

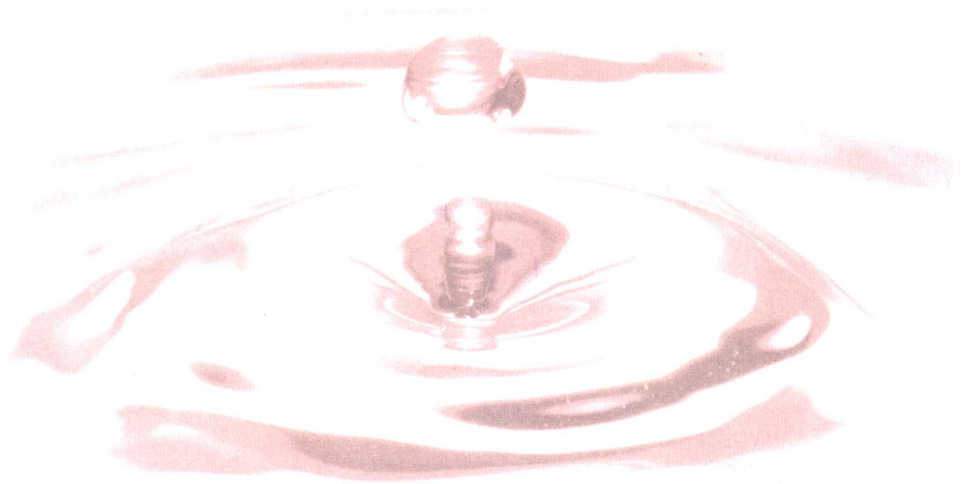
精·品·课·程·立·体·化·教·材·系·列



普通高等教育“十一五”国家级规划教材

# 情报学基础教程

叶 鹰 潘有能 潘 卫 编著



 科学出版社  
[www.sciencep.com](http://www.sciencep.com)

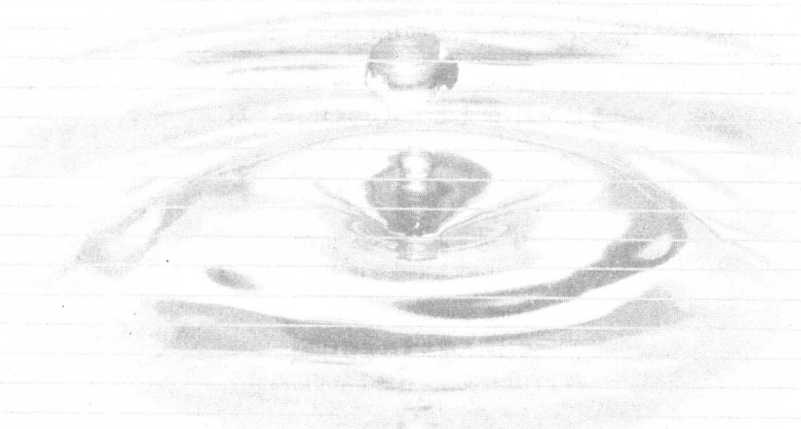
精·品·课·程·立·体·化·教·材·系·列



普通高等教育“十一五”国家级规划教材

# 情报学基础教程

叶 鹰 潘有能 潘 卫 编著



科学出版社

北 京

## 内 容 简 介

本书是普通高等教育“十一五”国家级规划教材。作为学习情报学的大学生、研究生以及科技情报、社科情报、图书情报工作人员编写的通识教材,本书的内容覆盖情报学理论与方法、情报管理与服务各相关领域,简明扼要地阐述了情报源、情报管理、情报检索、情报分析、情报技术等情报学核心分支的基本知识和新近发展。

本书配备多媒体教学课件等教学支持,适合作为大学生和研究生的教材及专业培训教材,也可供同等学力人员申请硕士学位学习参考。

### 图书在版编目(CIP)数据

---

情报学基础教程 / 叶鹰,潘有能,潘卫编著. —北京:科学出版社,2006  
普通高等教育“十一五”国家级规划教材

ISBN 7-03-018043-7

I. 情… II. ①叶…②潘…③潘… III. 情报学-高等学校-教材  
IV. G350

中国版本图书馆 CIP 数据核字(2006)第 108945 号

---

责任编辑:陈 亮 / 责任校对:邹慧卿

责任印制:张克忠 / 封面设计:陈 敬

科 学 出 版 社 出 版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

信浩彩色印装有限公司印刷

科学出版社发行 各地新华书店经销

\*

2006 年 10 月第 一 版 开本:B5(720×1000)

