DOI: 10.16353/j.cnki.1000-7490.2015.02.006

● 吕璐成，刘　娅，杨冠灿（中国科学技术信息研究所　战略研究中心，北京　100038）

# 基于决策树方法的专利被引影响因素研究

**摘　要：** 专利引用对于专利质量评价具有重要作用。文章基于决策树方法对可能影响专利被引的12个影响因素与专利是否被引的潜在关系进行分析。研究发现，专利的最早优先权年是其中影响最为显著的因素，而后依次是权利要求数量、专利权人数量、是否转让、平均引用时滞、优先权国家，而其他6个指标的影响效果并不明显。

**关键词：** 决策树；专利引用；数据挖掘

**Abstract:** Patent citation plays an important role in patent quality evaluation. Based on the decision tree method, this paper analyzes the prospective relationship between 12 influencing factors and patent citation. The research finds that the earliest priority year of patent is the most significant factor, and the quantity of patent claim, quantity of patentees, transfer of patent, mean citation lag and priority country are the following influencing factors, and the other 6 factors are less influential.

**Keywords:** decision tree; patent citation; data mining

专利逐渐成为衡量国家、地区乃至企业发展现状和创新能力的重要指标。因此，专利质量的评价问题也成为学术界的研究热点。自20世纪80年代以来，以专利引用指标为代表的质量评价指标不断被提出和应用[1]。学者们也纷纷开始进行专利引用相关的研究。目前，专利引用的研究较多是针对专利引用行为动机的研究[2]、专利高被引影响因素的研究等[3]。

总结已有研究可以发现：目前针对专利引用影响因素的研究主要针对引用数量及高被引的研究；基于多元回归统计研究专利高被引成因需要建立在各个指标相互独立的基础上。但是，对于专利引用数量的研究忽略部分专利引用形成的潜在关系；另外，基于指标相互独立的假设进行研究是不准确的。因此，本研究对于专利被引行为进行研究，并选取了一种不需要建立在变量独立假设条件下使用的方法——决策树方法，试图发现与专利被引而并不单单是“高被引”形成的影响因素以及各影响因素的重要程度。这将对于专利发明人或专利研究人员和机构认识和利用专利引用提供更为深入的参考和借鉴。

通过文献综述，本研究提取了12个可能对专利引用产生影响的因素，见表1。

本研究基于决策树（Decision Tree）方法展开研究。决策树是一种常用于分类预测的算法，通过大量数据有目的地分类，从中找到一些具有商业价值的、潜在的信息[10]，它在分析用户购买行为的潜在决定因素中已有较好的运用。本研究借鉴这种思想研究专利是否被引事件的潜在规则信息，从而探究可能对专利被引产生影响的因素的重要程度。

表1　专利被引影响因素

| 影响因素 | 简介 | 可能对专利被引事件产生的影响 |
|---|---|---|
| 引用专利 | 指观测专利所引用的专利文献 | 专利引用量对于专利质量的影响因素没有定论。一些学者认为专利参考文献是专利引文的主体部分，在很大程度上影响观测专利的新颖性和创造性，限制观测专利的权利范围，所以专利参考文献数量与专利质量负相关；也有部分学者认为一个技术领域发展越成熟，参与的发明人就越多，因而大量引用专利文献的专利越有可能包含高质量技术[4]；还有研究发现专利引用量跟专利质量之间并无关系[1] |
| 引用NPL（非专利文献） | 指观测专利所引用的除专利文献以外的其他文献 | 有学者以新西兰企业1976—2004年在美国授权的850件专利为样本，通过回归分析发现非专利参考文献数量与被引次数负相关[5] |
| 优先权国家 | 优先权国家是指专利首次提出申请所在的国家 | 研究证明，专利引用行为与地理因素存在关系[6] |
| 专利权人 | 专利权人是专利权的所有人及持有人的统称 | 专利的专利权人数量可以在一定程度上反映专利的合作情况，合作产生的专利可能投入更多的资金、人力，应该具有更高的价值，也可能更容易被引用 |
| 发明人 | 发明人指为发明创造的实质性特点作出创造性贡献的人 | 理论上讲，参与专利的发明人数量越多，发明凝结的技术就越多，创新性可能更高，也较为容易被引用。目前已有较多研究实证发现发明人数量与专利技术质量高度正相关[7] |
| 专利转让 | 指专利权人作为转让方，将其发明创造专利的所有权或将持有权移转受让方 | 被转让过的专利质量往往较高，被引用的可能性更大 |

