

改革开放 30 年来我国情报学研究的 回顾与展望(二)

——情报学研究论文的作者分析

邱均平 杨思洛 王明芝

(武汉大学中国科学评价研究中心 武汉 430072)

摘要 对改革开放 30 年来我国情报学研究论文的作者进行分析。首先以五年为段对各年段作者数量、人均发文量、合作度和合作率进行介绍。然后对各年段的核心作者进行统计,包括总年度发文最多的作者、各年段新出现和退出的作者,并对各年段进行洛特卡定律公式的拟合。最后,对核心作者进行共被引分析,包括分层聚类和多维尺度分析,并用 Pajek 进行可视化。

关键词 情报学 洛特卡定律 作者共被引分析 核心作者

分类号 G250

Review and Reflection of the Information Science Research in China in the Past 30 Years (II): An Analysis of the Authors of the Articles in Information Science

Qiu Junping, Yang Siluo, Wang Mingzhi

Research Center for Chinese Science Evaluation, Wuhan University, Wuhan 430072, China

Abstract The authors of the articles in information science in China in the past 30 years are studied in this paper. First, the number of the authors, the average of the articles per author, and the cooperation degree and rate are analyzed for every year on a five-year basis. And then the core authors in each period, including authors who are on the top list in the number of published papers, the new authors and those who have stopped published papers in this field, are analyzed. At the same time, the fitting of Lotka formula is made for each period. Finally, Author Co-citation Analysis of the core authors is conducted and visualized with Pajek.

Key words information science; Lotka law; ACA; core author

研究者是推动学科发展的力量。探讨作者的著述规律及其数量关系,可以发现和预测研究论文数量的增长和内容的分布及发展动向;从文献计量的角度出发也可预测科学家数量的增长和科学发展的规模及趋势等。

1 各年段作者数量统计

30 年来情报学论文总作者数为 46985 人, 92259 人次。其中无作者论文 215 篇,署名为“编辑部”的 105 篇及署名为“课题组”的 25 篇,对于这些论文以独著处理,没有考虑作者同名异人的情况,因为这些特殊论文占全部论文的比例很少,可认为不会对结果造成影响。具体各年段数见表 1。对于作者总体发文情况可

以作者平均发文量来表示,人均发文量 = 论文总数/作者数。各年段人均发文量在 1 篇左右,后面的年段人均发文量比较少,主要原因是合作人数在增加,另外每年段出现了大量新人。因为著者发文的累加性,所有年的人均发文量最大,达到 1.29。

科研是一项集体活动,一篇论文有适量的合作者,可以充分发挥群体智慧,可以在知识结构等方面相互取长补短,提高研究成果水平。对作者合作情报进行分析主要有两种指标:①合作度 = (一定时期内相关文献)作者总人次/(一定时期内相关文献)论文总数;②合作率 = (一定时期内相关文献)合作论文数/(一定时

期内相关文献)论文总数^[1]。从表1来看合作度是逐年段上升,到2003-2007年段平均每篇论文有1.65个作者。论文的合作率更是增长显著,在1978-1982年段超过83%的论文是独著的,但到2003-2007年段单作者论文仅占不到30%。总体上有超过一半的论文有两个或两个以上作者。随着科学不断向深度和广度发展,各学科相互交叉、相互渗透,科研难度加大,许多研究项目必须相互协作,联合攻关,依靠多学科集体的力量和智慧才能完成研究课题;同时计算机网络技术的发展也方便了人们的合作与交流,因而合著的论文越来越多,相信合作度和合作率在未来还会有所提高。

表1 情报学论文作者分布

年段	1978-1982	1983-1987	1988-1992	1993-1997	1998-2002	2003-2007	总年度
作者数	986	3190	5680	9170	15572	23779	46985
论文总数	1202	3907	5991	9264	15337	25054	60755
人均发文量	1.22	1.22	1.05	1.01	0.98	1.05	1.29
作者总人次	1443	4683	8011	13179	23710	41233	92259
独著论文数	1029	3313	4594	6563	9784	14678	39961
合作度	1.2	1.2	1.34	1.42	1.55	1.65	1.52
合作率(%)	16.81	17.93	30.41	41.16	56.76	70.69	52.04

