

基于改进灰关联分析的专利影响因素评估模型

林维勇, 张小平, 房俊华

(贵州大学 计算机科学与信息学院, 贵州 贵阳 550025)

摘要: 采用改进的灰关联分析方法, 构建了专利发展影响因素评估模型。研究并发现对贵州省专利发展影响较大的因素, 为相关部门或人员决策提供了科学的参考依据。

关键词: 灰关联分析; 专利; 灰系统理论; 灰关联系数; 影响因素

中图分类号: TP399

文献标识码: A

文章编号: 1674-7720(2012)09-0066-03

Factors assessment model affecting patent based on improved grey relational analysis

Lin Weiyong, Zhang Xiaoping, Fang Junhua

(College of Computer Science and Information, Guizhou University, Guiyang 550025, China)

Abstract: This paper researched the factors affecting the development of patent was necessary, especially the technology and economy was relatively backward in western regions, such as Guizhou Province, used an improved method of gray relational analysis to build the model of the factors affecting the patent development, researched and discovered the greater impact factors of development of patent in Guizhou Province, to provide a scientific frame of reference for related departments or personnel decisions in Guizhou Province.

Key words: grey relational analysis; patent; grey system theory; gray relational coefficient; factors

专利是受法律规范保护的发明创造, 它是指某项发明创造向国家审批机关提出专利申请, 经依法审查合格后向专利申请人授予的在规定的时间内对该项发明创造享有的专有权^[1]。专利能激励并保障人们的创新和发明, 是推动经济增长的一个重要因素。提高专利水平, 完善有关专利的知识产权机制是促进经济增长的一个重要途径^[2]。专利具有重要的价值, 是国家或地区宝贵的科技资产, 专利的拥有量在一定程度上反映了国家或地区的科技实力水平。因此, 注重专利的发展情况, 挖掘并分析影响专利产出的各种相关因素, 为政府制定科技和经济政策提供理论依据, 对促进国家或地区的科技创新和经济发展具有重要的意义。2010年贵州省的专利申请量为4336件, 专利授权量为3064件。而国家的专利申请量为122.2万件, 专利授权量为81.5万件^[3-4]。贵州省的专利申请量和专利授权量分别仅占全国的0.354%和0.375%。可见, 贵州省的专利发展水平严重滞后, 在一定程度上制约了国家或地区的科技和经济发展。因此, 加强对贵州省专利发展方面的研究势在必行。本文

运用灰系统理论的灰关联分析法^[5], 以专利申请量为产出指标, 对专利申请量及其可能影响因素进行研究。通过采用适当的无量纲化处理方法和合理的分辨系数动态取值方法^[7-8], 改进了以往通常使用的灰关联分析法, 计算出更准确的影响专利发展因素的灰关联度, 挖掘出影响较大的因素, 为贵州省相关部门或人员制定相关政策提供参考依据, 从而促进贵州省科技和经济的进一步发展。

1 指标选择及数据来源

1.1 指标选择

专利的考核指标有专利申请量和专利授权量, 由于专利申请量和专利授权量具有信息相关性及专利授权量的滞后性, 本文选取贵州省专利申请量作为贵州省专利发展的考核指标, 即在进行灰关联分析时以专利申请量作为系统特征行为序列。影响专利发展的因素很多, 根据以往国家和地方统计年鉴资料和经验数据, 本文选取了影响可能性较大的因素指标。分别为贵州省生产总值GDP(Gross Domestic Product)、贵州省人均生产总值、

技术与方法 Technique and Method

贵州省财政支出、贵州省科技经费支出、贵州省 R&D (Research and Development) 经费支出、贵州省 R&D 经费投入强度 (R&D 经费支出占 GDP 的比重) 和签订技术合同数, 共 7 个指标。由于生产总值决定了科技投入的总体规模, 生产总值既是产出指标, 又可作为投入指标。显而易见, 其他指标都是投入指标。

