

文章编号:1006-2467(2020)07-0705-13

DOI: 10.16183/j.cnki.jsjtu.2020.99.007

采用机器学习算法的技术机会挖掘模型及应用

包清临, 柴华奇, 赵嵩正, 王吉林

(西北工业大学 管理学院, 西安 710129)

摘要: 现有技术机会挖掘结果的应用性较低,究其原因,一是样本量较小,二是挖掘过程缺乏对技术应用前景的评估.为解决这一问题,以提升挖掘结果的应用性为目标,以海量专利为样本,在现有研究的基础上,加入对技术应用前景的评估,提出三维的专利预测模型.采用机器学习下的 PL-SA 算法,结合 Hadoop 平台下的 MapReduce 计算框架,运用专利文本挖掘,构建专利预测模型的技术维和功效维;采用熵权和 TOPSIS 法构建专利预测模型的价值维;基于 MapReduce 计算框架填充专利预测模型的单元项.并以 DII 数据库中钛领域 1999~2018 年 133 508 例专利文本为样本应用了专利预测模型.结果显示,该模型在钛领域内共挖掘出了 3 个优先级和 2 个次级的技术机会,可以按优先顺序对技术机会进行开发.该模型丰富了技术机会挖掘的方法,为创新主体指明了更为准确和前景化的技术研发方向.

关键词: 机器学习算法;专利文本挖掘;技术机会;钛

中图分类号: TP 391

文献标志码: A

Model of Technology Opportunity Mining Using Machine Learning Algorithm and Its Application

BAO Qinglin, CHAI Huaqi, ZHAO Songzheng, WANG Jilin

(School of Management, Northwestern Polytechnical University, Xi'an 710129, China)

Abstract: The applicability of the existing technology opportunity mining results is relatively low owing to the small sample size and the lack of evaluation on the technology application prospects in the mining process. In order to solve this problem, with the goal of improving the applicability of mining results, based on the existing research, this paper proposes a three-dimensional patent prediction model by taking into account a large number of patents and adding an assessment of the prospects for technology applications. Using the PLSA algorithm in machine learning and combining with the MapReduce computing framework under Hadoop, it uses patent text mining to construct the technology and function dimensions of the patent prediction model, adopts entropy weight and TOPSIS method to construct the value dimension of the patent prediction model, and fills the element items in the patent forecasting model based on the MapReduce computing framework. Then, it applies a patent prediction model to 133 508 patent texts in the titanium field in the DII database from 1999 to 2018. The results show that the model has identified a total of 3 priority and 2 secondary technology opportunities in the titanium field, and these technology opportunities can be developed in order of priority. This model enriches the method for

收稿日期: 2020-03-05

基金项目: 国家自然科学基金资助项目(71971170)

作者简介: 包清临(1990-),女(藏族),甘肃省天水市人,博士生,现主要从事数据挖掘与知识管理研究.

电话(Tel.): 029-84766218; E-mail: 499195647@qq.com.

technology opportunity mining and provides a more accurate and prospective technology research and development direction for innovation subjects.

Key words: machine learning algorithms; patent text mining; technology opportunity; titanium

技术机会是开展技术创新活动的前提和条件^[1-2]. 挖掘并抢占技术机会, 为创新活动的展开赢得先机^[3-5]. 近年来, 以专利驱动挖掘技术机会受到学术界的广泛关注. 专利记载了核心技术信息, 包含了前沿技术创意, 充分利用专利, 可以大大缩短技术机会挖掘时间, 节省 40% 的研发费用^[6-7].

根据技术机会相关研究中专利样本量的大小, 可以把现有专利辅助挖掘技术机会的研究分为两类. 一类研究是以少量的专利样本为出发点, 结合技术的应用前景挖掘技术机会. 江屏等^[8]在确立研发目标后, 以封隔器领域的 673 件专利为样本, 根据 TRIZ 理论为技术机会构建待规避专利技术的原理解模型, 使其规避可能存在的专利风险. Wang 等^[9]认为科学论文和专利记录了科学发现和技术发展中核心的知识, 通过二者可以发现并挖掘潜在的技术机会, 因此采用文本挖掘和聚类算法对微藻类生物燃料领域的技术机会进行了探索. Jia 等^[10]选择目标类专利, 采用类比设计方法, 对技术机会进行发掘, 寻找创新设计理念. 此类研究能较好地实现特定目标, 但由于专利样本量较小, 计算结果与设计目标容易背离, 技术机会挖掘结果应用性低. 另一类研究以海量专利样本为出发点, 通过数据挖掘技术, 结合统计方法对专利文本的结构化信息进行处理. Cheng^[11]以大量专利为样本, 采用文本挖掘技术, 构建专利技术功效矩阵 (PTFM), 通过对 PTFM 中技术呈现出的密集区及真空区的分析, 得到技术机会挖掘结果. Jin 等^[12]从专利信息出发, 通过市场和产品两个维度, 搭建 PTFM 并绘制专利技术功效图, 以识别太阳能照明领域的创新技术. Joung 等^[13]通过专利文本挖掘, 提取了技术关键词和功效关键词, 从而构建 PTFM, 对生物传感器中电子转移领域的新兴技术进行了识别. 相较于前一类研究, 非目标类研究对专利的人工分类干涉较少, 通过更广的专利覆盖, 确定出更加前沿的技术方向, 但是该类研究较少涉及了创新目标, 得出的技术机会的应用性较低.

由此可以看出, 少量专利样本或缺乏创新目标下的研究, 缺乏对技术前景的评估, 由此形成的技术机会必然对其应用性产生影响, 从而降低技术机会挖掘结果的可行性. 因此, 本文以提升结果的应用性为目标, 在技术机会挖掘过程中加入对技术应用前

景的评估, 采用机器学习算法, 以海量专利数据为样本, 结合文本挖掘、MapReduce 计算框架和熵权-TOPSIS 法, 使用 Python 编程语言, 构建三维的专利预测模型. 并在钛领域内应用该模型, 为其挖掘潜在的技术机会.

1 相关研究综述

1.1 文本挖掘与机器学习算法

文本挖掘亦称文本数据挖掘, 指使用数据挖掘、机器学习、自然语言处理等技术, 结合分类、聚类、偏差检测、关系分析等方法, 解决文本集合或语料库信息过载的问题, 以发现其中隐含的信息或模式的过程. 一般来说, 文本挖掘包括信息检索、信息抽取、信息挖掘以及信息解释 4 个步骤, 其中信息挖掘最为关键, 主要包括文本预处理、文本聚类或文本关联、特征提取以及结果可视化^[14].

最开始研究抽取关键词的模型是空间向量模型 (VSM), 但由于该模型方式不能对语意实现完全描述^[15], 便逐步涌现出了通过给定潜在变量, 经过建模主题实现关键词共现的主题模型. 在主题模型中, 概率潜在语义分析 (PLSA) 算法较为主流. PLSA 算法是在 1999 年由 Hofmam 提出的, 能够解决潜在语义分析 (LSA) 算法在文本挖掘过程中对大样本处理不足的问题, 同时, 相较于奇异值矩阵分解 (SVD) 降维方法, PLSA 算法能更清晰地推断文档与词语间的关系模型. 现有研究显示, 相较于 LSA 算法, PLSA 算法能显著提升关键词抽取的效率和精确度, 减少人工抽取的误差, 适于在大样本环境中运行. 它通过建立文档、主题与词语三者之间的映射关系, 基于双模态和共现准确抽取关键词^[16-17].

