基于主成分分析的企业专利实力评价研究

齐 萍, 凌 锋, 刘会景

(中国科学技术信息研究所, 北京 100038)

摘要:企业的专利实力反映了企业的技术创新能力、市场竞争力和企业发展的活力。对企业专利综合实力进行评价,能使企业把握科技发展的动态,了解技术创新中的优势和不足,并科学地调整企业的创新策略。本文以发明专利作为研究对象,选取了有代表性的专利评价指标,从申请类、授权类、专利维持、专利范围、专利引证五个维度构建了企业专利实力评价指标体系,并运用主成分分析方法,对2001至2005年美国专利申请量排名前20的企业专利综合实力进行评价。

关键词:专利;实力;指标体系;专利评价;主成分分析方法

中图分类号:D923.49 **文献标志码:**A **文章编号:**1671-1807(2015)01-0089-05

众所周知,专利实力是一个企业乃至一个国家技术创新能力和科技竞争力的重要表现。企业竞争,实质上是科技的竞争。谁拥有先进的技术、拥有较多的新技术,谁就有较强的竞争力,并在竞争中取胜。专利权是新技术的专有权,企业的竞争也表现在企业拥有质量好、数量多的专利上的竞争。依靠专利,企业的发展可以从无到有,从小到大,从弱到强;依靠专利,进行企业产品的保护,产品的专利权,不仅意味着该产品在技术上的专有权,也意味着在市场上的专有权。对企业专利综合实力做出科学的比较和评价,能使企业及时把握技术创新的节奏,从总体上了解技术创新中的优势和不足,并科学地制定相应的创新政策,从而提高企业的市场竞争力,获得更大的经济效益。

近几年,专利指标体系的研究得到越来越多学者的关注,如黄庆等^[1]将专利评价指标分为数量类、质量类和价值类三类;唐炜^[2]从总量指标、按时间分布研究的指标、按地理分布研究的指标、按人的要素分析的指标和专利引文指标几个横向层面对专利分析所使用的指标进行整理、定义;于晶晶、谭思明^[3]构建了由表征企业专利活动的数量类指标、表征企业专利的经济技术价值的质量类指标和通过各指标加权得出的综合类指标三部分构成的专利组合分析评价指标体系,等等。

本文采用 ISTIC 美国专利数据库中的原始数据,从申请类、授权类、专利维持、专利范围、专利引证五个维度构建了企业专利实力评价指标体系,并运用主成分分析方法^[4],对2001至2005年美国专利申请量排名前20的企业专利综合实力进行评价。

1 专利实力评价指标的构建

本文在评价时以发明专利作为研究对象,主要原因在于:

1)发明专利经过了最为严格的审查程序,其技术创新程度和权利稳定性都最高,市场经济价值最大,评价结果最可信。

2)便于进行国际比较,国际上对专利的认识是基于发明专利的,因此从发明专利出发比较国际间专利水平的标准基本上是统一的,有可比性。

1.1 发明专利的申请类指标

该类指标主要反应了专利意识和对专利的关注程度,主要用发明专利申请情况来表征:发明专利申请量,即以相同年为单位统计的发明专利申请量。

1.2 发明专利的授权类指标

发明专利申请要经过严格的实质审查才能获得授权,美国发明专利授权一般晚于其申请2.5年至3.5年不等。授权类指标可以反映专利申请中获得授权的情况,也在一定程度上反应专利申请文件的撰

收稿日期:2014-10-16

基金项目:国家科技部科研院所技术开发研究专项课题项目(2014EG146067)研究成果之一。

作者简介:齐萍(1986—),女,山东高密人,中国科学技术信息研究所万方数据股份有限公司研究员,硕士,研究方向:专利分析、专利数据挖掘;凌锋(1977—),男,硕士,研究方向:信息咨询与服务等;刘会景(1984—),女,硕士,专利分析师,研究方向:专利分析、专利数据挖掘。