2006 年 10 月第一次印刷 印张:19 1/2

印数:1—4 000 字数:377 000

定价:28.00 元

(如有印装质量问题,我社负责调换(环伟))

# 前言

情报学作为一个独立的研究领域，是 20 世纪 50 年代开始确立的：50 年代初法语世界首先使用 L'informatique，1959 年英语世界正式提出 Information Science，从此，情报学逐步发展起来，至今仍处于动态发展之中。

如果在信息概念基础上认识情报，则情报是具有传递价值的信息；如果在知识概念基础上认识情报，则情报是具有传递价值的知识。因此，具有传递价值是情报的根本属性。也就是说，情报是针对特定对象的需要而提供的、在科学研究和各种活动中起继承、借鉴或参考作用的有价值的信息或知识。

情报学就是研究情报的产生、传递和利用规律的一门学科，现代情报学尤其关注利用现代化信息技术与手段，使情报交流过程、情报系统保持最佳效能状态，并帮助人们充分利用信息技术和手段提高情报产生、加工、贮存、检索、交流、利用的效率。

当然，国内外情报界对情报和情报学的认识多种多样，尚无统一标准，认为情报学“是研究情报传递工作的理论、方法和原理的学科”，“是研究如何开发、利用信息资源的规律及其技术、方法的学科”或“是研究人类如何创造、利用和交流各种形式的情报的学科”等观点均有代表性。但总的来看，情报学可以说是以情报工作的理论、原则和技术、方法为研究对象的学科。

作为一门近半个世纪才发展起来的新兴学科，情报学的主要研究任务包括：①揭示情报和情报工作的本质与特性；②探索情报工作的产生和发展规律；③探讨情报工作的内容环节及其在情报工作中的地位、作用，它们之间的内在联系以及各个环节的原则和方法；④研究各个时期情报工作的方针、政策、标准化的内容，以及现代化理论和技术；⑤阐明情报工作的科学管理理论与方法；等等。因此，情报学研究具有交叉学科或跨学科特点，是一门综合性的横断学科。中国情

报界也有人把“情报学”和“情报科学”看作两个不同的概念，认为情报学是研究情报工作的原理、工作方式与技术方法，情报的组织、管理、服务、评价及情报事业发展历史的学科。而情报科学则是综合研究由语言学、文字学、语文学、符号学、目录学、文献学、档案学、图书馆学、情报学等多学科知识的科学。

在中国，科技情报工作虽然早在 20 世纪 50 年代就与世界同时起步，但情报学作为一个独立的专业研究领域，却是 1978 年后随着科技情报专业教育的发展才得以确立的。因此，情报专业教育可以说是情报学的引擎。

为持续推动情报专业教育发展，浙江大学和上海交通大学的有关专业教师在吸收前辈老师和同辈学者的众多研究成果基础上，合作编写了这本《情报学基础教程》。

为避免繁杂的概念讨论，本书以三个简明扼要的科学化概念为基础：把信息作为物理实在并定义为信息密度的积分，将情报定义为有价值的信息（有效信息），再将知识定义为有效信息的积分。

在采用新的概念系统的同时，本书在体系组织方面的构造思路如下：

学习情报学，首先应当建立科学思维和量化观念，故将情报学经验规律作为第 1 章内容，这是文献计量学的一个子集，主要陈述布拉德福定律、洛特卡定律、齐夫定律、文献增长律、文献老化律的基本内容，并简要介绍其机理研究。

作为一门学科，应当有自己独立的基础理论体系，故第 2 章在扼要介绍国内外代表性情报学基础理论基础上，尝试用“一个轴心、三条原则、一组方程”构造情报学的基本理论轮廓和体系框架，使学生或读者能对情报学的总体结构有一个简明的整体把握。