PLSA 算法通过计算“文档-主题”和“词语-主题”的概率矩阵, 对共现概率进行排序, 表现基于共同的观察数据的隐藏变量关系, 经过相似度计算和聚类分析, 达到筛选关键词的目的. 因此可以说 PLSA 算法是构建概率主题模型的最佳选择. PLSA 算法主要基于以下假设^[18]: ① 在同一潜在主题集合 $T = \{t_1, t_2, \dots, t_s\}$ 下, 文档集合 $D = \{d_1, d_2, \dots, d_n\}$ 和词语集合 $W = \{w_1, w_2, \dots, w_m\}$ 之间是独立的; ② 词语 $w_i (i = 1, 2, \dots, m)$ 和潜在主题 $t_k (k = 1, 2, \dots, s)$ 之间也是独立的.

由此,选定任一文档 $d_j (j = 1, 2, \dots, n)$ 的概率为 $P(d_j)$, 在 PLSA 算法中, 该文档所属任一潜在主题 t_k 的概率服从多项分布, 表示为 $P(t_k | d_j)$; 进一步在该潜在主题下, 任一文档 d_j 生成词语 w_i 的概率也服从多项分布, 表示为

$$P(d_j, w_i) = \sum_{k=1}^s P(w_i | t_k) P(t_k | d_j) = \sum_{k=1}^s P(t_k) P(d_j | t_k) P(w_i | t_k) \quad (1)$$

式中: t_k 为潜在主题集合 T 中的任一潜在主题; d_j 为文档集合 D 中任一文档; w_i 为词语集合 W 中任一词语; $P(d_j)$ 为选中文档 d_j 的概率; $P(w_i | t_k)$ 为在确定潜在主题 t_k 下选中词语 w_i 的条件概率; $P(t_k | d_j)$ 为在确定文档 d_j 后选中潜在主题 t_k 的条件概率; $P(t_k)$ 为选中同一潜在主题 t_k 的概率; $P(d_j | t_k)$ 为在确定潜在主题 t_k 下选中词语 w_i 的条件概率。由于 $P(t_k | d_j)$ 和 $P(w_i | t_k)$ 符合多项分布, 所以可以通过最大期望(EM)算法优化目标函数, 通过循环迭代得到式(1)的最大似然解, 有效解决数据过拟合问题。EM 算法下的计算分为 E 步和 M 步。E 步为期望步骤; M 步为最大化步骤。根据 E 步和 M 步的计算规则, 所得如下。

E 步: 计算给定观测对 (d_j, w_i) 时选中潜在主题 t_k 的概率。

$$P(t_k | d_j, w_i) = \frac{P(t_s) P(d_j | t_s) P(w_i | t_s)}{\sum_{k=1}^s P(t_k) P(d_j | t_k) P(w_i | t_k)} \quad (2)$$

M 步: 在确定 $P(t_k | d_j, w_i)$ 的条件下计算 $P(w_i | t_k)$ 和 $P(t_k | d_j)$ 的概率。

$$P(w_i | t_k) = \frac{\sum_{j=1}^m m(d_j, w_i) P(t_k | d_j, w_i)}{\sum_{j=1}^n \sum_{i=1}^m P(t_s | d_j, w_i)} \quad (3)$$

$$P(t_k | d_j) = \frac{m(d_n, w_m) P(t_k | d_n, w_m)}{\sum_{i=1}^m m(d_n, w_i) P(t_k | d_n, w_i)} \quad (4)$$

式中: $P(t_s)$ 为潜在主题取得最大值 t_s 时的概率; $P(d_j | t_s)$ 为在概率主题取得最大值 t_s 时选中文档 d_j 的条件概率; $P(w_i | t_s)$ 为在概率主题取得最大值 t_s 时选中词语 w_i 的条件概率; $P(t_k)$ 为任选某一潜在主题 t_k 的概率; $P(d_j | t_k)$ 为任选某一潜在主题 t_k 时选中文档 d_j 的条件概率; $P(w_i | t_k)$ 为任选某一潜在主题 t_k 时选中词语 w_i 的条件概率; $m(d_j, w_i)$ 为词语 w_i 与文档 d_j 的共现概率; $P(t_k | d_j, w_i)$ 为在给定观测对 (d_j, w_i) 时选中任一潜在主题 t_k 的条件

概率; $P(t_s | d_j, w_i)$ 为在给定观测对 (d_j, w_i) 时选中潜在主题最大取值 t_s 时的条件概率; $m(d_n, w_m)$ 为文档取得最大值 d_n 与词语取得最大值 w_m 共现概率; $P(t_k | d_n, w_m)$ 为在文档取得最大值 d_n 和词语取得最大值 w_m 后选中任一潜在主题 t_k 的条件概率; $m(d_n, w_i)$ 为文档取得最大值 d_n 与任一词语 w_i 的共现概率; $P(t_k | d_n, w_i)$ 为文档取得最大值 d_n 并任选词语 w_i 后选中任一潜在主题 t_k 的条件概率。用随机数初始化之后, 交替执行 E 步和 M 步进行迭代计算。PLSA 算法以专利文本挖掘来实现对技术机会的挖掘, 首先基于自然语言处理(NLP)对专利摘要文本进行预处理, 包括对专利文本的词性标注; 然后根据词性的不同, 抽取专利摘要中的技术关键词和功效关键词; 最后通过“专利摘要文本-词语”概率矩阵对技术关键词和功效关键词的词频进行统计。

1.2 Hadoop 平台下的 MapReduce 计算框架

PLSA 算法的实现, 需要把“文档-主题”和“词语-主题”二维矩阵映射到分布式处理框架中。Hadoop 是较为常见的大数据分布式处理平台, 它包括分布式文件系统(HDFS)和 MapReduce 计算框架。MapReduce 计算框架通过主节点的 Job Tracker 来调度和分发所有任务, 无需人工干预, 能较好地实现计算目的。

MapReduce 计算框架主要包含 Map 阶段和 Reduce 阶段。Map 阶段主要负责对数据的分析处理, 最终转化为 $\langle \text{key}, \text{value} \rangle$ 数据对; Reduce 阶段主要获取 $\langle \text{key}^*, \text{value}^* \rangle$ 数据对, 通过对其进行统计处理, 得到相应结果。

1.3 专利技术功效矩阵

在传统技术机会挖掘的模型中, PTFM 较为普遍。PTFM 以技术维为 x 轴, 功效维为 y 轴, 形成二维平面坐标^[19]。技术维与功效维的交点为专利单元项, 它的大小与对应具有相应技术关键词和功效关键词的专利数量一致。在 PTFM 中, 单元项越大, 表示专利组合数量越多, 该区域专利布局越密集; 单元项越小, 表示专利组合数量越少, 该区域的专利布局越空白^[20-21]。不少学者以 PTFM 为依据, 对技术机会进行挖掘与决策^[22-24]。

然而, 仅运用 PTFM 作为分析工具, 是不能实现对技术机会的完全预测。这是因为技术关键词和功效关键词仅是对技术特征及功能特征的描述, 不能对技术的运用前景和价值进行预测。因此, 需要对 PTFM 进行改进, 使其更加适应创新主体对前沿技术识别的需要。

1.4 专利价值与技术的应用前景

最早提出专利价值的是 Pakes 等^[25],他们指出专利价值是用于衡量专利技术存续期的重要指标. Reitzig^[26]认为专利价值与受专利保护的技术产品的市场指数呈正比. Bakker^[27]指出专利价值与创新主体的技术计划有关,是技术战略的反映. Bessen 等^[28]提到专利价值是专利带给专利权人的市场收益. Hiller 等^[29]认为专利价值是企业专利投资、转让、许可使用、质押等资产运营行为的重要价值参考. 由此可知,本文认为专利价值体现为专利带给专利权人的收益,该收益与技术机会的应用前景直接相关. 因此可以采用专利价值指标,对技术的应用前景进行评估.

