



情报杂志
Journal of Intelligence
ISSN 1002-1965, CN 61-1167/G3

《情报杂志》网络首发论文

题目：基于科技差距评估 GTM 专利地图识别的技术机会
作者：李子彪，孙可远，陈迪，聂进，鲁雪
网络首发日期：2023-03-06
引用格式：李子彪，孙可远，陈迪，聂进，鲁雪. 基于科技差距评估 GTM 专利地图识别的技术机会[J/OL]. 情报杂志.
<https://kns.cnki.net/kcms/detail/61.1167.G3.20230303.0940.020.html>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

基于科技差距评估 GTM 专利地图 识别的技术机会*

李子彪¹ 孙可远^{1,2} 陈迪¹ 聂进¹ 鲁雪¹

(1. 河北工业大学 经济与管理学院 天津 300401;

2. 天津商业大学 管理学院 天津 300133)

摘要: [研究目的] 识别具有科学可行性的技术创新机会, 对企业降低研发失败风险和制定创新战略具有重要价值。[研究方法] 提出一种融合式的识别科学可行的技术创新机会的方法。首先, 选取 GTM 专利地图对专利数据进行文本挖掘可视化表达, 识别技术空白并逆向解读; 然后, 对科学文献进行文本挖掘和 ORCLUS 聚类, 获取科学知识主题; 最后, 通过 TF-IDF 向量的余弦相似值评估潜在技术机会与科学主题的相似性, 筛选出科学可行的技术创新机会。[研究结论] 对质子交换膜燃料电池进行了实证研究, 识别出 8 项科学可行的技术创新机会, 证明了该模型的有效性, 可以为企业制定研发战略提供决策支持。

关键词: GTM 专利地图; 科技差距; 机会识别; 技术创新机会; 质子交换膜燃料电池

中图分类号: F273.1; G306

Evaluating the Technology Opportunities of GTM Patent Map Recognition Based on Science and Technology Gap

Li Zibiao¹ Sun Keyuan^{1,2} Chen Di¹ Nie Jin¹ Lu Xue¹

(1. School of Economics and Management, Hebei University of Technology, Tianjin 300401;

2. School of Management, Tianjin University of Commerce, Tianjin 300133)

Abstract: [Research purpose] Identifying technological innovation opportunities with scientific feasibility is of great value for enterprises to reduce the risk of R&D failure and formulate innovation strategies. [Research method] This paper proposes a fusion method to identify scientifically feasible technological innovation opportunities. Firstly, GTM patent map is selected to conduct text mining visual expression of patent data, identify technical gaps and reverse interpretation. Then, the scientific literature is mined by text mining and ORCLUS clustering to obtain scientific knowledge topics. Finally, the cosine similarity value of TF-IDF vector is used to evaluate the similarity between potential technological opportunities and scientific themes and scientific and feasible technological innovation opportunities are screened out. [Research conclusion] The empirical study on proton exchange membrane fuel cell has identified 8 scientific and feasible technological innovation opportunities, which proves the effectiveness of the model and can provide decision-making support for enterprises to formulate R&D strategies.

Key words: GTM patent map; technology gap; opportunity identification; technological innovation opportunities; proton exchange membrane fuel cell

基金项目: 河北省自然科学基金项目“制造业创新链“断链”风险管理及科技金融“修复”机制研究”(编号: G2021202001); 河北省教育厅人文社会科学研究重大课题攻关项目“创新生态系统发展理念引领的河北省高质量发展机制研究”(编号: ZD202004); 河北省社会科学基金重大项目“新时代河北省区域创新驱动发展机制建设研究”(编号: HB19ZD03); 河北省创新能力提升计划项目软科学研究专项“河北省钢铁产业碳中和技术创新路径研究”(编号: 225576190D)。

