

· 实证研究 ·

基于多重共现揭示高校图书馆与核心期刊间的 发文关联关系研究

庞弘燊

(中国科学院国家科学图书馆成都分馆 四川成都 610041)

〔摘要〕 本论文从多重共现的分析角度出发,通过对 Morris 交叉图技术进行改进,使其适用于多重共现的分析,同时综合运用多重共现的分析方法及多重共现的交叉图技术对高校图书馆与核心期刊间的发文关联关系进行了分析。通过研究高校图书馆-核心期刊-关键词的多重共现关系发现:(1)高校图书馆与核心期刊的发文关系具有一定的地域性特点;(2)高校图书馆较为注重用户服务和资源环境建设等方面的研究;(3)各核心期刊所刊载的高校图书馆论文的研究主题各具特色;(4)各高校图书馆除了会在某些核心期刊上发表关于某几个特定主题的文章之外,在各类期刊中发文的主题可以说是百花齐放。

〔关键词〕 共现 多重共现 发文关系 关键词分析

〔分类号〕 G353.1

1 引言

文献计量学者很早就注意到论文共现(co-occurrence and occurrence)现象,通过分析共现现象可以从多个角度解释、挖掘隐含在论文中的各类信息,揭示论文与论文之间的内容关联和逻辑关联。由于共现现象可以转换为形式化的表述方式(如共现矩阵、共现关系图)加以定量测度,尤其是在计算机技术的辅助下,共现分析以其方法的简明性和分析结果的可靠性,成为支撑信息内容分析研究过程的重要手段和工具,受到了研究者的关注并进行了大量理论探讨与应用研究。而学术期刊上单独成篇发表的论文数据看似孤立,实则有着千丝万缕的关联。每一篇论文都由若干个特征项(entities)组成,包括关键词、作者、机构、发表期刊等^[1]。这些特征项结合在一起构成了一篇论文的重要特征,也是论文之间相互区别的重要特质。在文献计量研究中,通常用分析特征项之间关联的方法去探索论文的关联,进而映射科学领域在不同方面的关联结构,揭示科学活动的发展规律。

要实现海量论文数据的量化分析,就必须对文献数据进行特征提炼,抽取可以定量分析的结构化数据。揭示某种特征项内部的关联结构,是目前大部分的可视化技术所能实现的,在科学计量研究中被广泛应用,比如为了揭示两种特征项之间的关联,美国科学计量专家 Morris^[2,3]借助两个共现矩阵相同特征项之间的关联,开发了交叉图和时间线技术并进行了应用研究,两种技术可以很好地弥补目前可视化技术不能揭示两种特征项关联的缺陷。Leydesdorff^[4]把“异质网络”的思想进一步扩展到了 3-mode 网络,他把作者-期刊-关键词的特征项联系起来,通过不同类型节点在同一网络中的展现,不仅有利于分析同一类型节点间以及不同类型节点间的关系,而且也是研究网络更加真实的反映。胡琼芳和曾建勋^[5]从文献共引、耦合、共篇 3 个维度出发,提出并实现了一种综合三特征项的文献相关度判定方法。其研究相当于使用论文中的引文-被引文-关键词三个共同出现的

特征项进行匹配对比,发掘各论文之间的相关性。

基于多个不同特征项关系的可视化方法与基于两个特征项关联的科学知识图谱可视化方法(包括主成分分析、聚类分析和多维尺度分析等多元统计分析方法)相比,在反映科学活动规律和科学知识领域方面增加了多个分析角度和信息来源,因此,其中蕴含的信息量也大幅度增加,有很大的挖掘和探索价值。本研究基于论文中多个不同特征项的关联关系来探讨高校图书馆与核心期刊之间的发文关系,将有助于发掘和揭示高校图书馆在核心期刊当中的发文规律及发文主题,同时本研究着重从多重共现的角度来对其进行分析,能够提供多维度的视角,能揭示更为深入的信息内容。

2 研究方法与数据来源

2.1 研究方法

目前许多关于共现的研究都只对两个共现项进行分析,因此对多个共现项进行分析显得很有意义,因为多个共现项能够揭示出更为全面而具体的内容。比如通过构建作者-发表期刊-关键词之间关系的三维矩阵,分析作者-发表期刊-关键词的共现关系图,可以形象地揭示哪类作者在哪些期刊发表哪类主题的论文。

