



图书情报与信息管 理实验教材

信息计量分析工具 理论与实践

Informatics Analysis Tool: Theory and Practice

赵蓉英 编著



WUHAN UNIVERSITY PRESS
武汉大学出版社



图书情报与信息实验教材

信息计量分析工具 理论与实践

Informetrics Analysis Tool: Theory and Practice

编著 赵蓉英

参编 王 嵩 魏绪秋 全 薇 冯雪峰



WUHAN UNIVERSITY PRESS

武汉大学出版社

图书在版编目(CIP)数据

信息计量分析工具理论与实践/赵蓉英编著. —武汉: 武汉大学出版社, 2017. 11

图书情报与信息管理实验教材

ISBN 978-7-307-19665-0

I. 信… II. 赵… III. 文献计量分析—高等学校—教材
IV. G250.252

中国版本图书馆 CIP 数据核字(2017)第 220738 号

责任编辑: 韩秋婷

责任校对: 汪欣怡

版式设计: 韩闻锦

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: 湖北省京山德兴印务有限公司

开本: 720 × 1000 1/16 印张: 22.25 字数: 397 千字 插页: 1

版次: 2017 年 11 月第 1 版 2017 年 11 月第 1 次印刷

ISBN 978-7-307-19665-0 定价: 34.00 元

版权所有, 不得翻印; 凡购我社的图书, 如有质量问题, 请与当地图书销售部门联系调换。

序

定量化是科学发展的必然趋势。21 世纪以来，信息计量学作为一门采用定量方法描述和研究信息的现象、过程和规律的学科，已经形成一整套完善的理论、方法和工具体系，并在信息管理、科学管理、科学交流以及科学决策等应用领域中发挥着日益重要的作用。大数据时代，具有迭代特性的知识发现过程，是建立在原有的科学知识、日常知识和信念之上的，通过将知识库中零散的各类知识组合化、系统化，以定量化的方式进行知识融合可以产生新的知识，而对知识库中科学知识进行定量化分析，则是其中最基础也是最关键的一步。

“信息计量学”这个术语最初是由德国学者奥托·纳克(Otto Nacke)在 1979 年的《文献工作通讯》杂志上提出来的，试图用以概括数学在情报学所有领域中的应用。我国著名信息计量学家邱均平认为：信息计量学应分为“广义信息计量学”与“狭义信息计量学”，前者是探讨以信息论为基础的广义信息的计量问题，范围非常广泛；后者才是通常意义上的信息计量学，主要研究情报信息(或语义信息)的计量问题，其主要内容是应用数学、统计学等定量方法来分析和研究信息的动态特性和内在规律。

信息计量学的主要研究内容是应用数学、统计学等定量方法来分析和处理信息过程中的种种矛盾，从定量的角度分析和研究信息的动态特征，并找出其中的内在规律。因而，信息计量分析的理论与方法基础，则包含了信息计量学理论、网络分析理论和可视化分析理论等。以此为基础，我们提出了信息计量分析的方法与步骤，并构建了信息计量分析模型。

信息计量工具是信息计量学为了达成一定研究目的所应用的具体手段和方式。而信息计量分析工具的研究对象实质上是由信息计量学的研究内容决定的。1997 年，国内学者赵红州认为，科学计量学、文献计量学和情报计量学的研究内容主要集中在变量数据的经验和概率分布拟合，学科地理分布图绘

制,绩效评价、影响分析,以及其他定量四大类;Schoepflin 和 Glanzel 将信息计量学划分为四大研究领域:文献计量学理论,案例研究,实证论文(包括应用程序、指标设计和数据显示的方法学论文),以及有关文献计量学和研究政策的社会学方法。这些对信息计量学研究内容的划分和定义,为信息计量分析工具的研究对象指明了方向。