表 1　　续表

| 影响因素 | 简介 | 可能对专利被引事件产生的影响 |
|---|---|---|
| PCT（Patent Cooperation Treaty）申请 | 专利申请人可以通过 PCT 途径递交国际专利申请，向多个国家申请专利 | 专利价值较高时，专利权人才会进行 PCT 申请，以获得更多的专利保护。因此，进行过 PCT 申请的专利可能更容易被引用 |
| 技术覆盖范围（Technology Scope） | 指专利所涉及的技术领域范围，通常可以用专利的 IPC 分类号进行区分[1] | 曾有学者通过实证研究发现专利的前四位 IPC 号数量与专利被引次数成高度正相关[8]。这可能源于专利技术覆盖范围越广，表明专利越具有基础性、一般规律性，更容易被引用 |
| 平均后向引用时滞（Mean Backwards Citations Lag） | 指引用专利与被引用专利的申请或授权年的差值的平均值[9] | 平均后向引用时滞能够反映专利的技术新颖性。通常其越大，可能表明专利是对较早技术的改进或颠覆，具有较高的价值，也更容易被引用 |
| 权利要求 | 指专利或专利申请中除说明书部分之外，由一系列名词词组组成的部分 | 一般而言，权利要求数量越多，专利的保护范围越大，说明专利的原创性越高，专利质量越高。韩国学者曾运用“零膨胀模型”证明权利要求数量和被引次数高度正相关[10] |
| 同族专利 | 指同一专利权的不同形式组合，既包括同一专利权在不同国家的布局，也包括该专利后续的延伸发明 | 同族专利数量便是指该专利所属专利家族包含的专利数量。同族专利数量越大，专利的价值也应当越高，也更可能被引用 |
| 专利优先权年 | 指专利申请过程中优先权的年份 | 随着时间的推移，优先权年越早的专利更可能被引用 |

注：平均后向引用时滞的计算公式为：平均后向引用时滞 = Σ（施引专利申请时间 - 被引参考文献公开时间）/引用数量。

## 1　数据来源与指标选取

### 1.1　数据来源

本研究选取光盘技术领域的美国专利数据作为研究对象进行实证研究。之所以选取光盘技术作为研究对象，是因为光盘技术在近 50 年的发展历程中，经历了四代的技术更迭和变更过程，其中诞生了 CD（1982）、DVD（1995）、UDO（2003）、UMD（2004）、Hi-MD（2004）、BD（2006）、HD DVD（2006）以及代表未来趋势的 Holographic Versatile Disc、LS-R、Protein-Coated disc 技术[11]等。因此，采用光盘技术领域的专利数据作为研究对象可以帮助我们理解在技术变革背景下专利被引各影响因素的作用。另外，由于美国专利数据的引用信息更为规范化，因此，选取美国的专利数据进行研究。

本研究数据来源于美国专利商标局（USPTO）数据库。根据美国专利分类号，光盘技术当前被分类在 US-PC720 主分类号中。据此检索获得专利数据 4402 条，通过去重等清洗操作之后，获得标准化数据共 4388 条。

### 1.2　指标选取

根据可能对专利被引产生影响的因素列表，拟定决策树的分析指标共 12 个，分别是：权利要求数量、专利引用量、NPL 引用量、同族专利数量、最早优先权年、专利权人数量、发明人计数、IPC 广度、PCT 申请、是否转让、平均引用时滞以及优先权国家，对其中前 11 个指标分别进行描述性统计分析，见表 2，而对优先权国家统计发现优先权国为“US”的专利数量 731 条，“非 US”的 3657 条。