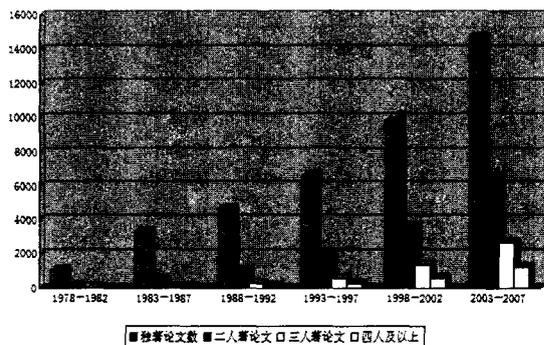
对每篇论文作者数量进行统计,可以更详细地分析作者的合作团体数量。根据表2,每年段中独著论文占的比重最大。各年段二人、三人、四人及以上著论文数量依次减少,但各年段的比重随着时间的发展有所增大。

表3 各年段高频作者分布

1978-1982		1983-1987		1988-1992		1993-1997		1998-2002		2003-2007	
作者	频次										
王洵	10	樊松林	21	樊松林	22	杨则正	23	邱均平	50	邱均平	76
吴建中	9	袁锁鸿	20	王崇德	21	张琪玉	21	张晓林	39	王知津	66
赵国琦	9	王秀成	19	周毅	19	樊松林	18	夏旭	28	杨华	53
洪传科	8	王崇德	17	胡昌平	18	秦惠基	18	马海群	27	范文田	50
莫作钦	8	刘东维	16	罗式胜	16	汪冰	18	倪波	23	侯汉清	38
王松益	8	黄祥喜	13	关金琦	15	夏旭	18	王知津	23	彭靖里	38
严怡民	8	邓璐华	12	刘植惠	14	谢新洲	18	方平	22	张玉峰	32
杨沛霆	8	陈光祚	11	吴国恩	14	柯平	17	盛小平	20	周宁	30
李毅	7	龚国伟	11	严怡民	14	毕强	16	张福炎	20	樊治平	29
王崇德	7	贺克毅	11	刘建明	13	罗式胜	16	张琪玉	19	胡沛泉	29
包昌火	6	胡昌平	11	邹荫生	13	霍国庆	15	程刚	18	梁战平	28
高兴民	6	匡华	11	陈文勇	12	谢阳群	14	王伟	18	秦铁辉	28
惠世荣	6	沈家模	11	陈光祚	11	张帆	14	焦玉英	17	苏新宁	27
孟广均	6	徐网大	11	谢先江	11	周智佑	14	范文田	16	张敏	27
吴尔中	6	秦铁辉	10	谢新洲	11	曹剑奇	13	胡昌平	16	韩毅	26
辛希孟	6	邱均平	10	徐身新	11	何荣利	13	刘磊	16	马海群	26
庄义逊	6	毕强	9	包昌火	10	张晓林	13	徐建华	16	王平	26

表2 每篇论文作者数量

年段	1978-1982	1983-1987	1988-1992	1993-1997	1998-2002	2003-2007	总年度
独著论文数	1029	3313	4594	6563	9785	14678	39962
二人著论文	132	482	1006	1850	3617	6480	13567
三人著论文	27	71	277	611	1360	2649	4995
四人及以上	14	41	114	240	575	1247	2231



2 核心作者分布统计

作者与论文数量分布具有不均衡性,这可通过作者在一定时间撰写论文数量来衡量。表3列出了各年段发表论文最多的前二十位作者,其中省略了匿名作者。在30年中既有持续的高产作者,也有阶段性的多产作者;有许多老一辈专家的退出,也有一批批新人的出现。各年段的作者都是情报界的精英,特别是总年度的前105位作者更为我们所耳熟能详,为我国情报学研究作出了重要贡献(见表4)。

陈光祚	5	迟忍	9	陈爱芬	10	霍忠文	12	王永成	15	文庭孝	26
陈界	5	关金琦	9	戴维民	10	罗曼	12	邹志仁	14	胡昌平	25
陈军	5	黄立军	9	何佳讯	10	马海群	12	查先进	13	王惠翔	25

