

国际科技合作对中国的科技能力影响分析*

□ 袁军鹏 / 中国科学技术信息研究所 北京 100038

薛澜 / 清华大学公共管理学院 北京 100084

宿洁 / 中央财经大学管理科学与工程学院 北京 100081

摘要: 文章利用科学计量学、数据挖掘技术,以1994—2006年Web of Science (含SCIE, SSCI和A&HCI)收录的中国发表论文为样本,从论文的机构类型分布、引用影响分布、国际科技合作与全部中国文献引用影响的对比、中国与国际著名机构的合作分布进行了研究,从科学计量学的视角描绘了1994年以来的中国参与国际科技合作对中国科技能力的影响以及中国自主创新能力的演变趋势,进而提出了中国各机构加强国际合作、提高自主创新能力应采取的对策思路。

关键词: 国际科技合作, 科学计量学, 数据挖掘, Web of Science

DOI: 10.3772/j.issn.1673-2286.2011.08.010

1 引言

近年来,随着经济全球化的不断发展,一些研究人员开始关心不断发展的科技全球化进程对中国科技能力的影响^[1-2]。任何一项科学研究和技术创造,都要以撰写必要的科学文献为其最后阶段,科学文献的数量和质量无疑是对科技能力的一种量度^[3]。

学术论文、专著作为传递新学术思想、成果的主要物质载体,它们之间的关系并不是孤立的,而是相互联系的,突出表现在科学文献之间的相互引用,因此,引文分析就成为文献计量学、情报计量学、科学计量学的重要组成部分。引文分析(Citation Analysis)就是利用图论、数理统计及其他数学、逻辑

思维方法,对科技文献的引用或被引用现象和规律进行分析,以便揭示出它们所蕴含着的研究对象具有的特征或者对象之间的关系^[4]。通过深入分析文献间互相引用的关系,引文分析在评价研究团队和个人的科研绩效、揭示学科特点和结构以及反映科学研究的焦点领域和发展态势等方面都发挥着重要的作用^[5]。澳大利亚政府工业与科学技术的工业经济局(BIE),曾于1995年对该国的科技体制和科学研究水平进行了评价。对其科研活动的绩效评价主要根据1981-1994年SCI数据库中的指标,通过对科学论文与引文进行国际比较以及对合作论文的考察,分析该国在若干学科领域中的优势与劣势以及国际合作状况(包括与不同地域、在不

同研究领域的合作),进而对国家基础研究的整体水平进行评估^[6]。美国兰德公司在2008年发布的著名研究报告《美国科技竞争力》中,采用SCI论文的被引次数开展了美国与其他国家的科技竞争力的比较分析^[7]。Ding Y. (2010)利用高产作者、高被引作者、高影响力论文等科学计量学指标评估了1960-2009年间语义网领域的研究现状^[8]。Persson (2010)比较了自主论文与国际合作论文的引用影响,发现在高被引论文中,国际合作论文并不如想象中的多,但是对于科研小国而言,国际合作论文占据了其发表的高被引论文的绝大部分数量^[9]。近年来,国外的研究者对中国科学异军突起表现出极大的关注^[10-11],David A. King用引文分析的方法对

* 基金项目: 国家自然科学基金科研课题(70973118) 中国国家博士后基金(20060390049), 中信所预研基金。

31个国家和地区的科学影响进行了分析,其中中国在归一化的篇均引文指标中位于倒数第四^[11]。英国皇家协会在2011年3月公布的一份报告中称^[12],衡量论文质量的重要指标是其他科学家在其工作中对其引用次数的多少,中国在这方面的表现落后于论文数量的增长速度。同样,2011年4月的西班牙《公共报》中也指出,中国目前的论文数量虽然呈现出快速增长趋势,但这些研究成果要达到质量的同等提升,还有很长一段路要走^[13]。

本文在上述研究基础上,利用数据挖掘、科学计量学技术,以1994-2006年Web of Science (含SCIE、SSCI和A&HCI)收录的中国发表论文为样本,对中国机构与国际各类型机构合著WoS论文的机构类型分布、引用影响分布、国际科技合作与全部中国文献引用影响的对比、中国与国际著名机构的合作分布进行研究,试图描绘1994年以来的中国

