



开放科学  
(资源服务)  
标识码  
(OSID)

# 基于机器学习的高校专利价值评估体系构建

何向 李莉 王小绪

南京理工大学经济管理学院 南京 210000

**摘要:** 高校专利价值评估机制不够成熟, 严重影响了专利转化的顺利推进。因此, 建立一个有效的高校专利价值评估体系能够客观地评判专利价值的高低。研究结合了 1000 条高校专利数据, 从其经济价值、技术价值以及法律价值出发, 共选取了 11 个可测度评价指标, 并借助机器学习的方法进行检验。最终共有 10 个指标通过了随机森林模型与支持向量机的分类检验, 构建了科学性较强、操作性较高的高校专利价值评估通用体系。然而, 本研究仅对体系中各指标项的合理性进行检验, 未来研究可针对指标项的具体权重展开讨论。该指标体系实现了对高校专利价值进行直接定量评估, 避免了定性评估有可能会出现的客观性缺陷, 为高校专利评价决策及科研管理提供新的思路与方向。

**关键词:** 专利价值; 专利价值评估体系; 机器学习

**中图分类号:** G306

## The Construction of Assessing College Patent Value System Based on Machine Learning

HE Xiang LI Li WANG Xiaoxu

School of Economic and Management, Nanjing University of Science and Technology, Nanjing 210000, China

**Abstract:** The evaluation mechanism of patent value in universities is still not mature enough, which seriously affects the progress of patent conversion. Therefore, an effective patent value evaluation system could evaluate the patent value objectively. The research based on economic value, technical value and legal value of 1,000 college patents, this paper selects 11 measurable evaluation dimensions and tests the effectiveness by machine learning. Finally, 10 dimensions are selected and have passed the classification test of random forest model and support vector machine successfully. While, this study only tests the rationality

**基金项目:** 国家重点研发计划子项目“知识产权大数据挖掘技术、智能推送技术”(2017YFB1401903); 国家自然科学基金面上项目“数据驱动的网络广告效应测评与投资决策研究”(71771122)。

**作者简介:** 何向(1992-), 博士研究生, 研究方向: 数据挖掘与知识管理, E-mail: hexiang9360@163.com; 李莉(1970-), 教授, 博士生导师, 研究方向: 数据挖掘与知识管理; 王小绪(1971-), 研究员, 研究方向: 知识产权管理。

of each dimension, the future researches would discuss the weight of each dimension. This method evaluates the patent quantitatively and provide new ideas for college patent evaluation decision-making and scientific research management

**Keywords:** Patent value; patent value assessment system; machine learning

## 引言

在科技创新的浪潮下,各国正在不断实施专利战略为国家科技发展提供坚实而牢固的技术基础。现有数据表明,截至2016年9月,我国有效专利累计达620余万,占世界有效专利比重的89%<sup>[1]</sup>。其中,高校凭借其优越的研发资源、科研团队和科研环境,于最近几年,其专利申请量和授权量得到快速增长,高校专利已成为我国专利体系的重要组成部分。尽管如此,由于申请人对专利价值的市场价值认知不明确,导致专利低转化与其高产出形成了鲜明对比<sup>[2]</sup>。已有数据指出,我国高校已实现成果转化的项目不到10%,真正实现产业化的不足5%,难以满足国家知识产权战略的需求<sup>[3]</sup>。其中,高校专利申请人对于专利价值评估不明确是影响高校专利转化的重要原因。因此,准确、客观、科学地衡量专利价值以促进高校专利的成功转化,是高校急需解决的专利技术交易的难题。

高校专利价值需构建一个有效的专利价值评估体系。然而,由于高校专利覆盖了化工、计算机、机械、生物等各个学科,涉及领域宽泛,且不同领域具有不同的技术复杂性,难以设置出统一的评估标准。同时,高校的专利往往是基础研究的副产品,而并非以市场为导向、用以解决实际问题,导致现有评价体系中难以体现其商业转化价值<sup>[4]</sup>。为此,研究和开发新型

