

面向高校科技成果转化效率评估的 “科学—技术”关联性实证研究^{*}

□董珏 杨眉 郭晶 宋海艳

摘要 通过对科学研究和技术应用的产出关联性和引用关联性分析,可以反映科学研究与技术应用间的相互扩散、转移及贡献程度,从而为合理地指导科学研究与技术应用资源配置和绩效评价提供依据。选取教育部自然科学类高校排名前3的国内理工类高校近30年(1985—2014年)的论文和专利产出、专利引用关系等7个数据变量共21种组合关系,采用关联性分析对现有的分析方法和指标进行修正,有效反映科学—技术在时间维度上的关联性。结果表明,关联性分析在实证研究中具有较好的适应性,其分析结果具有统计意义,能够消除突变噪声,为管理决策提供合理依据。

关键词 学科—技术关系 关联性分析 转化效率

分类号 G306

DOI 10.16603/j.issn 1002-1027.2016.05.006

1 引言

“科学—技术”关系的实践研究发展于上世纪90年代,弗朗西斯·纳林(Francis Narin)通过专利对论文引用在来源和时间方面的关联性,证明在高科技领域,科学与技术之间有着紧密的相互作用^[1]。利用论文和专利数据可以衡量国家、地区或机构的创新能力测度以及相互间知识转移、技术转移途径的研究,主要是基于宏观数据进行统计分析^[2-5]。基于专利的科学—技术关系的研究,主要包括定量指标方法、数理模型方法、科学—技术映射模型及社会网络分析方法。指标方面主要有科学联系度(Science Linkage)及其标准化后的产业标准化指数、当前影响指数(Current Impact Indicator)及衍生的科学强度、技术循环时间(Technology Cycle Time)、技术扩散系数、相对强弱指数(RSI)等^[3-6]。数理模型方面,有基于熵值算法学校—产业—政府的三螺旋模型与算法(已由Mode-1进化到Mode-2)^[7-12],技术成长曲线及其修正^[8],以及用于科研投入与产出绩效间的灰色关联分析^[9];科学—技术

映射模型方面,国内学者尝试建立科学学科分类与专利IPC分类的映射模型^[10-16]等;社会网络分析方法,如高继平提出的专利/论文的混合共被引网络分析、聚类分析和聚类自动标引^[11],基于社会网络分析工具的研究等^[12]。定量指标方法难以避免突发因素造成的数据噪音,例如论文或专利的数量的突发性增长或引用;模型研究和社会网络分析其关注的指标较为局限,例如模型研究主要是宏观的规模性产出数据;社会网络分析则要求变量数据间有共现关系。文章涉及7个论文和专利的数据变量,得到21种关系组合,扩展了定量指标体系,同时发现了一些原本弱相关或无关的数据变量去除时滞影响后呈现出较强或极强的相关性。

“科学—技术”相互作用分析能体现出科学研究与技术应用间的协同发展、扩散转化和相互贡献程度,从而为支持科技领域布局、制定技术创新策略、合理配置科研资源等提供决策依据。文章以国内高等学校“科学—技术”发展现状为研究对象,通过高校的专利数据和论文数据关联性分析,挖掘影响科

^{*} 上海交通大学文科科研创新培育基金项目“高校图书馆战略情报服务模式与体系研究”(编号:12QN03)的研究成果之一。

通讯作者:董珏,ORCID:0000-0002-0752-4021,jdong@lib.sjtu.edu.cn。

学和技术发展的关联性因素,探讨科学与技术相互转化、扩散、相互促进的时间效率。一项研究,或某一领域的研究,在研究发表后需要多长时间能被关注(被引)、被传播、以及促进技术应用的出现,通过时间度量来揭示该问题,能够为高校科技管理部门开展科技资源布局提供有效的决策支持的定量分析方法体系和策略。

2 研究方法

关联性分析方法在很多学科领域得到了应用,在挖掘“科学—技术”关联性分析方面,可以分析研究经费的投入产出,但其涉及的变量局限于产出数据,而忽略了引文关系,即科学与技术间的转化和扩散效率^[13];可以分析科研项目与专利产出间的关联性,但其忽略了科研与技术之间的时间转化差异,其相关性系数均低于 0.9^[14]。

对于科学与技术的关系,通常认为:先有科学研究再有专利产出,即研究总是走在应用的前面;成果公开之后会经过一段时间得到关注、扩散、被利用。那么一个机构、一个学科、一个技术领域甚至单个的研究内容发表之后,究竟要多长时间才能爆发出大量的技术应用成果?文章扩展了“科学—技术”常用

