

共词分析与共引分析方法的比较研究

伍若梅 孔悦凡 (东北师范大学传媒科学学院 长春 130117)

摘要 文章主要从共词分析和共引分析的起源、研究对象及性质、前提假设、影响因素、应用过程及范围以及存在的问题几个方面对这两种方法进行了比较,并得出若干结论。

关键词 文献计量 共词分析 共引分析

A Comparison on Co- word Analysis and Co- citation Analysis

Wu Ruomei Kong Yuefan (School of Media Science, Northeast Normal University, Changchun, 130117)

Abstract The author compares the co- word analysis and co- citation analysis from aspects such as the origin, research objects, supposes, factors and application range, and have drawn a series of related conclusions.

Keywords bibliometrics, co- word analysis, co- citation analysis

共词分析(co- word)和共引分析(co- citation)是随着电子文献形式和网络文献形式的出现而发展起来的两种重要的文献计量学方法。近年来随着科学知识图谱逐渐成为科学计量学研究的热门领域,共词方法和共引分析方法也成为人们所关注的焦点。作为情报学研究方法的重要组成部分,二者均可广泛应用于科学评价、科技管理等诸多领域。尤其是这两种方法都可以应用于学科范式可视化研究,但有的学者利用共词分析方法来研究,有的学者利用共引分析方法来对某学科或领域的范式进行研究。那么两者之间有何异同呢?为了更好地理解并恰当的利用这两种方法,本文试图从起源、概念、方法和过程等几个方面对二者做简要的比较分析。

1 共词分析与共引分析概述

共词分析法最早是由法国文献计量学家 Callon 提出的。利用共词分析法的基本原理不仅可以揭示研究领域内的研究热点,横向和纵向分析领域学科的发展过程、特点以及领域或学科之间的关系,而且还可以反映某个专业的科学研究水平及其发展历史的动态和静态结构,其最终分析研究结果可以作为高层决策者制定战略决策和长远发展规划的基础。

共引分析法始于 Small 于 1973 年提出的以文献为单位的共引分析,是将一批文献(或著者、期刊)作为分析对象,利用聚类分析、多维标度等多元统计分析方法,借助计算机,把众多的分析对象之间错综复杂

的共引网状关系简化为数目相对较少的若干类群之间的关系并直观地表示出来的过程^[1]。之后,共引概念推广到与文献相关的各种特征对象上,从而形成各种类型的共引概念,如词的共引、文献共引、著者共引、期刊共引、主题共引和类的共引等。自 White 和 Griffith 1981 年提出著者共引分析以来,其理论发展也已经比较成熟,除了可以反映科学的知识结构外,还被用来研究科学交流模式和信息检索中知识结构的可视化。另外,期刊共引的引入并运用于期刊及学科领域的研究,主题和类的共引的引入并运用于领域分析,乃至利用共引理论来探讨科学范式等等,都将共引分析理论研究提高到新的高度。

2 共词分析与共引分析比较

2.1 起源

共词分析方法最早是在 20 世纪 70 年代中后期由法国文献计量学家提出的。1986 年,法国国家科学研究中心 CNRS(Centre National de la Recherche Scientifique) 的 Callon, Law 和 Rip 出版了 *Mapping The Dynamics Of Science And Technology*^[2]。此书的出版使共词分析方法得以广泛推广。而共引分析是在 1973 年由美国情报学家 Small 最早提出的,他在《美国情报学会杂志》(Journal Of American Society For Information Science)发表了 *Co- Citation In The Scientific Literature: A New Measure Of The Relationship Between Two Documents* 一文,提出了“共引”的概念^[2]。同年,Marshakova 在她

关于利用参考文献进行文献分类的论文中,也提出了共引的概念。

2.2 研究对象及性质

共词分析的原理主要是对一组词两两统计它们在同一篇文献中出现的次数,以此为基础对这些词进行聚类分析,从而反映出这些词之间的亲疏关系,进而分析这些词所代表的学科和主题的结构变化。它利用大量文献中共同出现的关键词对有效地反映文本关键词之间的关联强度,减少了关键词的空间,用一套结构图有效地展示了关键词之间的关联^[3]。

而共引分析的基本原理主要是以具有一定学科代表性的一批文献为分析对象,利用聚类分析、多维尺度等多元统计分析方法,借助计算机,把众多的分析对象之间错综复杂的共引网状关系简化为数目相对较少的若干类群之间的关系并直观地表示出来,在此基础上分析研究分析对象所代表的学科及文献的结构和特点^[1]。

