

doi:10.3772/j.issn.2095-915x.2016.02.003

国际专利引文研究的计量分析

滕立^{1,2}, 黄兰青³

(1. 郑州航空工业管理学院信息科学学院 郑州 45001; 2. 航空经济发展河南省协同创新中心 郑州 450015;
3. 深圳实验学校 深圳 518049)

摘要: 本文以共被引和共词分析方法,以SCI中2000年至2013年的有关专利引文的研究文献为对象,分析了其知识基础与研究主题。结果表明,专利引文研究主要是建立在Jaffe AB(1993)、Narin F(1997)、Griliches Z(1990)和Trajtenberg M(1990)等人研究基础之上,核心研究领域集中在企业/产业的创新能力与绩效评价,涉及知识溢出、技术扩散、竞争优势等内容。关键词聚类在战略坐标中的象限分布则表明专利引文研究处于既定范式下的深度研究,缺少对范式扩展的广度研究。另外,研究还发现专利引文研究较少关注基础研究与应用技术间的互动关系。

关键词: 共被引分析, 共词分析, 专利引文, 知识基础, 研究主题

中图分类号: G306

文献代码: A

Informetrics Analysis of International Patent Citation Study

TENG Li^{1,2}, HUANG LanQing³

(1. Information Science Department, Zhengzhou University of Aeronautics, Zhengzhou 450015, China;
2. Collaborative Innovation Center for Aviation Economy Development of Henan Province, Zhengzhou 450015, China;
3. Shenzhen Experimental School, Shenzhen 518049, China)

Abstract: Articles of patent citation from SCI between 2000 and 2013 were analyzed in this paper by co-citation and co-word analysis methods in order to explore their knowledge base and research topics. The results have shown that four classical papers written by Jaffe Ab in 1993, Trajtenberg M in 1990, Griliches Z in 1990, Narin

基金项目: 国家知识产权局2016软科学项目“基于专利引文的发明专利质量指标评价体系研究”,青年学者优秀论文奖励计划。

作者简介: 滕立(1970-),汉族,四川内江人,博士,航空经济发展河南省协同创新中心副研究员。研究方向:科学计量学、情报学、科学技术政策等。联系方式:tengli1970@163.com;黄兰青(2000-),汉族,河南郑州人,学生,在本文中负责收集文献资料、数据下载与处理、文字校对等工作。

F in 1997 respectively formed the key knowledge base for patent citation field. The main research topics concentrated on innovation, performance evaluation of industries or firms, including content of knowledge spillovers, technology diffusion, competition advantage, etc. Distribution of keywords clusters implied that patent study was deep research under given paradigm while lack of study on expanded paradigm. Moreover, the results of this paper indicated that few studies focused on relationships between basic research and applied technologies of patent citation study.

Key words: Co-citation analysis, co-word analysis, patent-citation, knowledge base, research topics

引言

专利引文指已获授权的专利所引用的参考文献,包括引用的专利文献和非专利文献。1987年,纳林(Narin F)^[1]等人就提出专利引文可以有效地衡量企业技术竞争优势。随后,专利引文在国家、地区及产业层面的创新能力与绩效评估、竞争情报、技术跟踪与技术发展趋势预测等领域获得了不同程度的应用,并取得了一定成效。经过20多年的发展,国内外对专利引文研究的知识基础及其主要研究内容却少有人从整体上关注与研究。那么,国际上对专利引文的研究处于什么状态?专利引文研究是建立在怎样的知识基础之上?其主要研究内容是什么,有什么特点?对这些问题的回答与分析,有助于研究人员更好地认识与把握专利引文研究的相关理论与概念,了解专利引文研究中的热点与发展趋势,并为确定其研究方向提供基础决策信息。本文以SCI数据库中2000年至2013年所发表的有关专利引文的科学论文为研究对象,利用科学计量学中成熟的共被引分析方法和共词分析方法,对上述问题加以研究和解答。

1 研究方法介绍与指标体系

1.1 主要研究方法

两篇文献同时被其他文献引用,则这两篇文献间就存在共被引关系^[2]。在给定的文献集合中,参考文献通过共被引关系构成了文献共被引网络。文献共被引网络反映了对应研究领域的知识基础。陈超美^[3]的研究表明,在共被引网络中具有较高中介中心度^[4]的节点文献往往是这一领域中的经典文献,是构成其知识基础的主要成份。