的分析指标涉及的数据变量(见表 1),考量了科学与技术发展的时间差异性,一方面挖掘了更多的可用指标;一方面将原本弱相关的关联性通过时间关系处理得到了较强或极强的关联性,并分析该时间差异对科技转化效率的影响。

从表 1 看出现有文献的研究方法主要有两种:一是考察单一变量的变化趋势或者分布情况;二是对技术相关变量(如专利被引次数与专利公开量、引用论文文献和专利公开量或论文发表量)间通过逻辑运算得到的数值进行分析,形成衡量科学—技术关联性、活跃度和影响力的常见指标体系,运算方法主要包括平均数值、标准化数值、引用关系时间间隔等。但其存在的共性问题:变量关系单一,且依赖于指标的建立来对数据进行处理,无法消除由突发因素引起的数据噪音。

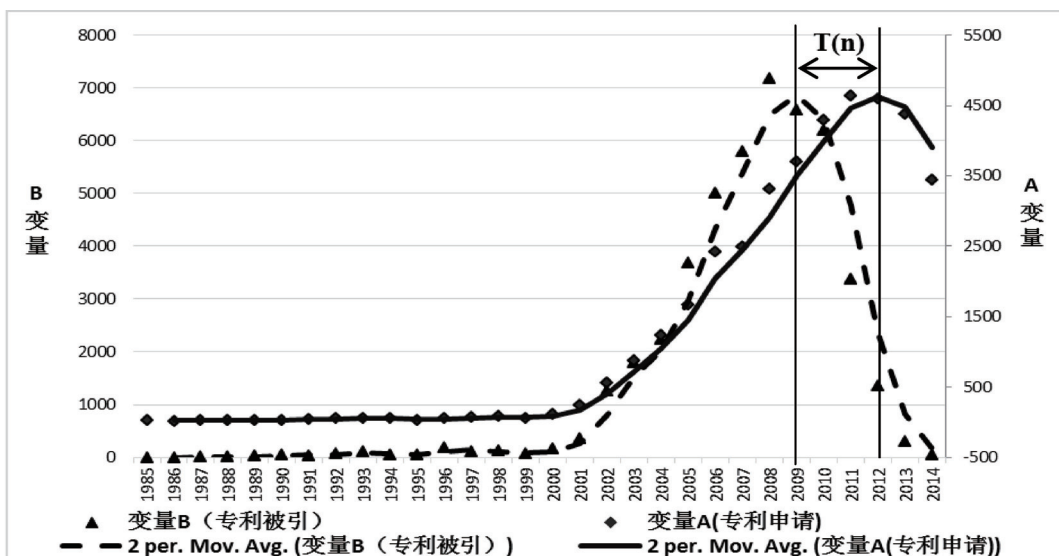
2.1 关联性分析方法

在不同的变量间可能存在着线性相关或曲线相关关系,可以是正相关也可以是负相关,不同类的相关又可分为强相关、中等相关、弱相关或者无关等几种关联程度^①。而曲线相关大多可以转换线性相关进行研究,衡量变量间的线性关系常用的系数为皮尔森(Pearson)相关系数。假设:变量 A(专利申请)

表 1 相关研究指标

指标名称	计算方法	分析意义
平均被引数量	专利被引次数/专利总数	技术影响力
当前影响指数	(5 年内)专利被引次数/专利被引次数	技术创新质量
技术强度	专利申请/授权量 * 当前影响指数	总体技术水平
技术循环时间	专利引用论文文献与发明专利之间的时间间隔(中值)	不同领域知识转移和技术转移的速度
科学关联度	专利引用论文科学论文/专利总数	技术创新对基础科学研究的依赖程度
科学强度	专利申请/授权量 * 科学关联度	专利与科学研究关联度活跃性的总体水平
产业标准化值	某产业科学联系度/所有产业科学联系度均值	对专利引用时滞问题的修正
相对研发能力	专利数 $\times 1.0$ + 专利被引用次数 $\times 1.2$ + 专利自我引用次数 $\times 0.8$	整体的专利水平和专利质量
技术独立性	引用自己专利的次数/该专利被引用次数总和(专利被引用次数 + 专利自我引用次数)	研发自主性强
技术扩散速度	专利申请与专利被引的平均时间间隔	技术活跃度
相对强弱指数	主体 A 的专利引用主体 B 论文数量与主体 A 论文被引总数的比值/主体 B 的论文数量与所有论文数量的比值	某一技术的专利与所有技术类别专利的平均引文水平的差异
科学集中度	某一专利分类的非专利引用数量/该分类专利中总的非专利引文数量	—