情报学研究需要采用各种方法，除成熟的科学通用方法如逻辑分析法、数学分析法等外，也有本学科特有的引文分析法等独特方法，这构成本书第 3 章的核心内容。

情报源是情报工作和情报学研究的物质基础，故作为本书第 4 章。

围绕分类编目、主题标引、文摘索引等情报工作形成的情报管理不仅是情报工作的中心内容，而且是学习和研究情报学必不可少的基础，故在本书第 5 章进行系统陈述。

情报检索不仅是情报学的核心内容之一，而且是当代大学生、研究生、科研人员应当掌握的基本技能，故本书第 6 章对其理论基础和主要操作技能进行介绍。

情报分析也是情报学的核心内容之一，是最能体现情报学“研究”价值的分支，具有开拓潜力和发展前景，值得重视，这就形成本书第 7 章的实体内容。

伴随情报学进化，情报技术也日新月异地发展，尤其是在计算机技术、网络技术、通讯技术等组合推进下，情报技术正发生天翻地覆的变化，本书第 8 章

概要介绍现代主流情报技术。

此外，情报系统与情报服务、情报学与其他学科交叉形成的研究领域等内容也是情报学的重要组成部分，这就构成本书最后两章。

本书具有一定逻辑结构层次的编排希望能为初学情报学的学生或读者提供一个比较明晰的情报学概貌，适合用作研究生、大学生、专升本等类型学生的“情报学基础”或“情报学概论”类课程教材，特别适合具有其他学科背景的学生尤其是具有理、工、农、医专业背景的学生学习情报学使用，也可供图书情报工作人员参考。全书以业界公认的知识内容为主体，也适当增加了独立探索研究成果。本书特色是带有较强的科学实证思维方式，这也正是浙江大学和上海交通大学的学术传统，希望能有一定的启发价值。

为方便教师教学，本书配备了多媒体教学课件等立体化教学支持，凡选用本书作为教材的教师可与出版社取得联系，以获得相关支持。

# 目录

## 前言

### 第1章

情报学经验规律 .....	1
1.1 布拉德福定律 .....	1
1.2 洛特卡定律 .....	7
1.3 齐夫定律 .....	10
1.4 文献增长律 .....	12
1.5 文献老化律 .....	14
1.6 情报学经验规律的统一模型和机理研究 .....	15
习题和思考题 .....	17
主要参考文献 .....	17

### 第2章

情报学基础理论 .....	18
2.1 情报学概念系统 .....	19
2.2 国外代表性情报学基础理论 .....	29
2.3 国内代表性情报学基础理论 .....	34
2.4 情报学基础理论新探 .....	36

2.5 情报学理论原则 .....	39
2.6 情报学相关学科 .....	43
习题和思考题 .....	44
主要参考文献 .....	45

### 第3章

<b>情报学方法论</b> .....	46
3.1 情报学方法论概述 .....	46
3.2 通用科学方法 .....	48
3.3 情报学专门方法 .....	70
习题和思考题 .....	84
主要参考文献 .....	84

### 第4章

<b>情报源</b> .....	86
4.1 情报源导引 .....	86
4.2 文献情报源及其数字化资源 .....	87
4.3 非文献情报源 .....	103
4.4 情报源评价 .....	109
习题和思考题 .....	111
主要参考文献 .....	112

### 第5章

<b>情报管理</b> .....	113
5.1 分类编目 .....	113
5.2 主题标引 .....	123
5.3 文摘索引 .....	131
5.4 元数据 .....	135
5.5 知识组织 .....	143
习题和思考题 .....	147



主要参考文献 .....	147
--------------	-----

## 第 6 章

---

<b>情报检索</b> .....	148
6.1 情报检索理论基础和检索评价参数 .....	148
6.2 全文数据库的发展及其检索利用 .....	151
6.3 重要手工检索工具及其对应数据库 .....	154
6.4 重要联机检索系统 DIALOG 及其操作 .....	161
6.5 网络信息检索技术和主要搜索引擎 .....	173
习题和思考题 .....	183
主要参考文献 .....	184

