

科学计量学理论与 实例

杨良斌◎著



科学技术文献出版社
SCIENTIFIC AND TECHNICAL DOCUMENTATION PRESS

科学计量学理论与 实例

杨良斌◎著



科学技术文献出版社
SCIENTIFIC AND TECHNICAL DOCUMENTATION PRESS

· 北京 ·

图书在版编目 (CIP) 数据

科学计量学理论与实例 / 杨良斌著. —北京: 科学技术文献出版社, 2014. 5

ISBN 978 - 7 - 5023 - 8785 - 3

I. ①科… II. ①杨… III. ①科学计量学 IV. ①G301

中国版本图书馆 CIP 数据核字 (2014) 第 067455 号

科学计量学理论与实例

策划编辑: 肖 朋 责任编辑: 孙洪利 责任校对: 赵 媛 责任出版: 张志平

出版者 科学技术文献出版社

地 址 北京市复兴路 15 号 邮编 100038

编 务 部 (010) 58882938, 58882987 (传真)

发 行 部 (010) 58882868, 58882874 (传真)

邮 购 部 (010) 58882873

官方网址 www.stdp.com.cn

发 行 者 科学技术文献出版社发行 全国各地新华书店经销

印 刷 者 北京天正元印务有限公司

版 次 2014 年 5 月第 1 版 2014 年 5 月第 1 次印刷

开 本 710 × 1000 1/16

字 数 229 千

印 张 16

书 号 ISBN 978 - 7 - 5023 - 8785 - 3

定 价 48.00 元



版权所有 违法必究

购买本社图书, 凡字迹不清、缺页、倒页、脱页者, 本社发行部负责调换

国关文库出版资助

国际关系学院中央高校基本科研业务费专项资金资助
(项目编号3262014T16)

序 言

科学计量学是对科学活动与管理实施量化评估、刻画和预测的科学学或情报学分支学科，用于科学学或情报学研究中定量方面的问题。在我国，虽然科学计量学研究起步较晚，但是已经涌现出了一大批堪称国际一流的计量学研究专家和学者，如赵红州、蒋国华、金碧辉、梁立明、武夷山、邱均平、山石、王崇德、罗式胜、刘则渊、朱东华等，他们在不同的课题领域，诸如布拉福德定律研究、互引分析、h 指数扩展、大学科学水平评价研究和我国科技发展定量规律研究等方面，均做出了很有意义的研究，并取得了重要成果。这些前辈学者的研究构成了我国科学计量学的主要研究成果。

文献计量学、信息计量学、情报计量学、科学计量学和网络计量学统称为“五计”，它们是分别具有特定研究对象和研究方法的学科，既相互区别又相互联系，今后还会相互依存共同发展下去。科学计量学是一个非常开放的学科，所以由“五计”等学科组成的学科群必将是一个生机勃勃的研究领域。科学计量学的基础理论和研究方法可以通过适当努力而掌握，这些理论和方法既不像自然科学那样深奥难懂，也不像大部分人文社会科学那样需要浸淫几十年才能有所得。很多其他领域的科学家也可以通过努力在科学计量学的理论与方法成果的长河中分得一杯羹。匈牙利著名科学计量学家布劳温原本从事分析化学研究，转向科学计量学研究后，获得了很多重要成果。h 指数的提出者赫希是物理学家，也没有进行科学计量学方向的研究，却因为 h 指数的提出在科学计量学史上写下了浓重的一笔，他们都是分得一杯羹的人。但是，要在科学计量学上取得成就又是很艰难的，恰恰因为其学科的开放性，进行科学计量学的研究需要掌握逻辑学、数学、

计算机科学、图形学、情报学、信息学等众多分门别类的知识,掌握这些知识需要大量的经验和时间。因此可以说,科学计量学门槛较低,但要取得一定的成果,难度很大,这是作者对科学计量学的一点不成熟的看法。