目前,信息计量分析工具经历了手工处理阶段、系统支持阶段、机助数据处理阶段、网络化专用数据处理软件阶段和可视化人工智能阶段。出现了诸如谷歌学术、百度学术、超星系列学术搜索引擎等国内外众多优秀的学术搜索分析引擎,为用户提供跨数据库平台的检索与发现带来了极大的便利。SCIE、CSSCI 和 CNKI 等定量数据收集与分析工具,则能为用户提供最科学、准确的分析数据,以了解最新的研究进展与思路。而 SPSS、Bibexcel 和 CiteSpace 等数据分析与可视化分系统工具,能够直观地传达关键的方面与特征,从而实现对于相当稀疏而又复杂的数据集的深入洞察。

本书对信息计量领域的应用工具进行了理论梳理和应用分析研究。首先对信息计量分析单元进行阐述,并介绍了信息计量分析工具的内涵和特点、类型及发展历史与现状;其次阐述了信息计量分析理论与方法,并构建了信息计量分析框架与模型;再次辅之以各种类型信息计量工具的介绍与实例演示;最后通过 6 篇经典实例分析,展示了信息计量工具的应用领域与应用方法。全书由赵蓉英、王嵩拟定提纲,并由王嵩撰写第 1、2 章,魏绪秋撰写第 3、5、6 章,冯雪峰、王嵩共同撰写第 4 章,全薇撰写第 7 章。全书由王嵩统稿并由赵蓉英修改定稿。本书在出版过程中得到了武汉大学出版社和有关院系领导的支持和帮助,责任编辑等编校人员为之亦付出了辛勤劳动。在此,我们谨向以上单位和个人表示最诚挚的谢意!

本书在写作过程中参考引用了国内外众多文献,谨在此对这些作者表示感谢。

信息计量分析工具目前正处于快速发展与变革之中,内容纷繁复杂且涉及面广,虽经我们竭力整理、总结国内外相关研究的理论与实践成果,以把研究做好、把书写好,但我们的学识和精力有限,且由多人分头执笔,书中难免还存在有些问题研究不够透彻甚至不妥之处,恳请专家和广大读者批评、指正,以便后续研究与完善。

赵蓉英

2015 年 7 月于武汉大学

目录

1 信息计量分析工具概述	1
1.1 信息计量分析单元	1
1.2 信息计量分析工具内涵	2
1.2.1 信息计量学的概念	2
1.2.2 “工具”的概念	2
1.2.3 信息计量分析工具的含义解析	3
1.2.4 信息计量分析工具的研究内容	3
1.2.5 信息计量分析工具发展的特点	4
1.3 信息计量分析的数据来源	6
1.4 信息计量分析工具类型	9
1.4.1 数理统计工具	9
1.4.2 引文索引工具	11
1.4.3 内容分析工具	17
1.4.4 网络信息计量工具	19
1.5 信息计量分析工具的发展历史与现状	20
1.5.1 手工处理阶段	20
1.5.2 系统支持阶段	21
1.5.3 机助数据处理阶段	25
1.5.4 网络化专用数据处理软件阶段	27
1.5.5 可视化人工智能阶段	27
2 信息计量分析工具理论与方法	29
2.1 文献信息统计分析法	29

2.1.1 文献计量学的基本原理	30
2.1.2 文献计量学的基本内容	30
2.2 引文分析法	31
2.2.1 引文分析基本概念	31
2.2.2 引文分析的基本步骤及工具	32
2.3 共现分析法	32
2.3.1 共词分析	33
2.3.2 共引分析	34
2.3.3 耦合分析	37
2.4 词频分析法	41
2.5 网络分析理论	44
2.5.1 网络科学理论	44
2.5.2 复杂网络理论	46
2.5.3 社会网络理论	49
2.5.4 知识网络理论	51
2.6 可视化分析理论	52
2.6.1 视觉感知和认知理论	53
2.6.2 信息可视化	56
2.6.3 可视化分析学	57
2.6.4 可视化流程	58
2.7 数据科学理论	61
2.7.1 数据科学内涵	62
2.7.2 数据科学的研究内容	64
2.7.3 数据科学分析及其工具	67
3 信息计量分析框架与模型	70
3.1 信息计量分析框架	70
3.1.1 时间维度	70
3.1.2 空间维度	72
3.1.3 主题维度	73
3.2 信息计量分析方法与步骤	75
3.2.1 信息计量分析的主要方式	75