## 第 7 章

---

<b>情报分析</b> .....	185
7.1 情报分析导引 .....	185
7.2 典型社科情报分析——竞争情报分析 .....	198
7.3 典型科技情报分析——专利情报分析 .....	213
习题和思考题 .....	221
主要参考文献 .....	221

## 第 8 章

---

<b>情报技术</b> .....	223
8.1 计算机技术基础 .....	223
8.2 网络技术基础 .....	228
8.3 数据库技术 .....	235
8.4 数据仓库技术 .....	242
8.5 数据挖掘技术 .....	247
8.6 其他情报技术 .....	252
习题和思考题 .....	257
主要参考文献 .....	257

## 第9章

---

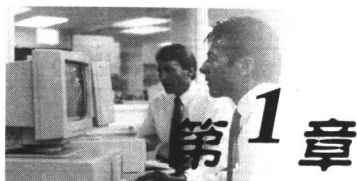
情报系统与情报服务 .....	258
9.1 从 MIS 到 ERP .....	258
9.2 从 DSS 到 CIS .....	266
9.3 从 OAS 到电子政务系统 .....	271
习题和思考题 .....	279
主要参考文献 .....	279

## 第10章

---

情报学扩展 .....	280
10.1 比较情报学 .....	280
10.2 信息管理学 .....	283
10.3 信息经济学 .....	285
10.4 信息法学 .....	289
10.5 信息伦理学 .....	291
10.6 情报学发展方向 .....	296
习题和思考题 .....	298
主要参考文献 .....	299

后记 .....	300
----------	-----



# 情报学经验规律

经验规律是学科赖以建立的基石，也是量化研究的起点。情报学虽然是一门“年轻”的学科，但在经验规律方面已有一定积累，其基础主要由下列文献计量规律（Egghe and Rousseau, 1990）奠定。

## 1.1 布拉德福定律

布拉德福定律是描述专业论文在期刊中分布情况的经验规律，由英国著名文献学家布拉德福（S. C. Bradford, 1878—1948）于1934年提出。

### 1.1.1 基础数据和规律表述

布拉德福毕业于英国伦敦大学，所学专业是化学，1922年获科学博士学位，但他热爱图书馆工作和图书馆事业，1925~1937年曾长期担任英国南肯辛顿科学图书馆馆长。在工作中，他发现一个学科的专业论文常分散发表在各种期刊杂志上，于是他以应用地球物理学和润滑专业为实例，研究了期刊论文分布的情况，提出了一个描述文献分散的经验定律。布拉德福采集的1929~1932年应用地球物理学原始数据见表1-1。

表 1-1 布拉德福原始数据

期刊数	相关论文数	累积期刊数	累积论文数	期刊数	相关论文数	累积期刊数	累积论文数
1	93	1	93	1	56	3	235
1	86	2	176	1	48	4	283

续表

期刊数	相关论文数	累积期刊数	累积论文数	期刊数	相关论文数	累积期刊数	累积论文数
1	46	5	329	5	10	27	662
1	35	6	354	3	9	30	689
1	28	7	392	8	8	38	753
1	20	8	412	7	7	45	802
1	17	9	429	11	6	56	868
4	16	13	493	12	5	68	928
1	15	14	508	17	4	85	996
5	14	19	578	23	3	108	1065
1	12	20	590	49	2	157	1163
2	11	22	612	169	1	326	1332

数据来源: Bradford S C. Documentation. Washington, D. C. : Public Affairs Press, 1950; 转引自 Garfield E. Bradford's law and related statistical patterns. Current Contents, 1980 (19): 5-12

如果将以上数据分成如下 3 个区域:

分区	期刊数	相关论文数	每刊年载相关论文数
C	9 ( $n_c$ )	429	>4
1	59 ( $n_1$ )	499	1~4
2	258 ( $n_2$ )	404	1

则近似有:

$$n_c : n_1 : n_2 = 1 : a : a^2, \quad a \approx 5$$

显然这只是一条近似规律。

用文字表述就是: 如果将期刊按其刊载某专业论文数量的多寡以递减顺序排列, 则可分出一个核心区 and 相继的几个领域, 当每区刊载的论文量相等时, 核心期刊数  $n_c$  和外围一区期刊数  $n_1$ 、外围二区期刊数  $n_2$  成  $n_c : n_1 : n_2 = 1 : a : a^2$  关系。其中,  $a$  称为布拉德福常数。