由两种方法的基本原理可以分析出,共词分析属于内容分析方法的一种。它是先探讨词与词之间的亲疏关系,通过分析词之间的关系推断学科的研究热点或研究范式。其研究对象是词或词组。由于高频词往往反映的是热点,所以共词分析适于研究新学科的研究范式或者成熟学科的研究热点。而共引分析的理论基础是同时被引用的文献在主题上具有或多或少的相似性,因此同被引次数可以测度文献在内容方面的相关度,其属于引文分析方法的一种。它是先探讨文献与文献之间的亲疏关系,通过以往发表的高被引的文献来反映学科的结构,其研究对象是文献。由于高被引文献的形成需要较长的时间,而新兴学科的研究往往人数众多而不集中,作品比较分散,被引用情况不稳定。因此,由共引分析得出的结论往往不能反映新文献所代表的研究主题,从而其更适于研究成熟学科的研究范式,不适于学科的研究热点研究。

2.3 假设前提

可靠性是人们选择研究方法的一个很重要的指标,而可靠性的测度除了考察方法的实施过程,还可以从其实施的假设前提来分析。共词分析的假设前提最早是由 Whittaker 于 1989 年提出的,而共引分析的前提假设最早由 Small 于 1974 提出。

从表 1 中可以看出,二者的共同点在于两种方法的实施都基于一定的假设前提之上。不同点在于共词分析实施所需要的假设多于共引分析所需要的假设。共词分析的假设前提主要是作者选择的词汇能够在一定程度上反映主题内容,而共引分析的假设前提主要是同被引的文献在主题上具有一定程度的相似性。一个是基于词的假设,一个是基于文献的假设。一般来说,一种方法的实现所需要的假设越多,即所需

表 1 共词分析与共引分析假设前提比较表

比较	共词分析 ^[4]	共引分析 ^[5]
假设前提	① 作者都是很认真的选择他们的术语; ② 同一篇文章中使用不同的术语,意味着术语之间的关系是被作者认可的; ③ 如果有足够的不同作者对同一种关系认可,那么这种关系对他们关注的科学领域具有一定的意义; ④ 如果用关键词来分析,那么标引论文的关键词可以反映科学研究的现状。	① 共引可以反映文献主题内容方面的相似性; ② 对共引关系的测度可以作为揭示科学结构的一种有效方法。

要的条件越多,那么由该方法得出的结论就越不可靠,并在使用过程中越容易出现问題。另外,由于汉语词汇本身所具有的语义模糊性以及词与词之间关系的不确定性,同时鉴于共引分析的文献间所具有实在的相互引用关系比共词分析的词与词之间的关系更为确定,所以在研究同一问题时共引分析方法比共词分析方法较为可靠。

2.4 影响因素

共词分析与共引分析两种方法经过 20 多年的发展,各自都在不断的发展、完善,共词分析主要经历三个阶段,共引分析的发展经历了两个方面的改进,其发展过程见表 2。

表 2 共词分析与共引分析演进过程比较表

比较	共词分析 ^[3]	共引分析 ^[5]
演进过程	① 第一代是基于包容指数和临近指数的共词分析方法。包容指数主要用来计算主题领域的层次,临近指数主要用来计算潜在的领域。在两种指数的基础上利用包容地图和临近地图的方式把关键词或主题词的聚类网络表现出来。包容地图用于揭示领域内的中心主题,描述低频词、关键词之间的关系,更多涉及某个主题的信息。临近地图用于揭示隐藏在中心主题之中的较小领域的关系,更多涉及主题之间的关联。 ② 第二代是基于战略坐标的共词分析方法。战略坐标是在建立主题词的共词矩阵和聚类的基础上,用可视化的形式表示产生的结果。用两个维度,即向心度和密度来表示某主题目前所处的研究状态。x 轴为向心度,表示领域间相互影响的强度;y 轴为密度,表示某一领域内部联系强度。坐标的四个象限分别表示不同研究主题所处的不同状态。 ③ 第三代是基于数据库内容结构分析的共词分析方法,也是新一代共词分析方法。该方法由 Kostoff 等人于 1995 年提出,用于分析大量的数字化文本资源的系统。	① 分析对象选择中的方法改进。分析对象的选择一般是以高被引频次为依据,选择被引频次高于某数值的文献集作为分析对象。由于不同学科在引文数量上的差异,Small 提出用改进的指标分数——引文量选择文献,即用每一篇引文用引用它的来源文献的引文长度进行加权。 ② 聚类分析中的方法改进。对共引关系形成后的文献集进行聚类需要聚类方法,因为聚类方法影响着分析结果,所以方法的选择对于共引分析来说依然重要。聚类的方法从简单的等级聚类法开始,其后产生了单一联结法、多维标度分析法、k-means 方法、因子分析法及自组织图法和路径搜寻法。