① 皮尔森相关性系数范围为(-1,1),负值表示负相关,正值表示正相关。绝对值为 0.9—1.0 极强相关,0.8—0.9 强相关,0.5—0.7 中等程度相关,0.2—0.5 弱相关,0.0—0.2 极弱相关或无相关。且需满足 95% 及以上的置信区间。

图1 变量A和B的时间序列趋势图^②

与变量B(专利被引)之间存在潜在的关联性,当 $N=30$ (1985—2014年),A与B两个变量曲线走势见图1,但两条曲线并不完全重合,可以首先计算得到一个皮尔森相关系数: $P_0(A_1^N B_1^N)$ 。

$$P_0(A_1^N B_1^N) = \frac{N \sum_{i=1}^N A_i B_i - \sum_{i=1}^N A_i \sum_{i=1}^N B_i}{\sqrt{N \sum_{i=1}^N A_i^2 - (\sum_{i=1}^N A_i)^2} \sqrt{N \sum_{i=1}^N B_i^2 - (\sum_{i=1}^N B_i)^2}} \quad (1)$$

对于同一年份的数据来说,专利被引相对于专利申请有所滞后。这一规律在图1中表现为变量B的峰值在时间上较变量A提前出现。因此,在计算相同年份的变量A和B的相关性之后,尝试寻找变量B的峰值相对变量A峰值前移的一般规律,计算不同年份的变量A和B之间的相关性;当 P_n 达到最大值时, n 即为变量B相对于变量A的峰值前移时间差,也可以理解为变量B相对于变量A的时滞。新的线性相关系数计算公式如下:

$$P_n(A_n^N B_1^{N-n}) = \frac{(N-n) \sum_{i=1}^{N-n} A_{i+n} B_i - \sum_{i=1}^{N-n} A_{i+n} \sum_{i=1}^{N-n} B_i}{\sqrt{(N-n) \sum_{i=1}^{N-n} A_{i+n}^2 - (\sum_{i=1}^{N-n} A_{i+n})^2} \sqrt{(N-n) \sum_{i=1}^{N-n} B_i^2 - (\sum_{i=1}^{N-n} B_i)^2}} \quad (2)$$

当 P_n 达到最大值时, n 即为A变量对B变量产生作用的时滞度量。

$$T(n) = \text{MAX}(P_n(A_n^N B_1^{N-n})) \quad (3)$$

② 图中线条均为趋势线,不代表具体数值。

③ R方值表明了模型对样本数据的拟合程度,其值越高说明模型对样本数据拟合度好。R方值为0—1间,若大于0.8说明线性关系较好;若在0.7—0.8间,说明还有其他因素影响线性关系;若低于0.7则说明线性关系不强。

关联性分析主要解决两个问题:(1)对于相互之间没有直接关联性的变量,通过考察其时间序列变化情况,挖掘其在时滞效应和转化速度方面的关联性,以考察其预测性。(2)消除单个数据带来的噪声影响,使分析结果具有较好的稳定性和合理性。

2.2 方法验证

通过曲线拟合,比较 A_1^N 和 B_1^N ,以及 A_n^N 和 B_1^{N-n} 两组曲线的皮尔森系数和R方值^③,若皮尔森系数R方值得到改善或有明显提高,且通过SPSS分析软件得到的置信区间均在95%及以上,则说明该方法对于该变量组合具有统计意义。

文章通过列举两个例子来进行验证说明:(1)选择机构G的30年间(1985—2014年)论文发表(A1)和专利申请量(B1),并假设论文研究能带动专利的申请,分析在当前的发展趋势下,预测专利申请要经过时间 n 才能达到当前科学研究水平带来的相应技术应用规模;(2)选择机构G的30年(1985—2014年)专利申请量(A2)和专利被引量(B2)进行关联性分析,与现有的“技术扩散速度”指标进行参照分析。两组变量30年时间序列数据的趋势分布与关联性见表2和图2。

表 2 两组变量 0—30 年的关联性

时间范围	A1 与 B1	A2 与 B2
1—30	0.900 **	0.681 **

注：** 相关性在 0.01 层上显著；* 相关性在 0.05 层上显著^①。

经公式(2)计算得到表 3。可见,G 单位专利申请与论文发表间的关联性,在时间差值 $n_1=2$ 年时达到最强,预测 2 年后在当前的总体研究态势下,会引发技术引用的大量出现;专利申请与专利引用之间原本呈现的弱关联性,在时差 $n_2=3$ 年时获得较强的关联