若以累积期刊数  $n$  的对数为横坐标、累积相关论文数  $R(n)$  为纵坐标, 可作出如图 1-1 所示的曲线。

因此, 后来布鲁克斯 (B. C. Brookes) 用公式将布拉德福定律表述为:

$$R(n) = \begin{cases} an^\beta, & (1 \leq n \leq n_c) \\ k \lg(n/s), & (n_c \leq n \leq N) \end{cases} \quad (1-1)$$

其中,  $R(n)$  为相关论文累积数;  $n$  是期刊累积数;  $a$  是  $n=1$  时的  $R(n)$ ;  $\beta, k, s$

为参数； $N$  为期刊总数。曲线右上端的弯折称为格魯斯 (Q. V. Groos) 下降。

### 1.1.2 改进和扩展

布拉德福 1934 年提出上述经验规律后，并未引起人们注意。1948 年，其专著《文献学》(Documentation) 出版，在第 9 章中全文收录了 1934 年的发现，才引起另一位英国文献学家维克

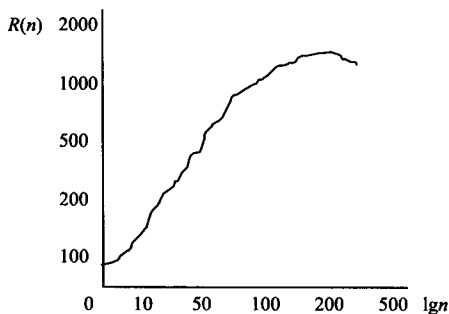


图 1-1 布拉德福分布曲线示意

利 (B. C. Vickery) 的重视。维克利对之进行研究后，将外围区扩展到多个区域，并首次用“布拉德福分散定律”命名这一经验规律 (Vickery, 1987)：

把期刊按照刊载某学科相关论文的数量排列，可得到专门登载该学科论文的核心区期刊和若干外围区期刊，当各区所含该学科相关论文相等时，各区期刊数成

$$n_c : n_1 : n_2 : \dots = 1 : a : a^2 : \dots \quad (a > 1) \quad (1-2)$$

其中， $a$  是布拉德福常数。

还有一个改进称为维克利推论：

$$n_c : (n_c + n_1) : (n_c + n_1 + n_2) : \dots = 1 : b : b^2 : \dots \quad (b > 1) \quad (1-3)$$

其中， $b$  叫做维克利系数。

1977 年，苏联情报学家斯莫里科夫 (И. А. Смольков) 对布鲁克斯的公式描述进行研究后，提出用一个统一的方程来表述布拉德福定律：

$$R(n) = K \lg(q_1 n + q_2 e^{-\beta n}) \quad (1-4)$$

其中， $R(n)$  为相关论文累积数； $n$  是期刊累积数； $K$ ， $\beta$ ， $q_1$ ， $q_2$  为参数。

### 1.1.3 应用提示

布拉德福定律的主要用途是用于确定核心期刊，以指导期刊订购和期刊利用，并由此扩展到核心馆藏维护、核心检索工具选择等。有人认为布拉德福定律可以看作是社会科学中普遍存在的“二八律”的一种表现：20% 的核心期刊上刊载了 80% 的重要论文。此外，布拉德福定律还可应用于核心出版社的研究等。

上海交通大学王国龙等通过对 SCI 和 SSCI 引用期刊的研究，发现影响因子 (IF, impact factor) 的分布也符合布拉德福定律。他们的数据如表 1.2~表 1.5 所示<sup>①</sup>。

<sup>①</sup> 王国龙，李佩. SCI 源期刊与 JCR. 上海交通大学学报，2003 (增刊)：287-290；王国龙. SSCI 源期刊概述. 中国索引，2004，2 (2)：20-25

SCI 源期刊数据见表 1-2。

表 1-2 SCI 源期刊数据 (2001 年)