的专利价值评估方法已成为业内的研究热点<sup>[5]</sup>。

目前的一些指标法所建立的专利价值评价体系多数仅停留在理论层面,或由于指标难以获取而缺乏数据支持,其实际可操作性还值得推敲<sup>[6]</sup>。为解决上述问题,本研究以高校专利文本为基础,旨在构建一个更为科学的高校专利价值评估指标体系,并用机器学习的方法验证其合理性,为知识产权评估与高校科技成果转化提供重要的借鉴意义。

## 1 文献综述

对于高校而言,通过将专利转让或许可给下游企业的方式来实现专利市场价值的前提是有一个客观的价值评估体系<sup>[3]</sup>。目前,高校所能提供的专利服务内容较为简单,缺乏对专利深层次的挖掘与分析工作<sup>[7]</sup>,且由于专利价值的影响因子多、评估成本高等因素,关于高校专利价值评估体系的建立一直都比较困难<sup>[8]</sup>。而近年来科研人员对于专利信息服务的强烈需求依旧引发了众多学者对于专利价值评估体系的思考<sup>[9]</sup>。

当前,多数学者都以传统的专利价值评估体系为中心,开展对评估指标的细化或整体体系优化的探索。从指标细化的角度看,研究多围绕经济价值、技术价值与法律价值三个维度进行思考。针对专利的经济价值,国外学者通

常使用实物期权法<sup>[10]</sup>和“Cinart-Park 指数法”<sup>[11]</sup>来代替计算,以解决专利交易市场不完善,无法用专利交易价格作为经济价值的参考这一缺陷。国内学者在此基础上将传统评估体系与实物期权法所采用的参数对专利的市场价值进行了结合分析<sup>[12]</sup>,上述这些方法在一定程度上衡量了专利的经济价值。但由于高校研究人员与市场联系不够紧密以至于不能及时衡量专利的市场价值,因此上述评判指标的可行性值得商榷。对于专利的技术价值,由于其对相应领域专家的依赖程度较高、工作量大,评价结果过于简单<sup>[13]</sup>,不少研究者开始将文本挖掘引入专利技术价值的分析领域,他们借助语义分析技术,将专利文本的非结构化数据转化为结构化数据<sup>[14]</sup>。通过挖掘专利文本中的信息,借助专利的引证关系来评估专利的技术关联性<sup>[15]</sup>;借助计算文档相似度来定义专利的技术新颖度<sup>[16]</sup>。文本挖掘技术的引进使得研究人员可以直接从专利文本内容的角度对其技术价值进行评估,减轻了专家团队评价的主观性,是构建合理的专利价值评估体系的坚实基础。然而,当前对文本挖掘的应用多集中于对单一领域或单个专利样本的计算,鲜有研究通过各个领域的真实专利数据综合评估各维度的科学性检验。

除了创新评价的指标方法外,也不乏有学者对专利价值评估体系做模型补充研究。组合预测模型<sup>[17]</sup>基于传统的评估三要素对专利价值进行了线性预测与非线性预测,为专利价值的评估提供了新思路,但其预测受阻于实证分析,因此所给结论缺乏可行性依据。四要素专利价值评估法<sup>[18]</sup>则弥补了上述缺陷,对专利技术、市场价值、技术可替代性以及专利保护强度方

面做了定量计算,并做了模拟实证计算。然而我国现有的专利数据库对各项评价指标没有准确计量,该模拟计算缺少大样本检验,结论的有效性需要进一步的检验。

目前的研究成果表明,尽管高校专利价值的评估工作起步较晚、经验不够丰富,尚未形成规范的评估标准<sup>[19]</sup>,但随着研究方法的日益丰富,专利评价指标体系的不断细化,一个合理客观的高校专利价值分析体系蓄势待发。为了弥补上述由于专利样本的缺乏而导致的有效性检验不足,本研究拟基于机器学习的方法,旨在构建科学合理的高校专利价值评价指标体系,有助于科研管理者在宏观上对专利的布局、发展方向进行综合性评估,统筹考虑科研成果的优势和劣势。