共词分析的方法是1983年由卡龙(Callon M)^[5-7]等人提出并加以完善的一种对文献内容进行分析的方法。该方法首先利用文献关键词中存在的共现关系,构建相应的共词网络。其次,利用相应的聚类方法对网络中的关键词进行聚类,并使用密度^[7]、中心度^[7]与战略坐标^[7]等指标,构建各聚类间的层次关系。位于战略坐标第一象限(密度与中心度均大于均值)的关键词聚类代表了领域中的核心研究,位于第二象限(中心度大于均值、密度小于均值)的则为活跃主题,处于第三象限(密度大于均值、中心度小于均值)的为老化主题,第四象限(密度与中心度均小于均值)中的主题为边缘研究。

关键词的聚类依据卡龙^[7]提出方法进行。

1.2 主要测度指标

1) 中介中心度

网络中节点中介中心度指标的含义是在网络中,任一节点占有网络中最短路径的能力。该指标为弗里曼^[4](Freeman LC)于1979年提出。在文献共被引网络中,高中介中心度的节点在网络不同部分起到联通的功能。因此,在本文中,高中介中心度节点文献可被认为是领域知识基础的关键节点,起到整合领域知识基础的作用,并代表了相关研究领域知识基础的主要成份。

2) 共现强度

共现强度表征了节点间的连接强度。本文以余弦指数对其进行计算,其公式如下:

$$\text{Cosine} = \frac{F(A, B)}{\sqrt{F(A) \times F(B)}} \quad (1)$$

其中 $F(A)$ 、 $F(B)$ 表示关键词/文献A、B在给定文献集中出现的次数; $F(A, B)$ 为关键词/文献A、B共现次数。该指数的取值范围在0至1之间,值越大,表明关键词/文献A、B间的共现强度越高。

3) 共词聚类的密度、中心度与战略坐标象限的划定

共词聚类的密度反映了聚类的凝聚程度,以其内部成员间的共现强度和表征。密度越大,聚类内部成员间的联系越紧密;中心度反映了聚类间的联系,以其成员与其他聚类成员的共现强度和表征。一个聚类的中心度越大,说明其与其他聚类间的联系越紧密结合。在战略坐标中,横轴表示聚类的中心度,纵轴表示聚类的密度,其坐标原点由各聚类的密度与中心度的平均值确定。

4) 外部连接度

外部连接度反映的是关键词间的联系。在对

关键词共现网络进行聚类后,每一个关键词都属于唯一一个聚类。其中,与聚类外部的关键词联系的最紧密的关键词起到聚类间知识沟通的主渠道作用。其计算方法为统计指定聚类中某一关键词与除本聚类内部的关键词外其他关键词的共现强度和。聚类中该值最大的关键词代表了此聚类与其他聚类间知识上的联系主要通道。

2 数据来源及其分布

基于SCI数据库在科学文献研究领域中的权威性,本文以SCI数据库为数据源,检索式设定为:(1)TS="patent citation*" (主题精确检索);(2)时间范围设定为2000年至2013年;(3)文献形式设定为"Article(论文)";(4)语言设定为"All language(所有语言)";(5)数据库选择SCI、SSCI及会议论文数据库。检索结果显示共有446篇文献。这些文献的时间分布情况见图1。

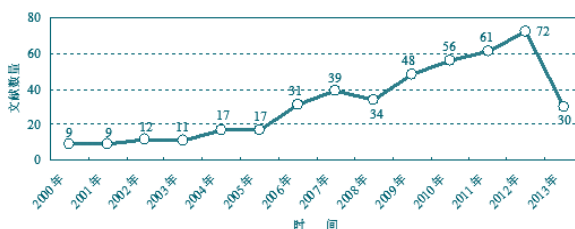


图1 国际专利引文研究文献年度分布图

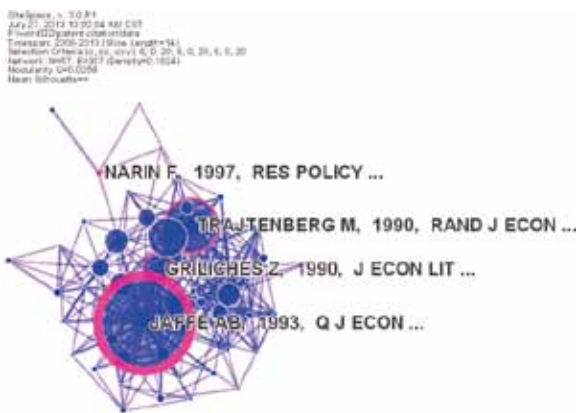
从图1中可以看出,国际专利引文研究论文随着时间呈现出不断增长的趋势:由2000年的9篇增加到2012年的72篇。2013年由于数据录入不全,故只检索到30篇。文献量的不断增加表明国际专利引文研究正在得到越来越多的研究者的重视与参与。