3.2.2	信息计量分析方法	77
3.2.3	信息计量分析步骤	80
3.3	信息计量分析模型	82
3.3.1	学科结构分析模型	82
3.3.2	学科演进分析模型	99
4	文献型数据源与分析工具	114
4.1	SCIE	114
4.1.1	SCIE 简介	114
4.1.2	资源概况	114
4.1.3	操作方法	115
4.1.4	计量分析	117
4.2	Scopus	118
4.2.1	Scopus 简介	118
4.2.2	资源概况	119
4.2.3	Scopus 功能	119
4.3	Springer	120
4.3.1	Springer 简介	120
4.3.2	资源概况	120
4.3.3	操作方法	121
4.4	CNKI	124
4.4.1	CNKI 简介	124
4.4.2	资源概况	124
4.4.3	操作方法	124
4.4.4	计量分析	128
4.5	CSSCI	130
4.5.1	CSSCI 简介	130
4.5.2	资源概况	131
4.5.3	操作方法	131
4.6	CCD	133
4.6.1	CCD 简介	133
4.6.2	检索方式	134

4.6.3 计量分析	135
5 网络数据计量分析工具	137
5.1 网络日志分析	137
5.1.1 概念、意义	137
5.1.2 分类和操作步骤	138
5.1.3 主要分析工具	139
5.2 网络链接分析	145
5.2.1 概念	145
5.2.2 机理	145
5.2.3 主要分析工具	146
5.3 域名分析法	152
5.3.1 定义与原理	152
5.3.2 域名分析法的流程	153
5.3.3 网络流量分析	154
5.3.4 网络引文分析	156
5.4 Altmetrics 及其研究工具	159
5.4.1 Altmetrics	160
5.4.2 Altmetrics 工具	163
5.5 百度学术及其功能	175
5.5.1 百度学术	175
5.5.2 主要功能	176
5.6 Google Scholar 及其功能	179
5.6.1 Google Scholar	179
5.6.2 主要功能	180
5.7 Microsoft Academic Search 及其功能	183
5.7.1 Microsoft Academic Search	183
5.7.2 主要功能	184
5.8 超星系列学术搜索引擎	187
5.8.1 读秀学术搜索	188
5.8.2 百链	190
5.8.3 超星发现	194

5.9 其他学术搜索引擎	199
5.9.1 CiteSeer ^x	199
5.9.2 OAIster	201
5.9.3 Base-Search	201
5.9.4 cnpLINKer	202
6 通用与专用数据分析工具	204
6.1 通用数据分析工具	204
6.1.1 SPSS	204
6.1.2 Excel	208
6.2 专用数据分析工具	213
6.2.1 Bibexcel	213
6.2.2 Bicomb	217
7 可视化分析工具	223
7.1 CiteSpace	223
7.1.1 CiteSpace 简介	223
7.1.2 CiteSpace 的基本原理	224
7.1.3 CiteSpace 的操作步骤	225
7.1.4 CiteSpace 在知识计量中的应用	227
7.2 Ucinet	229
7.2.1 Ucinet 简介	230
7.2.2 Ucinet 的使用	231
7.3 Pajek	234
7.3.1 Pajek 简介	234
7.3.2 Pajek 主要功能	234
7.3.3 Pajek 的特点	235
7.3.4 Pajek 在信息计量中的应用	236
7.4 其他	240
7.4.1 Bibexcel	241
7.4.2 ColPalRed	241
7.4.3 IN-SPIRE	242

