

90年代的文献计量学研究与应用

邱 均 平

本世纪90年代以来,文献计量学获得了全面发展。本文仅对近几年来国内外文献计量学研究的主要进展和成果进行综述与评价;同时还集中讨论了当前文献计量学应用的特点和主要领域的应用问题。

文献计量学是文献学情报学与数学、统计学等相互交叉和结合而产生的边缘学科,也是情报信息科学体系中一个新的重要分支学科。就世界范围而言,现代文献计量学研究是从1969年为起点的。当年,英国著名情报学家A. 普里查德(A. Pritchard)正式提出了“文献计量学”(Bibliometrics)名称,标志着这一学科进入了规范化的新的发展阶段。目前,文献计量学已被公认为图书情报领域内最活跃的一个分支学科,成为情报科学研究的主流,体现了当代学科定量化的趋势。美国著名情报学家H. D. 怀特(H. D. White)和B. C. 格里菲斯(B. C. Griffith)认为,包括引文分析在内的文献计量学几乎构成了整个情报科学的一半,而且是生气勃勃的一半。它引起了国际学术界的普遍关注和高度重视。特别是90年代以来,国内外文献计量学无论是理论研究、教育普及和学术交流等方面,还是计算机辅助计量研究、信息计量工具建设和具体应用领域,都取得了较大成就,获得了全面发展。但限于篇幅,本文仅对90年代以来国内外文献计量学的研究与应用发展问题进行综述和评价,旨在弄清和把握文献计量学的发展脉络与方向,进一步推动文献计量学的健康发展。

1. 文献计量学研究的主要进展

90年代以来,文献计量学研究的范围不断拓宽,研究深度大大加强,既有理论方法的探讨,又有广泛的应用研究,而且都取得了较大进展。这主要表现在以下几个方面:

1.1 关于文献计量学诸论的研究

文献计量学诸论主要包括文献计量学的概念、研究对象、学科性质、内容结构、产生和发展以及与相关学科的关系等基本问题。近几年来,在这方面发表了一系列颇有见解的文章,清晰地勾画出了文献计量学的基本轮廓,回顾了文献计量学的发展历程,评述了各个阶段的代表性成果和关键人物,阐明了文献计量学与科学计量学、信息计量学乃至图书馆计量学的密切关系等。这对于明确学科范围和地位,统一学术认识,促进文献计量学的学科建设起了

一定作用。关于文献计量学的概念、术语、计量单元等的讨论不断深化,从探讨物理单元或交流媒介的测度原理的角度,将文献计量学视为整个情报学的理论基础和方法论工具。例如,R. Broadus 在系统评价文献计量学的不同定义之后认为,文献计量学是定量研究物理出版单元、目录单元或二者替代物的学科,并进一步解释说,文献计量学的计量分析单元就是出版物,这种计量可能是针对物理单元,如馆藏数等,也可能是针对知识单元,如题名或引文信息等。有的学者则强调,文献计量学是研究信息交流中关于测度和评价交流媒介及其可获得性的一种方法。W. C. Rorick 在应用这一概念时,还提出了“Discometrics”这一术语,反映了利用文献计量方法研究不同交流媒介和物理出版单元的不同特征。陈光祚教授则提出,电子出版物作为一种全文数据库,能使文献计量学的计量单元从一篇篇文章深入到文献中的各个知识单元,甚至单字一级。这样就能大大提高文献计量分析的深度,使文献计量学有可能发展到信息计量阶段。

1.2 关于文献增长与老化规律的研究

关于科学文献增长与老化规律的研究,虽然所发表的文献量相对较少,但也取得了一定进展。有的学者利用灰色系统理论建立了文献增长与老化的 GM 模型,预测了文献增长趋势。L. Egghe 指出,离散计算的 $f(t+1)/f(t)$ 对于连续函数并不适用,提出应用连续速度来取代,并将其应用到三篇著名的有关增长与老化模型的论文中,使其结论得到了验证。朱西传对老化研究中的同时法与历时法这两类不同的引文分析方法进行了对比研究;刘文等则探讨了文献老化规律的泊松分布问题。笔者从我国的实际统计数据出发,对科学文献老化问题展开了系统研究,提出了引文半衰期、文献半衰期的新概念,并对半衰期和普赖斯指数进行了具体计算和比较分析。有的学者还利用文献计量方法从文献生产、传播和利用的宏观角度,综合研究科学生产力对文献增长与老化的直接或间接影响等。