作者简介: 李子彪, 男, 1979 年生, 博士, 教授, 博士生导师, 研究方向: 技术创新管理; 孙可远, 男, 1993 年生, 博士研究生, 研究方向: 技术创新管理; 陈迪, 女, 1997 年生, 博士研究生, 研究方向: 技术创新管理; 聂进, 男, 1984 年生, 博士研究生, 研究方向: 技术创新管理; 鲁雪, 女, 1993 年生, 博士研究生, 研究方向: 技术创新管理。

通信作者: 孙可远

0 引言

技术机会是技术进步以增强产品功能或产品生产的一系列可能性,探索和评估技术发展的风险和机会的技术机会分析是一种技术预测,因为它旨在预测技术发展的未来情景^[1]。技术机会分析衍生的新技术思想将极大地促进企业创新的增长和成功^[2],越来越多的技术开发人员和学者投入到技术机会分析的研究中^[3-5]。

但是目前在技术机会识别的方法上仍有待引进新兴的计算机技术以提高识别结果的客观性、精确性和可行性。在定性识别技术机会方面,德尔菲法、专家经验、问卷调查和 TRIZ 具有易于验证的优点,但是操作成本高、耗时长^[6]。在量化识别方面,基于文献计量、文本挖掘和社会网络等方法已经发展成为了较为丰富和完善的体系,但是在识别结果的精确性、客观性以及可行性等方面,仍有待进一步优化。尤其是随着大数据时代的到来和机器学习技术的蓬勃发展,技术机会识别可以引入更智能的方法,例如利用文本挖掘、数据挖掘和机器学习等方法定量地、客观地对技术机会进行识别和评估,以逐步摆脱对专家知识的依赖^[7]。

因此,有必要丰富和拓展现有技术机会识别、评估与筛选的智能分析方法。围绕以往技术机会识别存在的结果呈现不够直观、解读不够客观及未对研发的难易性进行评估等不足,本文基于技术机会识别与可行性评估双重视角,提出了一套融合文本挖掘、专利地图、主题聚类、语义相似度技术的组合式分析方法,以期识别出具备科学可行性的技术创新机会。

1 理论基础

1.1 基于专利地图的技术机会识别

随着技术竞争的日趋激烈,专利地图已经成为技术挖掘的有效工具。技术地图是由技术数据信息构成,通过算法将原始数据从高维数据空间映射到低维正则网格上,地图中空点表示该点并没有与之对应的技术组合,即为技术空白点,通过对技术空白点的分析可以更为有效的识别新兴的技术创新机会的^[8]。

以往研究基于专利地图以识别技术机会的方法大致可分为三类:(1)自组织映射网络(SOM)的专利地图将专利之间的复杂关系和专利的动态模式可视化。这种方法基于无监督学习神经网络形成目标网络节点,但对技术要求较高,不能反向解释地图^[6]。(2)主成分分析(PCA),它可以用更少的变量解释更多的原始变量,但可视化程度较低,无法向后解释地图^[9]。(3)生成式拓扑映射(GTM)是一种通过贝叶斯理论将多维数据空间映射到低维潜在空间的概率方法^[10],可

视化效果优且可逆向解释技术空白。

基于专利地图的新兴技术机会挖掘的关键在于绘制科学的专利地图和技术空白的客观解释。虽然以上3种方法都能够客观科学地展现专利数据,而在技术空白的解释方面,PCA和SOM识别出的专利地图都需要依靠科研人员主观地进行解释^[11],GTM却能够自动识别地图中的空白区域且还可以通过逆映射自动客观地提取原始主题词矩阵来解释技术空白^[12]。正是由于GTM在识别空白技术的同时还可以客观的解读技术空白,故本文选取GTM专利地图法来识别质子交换膜燃料电池潜在的技术创新机会。