写质量和技术新颖性、创造性、实用性,用以下几个指标来表征:

1)发明专利授权量,即以相同年为单位统计的发明专利授权量。

2)发明专利授权率,即以相同年为单位统计的发明专利授权量/以相同年为单位统计的发明专利申请量。

1.3 发明专利维持指标

专利维持^[5]是指通过定期缴纳一定费用保持专利有效性的行为。根据市场法则,只有当专利价值高于维持专利权的成本时,专利权人才会继续维持专利,因此可以用专利的存活量和存活率作为表征专利价值的指标。美国发明专利年费,自授权日起计算:在办理专利授权登记时所缴付的费用,已经包括 1—4 年的年费;在第 3.5 年前,缴付第 5—8 年的维持年费;第 7.5 年前,缴付第 9—12 年的维持年费;在第 11.5 年前,缴付所余专利权年限的维持年费。美国交付年费的正常时间是 1 年期,分别是 3—4 年、7—8 年、11—12 年,如从第 3—3.5 年、7—7.8 年、11—11.5 年是按照以上费用缴付,但在 3.5—4 年、7.5—8 年、11.5—12 年的时候,则属于宽限期,需额外补交滞纳金,同时缴付所欠的维持年费。而美国发明专利申请时间为 2.5—3.5 年不等,因此本文使用“申请日起第 8 年是否维持”来评价企业的专利质量,用以下几个指标来表征:

1)维持 8 年以上的专利数量,即以相同年为单位统计的从申请日起维持 8 年以上的发明专利的数量。

2)维持 8 年以上的专利比例,以相同年为单位统计的从申请日起维持 8 年以上的发明专利的数量/以相同年为单位统计的发明专利授权量。

1.4 专利范围指标

专利范围在此仅指专利获得保护的地域范围和权利范围,分别用专利族的规模和权利要求数量来衡量。

专利族^[6]狭义上是指同一发明在不同国家提交专利申请或获得专利授权的集合。就一个发明向多国申请专利,不仅需要花费更多的资金,还表明该发明具有较强的国际竞争力;如果最终被多国授权,说明该发明的可专利性十分稳定,经得起考验,技术含量较高。因此,专利族的规模可以反映发明专利的质量。

权利要求书是确定专利保护范围的法律文书,明确记载了受法律保护的技术特征。权利要求数量^[6]可以在一定程度上反映专利保护的“范围”或“宽度”,权利要求数量越多,权利保护范围越大,说明专利的原创性越高,专利质量越高。目前,权利要求数量已成为常用的专利质量指标。

在评价企业的专利价值时,主要用到以下几个指标:

1)存在专利族的专利数量,即以相同年为单位统计的存在专利族的发明专利数量;

2)存在专利族的专利比例,即以相同年为单位统计的存在专利族的发明专利数量/以相同年为单位统计的发明专利申请量;

3)平均专利族数量,即以相同年为单位统计的发明专利的专利族总数量/以相同年为单位统计的存在专利族的发明专利数量。

4)平均权利要求数量,即以相同年为单位统计的发明专利的权利要求总数量/以相同年为单位统计的发明专利授权量。

1.5 专利引证指标^[7]

专利被引次数指观测专利被后续专利引用的次数,目前已成为认可程度最高、应用范围最广的专利引证指标。被引次数反映专利质量信息的原理主要体现在两方面:一是对后续专利技术创新的参考价值,被引次数越高,说明专利技术可被参考的价值越大,即基础专利;二是对后续专利的权利限制,被引次数越高,说明专利的限制作用越大,即核心专利。本文使用专利平均被引次数作为专利评价指标:专利平均被引次数,即以相同年为单位统计的发明专利的总被引次数/以相同年为单位统计的发明专利申请量。