1.3 关于“三大定律”的研究

文献计量学的“三大定律”是它的重要理论基础,无论是在国外还是国内一直都是研究的热点。学者们围绕布拉德福定律、齐普夫定律和洛特卡定律(即“三大定律”)进行了深入探讨,并获得了一些令人注目的结论,比较突出的成果集中在经验定律及其各种导出形式的分类、归纳和统一模型的研究方面。例如,有的将经验定律及其推论分为两类,并给出了每类的统一表达式。在“三大”经验定律中,布氏定律仍然占有重要的地位,其研究也最为集中和富有成果。有的广泛探讨了布拉德福定律的原理、描述和应用等。例如,R. Rousseau 以“布拉德福曲线”为题进行了详细而深入的研究,论证了莱姆库勒统一方程式可以对布氏曲线的各种变形进行拟合,包括格鲁斯下降和尾部上翘的现象。有的则从 6 个社会科学书目所获得的数据出发,进一步探讨了布氏曲线的普遍适用性问题,发现不同类型的书目数据的布氏分布曲线的形状是不同的,而且书目中论文/期刊的密度越低,布氏曲线的凹度就越大。关于齐普夫定律,Meadow, C. T, et al 在回顾词频分布规律研究的基础上,对齐普夫公式与芒代尔布罗公式进行了比较,计算了有关参数、检验了对应的模型,并提出应用词频分布规律及其参数,可以预测关键词的数量,计算各词汇的重要程度,识别人工(命令)语言及不同人群使用这些语言的不同方式。这就使得计算机的人机交互界面能以一种适用于用户背景知识及其技能的方式来反映和服务用户等。有的采用聚类分析法对词频数据及其分布规律进行了探讨。目前,学术界对于洛特卡定律的研究主要集中在两个方面,一是洛特卡定律广义公

式(一般公式)的推导、验证和参数估计问题;二是洛特卡定律的机理及适用性的研究。作者集合的大小、统计的时间跨度、合著者和高产著者的数据处理等对洛特卡分布的影响与解释,仍然是许多学者热情关注和探讨的重要课题之一。由于一篇论文只能刊登在某一期刊上,而一篇论文却可以有多个作者,因此,作者一出版物体系与期刊一论文体系之间显然存在着结构性差异。Egghe, L 还对按文献第一作者与按部分作者计算的情况作了比较,并提出了洛氏定律的推论。值得注意的是,近几年来对文献计量学中三个新的定律(Price 的科学生产率平方根定律、Garfield 的引文集中定律、Sengupta 的扩展定律)的研究及其应用呈增长趋势,获得了一些有趣的结论。

1.4 关于集中与离散分布理论的研究

在长期的研究过程中,人们发现文献计量学的以上三个定律实质上揭示的都是文献标识特征分布的集中与离散现象,即某一具体对象(如论文)在其主体来源中(如期刊)的分布规律。在人类对知识或语言的选择过程中,在全部文献范围内对这种选择进行统计分析时,人们常常可以发现文献标识的某种数据项能引起相关数据项的集中与离散。这种文献计量数据分布的集中性与离散性贯穿于文献计量学的全部领域中,一直是学者们探讨的热点问题。对于这些不同类型的文献计量集中与离散现象及其布一齐聚一洛定律,有的研究其共同的数学基础,寻求统一的方程表达式;有的对其形成机理进行探讨,指出累积优势分布(亦称马太效应)将导致集中,而多样化、分散化则导致离散;或者用复杂的数学工具去模拟这一选择过程,导出各种分布模型。例如,有的论文讨论了逆高斯—泊松模型的真正作用及其可操作性。J. A. Stewart 提出的文献计量学、科学计量学分布的泊松—朗格曼复合模型,对与三个定律有关的经验数据的拟合程度很高;同时还对可能包含的随机过程及文献计量、科学计量的进一步推论作了研究。Basu, A 等深入探讨了有关分布模型与定律之间的关系及其在文献计量学研究中的应用问题,都是有关布氏分布理论及其统一方程研究方面的代表作。