本书是对作者从2006年从事科学计量学研究以来的主要研究成果的一次小结。本书虽名为《科学计量学理论与实例》,但主要重在“实例”,只有很少的“理论”,或者可以狡辩地说,本书的“理论”是被包含在“实例”中的。本书主要包括科学计量学相关问题的思考、科学计量学与可视化的关系、世界化学史计量分析、数字图书馆的发展态势分析、物联网领域研究现状分析、网络舆情预警指标体系的构建、跨学科理论与跨科学测度的实证分析、学术机构科研活动分析、g指数理论与证明、纳米科技领域的多学科性和跨学科性、专利和论文可视化的比较、国际情报学领域研究进展等内容,其中第九、十和十一章内容主体是译自国外科学计量学期刊上的重要论文,目的是让读者了解和掌握国外学者的有关科学计量学研究方法。除第六、第七章关系较为紧密外,其余各章均是各自成篇,读者可以打乱篇章顺序抽取部分章节阅读。因作者学术能力和学术成果有限,不能为读者贡献足够多的科学计量学知识和成果,作者深感愧疚和遗憾,只能期待今后的学术科研生涯中,能有更多更好的科研成果回报读者。

特别感谢国际关系学院“国关文库出版资助”提供的出版资金,特别感谢中国科学院文献情报中心杨立英研究员,对本书提出的一系列宝贵意见与建议;特别感谢国际关系学院科研处和信息科技系领导及同事,对本书的出版提供了大量的帮助和支持。

此外,感谢选修《情报学基础》课程的国际关系学院信息科技系本科生的帮助和支持,感谢国际关系学院中央高校基本科研业务费专项资金资助,感谢科学技术文献出版社对本书的出版。

限于作者水平和时间有限,错误疏漏不足之处在所难免,恳请读者批评指正。

杨良斌

2013年10月24日于国际关系学院

目 录

CONTENTS

第一章 绪论 1

1. 1 引言 1

1. 1. 1 科学计量学的概念 1

1. 1. 2 科学计量学的基本理论与研究内容 2

1. 1. 3 科学计量学的研究工具和方法 2

1. 2 关于科学计量学相关问题的思考 3

1. 2. 1 科学计量学研究方法 4

1. 2. 2 普赖斯科技文献增长模型 5

1. 2. 3 科学计量指标与同行评议 6

1. 3 结语 9

第二章 科学计量学与可视化 11

2. 1 信息可视化与知识可视化 12

2. 2 科学计量学与可视化的关系 14

2. 2. 1 可视化与科学计量学的相互促进 14

2.2.2	可视化的局限性	15
2.3	可视化在科学计量学中的实例	16
2.4	可视化在金融工程领域的应用	17
2.4.1	数据与研究方法	18
2.4.2	结果分析	20
2.5	结语	27
第三章 科学计量学方法的简单应用		30
3.1	世界化学史分析	30
3.1.1	20 世纪化学研究中心的转移	31
3.1.2	20 世纪化学研究内容的变迁	34
3.1.3	小结	36
3.2	数字图书馆的发展态势分析	37
3.2.1	材料与方法	38
3.2.2	结果与分析	41
3.3	我国物联网领域研究现状分析	45
3.3.1	研究方法和数据来源	46
3.3.2	统计与分析	47
3.3.3	重点论文作者和期刊	49
3.3.4	高被引论文	51
3.3.5	热点关键词分析	53
3.4	基金资助和研究层次分布	54
3.5	结语	56
第四章 网络舆情预警系统指标体系的构建分析		60
4.1	网络舆情生命周期	61
4.2	指标体系的构建	63
4.2.1	指标框架的确定	63

4.2.2 指标说明与赋值	65
4.2.3 指标权重的确定	67
4.3 结语	68
第五章 跨学科测度的研究综述	70
5.1 引言	70
5.2 跨学科的概念、性质和发展状况	71
5.3 跨学科的学科间信息转移理论及转移模式	73
5.4 跨学科研究方法:共类分析和共词分析	75
5.5 跨学科的学科交叉度	76
5.5.1 布里渊指标	76
5.5.2 基于引文的学科交叉度指标	77
5.5.3 科学家个人的学科交叉度指标	78
5.6 结语	78
第六章 跨学科理论及其测度指标的构建分析	84
6.1 跨学科研究的驱动力	84
6.1.1 跨学科研究的内在驱动力	84
6.1.2 跨学科研究的加速器	86
6.2 倡导跨学科研究的原因	88
6.2.1 跨学科研究的障碍	88
6.2.2 跨学科研究的重要性	89
6.2.3 跨学科研究的现实性	90
6.3 基于文献计量的跨学科学	92
6.3.1 期刊论文与学科交叉的关系	92
6.3.2 跨学科学研究的基本问题	94
6.4 多学科度测度分析	94
6.4.1 信息熵测度公式	95