参与国际科技合作对中国科技能力的影响以及中国自主创新能力的演变趋势。该研究将为我国更好地整合和利用全球优势资源,在更大范围、更广领域和更高层次上参与国际科技合作与竞争,提高自主创新能力和建设创新型国家提供科学决策支持。

2 数据来源与数据库的建立

科学引文索引(SCI)数据库历来被公认为世界范围最权威的科学技术文献的索引工具和科学计量、科学评价的重要工具,是因为它收录期刊的标准高,代表性强以及独特的编排方法所带来的特殊作用。为了获得更全面更权威的数据,我们将SCI数据库和Thomson Scientific公司ISI开发的另外两个同等重要的数据库——社会科学引文索引数据库(SSCI)和艺术与人文科学引文索引数据库(A&HCI)作为研究的数据来源。

2006年10月25日,在清华大学购买的WoS中,以“CU=Peoples R China”为检索对象^①,检索出WoS中所有中国发表的文献,得到结果为441769篇。同日在Thomson Scientific公司的WoS以同样的条件得到结果为442697篇,其查全率(Recall)为99.79%,所以该数据是可靠的。利用信息抽取和自动监测技术将检索到的文献信息建立基于SQL Server的ChinaSCI数据库,得到441769条文献记录。我们定义中国国际合著关系为一篇文章的合著作者中至少有一位来自中国并且至少有一位来自其他国家。以此为依据,设计国际科技合作自动识别算法,得到所有的中国的国际科技合作的WoS文献(SCI、SSCI和A&HCI)为103585篇。我们从著者地址出发,将涉及的44149个机构分为大学、研究机构、企业、医院、政府与国际组织、其他几种类型^②,然后分析各类型机构的文献情况。

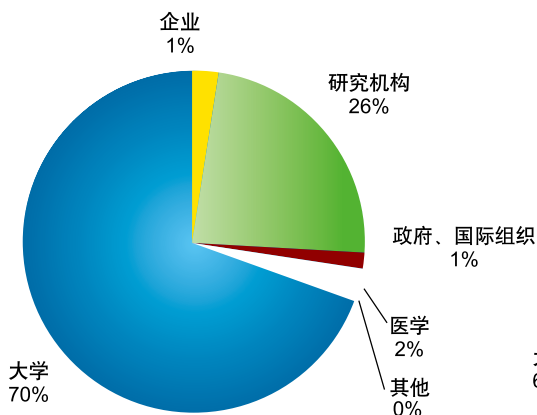


图 1a 中国各类型机构发表WoS文献数量分布图

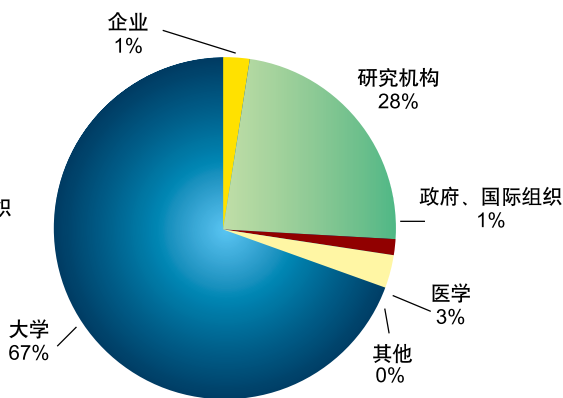


图 1b 中国参与国际合作的机构发表WoS文献数量分布图

图1 中国的各类型机构WoS文献数量分布图

数据来源: 本文中未作说明的图表均来自于Web of Science (清华大学镜像), 2006年10月25日

① 清华大学的web of Science回溯到1994年,所以Time span =1994-2006,另外设定Doc Type=All document types; Language=All languages; Databases=SCI-EXPANDED, SSCI, A&HCI。
② 由于机构名称书写不规范,加上很多大学改名,所以,机构数量存在较大的误差,分类后则可以避免这个问题。