1.2 基于科技差距筛选技术机会

虽然专利分析是技术机会识别的代表性方法,但科学与技术知识的结合滋养了探索新技术机会的过程^[13],科学知识为技术创新提供了基础。科学研究的主要目标是探索未知领域并为公众创造知识,而不是为私营企业创造新技术和新产品,但是基础研究的成果则往往为工业界带来了技术创新的种子^[14]。由于科学发现可能具有科学家不熟悉的工业应用,而企业管理人员往往没有意识到哪些科学发现可以促进他们的技术创新^[15],因此,识别那些还未被工业界重视的基础研究一直以来都是研发人员关注的焦点。

科学与技术之间的差距则意味着技术创新的潜力,有助于探索新的技术机会^[16]。Cristian和Shen等人^[15,16]正是基于此,通过将专利文本的聚类结果与科学文献的聚类结果进行比较分析,确定了科学与技术之间的差距,并将对比结果分为四个不同的类别^[15],如图1所示。C区间由学术界的基础研究组成,但还没有被商业化,没有被企业视为专利进行开发活动,这可能意味着该区域存在着潜在的技术创新机会。本文正是结合科学界和工业界的研究差距,以这四种关系为基础,通过科学知识主题聚类对技术空白机会进行评估和筛选,以获得C区间的技术创新机会,即具有科学支撑的科学可行的技术创新机会。

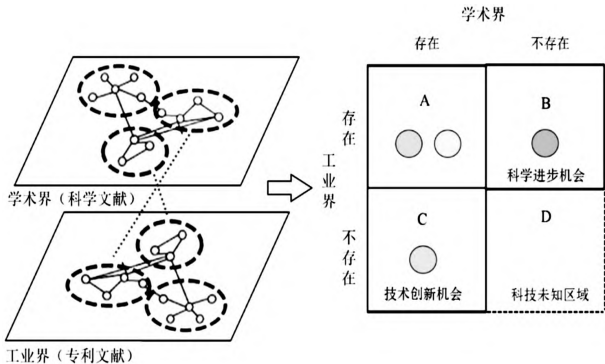


图1 基于科技差距对技术空白机会的识别筛选

1.3 文献评述与主要研究内容

现有关于技术机会的研究在分析方法、性能和规

范方面存在一些局限。第一,大多数现有方法都基于定性分析,对专家意见依赖过重。第二,虽然已有研究人员采用可视化的方法来将未开发的技术空白作为技术机会,但对其的解读依赖于专家知识、不够客观,未能深入解读技术空白的真正含义。第三,即使在定量方法中,对技术机会的识别也并不系统与完善,较少对识别结果的研发突破难易程度进行评估,继而筛选出具有科学可行的技术创新机会。第四,以往研究大多是对专利或论文单一维度数据源进行技术机会分析,缺乏对多源文献的综合运用。

为了弥合这些研究不足,本研究拟对以往识别方法进行融合优化,以发现技术创新的可行性机会。本文从技术机会识别与评估的双重视角出发,构建了一种融合 GTM 专利地图和 ORCLUS 和 tf-idf 向量的余弦相似度方法的优点的技术机会识别框架,以帮助获得科学可行的技术创新机会。首先,选取 GTM 专利地图方法对专利数据进行文本挖掘,可视化表达技术领域的专利布局,识别出现存的专利空白,并逆向解读出潜在的技术创新机会;其次,对科学文献数据进行文本挖掘,然后采用 ORCLUS 主题聚类,获得目标领域的科学知识主题;然后,通过 TF-IDF 向量的余弦值来测量潜在技术创新机会和科学知识主题的语义相似性,以评估和筛选出具有科学可行性的技术创新机会。最后,选取质子交换膜燃料电池技术的专利和论文数据进行实证检验。

2 研究设计

2.1 数据收集及预处理

选择智慧芽 (PatSnap) 专利数据库和 WOS 中的 SCI-Expanded 数据库作为文献数据检索来源。构建检索表达式获取相关数据,然后对专利文献和科学文献数据进行清洗,去除信息缺失的数据,确立初始专利数据库。