6.4.2	统计熵测度公式	95
6.4.3	信息熵与统计熵测度公式的比较	96
6.5	专业度测度分析	97
6.5.1	文章与作者专业度公式	98
6.5.2	专业度公式的数学论证	98
6.6	学科交叉度测度分析	99
6.6.1	交叉度测度公式	100
6.6.2	交叉度公式的理论说明	101
6.6.3	共信息度测度公式	102
6.7	合作度测度分析	103
6.8	多学科度与学科交叉度的关系	103
6.9	结语	105
第七章 基于跨学科测度指标的领域发展态势分析		111
7.1	研究领域的选择与数据准备	111
7.1.1	学科分类研究	112
7.1.2	数据准备及研究流程	112
7.2	结果分析	113
7.3	数据集构建及指标计算	119
7.3.1	数据集构建方法与过程	119
7.3.2	跨学科测度指标的计算	120
7.4	结果分析	122
7.4.1	两个领域的概况	122
7.4.2	交叉学科领域分析——以超敏反应领域为例	123
7.4.3	跨学科领域分析——以燃料电池领域为例	128
7.4.4	研究领域学科交叉状态的判定	132
7.5	结语	133

第八章 学术机构科研活动分析 136

- 8.1 引言 136
- 8.2 机构合作模式分析 137
 - 8.2.1 基本思路 137
 - 8.2.2 数据与方法 138
 - 8.2.3 结果与解释 140
- 8.3 优势机构研究主题的揭示 143
 - 8.3.1 交叉图技术概述 144
 - 8.3.2 数据处理 144
 - 8.3.3 结果与解释 145
- 8.4 结语 149

第九章 h 指数的改进—g 指数 152

- 9.1 介绍 153
- 9.2 g 指数的数学理论 156
- 9.3 普赖斯奖章获得者 g 指数和 h 指数的计算和比较 163
- 9.4 结论与开放性问题 166

第十章 多学科、跨学科和纳米科技的合作研究模式 169

- 10.1 简介 170
- 10.2 “纳米标题论文”的增长 171
- 10.3 具体方法论 175
 - 10.3.1 纳米的研究范围:“纳米期刊论文”和“纳米标题论文” 175
 - 10.3.2 衡量跨学科的方法 179
 - 10.3.3 合著者分析 182
 - 10.3.4 跨学科方式和指标 184
 - 10.3.5 建立在分子图形上的可视化跨学科研究 185

10.4	数据分析与输出	188
10.4.1	跨学科对比多学科	188
10.4.2	多学科的模式	190
10.4.3	机构间合作和跨洲际合作	195
10.5	结语	203
 第十一章 太赫兹技术的专利和论文的可视化比较研究……		206
11.1	引言	206
11.2	方法论	209
11.3	结果分析	210
11.4	结语	229
 第十二章 国际情报学领域研究进展分析		230
12.1	引言	230
12.2	数据准备及基本统计结果	231
12.2.1	数据源	231
12.3	文献共被引和作者共被引图谱及其分析	235
12.3.1	文献共被引图谱的生成及分析	235
12.3.2	作者共被引图谱的生成及分析	237
12.4	关键词共现图谱和突现词的时区分布图分析	238
12.4.1	关键词共现图谱的生成及分析	238
12.4.2	突现词的时区分布图及分析	240
12.5	结语	243

第一章

绪 论

1.1 引 言

1.1.1 科学计量学的概念

科学计量学 (Scientometrics) 是应用数理统计和计算技术等数学方法对科学活动的投入 (如科研人员、研究经费)、产出 (如论文数量、被引数量) 和过程 (如信息传播、交流网络的形成) 进行定量分析, 从中找出科学活动规律性的一门分支学科。