3 结果分析

3.1 各机构类型总体数量分布分析

1994.1-2006.10期间,中国共有26627个机构发表过WoS文献^②,总体情况如图1所示。

在中国,学校和研究机构是WoS文章发表的主要类型,占到中国所有文章的96%,而且在参与国际合作的机构中,学校和研究机构的文章数量也占到了95%。对比图1.a和1.b可以看出,在国际合作的文献中,研究机构的比例上升为28%,医院由2%上升为3%,这二者的上升恰恰是由于学校的比例下降所导致,其他类型的机构发表的文献所占的比例没有变化。从中可以看出,学校是发表WoS的绝对主要类型,其次是研究机构,医院排在第三位,企业等在其中分别贡献1%的文献,这也符合Jaffe & Trajtenberg在2002年的判断^[14]:企业并不是科技文章的主要发表者,其创新的成果多在专利中体现。

3.2 引用影响因子分析

这里有一个公理性的潜在理念是:引用是质量的关系函数,即一般情况下,被引用频次反映的是某种学术思想的影响程度,一项研究成果被引用的次数越多,表明这项成果的影响力越大^[15],所以,我们定义引用影响因子(Citation Impact)为

$$CI=CT/N \quad (1)$$

其中,CT=总被引用次数,N=总文献数量。

(1) 各类型机构的引用影响

因子分析

利用公式(1),计算各类型机构的历年的引用影响因子,得到图2。

从图2可以看出,就引用影响因子来看,由于医院基本只发表特定学科的文章,而这些学科的引用影响因子较高,所以导致医院的一直较高,而且在1997、1998、2003年明显高于其他类型的机构。学校和研究机构不但文献数量较多,而且二者的引用影响因子也较高。从年代的变化趋势来看,医院的引用影响因子自1994年的8.70开始下降到1995年的7.63,然后又上升到1996年的8.13,随后有个快速上升期,到1998

年的18.41,然后开始呈现下降趋势,至2002年的8.15,2003年上升到13.37,其后一直呈下降趋势。学校和研究机构的引用影响因子历年基本保持持平状态,特别是在1997年以后,二者的趋势基本完全一致。政府、企业的引用影响因子历年也呈现波动的趋势,具体趋势见图2。

(2) 各类型机构的国际科技合作的引用影响比较分析

为了比较各类型机构参与国际科技合作与全部的WoS文献的引用影响因子的区别,我们首先以历年的国际合作文献的CI除以历年全部文献的CI,然后按照不同区间统计其分布频次,得到图3。