3 国际专利引文研究的知识基础

在CitespaceII¹软件中,设置分析对象为“参考文献”,时间范围设定为2000年至2013年、

1 <http://cluster.cis.dvexel.edu/cchm/citespace/>

时间段取 14 年、阈值设定为 (c,cc,ccv)=(6,0,20) (表示分析对象为出现 6 次或以上且与其他参考文献的共现强度大于等于 0.2 的参考文献)。运行结果为一由 57 个节点与 307 条边构成的文献共被引网络 (见图 2)。图 2 中粉色外圈的节点表示其在网络中具有较高的中介中心度 (≥ 0.1)，黑色字体标示出了这些节点的相关信息，如作者、发表时间和发表的期刊等。表 1 列出了其相关的统计指标。



这些文献的主要研究内容分别为：

1) Jaffe AB 等人在 1993 年发表的论文 *Geographic localization of knowledge spillovers as evidenced by patent citations* 中研究了被引专利收到的专利引文的地理分布情况，并以此来描述知识溢出与地理位置的关系。

2) Trajtenberg M 在 1990 年发表的论文 *A penny for your quotes: patent citations and the value of innovations* 中指出简单的统计专利数量并不能很好地评价创新性产出，并提出利用专利引文做为指标来评价创新价值。

3) Griliches Z 在 1990 年在发表的 *Patent statistics as economic indicators: a survey* 一文中通过分析总结专利数据的特征，认为专利数据是研究技术变迁的重要途径。

4) Narin F 与合作者 1997 年发表的 *The increasing linkage between US technology and public science* 一文中，利用专利引文，研究了美国产业中技术对基础研究的依赖情况。通过分析美国专利对科学论文的引用，纳林等人发现来自公共部

门，尤其是来自顶级大学与实验室和受到 NSF、NIH 资助的基础研究，对美国的技术进步产生了很大的影响。

4 国际专利引文研究的主题结构

提取 446 篇文献中出现次数大于等于 3 次的关键词，去噪并标准化后，共得到 77 个关键词。其共现网络见图 3，相关节点统计指标见表 3。



图 3 国际专利引文研究关键词共现网络
(注：图中显示的为部分中介中心度大于 0.07 的关键词)

表 3 国际专利引文研究关键词指标统计（中介中心度 ≥ 0.04）

序号	关键词	频次	中介中心度	序号	关键词	频次	中介中心度
1	clusters	9	0.11	12	increasing returns	10	0.07
2	bayh-dole act	8	0.1	13	organization	7	0.07
3	pharmaceutical-industry	9	0.09	14	agglomeration economies	14	0.06
4	firm performance	8	0.09	15	inventors	11	0.06
5	search	14	0.08	16	statistics	9	0.06
6	invention	13	0.08	17	multinational-corporations	8	0.06
7	industry	57	0.07	18	research-and-development	127	0.05
8	biotechnology	32	0.07	19	knowledge flows	23	0.05
9	mobility	15	0.07	20	product development	12	0.05
10	linkage	13	0.07	21	collaboration	10	0.05
11	intellectual property	11	0.07	22	knowledge	56	0.04

在 77 个关键词中，共有 4 个关键词的频次超过 100 次，分别是“专利引文（patent citations）”的频次为 310 次、“创新（innovation）”为 205 次、“知识溢出（knowledge spillovers）”为 156 次、“研发（research-and-development）”127 次。整个关键词网络的密度