综合以上分析,在算法选择上,本文认为 PLSA 算法能更加突出地解决大样本下专利文本的词性标

注和关键词抽取问题,相较于其他方法更加高效,同时通过文档-词语间的概率转换关系,能最大程度确保关键词抽取的精确度;在框架选择上,本文选择 MapReduce 计算框架完成关键词抽取,因此能不受人工干预完成实验内容;在模型选择上,本文将改进 PTFM,构建以技术应用前景为指引的技术机会挖掘模型.

2 专利预测模型的提出及构建

2.1 专利预测模型的内涵

在 PTFM 的基础上,创新性的添加第三维——价值维. 即价值维与技术维和功效维一起共同构成三维的专利预测模型. 专利预测模型是基于专利的、以价值为导向的、提供技术机会前景预测的技术应用模型. 专利预测模型的构建流程如图 1 所示.

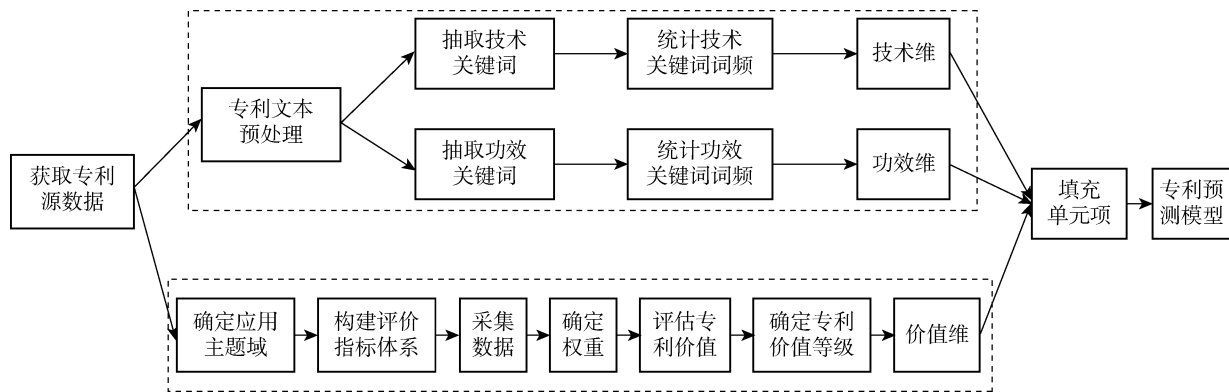


图 1 专利预测模型构建流程

Fig. 1 Construction process of patent prediction model

实验环境由 3 台计算机搭建,在 Java 8.1 环境中使用 Python 语言,采用 PLSA 算法,基于 MapReduce 计算框架对专利文本进行挖掘.

2.2 技术维和功效维的构建

专利预测模型的 x 轴是技术维,其坐标上为技术关键词,技术关键词是领域内高频技术方案和技术特征的展示,代表确定领域密集技术方向,为未来技术发展提供预见依据. 专利预测模型的 y 轴是功效维,其坐标上功效关键词,功效关键词是领域内专利技术解决方案的展示,代表确定领域的技术功能,为未来技术应用功能的创新提供预见依据. 技术维和功效维的构建步骤如下:

步骤 1 专利源数据获取. 德温特专利索引数据库(DII)中的专利已是二次加工的数据,具有格式规范、噪音小的特点^[30]. 因此选择 DII 中的专利摘要文本为源数据.

步骤 2 基于自然语言处理工具(NLTK)的文

本预处理. 专利摘要文本为自然语言,在计算机直接处理前,需通过预处理,对文档进行分词,除去停用词,标注词性,并确定线索词. ① 词性标注. 技术关键词和功效关键词一般为独立名词或名词与名词的组合,因此可以通过词性标注来提高关键词抽取的有效性. 本文选择对专利摘要文本中“USE”和“ADVANTAGE”两部分的文本进行词性标注,技术关键词为名词,功效关键词为形容词. ② 确定功效关键词的线索词. 本文使用同义词典中的特定动词对功效关键词定位,这些特定动词就是确定功效关键词的线索词,使用线索词,可以显著提高专利功效关键词挖掘的速度和正确率. 功效关键词的线索词如表 1 所示.

步骤 3 在 MapReduce 计算框架之中,基于 PLSA 算法,抽取并计算技术关键词的词频,同时将结果输出;抽取并统计功效关键词的词频,同时将结果输出.

表 1 确定功效关键词的线索词

Tab. 1 Clue words for determination of function keywords

序号	线索词(v.)
1	防止/妨碍/中止/阻塞/阻碍/阻止(prevent/impede/stem/stop/obstruct/discourage)
2	提高/增强/上升/升级/拉升/升起(improve/enhance/raise/escalate/lift/hoist)
3	减少/减轻/削减/缩减/降低(reduce/lessen/cut/diminish/decrease)
4	改变/转换(change/alter)
5	确保/保证/保障(ensure/guarantee/safeguard)
6	优化/加强/巩固(optimize/intensify/strengthen)

2.3 价值维的构建

本文采用专利价值来表示对技术应用前景的评估.在评估专利价值方法选择上,熵权-TOPSIS 法

具有样本容量大、不受参考序列选择的干扰,信息失真小、运算灵活及应用领域广等优点^[31],因此选择该方法对专利价值进行评估.

2.3.1 基于 DMC 的应用主题域确定 德温特手工代码(DMC)是 DII 数据库中的专业人员按照严格标引准则分配给每例专利的分类代码,本文使用 DMC 作为应用主题域.由于专利被引次数和引用关系的相关研究是目前预测技术前景趋势的重要指标^[32-34],本文选择确定领域内前 1000 例高被引专利,统计其 DMC,并筛选出频率最高的 25 个 DMC,把 DMC 转化为所属应用主题域.

2.3.2 专利价值评价指标体系构建 基于前人研究,本文从专利技术价值、专利市场价值和专利权利价值三个方面来构建专利价值评价指标体系^[35-36].该指标体系共有 3 个一级指标、9 个二级指标,具体指标项、指标含义及作用方向如表 2 所示.

表 2 专利价值评价指标体系

Tab. 2 Evaluation index system of patent value

评价目标	一级指标	二级指标	二级指标的描述	作用方向
专利价值	专利技术价值 PTV	技术新颖度 PTV ₁	专利技术的新兴程度	+
		技术核心度 PTV ₂	专利技术的核心程度	+
		技术应用度 PTV ₃	专利技术的应用程度	+
	专利市场价值 PMV	市场广泛度 PMV ₁	专利在相关市场中分布的广泛程度	+
		市场适应度 PMV ₂	专利适应市场环境,及价值实现程度	+
		市场竞争度 PMV ₃	专利面对市场竞争的反应程度	+
		市场垄断度 PMV ₄	专利垄断相关市场的程度	-
	专利权利价值 PRV	专利剩余有效程度 PRV ₁	法律规定年限内专利剩余有效期限	+
		专利诉讼程度 PRV ₂	专利参与法律诉讼的程度	-

2.3.3 指标打分及数据采集 聘请熟悉某领域的学者、管理人员和专利知识产权法律工作者进行打分,采用 10 分制对每个二级指标进行打分,其中 10 分为最高,1 分为最低.