表 2　分析指标描述性统计分析

| | 权利要求数量 | 专利引用量 | NPL 引用量 | 同族专利数量 | 最早优先权年 | 专利权人数量 | 发明人计数 | IPC 广度 | PCT 申请 | 是否转让 | 平均引用时滞 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 平均 | 12.05 | 14.78 | 0.71 | 6.83 | 1999.52 | 1.13 | 2.46 | 3.05 | 0.09 | 0.83 | 10.86 |
| 标准误差 | 0.14 | 0.27 | 0.04 | 0.14 | 0.09 | 0.01 | 0.03 | 0.04 | 0 | 0.01 | 0.06 |
| 中位数 | 10 | 11 | 0 | 5 | 2000 | 1 | 2 | 2 | 0 | 1 | 10.4 |
| 众数 | 6 | 8 | 0 | 4 | 2003 | 1 | 1 | 1 | 0 | 1 | 8 |
| 方差 | 85.85 | 312.43 | 7.64 | 91.97 | 32.39 | 0.41 | 2.87 | 6.86 | 0.08 | 0.14 | 14.33 |
| 最小值 | 1 | 0 | 0 | 0 | 1983 | 1 | 1 | 1 | 0 | 0 | 0 |
| 最大值 | 110 | 581 | 97 | 171 | 2012 | 15 | 15 | 31 | 1 | 1 | 55.66667 |
| 求和 | 52888 | 64851 | 3097 | 29955 | 8773875 | 4937 | 10790 | 13396 | 406 | 3656 | 47640.68 |

## 2　SQL Server 2012 BI 平台及决策树方法简介

### 2.1　SQL Server 2012 BI 平台

SQL Server 2012 Business Intelligence（SQL Server 2012 BI）平台是微软最新的商业智能解决方案和数据服务平台，在原有的 SSIS（SQL Server Integration Service）、SSAS（SQL Server Analysis Service）、SSRS（SQL Server Report Service）基础上，新增了 BI 语义模型（Business Intelligence Semantic Model，BISM）、字段存储索引、数据质量服务、强力视图、语义搜索等特性[13]，并且提供了一系列大数据解决方案，如 Apache Hadoop 连接器、开源分布式计算架构、能够存储并处理结构化和非结构化数据等[14]。借助 SQL Server 2012 BI 平台，情报分析人员可以较为方便快捷地分析和解决一系列问题。本研究借助 SSAS 中集成的 Microsoft 决策树算法，对专利被引影响因素进行研究。

### 2.2　决策树方法

决策树是一种快速而直观的分类技术，也是目前最受欢迎的数据挖掘分析技术之一，主要用于分类和预测。决策树是一棵有向无环树，在外观上很类似于流程图。树中的任一个非叶节点对应着数据集中某个属性，叶节点则对应着分类结果，树中每个分支对应其所连接属性的节点所对应属性的某个数值[15]。每一条从根节点到叶节点的路径就是目标变量的一条规则[16]，其中越靠近根节点的属性对于分类结果的影响越重要[17]。借助这种思想，可以对专利被引影响因素的重要程度进行研究，发现形成专利被引的影响因素规则及影响因素中较为重要的因素。

## 3 数据的清洗与转换

因直接从数据库中得到的专利数据不规范，所以需要对其进行清洗与转换。本研究对于数据的预处理采用的工具是 Microsoft Excel VBA，从原始数据中逐个提取研究需要用到的 12 个指标。以平均引用时滞的处理为例，首先对原始数据“引用专利”中的引用时间进行提取，核心代码如下:

```
For i =2 To 4388
    sWord = 专利引用信息
    k1 = InStr( 1 ,sWord," ( Examiner) ")
    k2 = InStr( 1 ,sWord," ( Applicant) ")
    While k1 < >0 Or k2 < >0
        Print 引用时间
    Wend
Next
```

如此便可得到每一条专利所引用其他文献的时间，然后再利用专利平均后向引用时滞的计算公式即可计算得到每一条专利的平均后向引用时滞。

将原始数据进行预处理之后，即可导入到 Microsoft SQL Server Management Studio 中。此时便可对某些仍旧不标准的数据进行转换，在这里将“优先权国家”不是“US”的均更改为“非 US”。

update cleanUS set 优先权国家 =´非 US´ from Test_OLAP_DecisionTree cleanUS where 优先权国家 < >´US´;