本研究拟通过分析机构-发表期刊-关键词的这三个特征项的共现关系,来发掘它们之间存在的关联信息。并且笔者把这种出现两个以上共现项的情况称作多重共现现象,通过分析多重共现现象可以揭示出更为深入的信息内容,比如在分析机构-发表期刊-关键词这三个共现项的关系中,如果从机构的视点出发,可以发掘出哪些机构偏向于在哪类期刊上发表某类研究主题的文章;从期刊的视角出发,可以揭示某类期刊上所具有的稳定机构作者群,还有该机构在期刊上所发表论文的主题方向;而从关键词的角度出发,可以找出发表关于某类主题论文的机构群体和期刊集合。可见多重共现现象比一般的共现现象更为复杂,并且能揭示出更为详细的信息内容。

同时本研究为了能更有效地显示出多重共现中各特征项的关联关系,笔者借鉴了 Morris 的交叉图显示方式,并对其作出改进,如图 1 所示是 Morris 的交叉图,而图 2 则是笔者对其作出改进后的交叉图显示样例。Morris 的交叉图技术分别用 x 轴、y 轴表示两种不同类型的特征项,首先利用 x 轴、y 轴轴线揭示每种特征项内部的关联结构,然后在 x 轴、y 轴的交叉点用节点(颜色、大小)显示两种特征项关联强度。可见,一张交叉图不仅包括了两个不同特征项的关联关系,还包含了这两个特征项自身各自的聚类关系。而为了能够显示出三个特征项之间的关联关系,笔者在 x 轴、y 轴的交界处加入了另一特征项来显示其与 x 轴、y 轴上两种特征项的多重共现关系(如图 2 所示),同时在坐标的外围处,笔者也加入另一特征项来分别显示其与 x 轴、y 轴特征项的共现关系。从图 1 中可以看出, Morris 的交叉图着重于显示两个特征项之间的关联关系,如图 1 中机构合作与研究主题交叉图可以考察哪些机构合作研究了哪些相关的研究主题;笔者改进后的交叉图除了可用于显示两个特征项之间的关联关系之外,还可以显示三个特征项之间的共现关系,如图 2 中机构-期刊-关键词多重共现交叉图可以考察哪些机构在哪些期刊上发表某类研究主题(关键词)的论文。多重共现的交叉图可以揭示更多的信息,挖掘出原有交叉图可视化技术无法揭示的信息。

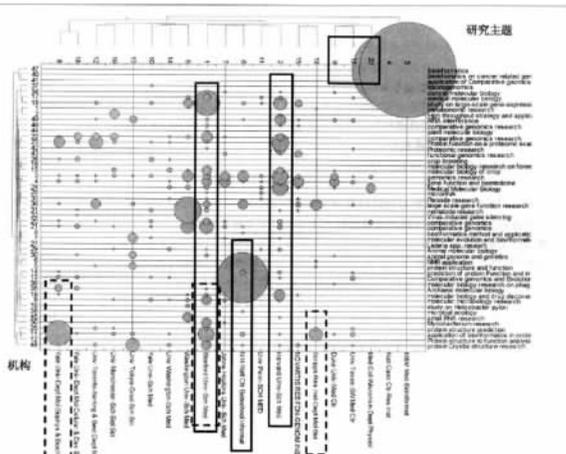


图 1 Morris 机构-研究主题交叉图⁽⁶⁾



图 2 机构-期刊-关键词多重共现交叉图

图 2 中圆圈所示代表发文量的多寡,发文越多,圆圈越大。高频关键词区域则可显示某机构在某期刊上发表论文所使用的高频关键词状况,按照其所使用的关键词频次高低,图中数字以 1-5 进行排序,并标以不同的颜色深浅和字

号大小作为区别,关键词频次越高,其标识的颜色越深并且字号也越大。

在基于共现分析方法以及多重共现可视化技术改进的基础上,本研究拟通过以下的分析模型对高校图书馆与核心期刊的发文关系进行分析。



图 3 机构-期刊-关键词多重共现分析模型

2.2 数据来源

为了能更有效地揭示高校图书馆和核心期刊的论文主题研究情况以及发掘出高校图书馆与核心期刊之间的发文关系,本研究需要限定所选取的高校图书馆和核心期刊的样本数量。因此笔者选取中国校友会网 2011 年中国大学排行榜 20 强高校⁽⁷⁾的图书馆作为机构分析样本,在期刊样本上则选择了 CSSCI(2010-2011 年)来源期刊“图书馆、情报与文献学”中的 18 种图书馆学、情报学期刊⁽⁸⁾(以下简称 18 种期刊)。然后根据所选取的图书馆样本和期刊样本在 CNKI 的中国学术期刊网络出版总库数据库搜索相关的论文,论文发表年份限定为 2006-2010 年(检索日期:2011 年 3 月 12 日),共检索出 1143 篇论文(如表 1 所示),基本可以认为检索出的论文集合代表着目前高校图书馆的主要研究方向。