2.2 文献-关键词矩阵构建

对文献的标题和摘要字段进行关键词提取,在此之前通过分词和去停用词对数据进行预处理。本文选择词频-逆文档频率 (TF-IDF) 方法来提取专利的特征关键词,将这些特征关键词集成表征文献的语料库,进一步生成文献-关键词矩阵。

2.3 制定 GTM 专利地图并逆向解读技术机会

生成式拓扑映射 (GTM) 是根据概率分布将潜在变量转换到数据空间,使用高斯混合分布作为概率分布,以便根据变换后的潜在数据 $y(x)$ 和观察数据 t 之间的距离来确定接近度,如等式(1)。如果数据集中的数据 t 越接近 $y(x)$,则 $p(t|x)$ 的概率越高^[1]。

$$p(t|x, W, \beta) = N(y(x, W), \beta) =$$

$$\left(\frac{\beta}{2\pi}\right)^{-D/2} \exp\left\{-\frac{\beta}{2} \sum_d (t_d - y_d)(x, W)\right\} \quad (1)$$

其中, y 是转换函数, x 是潜在变量, t 是数据变量, D 是 t 的维度, $y(x, W)$ 是将 x 转换为数据集,而 b 则是噪音。

根据全概率定律,公式(2)中 t 空间中的数据分布由 x 分布的积分表示:

$$p(t|W, \beta) = \int p(t|x, W, \beta)p(x)dx \quad (2)$$

但很难推导出 $p(t)$, 因为它是连续分布的。为克服这个问题,将 δ 函数应用于离散 $p(t|x)$, 制造了所谓的 GTM 网格,并允许在预定规则网格的离散节点上定位数据^[12], 如公式(3)。

$$p(t|W, \beta) = \frac{1}{K} \sum_k p(t|x_k, W, \beta) \quad (3)$$

其中, K 是网格指针的数量, x_k 指的是潜在空间中的网格点。

权重矩阵和噪音的参数由期望最大化 (EM) 算法来估计^[12]。在将 GTM 拟合到数据集后,可以通过估计数据点的概率来将观察到的数据点标记为潜在变量的符号,如公式(4)。

$$p(x_k|t_n, W, \beta) = \frac{p(t_n, W, \beta)p(x_k)}{\sum_k p(t_n|x_k, W, \beta)p(x_k)} \quad (4)$$

最后,观察到的数据可以移到潜在空间;反之亦然,将地图中的空白技术点反向映射到原始技术空间,进而获得空白技术的解释矩阵。

$$y(x, W) = \varphi(x)W \quad (5)$$

上式中 $\varphi(x)$ 是 M 个潜在变量的固定基函数, W 为 $D \times M$ 矩阵。

综上,基于 GTM 的专利技术空白识别框架如图 2 所示。

2.4 基于 ORCLUS 聚类识别科学知识主题

对科学文献进行分组聚类的方法主要有两种,即基于引用的方法和基于文本的方法^[16]。但是 Wang 等人^[13]认为某些新兴技术可能还不发达,其基础知识库和引文网络还未建立。此外,在一篇论文或专利收到后续论文或专利的引用之前,最多可能需要两年时间^[17],因此,基于文本的聚类方法可能比基于引文的聚类方法更适合用来分析具有前沿性的科学知识主题和技术创新机会。而聚类算法方面,在 Müller 等人^[18]的评估措施中发现,PROCLUS 在几个方面要优于其他算法,但当 PROCLUS 应用于 Shen 等^[16]的研究数据时,得到的子空间的维数非常不平衡,这意味着大多数簇存在于只有两到三个维度的子空间中,一个簇占据了高维空间。因此,本文选择了 ORCLUS 聚类算法

(即 PROCLUS 算法的扩展版)对高维数据进行聚类以识别 PEMFC 技术的科学知识主题。

在获得了科学文献数据的基础上,利用术语度计算提取主题词,并生成科学文献的主题词-文档矩阵,

采用 t-SNE 算法将高维空间样本投影到二维空间,使用 ORCLUS 方法对论文的向量化矩阵进行主题聚类分析。采用 ITGInsight 软件^[19]和 R 语言中的 ORCLUS 包进行科学知识主题的聚类分析,框架如图 3:

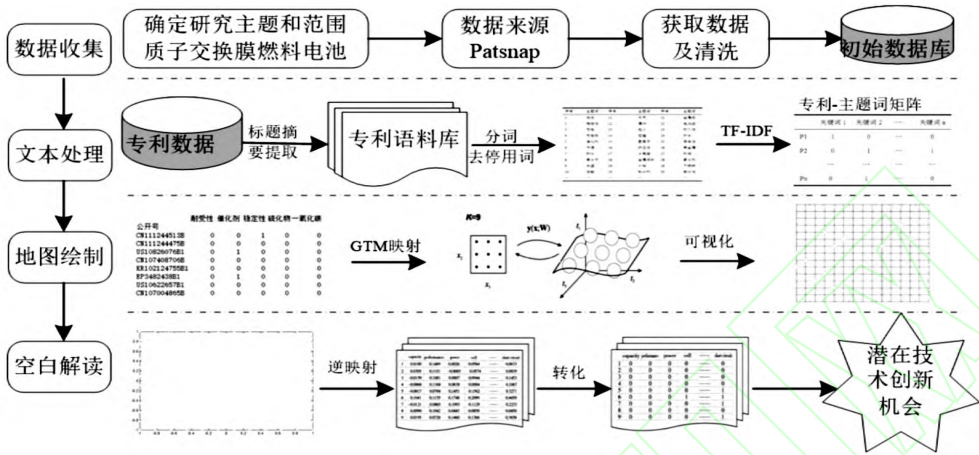


图2 基于 GTM 的潜在技术创新机会识别框架

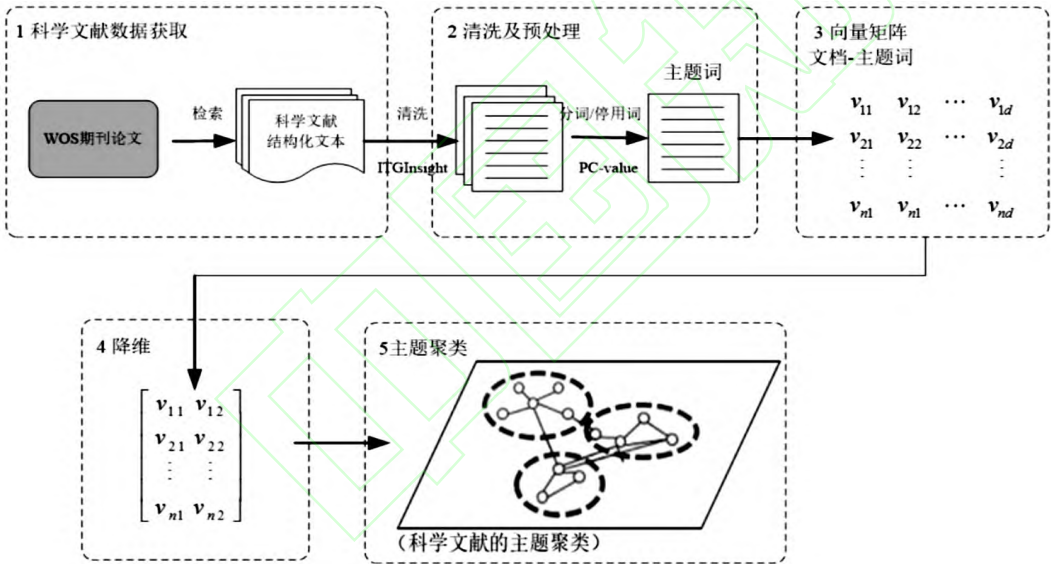


图3 基于论文的科学知识主题聚类框架

2.5 挖掘科技差距识别科学可行的技术机会

关于衡量科学和技术之间差距的方法,以往研究中 Wang 等人^[20]的研究中使用专家的主观判断来评估科学和技术之间的差距。但由于专家的判断受其领域知识的限制,且花费了大量的时间和精力。为了改善专家判断主观性的劣势,Wang 等人^[20]建议应用 Shibata 等人^[21]证明的 TF-IDF 向量的余弦相似度来评估科技之间的差距。Shibata 等人^[21]在研究中比较了 Jaccard 系数、TF-IDF 和 Log-TF-IDF 向量的余弦相似度以探索科学论文集群和专利集群之间的语义相似度,结果发现 TF-IDF 向量在检测科学集群和专利集群之间的相似性方面表现最好。因此,本文采用 TF-IDF 向量的余弦相似度来评估科学知识主题与技术创新机会之间的相似度。

TF-IDF 向量的余弦相似度假设两个簇的常用词

数越大,两个文档簇就越相似。科学集群 s 和专利集群 t 之间的余弦相似度 $\cosine(s, t)$ 用术语频率加权因子 $FreW$ 定义为:

$$\cosine(s, t) = \frac{\sum \overrightarrow{FreW_{si}} \times \overrightarrow{FreW_{ti}}}{\sqrt{(\sum \overrightarrow{FreW_{si}})^2} \sqrt{(\sum \overrightarrow{FreW_{ti}})^2}} \tag{6}$$

$$FreW_{si} = \frac{n_{si}}{n_s} \times \log\left(\frac{N_s}{N_i}\right) \tag{7}$$