1.2 数据来源

本文研究的数据来源于《2010 贵州统计年鉴》、《2005~2010 年贵州省国民经济和社会发展统计公报》、《2005~2010 年全国科技经费投入统计公报》。通过对以上资料包含的相关数据进行统计、整理和分析, 得到相关指标数据统计表, 见表 1。

表 1 2005~2010 年贵州省专利申请量及相关指标数据统计表

指标	2005	2006	2007	2008	2009	2010
全省生产总值/亿元	1942	2338.98	2884.11	3561.56	3912.68	4593.97
人均生产总值/元	4957	5932	7273	9428	10309	13221
财政支出/亿元	520.49	610.64	795.4	1055.39	1372.27	1640.17
科技经费支出/亿元	3.04	3.22	9.84	12.93	14.07	16.56
R&D 经费支出/万元	110000	145000	137000	189000	264134	300000
R&D 经费投入强度/%	0.57	0.64	0.5	0.57	0.68	0.65
签订技术合同/项	466	263	457	579	989	650
专利申请量/件	2226	2674	2759	2943	3709	4336

2 专利发展影响因素的灰关联分析

2.1 灰关联分析

灰关联分析就是对系统的因素分析, 是对系统发展变化态势的定量比较和反映, 通过灰色关联度分析和确定系统因素间的影响程度或因素对系统主行为的贡献测度的一种方法。灰关联分析的基本思想是根据序列曲线几何形状的相似程度来判断其关联程度。若曲线越接近, 则相应序列之间关联度越大, 反之就越小^[9]。

2.2 建立灰关联分析评估模型

2.2.1 确定分析序列

贵州省专利申请量为参考序列 (又称母序列), 定义为 $y = \{y(k) | k=1, 2, \dots, m\}$ 。其中, y 表示专利申请量; k 表示年份, 2005 年~2010 年按顺序替换为 1~6, 根据表 1, $m=6$; $y(k)$ 表示第 k 年专利申请量。

贵州省生产总值、人均生产总值、财政支出、科技经费支出、R&D 经费支出、R&D 经费投入强度和签订技术合同数为比较序列 (又称子序列), 定义为 $x_i = \{x_i(k) | k=1, 2, \dots, m\} i=1, 2, \dots, n$ 。其中, x_i 表示影响专利发展的第 i 个指标, 根据表 1, $n=7$; k 表示年份, 根据表 1, $m=6$; $x_i(k)$ 表示影响专利发展的第 i 个指标在第 k 年的指标值。

2.2.2 指标值的无量纲化处理

由于各个指标计量单位和数量级存在差异, 使各指标间不具有综合性, 不能直接进行综合分析。因此必须

采用某种方法对各指标数值进行无量纲化处理, 解决各指标数值不可综合性问题^[10-12]。

目前常见的无量纲化处理方法主要有极值化、标准化、均值化以及标准差化方法。本文采用均值化方法。

均值化方法的基本原理是每一变量值除以该变量的平均值。标准化后各变量的平均值都为 1, 标准差为原始变量的变异系数。该方法在消除量纲和数量级影响的同时, 保留了各变量取值差异程度上的信息, 差异程度越大的变量对综合分析的影响也越大。该无量纲化方法在保留原始变量变异程度信息时, 并不是仅取决于原始变量标准差, 而是原始变量的变异系数, 保证了保留变量变异程度信息的同时数据的可比性^[6]。采用均值化方法对以上各指标量进行无量纲化处理, 计算公式为:

$$X_i(k) = \frac{x_i(k)}{X_i}; Y(k) = \frac{y(k)}{y}$$

其中, $X_i = \frac{1}{m} (\sum_{k=1}^m x_i(k))$, $y = \frac{1}{m} (\sum_{k=1}^m y(k))$, $k=1, 2, \dots, m$; $i=1, 2, \dots, n$ 。根据表 1, $n=7, m=6$ 。 $X_i(k)$ 表示无量纲化处理后的影响专利发展的第 i 个指标在第 k 年的指标值。 $Y(k)$ 表示无量纲化处理后的第 k 年专利申请量。