2.3.4 熵权法下权重的计算 熵权法是一种有效确定各评价指标权重的方法.设有 m 个应用主题域, n 个评价指标,则对应的评价指标决策矩阵可表示为

$$\mathbf{X} = (x_{ij})_{m \times n} = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & & \vdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix} \quad (5)$$

各评价指标与专利价值呈正比,专利指标的最大为最优值.记 x_j^* 为第 j 列的最大值,因此 x_{ij} 与 x_j^* 的之间关系可表示为

$$D_{ij} = \frac{x_{ij}}{x_j^*} \quad (6)$$

对 D_{ij} 进行归一化处理,即

$$d_{ij} = D_{ij} / \sum_{i=1}^m \sum_{j=1}^n D_{ij} \quad (7)$$

计算第 j 个指标的熵值可表示为

$$e_j = - \frac{1}{\ln n} \sum_{j=1}^n (f_{ij} \ln f_{ij}) \quad (8)$$

式中: $f_{ij} = \frac{d_{ij}}{\sum_{j=1}^n d_{ij}}$,若 $f_{ij} = 0$,则

$$\lim_{f_{ij} \rightarrow 0} (f_{ij} \ln f_{ij}) = 0$$

因此第 j 个指标的权重系数可表示为

$$\omega_j = j(1 - e_j) / \sum_{j=1}^n e_j \quad (9)$$

2.3.5 基于 TOPSIS 法的专利价值评估 TOPSIS 法是由 Hwang 等^[37]于 1981 年首次提出的,该方法的核心是逼近理想解的排序,基本思路是衡量

被评价指标现实状态与理想状态的欧氏距离,从而对其实现评估. TOPSIS 法已被广泛应用于对有限方案进行多目标评价决策领域^[38-40].

以式(9)为基础构建标准化权重矩阵:

$$\mathbf{X}^* = \omega_j x_{ij}^* = \begin{bmatrix} \omega_1 x_{11}^* & \cdots & \omega_n x_{1n}^* \\ \vdots & & \vdots \\ \omega_1 x_{m1}^* & \cdots & \omega_n x_{mn}^* \end{bmatrix} \quad (10)$$

确定正理想解 \mathbf{R}_j^+ , 正理想解为设想的最优解, 是由 \mathbf{X}^* 每列中最大值组成的 $1 \times n$ 阶矩阵:

$$\mathbf{R}_j^+ = [R_1^+ \quad R_2^+ \quad \cdots \quad R_n^+] = [\max x_{i1} \quad \max x_{i2} \quad \cdots \quad \max x_{in}] \quad (11)$$

确定负理想解 \mathbf{R}_j^- , 负理想解是由 \mathbf{X}^* 每列中最小值组成的 $1 \times n$ 阶矩阵:

$$\mathbf{R}_j^- = [R_1^- \quad R_2^- \quad \cdots \quad R_n^-] = [\min x_{i1} \quad \min x_{i2} \quad \cdots \quad \min x_{in}] \quad (12)$$

计算每个应用主题域的各指标值与正理想解间的欧式距离为

$$d_i^+ = \sqrt{\sum_{j=1}^n (x_{ij} - R_j^+)^2} \quad (13)$$

计算每个应用主题域的各指标值与负理想解间的欧式距离为

$$d_i^- = \sqrt{\sum_{j=1}^n (x_{ij} - R_j^-)^2} \quad (14)$$

则各应用主题域与正负理想解的相对接近程度为

$$E_i = \frac{d_i^-}{d_i^+ + d_i^-} \quad (15)$$

2.3.6 专利价值等级的划分标准 按相对接近度的大小将专利价值进行 10 分制转换, 把最高的相对接近度定为 10, 对专利价值得分 t_j 依次折算并取整:

$$t_i = \text{int}\left(\frac{10E_i}{\max E_i}\right) \quad (16)$$

在确定应用主题域得分后, 将该得分按照专利价值划分标准进行划分, 可以得到专利价值所属区域. 专利价值的划分标准、等级评定及所属区域如表 3 所示.

表 3 专利价值等级划分

Tab. 3 Delimitation of patent value grades

专利价值 所处得分区间	专利价值等级描述	所属专利价值等级
(6, 10]	专利价值高	一级(Level I)
(3, 6]	专利价值一般	二级(Level II)
(0, 3]	专利价值低	三级(Level III)

2.4 单元项的填充

专利预测模型中的单元项指处于相应技术关键词、功效关键词和 DMC 交叉点上的专利组合. 单元项填充, 就是对具有相应技术关键词、功效关键词和 DMC 坐标点上的专利组合数量进行的聚类统计. PTFM 中单元项填充一般由技术关键词和功效关键词的共现矩阵确定, 因此本文在沿用共现矩阵进行专利聚类的基础上, 通过先遴选共现矩阵再以 DMC 筛选的方式, 填充专利预测模型中的单元项. 基于 MapReduce 计算框架的单元项填充流程如图 2 所示.

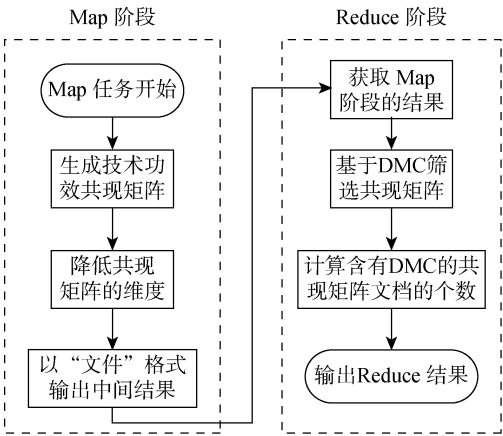


图 2 基于 MapReduce 计算框架的单元项填充流程
Fig. 2 Element item filling process based on MapReduce computing framework

2.5 专利预测模型的可视化

综合以上分析, 得到专利预测模型的技术维、功效维和价值维, 分别按照 x 轴、 y 轴和 z 轴的顺序置于三维坐标系中, 同时结合专利预测模型的单元项, 通过 Origin 2019b 软件, 可视化为三维的专利预测模型.

2.6 基于专利预测模型的技术机会挖掘步骤和标准

步骤 1 绘制专利预测地图. 专利组合是具有同一技术关键词和同一功效关键词的专利所组成的集合, 某区域内专利组合越多, 说明专利组合越密集. 以专利组合的密集程度为专利预测地图的 x 轴, 以专利价值程度为专利预测地图的 y 轴, 绘制专利预测地图.