最后，便可得到待处理的数据源表 Test_OLAP_DecisionTree，见图 1。

| | 公开号 | 权利要求数量 | 专利引用量 | 前向引用量 | 同族专利数量 | 优先权国 | 最早优先权年 | 专利权人数量 | 发明人计数 | IPC广度 | PCT申请 | 是否转让 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | US8528011132 | 2 | 15 | 3 | 4 | 非US | 2010 | 3 | 2 | 1 | 0 | 1 |
| 2 | US8528010082 | 5 | 4 | 0 | 3 | 非US | 2011 | 3 | 2 | 1 | 0 | 1 |
| 3 | US8528009982 | 10 | 7 | 0 | 5 | 非US | 2011 | 2 | 1 | 2 | 0 | 1 |
| 4 | US8522262911 | 4 | 8 | 0 | 3 | 非US | 2012 | 2 | 1 | 1 | 0 | 1 |
| 5 | US8521005321 | 13 | 2 | 4 | 2 | US | 2006 | 1 | 1 | 1 | 0 | 0 |
| 6 | US8516513821 | 2 | 5 | 2 | 2 | US | 2009 | 8 | 7 | 2 | 0 | 1 |
| 7 | US8516512921 | 9 | 2 | 1 | 3 | 非US | 2011 | 4 | 3 | 1 | 0 | 1 |
| 8 | US8516511321 | 12 | 3 | 0 | 5 | 非US | 2011 | 4 | 3 | 1 | 0 | 1 |

图 1 待分析数据源表

## 4 挖掘模型的创建

在 SQL Server Data Tools 中新建一个 Analysis Services 多维和数据挖掘项目之后，便可进行挖掘模型的构建。

首先，新建数据源与 SQL Server Management Studio 中的 PatentOLAP 数据库对接。“数据源”是一种数据连接，包含源数据所在的服务器和数据库的名称。

其次，创建数据源视图，与数据表 Test_OLAP_DecisionTree 建立关系。数据源视图是基于数据源生成的，定义用来填充数据仓库的数据的子集。通过数据源视图，可以选择与特定项目相关的表，建立表之间的关系，并添加计算列和命名视图，而不必修改原始的数据源。

最后，创建挖掘结构 Test OLAP_DM。首先设定使用的数据挖掘算法为“Microsoft 决策树算法”，然后选择 key“专利公开号”、用于分析的 12 个列和用于预测的列“是否是被引专利”，最后将本研究用于结果准确性测试的数据集实例数设定为 800 个。这样就成功创建了挖掘模型。

## 5 挖掘模型的准确性评估

### 5.1 分类矩阵

本研究预留 800 条专利作为测试集对生成的挖掘模型进行测试并将测试结果以分类矩阵的形式进行展示，见表 3。

表 3 决策树模型的分类矩阵

| | 未被引用 | 被引用 |
|---|---|---|
| 未被引用 | 56 | 28 |
| 被引用 | 149 | 567 |

从表 3 中可以看到，预测未被引用为真的专利数有 56 条，预测未被引用为假的专利数有 28 条，预测被引用为假的专利数有 149 条，预测被引用为真的专利数有 567 条。

准确率为: (56 +567) /(56 +28 +149 +567) =78%。这表明模型较为可靠。

### 5.2 挖掘提升图

提升图是按照测试数据集中可预测列的已知值来绘制从该测试数据集进行预测查询的结果，其不仅显示挖掘模型的结果，而且还给出理想模型和随机模型的显示结果。针对随机模型线所做的任何改进均称为“提升”。提升图中挖掘模型所演示的提升量越多，模型也就越有效。