$$FreW_{ti} = \frac{n_{ti}}{n_t} \times \log\left(\frac{N_t}{N_i}\right) \tag{8}$$

其中, i 是术语, s 是集群, t 是专利数据集。 n_{si} 和 n_{ti} 分别是术语 i 在集群 s 和专利数据集 t 中出现的次数。 n_s 和 n_t 分别是集群 s 和专利数据集 t 中的术语总数。 N_s 和 N_t 分别是集群 s 和专利数据集 t 中的文档总数。 N_i 是

集群 s 或专利数据集中包含术语 i 的文档数。

最后,对识别出的技术空白和科学知识主题进行语义相似度分析,评估和筛选出具有较高科学知识支撑的技术空白,即具有较高科学可行性的技术创新机会。

3 实证研究

质子交换膜燃料电池 (PEMFC) 是我国正面临被“卡脖子”困境的 35 项关键核心技术之一,识别其技术创新机会对于我国突破发达国家的技术封锁、实现技术赶超以及保障国家安全具有重大的意义。同时,质子交换膜作为燃料电池汽车的核心技术之一,对交通和环境的可持续发展以及“双碳”目标的实现均具有重要的价值。因此,本文选取质子交换膜燃料电池技术作为研究对象。

3.1 数据收集及预处理

专利检索时间范围设置为 2001–2021 年,检索日期为 2021 年 10 月 1 日,具体检索表达式为:TA: (“proton exchange membrane fuel cell” or “PEMFC” or “PEM fuel cell” not (“polymer electrolyte membrane fuel cell” or “polymer exchange membrane fuel cells” or “proton electrolyte membrane fuel cell”)),共检索到申请发明专利 11009 件,同族合并后 8602 件发明专利。科学文献共检索到 14052 篇研究论文。