步骤 2 挖掘优先级的技术机会. 价值原则下的优先级技术机会挖掘标准为: 在同一专利价值等级区域内比较各专利价值得分下的专利组合总数, 专利组合总数最少的专利组合所代表的技术方向即为优先级技术机会. 在一级、二级、三级专利价值区域中, 分别选择专利组合总数最少的专利组合所代

表的方向为优先级的技术机会。

步骤 3 挖掘次级的技术机会。在价值原则下的次级挖掘标准如下:比较不同的专利价值等级区域中专利组合的总数,专利价值比较高的专利组合即为次级的技术机会。也就是说,对一级和二级专利价值区域、二级和三级专利价值区域内的专利组合分别进行两两比较,选择专利组合总数相近但是专利价值更高的专利组合所代表的方向为次级技术机会。

步骤 4 对已挖掘出的技术机会排序并描述。遵循价值原则的技术机会研发顺序,即根据专利价值的高低,首先对优先级的技术机会进行研发创新,再对次级的技术机会进行研发创新,最后对所有的技术机会进行描述。

3 专利预测模型的应用:以钛领域为例

3.1 钛领域的概况

钛及钛合金因其优良的超导性、形状记忆性和生物相容性而被广泛地应用于航空、航天、航海、兵器、石油、化工、冶金、电力、医疗、制药、建筑、海洋工程、体育休闲等领域。它既是航空航天、舰船兵器、海洋工程等军工高技术领域不可或缺的关键性结构材料,也是现代医疗、制药、建筑、体育休闲等民用领域至关重要的拓展性新兴材料。目前,钛的应用量和钛工业的发展水平已经成为衡量国家实力的重要标志之一。

3.2 钛领域专利预测模型的技术维和功效维构建

3.2.1 专利源数据获取 专利有效期为 20 年,为更加全面展示钛领域的专利技术创新点和发展方向,本文在 DII 数据库按以下检索策略进行检索,如表 4 所示。

表 4 构建技术维和功效维的专利检索策略

Tab. 4 Patent search strategy in technology and function dimensions

检索项设置	检索方式
数据库名称	德温特专利检索数据库(DII 数据库)
检索式	TI=titanium
检索时间段	1999 年 1 月 1 日至 2018 年 12 月 31 日
检索时间	2019 年 3 月
检索结果	133 508

3.2.2 使用 NLTK 的源数据预处理 使用 NLTK 对专利摘要文本进行源数据预处理,并进行清洗。为使技术关键词和功效关键词抽取的更加准确,本文使用 NLTK 对钛领域专利摘要文本进行分隔以及词性标注。

3.2.3 基于 MapReduce 的技术关键词和功效关键词抽取及词频统计 基于 MapReduce 计算框架对钛领域专利预测模型的技术关键词及词频进行抽取和计算。把技术关键词抽取结果以 .xlsx 格式输出,得到钛领域专利技术关键词。为集中体现钛领域专利布局,也为使专利预测模型更直观可见,本文统计出现频率前 25 高的技术关键词,如表 5 所示。

表 5 技术关键词的抽取及词频统计

Tab. 5 Extraction and word frequency statistics of technology keywords

代号	技术关键词	词频	代号	技术关键词	词频
TK1	有色合金	968	TK14	分离式元件	263
TK2	化合物	869	TK15	聚氯乙烯聚四氟乙烯相关	263
TK3	半导体材料与工艺	705	TK16	聚合物	261
TK4	涂层材料	628	TK17	口腔相关	259
TK5	电化学储存	607	TK18	注射器	253
TK6	钢铁冶金	543	TK19	涂层抛光	249
TK7	上色抛光	473	TK20	医疗假肢	244
TK8	粉末冶金	438	TK21	集成电路	238
TK9	灭菌洗涤剂	408	TK22	多层产品	237
TK10	耐火材料	391	TK23	电解合成	231
TK11	加热工艺	383	TK24	粒子加速器相关	226
TK12	电气应用	367	TK25	金属冶金	218
TK13	电气设备	291			

同样基于 MapReduce 计算框架对钛领域专利的
功效关键词及词频进行抽取和统计. 把抽取结果
按 .xlsx 格式输出,得到基于 MapReduce 计算框架
的钛领域功效关键词的词频计算结果,本文统计前
25 个功效关键词,如表 6 所示.

表 6 功效关键词的抽取及词频统计
Tab.6 Extraction and word frequency statistics of function
keywords

代号	功效关键词	词频	代号	功效关键词	词频
FK1	成本	1 038	FK14	抗菌性	513
FK2	环保性	977	FK15	抗氧化性	496
FK3	综合性能	913	FK16	节能效果	417
FK4	热稳定性	864	FK17	离子电导率	385
FK5	抗腐蚀性	839	FK18	接触电阻	359
FK6	防水性	801	FK19	易于整合	341
FK7	强度	762	FK20	保质期	302
FK8	蚀刻速率	751	FK21	反射效率	173
FK9	进程时长	749	FK22	循环特性	159
FK10	机械强度	731	FK23	焊接性能	127
FK11	光催化效率	726	FK24	电压效果	98
FK12	强度和工作性	573	FK25	迁徙率	75
FK13	进程统一性	539			

表 7 前 25 个高频 DMC 对应的应用主题域
Tab.7 Subject areas from top 25 high frequency DMC

代号	DMC	应用主题域	代号	DMC	应用主题域
SA1	T03-A03J5	钛磁头	SA14	J04-F02	钛纳米材料
SA2	L04-C10K,U11-C05C	钛合金导电层	SA15	A12-E06A, L03-E01B5B, L03-E08B,X16-B01F1	钛酸锂电池
SA3	D11-D01H,D11-D07B,	钛洗涤剂/添加剂	SA16	L02-K02	口腔医学用钛
SA4	M24-D01A	钛合金钛热处理	SA17	J04-F02	建筑用钛
SA5	U12-A01A, L03-E08B	钛电极制造	SA18	A12-E06A, L03-E01B5B, L03-E08B,X16-B01F1	贵金属合金
SA6	L04-A01A	半导体材料	SA19	L02-K02	钛耐火材料
SA7	L03-H04D	等离子弧焊接相关	SA20	L02-C02	碳化钛陶瓷
SA8	L04-A08	粒子加速器相关	SA21	M26-B09R,M26-B09T, M26-B09Z,M26-B09M	钛等离子蚀刻技术
SA9	L01-A03C,L01-A05,L01-A03A	钛化玻璃	SA22	L02-A03	手持医学成像
SA10	D04-A01	净化水质	SA23	L02-G04	钛催化剂
SA11	S05-Y03,A12-V02	植入式医疗器械	SA24	U11-C07C3,U11-C07A1	图案化钛涂层制备
SA12	A11-B16	激光立体成形	SA25	W01-C01B8H,U13-A01	钛废气处理装置
SA13	M27-B01	钛复合材			

3.3 钛领域专利预测模型的价值维

3.3.1 专利源数据获取 为最大范围地评价专利
价值,本文进行源数据获取时选定检索期限为 20
年,使用的专利检索策略与表 4 相同,在数据下载
时,按“被引频次”降序排列,下载 1 000 项最高被引
专利.

3.3.2 基于 DMC 的应用主题域确定 提取检索
结果,统计前 25 个高频 DMC,并将 DMC 转化为应
用主题域^[41],如表 7 所示.

3.3.3 指标打分及数据采集 清华大学机械工程
学院、西安交通大学材料科学与工程学院、西北工业
大学管理学院、西北工业大学材料学院、西安理工大
学材料科学与工程学院的专家学者,宝鸡工业技术
创新有限公司、宝钛集团、西北有色金属研究院中层
以上技术经理、产品经理、市场经理以及涉钛知识产
权律师为二级指标打分. 针对钛领域应用主题域专
利价值情况,采用线下与线上相结合的发放问卷的
方式,一共发放 600 份调查问卷,回收 491 份,有效
问卷 473 份,问卷回收有效率达 78. 83%.

3.3.4 二级指标权重的计算 根据 2. 3. 2 节中构
建出的专利价值指标体系,以及式(8)和(9),计算可
得各二级指标的熵值和权重,计算结果如表 8 所示.