从表 2 中两种方法的演进过程可以看出,共词方法与共引方法的发展影响因素不同。影响共词分析方法的因素有:一是共词分析结果的可视化需要,如包容地图和战略坐标都是结果可视化的不同表示;二是分析对象形式的改变,即从以往需要人为集中选择的文献到现在的以数据库结构形式为主的分析对象。而共引分析方法的发展两个影响因素是:一是对分析对象选择的精确性要求,即对高被引文献选择的不同要求;二是对分析对象进行聚类的准确性要求,如聚类方法研究进展。由此可以看出,共词分析更注重结果的可视化和分析对象的形式,而共引分析更注重对分析对象的精确性处理。但无论如何发展,两种方法的最终追求都是通过对分析对象的处理来准确地反映客观事实。

2.5 应用过程及范围

表3 共词分析与共引分析实施过程比较表

比较	共词分析 ^[6]	共引分析 ^[1]
实施过程	①确定分析领域,即确定拟研究的学科领域,选择学科范围。 ②确定分析单位,即确定关键词和主题词等。 ③选择分析对象,一般以词出现的频率为依据。 ④计算词共现的频率并形成矩阵。 ⑤将原始共词矩阵转化为相关矩阵。 ⑥数据的综合处理分析,即共词的聚类处理和结果的可视化处理等。 ⑦结果的分析与解释。	①确定分析领域。 ②确定分析单位,即选择文献、作者、期刊或者其他的特征对象。 ③选择分析对象,一般以被引次数为依据,作者的选定可以咨询专家。 ④搜集引文数据并生成共引矩阵,一般选择引文数据库,国外的有SCI、SSCI和AHCI,国内主要有CSCD、CSSCI和CNKI等。 ⑤将原始共引矩阵转化为相关矩阵。 ⑥数据的综合处理,多采用聚类分析、因子分析和多维尺度分析等。 ⑦结果的分析与解释。

表4 共词分析与共引分析应用范围比较表

	应用范围
共词分析	①概述研究领域研究热点;②横向和纵向分析领域学科的发展过程、特点以及领域或学科之间的关系;③反映某个专业的科学研究水平及其发展历史的动态和静态结构;④拓展信息检索领域帮助用户检索信息 ^[3] 。
共引分析	①揭示科学结构并描绘科学结构的发展历程;②用于科学研究前沿分析;③用于信息检索;④用于科研管理;⑤用于科学交流模式的展示 ^[1] 。

同为两种文献计量学方法,共词分析与共引分析的应用既有相同点又有不同点。从表3和表4可以看出,二者的相同点在于:(1)应用过程都需要确定分析领域、将原始矩阵转化为相关矩阵、数据的综合处理所用的方法等。(2)都可以用于揭示研究领域的发展过程,信息检索和反映学科结构。而不同之处在于:(1)数据的获取方式不同。共词分析数据直接从选择的文献中抽取词汇,不需要外在的数据库,而共引分析数据的获得需要一些引文数据库。(2)应用范围不同。共词方法的应用范围比共引方法小。一是因为分析对象不同。共词方法分析基础是词或词组,共引分析的对象是文献。而两个较大领域之间词共现的几率比较小,所以利用共词方法时选择高频词会很困难。再者,两个领域之间知识输入与输出的关系往往反映在文献的相互引用上,而不一定反映在词共现上。因此,共引方法可以选择多领域进行分析,而共词方法的选择领域应该尽可能小一些。另外,共引方法可以揭示微观的文献所代表的主题关系、中观的学科结构和宏观的科学结构,从而将文献信息所反映的科学的、领域的、知识的关系可视化,其不仅用于传统的文献计量学领域,更广泛应用于科学学研究,揭示科学的发展规律与趋势。而共词方法往往只可以用于某具体的学科结构或研究领域分析。共引方法还可以用于科学交流模式的分析,而共词方法多用于具体学科研究热点的分析。因此,理论上共引方法的应用范围比共词方法广。(3)使用的关键因素不同。对共词分析来说,词对出现频率是共词方法数量化处理的基础对象,对不同的词对处理方式产生不同的共词分析类型。如共词聚类分析法、共词关联分析法、共词词频分析法和突发词监测法等^[6]。而共引方法

始于以文献为单位的共引分析,之后经过共引概念的推广而形成了各种类型的共引方式,如:词的共引、文献共引、著者共引、期刊共引、主题共引和类的共引等。其中文献共引、作者共引和期刊共引是共引分析方法最常用的三种类型。从两种方法的类型来看,共词分析的类型是从处理词对的方法形成的,共引分析的类型是从共引概念在各种特征对象上的推广而形成的。从这个方面看,共词分析使用的关键在于选择处理词对的方法,而共引分析使用的关键在于选择恰当的特征对象。

2.6 存在问题

由于两种方法实施所需要的假设前提具有一定的理想性,共词分析和共引分析均不可避免的存在着一些问题。共词分析是基于词的分析,而词的选择很容易出现问题,尤其关键词的选择具有很强的主观性和随意性,由此造成同一主题内容由不同词表示或者同一词表示不同的主题内容,因此,得出的结论不能完全符合客观事实。