性,且推测得出 G 单位的技术扩散速度为 3 年。

经过该处理后,通过 SPSS 以 A 为自变量,B 为因变量进行曲线拟合,R 方值都得到了明显改善(见表 4)。证明该关联性方法具有统计意义,能够表征两个变量间的关联程度。

同样,通过数据验证,关联性分析 7 个变量(论文发表、论文被引、专利申请、专利授权、专利被引、引用专利、引用论文)的 21 类组合关系都呈现出较好的应用可行性。

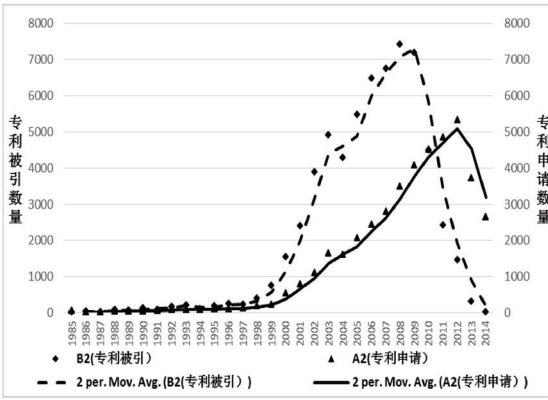
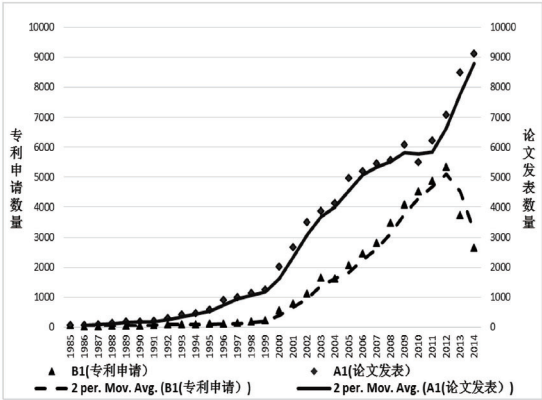


图 2 两组数据的趋势分布

表 3 不同变量组合的差值关联性分析

n 值 变量	0	1	2	3	4	5	6	7	8	
A1 与 B1	0.900 **	0.933 **	0.938 **	0.925 **	0.91 **	0.892 **	0.875 **	0.854 **	0.833 **	
A2 与 B2	0.681 **	0.816 **	0.921 **	0.963 **	0.914 **	0.825 **	0.770 **	0.747 **	0.702 **	

注：** 相关性在 0.01 层上显著；* 相关性在 0.05 层上显著。

表 4 经时差处理后两组变量的曲线拟合度比较

R 方	线性	对数	倒数	二次曲线	三次曲线	复合	次方	S	成长	指数	Logistic
A1 与 B1 ($n_1=0$)	0.809	0.619	0.220	0.824	0.933	0.863	0.925	0.487	0.863	0.863	0.863
A1 与 B1 ($n_1=2$)	0.879	0.565	0.230	0.937	0.958	0.937	0.893	0.519	0.937	0.937	0.937
A2 与 B2 ($n_2=0$)	0.350	0.526	0.263	0.649	0.650	0.343	0.592	0.466	0.343	0.343	0.343
A2 与 B2 ($n_2=4$)	0.836	0.592	0.298	0.844	0.861	0.841	0.891	0.663	0.841	0.841	0.841

④ ** 相关性在 0.01 层上显著,表示置信区间为 99%;* 相关性在 0.05 层上显著,表示置信区间为 95%。

表 5 三所高校变量间关联情况

A	专利申请	专利授权	专利授权	专利授权	专利授权	论文发表	论文发表	专利申请	论文被引	论文发表	论文发表	论文发表	论文发表	专利申请	专利申请	引用专利	引用专利	专利申请	引用论文	引用论文	引用专利
B	专利授权	论文被引	引用专利	引用论文	专利被引	专利授权	专利申请	论文被引	专利被引	论文被引	专利被引	引用专利	引用论文	专利被引	引用专利	论文被引	专利被引	引用论文	论文被引	专利被引	引用论文