表 8 二级指标的熵值和权重

Tab. 8 Entropy and weights of secondary targets			
二级指标		熵值	权重
PTV ₁		0.025 0	0.167 5
PTV ₂		0.075 7	0.263 3
PTV ₃		0.094 0	0.054 7
PMV ₁		0.348 3	0.023 2
PMV ₂		0.148 3	0.113 1
PMV ₃		0.097 5	0.147 3
PMV ₄		0.093 0	0.092 3
PRV ₁		0.031 0	0.014 1
PRV ₂		0.067 4	0.044 1

3.3.5 专利价值评估 根据式(15)计算各应用主题域的相对接近度 E_i ,同时对相对接近度进行 10 分制转换并取整,得到各应用主题域专利价值的得分 t_i ,并对其排序,如表 9 所示。

表 9 应用主题域的相对接近度和专利价值得分
Tab.9 Closeness coefficient and patent value score of each subject area

代号	相对接近度	得分	排序	代号	相对接近度	得分	排序
SA1	0.701	5	13	SA14	0.791	8	2
SA2	0.587	4	16	SA15	0.778	7	3
SA3	0.245	1	25	SA16	0.722	5	8
SA4	0.537	4	18	SA17	0.396	3	21
SA5	0.624	5	15	SA18	0.544	4	17
SA6	0.663	5	14	SA19	0.713	5	10
SA7	0.754	6	5	SA20	0.341	2	23
SA8	0.764	6	4	SA21	0.735	6	7
SA9	0.293	2	24	SA22	0.748	6	6
SA10	0.715	5	9	SA23	0.416	3	20
SA11	0.873	10	1	SA24	0.360	2	22
SA12	0.711	5	12	SA25	0.712	5	11
SA13	0.463	3	19				

3.3.6 专利价值等级确定 根据表 3 可以得到钛领域应用主题域的专利价值等级,划分结果如表 10 所示。

3.4 钛领域专利预测模型的单元项填充

3.4.1 专利源数据获取 钛领域的专利预测模型中单元项填充的专利检索策略与表 4 相同。

3.4.2 基于 MapReduce 计算框架的单元项填充 根据单元项填充流程,本文以技术关键词和功效关键词生成技术功效共现矩阵,采用 MapReduce 计算

表 10 钛领域应用主题域的专利价值等级划分

Tab.10 Determination of patent value grade of each subject area for titanium field		
所属专利价值等级	应用主题域	个数
一级	植入式医疗器械,钛纳米材料,钛酸锂电池	3
二级	钛磁头,钛合金导电层,钛合金钛热处理,钛电极制造,半导体材料,等离子弧焊接相关,粒子加速器相关,净化水质,激光立体成形,口腔医学用钛,贵金属合金,钛耐火材料,钛等离子蚀刻技术,手持医学成像,钛废气处理装置	15
三级	钛洗剂/添加剂,钛化玻璃,钛复合材,建筑用钛,碳化钛陶瓷,钛催化剂,图案化钛涂层制备	7

框架对技术功效共现矩阵进行降维处理,输出包含相应技术特征词和功效特征词的专利文档,并统计专利文档的个数.此后,采用 MapReduce 计算框架,以 25 个 DMC 为基准对已筛选出的专利文档再次筛选,得到的专利文档即为目标文档,其个数即为专利预测模型的单元项。

3.5 钛领域专利预测模型的可视化

根据已确定出钛领域的技术维、功效维、价值维以及单元项填充结果,以技术维为 x 轴,功效维为 y 轴,价值维为 z 轴,结合单元项填充结果建立空间直角坐标系搭建专利预测模型,并通过 Origin 2019b 软件实现可视化.由于在 x 轴和 y 轴上使用全文字标注会导致三维专利预测模型不够清晰,所以本文使用代号表示相应坐标文字,代号与坐标文字的一一对应关系如表 5 和 6 所示.钛领域的专利预测模型如图 3 所示。