1.5 关于以“引文分析法”为核心的方法论研究

近几年来,学术界对以引文分析法为核心的文献计量学方法论的研究日趋活跃,提出了许多基本概念和测度指标,创造了一些新的分析方法,逐渐发展成为一整套有特色的方法体系,而且具有较强的渗透性和辐射力,使得情报学在大量借鉴和移植其它学科方法的同时,也向别的学科提供具有鲜明的情报学特色的研究方法。因此,文献计量学方法论研究不仅是文献计量学内容体系的重要组成部分,而且在整个情报科学领域也是不可忽视的。有的论文提出了文献计量学方法体系的构成,论述了文献计量学作为情报学特征方法的意义和作用,并对这方面的研究作了综述和评价。关于引文分析的研究,虽然也有引证规律探讨的理论成分,但其意义主要集中在引文分析方法的完善及其应用方面。它在文献计量学中占有举足轻重的地位,是研究最为活跃、发表文章最多的一个专门领域。在汪冰、马费成等的统计中,引文分析方面的论文数分别占所统计文献量的 52.87% 和 41.57%。纵观近年的文献及成果,当前的引文分析研究基本上是沿着两个方面展开的,一是对引文分析本身的理论方法研究。探讨其概念、原理、测度指标、方法程序、引文结构、方法评价、适用范围等,并深入讨论其内在机制、引用行为、引用心理及影响因素等,使之在原有理论和方法上更加完善。相对其应用来讲,这方面的论文不是太多,但也不乏影响较大的佳作。有的研究和分析了引文的重要性,认为引文分析是评价个人、机构、期刊的学术影响的一种有效方法,提出了引证作者引

用文献的 17 种原因；还对谁引用了加菲尔德的论著、为什么会引用等问题进行了深入研究，发现加菲尔德是被引率最高的情报科学家，同时也是一位高自引者。有的就引证文献与被引文献的关系、引文网络中级序的测量、引文的年代分布和学科分布等进行了研究，取得了一些进展。二是引文分析方法的应用研究。主要集中于评定各学科专业的核心期刊、核心著者、聚类分析以及某一类文献的引文特征分析等，涉及的专业领域非常广泛。其次，引文分析还被广泛应用于科学技术评价、预测、政策优化、人才评价以及科技管理的其他方面，说明文献计量学方法已强有力地渗透于科学技术的各个领域，能够为科技管理和决策提供支持依据。由此可见，当前引文分析方法的应用范围越来越广，从自然科学到社会科学、人文科学、管理科学、工程技术等各个学科领域，只要与引文数据信息有关，就都采用引文分析直接或间接地进行研究，使引文分析成为整个科学技术领域普遍重视的方法。同时，利用的数据量越来越大，研究内容深度不断增强，较多地采用了定量化、模型化的分析方法，使之日趋规范化。随着文献计量学研究的不断深入和扩展，在引文分析研究的带动下，对文献统计分析方法、数学模型分析方法等也常有论文发表，逐步形成了文献计量学的方法论体系。