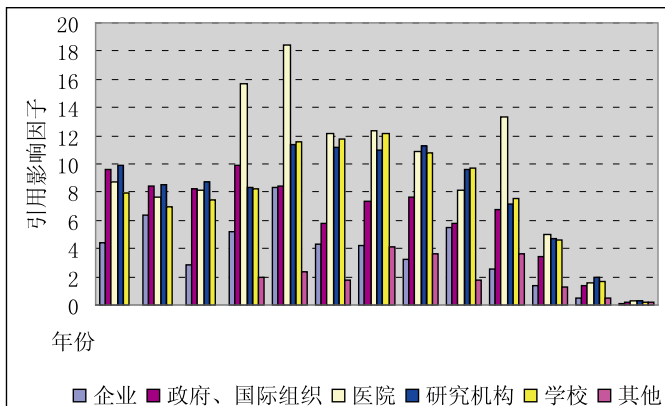


图2 中国的各类型机构发表的全部WoS的引用影响因子按年份分布

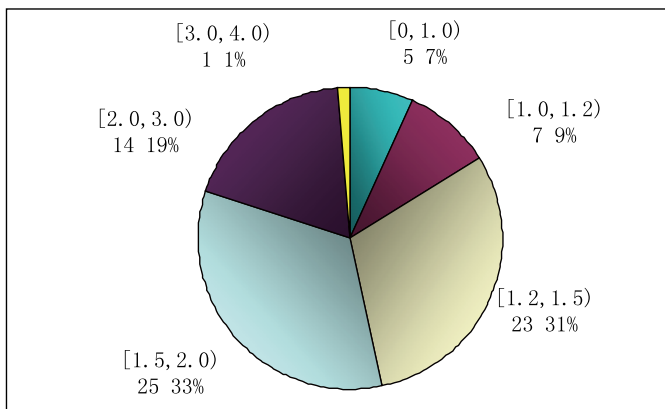


图3 中国的各类型机构发表的合作文献与全部WoS的引用影响因子比值的频次分布

注:图中,[0,1.0)表示比值分布在大于等于0且小于1.0的区间中,5表示频次为5个,7%表示这5个频次占有所有频次的比例为7%

^② 不包括港、澳、台的505个机构,同样,机构数量有一定误差。

从图3来看, 小于1.0的比值只有5个, 占有比值的7%, 说明只有5次(7%)是国际合作文献的CI小于全部文献的CI, 其余的93%全部是国际合作文献的CI高于全部文献的CI, 而且, 有53%的比例分布在[1.5,4.0)区间, 说明有40%的国际合作文献的CI是全部文献的CI的1.5倍以上, 这充分说明参与国际科技合作有助于提高我国论文的引用影响力, 有助于提高我国的科技能力。有40%的比例分布在[1.0,1.5)区间, 说明有些年, 有些机构的中国的全部文献与合作文献的引用影响力因子差距并不是很大, 中国在充分利用国际科技合作、加强自主创新能力后, 完全有可能提高我国论文的引用影响。

由于统计源的学科结构差别, 以及各个学科自身发展的特点和特有引文行为的不同, 如科学家研究行为的社会性、学科间交叉渗透的程度、学科发展所处的阶段等, 引用影响力因子在各个学科之间具有较大的差异性, 由此产生了不同学科论文之间影响力因子的不可比性, 而且, 引文中还有转引、崇引、不恰当的自引、伪引等自身的局限性^[16], 所以, 仅仅使用引文分析的方法不能充分说明问题, 下面我们进一步进行分析我国机构与国外知名机构的合作情况。与国际知名机构合作的文章数量变化趋势可以从一个方面表明中国的研究质量和科技能力的变化。

3.3 与国际知名机构的合作情况

(1) 国际著名大学

依据美国新闻周刊全球前100著

表1 全球前30著名大学

哈佛大学	密歇根大学	苏黎世瑞士联邦理工学院
斯坦福大学	加州大学洛杉矶分校	西雅图华盛顿大学
耶鲁大学	宾夕法尼亚大学	牛津大学
加州理工学院	杜克大学	约翰·霍普金斯大学
加州大学伯克利分校	普林斯顿大学	伦敦大学学院
剑桥大学	东京大学	多伦多大学
麻省理工学院	帝国理工学院	德州大学奥斯丁分校
加州大学圣地亚哥分校	洛桑瑞士联邦理工学院	威斯康辛大学麦迪逊分校
加州大学旧金山分校	康奈尔大学	京都大学
哥伦比亚大学	芝加哥大学	明尼苏达大学

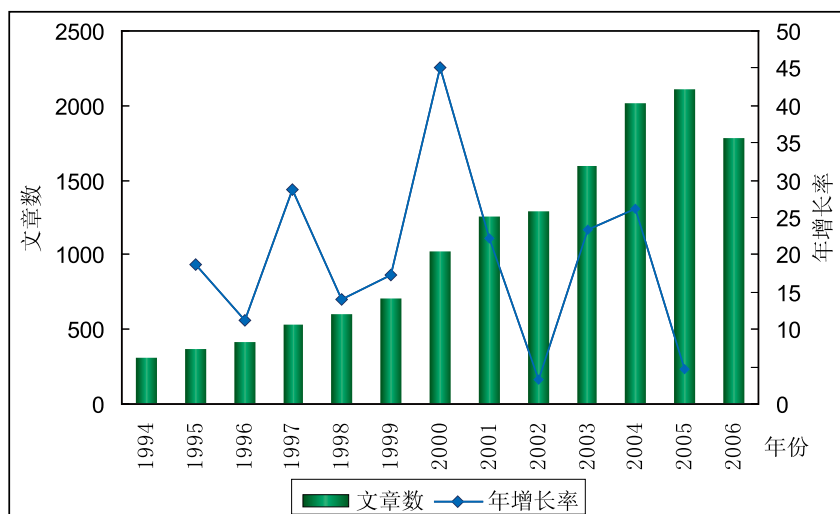


图4 中国与国际著名高校合作历年文章数变化趋势

名大学最新排名选取其中的前30名, 计算中国与这些大学的历年合作趋势, 得到图4。

如图4所示, 中国与这30所著名高校合作的文章数量逐年增长, 由1994年的309篇一直增加到2005年的2101篇, 年增长率一直大于0^①。但是, 从增长速度来看是不均衡的: 1996-1997、1999-2000、2003-2004期间的增长率 (IR, Increasing Rate) 较高, 尤其是2000的IR达到45.16%, 呈现三个波峰; 1996、