7.4.4	Leydesdorff	242
7.4.5	Network Workbench	242
7.4.6	Science of Science(Sci2) Tool	243
7.4.7	Gephi	243
7.4.8	VOSViewer	244
8	应用案例分析	246
8.1	Visualizing the Research on Pervasive and Ubiquitous Computing	246
8.1.1	Background	247
8.1.2	Data and Methodology	248
8.1.3	Results	249
8.1.4	Conclusion	269
8.1.5	References	270
8.2	文献计量学发展演进与研究前沿的知识图谱探析	272
8.2.1	数据来源与研究方法	272
8.2.2	文献计量学发展演进的知识图谱	273
8.2.3	文献计量学研究热点与前沿的知识图谱	278
8.2.4	结论	282
8.3	基于知识图谱的国内外企业知识共享比较研究	283
8.3.1	数据来源和研究方法	284
8.3.2	国内外有关企业知识共享的对比研究	284
8.3.3	结论	293
8.4	Exploring Scientists' Working Timetable: Do Scientists Often Work Overtime	294
8.4.1	Introduction	294
8.4.2	Data and Methods	295
8.4.3	Result	296
8.4.4	Discussions	303
8.4.5	Note	303
8.5	中国新闻传播学高被引论文分析——基于 CSSCI、CNKI 两个主流引文数据库的研究	304
8.5.1	问题的提出	304

8.5.2 数据收集与处理方法	305
8.5.3 数据分析	306
8.5.4 结语	316
参考文献.....	318



1 信息计量分析工具概述

1.1 信息计量分析单元

信息计量学可以从多个角度进行研究,信息计量工具的分析单位可以是一维、二维、三维甚至 n 维。Leo Egghe 认为,未来的信息计量学则主要聚焦三维计量学单元的研究:期刊-引文-作者、期刊-论文-引文和参考文献-论文-引文^①。Loet Leydesdorff 从科学研究的角度出发,认为信息分析单元可以建构成一个由文本-科学家-认知组成的三维立体层次模型^②。我国学者梁立明则认为,信息计量研究中的计量单元,可以分为国家、地区、机构、学科、期刊和个人^③。邱均平认为,信息计量学中的文献计量不再仅仅停留在篇、册、本等单元上,而是深入到与文献相关的内部知识单元上去,如题名、主题词、关键词、词频、知识项和引文信息等都已成为计量的对象^④。

综合可知,信息计量工具的分析单元涉及文献单元、知识单元、主题单元、学科单元和科研共同体单元,如表 1-1 所示。文献单元是出版物的篇、册、本等单元;知识单元主要表现为题名、主题词和关键词等单元;主题领域单元指的是学科内的研究领域和方向;学科单元则是以学科作为测量单元,侧重学科之间关系的动态演化研究;科研共同体单元包括无形学院的研究者和作者、机构(如公司、大学和研究院等)、地区和国家单元。

① Egghe L, Rousseau R. Introduction to Informetrics: Quantitative Methods in Library, Documentation and Information Science [M]. New York: Elsevier Science Publishers, 1990: 1.

② Leydesdorff. The Challenge of Scientometrics [M]. Leiden: The Netherlands, 1995: 1-6.

③ 梁立明. 科学计量学·指标·模型·应用 [M]. 北京: 科学出版社, 1995.

④ 邱均平. 信息计量学(一): 信息计量学的兴起和发展 [J]. 情报理论与实践, 2000 (1): 75-80.

表 1-1 信息计量工具的分析单元列表

文献单元	知识单元	主题单元	学科单元	科研共同体单元
专著、论文、 专利等的外部 特征单元	题名、主题词 和关键词等	学科内研究领 域和方向	学科知识体系 单元	作者、机构、 地区和国家 单元

1.2 信息计量分析工具内涵

1.2.1 信息计量学的概念

“信息计量学”这个术语最初是由德国学者奥托·纳克在 1979 年的《文献工作通讯》杂志上提出来的，试图用以概括数学在情报学所有领域中的应用。Leo Egghe 在 1990 年提出的信息计量学学科范式框架中认为，信息计量学是借助物理学、数学、计算机科学和其他计量学学科的工具(技术和模型)进行图书馆管理、科学社会学、科学史、科学政策和信息检索领域的研究。Leo Egghe 在 2005 年认为，信息计量学是所有与信息科学相关的计量学科的统称，包括文献计量学(书目、图书馆)、科学计量学(科学政策、引文分析、研究评价等)和网络计量学(网络、INTERNET 或其他社会网络，如引文网络、合作网络)。