1.6 关于计算机辅助的信息计量研究

文献计量学是一门定量性、实用性很强的学科，无论是理论研究还是实际应用都必须有一定规模的数据支持。例如，美国、匈牙利等许多国家都十分重视利用计算机建立规范化的文献数据工具和原始数据的获取系统，力图实现文献计量学研究的规模化、模型化和现代化。在国外，文献计量学研究虽然也还有以手工方式进行的，但大多数的研究一般都是利用《科学引文索引》(SCI)、《社会科学引文索引》(SSCI)、《期刊引证报告》(JCR)等文献信息计量工具，或者从大型数据库或专题文档中获取数据，然后再进行计算机辅助的信息计量分析的。计算机辅助的信息计量研究是在当今新的形势下实现文献计量学突破的重要途径。在这方面，我国的差距很大，但在这种国际趋势和实际需求的推动下，我国近年来对这方面的研究也开始重视起来，并有了良好的开端。对这一新兴领域，有的学者率先进行了探讨，提出用计算机辅助计量研究来实现文献计量学研究的规模化、模型化和现代化，并讨论了计算机辅助文献计量分析的系统设计和实现途径。即 a. 建立主要用于文献计量学研究的数据库系统；b. 利用和改进现有的情报检索系统，增加有关的数据记录，扩大统计范围，使之具有定量分析功能，以适应文献信息计量研究的需要；c. 利用套录技术建立专用于文献计量研究的数据库，目前已经有了专门利用套录技术建立数据库的软件，如 SCI-NATE 就是其中之一。李晓红、任安良也以此为题撰写了硕士论文，进行了有益的探索。李胜强采用编程和原有功能相结合的方法，为 ISTIC/ISIS 增加了文献计量统计功能，可实现机助方式的著者发文量、关键词词频等的文献计量统计分析。近年来，中国科技信息研究所及其重庆分所、中国科学院文献情报中心、武汉大学、北京大学等单位，在建立中文科学引文数据库和其他可供统计的信息数据库方面都作了不少工作，并取得了一定进展，为以后的文献计量工具建设和研究手段的现代化奠定了必要的基础。

2. 关于文献计量学的应用问题

90 年代以来，国内外图书情报界在大力加强文献计量学理论研究的同时，还十分重视其应用研究与实践，出现了一些新的特点和趋势。

2.1 重视应用问题的研究和计量指标的运用

文献计量学应用问题的研究是其内容体系的重要组成部分,近年来特别受到广大研究者的重视。从所发表的文献内容来看,主要涉及两个方面:一是对文献计量学的方法、技术的研究和应用,如等级排列技术、区域分析法、图像分析法等;二是文献计量学在有关领域实际应用的探讨,主要包括在图书馆、情报与文献学等本体领域和在科学学、科技管理、科技预测甚至科学技术领域中的应用。目前,文献计量学应用的显著特点之一就是文献信息计量指标广泛应用于评价科学生产率,评价人才、成果质量、科研机构(包括大学、研究所、科技产业公司等)乃至整个国家的科技水平与影响力等,从而使决策者能够进行有效的定量化管理。例如,美国R&D的有关文献计量学数据就被列入“科学指标丛书”中,供学者及管理者评价使用。英国等欧洲国家也非常重视利用文献计量学指标去评价大学和专门研究中心的科研水平。Van Rean编辑的“科学定量研究手册”和T. Braun等人编辑的“32个国家科学计量学指标体系”是这方面研究工作非常有用的工具书。

2.2 应用的规模和数据量越来越大

在当前文献计量学的应用中,无论是利用美国《科学引文索引》(SCI)等计量工具还是通过其它数据来源渠道,研究者们都尽最大努力获取全面、充分的原始数据来支持其分析过程和结论,使利用的数据量和研究规模不断扩大。例如,匈牙利著名科学家T. Braun等人在进行科学计量学研究和比较分析时,所建立和利用的文献信息计量文档就包含了世界上96个国家和地区的114个主要科学领域中的2649种杂志及其所发表的论文。