2 主成分分析方法

主成分分析方法作为统计学中一种对于多元问题的数据处理方法,已在综合指标评价中得到广泛的应用。

2.1 主成分分析的基本思想

面对收集到的大量资料,人们希望把原始指标组合成较少的综合指标,仍能较好地反映样本各指标的变化。主成分分析是设法将原来众多具有一定相关性的指标(如 p 个指标)线性组合成一组新的互相无关的综合指标来代替原来的指标。最经典的做法就是用 F_1 的方差来表示,一般用 F_1 表示第一主成分, F_1 方差最大,包含的信息最多。如果 F_1 不足以表达原来 p 个指标的信息, F_1 已有的信息就不必再出现在 F_2 中。依次类推可以构造出第三、第四…、第 m 个主成分。

主成分模型如下:

$$\begin{cases} F_1 = u_{11}X_1 + u_{21}X_2 + \cdots + u_{p1}X_p \\ \cdots \\ F_m = u_{1m}X_1 + u_{2m}X_2 + \cdots + u_{pm}X_p \end{cases} \quad (1)$$

其中 $u_{1i}, \cdots, u_{pi}$ ($i = 1, \cdots, m$) 为 X 的协方差阵 R 的特征值所对应的特征向量, $X_1, X_2, \cdots, X_p$ 是原始变量经过标准化处理的值向量。

2.2 主成分分析的计算步骤

1)对样本数据进行标准化处理。

2)计算原始指标数据相关系数矩阵 R 。 R 可以发现各测评指标间的相关状况,从而能够看出指标间的信息重叠程度。

3)计算 R 的特征值 λ 、方差贡献率、累积方差贡献率 Q 和因子载荷矩阵 A , A 是反映主成分经济意义的重要指标,设第 i 个变量(因子)对第 j 个主成分的载荷量为 a_{ij} ,则 $a_{ij} = u_{ij} \sqrt{\lambda_j} (i = 1, 2, \dots, p; j = 1, 2, \dots, m)$ 。

4)选择 m 个主成分,计算相应的单位特征向量 $(u_1, u_2, \dots, u_m)$ 。对所选择主成分的经济意义给予恰当的解释。确定 m 有两种方法:① 选取特征值 $\lambda > 1$ 的主成分;② 用累积方差贡献率 Q 确定,一般 $Q \geq 80\%$, $1 \leq m \leq 3$ 。在各种数据确定的情况下,若 m/p

的值越小,则该资料使用主成分分析方法就越合适。

3 运用主成分分析方法评价

在确定对某一年的专利情况进行评价时,申请类指标可以取当年的申请量数据。由于本文选取的专利维持指标是自申请日起维持 8 年以上的专利数量和比例,因此采用 2001 至 2005 年美国发明专利申请量排名前 20 的企业数据。

第一,将表 1 中的原样本数据用 SPSS 软件标准化,消除量纲的影响。数据表中的 10 项指标申请量、授权量、授权率、存在专利族的专利数量、存在专利族的专利比例、平均专利族数量、维持 8 年以上的专利数量、维持 8 年以上的专利比例、专利平均被引次数、平均权利要求数量分别用 $X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}$ 来表示。所得出相关指标数据如表 1 所示。