1998、2002的IR处于波谷。进一步, 我们分析中国与其中前10名的特别优秀高校的合作情况, 得到图5。

如图5所示, 中国与这些著名高校合作的文章数量逐年增长, 由1994年的175篇一直增加到2005年的1163篇, 年增长率一直大于0。但是, 从增长速度来看是不均衡的: 1996-1997、1999-2000、2003-2004期间的增长率 (IR) 较高, 都超过34%, 呈现三个波峰; 1997-1998

^① 2006年数据只有10个月, 所以增长呈负数。

增长率达到25.83%；1998-1999、2000-2001、2002-2003期间的增长率为11.4%-18.5%；1994-1995、

1995-1996、2001-2002、2004-2005的增长率只有2.6%-8.6%左右。

(2) 国际优秀机构

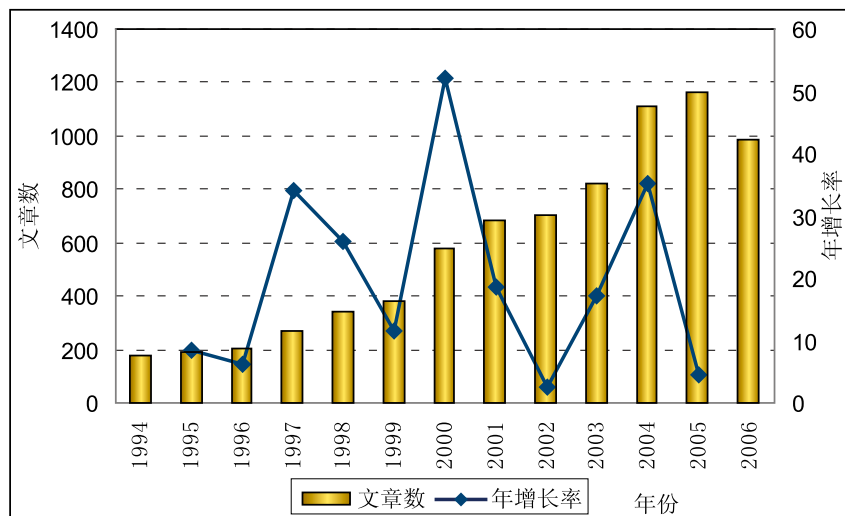


图5 中国与国际著名高校合作历年文章数变化趋势

对于优秀机构的认定目前没有一个权威的、统一的标准。我们利用ESI数据，按照发表论文数、被引用次数、篇均被引用次数、文章数大于500篇的篇均被引用次数、文章数大于1000篇的篇均被引用次数、

文章数大于2000篇的篇均被引用次数等标准取排名前20位的机构，对这120个机构去除重复的机构，得到83个机构，分析中国与这83个机构合作发表论文的情况，得到图6。