## 2 高校专利价值评估指标体系设计

本文拟从专利的经济价值、技术价值和法律价值三个维度尝试建立更加完善,指标更为科学的价值评估体系,并在文章第四部分进行实证研究来验证该体系的有效性。

### 2.1 经济价值类指标

专利的经济价值又被称为市场价值,是决定专利价值的外部变量<sup>[6]</sup>。外部环境受当下市场经济走势、供需客观条件等多方面影响,造成该指标难以进行量化分析。因此我们可借助对专利文本内容的分析来弥补这一缺失。

#### (1) 专利经济寿命 (Patent Economic Live)

该指标指专利在市场中可使用的期限,研究中采用剩余有效期 (Residual validity period

long, rpl) 来计算, 以衡量专利的经济时效性。

计算公式如下:

专利经济寿命 = 剩余有效期

### (2) 专利应用性 (Patent Application)

该指标通过专利文本中的关键词数量 (Keyword Number, kwn) 与主分类号数量 (Primary Class Number, pcn) 来反应本专利在市场中的应用领域和适用范围。关键词及主分类号是该专利主旨的体现, 其数量越大, 则表示该专利在市场中的应用范围越广, 专利应用性也越强。则专利应用性公式如下:

专利应用性 = 关键词数量 + 主分类号数量

## 2.2 技术价值类指标

对于专利技术价值的探究是众多学者重点关注的内容, 该类指标是专利质量的表征, 反映了专利的技术创新程度, 与高新技术、前沿科学的关联程度等情况<sup>[21]</sup>。

### (1) 科学关联度 (Science linkage)

科学关联度指的是专利在说明书中引用了专利及非专利的数量。较之企业而言, 高校研发人员更倾向于通过专利引证 (Cited Index, CI) 和其他引证 (Other Index, OI) 来体现专利之间的科学关联性, 而冯岭等<sup>[14]</sup>也证实了专利的直接和间接引用对专利价值产生显著影响, 能有效地评估各个专利的价值。计算公式如下:

科学关联度 = 专利引证 + 其他引证

### (2) 技术专业强度 (Technical Specialty Strength)

该指标指目标专利在专利说明中展开对该专利进行专业性描述的力度, 可反映专利在该领域的衍生性以及专利质量的成熟度。技术专

业强度越高通常伴随着权利要求数的增加, 因此我们用独立权利要求数 (icln)、从属权利要求数 (dcln) 来衡量专利的技术专业强度。计算公式如下:

技术专业强度 =  $w_1$  \* 独立权利要求数 +  $w_2$  \* 从属权利要求数

其中,  $w_1$ 、 $w_2$  均大于零, 且  $w_1 + w_2 = 1$ 。

### (3) 技术研发力度 (Technical development power)

该指标可测试专利的投入状况, 体现了研发团队对于该专利的重视程度, 从侧面反映了高校专利的技术质量。一般来说, 一项高价值的专利需要研发团队投入更多的人力、物力及财力。而物力与财力在文本中暂且无法体现, 因此本研究拟通过发明人数量 (incp) 来计算技术研发力度。技术研发力度计算公式如下:

技术研发力度 = 发明人数量

## 2.3 法律价值类指标

从法律价值维度来看, 专利权利人是否已经充分运用了专利策略来有效地保护专利是影响专利法律价值的一个重要因素<sup>[15]</sup>, 本文拟以专利稳定性和专利保护力度来判断专利受到法律保护程度的大小。

### (1) 专利稳定性 (Patent Stability)

该指标用来判断专利在市场中的发展是否持续稳定。专利受到法律的保护年限越长, 则其在市场中的发展越稳定。由于不同类型专利的保护年限并不相同, 我们考虑将专利类型 (Patent Type) 作为专利稳定性的底层指标。计算公式如下:

专利稳定性  $\begin{cases} \text{pt1=发明专利} \\ \text{pt2=实用新型专利} \\ \text{pt3=外观设计专利} \end{cases}$