2.3 重视现代化的文献信息计量工具的建设

文献计量学的应用是建立在大量数据的基础上的,因此必须利用计算机等现代化手段建立正规的文献信息计量工具,为应用提供大规模数据的获取渠道和来源。国外对此十分重视。早在60年代,美国就开始编制《科学引文索引》(SCI)。这一大型索引的出版和发行,为文献计量学研究提供了一种多功能的有力工具,一定程度上解决了文献计量学应用所必需的大量数据,有效地推动了文献计量学的全面发展,被誉为文献计量学史上具有划时代意义的研究成果。可以说,没有SCI就没有现代的文献计量学。在这方面,国内的差距很大,研究手段十分落后。我们应该认识到,如果不利用计算机等现代化的技术手段解决文献计量工具问题,我国的文献计量学就不可能提高到一个新的发展阶段。为此,笔者撰文强调:“开展我国文献计量工具和手段现代化的研究是我们面临的十分紧迫的任务,应当引起国内情报学界的高度重视”,并提出应该从三个方面着手进行:a. 引进和开发《科学引文索引》(SCI); b. 自编《中文科学引文索引》; c. 开展计算机辅助的文献计量分析的研究。近几年来,一些学者同仁在这些方面进行了大胆探索,取得了一些进展。例如,兰州大学图书馆靖钦恕、钱家秀研究馆员,在国内首次编制了《中文自然科学引文索引》。中国科学院文献情报中心经过多年的艰苦努力,终于建成了“中国科学引文数据库”,并于1995年开始印行《中国科学引文索引》。该索引收录了我国公开出版的315种重要期刊及其引文,内容涉及数、理、化、天、地、生、农、林、医及工程技术等领域,年报道量约3万条,包括引文索引、来源索引、机构索引和轮排主题词索引等,为我国文献计量学的研究和推广应用提供了一种现代化的基本手段和工具,具有重要的现实意义和应用价值。

2.4 重视文献计量学在科技管理和决策中的应用

近几年来,文献计量学的应用范围不断拓宽,应用研究的力度明显加大,特别是在科技

管理与决策中的应用越来越受到有关领导部门和研究者的重视。例如,美国、英国、匈牙利、印度等许多国家不仅把文献计量指标作为科学计量学指标体系中的重要组成部分,而且还从管理的高度,认为它是衡量一个国家科学文化水平乃至综合国力的一种重要途径和有效方法。我国著名学者赵红洲、蒋国华等人曾经利用文献计量方法,排出了我国主要大学发表论文的名次,并以“学术榜”的名义在报纸上公布后,引起了社会各界的强烈反响,并受到国家科委、中国科学院和国家自然科学基金会领导的高度重视。国家科委专门下达资助课题,要求中国科技信息研究所进行更大范围、更系统的文献信息统计分析,对我国科技水平在世界上所处的地位以及主要大学、科研院、所的科学生产能力和学术水平作出客观评价,并逐步形成制度,每年召开一次新闻发布会,公布有关统计结果。近几年来,国家自然科学基金会连续资助了 6 项文献计量学和科学计量学方面的研究课题,促使其研究上规模、上档次、上水平,从而有力地推动了文献计量学的深入发展。由于文献计量指标的评价功能与其它社会评价指标的功能是一致的,而且其研究成果和计量数据可以为有关部门的管理和决策提供定量依据与支持,因而越来越受到有关领导和管理部门的广泛重视。这一重要进展表明,文献计量学的某些内容和方法正在由课题研究向事业化方向发展,成为国家科技文化事业的一个组成部分。这有利于文献计量学冲破传统的局限,增强其渗透力和辐射力,大步进入“科技圈”、“管理圈”、“决策圈”,在更大的范围内充分发挥其作用,从而进一步得到社会各界的承认和重视。这是 90 年代以来文献计量学研究 and 应用发展的显著特点与趋势之一。

参考文献:

- 1 邱均平, 谢新洲译:《文献计量学发展史》(一)、(二),载《国外情报学》1992 年第 1、2 期。
- 2 王崇德, 宋力屏译:《文献计量学的演变与发展》,载《情报科学》1990 年第 1、2 期。
- 3 陈光祚:《电子出版物及其制作技术》,武汉大学出版社,1994 年版。
- 4 邱均平, 黄建东:《我国当代图书出版数学模型的研究》,载《图书情报工作》1990 年第 5 期。
- 5 李正吾:《科学文献增长和老化的灰色预测模型 (GM)》,载《情报学报》1990 年第 5 期。
- 6 Egghe L: A Theory of Continuous Rates and Applications to the Theory of Growth and Obsolescence Rates, Infor. Processing & Management, 1994, 30 (2): 279—292.
- 7 朱西传:《科技文献老化同时法与历时法的对比研究》,载《情报科学》1994 年第 4 期。
- 8 刘文, 王庆林:《科学文献老化规律 Poisson 分布质疑》,载《情报学报》1994 年第 5 期。
- 9 邱均平:《我国科技文献的时序分布与老化研究》,载《图书情报工作》1989 年第 4 期。
- 10 邱均平等译:《论文献计量学定律及其推论的分类、精确性和可能效用》,载《国外情报科学》1989 年第 4 期。
- 11 Burrell, Q. L: The Bradford Distribution and the GINI Index, Scientometrics, 1991, 21: 131—194.
- 12 孙斌等:《再谈布拉德福定律》,载《情报学报》1994 年第 2 期。
- 13 Rousseau, R: Bradford Curves, Information Processing & Management. 1994, 30 (2): 267—277.
- 14 Coleman, S. R. Bradford Distributions of Social—Science Bibliographies Varying in Definitional Homogeneity, Scientometrics, 1993, 27 (1): 75—91
- 15 Meadow, C. T, et al.: An Analysis of Zipf—Mandelbrot Language Measures and Their Application to Artificial Languages. J. of Infor. Science, 1993, 19: 247—258
- 16 王庆林、刘文:《文献中词频分布规律的研究》,载《情报学院》1994 年第 3 期。
- 17 李有梅:《用聚类分析法研究词频数据》,载《情报理论与实践》1994 年第 2 期。