科学计量学与文献计量学、信息计量学和网络计量学有一定的交叠。由于科学活动的产出和交流的主要形式之一是科学文献, 因此对这类文献进行的定量研究既是科学计量学研究, 又是文献计量学研究。同理, 用定量方法处理科学信息的产生、流行、传播和利用, 则既属科学计量学研究, 也属信息计量学研究。但科学计量学也有独特的研究领域, 如对科学创造最佳年龄结构的研究, 出重大科技成果时科学家年龄的频度分布规律的研究, 等等。

科学计量学的研究结论较为客观, 它有助于加深对科学发展内在规律的认识, 从而为科研管理工作和科技政策制定提供参考和指导。

1.1.2 科学计量学的基本理论与研究内容

科学计量学的基本理论包括：描述文献增长定律的普拉斯指数；描述文献老化规律的半衰期和普拉斯老化指数；描述论文作者分布规律的洛特卡定律；描述科技文献离散定律的布拉德福定律；描述词频分布规律的齐普夫定律；描述文献引用规律的齐曼定律；描述重大科技成果的威布尔分布模型以及引文分析、科学学科发展规律、科学中心转移规律和科学发现的当采模型等。这些都是经典的科学计量学的基本理论，是构建科学计量学理论架构的基石。

科学计量学根据已掌握的各种数据，从不同的侧面建立各种科学计量学指标模型，揭示科学发展的种种规律。科学计量学试图通过定量方法寻找科学活动的内在规律或准规律，并为更有效率地开展科学活动提供指导。典型的科学计量学问题和内容如下^[1]：

- 1) 科学研究的生产率问题；
- 2) 科研资金投入的最优化；
- 3) 通过科学计量学方法和指标预测学科发展趋势和确定资助重点；
- 4) 通过科学计量学方法和指标识别科学的不同学科之间，以及科学活动同技术活动之间的联系，从而为跨学科研究和理性的科技政策制定提供指导；
- 5) 通过科技产出指标进行科研绩效评估；
- 6) 描述科学活动规律和准规律的各种数学模型，如“成功导致成功”的数学模型、洛特卡定律、布拉德福定律、齐普夫 - 帕雷托分布，等等；
- 7) 用科学计量学方法和指标研究科技人才和科技教育问题。

1.1.3 科学计量学的研究工具和方法

科学计量学的研究工具主要有：SAS；Matlab；Excel；SPSS；Pajek；CiteSpace 等数据处理软件和可视化软件，对这些软件的理解掌握是科学

计量学研究的基本要求。

科学计量学的研究方法有：指标法；引文分析法；模型法；数理论证法等，这些方法是科学计量学方法论体系的重要组成部分。研究中主要数据来源有：SCI；SSCI；A&HCL；ESI；EI；CPCI；CSTPCD；CSCI；CSSCI等数据库，这些数据库是科学计量学的数据来源和研究基础。

1.2 关于科学计量学相关问题的思考

科学计量学是一门以科学自身为研究对象，进行定量研究的学科。这门学科借助科学计量学指标，运用数学方法计量科学研究的成果、描述科学体系的结构、分析科学系统的内在运行机制，揭示科学发展的时空特征，也探讨在整个社会大背景之下科学活动的定量规律性。自 20 世纪 60 年代初创立至今，科学计量学已被广泛应用于科研主体实力考察、学术期刊质量评估、科技发展规划制订以及科学基金项目管理等许多方面^[2]。近年来，人们认为，科学计量学是由基于引文分析的领域可视化研究领域派生出来的新学科，采用文献计量学手段分析科学文献，对科技交流进行数量级的研究^[3]。通过特殊的方法，科学计量学已经广泛地应用于评价科学进步与科学家行为的研究中。经过数十年的发展，科学计量学以大规模的学术数据库为基础，利用文献、专利、基金等信息作为重要的工具描绘大量学科领域的发展和当今的科学技术，已经取得了大量的成就^[4]。

科学计量学的发展，使科研绩效的评估向着更加科学化的进程迈进。在这方面，集中度定律、洛特卡定律、布拉德福定律、齐普夫定律、普赖斯定律、引文定律、文献老化定律、文献增长与冗余定律等文献计量学理论，为科学计量学在科研绩效评估中的应用提供了一个广阔的前景^[5]。