表 1 2001 至 2005 年美国发明专利申请量排名前 20 的企业指标数据

企业	X_1	X_2	$X_3(\%)$	X_4	$X_5(\%)$	X_6	X_7	$X_8(\%)$	X_9	X_{10}
IBM	24 335	19 359	79.55	18 472	75.91	2.36	17 185	88.77	10.35	14.15
三星	21 625	17 384	80.39	16 202	74.92	4.51	16 481	94.81	7.22	15.00
日立	15 044	13 596	90.37	12 504	83.12	5.40	12 537	92.21	8.51	9.43
松下	14 960	12 876	86.07	11 538	77.13	5.84	12 204	94.78	7.11	10.63
佳能	12 179	11 171	91.72	8 611	70.70	3.64	11 133	99.66	8.97	10.25
惠普	11 749	11 389	96.94	7 995	68.05	2.64	11 228	98.59	9.53	18.14
索尼	11 566	10 522	90.97	9 775	84.51	5.35	9 366	89.01	7.61	12.99
东芝	11 236	8 992	80.03	8 574	76.31	4.89	8 149	90.63	8.09	9.94
富士通	10 751	8 516	79.21	8 303	77.23	3.79	8 059	94.63	7.04	10.44
英特尔	10 453	9 803	93.78	7 266	69.51	3.27	9 445	96.35	9.94	17.17
西门子	9 314	8 175	87.77	6 795	72.95	4.52	7 705	94.25	5.62	12.81
三菱	8 965	7 632	85.13	6 559	73.16	5.41	7 002	91.75	6.25	7.26
镁光科技	8 863	8 566	96.65	7 627	86.05	5.21	8 540	99.70	9.43	17.50
飞利浦	8 817	5 522	62.63	7 633	86.57	6.47	4 459	80.75	6.24	10.26
NEC	8 427	6 504	77.18	6 326	75.07	3.97	6 174	94.93	6.95	12.21
富士胶片	7 307	5 651	77.34	4 665	63.84	3.45	5 463	96.67	4.75	10.12
精工爱普生	7 094	6 182	87.14	5 261	74.16	5.66	6 087	98.46	8.06	11.46
通用电气	7 061	6 163	87.28	5 040	71.38	3.88	5 762	93.49	7.80	15.06
英飞凌科技	5 603	5 248	93.66	3 641	64.98	3.54	5 076	96.72	5.58	12.36
诺基亚	5 403	4 214	77.99	4 494	83.18	5.34	4 107	97.46	11.95	22.81

第二,计算相关系数矩阵 R 的特征值 λ 、方差贡献率 Q 、累计方差贡献率 TQ ,如表 2 所示。

表 2 R 的特征值、方差贡献率和累积方差贡献率

主成分	特征值 λ	贡献率(%) Q	累计贡献率(%) TQ
1	4.117	41.174	41.174
2	2.589	25.893	67.066
3	1.756	17.559	84.625
4	0.853	8.526	93.151
5	0.319	3.195	96.346
6	0.237	2.373	98.719

主成分	特征值 λ	贡献率(%) Q	累计贡献率(%) TQ
7	0.122	1.222	99.941
8	0.005	0.053	99.993
9	0.000	0.005	99.998
10	0.000	0.002	100.000

第三,确定主成分的个数,根据主成分的因子载荷矩阵计算相应单位特征向量。

从表 2 可以看出,前 3 个主成分的累积贡献率为 84.625%,已经大于 80%,而且也符合主成分特征值大于 1 的原则。所以,取前三个主成分即可,即 $m =$

3,并用 F_1, F_2, F_3 来表示。前三个主成分的因子载荷矩阵 A 见表 3。

表 3 因子载荷矩阵

Z	主成分		
	F_1	F_2	F_3
Z (X ₁)	0.982	-0.116	-0.112
Z (X ₂)	0.991	0.026	-0.096
Z (X ₃)	0.087	0.732	0.010
Z (X ₄)	0.974	-0.214	-0.013
Z (X ₅)	0.158	-0.580	0.733
Z (X ₆)	-0.267	-0.674	0.407
Z (X ₇)	0.979	0.104	-0.099
Z (X ₈)	-0.158	0.848	0.010
Z (X ₉)	0.365	0.406	0.745
Z (X ₁₀)	0.053	0.555	0.683