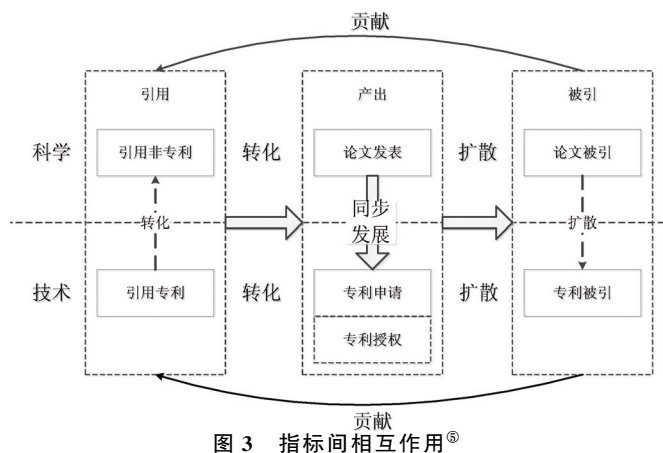
3 评价体系的构建

文章的数据包括专利和论文两部分。专利数据来源于 TI 数据库,以专利权人代码为检索条件,获得 1985—2014 年国内排名前 3 的理工科高校(高校 S、高校 Q、高校 Z)的专利公开和引用情况;论文数据来源于 InCites 平台中 3 所高校 1985—2014 年的论文发表和被引情况。

3.1 评价指标的确立

从论文和专利数据中提取出 7 类变量,通过前期的关联性分析后,得到 21 个组合形式并确定变量 A 对变量 B 具有促进或推动作用(表 5)。

根据主/客体(论文与专利)与行为(引用与被引)之间的时间维度关联性(图 3),15 对变量组合可以大致划分为四个类别:



(1) 表征产出规模驱动的时间效应,体现在不同主体具有同样的行为:论文发表—专利申请,即科学研究产出对技术应用产出的推动作用。

(2) 表征扩散时间效应,体现为不同主体的产出与被引行为的时间维度关联性,又可以细分为两类:一类包含论文发表—论文被引、论文发表—专利被引、论文被引—专利被引,即科学研究的扩散对科学研究和技术引用产出的时间影响;一类包含专利申请—专利被引、专利申请—论文被引,即技术应用的

扩散对新技术应用和科学研究扩散的时间影响。

(3) 表征转化时间效应,体现为不同主体的产出与其引用行为的时间维度关联性,分为两类:一类包含专利申请—引用论文、专利申请—引用专利、引用专利—引用论文,即技术转化对科学研究和新技术应用产出的作用;一类包含论文发表—引用专利、论文发表—引用论文,即科学研究转化对新的研究内容、技术应用和技术应用转化的影响。

(4) 表征贡献程度,体现为引用行为与被引行为的时间维度关联性,也可分为两类:一类是科学研究对新的科学研究转化和技术应用转化的贡献度,包括引用论文—论文被引、引用论文—专利被引;一类是技术应用对新的科学研究转化和技术应用转化的贡献度,包括引用专利—专利被引、引用专利—论文被引。

3.2 机构数据对比

通过公式(3)的计算处理后,我们得到以下结果(表 6)。比对国内 3 所水平较高的理工类院校的论文和专利的产出及引用关系数据,其论文数量与专利信息的产出规模、扩散效应、转化效应和相互贡献程度都具有极高的关联性。对比变量 A 对变量 B 的推动或促进作用在时间维度的间隔大小,可以挖掘各高校间的“科学—技术”间扩散、转化速度的快慢,以及相互贡献程度的时滞效应。

从关联性来看,高校 Q 除了专利申请与专利被引、专利授权与论文发表、论文发表与专利被引、论文发表与专利引用专利文献、论文发表与专利引用论文文献和专利被引与专利引用专利文献 6 种组合呈现出中等相关外,其他变量间都呈现出较强相关;高校 S 和高校 Z 在专利被引与专利引用专利文献、专利被引与专利引用论文文献 2 组变量无关联性。