## (2) 专利保护力度 (Patent Protect Power)

专利保护力度指专利能够受到法律保护范围的大小。随着专利保护意识的增强,高校开始不断开拓和保护国际市场<sup>[20]</sup>,即同族专利 (Patent Family Members, scn) 的规模大小能反映该项技术的重要程度。再者,现有的在线专利评估系统中,天弓系统 (<https://www.tiikong.com/>) 首次加入专利说明书中独立权利要求长度 (icll)、从属权利要求长度 (dcll) 的计算来进一步识别

专利的保护力度。一般来说,内容越长的独立权利要求与从属权利,其保护力度越高。因此,同族专利大小和权利长度共同反映了权利人对于目标专利的重视程度,其计算公式如下:

$$\text{专利保护力度} = w_1 * \text{同族专利数} + w_2 * \text{独立权利要求长度} + w_3 * \text{从属权利要求长度}$$

其中,  $w_1$ 、 $w_2$ 、 $w_3$  均大于零,且  $w_1 + w_2 + w_3 = 1$ 。

综上所述,本文拟建立的高校专利价值评估体系指标,以及体系中每个指标所包含的底层指标如表 1 所示。

表 1 高校专利价值评估指标体系

一级指标	二级指标	底层指标	标引代码
(E)经济价值	(E1)专利经济寿命	剩余有效期	rpl
	(E2)专利应用性	关键词数量 主分类号的数量	Kwn pcn
(T)技术价值	(T1)科学关联度	专利引证数	in
	(T2)技术研发力度	发明人数量	inco
	(T3)技术专业强度	独立权利要求数	icln
		从属权利要求数	dcln
(L)法律价值	(L1)专利稳定性	专利类型	pt1,pt2,pt3
	(L2)专利保护力度	独立权利要求长度	icll
		从属权利要求长度	dcll
		同族专利数	scn

### 3 实证研究

为验证上述所列指标体系的合理性,本文从中国高校最权威的知识产权运营平台——中高知识产权运营交易平台 (<https://www.cnuiip.com>) 中随机选取涵盖有电力、采矿业、信息服务业、制造业以及农林牧渔业等领域的 1000

条专利。其中,发生过转让的专利为 500 条,未发生过转让的专利也为 500 条。对于专利而言,只有已实施有收益的或者将会实施有潜在经济价值的专利才会发生专利转移<sup>[18]</sup>,一般地,相较于未转移的专利来说,已发生转移的专利更具有新颖性、创造性、实用性。因此发生过转移的专利能更真实地反映专利技术的实际情

况。本次实证以高校专利价值评估体系为标准展开对 1000 条专利进行转移与未转移的分类，结果更具有参考性。为确保实证结果的公平性，所挑选专利类别涉及机械、化学、生物医学等各个领域，研究的实证流程如图 1 所示。

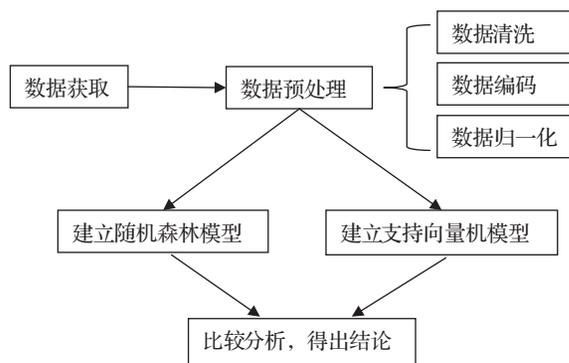


图 1 高校专利价值评估体系实证流程

### 3.1 数据获取与预处理

本文中所有实证过程均使用 R 语言在 Rstudio 中进行。我们首先借助中高专利知识平台 (<https://www.cnui.cn/>)，根据符合要求的检索式从数据库中获得 1000 个分属各个领域的高校专利信息样本，并依照表 1 中的指标要求进行各个维度的标引。