为 0.0991，处于比较低的水平。另外，高频关键词的中介中心度值普遍较低，高中介中心度值的关键词则普遍频次较低。

利用 CitespaceII 软件的矩阵提取功能，获得其共现矩阵并按 1.1 节的聚类方法对进行主题聚类，共得到 9 个聚类，各聚类内容见表 4。

表 4 国际专利引文研究关键词聚类

聚类编号	聚 类 内 容
1	patent-citations 专利引文；innovation 创新； knowledge-spillovers 知识溢出；research-and-development 研发；technology 技术 geographic-localization 地理定位；patents 专利；firms 企业；knowledge 知识；growth 增长 foreign-direct-investment 外商直接投资；United States 美国； multinational-corporations 跨国公司；absorptive-capacity 吸收能力；invention 发明； agglomeration-economies 集约经济；international-technology-diffusion 国际技术扩散； localization 本土化；local-search 局部检索；university-research 高校科研 Korea 韩国；Taiwan China 中国台湾； statistics 统计；brokerage 中介；patent-citation-network 专利引文网络；market 市场； semiconductor-industry 半导体产业；knowledge-diffusion 知识扩散；networks 网络； information 信息 bayh-dole-act 拜杜法案；academic-research 学术研究； firm-performance 企业绩效；intellectual-property 知识产权；increasing-returns 增长回报； empirical-analysis 实证分析；science 科学；China 中国；knowledge-flows 知识流； biotechnology 生物技术 flows 流；examiner-citations 审查员引文； inventors 发明人；models 模式；productivity-growth 生产率增长；exploration 探索； competences 能力；technological-search 技术检索；citation-analysis 引文分析 indicators 指标 organization 组织；mobility 移动性 capabilities 能力；clusters 集群；protection 保护；pharmaceutical-industry 制药产业； competitive-advantage 竞争优势；collaboration 合作；performance 绩效； product-development 产品开发 patent-analysis 专利分析；portfolios 档案； market-value 市值；industry 产业；perspective 视角 investment 投资；productivity 生产率； economic-growth 经济增长；development-spillovers 发展溢出；competition 竞争； Europe 欧洲；impact 影响 search 检索；evolution 评估 determinants 决定因素；linkage 连接
2	
3	
4	
5	
6	
7	
8	
9	

表4中各聚类加黑的两个词是该聚类的起始词,表征了该聚类的研究主题,其余成员表征了围绕该主题展开的相关研究内容。聚类1的研究主题为专利引文与创新,主要内容包括利用专利引文研究企业间的知识溢出及评估其创新能力;聚为2的主题为外商直接投资对美国的影响,涉及国际技术扩散、跨国公司等内容;聚类3的主题为韩国和中国台湾,主要是从专利引文网络和知识扩散的角度对比二者在半导体产业的竞争优势;聚类4为针对拜杜法案的相关研究,实证分析对象主要为生物技术,研究内容主要为知识产权与企业绩效;聚类5的主题为专利审查员引文,主要内容有引文分析、技术检索和指标体系,以及通过专利审查员引文分析企业的生产率增长模式;聚类6的主题为组织移动性,内容涉及制药产业的竞争优势、产品开发、企业间的合作与绩效等方面;聚类7的主题为专利分析,主要是从产业视角研究其市值;聚类8研究的主题为生产

率与投资的关系,研究对象为欧洲,涉及经济增长、竞争发展溢出等内容;聚类9的研究主题为检索与评估。

表5列出了根据1.2节的方法计算的各聚类的相关统计指标。从中可以看到,聚类1、2、4、6位于战略坐标的第一象限内,其余则位于第四象限。在第二和第三象限中没有主题分布。其分布情况见图4。4聚类个分布在第一象限,构成了专利引文研究领域的核心主题,5个聚类位于第四象限,形成领域中的边缘主题。整个专利引文研究领域没有老化的研究主题(位于第三象限)和新兴的研究主题(位于第二象限)。高频关键词集中在聚类1中,但是其中高中介中心性的关键词较少。与之对应的是高中介中心度的关键词主要集中在位于第四象限的聚类4和6中,以及位于第一象限的聚类2中。