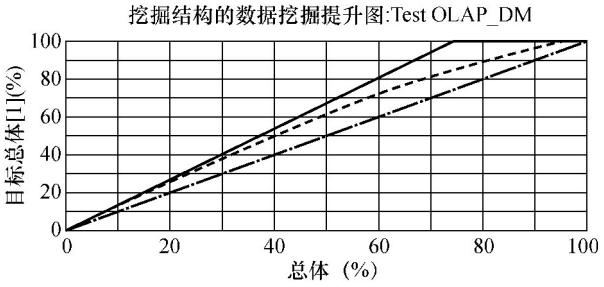


图 2 模型在测试集上是否被引 = “1”条件下的提升图

利用软件绘制模型在测试集上是否被引为“1”条件下的提升图来检验模型准确率，见图 2。其中 $X$ 轴表示从测试集中抽取的事例数所占测试集总体事例数的百分比，$Y$ 轴表示从测试集中抽取的事例数中满足可预测列是否被引 = “1”条件的专利数所占测试集总体专利数中满足可预测列是否被引 = “1”条件的百分比。图中蓝色的线

（计算机显示颜色，图中为最下面的线）表示随机预测模型，绿色的线（中间的线）表示理想预测模型，可以发现生成的模型对应的红色提升曲线（最上面的线）轨迹较之随机预测曲线有较大的提升，并且比较接近理想模型下对应的绿色提升曲线轨迹。结合它们分别对应的挖掘图例，见图3，可以看出该决策树模型具有较高的预测准确率。

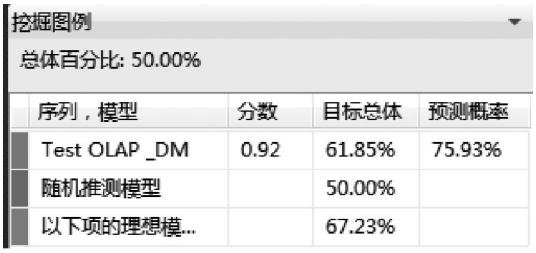


图3　挖掘图例

## 6　分析结果与讨论

挖掘模型生成的决策树分析结果见图4，抽取其中重要的规则，见表4。

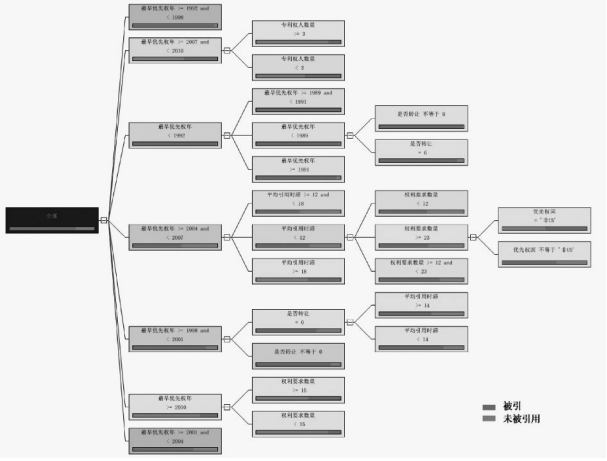


图4　决策树分析结果

规则1～规则7表明专利优先权年越早的专利更容易被引用。1992年之前的专利被引率达到98.5%，而2010—2013年的专利被引率仅仅为20.51%。根据专利被引率绘制被引率趋势图，见图5，可以发现，专利引用率逐步下降，这进一步证明了时间因素对于专利被引行为的影响作用。

从规则8～规则9可以看到美国的专利被引率要高于非美国的专利被引率，这表明美国专利较之非美国专利更容易被引用。从规则10～规则11可以看到在2010—2013年，权利要求数量小于15项的专利被引率仅仅为14.04%，而权利要求数量大于15项的专利被引率达到