序号	国家/地区	期刊种数	影响因子 (IF)	期刊所占比例 (%)	序号	国家名	期刊种数	影响因子 (IF)	期刊所占比例 (%)
1	美国	2220	1.83	38.60	30	比利时	13	0.43	0.23
2	英格兰	1139	1.54	19.80	31	斯洛伐克	12	0.38	0.21
3	荷兰	565	0.94	9.82	32	芬兰	11	0.75	0.19
4	德国	432	0.90	7.51	33	以色列	10	0.63	0.17
5	瑞士	154	1.03	2.68	34	克罗地亚	10	0.18	0.17
6	法国	148	0.59	2.57	35	墨西哥	7	0.37	0.12
7	日本	147	0.57	2.56	36	乌克兰	7	0.22	0.12
8	俄罗斯	99	0.31	1.72	37	智利	6	0.47	0.11
9	加拿大	76	1.32	1.32	38	罗马尼亚	6	0.23	0.11
10	意大利	69	0.76	1.20	39	阿根廷	6	0.04	0.11
11	丹麦	60	1.24	1.04	40	威尔士	5	0.42	0.09
12	澳大利亚	59	0.65	1.03	41	希腊	4	1.48	0.07
13	中国	57	0.37	0.99	42	委内瑞拉	4	0.15	0.07
14	印度	45	0.22	0.78	43	南斯拉夫	3	0.27	0.05
15	波兰	37	0.52	0.64	44	伊朗	3	0.18	0.05
16	挪威	34	0.97	0.59	45	土耳其	3	0.12	0.05
17	新加坡	29	0.67	0.50	46	沙特	3	0.11	0.05
18	苏格兰	26	0.87	0.45	47	孟加拉国	3	0.10	0.05
19	西班牙	26	0.54	0.45	48	斯洛文尼亚	3	0.04	0.05
20	奥地利	25	0.94	0.44	49	埃塞俄比亚	2	0.14	0.03
21	捷克	22	0.41	0.38	50	肯尼亚	2	0.13	0.03
22	新西兰	21	1.31	0.37	51	保加利亚	2	0.12	0.03
23	南非	19	0.39	0.33	52	巴基斯坦	2	0.08	0.03
24	瑞典	18	1.04	0.31	53	泰国	1	0.38	0.02
25	韩国	18	0.59	0.31	54	拉脱维亚	1	0.37	0.02
26	匈牙利	18	0.29	0.31	55	乌兹别克斯坦	1	0.32	0.02
27	爱尔兰	16	1.03	0.28	56	爱沙尼亚	1	0.26	0.02
28	中国台湾	16	0.54	0.28	57	牙买加	1	0.26	0.02
29	巴西	15	0.35	0.26					

续表

序号	国家/地区	期刊种数	影响因子 (IF)	期刊所占比例 (%)	序号	国家名	期刊种数	影响因子 (IF)	期刊所占比例 (%)
58	乌克兰	1	0.20	0.02	62	特立尼达和多巴哥	1	0.08	0.02
59	哥斯达黎加	1	0.10	0.02					
60	北爱尔兰	1	0.09	0.02	63	厄瓜多尔	1	0.04	0.02
61	科威特	1	0.08	0.02	64	古巴	1	0.02	0.02
	其他	3		0.05	总计		5752	1.44	100

根据表 1-2 数据归纳整理出的结论见表 1-3。

表 1-3 SCI 源期刊 IF 的布拉德福分布

布拉德福分区	核心区	准核心区	外围区
期刊 (种)	328	1173	4251
IF 的范围	4.14~46.23	1.56~4.13	<1.56
总影响因子 (TIF)	2760.3	2758	2759.7

SSCI 源期刊数据见表 1-4。

表 1-4 SSCI 源期刊数据 (2002 年)

序号	国家/地区	期刊种数	影响因子 (IF)		期刊所占比例 (%)
			总数	平均	
1	美国	1003	968	0.97	58.69
2	英格兰	386	291.6	0.76	22.59
3	荷兰	99	75.4	0.76	5.79
4	德国	54	28.5	0.53	3.16
5	加拿大	29	13.98	0.48	1.70
6	法国	18	5.6	0.31	1.05
7	瑞士	17	13.8	0.81	0.99
7	澳大利亚	17	9.55	0.56	0.99
9	日本	9	2.30	0.26	0.53
10	俄罗斯	7	1.62	0.23	0.41
11	挪威	6	2.97	0.50	0.35
12	苏格兰	6	2.61	0.44	0.35
13	捷克	4	0.74	0.18	0.23