考察各聚类的最高外部连接度关键词(表5),可以看到关键词“专利引文”在其中外部连接度

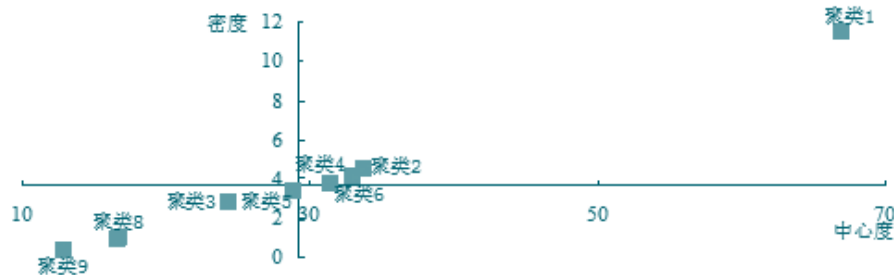


图4 国际专利引文研究关键词聚类在战略坐标象限中的分布

表5 国际专利引文研究关键词聚类指标统计

聚类编号	中心度	密度	战略坐标象限	各聚类最高外部连接度的词	外部连接度
1	66.7778	11.5023	I	patent-citations 专利引文	10.8213
2	33.5739	4.5703	I	local-search 局部检索	4.8106
3	24.2004	2.8850	IV	networks 网络	6.2890
4	32.7600	4.2527	I	biotechnology 生物技术	5.1983
5	28.7676	3.4434	IV	models 模式	4.9165
6	31.3444	3.8557	I	performance 绩效	5.2102
7	16.5575	1.0715	IV	industry 产业	6.6747
8	16.7228	1.1098	IV	productivity 生产率	3.6820
9	12.7812	0.4837	IV	determinants 决定因素	4.5318
平均值	29.2762	3.6861			

最高,是各聚类知识联系上的主要节点,在整个网络中居于统治地位。外部连接度大于 5 的词还有“网络”、“生物技术”、“绩效”、“产业”。这些词与“专利引文”共同构成了连接各聚类的骨干节点。

5 结论

通过对国际专利研究的论文进行分析,发现其知识基础主要是建立在四篇经典文献之上。这四篇文献中通过对专利数据的特征进行分析,Griliches Z 确立的专利数据在研究中使用上的可靠性,Trajtenberg M 提出利用专利引文做为研究企业创新的指标,Jaffe AB 利用专利引文的地理分布研究了企业/产业间的知识溢出,Narin F 从专利引文的角度探讨了基础研究对技术进步的影响。这些开创性的研究奠定了专利引文研究的范式基础。

对比相应的专利引文研究主题,可以看到四个位于战略坐标第一象限的核心主题与知识基础具有明显的对应性。这些主题涉及产业间的知识溢出、创新能力、企业绩效、技术竞争优势等内容,其共性是利用专利引文做为指标从宏观视角对产业创新能力进行评估,并探讨影响创新与绩效的因素。结合关键词聚类在战略坐标中的分布、关键词网络密度和中介中心度的定义,可以合理地推论出专利引文分析的研究主题聚集于聚类 1 和 2,但是主题间的联系不紧密,即主题间只通过少数关键词进行联系。这种现象表明整个专利引文研究仍处于新兴生长阶段,研究者关注的问题比较集中,整个领域处在既有范式下的深度研究,缺少对范式扩展的广度研究。

值得注意的是,Narin F 开创的利用专利引文研究基础研究对技术进步的影响在研究主题中却没有出现。这表明虽然专利引文在企业能力评价中获得广泛应用,对基础研究与应用技术间互动

关系的研究还没有得到相应的重视。笔者认为,从理论与实践上对基础研究与应用技术间互动关系的探究,总结其互动规律,既有助于产业/企业更好地吸收利用基础研究的成果,增强其创新能力,也有助于宏观科学技术政策的制定,并能使我们更深刻地认识专利引文所表征的含义,完善其应用。相信这方面的研究会是未来发展的一个方向。

参考文献

- [1] Narin, Francis, Elliot, et al. Patents as Indicators of Corporate Technological Strength[J]. Research Policy, 1987, 16(2-4):143-155.
- [2] Small H. Cocitation in scientific literature: new measure of relationship between 2 documents[J]. Journal of the American Society for Information Science, 1973, 24(4): 265-269.
- [3] Chen C, Chen Y, Horowitz M, et al. Towards an explanatory and computational theory of scientific discovery[J]. Journal of Informetrics, 2009, 3(3):191-209.
- [4] Freeman L C. Centrality in social networks conceptual clarification[J]. Social networks, 1979, 1(3): 215-239.
- [5] Callon M, Law J, Tuner W, et al. From Translations to Problematic Networks: an introduction to co-word analysis[J]. Social Science Information, 1983(22): 191-235.
- [6] Callon M, Law J, Rip A. Mapping the dynamics of science and technology : sociology of science in the real world[M]. Macmillan, 1986:10.
- [7] Callon M, Courtial J P, Laville F. Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry[J]. Scientometrics, 1991, 22(1):155-205.
- [8] 滕立, 沈君, 高继平. 共词知识网络中的认知结构: 理论、方法与实证 [J]. 情报学报, 2013, 32(9):976-989.