- 18 Egghe, L.: Consequences of Lotka's Law in the Case of Fractional Counting of Authorship and of First Author Counts. *Mathematical and Computer Modelling*, 1993, 18 (9): 63—77.
- 19 周平译:《逆高斯—泊松模型的真实作用及其可操作性》,载《情报科学》1994年第6期。
- 20 Basu, A.: Hierarchical Distributions and Bradford's Law. *JASIS*, 1992, 43: 494—500.
- 21 邱均平:《文献计量学》,科学技术文献出版社,1988年。
- 22 秦季章:《论作为情报学特征方法的文献计量学》,载《情报科学》1992年第1期。
- 23 汪冰:《我国文献计量学十年研究与发展》,载《情报科学》1992年第1期。
- 24 马费成:《文献计量学的现状与趋势》,载《情报科学》1992年第5期。
- 25 Baird, L. M., Oppenheim, C. D.: Citations Matter? *J. of Infor. Science*, 1994, 20 (1): 2—15.
- 26 Baird, L. M.: Who Cites Eugene Garfield and Why? *MSC Dissertation*, Uni. of Strathclyde, 1993.
- 27 吴显庆译:《图书和情报科学期刊中的被引文献和引文之间的语义关系》,载《情报科学》1994年第6期。
- 28 Doreian, P. A.: Measure of Standing for Citation Networks Within a wider Environment, *infor. Processing & Management*, 1994, 30 (1): 21—31
- 29 朱西传:《关于引文年代分布数学模型的探讨》,载《情报学报》1994年第6期。
- 30 刘廷元、彭雪雪:《中外文期刊引用的年代分布和学科分布比较研究》,载《情报理论与实践》1994年第4期。
- 31 徐光宇:《中美地震学核心期刊文献近期引文特征分析比较》,载《情报科学》1994年第4期。
- 32 邱均平等:《采用三种定量方法综合评价我国物理学核心期刊》,载《中国科技期刊研究》1995年第2期。
- 33 李咏洲:《高校研究生学位论文的引文统计分析》,载《图书情报知识》1995年1期。
- 34 邱均平:《文献统计方法及其应用》,载《图书情报知识》1987年第1期。
- 35 罗式胜:《科学文献关键词键的概念——一种统计分析方法》,载《情报学报》1994年第2期。
- 36 邱均平、谢新洲:《文献计量学研究的新趋势——计算机辅助的文献计量研究》,载《情报学刊》1990年第3期。
- 37 李晓红:《计算机辅助引文分析的研究》,武汉大学硕士论文,1993年。
- 38 任安良:《计算机辅助文献计量分析数据库管理系统的设计》,西北农业大学硕士论文,1991年。
- 39 李胜强:《为ISTIC/ISIS增加文献计量统计功能》,载《情报理论与实践》1994年第2期。
- 40 Schubert, A., Braun, T.: Scientometric Datafiles Juppementary Indicators on 96 Countries 1981—1985, *Scientometrics*, 1990, 18 (3): 172—178.

(责任编辑 江 平 谭辉霞)