从图6可以看出，我国与这

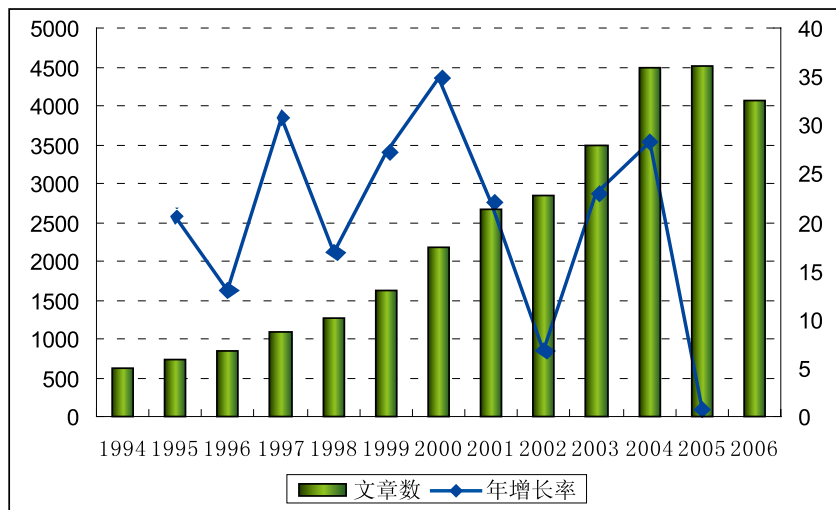


图6 中国与国际优秀机构合作历年文章数变化趋势

些优秀机构的合作文献数量呈现逐年增长的趋势，其文章数量由1994年的612篇增加到2005年的4515篇。但其增长速度也是不均衡的：1997、2000年的IR分别达到30.82%和34.77%；2004年的IR达到28.18%，这三年形成3个波峰，对比与国际著名大学的趋势来看，这三个波峰是同一年达到的；1996、1998、2002、2005年的IR形成4个波谷，与国际著名大学的趋势对比可以发现，形成波谷的年份也很类似，仅在1998年时与前10名高校的有区别。

考虑到以上83个机构过多，尤其是按照篇均被引用次数引入了发表篇数仅为1-2的机构，这些机构是否为优秀机构不是很稳定，所以作者取这83个机构中在上述方法中出现2次以上的为比较优秀机构，得到30个机构。分析中国与这30个机构合作的情况，得到图7。

从图7可以看出，我国与这30个比较优秀机构的合作文献数量与前30名高校一样，也呈现逐年增长的趋势，而且其增长速度也基本与前30名高校的一致。最终2005年的文章数量达到1467篇，较前30名高校的2101篇少了634篇，总体来说，中国与前30名高校的合作文章数量达到13934篇，而和这30个机构的合作总文章数只有9568篇，差距为4366篇。所以，对比这30个机构与30所高校可以看出，中国已经与这些优秀的机构和高校开展了合作，而且其增长趋势基本一致，但是与机构合作的文章数量较纯粹与高校合作的数量少很多。

为了与前10名高校对比，我们取这83个机构中在上述方法中出现3次以上的为特别优秀的机构，

表2 国际比较优秀的机构

AMGEN	GENENTECH INC	UNIV TEXAS	STANFORD UNIV	UNIV TOKYO
BURNHAM INST	UNIV MICHIGAN	NHGRI	NIAID	UNIV TORONTO
CANCER RES UK	HARVARD UNIV	NHLBI	UNIV PENN	UNIV WASHINGTON
NETHERLANDS CANC INST	HOWARD HUGHES MED INST	UNIV CALIF BERKELEY	GLAXO WELLCOME INC	UNIV WISCONSIN
DANA FARBER CANC INST	JOHNS HOPKINS UNIV	ROCKEFELLER UNIV	UNIV CALIF LOS ANGELES	WELLCOME TRUST SANGER INST
EUROPEAN MOLEC BIOL LAB	MAX PLANCK SOCIETY	SCRIPPS RES INST	COLD SPRING HARBOR LAB	SALK INST BIOL STUDIES