续表

序号	国家/地区	期刊种数	影响因子 (IF)		期刊所占比例 (%)
			总数	平均	
13	新西兰	4	0.93	0.23	0.23
13	爱尔兰	4	2.40	0.60	0.23
13	墨西哥	4	1.18	0.30	0.23
17	瑞典	3	0.82	0.27	0.18
17	中国	3	0.05	0.02	0.18
17	南非	3	1.08	0.36	0.18
17	奥地利	3	1.23	0.41	0.18
17	印度	3	1.40	0.37	0.18
22	克罗地亚	2	0.40	0.20	0.12
22	比利时	2	1.11	0.56	0.12
22	丹麦	2	2.32	1.16	0.12
22	韩国	2	0.27	0.14	0.12
22	巴西	2	0.50	0.25	0.12
22	西班牙	2	1.25	0.63	0.12
22	斯洛伐克	2	0.24	0.12	0.12
29	中国台湾	1	0.29	0.29	0.06
29	斯洛文尼亚	1	0.30	0.30	0.06
29	新加坡	1	0.04	0.04	0.06
29	芬兰	1	0.08	0.08	0.06
29	以色列	1	0.30	0.30	0.06
29	意大利	1	0.78	0.78	0.06
29	土耳其	1	0.28	0.28	0.06
29	哥伦比亚	1	1.10	1.10	0.06
29	智利	1	0.47	0.47	0.06
29	葡萄牙	1	0.29	0.29	0.06
29	阿根廷	1	0.10	0.10	0.06
29	南斯拉夫	1	0.14	0.14	0.06
	其他	1			0.06
总计		1709	1436.1	0.84	100

根据表 1-4 数据归纳整理出的结论见表 1-5。



表 1-5 SSCI 源期刊 IF 的布拉德福分布

布拉德福分区	核心区	准核心区	外围区
期刊(种)	197	436	1076
IF 的范围	1.676~11.622	0.789~1.670	<0.789
总影响因子(TIF)	478.8	478.6	478.7

这一研究将布拉德福定律的应用范围扩展到影响因子,具有参考意义。

## 1.2 洛特卡定律

洛特卡定律是描述作者与论文数量之间关系的经验规律,由美国情报学家洛特卡(A. J. Lotka, 1880—1949)于1926年提出。

### 1.2.1 基础数据和规律表述

洛特卡的原始研究是对物理学和化学两大学科中科学家们发表论文的情况进行统计分析。对于物理学,他使用德国奥尔巴赫(Aürbach)《物理学史一览表》(Geschichtstafeln der Physik, 1910)的人名索引中,对其全部1325位作者进行统计分析,依次列出发表1篇、2篇和多篇论文的作者数;对于化学,他采用《化学文摘》(Chemical Abstracts, 1907~1917)十年累积索引中的姓氏以A、B开头的6891位作者进行了统计,分别列出发表1篇、2篇、3篇直至346篇论文的作者数。结果发现物理学1325位作者中,发表1篇论文的为784人,占59.2%;化学A字母姓氏作者1543人,发表1篇论文的为890人,占57.7%;B字母姓氏作者5348人,发表1篇论文的为3101人,占57.98%。于是他提出作者的百分比分布应符合如下公式:

$$f(x) = \frac{c}{x^a} \quad (1-5)$$

其中, $f(x)$ 是发表 $x$ 篇论文的作者占作者总数的百分比(作者频率),常数 $a > 1$ 。上式取对数,有

$$\log f(x) = -a \log x + \log c \quad (1-6)$$

故以 $\log f(x)$ 为纵坐标、 $\log x$ 为横坐标作图应是一斜率为 $-a$ 、截距为 $\log c$ 的直线。

洛特卡的原始数据中,Aürbach数据部分见表1-6(计算百分比是根据洛特卡平方反比律计算所得)。