表 1 高校图书馆在核心期刊上的发文情况(2006-2010 年)

机构	总发文量(A)	在 18 种期刊上的发文量(B)	比例(B/A*100%)
北京大学图书馆	147 篇	107 篇	72.79%
清华大学图书馆	179 篇	135 篇	75.42%
浙江大学图书馆	108 篇	74 篇	68.52%
复旦大学图书馆	141 篇	47 篇	33.33%
南京大学图书馆	88 篇	29 篇	32.95%
上海交通大学图书馆	83 篇	54 篇	65.06%
武汉大学图书馆	160 篇	69 篇	43.13%
中国人民大学图书馆	58 篇	29 篇	50%
华中科技大学图书馆	159 篇	25 篇	15.7%
中山大学图书馆	211 篇	122 篇	57.82%
吉林大学图书馆	145 篇	32 篇	22.07%
四川大学图书馆	143 篇	38 篇	26.57%
北京师范大学图书馆	208 篇	88 篇	42.31%
南开大学图书馆	144 篇	107 篇	74.31%
中南大学图书馆	91 篇	30 篇	32.97%
山东大学图书馆	147 篇	27 篇	18.37%
哈尔滨工业大学图书馆	107 篇	36 篇	33.64%
中国科技大学图书馆	2 篇	0 篇	0%
西安交通大学图书馆	105 篇	47 篇	44.7%
厦门大学图书馆	80 篇	47 篇	58.75%
总计	2506 篇	1143 篇	45.61%

3 基于多重共现揭示高校图书馆与核心期刊间的发文关联

关系

本研究从20所高校图书馆在18种期刊中所发表的论文当中提取关键词,并导入到自编的软件当中绘制出高校图书馆-核心期刊-关键词多重共现交叉图(如图4所示),该图可以直观地显示出高校图书馆在核心期刊中的发文关系,以下将基于多重共现交叉图来深入分析高校图书馆与核心期刊间的发文关联关系。

3.1 发文量数据分析



图4 高校图书馆-核心期刊-关键词多重共现交叉图

如图4所示,图中圆圈大的地方代表某图书馆在某期刊上发表了较多的文章,如北京大学图书馆在《大学图书馆学报》上发表了大量的文章,中山大学图书馆在《图书馆论坛》上发表了较多的文章,清华大学图书馆在《现代图书情报技术》上发表了较多的文章,南开大学图书馆在《图书馆工作与研究》上发表了较多的文章。从中可以看出高校图书馆与核心期刊的发文关系具有一定的地域性特点,即高校图书馆的研究人员会较为集中地在其所处地域(或邻近地域)的核心期刊上发表较多的文章。

同时从图4中也可看出《大学图书馆学报》、《图书情报工作》、《图书馆杂志》和《图书馆工作与研究》刊载了20所高校图书馆大量的论文,根据数据统计这四个期刊各刊载了差不多100多篇20所高校图书馆发表的论文。另外如图4和表1所示,清华大学图书馆、中山大学图书馆、北京大学图书馆、南开大学图书馆在18种期刊中发文量居于前列,都发表了超过100篇的论文,同时清华大学图书馆、南开大学图书馆、北京大学图书馆在18种期刊中发文量占总发文量的比例也是居于前列,其比例远远超过所有高校的平均比例(45.61%),这三所高校的图书馆发表了较多高质量的研究论文。在数据统计当中,笔者发现中国科技大学图书馆发文量较少,笔者猜测其发表文章的署名单位可能是只使用了中国科技大学的名称,导致在数据库中只能检索出其图书馆极少量的发文数据。

3.2 高校图书馆高频关键词分析

高校图书馆在核心期刊中发文的高频关键词,可以折射出高校图书馆的主要研究方向。20所高校图书馆在18种期刊上所发表论文的高频关键词如表2所示,从中可以看出高校图书馆非常关注服务,许多研究都是围绕用户服务展开,比如信息服务、读者服务、图书馆服务、学科服务等的一些高频词。同时高校图书馆也很注重资源环境建设等方面的研究,如关键词:数字图书馆、电子资源、网络环境、数据库、数