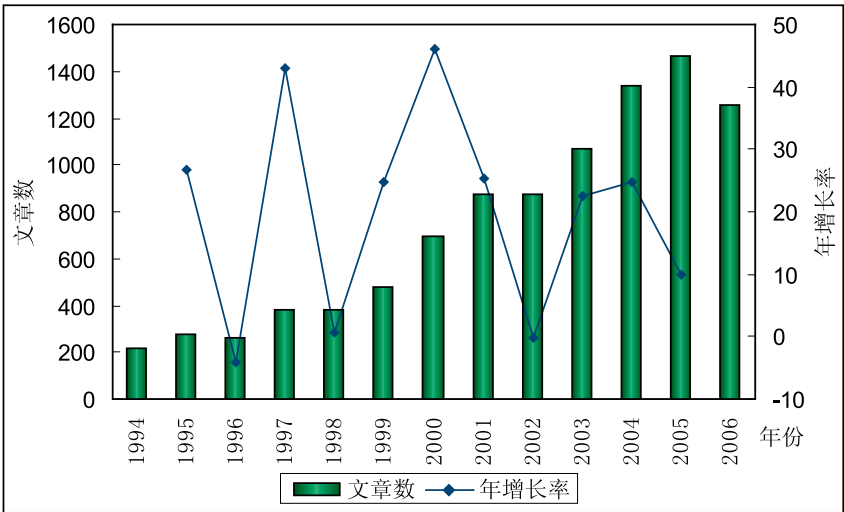


图7 中国与国际比较优秀的机构合作历年文章数变化趋势

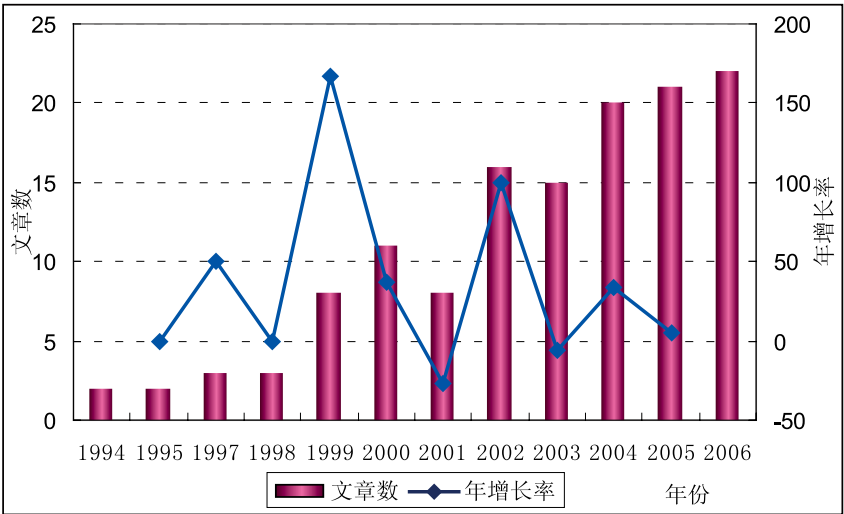


图8 中国与国际特别优秀的机构合作历年文章数变化趋势

得到7个机构：BURNHAM INST、CANCER RES UK、EUROPEAN MOLEC BIOL LAB、GENENTECH INC、HOWARD HUGHES MED INST、ROCKEFELLER UNIV、SALK INST BIOL STUDIES，分析中国与这7个机构合作的情况得到图8。

从图8可以看出，中国与这些机构虽然有合作的文献，但文献数量较少，总量只有131篇文献，数量基本保持一定的增长，但是2000-2001、2002-2003年的增长率为负值，其增长率也是不均衡的。对比前10名高校的数据可以看出：虽然二者都有三个波峰，但是三个波峰的时间点基本不一致，除了1997年外。在数量上来看，与前10名的高校合作的文章数量达到7614篇，较与7名机构的131篇高出57倍多，更加证明了30个机构的结论。

4 结论与政策建议

对中国机构与国际各类型机构合著WoS论文的机构类型分布、引用影响分布、国际科技合作与全部中国文献引用影响的对比、中国与国际著名机构的合作分布的考察，

从科学计量学和数据挖掘的视角了解了1994年以来中国参与国际科技合作对中国科技能力的影响以及中国自主创新能力的演变趋势。从引用影响上来看,医院最高,学校和研究机构这两个发表文章最多的机构保持比较稳定的态势,而且国际科技合作有助于加强我国论文的引用影响,并且,中国已经开始与国际著名机构开展合作,而且基本呈现逐年上升的态势,但是其增长率却是不均衡的,

从数量上来看,中国与国际著名高校合作的论文数量远远大于与优秀研究机构合作的数量。

这些发现对于进一步加强我国参与国际科技合作有着重要的意义。首先,我们应该进一步鼓励中国学者积极参与国际科技合作,这有助于提高我国学者的论文的引用影响力。同时,国家在鼓励国际科技合作的总体政策框架下,也应当注意到不同机构类型和不同机构层次的差异,通过具体的合作政策,