### 3.2 Random Forests 随机森林模型

Random Forests 随机森林模型是机器学习领域中一个普遍使用的分类器，本次实证拟定以表 1 中的 12 个底层指标为初始自变量，专利最终是否实现转移为因变量（已实现转移的专利设定类别为‘1’，未实现转移的专利类别为‘0’）建立模型。我们从 1000 个专利样本中随机抽取 600 个作为训练集训练中我们所需模型，随后用另外 400 个样本作为测试集进行精

确度检验。模型实现过程中的部分代码以及结果准确性如下：

```

> plot(model)
> pred=predict(model,testdata[,-13])# 对测试数据进行预测
> accuracy=sum(pred==testdata$ 是 否 )/nrow(testdata)# 计算准确率
> accuracy
[1] 0.696
  
```

各指标在随机森林模型中的重要性如表 2 所示。

表 2 高校专利价值评估体系指标重要性——基于随机森林模型

评估指标	指标重要性
剩余有效期rpl	70.441158
关键词数量kwn	25.667920
发明人数量inco	25.399781
独立权利数icln	9.309576
从属权利数dcln	26.586138
独立权利长度icll	45.665237
从属权利长度dcll	41.795221
同族专利数scn	5.458101
专利类型pt1	4.922319
专利类型pt2	7.840561
专利类型pt3	1.256812
专利引证数in	14.572025

从建模结果中发现，上述指标的预测准确度达到 69.6%，结果较为理想。尤其我们注意到，体系中各个指标的重要性都为正，且剩余有效期、关键词数量、发明人数量、独立权利长度、从属权利长度以及专利引证数的指标重要性相对较高，即这些指标与评估体系的相关性更强，

而独立权利数、同族专利数、专利类型与体系之间的相关性相对较弱。

### 3.3 SVM支持向量机模型

SVM 支持向量机属于监督式学习方法，它广泛应用于统计分类以及回归分析中。在本次实证中，我们重新将 1000 个样本按比例随机分为训练集和预测集，其中训练集样本量为 700 个，预测集样本量为 300 个。实证过程中依旧按照表 1 提出的底层指标作为自变量，以专利是否转移为因变量进行分类预测。预测过程的部分代码及精确度结果如下：

```
> plot(model)
> pred=predict(model,testdata[,-13])# 对测试
数据进行预测
> accuracy=sum(pred==testdata$ 是否)/nrow
(testdata)# 计算准确率
> accuracy
[1] 0.92
```

预测结果显示，在 300 个样本的测试集中，预测正确率为 92%。该模型准确率高于随机森林模型（92%>69.6%），这不仅说明 SVM 支持向量机模型构建的分类器可以很好的适用于专利分类，也再一次证明了上述指标体系的可用性。

### 3.4 实证结果

本次实证基于高校专利文本信息，借助机器学习的方法验证了同一指标体系在不同的两个分类模型中的有效性。分类模型均以 11 项底层指标为自变量，专利是否发生转移为因变量进行架构。由于实证数据中主分类号数量的实测值相等（pcn=10），无法服务于模型的建立，

因此将这一指标暂时剔除。实证最终结果表明，在本文选取的 11 项底层指标中提示有 10 项指标（关键词数量，发明人数量，非专利引文数，引证专利数，主分类号的数量，独立权利要求数，从属权利要求数，专利类型，独立权利要求长度，从属权利要求长度，同族专利数）通过了随机森林模型和支持向量机模型的检验（准确率分别为 69.6 和 92%），证实了本文提出的高校专利价值评估指标体系可以有效作为判断高校专利是否能够发生转移的可靠参考，可在后续专利价值评估中被广泛应用，最终评估体系如图 2 所示。

## 4 结论与启示

本文立足于高校专利本文提出高校专利价值评估体系，并基于文本挖掘技术验证了该方法的可靠性和实用性。该体系为高校知识产权的评估与定价提供了理论依据、算法支持和应用指导。在此基础上，本文的最终结果如下：