表4　专利被引影响因素决策树模型抽取的部分规则

| 序号 | 规则内容 |
|---|---|
| 1 | 最早优先权年 < 1992 = > 是被引专利（404/398）置信度98.5% |
| 2 | 最早优先权年 > = 1992 and < 1998 = > 是被引专利（819/782）置信度95.48% |
| 3 | 最早优先权年 > = 1998 and < 2001 = > 是被引专利（559/484）置信度86.58% |
| 4 | 最早优先权年 > = 2001 and < 2004 = > 是被引专利（848/603）置信度71.11% |
| 5 | 最早优先权年 > = 2004 and < 2007 = > 是被引专利（611/350）置信度57.28% |
| 6 | 最早优先权年 > = 2007 and < 2010 = > 是被引专利（269/109）置信度40.52% |
| 7 | 最早优先权年 > = 2010 = > 是被引专利（78/16）置信度20.51%<br>证明专利优先权年的重要性 |
| 8 | 最早优先权年 > = 2004 and < 2007 and 平均引用时滞 < 12 and 权利要求数量 > = 23 and 优先权国 不等于 ´非 US´ = > 是被引专利（12/8）置信度66.67% |
| 9 | 最早优先权年 > = 2004 and < 2007 and 平均引用时滞 < 12 and 权利要求数量 > = 23 and 优先权国 = ´非 US´ = > 是被引专利（10/5）置信度50%<br>以上两条规则可以看出，美国本国的专利更容易被引用。 |
| 10 | 最早优先权年 > = 2010 and 权利要求数量 < 15 = > 非被引专利（57/49）置信度85.96% |
| 11 | 最早优先权年 > = 2010 and 权利要求数量 > = 15 = > 非被引专利（21/13）置信度61.9% |
| 12 | 最早优先权年 < 1989 and 是否转让 = 0 = > 是被引专利（36/34）置信度94.44% |
| 13 | 最早优先权年 < 1989 and 是否转让 不等于 0 = > 是被引专利（74/74）置信度100% |
| 14 | 最早优先权年 > = 1998 and < 2001 and 是否转让 不等于 0 = > 是被引专利（483/429）置信度88.66% |
| 15 | 最早优先权年 > = 1998 and < 2001 and 是否转让 = 0 = > 是被引专利（76/55）置信度72.36% |
| 16 | 最早优先权年 > = 2007 and < 2010 and 专利权人数量 > = 3 = > 非被引专利（24/21）置信度87.5% |
| 17 | 最早优先权年 > = 2007 and < 2010 and 专利权人数量 < 3 = > 非被引专利（245/139）置信度56.73% |
| 18 | 最早优先权年 > = 2004 and < 2007 and 平均引用时滞 < 12 = > 是被引专利（341/215）置信度63.05% |
| 19 | 最早优先权年 > = 2004 and < 2007 and 平均引用时滞 > = 12 and < 18 = > 是被引专利（239/123）置信度51.46% |
| 20 | 最早优先权年 > = 2004 and < 2007 and 平均引用时滞 > = 18 = > 是被引专利（31/12）置信度38.71% |

38.1%，这表明权利要求数量越少的专利，更不容易被应用。从规则 12 ~ 规则 15 可以发现经过转让的专利具有更高的被引率，这证明被转让的专利更容易被引用。规则 16 ~ 规则 17 证明专利的专利权人数量越多，并不会对专利被引产生促进作用。规则 18 ~ 规则 20 证明平均引用时滞越小，专利被引概率越大，而引用时滞越长的专利被引率却低得多。

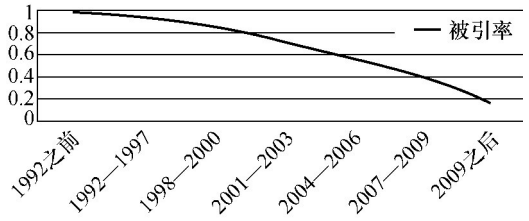


图5 优先权年的影响效果

根据决策树预测模型的依赖关系网络绘制专利被引重要影响因素示意图，见图 6。可以发现，能够对预测属性产生影响的属性由强到弱依次是：最早优先权年、权利要求数量、专利权人数量、是否转让、平均引用时滞、优先权国家。因此在本研究选取的 12 个影响专利被引的因素中，专利的最早优先权年对被引影响最为显著的因素。而引用专利数量、发明人数量等 6 个指标的影响效果并不显著。

