学交流的重要载体。随着科学的发展与人类知识的积累，学术期刊及其刊载论文的数量一直保持快速增长的趋势。据统计，1973 年大不列颠图书馆外借分部收藏的各国期刊数量达到 25 000 种①，内容涉及物理学、生物学、工程与数学等九个领域，比 60 年前 Hulme② 统计的数量翻了一番。英国学者 Archibald

① Carpenter, M. P., Narin, F. The subject composition of the world's scientific journals. Scientometrics, 1980, 2(1): 53-63.

② Hulme, E. W. Statistical bibliography in relation to the growth of modern civilization, London: Grafton, 1923.

和 Line① 调查了 20 世纪 50 年代至 80 年代社会科学、数学、生命科学、文献学等十个主题领域的 190 种期刊的论文数量，发现大多数主题领域的期刊论文数量呈快速增长的趋势，尤其是在 1970 年以前。虽然从 1976 年开始，期刊论文数量的增长速度有所放缓，但是在截至研究日期之前的十年期间，期刊总数仍然在上升。

期刊数量的不断增长必然导致期刊内容的交叉重复，同一学科或研究领域内可能包括许多期刊，如何从主题的角度有效地收藏、利用并管理学术期刊受到许多机构与个人的关注。人们对期刊数量的关心也逐渐转化为对期刊主题内容的注意。具体来说，期刊主题研究的意义主要体现在以下几个方面：

(1)为图书馆等收藏机构有效地选择采购学术期刊提供参考

为了有效地选择高质量的期刊进行采购与收藏，图书馆等收藏机构通常利用美国 ISI 公司发布的期刊引证报告(JCR)②，根据其中的影响因子来判断期刊质量的高低，实际上是判断期刊的被引用率。这种方法在判断同一学科领域内期刊“质量”的高低时能够发挥一定的作用，但是据此选择采购该学科领域的高“质量”期刊则可能存在一些问题。原因在于，科学的发展使得研究分支越来越细化，即使在同一学科领域内，各期刊的研究主题也可能有其侧重点，如果仅挑选影响因子较高的期刊，则可能这些期刊的主题集中在某些领域，而还有一些主题被遗漏。因此，有必要研究期刊的主题，将同一学科领域内的期刊划分为更小的主题类别，然后在不同的主题类别中选择高质量的期刊进行采购，这些既能保证被采购期刊的质量，又能够全面覆盖各个主题，从而提高期刊收藏的效率。

(2)为新进入的研究者选择研究方向提供参考

新进入的研究者通常对本领域的研究重点与发展方向不太了

① Archibald, G., Line, M. B. The size and growth of serial literature 1950-1987, In terms of the number of articles per serial. Scientometrics, 1991, 20 (1): 173-196.

② Journal Citation Reports. http://www. thomsonreuters. com/products _ services/scientific/Journal_Citation_Reports. Oct 1st, 2008.

解，尤其是在宏观层次上。广泛阅读相关期刊虽然可以提供一定的帮助，但是需要耗费大量的时间与精力。由于知识共享的各种障碍，向有经验的研究者请教也不是一件容易的事，而且这些隐性知识可能并不全面或准确。期刊主题的研究，例如期刊热点主题与发展趋势的分析有助于新进入的研究者选择感兴趣且能够发挥自身优势的主题进行研究，既符合国内外研究的前沿与热点，又能够满足研究资助机构的主题偏好。

(3)为研究者选择与其研究内容相关的期刊进行投稿提供参考

学术期刊数量的增长使得研究者在投稿时面临更多的选择，同时也更容易对选择哪种期刊进行投稿产生困惑。影响因子等揭示期刊外在特征的指标可以将某学科领域内的期刊进行“质量”排序，实际上是被引用率排序，但是不能有效地帮助研究者准确定位到与自身研究内容相关的期刊。期刊主题的研究，例如分析特定期刊的主题构成及其发展趋势有助于研究者在主题相关的期刊中选择高质量的期刊进行投稿，提高投稿的命中率。

(4)为学术期刊制定相应的发展策略提供参考

学术期刊要想在同类期刊中保持领先地位，需要了解该学科领域的研究前沿与热点，掌握其他同类期刊的主题现状及变化趋势，探索一条能够突出自身特色的发展道路。此外，随着时间的推移及学科的发展，学术期刊需要评估自身的主题变化是否符合学科发展的规律。以上目标可以通过期刊主题的研究来实现，例如根据主题对期刊进行聚类分析，使某期刊主编了解与本期刊主题相似的其他期刊有哪些；期刊热点主题分析可以帮助期刊主编相应地调整录稿政策，对热点主题有所倾斜；而期刊主题的时序分析可以帮助期刊主编了解本期刊的主题发展过程，分析自身与所在学科的研究主题发展趋势之间的吻合程度等。实际上，在期刊编辑与出版行业，越来越多的从业人员认识到主题策划在期刊发展中的重要性①与诸多

① 许丽燕. 让主题策划在期刊编辑中唱“主角”. 出版参考，2007(21)：24.