信息计量学是对任何形式、任何组织的信息的数量方面的研究。我国著名信息计量学家邱均平认为：信息计量学应分为“广义信息计量学”与“狭义信息计量学”，前者是探讨以信息论为基础的广义信息的计量问题，范围非常广泛；后者才是通常意义上的信息计量学，主要研究情报信息(或语义信息)的计量问题，其主要内容是应用数学、统计学等定量方法来分析和研究信息的动态特性和内在规律。在信息计量学中，建立一套定量描述情报信息现象、过程和规律的测度指标和单位，寻找新的适合情报信息特征的计量化途径和有效方法，是非常有必要的。

1.2.2 “工具”的概念

现实中，“工具”原指生产过程中运用的器具，后引申为达到、完成或促进某一事物的手段。这里的手段包括人、物、事和为实现某种目标所借助的各

种物质和非物质形态。因而在现代汉语中,“工具”一词具有两种含义,一是指劳动的器具,是一个静态的、相对于人而言的客体性的概念;二是指达成目标的事物,既包括器具,亦包括器具的运行方式、方法,是一个客体及其运行方式合一的概念。

1.2.3 信息计量分析工具的含义解析

结合“信息计量学”和“工具”的含义,可知,信息计量工具是信息计量学为了达成一定研究目的所应用的具体手段和方式。从学科范围来看,信息计量工具有两种来源渠道,一是来自其他学科领域的定量分析工具;二是来自信息科学内与其相关的计量学学科工具。信息计量学工具是与信息科学相关的计量学科工具和其他相关学科工具的集合。从学科性质来看,信息计量工具必须具备基础的定量分析技能。信息计量学工具可以分为三个层次:一是具备基本统计分析功能的工具;二是融合统计分析功能和信息计量学学科理论于一体的工具;三是在第二种类型工具的基础上,融合其他多学科先进方法与技术的工具。

1.2.4 信息计量分析工具的研究内容

信息计量分析工具的研究对象实质上是由信息计量学的研究内容决定的。1997年,学者赵红州认为,科学计量学、文献计量学和情报计量学的研究内容主要集中在变量数据的经验和概率分布拟合,学科地理分布图绘制,绩效评价、影响分析,以及其他定量四大类。2003年,王宏鑫^①认为,信息计量学主要是对信息传播过程中的信息规律、人的认知过程中的信息规律和决策过程中的信息规律开展计量分析。国外也有不少学者对信息计量学的研究内容进行过探索。2001年, Van den Besselaar 使用因子分析方法,将科学的定量研究界定为政策导向的科学计量学、实证研究、共词研究、科学计量分布、科学计量学、专利研究和技术变革七大分支^②。Schoepflin 和 Glanzel 将信息计量学划分为四大研究领域:文献计量学理论,案例研究,实证论文,包括应用程序、指

① 王宏鑫. 信息计量学的内容与形式结构[J]. 情报科学, 2003, 21(8): 792-795.

② Besselaar V D. The Cognitive and the Social Structure of STS[J]. Scientometrics, 2001, 51(2): 441-460.

标设计和数据显示的方法学论文,以及有关文献计量学和研究政策的社会学方法^①。2005年,Glenisson等人根据信息计量学取得的最新进展,将信息计量学分为科学计量发展、实证论文/案例分析、数学模型、政治问题、社会方法以及信息计量学/网络计量学^②。2006年,Janssens等学者利用共词分析绘制出图书情报领域的研究结构,确定了五大研究点:文献计量学、专利、信息检索、网络计量学和信息科学的社会学方面^③。