根据我们的研究,非期刊高产作者成为高h指数作者的概率极低,即高发文量是作者影响力的基础,故发文量在很大程度上说明作者

在学科中的地位,这从高产作者与高h指数作者的吻合上也说明这一问题^[2]。

表4 总年度发文量前105位作者

作者	频次	作者	频次	作者	频次	作者	频次	作者	频次	作者	频次	作者	频次
邱均平	155	彭靖里	52	盛小平	43	谢阳群	36	刘东维	32	刘春茂	29	甘利人	27
王知津	109	方平	50	张帆	42	靖继鹏	35	文庭孝	32	刘伟	29	关金琦	27
胡昌平	81	周宁	50	梁战平	41	倪波	35	樊治平	31	秦珂	29	贺克毅	27
樊松林	74	罗式胜	49	赖茂生	40	韩毅	33	黄敏	31	王兰成	29	胡德华	27
张晓林	74	马费成	46	周智佑	40	李莉	33	霍国庆	31	张颖	29	刘学和	27
张琪玉	71	王军	46	王永成	39	李敏	33	李勇	31	李娜	28	王勇	27
范文田	70	王伟	46	黄晓斌	38	王宏鑫	33	刘莉	31	庞景安	28	杨则正	27
马海群	69	包昌火	45	刘植惠	38	王平	33	汪冰	31	王辉	28	周九常	27
王崇德	68	柯平	45	秦惠基	38	张敏	33	张静	31	王敏	28	缪其浩	26
毕强	58	张玉峰	45	陈文勇	37	周毅	33	赵捧未	31	王颖	28	张建华	26
侯汉清	58	陈光祚	44	范并思	37	邹志仁	33	崔雷	30	张莉	28	张进	26
秦铁辉	58	苏新宁	44	柳晓春	37	陈伟	32	胡沛泉	30	张燕	28	郑建明	26
夏旭	58	谢新洲	44	焦玉英	36	黄如花	32	李颖	30	周晓英	28	朱庆华	26
杨华	57	严怡民	44	李静	36	李玲	32	王秀成	30	查先进	27	董慧	25
王惠翔	56	何荣利	43	刘磊	36	李艳	32	冷伏海	29	程刚	27	华薇娜	25

表5 各年段新出现的作者

1983-1987		1988-1992		1993-1997		1998-2002		2003-2007	
作者	频次								
袁锁鸿	20	吴国恩	14	夏旭	15	盛小平	18	张永军	15
樊松林	19	陈文勇	13	曹剑奇	13	程刚	17	吴晓伟	15
王秀成	17	刘建明	13	霍国庆	13	刘家真	13	侯长来	13
刘东维	16	谢先江	11	谢阳群	13	冷伏海	10	郭海明	13
陈光祚	11	徐身新	11	张晓林	13	丁蔚	10	龚蛟腾	12
邓路华	11	何佳讯	10	胡明	11	徐建华	10	陆伟	12
黄祥喜	11	吴宏	10	胡时俊	11	范晓虹	9	李宗红	12
匡华	11	江治平	8	罗曼	11	黄奇	9	刘红泉	11
胡昌平	10	刘焕成	8	唐浩	11	周薇	9	杨峰	11
黄立军	9	刘瑞兴	8	吴永臻	10	郝亚玲	8	王平	11
姜岳	9	郭锐	7	杨明华	10	刘玉照	8	王丽华	10
蒋先治	8	黄建东	7	周全明	10	陆宝益	8	高峰	10
李志生	8	李道军	7	吴新年	10	沈丽容	8	张新民	10
林春山	8	李宁	7	郑士贵	9	王均林	8	杨彩霞	10
邱均平	8	刘源甫	7	白崇远	9	夏火松	8	王晋生	10

为反映作者变化情况,我们把六个年段进行了第一作者对比分析,每相临两个年段进行两两对比。得出了表5和表6的结果。在表5中列出了各年段相对于前年段新出现的作者,由于篇幅限制,我们仅列出了发文频次最高的前15位作者,如1983-1987年段相对前一年段新出现的作者中,樊松林以第一作者发文19

篇。我们认为如果一个作者在五年内都没有以第一作者发表论文,那他可能已经隐身或退出了学术界或转向其它学术领域,表6就是关于作者退出情况的统计。表6列出相对于后一年段不发文的作者中,在该年段发文最高的前15位作者。需要说明的是作者是以第一作者算的频次,所以可能会少于前面表中的数据。