因子 $Z(X_1) \sim Z(X_{10})$ 是原样本因子 $X_1 \sim X_{10}$ 数据标准化后的对应值。从因子载荷矩阵 A 中可以看出, F_1 中第 1、2、4、7 项的载荷量比较大,说明 F_1 是反应申请量、授权量、存在专利族的专利数量、维持 7 年以上的专利数量这 4 个“量”的指标状况的综合指标; F_2 中第 3、8 项的载荷量比较大,即授权率和维持 7 年以上的专利比例这两个指标比较显著,说明 F_2 是从“率”的状况反应专利授权和维持专利权有效性的综合指标; F_3 中第 5、6、9、10 项的载荷量比较显著,说明 F_3 是反应存在专利族的专利比例、平均专利族数量、专利平均被引次数、平均权利要求数量 4 项“范围”类指标的综合指标。总体上看 F_1, F_2 和 F_3 对原有指标信息做了比较完善的综合。

根据因子载荷矩阵 A , 及 $a_{ij} = u_{ij} \sqrt{\lambda_i} (i = 1, 2, \dots, p; j = 1, 2, \dots, m)$, 求得主成分得分系数矩阵(主成分的单位特征向量矩阵)如表 4 所示:

表 4 主成分得分系数矩阵

Z	主成分		
	u_1	u_2	u_3
Z (X ₁)	0.484	-0.072	-0.085
Z (X ₂)	0.488	0.016	-0.072
Z (X ₃)	0.043	0.455	0.008
Z (X ₄)	0.480	-0.133	-0.010
Z (X ₅)	0.078	-0.360	0.553
Z (X ₆)	-0.132	-0.419	0.307
Z (X ₇)	0.482	0.065	-0.075
Z (X ₈)	-0.078	0.527	0.008
Z (X ₉)	0.180	0.252	0.562
Z (X ₁₀)	0.026	0.345	0.515

根据表 4 可得主成分 F_1, F_2 和 F_3 的表达式如下:

$$F_1 = 0.484 Z(X_1) + 0.488 Z(X_2) + 0.043 Z(X_3) + 0.480 Z(X_4) + 0.078 Z(X_5) - 0.132 Z(X_6) + 0.482 Z(X_7) - 0.078 Z(X_8) + 0.180 Z(X_9) + 0.026 Z(X_{10});$$

$$F_2 = -0.072 Z(X_1) + 0.016 Z(X_2) + 0.455 Z(X_3) - 0.133 Z(X_4) - 0.360 Z(X_5) - 0.419 Z(X_6) + 0.065 Z(X_7) + 0.527 Z(X_8) + 0.252 Z(X_9) + 0.345 Z(X_{10});$$

$$F_3 = -0.085 Z(X_1) - 0.072 Z(X_2) + 0.008 Z(X_3) - 0.010 Z(X_4) + 0.553 Z(X_5) + 0.307 Z(X_6) - 0.075 Z(X_7) + 0.008 Z(X_8) + 0.562 Z(X_9) + 0.515 Z(X_{10})$$

由 F_1, F_2 和 F_3 及其方差贡献率 Q_1, Q_2, Q_3 构造综合评价函数 $F = 0.41174F_1 + 0.25893F_2 + 0.17559F_3$, 得到 20 个企业的综合评价值 F , 按 F 由大到小排序见表 5。

表 5 2001 至 2005 年美国发明专利申请量排名前 20 的企业专利实力排名

企业	F_1	F_2	F_3	F	排名
IBM	5.457	-0.086	-0.214	2.187	1
三星	3.903	-0.339	-0.455	1.440	2
惠普	0.895	2.997	0.044	1.153	3
英特尔	0.252	2.226	0.413	0.753	4
日立	1.995	-1.072	0.382	0.611	5
镁光科技	-0.251	1.101	2.287	0.583	6
佳能	0.735	1.485	-0.733	0.558	7
松下	1.415	-0.900	-0.246	0.306	8
诺基亚	-1.907	0.844	3.814	0.103	9
索尼	0.539	-1.187	0.878	0.069	10
富士通	-0.274	-0.477	-0.607	-0.343	11
东芝	-0.088	-1.170	-0.152	-0.366	12
通用电气	-1.543	0.765	-0.029	-0.442	13
西门子	-0.904	-0.007	-0.831	-0.520	14