优点①，建议按照主题进行组稿，同时突出重点主题的稿件②，以便更加主动、有针对性且深入地为读者服务③。

(5)为科研政策与资助计划的制定提供参考

科研政策为科学研究向科学发展与社会实践需要的方向发展提供引导，使科学研究产生最大的学术与社会价值。科学研究需要一定的经费资助，反过来，恰当的科研资助能够促进科学研究的顺利开展，使其产生更多优秀的科研成果。科研政策与资助计划的制定需要在全面深入地掌握科学研究的现状与发展趋势的基础上进行，而期刊主题研究则可以实现这一目标。例如通过期刊热点主题分析与期刊主题发展趋势分析，科研政策与资助计划的制定者可以了解某学科领域的研究热点与未来的发展方向，并在制定相关科研政策与资助计划时向这些研究主题倾斜。

综上所述，期刊主题研究具有重要的学术意义与社会价值。然而，学术期刊通常涉及大量的主题，这种高维数据的特点使得期刊主题研究开展起来不太容易。鉴于此，本书将采用一种可视化的降维方法，即自组织映射(SOM)人工神经网络方法来研究期刊主题，使高维的期刊主题数据显示在低维的 SOM 空间中，便于研究者观察期刊主题的特点。本书将建立一套较为完整的利用 SOM 算法进行期刊主题研究的方法，并以图书情报领域的英文期刊及主题数据为例来验证这一方法的有效性，其研究结果可用于建立该领域的等级式主题目录，为期刊分类、采购与利用提供参考，使研究者及相关机构了解该领域期刊的热点主题，掌握特定期刊的主题发展趋势。此外，本书的研究思路与方法还可用于研究其他学科领域的期刊主题，为其他学科的主题组织、期刊分类、采购与利用提供参考。

---

① 沙培宁. 期刊主题化运作的“优”与“思”. 编辑之友，2005(2)：50-51.

② 杨勇，孟利宁. 关于珠算类期刊主题的思考. 新理财，1999(1)：24-26；李恩昌. 探索期刊主题策划的新路子. 报刊之友，2000(4)：36-38.

③ 李建国. 试论主题策划在现代医学科普期刊中的运用. 中国健康教育，1999，15(1)：29-30.

0.2　国内外相关研究

0.2.1　国外相关研究

国外的期刊主题研究主要有以下四个方面：

（1）期刊主题聚类分析

期刊主题聚类分析主要是对期刊论文的主题进行聚类，生成等级式的主题目录，便于用户查找相关主题及其文献，或为用户修改搜索术语提供有价值的建议。期刊主题聚类分析的数据来源既可以是某种特定的期刊，也可以是某研究领域内的期刊群。前者能够揭示特定期刊的主题结构，后者则能揭示某研究领域内的主题构成。例如，文献①基于1997—2002年Group Dynamics：Theory，Research，and Practice期刊刊载的97篇论文摘要的内容，用潜在语义分析推导成对相似性排序，然后用聚类分析来处理相似性矩阵，识别这些论文的主题，从而找出凝聚力与集团识别、集团中的权能与感觉、集团中的领导与绩效、集团成员间的权力与关系、集团中的知识与认知过程以及集团心理疗法这6个主题。

（2）基于主题的期刊聚类分析

长期以来，人们倾向于按照线性的排序模式来评价期刊，即笼统地判断期刊“质量”的高低，实际上往往是判断期刊被引用率的高低。然而，这种方法没有揭示期刊的实质内容、侧重点或目标读者。理想的研究目标是研究期刊在主题上的相似性与差异程度，按照主题对期刊进行分类与聚类。然而，由于主题类别的模糊性以及学术期刊往往属于多个学科②，有学者认为按照主题对期刊进行明

① Kivlighan Jr，D. M. and Miles，J. R. Content Themes in Group Dynamics：Theory，Research，and Practice，1997-2002. Group Dynamics，2007，11(3)：129-139.

② Leydesdorff，L. and Bensman，S. J. Classification and powerlaws：The logarithmic transformation. Journal of the American Society for Information Science and Technology，2006，57(11)：1470-1486.