表6 各年段情报学论文作者退出情况

1978-1982		1983-1987		1988-1992		1993-1997		1998-2002	
作者	频次								
王洵	10	杨学山	9	刘建明	13	曹剑奇	13	王纯	11
李毅	7	林春山	8	孙绍荣	10	胡明	11	王崇德	11
高兴民	6	纪晏平	7	吴宏	10	胡时俊	11	樊松林	10
何吉成	6	肖国英	7	王秀成	9	唐浩	11	周智佑	9
孟广均	6	叶千军	7	曾祥丰	9	康焕龙	10	严怡民	9
辛希孟	6	张健	7	刘焕成	8	刘廷元	10	黄奇	9
赵国琦	6	曹剑奇	6	吴天俊	8	周全明	10	范晓虹	9
鱼际元	5	丁辛	6	袁国政	8	程磊	9	伍宪	8
林守一	4	樊松金	6	郭锐	7	时文生	9	王均林	8
陆伯华	4	葛耀良	6	黄建东	7	王益明	9	倪莉	7
钱宝峰	4	李道隆	6	李道军	7	白晓煌	8	孟连生	7
石索安	4	刘士诚	6	刘源甫	7	卢晓宾	8	马红静	7
王连安	4	刘毅夫	6	汪冰	7	麻智辉	8	柳晓春	7
王勇	4	秦蕞	6	肖永鑫	7	汤兆魁	8	刘艳苏	7
张振宜	4	常明	6	周智佑	7	熊第志	8	康世云	7

表 7 1978 - 1982 年段论文作者分布

x	y	X = lgx	Y = lgy	XY	XX
1	759	0	2.880242	0	0
2	130	0.30103	2.113943	0.63636	0.090619
3	44	0.477121	1.643453	0.784126	0.227645
4	22	0.60206	1.342423	0.808219	0.362476
5	12	0.69897	1.079181	0.754315	0.488559
6	8	0.778151	0.90309	0.702741	0.605519
7	2	0.845098	0.30103	0.2544	0.714191
8	5	0.90309	0.69897	0.631233	0.815572
9	2	0.954243	0.30103	0.287256	0.910579
10	1	1	0	0	1
14	1	1.146128	0	0	1.313609

$N = 6$ $\sum_{x=1}^N X = 2.857332$ $\sum_{x=1}^N Y = 9.962332$
 $\sum_{x=1}^N XY = 3.685761$ $\sum_{x=1}^N XX = 1.774818$

1984 年帕欧在洛特卡原始处理过程的基础上,通过修改,总结出一套处理洛特卡定律数据的一般步骤和规范标准,得到学界的广泛认可。其具体步骤包括:收集数据、建立数学模型、斜率求解、C 值计算、进行预测,还有结果的检验^[3]。收集的数据按表 7 进行计算,其中建立数学模型就是确定广义洛特卡定律中选定参数,广义洛氏定律公式为: $F(x) = C/x^a$ 。对 1978 - 1982 年度进行计算。首先去掉了 1% 的高产作者,按照处理过程分别计算出 C 和 N

表 8 拟合洛特卡定律的 K - S 检验结果

年段	作者总数	选取作者量	Dmax	K - S 检验值	C 值	a 值
1978 - 1982	986	975	0.021796	0.052202	2.55624	0.76076
1983 - 1987	3190	3157	0.019865	0.02901	2.63842	0.776144
1988 - 1992	5680	5647	0.01574	0.021691	2.89667	0.81829
1993 - 1997	9170	9096	0.01454	0.017091	2.81267	0.805552
1998 - 2002	15572	15263	0.033091	0.013194	2.49767	0.749161
2003 - 2007	23779	23518	0.011994	0.010629	2.47839	0.745224

3 作者共被引分析

1981 年, White 和 Griffith 正式提出作者共被引分析 (Author Co - citation Analysis, ACA), 对于探讨学科结构有着积极的开创意义。ACA 分析现在已有较固定的一般步骤,包括作者选定、构造作者共被引矩阵、转化矩阵、进行聚类与多维尺度分析等^[5]。