字资源、CALIS等就有所反映。

表2 样本论文中的高频关键词

序号	关键词	频次	序号	关键词	频次
1	高校图书馆	302	11	图书馆服务	36
2	图书馆	281	12	图书馆管理	35
3	数字图书馆	102	13	数据库	31
4	大学图书馆	56	14	数字资源	29
5	信息服务	50	15	CALIS	29
6	读者服务	50	16	科技查新	29
7	电子资源	46	17	学科服务	26
8	学科馆员	42	18	服务	26
9	网络环境	36	19	博客	25
10	图书馆员	36	20	资源整合	25

笔者拟定了一个高频关键词相似度计算公式,用以分析各高校研究论文关键词的相似度,公式如下:

$$\text{高频关键词相似度} = \frac{\text{某机构高频关键词与所有机构高频关键词的相同个数}}{\text{所有机构的高频关键词数}} \times 100\%$$

相似度越高,说明该图书馆与其它图书馆的论文研究关键词越相似,其研究主题和整体比较一致,如果相似度低,说明该图书馆的研究主题比较独特,和其它的研究方向不大一样。笔者把高频关键词数限定在20个,通过统计和对比各高校图书馆前20的高频关键词与总样本的前20高频关键词,得出各高校图书馆的高频关键词相似度(如表3所示)。

表3 各高校图书馆高频关键词相似度

序号	机构	高频关键词相似度	序号	机构	高频关键词相似度
1	南开大学图书馆	55%	11	中山大学图书馆	25%
2	清华大学图书馆	45%	12	华中科技大学图书馆	25%
3	北京大学图书馆	35%	13	上海交通大学图书馆	25%
4	西安交通大学图书馆	30%	14	复旦大学图书馆	25%
5	哈尔滨工业大学图书馆	30%	15	厦门大学图书馆	15%
6	中南大学图书馆	30%	16	山东大学图书馆	15%
7	北京师范大学图书馆	30%	17	吉林大学图书馆	15%
8	武汉大学图书馆	30%	18	中国人民大学图书馆	15%
9	浙江大学图书馆	30%	19	南京大学图书馆	10%
10	四川大学图书馆	25%	20	中国科技大学图书馆	0%

从表3中数据以及图4中最右侧的关键词区域可以看出,南开大学图书馆、清华大学图书馆、北京大学图书馆的高频词相似度最高。南开大学图书馆的高频词包括有高校图书馆、学科馆员、信息服务等;清华大学图书馆的高频词包括高校图书馆、电子资源、学科馆员、学科服务等;北京大学图书馆的高频词包括数字图书馆、数字图书馆门户、CALIS、数字环境等。这三所高校图书馆的论文研究多集中于信息服务及数字图书馆建设等领域,它们代表并引领着高校图书馆研究的主要方向。而南京大学图书馆的高频词相似度较低,代表其研究方向比较独特,和其它图书馆的主要研究方向不大一样。

3.3 核心期刊高频关键词分析

通过对核心期刊刊载高校图书馆论文的关键词进行分析,可以找出各种期刊偏重于录用高校图书馆关于哪类主题的研究论文。从图4中可以看出,18种期刊刊载了许多以高校图书馆、数字图书馆、大学图书馆为关键词的论文,其中发现某些期刊所刊载论文聚集在某几类的关键词,如《大学图书馆学报》刊载了许多以CALIS、学科导航、规范、标准为关键词的论文,《图书馆杂志》刊载了较多以学科服务、OPAC、图书馆2.0为关键词的论文,《图书馆工作与研究》刊载了较多以信息服务、采访、采购模式为关键词的论文,《图书馆建

设》刊载了较多以电子资源、集团采购、图书馆服务为关键词的论文。除此以外,有一些期刊偏重于录用关于技术类主题的论文,如《情报科学》刊载了较多以网络、搜索引擎、个性化信息服务为关键词的论文,《现代图书情报技术》刊载了较多以 DSpace、数字图书馆、OPAC、数字图书馆门户为关键词的论文,还有如《国家图书馆学刊》偏重于刊载研究图书馆实务和发展的论文,其高频关键词包括有用户服务、图书馆管理、资源建设、发展趋势、战略规划。由此可见,各核心期刊所刊载的高校图书馆论文的研究主题各具特色。