从时间关系上来看,专利授权与其他变量的关系,和专利申请与其他变量的关系差值在 0—2 之间,整体抵消了我国专利申请到授权的 18 个月审核

^⑤ 箭头指向为由变量 A 到变量 B,即 A 对 B 有促进或推动作用;实线箭头表示变量间有直接关联,虚线箭头表示变量间有间接关联。

表 6 标杆机构关联性分析结果

分类	A	B	T(n)			P 值比较					
			高校 S	高校 Q	高校 Z	高校 S		高校 Q		高校 Z	
产出规模	论文发表	专利申请	2	2	1	.942 * *	.987 * *	.900 * *	.938 * *	.958 * *	.968 * *
	论文发表	专利授权	2	2	2	.722 * *	.973 * *	.720 * *	.940 * *	.789 * *	.968 * *
	专利申请	专利授权	1	1	1	.904 * *	.980 * *	.904 * *	.984 * *	.909 * *	.991 * *
扩散效应	论文发表	论文被引	4	3	3	.810 * *	.969 * *	.839 * *	.954 * *	.822 * *	.964 * *
	专利申请	论文被引	0	0	0	.918 * *	—	.896 * *	—	.873 * *	—
	论文被引	专利被引	2	2	1	.804 * *	.971 * *	.876 * *	.941 * *	.896 * *	.931 * *
	论文发表	专利被引	5	5	4	.442 *	.920 * *	.604 * *	.907 * *	.576 * *	.917 * *
	专利申请	专利被引	4	3	3	.622 * *	.905 * *	.681 * *	.963 * *	.663 * *	.971 * *
	专利授权	论文被引	0	0	0	.955 * *	—	.943 * *	—	.943 * *	—
	专利授权	专利被引	2	2	2	.800 * *	.917 * *	.800 * *	.946 * *	.808 * *	.973 * *
转化效应	论文发表	引用论文	2	3	2	.534 *	.859 * *	.653 * *	.884 * *	.449	.778 * *
	专利申请	引用论文	1	1	1	.671 * *	.783 * *	.871 * *	.933 * *	.723 * *	.730 * *
	论文发表	引用专利	2	3	2	.586 * *	.933 * *	.595 * *	.784 * *	.607 * *	.887 * *
	专利申请	引用专利	1	1	1	.711 * *	.863 * *	.905 * *	.925 * *	.812 * *	.900 * *
	专利授权	引用专利	0	0	0	.733 * *	—	.963 * *	—	.867 * *	—
	专利授权	引用论文	0	0	0	.720 * *	—	.958 * *	—	.803 * *	—
	专利引用	引用论文	0	0	0	.987 * *	—	.955 * *	—	.967 * *	—
贡献程度	引用论文	论文被引	1	1	4	.694 * *	.712 * *	.876 * *	.847 * *	.502	.691 * *
	引用专利	论文被引	1	0	1	.728 * *	.842 * *	.849 * *	—	.658 * *	.827 * *
	引用论文	专利被引	3	2	3	.205	.730 * *	.817 * *	.899 * *	.502	.809 * *
	引用专利	专利被引	3	2	2	.211	.875 * *	.693 * *	.920 * *	.426	.930 * *

周期的时间,因此两者可按需选择,文章以专利申请为主要变量开展分析。另论文被引和专利被引(2,2,1)在时间关系上同步于论文发表和专利申请(2,2,1),即一项/篇专利和论文被引用的同时,意味着另一项/篇专利和论文公开发表。因此两组变量具有相互验证的效果(表 6 中灰色部分为文章未选取的变量组合)。

(1)从产出规模来看:高校 Z 对科学—技术的产出发展较快,其变量间的时间间隔均小于高校 S 和高校 Q。

(2)从扩散效应来看:高校 S 与高校 Q 技术扩散对科学的影响时效长于高校 Z(论文被引—专利被引、论文发表—专利被引、专利申请—专利被引),而科学研究扩散对新的科学研究的促进(专利申请—论文被引)高校 S 要迟于高校 Q 和高校 Z,科学研究扩散对新的技术应用的促进(论文发表—论文被引)三所高校时效相同。

(3)从转化效应来看:在技术转化方面(专利申

请—引用专利),三所高校的效率相似;但高校 Q 的科学研究转化效率(论文发表—引用论文、论文发表—引用专利)略低于其他两所高校。

(4)从贡献程度来看:高校 Z 的科学贡献度较差(引用论文—论文被引),高校 Q 的科学贡献度和技术贡献度(引用论文—专利被引、引用专利—专利被引)较好,高校 S 的技术贡献度和科学贡献度(引用论文—专利被引、引用专利—专利被引)都略低于高校 Q。

总体来看,高校 Z 近年来的产出发展较快,其科学研究向技术应用方面的转化速度较快,但是技术与科学的关联度较低,科学贡献度较低,但技术影响力较高。高校 Q 近年来科技产出变缓,科学研究向技术应用扩散速度较快,高校 Q 的科学研究转化效率较高,科学与技术的关联度居中,技术循环时间变长,但其技术贡献度和影响力仍较高。高校 S 在产出发展速度上略缓于其他两所高校,其技术扩散效率不高,科学贡献度和技术贡献度的水平居中。