本文选择在 30 年中发文量大于或等于 43 篇的前 30 位作者。以中国学术期刊全文数据库武汉大学镜像站为统计源,利用该数据库系统引文检索中的著者同被引检索功能,在线检索出 30 位高产著者的同被引次数,共有 $(30 \times 29)/2 = 435$ 组不同的数据。具体方法是:在检

值,得到拟合公式: $F(x) = 0.76407/X^{2.57348}$ 。

设定检验显著水平为 0.01, K - S 检验值为 0.05781, $D_{max} = 0.02865$, D_{max} 小于检验值。可以认为具有较好的拟合度。表 8 为所有各年段的拟合结果,可看出前三年段 D_{max} 值均小于检验值。后二年段, D_{max} 值大于检验值,其原因主要是 K - S 检验的影响。计算过程中若取显著水平为 0.01, 临界值为: $1.63/\sqrt{n}$ 。n 为各级著者数总和,显然,临界值的大小只受著者总数的影响,即受著者群大小的影响。著者群越大,值越小,检验精度越高;著者群越小,值越大,检验精度越低。著者群也不能太大,因为著者群越大, K - S 检验的精度越高,越不容易通过检验。著者群只能在一定的范围内取值。既不能太小(著者群下限),也不能太大(著者群上限)。著者总数大于著者群上限时,通过 K - S 检验的概率小于 10%^[4]。经查作者总数与对应的 C 值,后二年段都超过了上限,所以不能通过检验。对于作者分布洛氏定律的验证与公式的研究主要集中在上世纪 90 年代,现在新的网络环境下,作者样本数量的大幅增长,这是一个值得深入研究的问题。

索路径中选择“参考文献”字段,在检索词中分别输入两位作者的姓名,二者为逻辑“与”关系,而后得出两位作者的同被引次数以及同被引的文献列表。检索过程为“精确”匹配,检索年限选择为 1978 - 2008 年。检索学科范围为全部。因为检索时间为 2008 年 11 月中旬且数据库有滞后性,所以 2008 年为不完全数据。

通过检索得到共被引矩阵,其中范文田和彭靖里的同被引次数过少,予以删除,另外王伟、王军、杨华等名字发文较多,但经验证这些名字对应了众多的不同作者,故也删除,最后得到 25 人组成的矩阵。初步处理后,利用 SPSS 中的相关分析,将著者同被引次数矩阵转化为

泊松相关矩阵,消除由著者被引次数差异所带来的影响。对这25位高被引作者进行了聚类

和多维尺度分析,并用Pajek进行可视化,得到了图1、图2和图3的结果。

表9 发文量最多的前30位作者共被引矩阵(部分)

	邱均平	王知津	胡昌平	樊松林	张琪玉	张晓林	张琪玉	范文田	马海群	王崇德	毕强	侯汉清	秦铁辉	夏旭	杨华	王惠翔	彭满里	方平	周宁	罗式胜	马费成
邱均平	97	77	6	87	10	0	179	283	26	13	26	28	39	73	11	40	11	143	147		
王知津	97	24	3	46	24	0	36	29	16	19	20	3	7	2	11	10	11	6	99		
胡昌平	77	24	8	76	4	2	19	24	17	1	11	0	2	2	2	6	6	8	104		
樊松林	6	3	8	0	0	0	7	9	2	1	6	2	0	0	12	0	1	4	9		
张琪玉	87	46	76	0	5	0	38	15	19	19	6	15	6	3	3	11	23	3	51		
张琪玉	10	24	4	0	5	0	1	17	8	190	2	1	1	3	0	17	10	2	17		
范文田	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2		
马海群	179	36	19	7	38	1	0	4	3	3	9	8	3	3	11	9	4	1	53		
王崇德	283	29	24	9	15	17	0	4	12	6	12	4	5	43	1	6	3	98	78		
毕强	26	16	17	2	19	8	0	3	12	3	2	1	1	0	0	2	13	2	36		
侯汉清	13	19	1	1	19	190	0	3	6	3	0	1	0	2	0	9	5	1	5		
秦铁辉	26	20	11	6	6	2	0	9	12	2	0	0	0	2	14	1	0	0	19		
夏旭	28	3	0	2	15	1	0	8	4	1	1	0	8	11	0	62	2	10	10		
杨华	39	7	2	0	6	1	0	3	5	1	0	0	8	7	2	35	5	16	4		
王惠翔	73	2	2	0	3	3	0	3	43	0	2	2	11	7	2	2	0	18	3		
彭满里	11	11	2	12	3	0	0	11	1	0	0	14	0	2	2	0	0	0	1		
方平	40	10	6	0	11	17	0	9	6	2	9	1	62	35	2	0	9	1	10		
周宁	11	11	6	1	23	10	0	4	3	13	5	0	2	5	0	0	9	0	28		
罗式胜	143	6	8	4	3	2	2	1	98	2	1	0	10	16	18	0	1	0	25		
马费成	147	99	104	9	51	17	0	53	78	36	5	19	10	4	3	1	10	28	25		