2.2.3 计算关联系数

$y(k)$ 与 $x_i(k)$ 的关联系数为:

$$\theta_i(k) = \frac{\min_i \min_k |Y(k) - X_i(k)| + \rho \max_i \max_k |Y(k) - X_i(k)|}{|Y(k) - X_i(k)| + \rho \max_i \max_k |Y(k) - X_i(k)|} \quad (1)$$

设 $\Delta_i(k) = |Y(k) - X_i(k)|$; $M = \max_i \max_k \Delta_i(k)$; $N = \min_i \min_k \Delta_i(k)$ 。则式 (1) 转换为:

$$\theta_i(k) = \frac{N + \rho M}{\Delta_i(k) + \rho M} \quad (2)$$

$k=1, 2, \dots, m; i=1, 2, \dots, n$ 。其中 $\rho \in (0, +\infty)$ 为分辨系数, 是直接关联分析分辨率的一个因子, 它的取值决定着关联系数的分布状况。通常灰关联分析的相关文献中分辨系数 ρ 取值为 0.5, 则关联系数 $\theta_i(k) \in [0.3333, 1]$ 。为了提高关联系数 $\theta_i(k)$ 对 $y(k)$ 与 $x_i(k)$ 的分辨率, 使 $\theta_i(k)$ 的取值范围 (关联系数的取值区间的长度) 尽量大, 分辨系数 ρ 应取值在 0.05 附近, 则关联系数 $\theta_i(k) \in [0.0476, 1]$, 大大扩展了取值范围, 更容易观察关联度的分辨率的变化^[7]。但是分辨系数 ρ 应根据实际观测序列的具体情况动态取值。当观测序列出现奇异值时, ρ 应取较小的值, 以克服奇异值的支配作用; 当观测序列比较平稳时, ρ 应取较大的值, 充分体现关联度的整体性^[8]。因此参考文献 [8] 给出了一种新的分辨系数动态取值方法, 解决了以往分辨系数取值难以量化的问题, 使关联分析更符合实际。本文将利用该方法确定分辨系数。则式 (2) 将变换为:

$$\theta_i(k) = \frac{N + \rho(k)M}{\Delta_i(k) + \rho(k)M}; k=1, 2, \dots, m; i=1, 2, \dots, n$$

技术与方法 Technique and Method

2.2.4 计算关联度

关联度 $\gamma_i = \frac{1}{m} \sum_{k=1}^m \theta_i(k)$; $k=1, 2, \dots, m$; $i=1, 2, \dots, n$ 。其

中 γ_i 表示第 i 个因素(指标)与参考序列(专利申请量)的关联度。

2.3 灰关联分析模型的求解

根据以上建立的灰关联分析模型以及表 1 中的数据,借助 Matlab 工具编程求解得到各个专利发展影响因素(指标)与专利申请量的关联系数和关联度,见表 2。其中采用参考文献 [8] 中的分辨系数动态取值方法,并借助 Matlab 工具编程求解,计算出本文实验数据相关的分辨系数 $\rho = \{\rho(1), \rho(2), \rho(3), \rho(4), \rho(5), \rho(6)\} = \{0.6895, 0.9258, 0.2138, 0.3523, 0.4868, 0.7547\}$ 。

表 2 专利发展影响因素的关联系数和关联度

序号	指标	关联系数					关联度	
		2005	2006	2007	2008	2009		2010
1	全省生产总值/亿元	0.7752	0.7963	0.9107	0.5425	0.9092	0.9171	0.8085
2	人均生产总值/元	0.7390	0.7567	0.7777	0.5494	0.9432	0.7267	0.7488
3	财政支出/亿元	0.6607	0.6719	0.5639	0.6403	0.5986	0.6279	0.6272
4	科技经费支出/亿元	0.4806	0.4872	0.5376	0.3548	0.5479	0.6062	0.5024
5	R&D 经费支出/万元	0.7312	0.8355	0.4102	0.8190	0.5851	0.7020	0.6805
6	R&D 经费投入强度/%	0.6219	0.7151	0.6768	1.0000	0.8101	0.5691	0.7322
7	签订技术合同/项	0.7837	0.5624	0.5904	0.7262	0.3279	0.6251	0.6026