共引分析是基于文献之间相互引用关系进行的分析,而文献间的引用本身存在着许多问题。引文中存在着引用动机的复杂性、引文数量随机性等问题,造成引文分析结构的不准确。由此造成共引分析方法也存在分析结果的不准确性。

3 小结

作为两种重要的文献计量学方法,由于共词分析和共引分析在利用过程中存在着交叉,人们在选择利用时会出现一些问题。而研究方法对研究问题和研究结果是至关重要的,所以本文对两种方法从几个方面进行了比较,以期减少利用中出现的问题,并得出以下几个结论:(1)起源上来说,共引分析方法稍早于共词分析。(2)研究对象与性质方面,共词分析研究对象是词或词组,属于内容分析方法的一种,共引分析研究对象是文献,属于引文分析方法一种。(3)从假设前提来看,共词分析多于共引分析,由此其可靠性弱于共引分析。(4)从方法演进上来看,两种方法发展的影响因素有所不同。共词分析是基于分析结果可视化的需要与分析对象形式的改变两种因素而发展的,共引分析发展的影响因素是分析对象选择的精确性要求和分析对象进行聚类准确性要求。(5)从应用上来比较,二者在应用过程和应用范围上有相同点,也有不同点。不同点在于数据获取方式、应用领域及应用关键因素三个方面。共词分析多用于比较微观的具体的学科或领域,而共引分析既可用于微观的学科与领域,也可用于中观和宏观的科学结构的分析。共词分析多用于研究新学科、领域的研究范式或者成熟学科的研究热点,而共引分析多用于研究成熟学科的研究范式或学科结构。两种方法均存在一定程度的问题。

人文社会科学研究成果评价指标及其理论基础

任全娥 (中国社会科学院文献信息中心 北京 100732)

摘要 人文社会科学研究成果评价的理论基础是价值论, 哲学价值论决定了评价问题首先应该是评价主体的选择问题, 经济学价值论则阐释了科研成果背后的研究基础、资金投入与人力资源投入等价值源头问题。创新思想是评价的核心标准, 研究规范是评价的基础标准, 学术价值与社会价值则是评价的重要标准。评价指标构成主要包括创新程度、完备程度、价值含量与研究投入四个方面, 其中前三项为成果质量评价指标, 结合研究投入指标的综合分值合成, 才是对研究成果的最后价值评价。

关键词 人文社会科学 研究成果 评价指标 价值论

Evaluation Index of Humanities and Social Sciences Research Achievements and Its Theoretical Base

Ren Quan'e (Center for Documentation and Information, Chinese Academy of Social Sciences, Beijing, 100732)

Abstract The evaluation theory element for research achievements of humanity and social science is axiology. Philosophy axiology shows us that evaluation at first should be the choice for evaluation subject. From economics axiology we can interpret the value source, such as research basic, fund and human resources. The evaluation standard involves innovation idea, research basic, academy value and social value. The components of evaluation index includes five aspects, such as innovational degree, completeness, academic or social value and research investment.

Keywords humanity and social science, research achievements, evaluation index, evaluation theory element

1 人文社会科学研究成果评价的价值论基础

科学研究成果评价的实质是一种价值判断^[1]。评价本质上属于价值论研究的命题, 是一种认识价值的观念性活动, 人文社会科学研究成果评价也不例外。人文社会科学研究活动本身就是社会认知与社会评

价过程的统一, 而同时它又是一种价值创造过程, 因此其研究成果与自然科学一样需要研究投入与付出代价, 只是前者研究过程具有更明显的价值非中立性。

所以, 人文社会科学研究成果的价值判断, 实际上包含着双层涵义: 哲学价值论与经济学价值论。也

参考文献

- [1] 耿海英. 共引分析方法及其应用研究[D]. 北京: 中国科学院研究生院, 2007: 3, 12-13.
- [2] Callon M, Law J, Rip A. Mapping the dynamics of Science and technology: Sociology of science in the real world[M]. Macmillan, 1986: 2-16.
- [3] 冯路, 冷伏海. 共词分析方法理论进展[J]. 中国图书馆学报, 2006(2): 88-91.
- [4] Qin He. Knowledge discovery through Co-word analysis[J]. Library Trends, 48, 1999(1): 133-159.

[5] 王建芳, 冷伏海. 共引分析理论与实践进展[J]. 中国图书馆学报, 2006(1): 85-87.

[6] 钟伟金, 李佳. 共词分析法研究(一)——共词分析的过程与方式[J]. 情报杂志, 2008(5): 70-72.

[作者简介] 伍若梅, 女, 1986年生, 东北师范大学传媒科学学院情报学研究生。

孔悦凡, 女, 1984年生, 东北师范大学传媒科学学院图书馆学研究生。

收稿日期: 2009-10-05