针对科学计量学的各种研究过程中，由于对科学计量学理论和方法上认识的不足而在应用过程中出现的各种问题，笔者基于此对科学计量学本身提出一些自己的看法。

1.2.1 科学计量学研究方法

1) 研究内容的思考

科学计量学包括三个部分：①国家科技政策分析，这部分需要结合文献计量、科技企业成长、GDP 等不同的社会经济指标，以宏观的角度对科技发展政策提出建议；②项目、机构、团队、学者评价，这部分需要与期刊影响因子、引文率、各类指数的结合分析，与同行评议比较，同时必须考虑专家意见；③对测量所用的指数本身的研究，如 H、A、G、RA 等指数的研究，分析指数的优缺点后，针对缺点进行原有指数的改良。无论是研究哪个部分，都不能不参考其他部分的研究结果。笔者认为，科学计量学建立在文献计量学和统计学的基础之上，讨论的内容和设计问题的方向虽有不同，但是运用的方法和思路有共通之处，就分析工具而言，科学计量也有专属的软件，与一般的统计软件大同小异但是更为侧重利用已有的数据库资料和指数作分析。可以认为，科学计量学的研究内容是较为成熟的，已经成为一个独立的学科体系。

2) 研究步骤的思考

目前图书馆学情报学的研究主要集中在三个子领域：用户、计量、系统。它们需要具有社会科学背景的人才去进行用户研究，具有统计知识背景的人才去进行书目计量、科学计量的研究，具有计算机编程背景的人才去进行数字图书馆技术的了解与尝试性开发，这三者都必须建立在实证的基础上，可以被其他研究者重复，或者进行修改；与用户研究相同：数据收集、数据整理、数据分析、形成结论四个步骤是必要的研究过程，然而科学计量学的研究必须还要加上第五个步骤“专家意见”才算合格。从科学计量学的实证研究可以看出，透过案例了解科学计量学，透过抽象归纳知道科学的量化研究的步骤与局限，找到指导自己研究工作原则的同时，也认识部分情报学的内容。

3) 关于普赖斯批判的思考

对普赖斯的《大科学小科学》内容的阅读和相关参考文献的讨论中

发现,由于他(对哲学观点的简单二分法下)是以英国经验主义为出发点^[6],沉浸在物理学杂志的书海中,归纳历史发展的规律,并且以数学形式进行表达,获得阶段性的突破与成果;必然遭到苏联科学家们以德国观念主义(如自然辩证法)的另外一种角度,对其不足之处产生分析与批判,这样的一种结果,又会遭到目前科学家的批判,而我们阅读的时候,又会再予以批判;这样持批判性眼光的阅读与思考过程最好是减少门户之见,而增加意见交流,最好的方式是阅读伴随着怀疑和猜测,然后激发自己想去实证的念头和动作。

4) 预测与科研的思考

科学研究的发展是渐进的还是突变的,从科学计量的角度来说,我们只能透过总结历史轨迹,来预测未来渐进发展的趋势,对于以几个天才的灵感造成科学发展突变的现象,是无法透过计算来预言的。而恰恰是基于这样两种观点和做法,科学计量学给我们的启示是,进行科学研究需要建立在前人基础上,发现或者挖掘新的现象和问题,世界上终归会有人把没有被注意到的事实给挖掘出来,要想保持科学领先,就必须时时关注别人的工作,让自己不处于模仿和重复的工作状态;至于是否能够有造成领域大变动的创造,则是在上述的基础上,可遇而不可求的事情。

1.2.2 普赖斯科技文献增长模型

普赖斯关于科技文献的增长模型,本质上是利用数学和统计学的方法对所获得的各类文献统计数据进行处理,最后得到描述文献增长的数学公式,即所谓的经验公式(empirical formula)。总的来说,普赖斯主要通过这种方法总结出了指数增长和逻辑斯蒂曲线增长、直线增长等模型。普赖斯当时是以世界范围的科技文献出版物为统计对象的,如《英国皇家学会哲学会刊》。普赖斯在得出科技文献的逻辑斯蒂增长模型后,认为科技文献的增长分为起始、发展和饱和三个阶段。

对于当今科技发展,已与普赖斯当时所处的世界显然有很大的不同。当今各国的科技发展是极为不平衡的,这必然反映到科技文献的不同上