2.4 结果分析

根据表 2 的结果数据,影响专利申请量(y)的贵州省生产总值(x_1)、人均生产总值(x_2)、财政支出(x_3)、科技经费支出(x_4)、R&D 经费支出(x_5)、R&D 经费投入强度(x_6)和签订技术合同数(x_7)分别对应的关联度为: $\gamma_1=0.8085$, $\gamma_2=0.7488$, $\gamma_3=0.6722$, $\gamma_4=0.5024$, $\gamma_5=0.6805$, $\gamma_6=0.7322$, $\gamma_7=0.6026$ 。可见 $\gamma_1 > \gamma_2 > \gamma_6 > \gamma_5 > \gamma_3 > \gamma_7 > \gamma_4$ 。各灰关联度都大于 0.5, 表明以上因素对贵州省专利发展都具有较大的影响程度。以上各因素反映了贵州省的经济发展水平、政府财力以及对科技的投入水平,这些对专利的发展影响很大。其中生产总值和 R&D 经费投入强度与专利产出的关联度最大,表明贵州省的经济水平和 R&D 经费投入比重是影响科技产出最关键的因素。为提高贵州省的专利发展水平,促进当地科技发展,必须不断提高生产力水平,加大 R&D 经费投入力度。

本文通过采用合适的无量纲化处理方法和客观的分辨系数动态取值方法,改进了以往通常使用的灰关联分析方法,并应用到贵州省专利发展影响因素的研究,建立相应的灰关联分析评估模型,挖掘出对贵州省专利发展影响较大的因素,对贵州省的科技决策具有理论参

考价值。

参考文献

- [1] 冯晓青,刘友华.专利法[M].北京:法律出版社,2010.
- [2] 徐竹青.专利、技术创新与经济增长:理论与实证[J].科技管理研究,2004,24(5):109-111.
- [3] 贵州省统计局,国家统计局贵州调查总队.贵州统计年鉴 2010[M].北京:中国统计出版社,2010.
- [4] 中华人民共和国国家统计局.中国统计年鉴 2010[M].北京:中国统计出版社,2010.
- [5] 邓聚龙.灰理论基础[M].武汉:华中科技大学出版社,2002.
- [6] 韩胜娟.SPSS 聚类分析中数据无量纲化方法比较[J].科技广场,2008(3):229-231.
- [7] 申卯兴,薛西锋,张小水.灰色关联分析中分辨系数的选取[J].空军工程大学学报(自然科学版),2003,4(1):68-70.
- [8] 东亚斌,段志善.灰色关联度分辨系数的一种新的确定方法[J].西安建筑科技大学学报(自然科学版),2008,40(4):589-592.
- [9] 郝鹏飞,王少凡,王晓林,等.基于灰色系统理论的我国高校专利申请影响因素分析[J].知识经济,2011(9):7-8.
- [10] 郭亚军,易平涛.线性无量纲化方法的性质分析[J].统计研究,2008,25(2):93-100.
- [11] 易平涛,张丹宁,郭亚军,等.动态综合评价中的无量纲化方法[J].东北大学学报(自然科学版),2009,30(6):889-892.
- [12] 李炳军,朱春阳,周杰.原始数据无量纲化处理对灰色关联度的影响[J].河南农业大学学报,2002,36(2):199-202.

(收稿日期:2011-12-14)

作者简介:

林维勇,男,1986年生,硕士,主要研究方向:数据库技术与应用系统。

张小平,男,1956年生,研究员,主要研究方向:计算机软件。

房俊华,男,1986年生,硕士,主要研究方向:数据库技术与应用系统。