1.2.5 信息计量分析工具发展的特点

(1) 集成化

交叉与综合是社会科学研究创新发展的新视野。由于不同研究方法之间具有较强的互补性和结果上的对比性,多种方法的综合与集成可以得到更加准确可靠的结果。研究方法的综合与集成,首先表现在定性方法与定量方法的结合。信息计量学研究的实质是利用数理统计的方法,通过定量分析揭示事物内在的发展规律。定量分析结果可以通过各种形式表现,如数值、图表,但研究目的不在于这些图表,而在于其中所隐含的规律。因此信息计量学的分析中也不能忽视定性分析的作用,应做到定量分析与定性分析相结合^④。目前已有学者对文献计量法与内容分析法进行了综合比较和应用研究。其次,表现为不同分析方法的结合应用。这种应用有两种不同的方向:一是方法组合,即将两种方法组合起来共同揭示某一现象;二是方法融合,即将不同关系融合成一种新的方法来解释分析单元之间的关系特征^⑤。调查发现,许多学者已经越来越多地考虑将多种不同的分析方法加以综合利用,例如共引聚类分析与词频分析的

① Schoepflin U, Glanzel W. Two Decades of "Scientometrics": An Interdisciplinary Field Represented by its Leading Journal[J]. *Scientometrics*, 2001, 50(2): 301-312.

② Glenisson P, Glanzel W, Janssens, F., De Moor, B. Combining Full Text and Bibliometric Information in Mapping Scientific Disciplines [J]. *Information Processing and Management*, 2005, 41(6): 1548-1572.

③ Janssens F, Leta J, Glanzel W, De Moor B. Towards Mapping Library and Information Science[J]. *Information Processing and Management*, 2006, 42(6): 1614-1642.

④ 杨炳儒,江亚东.基于大型数据库的KDD系统及应用研究[J].*世界科技研究发展*, 2001(1): 49-56.

⑤ 刘志辉.对信息计量学理论、方法和应用的思考[J].*图书情报研究*, 2009, 2(1): 38-41.

结合、共词分析与引文网络的结合以及同被引分析和文献耦合分析的结合等^①。

(2) 可视化

伴随着信息技术的高速发展,传统的文献计量方法和技术得到了前所未有的发展,并逐步从早期的简单计数统计过渡到目前的网络分析及可视化技术等热点领域。将可视化技术和文献计量中的引文分析技术进行有机结合,可以在客观的量化分析的基础上同时兼具全面、简洁和直观的优点,已经被越来越多的学者所研究。进入 21 世纪以来,作为引文分析与信息可视化相结合的产物——科学知识图谱,以其理论上的综合化、方法上的可视化、描绘上的形象化等诸多特征,获得迅猛的发展,一跃成为当代计量学领域的研究热点与最新前沿,研究极为活跃。

可视化工具不只是知识的一种图解方式,更需要人工智能、知识科学、计算语言学和认知语言学等学科的支持,作为连接认知心理学和人工智能研究的新桥梁,未来可视化计量工具有两个发展趋势:① 二维到多维的转变;知识可视化技术与虚拟现实空间相结合,向多维空间迈进;② 静态向动态、交互的转变,可视化工具将建立一种迭代的、协作的过程,使可视化过程动态发生^②。

(3) 智能化

计算机辅助计量分析经历了机助数据处理阶段、系统支持阶段和人工智能阶段三个阶段,如表 1-2 所示。机助数据处理阶段采用通用软件开展计量分析工作。系统支持阶段主要是通过研制专用数据库,实现计量分析工作的自动处理。人工智能阶段则是广泛利用数据挖掘、数据仓库、联机分析处理、知识发现等先进技术和手段帮助研究者进行大规模的数据抽取、深层次的信息分析以及多维、动态的分析结果展示,来实施知识层面的计量分析^③。由于信息计量从语法层次向语义和语用层次发展是一个基本的趋势,必然促进信息计量学软件开发向智能信息处理方向发展。

① 苏娜,张志强.科学计量学中多重关系融合方法研究进展及分析[J].情报科学,2010,28(9):1309-1313.

② 赵丹群.新世纪信息计量学研究进展评述[J].图书情报工作,2009(1):14-18.

③ 沙勇忠,肖仙,桃赵波.计算机辅助信息分析论略[J].情报杂志,2001(7):74-77.