3.4 高校图书馆-核心期刊-关键词多重共现关系分析

从图4中可以看出,北京大学图书馆在《大学图书馆学报》上发表了许多以数字图书馆、CALIS为关键词的论文,西安交通大学图书馆在《大学图书馆学报》上发表了许多以CALIS、学科导航为关键词的论文。除此以外,还发现有多所大学图书馆在多种核心期刊上发表了较多以高校图书馆、CALIS、图书馆为关键词的论文,这三个关键词在图4中的机构-期刊-关键词交叉区域频繁地出现,这标志着高校图书馆的研究以图书馆类研究为主,较少涉及情报类研究,并且都是以高校图书馆这一群体的视角出发,较多地关注于CALIS的研究。另外各个高校图书馆在各种期刊上也发表多种主题的研究论文,如厦门大学图书馆在《大学图书馆学报》上发表了较多以学科导航、标准、规范、CALIS为关键词的论文,上海交通大学图书馆在《图书馆杂志》上发表了较多以学科服务为关键词的论文,北京大学图书馆在《图书与情报》上发表了较多以阅读疗法为关键词的论文,北京师范大学图书馆在《图书馆杂志》上发表了较多以高校图书馆为关键词的论文,南开大学图书馆在《图书馆工作与研究》上发表了较多以信息服务、图书馆员、美国为关键词的论文,清华大学图书馆在《图书馆杂志》上发表了较多以学科服务、学科馆员为关键词的论文。由此可见各个高校图书馆除了会在某些期刊上发表关于某几个特定主题的文章之外,在各类期刊发文的主题可以说是百花齐放。

另外笔者还发现某些高校图书馆发表的论文会采用本身机构的名称作为关键词,比如北京大学图书馆、清华大学图书馆、南开大学图书馆、中山大学图书馆等。由此可以看出这几个高校图书馆其研究多从本馆的工作出发,注重以本馆实践来做研究。

4 结语

本研究通过对Morris交叉图技术进行改进,使其适用于多重共现的分析,同时综合运用多重共现的分析方法及多重共现的交叉图技术对高校图书馆与核心期刊间的发文关联关系进行了分析。通过研究高校图书馆-核心期刊-关键词的多重共现关系发现:(1)高校图书馆与核心期刊的发文关系具有一定的地域性特点;(2)高校图书馆较为注重用户服务和资源环境建设等方面的研究;(3)各核心期刊所刊载的高校图书馆论文的研究主题各具特色。(4)各高校图书馆除了会在某些核心期刊上发表关于某几个特定主题的文章之外,在各类期刊中发文的主题可以说是百花齐放。此外,该多重共现的分析方法以及多重共现交叉图技术也可用于分析论文中其它不同类型的特征项关联关系,根据不同的研究目的来分析不同的特征项之间的共现关系。由于交叉图技术的局限性,目前只能分析三个或以下特征项的共现关系,并且对于其它不同类型特征项共现关系的分析方法,仍有待进一步研究。

(来稿时间:2011年4月)

参考文献:

1. Morris S. A. Unified mathematical treatment of complex cascaded bipartite networks: the case of collections of journal papers. Oklahoma: OklahomaStateUniversity 2005
2. Morris S. A. etc. DIVA: a visualization system for exploring document databases for technology forecasting. Computers & Industrial Engineering 2002(43): 841-862
3. Morris S. A., Gary G. Yen. Crossmaps: Visualization of overlapping relationships in collections of journal papers. (2011-03-21). www.pnas.org/cgi/doi/10.1073/pnas.030760410
4. Loet Leydesdorff. What Can Heterogeneity Add to the Scientometric Map? Steps towards algorithmic historiography. (2011-03-21). http://arxiv.org/abs/1002.0532
5. 胡琼芳,曾建勋.基于多共现的文献相关度判定研究.情报理论与实践 2010,33(8):77-80
6. 杨良斌,杨立英,乔忠华.基因组学领域的学术机构科研活动分析.图书与情报 2010(1):93-98
7. 2011中国大学排行榜揭晓,北京大学实现4连冠. (2011-03-21). http://www.cuaa.net/cur/2011/
8. 图书馆、情报与文献学(20种). (2011-03-21). http://www.cssei.com.cn/lyk2010/tq.htm

Published Papers Relationship Between Academic Libraries and Core Journals Based on Multiple Occurrence Analysis

Pang Hongshen

(The Chengdu Branch of the National Science Library, Chinese Academy of Sciences)

(Abstract) This paper improved the Morris' cross map technology based on the view of multiple occurrence analysis, and used the analysis method and cross map of multiple occurrence to analyze the published papers relationship between academic libraries and core journals. By studying the multiple occurrence relationship among academic libraries-core journals-keywords, authors find that: (1) there is a regional characteristic in the published papers between academic libraries and core journals; (2) academic libraries focused on the research of user services and construction of resources and environment; (3) there are many specific research topics about academic library in each published journal; (4) some academic libraries published papers about some specific topics in several journals, besides, the topics are various in all journals.

(作者简介) 庞弘燊(1983-),男,中科院国家科学图书馆情报学博士研究生。