续表 5

精工爱普生	-1.729	0.226	0.271	-0.606	15
NEC	-1.300	-0.288	-0.404	-0.681	16
英飞凌科技	-2.362	1.458	-1.651	-0.885	17
三菱	-1.173	-1.211	-1.118	-0.993	18
富士胶片	-2.150	0.287	-2.401	-1.233	19
飞利浦	-1.509	-4.651	0.751	-1.694	20

通过表 5 中各主成分的得分情况,可以得到影响各企业专利实力的因素,例如三星的第一主成分得分比较高,说明专利的申请量、授权量和存在专利族的专利数量都比较高,第二和第三主成分相对较低,说明在授权率、专利维持及专利被引用等方面的指标比较低。结合表 1 中各企业的专利申请量排名可以看出,一个企业的专利实力不仅在于专利申请量的多少,后期的专利授权、专利维持、专利族的大小及专利被引用情况也体现了该企业专利的价值。例如诺基亚公司按申请量排名最末,但专利授权率比较高,并且注重专利有效性的维持,因此专利的综合实力排名就比较靠前。由此可见,本文建立的专利实力评价体系是比较科学和合理的。

4 结果分析

本文建立了基于申请量、授权量、授权率、存在专利族的专利数量、存在专利族的专利比例、平均专利族数量、维持 8 年以上的专利数量、维持 8 年以上的专利比例、专利平均被引次数、平均权利要求数量 10 项指标的专利实力评价体系,并利用主成分分析的方法对 2001 至 2005 年美国发明专利申请量排名前 20 的企业进行专利综合实力排名,与单纯申请量评价方

式相比,本文中的指标体系和评价方法是一种进步。本文中的指标体系不可避免的存在不够全面和需要深入研究的地方,随着专利评价指标体系的不断完善,评价结果会更加科学和准确,将会更加有效地促进企业专利水平的提高。

参考文献

- [1] 黄庆,曹津燕,瞿卫军,刘洋,石昱,肖云鹏. 专利评价指标体系(一)——专利评价指标体系的设计和构建[J]. 知识产权, 2004,14(5):25-28.
- [2] 唐炜. 面向战略决策服务的专利分析指标研究[D]. 北京:中国科学院研究生院(文献情报中心),2006.
- [3] 于晶晶,谭思明. 专利组合分析评价指标体系的构建[J]. 现代情报,2009,29(12):152-155.
- [4] 于秀林,任雪松. 多元统计分析[M]. 北京:中国统计出版社,1999.
- [5] 朱雪忠,乔永忠,万小丽. 基于维持时间的发明专利质量实证研究——以中国国家知识产权局 1994 年授权的发明专利为例[J]. 管理世界,2009(1):174-175.
- [6] 万小丽,朱雪忠. 国际视野下专利质量指标研究的现状与趋势[J]. 情报杂志,2009,28(7):49-54.
- [7] 马天旗,刘欢. 利用专利引证信息评价专利质量的改进研究[J]. 中国发明与专利,2013(1):58-61.

Evaluation of Enterprises Patent Strength Based on Principal Component Analysis

QI Ping, LING Feng, LIU Hui-jing

(Institute of Scientific and Technical Information of China, Beijing 100038, China)

Abstract: Enterprises patent strength reflects the technological innovation capability, market competitiveness and development dynamic. The evaluation of enterprises patent comprehensive strength enables companies to grasp the trend of technological development, understand the strengths and weaknesses of technical innovation, and adjust innovation strategy scientifically. This paper regards invention patent as research object and selects representative patent evaluation indexes, to build an indicator system for evaluating enterprises patent quality based on application, authorization, patent maintenance, patent scope and patent citations. At last, the paper uses principal component analysis to evaluate comprehensive patent strength of the companies whose U. S. patent application number ranked in the top 20.

Key words: patent; strength; index system; patent evaluation; principal component analysis