确的分类似乎是不可能的①。当然，还是有一些学者在这方面进行了尝试。例如，文献②根据专家调查和 ISI 的收录范围，收集了 122 种与信息系统相关的候选期刊，通过共引分析从中筛选出 100 种核心期刊，利用 Dialog 信息系统收集了 1990—1999 年这些核心期刊的共引数据，利用 SPSS 的聚类功能将这些核心期刊分为 7 个主题聚类，包括计算机科学、计算机网络、计算机工程、信息科学、软件工程、人机交互以及管理信息系统，并发现信息科学期刊处于面向技术和集中于应用的聚类之间的桥梁位置，而 ASIST 出版物、JASIST、ARIST 和 PASIS 在信息科学聚类中占据显著位置。

由于学术期刊通常包含大量的主题，实际收集这些主题需要耗费大量的时间与精力，因此有学者采用专家调查法来收集专业人士对期刊的主观感受，并利用这些主观判断的数据来分析期刊在主题上的相似性。例如，文献③采用认知映射方法来研究健康护理管理学院的教师对 34 种北美健康护理期刊的感受。他们请北美健康管理专业的老师按照自己的标准给期刊分类，填写重要性等级量表，并收到 147 位老师的反馈。作者用多维标度和等级聚类来分析数据，给出了一个包含七个聚类的三维图，其中维度 I 将应用管理实践的期刊与健康政策的期刊进行对比，维度 II 将具体领域的期刊与广度研究的期刊进行对比，维度 III 将健康护理服务的期刊和健康护理基础设施的期刊进行对比，根据反馈者给聚类中每个期刊确定的重要性给这七个被感觉相似的期刊聚类分配权重。该研究框架揭示了学者对期刊的认知，可以作为期刊排名的辅助手段。

---

① Bensman, S. J. Bradford's law and fuzzy sets: Statistical implications for library analyses. IFLA Journal, 2001, 27: 238-246.

② Marion, L. S. Wilson, C. S. and Davis, M. Intellectual structure and subject themes in information systems research: A journal cocitation study. Proceedings of the American Society for Information Science and Technology, 2006, 42 (1).

③ Shewchuk, R. M., O'Connor, S. J., Williams, E. S. and Savage, G. T. Beyond rankings: Using cognitive mapping to understand what health care journals represent. Social Science & Medicine, 2006, 62(5): 1192-1204.

(3)期刊主题的稳定性研究

期刊主题的稳定性研究主要是探索期刊发表的论文是稳定在一定的主题范围内还是转移到其他主题。Pratt 在文献①中定义了一个相对集中测量指标，其定义如下：设所有关于某个主题的论文集合被分为 N 个子主题，使得论文以某种方式分布在这 N 个子类中。设 $a_i(i=1,2,\cdots,N)$ 表示第 i 类的所有论文，并且认为 $i \leqslant j\ (i,j=1,2,\cdots,N)$，$a_i \geqslant a_j$。设现在有一个固定的期刊也在这个主题上发文。同样对于这个期刊，设部分 $\alpha_i(i=1,2,\cdots,N)$ 为所有来自这个期刊且属于第 i 类(按照主要分类)的所有论文。现在人们想知道该期刊是按照一般倾向发文(即在完美情况下，对于每个 i 都有 $a_i=\alpha_i$)，还是该期刊偏离这个倾向，在某种相对情况下，在某些子类中发文比其他子类要多。于是 Pratt 定义了一个相对集中测量指标 C_r，如等式(0.1)所示：

$$C_r = \frac{\sum_{i=1}^{N} i(a_i - \alpha_i)}{N-1} \tag{0.1}$$

为了说明他的测量指标效果很好，Pratt 考虑了三种极端的情况，但文献②指出 Pratt 的测量指标存在两个缺点，一是它在极端情况下不是极端的，二是在某些无差别的情况下，该测量指标给出了不同的值。于是 Egghe 基于 Pratt 的原始定义，定义了一种新的相对集中测量指标 C_r'，如等式(0.2)所示：

$$C_r' = \frac{\max\limits_{\phi \in \pi_N} \sum_{i=1}^{N} \phi(i)(a_i - \alpha_i)}{N-1} \tag{0.2}$$

其中 $\pi_N = \{\phi \parallel \phi$ 是$\{1, 2, \cdots, N\}$的排列$\}$。通过计算 Transaction of the American Mathematical Society 的所有数学出版物 1975 年、1980

① Pratt, A. D. A measure of class concentration. Journal of the American Society for Information Science, 1977, 28: 285-292.

② Egghe, L. The relative concentration of a journal with respect to a subject and the use of online services in calculating it. Journal of the American Society for Information Science, 1988, 39(4): 281-284.