首先，从研究方法上来看，本文立足于高校专利文本建立了一个具有代表性的高校专利价值评估体系，相较于其他专利价值评估模型来说，该体系首次从宏观层面上通过了机器学习方法中两种分类器的有效实证检验，是传统评估方法与机器学习相结合的一次创新性实践。

其次，由于模型提示主分类号数量与专利保护年限两项指标被剔除，则目前常用于企业专利价值评估中的这两项指标实际上并不适用于我国高校的专利数据分析与解读<sup>[21]</sup>，与传统观点和预期结果不一致。在往后的体系研究中，应慎重考虑这两项指标的可参考性。

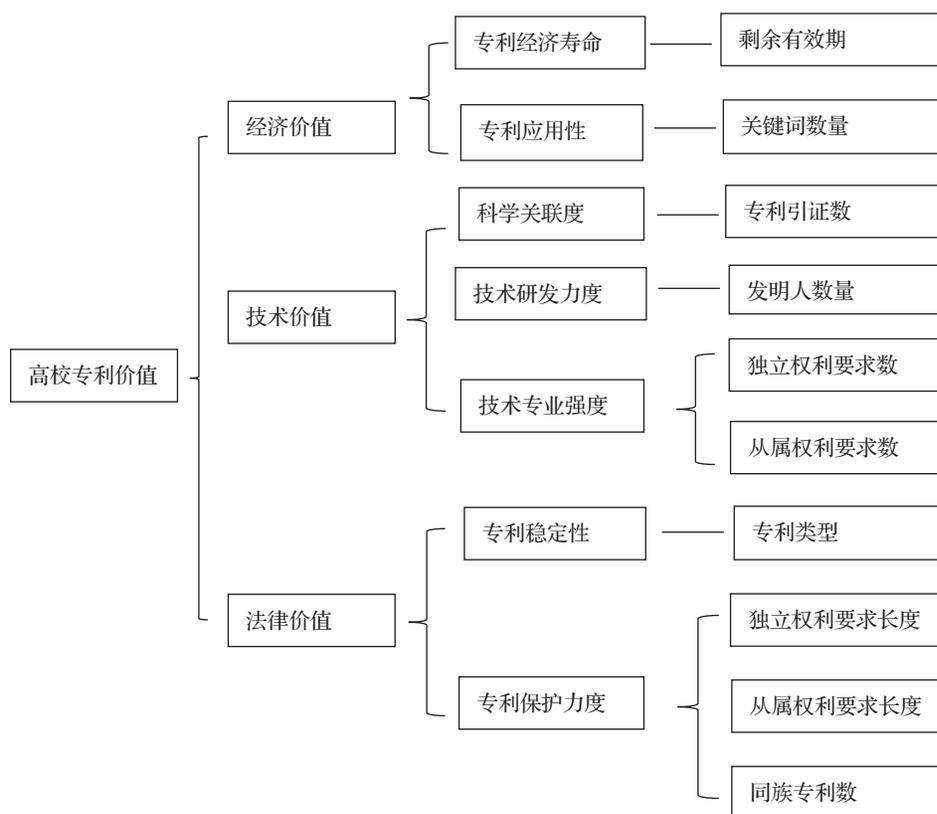


图2 高校专利价值评估指标体系

再者，作为高校专利特有的科学关联度指标，可在一定程度上反映分析对象在某专题或某领域研发的连续性、进展程度、集聚程度及其围绕技术核心的专利战略情况<sup>[16]</sup>。研发人员在申请专利时应充分利用这一特点来提高专利在本领域的技术价值，这既是相关专利技术知识基础积累和创新度的体现，又是本专利的引用优势和技术基础优势的展现<sup>[22]</sup>。

最后，模型赋予专利保护力度相对较高的的重要性。一方面指明高校应充分树立法律保护意识，不仅要维护好专利在国内的各项利益，也可通过申请同族专利来提高专利在国外的保护力度，扩大保护范围；另一方面也反映了高校应提高专利的独有性和特殊性，重视专利本文内容中独立权利与从属权利来提升专利的法