鼓励中国学者选择国际上优秀的机构和著名的学者进行合作,加强与世界一流机构的交流与合作,国家有关政策应该鼓励中国学者在这方面发挥更大作用。

本研究基于WoS论文,我们假定文献能够反映科研的实际水平和学者的主要思想,但其信息传递的可靠性和有效性问题需要进一步研究。另外,引文分析自身的局限性也使本文的结论受到限制。

参考文献

- [1] 中国科技发展研究报告(2000) 研究组. 中国科技发展研究报告(2000)——科技全球化及中国面临的挑战[M]. 北京: 社会科学文献出版社,2000.
- [2] 江小涓. 理解科技全球化——资源重组、优势集成和自主创新能力的提升[J]. 管理世界,2004(6): 4-13.
- [3] 毛莉菊, 席庆奎, 张彬, 等. 南农大SCI论文量与科研经费投入及科研激励机制的关系分析[J]. 农业图书情报学刊,2006(4).
- [4] GARFIELD E. Citation indexes for science. A new dimension in documentation through association of ideas (Reprinted from Science, vol 122, pg 108-11, 1955)[J]. International Journal of Epidemiology,2006, 35(5): 1123-1127.
- [5] 马楠, 官建成. 利用引文分析方法识别研究前沿的进展与展望[J]. 中国科技论坛,2006 (4): 110-114.
- [6] 龚旭. SCI、科研评价与资源优化配置[J]. 科技导报,2002(2): 36-39.
- [7] ALAMA T, HOSEK J. U.S. Competitiveness in Science and Technology [R]. RAND NATIONAL DEFENSE RESEARCH INSTITUTE,2008.
- [8] DING Y. Semantic Web: Who is who in the field - a bibliometric analysis [J]. Journal of Information Science,2010, 36(3): 335-356.
- [9] PERSSON O. Are highly cited papers more international? [J]. Scientometrics, 2010, 83(2): 397-401.
- [10] KOSTOFF R N. The (Scientific) wealth of nations[J]. Scientist,2004, 18(18): 10-10.
- [11] KING D A. The scientific impact of nations[J]. Nature, 2004, 430(6997): 311-316.
- [12] Knowledge, networks and nations Global scientific collaboration in the 21st century[R]. The ROYAL SOCIETY,2011.
- [13] 戴维·布劳纳特. 官僚主义和霸权阻碍中国科技腾飞[N]. 西班牙《公共报》,2011-04-09.
- [14] JAFFE A B, TRAJTENBERG M. Patents, Citations, and Innovations: A Window on the Knowledge Economy[M]. Cambridge MA/London: MIT Press,2002.
- [15] HARSANYI M A. Multiple authors, multiple problems bibliometrics and the study of scholarly collaboration - a literature review[J]. Library & Information Science Research,15(4): 325-354.
- [16] 庞龙, 张培富, 杨立英. 引文分析方法在国内外应用的比较研究[J]. 山西大学学报(哲学社会科学版),2006, 29(3): 134-137.

作者简介

袁军鹏 (1973-), 男, 博士, 副研究员, 研究方向: 科学计量学, 科技政策。E-mail: junpengyuan@gmail.com
薛澜 (1959-), 男, 博士, 教授, 博导, 研究方向: 公共政策、科技政策、创新管理、危机管理。
宿洁 (1975-), 女, 博士, 副教授, 主要研究方向: 优化理论与算法。

A Scientometrics Analysis of the Participation in International Science and Technology Cooperation on China's Scientific and Technical Ability

Yuan Junpeng / Institute of Scientific and Technical Information of China, Beijing, 100038

Xue Lan / School of Public Policy & Management, Tsinghua University, Beijing, 100084

Su Jie / School of Management Science and Engineering, Central University of Finance and Economics, Beijing, 100081

Abstract: Based on the published paper from 1994-2006 with a Chinese address samples, included in the Web of Science (including SCI, SSCI and A&HCI), this paper tries to adopt scientometrics and data mining approach to analyze patterns of Chinese international academic research collaboration., including disciplinary distribution of papers, institution types, ratio of international collaboration. This analysis revealed some interesting characteristics of China's research cooperation with the world. Policy recommendations are made on how to improve the productivity of basic research and international science and technology cooperation in China.

Keywords: International Science and Technology Cooperation, Scientometrics, Data Mining, Web of Science