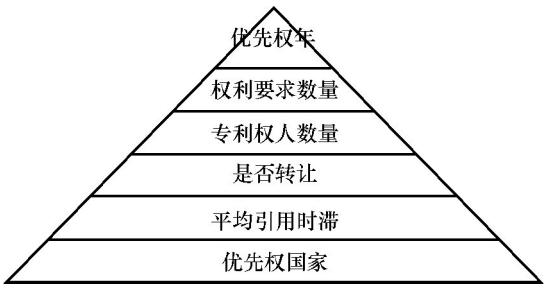


图6 专利被引重要影响因素（重要程度由下往上逐渐变强）

## 7 结束语

本研究利用 SQL Server 2012 BI 平台的决策树方法对专利被引的影响因素进行了分析，并运用分类矩阵和提升图对分析结果进行准确性评估。研究最终发现在影响专利被引的 12 个因素中，专利的最早优先权年是影响最为显著的因素，而后依次是权利要求数量、专利权人数量、是否转让、平均引用时滞、优先权国家，而其他 6 个指标的影响效果并不明显。本研究的结果在一定程度上证明了专利被引影响因素的重要程度，这将有助于专利发明人、专利权人以及研究人员更为深入地认识专利被引行为。但是，本文属于探索性研究，方法的适用性仍旧需要更多的实证研究，另外，运用其他的数据挖掘算法来比较验证分析结果将是下一步研究的重点。□

## 参考文献

[1] 万小丽. 专利质量指标研究 [D]. 武汉：华中科技大学，2009.

[2] MEYER M. What is special about patent citations? Differences between scientific and patent citations [J]. Scientometrics，2000，49 (1)：93-123.

[3] BORNMANN L，DANIEL H D. What do citation counts measure? A review of studies on citing behavior [J]. Journal of Documentation，2008，64 (1)：45-80.

[4] NARIN F，NOMA E，PERRY R. Patents as indicators of corporate technological strength [J]. Research Policy，1987 (16)：143-155.

[5] HE Z L，DENG M. The evidence of systematic noise in non-patent references：a study of New Zealand companies' patents [J]. Scientometrics，2007，72 (1)：149-166.

[6] SOPER M E. Characteristics and use of personal collections [J]. The Library Quarterly，1976：397-415.

[7] GUELLEC D，et al. Applications grants and the value of patent [J]. Economics Letters，2000，69 (1)：109-114.

[8] LERNER J. The importance of patent scope：an empirical analysis [J]. RAND Journal of Economics，1994 (25)：319-333.

[9] 蔡虹，吴凯，孙顺成，等. 基于专利引用的国际性技术外溢实证研究 [J]. 管理科学，2010，23 (1)：18-26. DOI：10.3969/j.issn.1672-0334.2010.01.003.

[10] LEE Y G，et al. An in-depth empirical analysis of patent citation counts using zero-inflated count data model：the case of KIST [J]. Scientometrics，2007，70 (1)：27-39.

[11] 唐华松，姚辉文. 数据挖掘中决策树算法的探讨 [J]. 计算机应用研究，2001，18 (8)：18-19.

[12] 杨冠灿，刘彤，李纲，等. 基于综合引用网络的专利价值评价研究 [J]. 情报学报，2013，32 (12)：1265-1277. DOI：10.3772/j.issn.1000-0135.2013.12.004.

[13] SHELDON R. SQL Server 2012 的五个商业智能特性 [EB/OL]. [2014-10-10]. http://www.searchdatabase.com.cn/showcontent_80440.htm.

[14] Mark Fontecchio. SQL Server 2012 将至全力打造大数据特性 [EB/OL]. [2014-10-10]. http://www.searchdatabase.com.cn/showcontent_58217.htm.

[15] 黄宇达，侯艳芳，王迤冉，等. 基于决策树技术和 SQL Server BI 平台的课程成绩分析 [J]. 电脑知识与技术，2012，8 (16)：3759-3763.

[16] MACLENNAN J，TANG ZhaoHui，CRIVAT B. 数据挖掘原理与应用 [M]. 董艳，程文俊，译. 2 版. 北京：清华大学出版社，2010.

[17] Microsoft. Microsoft 决策树算法 [EB/OL]. [2014-10-10]. http://msdn.microsoft.com/zh-cn/library/ms175312.aspx.

作者简介：吕璐成，男，1989 年生，硕士生。研究方向：专利情报分析。通讯作者。
刘娅，女，1970 年生，硕士，研究员。研究方向：科技政策与管理。
杨冠灿，男，1981 年生，博士。研究方向：专利分析，技术竞争情报。

收稿日期：2014 - 09 - 11