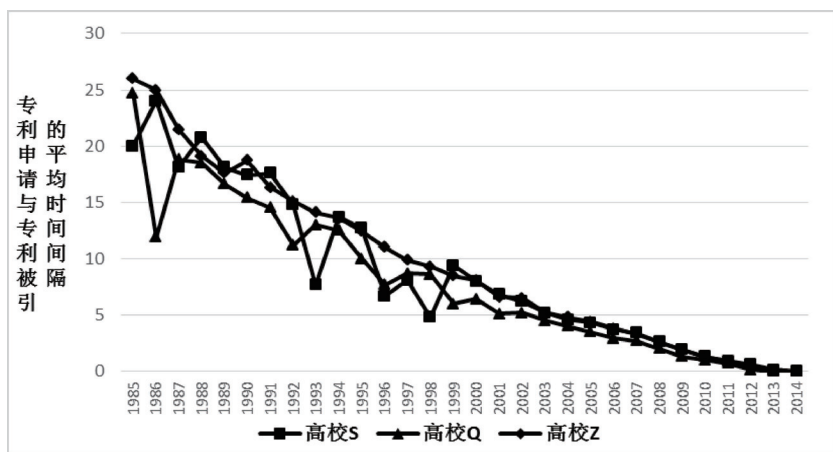


图4 高校S、高校Z、高校Q专利申请与专利引用趋势图

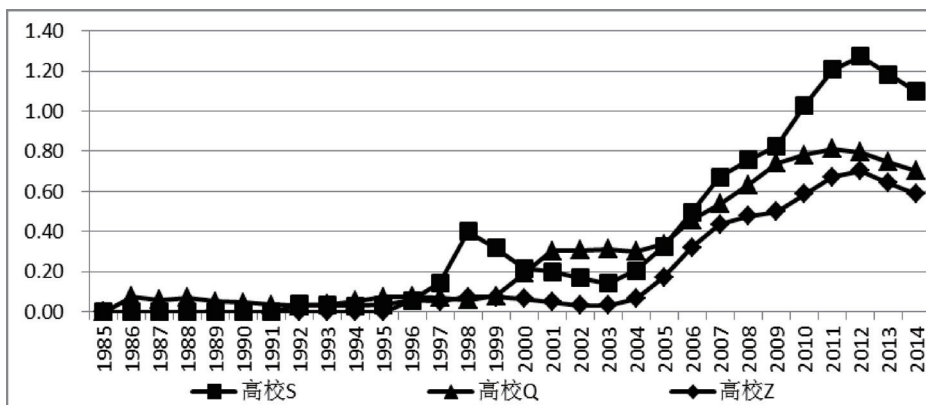


图5 高校S、高校Z、高校Q历年科学关联性趋势图

3.5 与传统指标的对比检验

传统指标受到时间变化以及突现的大数值个例影响,会导致指标结果具有数据样本的依赖性。通过分别比较扩散效应与“技术扩散速度”指标,转化效应关联性分析与“学科关联度”和“技术循环时间”指标(见表1)的数据值,对关联性方法的可行性进行检验。

(1) 扩散效应关联性与“技术扩散速度”指标。

在已有的科学—技术分析中,计量扩散效应的是(专利申请)与(专利被引)的平均时间间隔,即“技术扩散速度”这个指标来度量的,用这个办法得出来的一个高校的技术扩散速度,选取的时间范围距离现在越远则值越大,选取的时间范围距今越近值越小。例如:图4中,选取30年的数据时,校S、高校Q和高校Z的均值分别为9.1、8.2和9.9年,中值为8.7、6.2和8.3年;选取近十年的数据时,三者的技术扩散速度分别为2.3、1.8和2.4年,中值为2.2、1.6和2.2年。

关联性的分析则是对两个变量的总体趋势进行对比,根据表6的结果,在30年的数据范围内,科学研究的扩散对科学研究和技术引用产出的时间影响(论文发表和论文被引:4、3、3,专利申请和专利被引:4、3、3),以及技术应用的扩散对新技术应用、科学研究和科学研究扩散的时间影响(论文发表和专利被引:5、5、4)的时间效应更加的均衡和稳定。

(2) 转化效应关联性与“科学关联度”和“技术循环时间”指标。

传统指标中评价科研转化和技术转化的指标有科学关联度和技术循环时间两个指标。通过计算学科关联度得到图5,得到的2014年累积值分别为1.10、0.70、0.59,高校S的科学关联度较高,高校Q科学关联度居中,高校Z的科学关联度较低。该结果与三者的科学研究转化对新的研究内容、技术应用和技术应用转化关联分析结果(1,1,1)近似。

近五年(2010—2014年)技术循环时间方面的高校S的技术循环时间(4、5、3、3、3年)与高校Z(5、

4、4、3、2年)近似,较高校 Q(8、7、4、3、4年)略快,且年代越近,其技术循环时间越小。通过关联性分析,技术转化对科学研究和新技术应用产出的作用(2、3、2),高校 Q 略缓于高校 S 和高校 Z,与传统指标的结果相近,但受数据集合时间范围的影响更小。