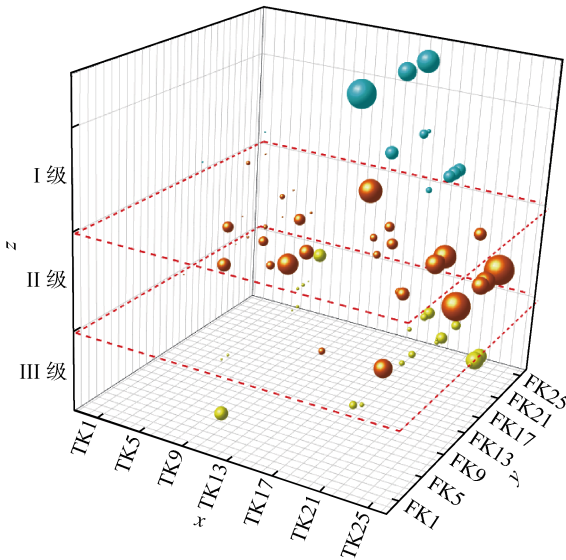


图 3 钛领域的专利预测模型
Fig. 3 Patent prediction model for titanium field

4 基于专利预测模型挖掘钛领域的技术机会

根据技术机会的挖掘步骤,结合图 3,可对钛领

域的技术机会进行挖掘。

4.1 钛领域的专利预测地图

根据图 3,绘制钛领域内的专利预测地图,如图 4 所示。

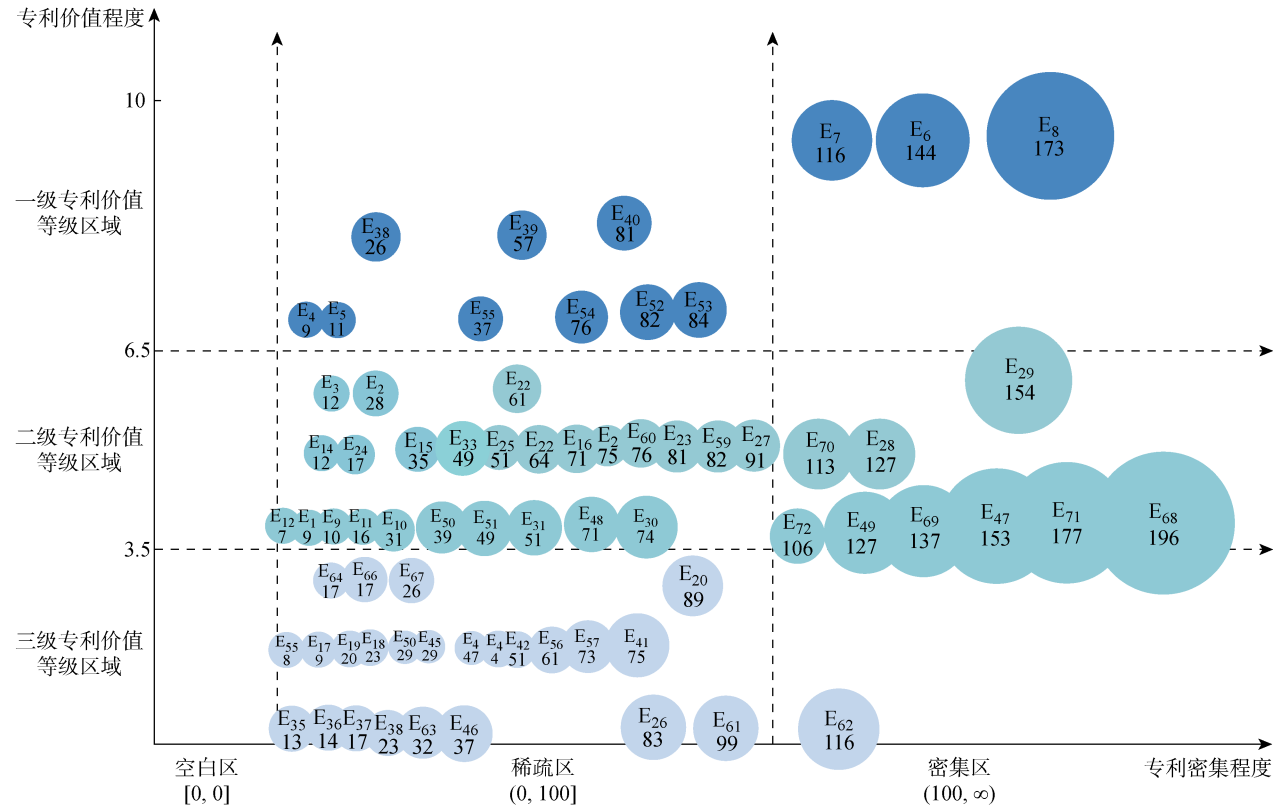


图 4 钛领域的专利预测地图

Fig. 4 Patent prediction map for titanium field

4.2 钛领域的技术机会挖掘

4.2.1 优先级的技术机会挖掘

(1) 一级专利价值等级区域中的技术机会. 该区域中专利组合有 896 个,专利价值为 10 分的专利组合有 433 个,专利价值为 8 分的专利组合有 164 个,专利价值为 7 分的专利组合有 299 个. 因此,在一级专利价值等级区域中,按照基于专利预测模型的技术机会挖掘步骤 2,标准①进行挖掘,选择专利价值得分 8 分中的专利组合为技术机会。

(2) 二级专利价值等级区域中的技术机会. 该区域中专利组合有 2 451 个,专利价值为 6 分的专利组合有 194 个,专利价值为 5 分的专利组合有 892 个,专利价值为 4 分的专利组合有 1 365 个. 因此,在二级专利价值等级区域中,按照基于专利预测模型的技术机会挖掘步骤 2,选择专利价值得分为 6 分的专利组合为技术机会。

(3) 三级专利价值等级区域中的技术机会. 该区域中专利组合有 1 057 个,专利价值为 3 分的专

利组合有 157 个,专利价值为 2 分的专利组合有 370 个,专利价值为 1 分的专利组合有 530 个. 因此,在三级专利价值等级区域中,按照基于专利预测模型的技术机会挖掘步骤 2,选择专利价值得分为 3 分的专利组合为技术机会。

综上,优先级的技术机会为得分是 8 分、6 分以及 3 分的专利组合。

4.2.2 次级的技术机会挖掘 比较不同专利价值得分区域,专利价值数量相近的专利组合为 7 分和 2 分的组合,以及 10 分和 1 分的组合. 按照基于专利预测模型的技术机会挖掘步骤 3,认为次级的技术机会为得分是 7 分和 10 分的专利组合。

4.2.3 技术机会排序并描述 按照优先级和次级的技术机会的顺序,对钛领域内的技术机会进行描述,如表 11 所示。

由表 11 可知,根据专利预测模型,在钛领域中共有 5 个技术机会,根据其优先级别给予研发顺序的建议. 其中①、②、③为优先级的技术机会,④和⑤

表 11 专利预测模型下钛领域的技术机会

Tab. 11 Technology opportunities in titanium field under patent prediction model

技术机会的优先级别	专利价值得分	钛领域技术机会的描述	研发顺序
优先级的技术机会	8	围绕钛的粉末冶金进行技术创新,同时提升其环保性能、综合性能以及进程时长	①
优先级的技术机会	6	加强钛在等离子焊接、粒子加速器和手持医学成像领域内的技术创新程度,提升其稳定性,控制成本	②
优先级的技术机会	3	加大钛在注射器和化合物中的技术创新力度,降低涉钛注射器的产品成本,提升钛化合物的抗腐蚀性、防水性以及电压效果,同时提升其光催化效率	③
次等的技术机会	10	扩大钛在医疗假肢领域的应用范围,提升产成品的成本、抗腐蚀性以及可操作性	④
次等的技术机会	7	扩大钛在电解合成以及电化学储存领域的技术创新,围绕综合性能、可操作性、环保性、热稳定性、蚀刻速率等进行功能性提升	⑤

为次级的技术机会.

事实上,把技术机会按照优先级和次级的顺序进行了排列,可以更好地为资源有限型创新主体提供导航式的研发建议.这主要是因为创新主体大多受制于资金、人员等资源,对领域内的技术机会进行全覆盖式的研发显然不现实.因此帮助创新主体在力所能及的范围内选择适合的技术机会,展开导航式的技术研发,有助于提升其技术创新效率,提高技术创新成功率.同时由于技术机会是基于已有专利展开的,所以还能有效降低专利侵权的风险,保障创新型钛领域自主知识产权主体的基本权益.

5 结语

本文以采用 PLSA 算法,基于 MapReduce 计算框架,以海量专利为样本,构建三维的专利预测模型以实现和技术机会的挖掘,并以钛领域为例进行了应用.该模型不仅能通过专利发现技术研发的热点和空白点,同时突出了专利价值对技术前景的引导功能,使研发主体能更加清晰地掌握具有应用前景的技术,明确技术研发的目标,更迅速地定位技术研发过程中的重点和难点,帮助其提高技术研发的效率,有效节约了资源成本和时间成本,深化技术机会的内涵,辅助决策者制定更加明确的技术研发战略.

本文也存在一定的局限性.在样本选取上,未来还可以考虑其他形式的专利,如国防专利,以满足国防技术创新需要.此外,在对专利预测模型的应用上,未来还可以考虑双专利预测模型,即通过比较分析方法,比较先进主体和落后主体在同一领域内的专利预测模型,通过寻找二者的差异,为落后主体挖掘潜在的技术机会.

参考文献:

[1] TRAPPEY A J C, TRAPPEY C V, CHANG A C,

et al. Deriving competitive foresight using an ontology-based patent roadmap and valuation analysis[J]. **International Journal on Semantic Web and Information Systems**, 2019, 15(2): 68-91.

[2] FRISHAMMAR J, RICHTNÉR A, BRATTSTRÖM A, *et al.* Opportunities and challenges in the new innovation landscape: Implications for innovation auditing and innovation management[J]. **European Management Journal**, 2019, 37(2): 151-164.

[3] ÖBERG C, ALEXANDER A T. The openness of open innovation in ecosystems: Integrating innovation and management literature on knowledge linkages[J]. **Journal of Innovation & Knowledge**, 2019, 4(4): 211-218.

[4] KARIA N. Knowledge resources, technology resources and competitive advantage of logistics service providers[J]. **Knowledge Management Research & Practice**, 2018, 16(3): 414-426.

[5] KUNCORO W, SURIANI W O. Achieving sustainable competitive advantage through product innovation and market driving[J]. **Asia Pacific Management Review**, 2018, 23(3): 186-192.

[6] CHEN C. Searching for intellectual turning points: Progressive knowledge domain visualization [J]. **PNAS**, 2004, 101(Sup. 1): 5303-5310.

[7] GRAFSTRÖM J. International knowledge spillovers in the wind power industry: Evidence from the European Union[J]. **Economics of Innovation and New Technology**, 2018, 27(3): 205-224.

[8] 江屏, 张瑞红, 孙建广, 等. 基于 TRIZ 的专利规避设计方法与应用[J]. **计算机集成制造系统**, 2015, 21(4): 914-923.

JIANG Ping, ZHANG Ruihong, SUN Jianguang, *et al.* Method and application of patent design around based on TRIZ[J]. **Computer Integrated Manufacturing Systems**, 2015, 21(4): 914-923.