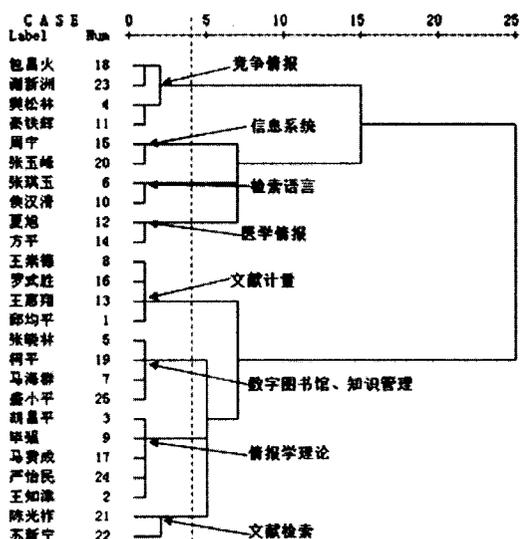


图1 多产作者多层聚类分析结果

图1为改革开放以来我国情报学高产作者的聚类情况,25位作者细分为8个方向,这与实际情况比较相合。①竞争情报。竞争情报研究萌芽于20世纪80年代,它在情报学中占有重要地位。上世纪末至今,竞争情报研究与实践进入新的阶段,纷纷对竞争情报的概念、基本特征、功能、意义和操作方法等基本问题以及竞争情报与知识管理的关系、反竞争情报、竞争情报教育等课题进行了深入的探讨。这几位高产作者是国内较早从事竞争情报理论研究和实践的著名学者。②信息系统。这是一个与计算机网络技术比较密切的领域,内容包括智能信息系统,智能检索系统,信息可视化系统等。③检索语言。检索语言在信息组织与查寻中十分重要,张琪玉和侯汉清在此领域赫赫有名,研究的

内容包括分类法和主题法、信息标引、自动抽词与自动分词、索引编制、语义检索等。④医学情报。医学情报在研究中往往被人们所忽视,其实它有自身独特的研究内容与领域,国内也有一些医学情报院系及系列的医学情报杂志。夏旭为南方医科大学教师,方平为中南大学湘雅医学院信息管理系教授,他们是医学情报方面的代表人物。⑤文献计量。历年来,文献计量都占情报学研究的大部分内容,先后经历了书目统计、文献计量、信息计量和网络计量等阶段。王崇德、罗式胜和邱均平等发表大量相关论文,都出版了文献计量学方面专著。⑥数字图书馆、知识管理。这个方向最为复杂,数字图书馆是图书馆学与情报学交叉领域,将情报学和图书馆学的理论、技术等运用到数字图书馆的服务、管理和利用当中。与国外不同,国内知识管理研究与图书馆学有着千丝万缕的联系。许多知识管理学者都从图书馆学起步或偏向图书馆的知识管理,如盛小平、柯平等。⑦情报学理论。该领域研究者数量最多,研究者之间的关系也较为密切。具体研究内容包括情报学学科基本理论、学科发展、信息用户与服务、信息经济等,具体作者包括严怡民、马费成和胡昌平等。⑧文献检索。情报检索理论和方法的探讨是情报学中最具特色的领域之一,一般认为与情报检索语言比较紧密,但两者侧重点不同,在图中两者显著分开。情报检索在国外情报学学科结构中占据了主导地位也是十分重要的研究内容。但国内这一领域的研究者数量较少,尚未形成较为成熟的学术集团。