律价值。其他指标中例如独立权利数，从属权利数等与结构指标的相关性相对不显著，提示它们作为目前专利技术价值指标的合理性不足。但是由于实证数量的局限性，在分类结果准确的前提下，仍可以作为技术附加值来表征。

综上所述，本文提出的高校专利价值评估体系将有利于推进高校专利技术的转化和应用，提升整体知识产权的应用和管理水平。同时，基于高校专利价值评估体系的分析结果，不仅可以用于比较不同高校之间的科研实力，也可以由此而分析出各个高校在不同专业领域的变化及优势，更科学地对专利研发进行规划，提高高校科研决策水平。必须要指出的是，本研究仅探讨了高校专利价值评估的各项具体指标，未来的研究可考虑对每个指标的权重加以分析，

进一步细化高校专利价值评估体系。

## 参考文献

- [1] 国家知识产权局.《中华人民共和国国家知识产权局专利统计年报2016》[EB/OL]. [2019-01-13]. <http://www.cnipa.gov.cn/tjxx/jianbao/year2016/indexy.html>
- [2] 饶凯,孟宪飞,徐亮.研发投入对地方高校专利技术转移活动的影响——基于省级面板数据的实证分析[J].管理评论,2013,25(5):144-154.
- [3] 袁晓东,张军荣,杨健安.中国高校专利利用的影响因素研究[J].科研管理,2014,35(4):76-82.
- [4] 王小绪.大学技术转移机构建设:现状、问题与对策[J].高等教育研究,2014(12):59-62.
- [5] 杨思思,戴磊,郝屹.专利经济价值度通用评估方法研究[J].情报学报,2018,37(1):52-60.
- [6] 吕晓蓉.专利价值评估指标体系与专利技术质量评价实证研究[J].科技进步与对策,2014,31(20):113-116.
- [7] 赵慧清.高校图书馆专利咨询服务现状调研分析与对策[J].图书馆理论与实践,2012(6):59-61.
- [8] 吴红,马永新,董坤,等.高校专利分级管理实现的障碍及对策研究[J].图书情报工作,2016(2):59-63.
- [9] 王丽萍,杨波,秦霞,等.高校图书馆专利信息服务内容、模式与趋势[J].图书情报工作,2015(6):113-119.
- [10] Myers S C. Determinants of corporate borrowing [J]. Journal of Financial Economics, 1977, 5(2):147-175.
- [11] Ginarte J C, Park W G. Determinants of patent rights: A cross-national study [J]. Research Policy, 1997, 26(3):283-301.
- [12] 李秀娟.专利价值评估的影响因子[J].电子知识产权,2009(5):64-67.
- [13] 刘玉琴,汪雪锋,雷孝平.基于文本挖掘技术的专利质量评价与实证研究[J].计算机工程与应用,2007,43(33):12-14.
- [14] 冯岭,彭智勇,刘斌,等.一种基于潜在引用网络的专利价值评估方法[J].计算机研究与发展,2015,52(3):649-660.
- [15] Chiesa V, Frattini F, Gilardoni E, et al. Searching for factors influencing technological asset value [J]. European Journal of Innovation Management, 2007, 10(4):467-488.
- [16] 肖国华,王春,姜禾,等.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):96-96.
- [17] 孙玉艳,张文德.基于组合预测模型的专利价值评估研究[J].情报探索,2010(6):73-76.
- [18] 李振亚,孟凡生,曹霞.基于四要素的专利价值评估方法研究[J].情报杂志,2010,29(8):87-90.
- [19] 梁燕,吴锡尧,张素娟,等.高校专利评估现状、影响因素及其对策研究[J].科学学与科学技术管理,2004,25(4):48-50.
- [20] 王嵘.我国高校专利战略信息分析及研究[J].现代情报,2007,27(10):40-45.
- [21] 李姝影,方曙,许海云.面向科研机构专利布局的专利组合分析模型及实证研究[J].情报杂志,2016,35(7):132-138.
- [22] 胡小君,陈劲.基于专利结构化数据的专利价值评估指标研究[J].科学学研究,2014,32(3):343-351.