- [9] WANG M Y, FANG S C, CHANG Y H. Exploring technological opportunities by mining the gaps between science and technology: Microalgal biofuels[J]. **Technological Forecasting and Social Change**, 2015, 92: 182-195.
- [10] JIA L Z, WU C L, ZHU X H, *et al.* Design by analogy: Achieving more patentable ideas from one creative design[J]. **Chinese Journal of Mechanical Engineering**, 2018, 31: 37.
- [11] CHENG T Y. A new method of creating technology/function matrix for systematic innovation without expert[J]. **Journal of Technology Management & Innovation**, 2012, 7(1): 118-127.
- [12] JIN G, JEONG Y, YOON B. Technology-driven roadmaps for identifying new product/market opportunities: Use of text mining and quality function deployment[J]. **Advanced Engineering Informatics**, 2015, 29(1): 126-138.
- [13] JOUNG J, KIM K. Monitoring emerging technologies for technology planning using technical keyword based analysis from patent data[J]. **Technological Forecasting and Social Change**, 2017, 114: 281-292.
- [14] 李凡, 鲁明羽, 陆玉昌. 关于文本特征抽取新方法的研究[J]. **清华大学学报(自然科学版)**, 2001, 41(7): 98-101.
LI Fan, LU Mingyu, LU Yuchang. Research about new methods of text feature extraction[J]. **Journal of Tsinghua University (Science and Technology)**, 2001, 41(7): 98-101.
- [15] 宋晓雷, 王素格, 李红霞, 等. 基于概率潜在语义分析的词汇情感倾向判别[J]. **中文信息学报**, 2011, 25(2): 89-93.
SONG Xiaolei, WANG Suge, LI Hongxia, *et al.* Word sentiment orientation discrimination based on PLSA[J]. **Journal of Chinese Information Processing**, 2011, 25(2): 89-93.
- [16] DEERWESTER S, DUMAIS S T, FURNAS G W, *et al.* Indexing by latent semantic analysis[J]. **Journal of the American Society for Information Science**, 1990, 41(6): 391-407.
- [17] HOFMANN T. Probabilistic latent semantic analysis [DB/OL]. (2013-01-23). <http://arxiv.org/abs/1301.6705>.
- [18] 李宁, 罗文娟, 庄福振, 等. 基于 MapReduce 的并行 PLSA 算法及在文本挖掘中的应用[J]. **中文信息学报**, 2015, 29(2): 79-86.
LI Ning, LUO Wenjuan, ZHUANG Fuzhen, *et al.* MapReduce based parallel probabilistic latent semantic analysis for text mining[J]. **Journal of Chinese Information Processing**, 2015, 29(2): 79-86.
- [19] HUANG J Y, HSU H T. Technology-function matrix based network analysis of cloud computing[J]. **Scientometrics**, 2017, 113(1): 17-44.
- [20] KIM G, BAE J. A novel approach to forecast promising technology through patent analysis[J]. **Technological Forecasting and Social Change**, 2017, 117: 228-237.
- [21] WANG H, CHI Y, HSIN P. Constructing patent maps using text mining to sustainably detect potential technological opportunities[J]. **Sustainability**, 2018, 10(10): 3729.
- [22] LIU K, YEN Y, KUO Y H. A quick approach to get a technology-function matrix for an interested technical topic of patents[J]. **International Journal of Arts and Commerce**, 2013, 2(6): 85-96.
- [23] TRAPPEY A J C, TRAPPEY C V, WANG T M, *et al.* Ontology-based technology function matrix for patent analysis of additive manufacturing in the dental industry[J]. **International Journal of Manufacturing Research**, 2017, 12(1): 64-82.
- [24] TARPØ M, FRIIS T, OLSEN P, *et al.* Automated reduction of statistical errors in the estimated correlation function matrix for operational modal analysis[J]. **Mechanical Systems and Signal Processing**, 2019, 132: 790-805.
- [25] PAKES A, SCHANKERMAN M. The rate of obsolescence of patents, research gestation lags, and the private rate of return to research resources[M] // R&D, patents, and productivity. Chicago: University of Chicago Press, 1984: 73-88.
- [26] REITZIG M. What determines patent value? Insights from the semiconductor industry[J]. **Research Policy**, 2003, 32(1): 13-26.
- [27] BAKKER J. The log-linear relation between patent citations and patent value[J]. **Scientometrics**, 2017, 110(2): 879-892.
- [28] BESSEN J, NEUHÄUSLER P, TURNER J L, *et al.* Trends in private patent costs and rents for publicly-traded United States firms[J]. **International Review of Law and Economics**, 2018, 56: 53-69.
- [29] HILLER R S, SAVAGE S J, WALDMAN D M. Using aggregate market data to estimate patent value: An application to United States smartphones 2010 to 2015[J]. **International Journal of Industrial Organization**, 2018, 60: 1-31.
- [30] MA J, PORTER A L. Analyzing patent topical information to identify technology pathways and potential opportunities[J]. **Scientometrics**, 2015, 102:

811-827.

[31] GINTING G, FADLINA, MESRAN, *et al.* Technical approach of TOPSIS in decision making[J]. **International Journal of Recent Trends in Engineering & Research**, 2017, 3(8): 58-64.

[32] HEGDE D, LUO H. Patent publication and the market for ideas[J]. **Management Science**, 2018, 64(2): 652-672.

[33] HARRIGAN K R, DI GUARDO M C, MARKU E. Patent value and the Tobin's Q ratio in media services[J]. **The Journal of Technology Transfer**, 2018, 43(1): 1-19.

[34] VAN RAAN A F J. Patent citations analysis and its value in research evaluation: A review and a new approach to map technology-relevant research[J]. **Journal of Data and Information Science**, 2017, 2(1): 13-50.

[35] 万小丽, 朱雪忠. 专利价值的评估指标体系及模糊综合评价[J]. **科研管理**, 2008, 29(2): 185-191.

WAN Xiaoli, ZHU Xuezhong. The indicator system and fuzzy comprehensive evaluation of patent value [J]. **Science Research Management**, 2008, 29(2): 185-191.

[36] 孙玉艳, 张文德. 企业专利量化评估系统的设计与实现[J]. **现代图书情报技术**, 2009, 25(11): 64-68.

SUN Yuyan, ZHANG Wende. Design and implementation of evaluation system of enterprise patent value [J]. **New Technology of Library and Information Service**, 2009, 25(11): 64-68.

[37] HWANG C L, YOON K. Methods for multiple attribute decision making[M]//Multiple attribute decision making. Berlin, Heidelberg: Springer, 1981: 58-191.

[38] CHEN C T. Extensions of the TOPSIS for group decision-making under fuzzy environment[J]. **Fuzzy Sets and Systems**, 2000, 114(1): 1-9.

[39] JAHANSHAHLOO G R, LOTFI F H, IZADIKHAH M. Extension of the TOPSIS method for decision-making problems with fuzzy data[J]. **Applied Mathematics and Computation**, 2006, 181(2): 1544-1551.

[40] BISWAS P, PRAMANIK S, GIRI B C. TOPSIS method for multi-attribute group decision-making under single-valued neutrosophic environment[J]. **Neural Computing and Applications**, 2016, 27(3): 727-737.

[41] Clarivate. Manual code lookup [EB/OL]. [2020-03-05]. <http://ips.clarivate.com/cgi-bin/mc/display-code.cgi>.

(本文编辑: 黄伟)