4 结论

在现有的科研产出绩效评价体系中,对于产出规模的考量远多于对于“科学—技术”。即重视“论文发表—专利申请”的数量发展,但是对于一个机构、一个学科的“科学—技术”发展状态的衡量,还要综合考虑科学研究与技术应用的扩散效应、转化效应和贡献度,从多种角度考察发展现状,选择有效的评价因子,制订合理的评价机制。文章提出的关联性方法,将现有指标化评价转化为从时间维度的关联性评估,有效地解决了突变因子的负作用,有利于更加客观、均衡地表征“科学—技术”间的发展状态。

关联性分析在大样本量的数据集合中,例如学校层面具有很好的表征作用,最大程度地挖掘了变量间的影响关系,且可以消除历史数据或突变数据带来的负面作用,更加合理地体现变量间相互作用的时间关系。但在高校内部学科层面的集合中,变量间的关联程度不够稳定,该问题主要来源于小规模数据集合的时间序列的不完整性,以及专利引用动机的复杂性。因此,可以通过两种手段进行进一步的修正,例如:对于时间序列的缺失,可以采用累积统计的办法消除时间序列空白,保证统计意义上的完整性和有效性;对于专利引用动机复杂性,可以通过专利信息中的引用来源标注(申请引用和审查

员检索引用)和关联性代码(对于引用文献是对技术背景的描述,还是技术新颖性判断等)进行进一步的数据清理和深层次关系的挖掘。

参考文献

- 1 Narin F. Patent bibliometrics[J]. Scientometrics, 1994, 30(1): 147—155.
- 2 杨中楷,沈露威. 国家关系视角下中国在美专利引用状况计量研究[J]. 图书情报工作, 2010, 54(22):80—84.
- 3 陈凯,徐峰,程如烟. 非专利引文分析研究进展[J]. 图书情报工作, 2015(5):137—144.
- 4 肖国华,王春,姜禾,郭婕婷. 专利分析评价指标体系的设计与构建[J]. 图书情报工作, 2008, 52(3):96—99.
- 5 韩宇. 知识转移及其分析方法研究[D]. 天津:天津大学, 2004.
- 6 刘倩楠. 基于专利引文网络的技术演进路径识别研究[D]. 大连:大连理工大学, 2010.
- 7 Leydesdorff L. The triple helix of university-industry-government relations[J]. Scientometrics, 2003, 58(2):191—203.
- 8 肖沪卫. 专利地图方法与应用[M]. 上海:上海交通大学出版社, 2011.
- 9 姚颖靖,彭辉. 专利资助政策功能的实证研究——基于 2007 年中国 30 个省区的灰色关联分析[J]. 现代情报, 2011, 31(7):20—26.
- 10 李睿. 专利引文分析法与共词分析法在揭示科学—技术知识关联方面的差异对比[J]. 图书情报工作, 2010, 54(6):91—93.
- 11 高继平,丁堃,滕立,庞杰. 专利—论文混合共被引分析法的实现及其应用:以德温特专利数据库为例[J]. 情报学报, 2012, 31(3):317—324.
- 12 殷媛媛,肖沪卫. 基于论文专利的科学技术互动发展趋势研究——以立体显示产业为例[J]. 情报杂志, 2011, 30(6):25—29.
- 13 逢淑媛,陈德智. 专利与研发经费的相关性研究——基于全球研发顶尖公司 10 年面板数据的研究[J]. 科学学研究, 2009, 27(10):1500—1505.
- 14 袁翠敏. 美国 NSF 科研项目立项与专利信息关联关系分析的案例研究[D]. 吉林:吉林大学, 2015.

作者单位:上海交通大学图书馆,上海,200240

收稿日期:2016 年 5 月 4 日

An Empirical Study on Science-Technology Relationship for University Technology Conversion Evaluation

Dong Jue Yang Mei Guo Jing Song Haiyan

Abstract: By analyzing output and correlation of citation activities between the scientific research and technology applications, it is expected to reflect the diffusion, conversion and contribution between the scientific research and technology application, and to provide data support for how to allocate the resources and evaluate performance of scientific research. The paper selects the top 3 Chinese universities in nature science, makes analysis on 7 variables in 30 years (1985—2014), draws abstract of 21 combinations of correlation between these variables, and modify the existing analytical methods based in indexes by using correlation analysis, in order to reflect the time efficiency of science and technology relevance effectively. The results show that correlation analysis is not only adaptable but also statistically significant in empirical research, while it is possible to eliminate noise mutation, and provide a reasonable basis for making decisions.

Keywords: Science-Technology Relationship; Correlation Analysis; Conversion Efficiency