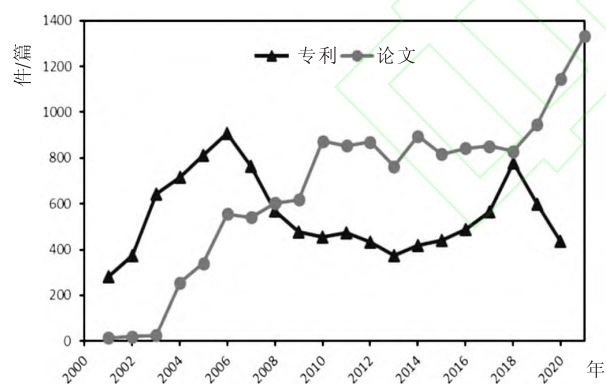


图4 PEMFC技术的发明专利申请与论文发表趋势

由图4可以看出,PEMFC的发明专利申请量在2006年达到了高峰,之后有段时间的下降趋势,说明在此期间PEMFC技术可能达到了一定的技术瓶颈期,但该趋势在2013年开始回升,并于2018年又再次达到了高峰,此时的专利申请量达到每年781件。而学术界对PEMFC技术的科学研究兴趣却一直持续高涨,尤其是“双碳”战略提出以来,学者们更加意识到PEMFC技术的价值。

选择论文和专利文本的标题和摘要字段进行主题词提取,在此之前通过分词和去停用词对文本数据进行预处理。借助文本挖掘软件 ItgInsight V2.0 提取主

题术语^[19],并根据专家知识剔除没有解释力的关键词,共提取了200个论文与专利共有主题词,生成两种主题词的词云图如图4所示。最后,使用Python软件将专利和论文文本向量化表征,生成了8602 * 200维的专利-主题词矩阵和14052 * 200维的论文-主题词矩阵。

3.2 基于GTM专利地图识别PEMFC的潜在技术创新机会

GTM模型由二维空间中的潜在点的12×12方形网格组成。该模型使用了81个高斯基函数,其中每个函数的中心位于潜在空间中的9×9方形网格上。两个网格都以潜在空间的原点为中心,基函数的共同宽度是两个相邻基函数之间最短距离的1.5倍,针对训练算法的10次迭代进行训练,控制训练期间的权重衰减程度的权重正则化因子为0.001。基于上述参数,使用MATLAB和GTM工具包开发基于GTM的专利地图^[6]。

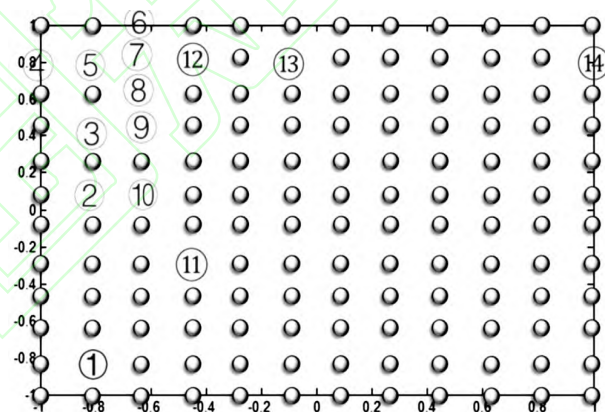


图5 PEMFC技术的GTM专利地图

由上图可知,本文共识别出14项PEMFC的技术空白。考虑到GTM的逆映射功能可通过阈值对每个主题词填充0或1来向量化表征专利地图中的空白坐标的主题内容,本文结合主题向量的方差水平,设置阈值为0.1,根据空白坐标的逆映射来解释技术空白的内容,以读取出这14项潜在技术创新机会的具体内容。

由表1对上述技术空白的逆向解读可知识别出的14项潜在技术创新机会的具体内容:技术空白T1是多嵌段共聚物的电解质膜制造,技术空白T2和T12是气体扩散层中的多孔结构与材料,技术空白T3、T4、T5、T6和T7涉及到提升催化剂的催化活性与新材料替代以提升性能的技术,技术空白T8和T9是燃料电池中双极板的保护材料制备技术,技术空白T11和T13则是电池内部的水分控制和温度控制等的辅助系统开发技术,T14是氧化还原反应中产生的CO的去除方法。

表 1 PEMFC 的潜在技术创新机会

空白	技术机会内容	代表性主题词
T1	具有磺化侧链的多嵌段共聚物作为电解质膜	电流密度、甲醇渗透率、聚苯磺酸、磺化侧链、多嵌段
T2	多孔材料的低成本制备	材料、多孔结构、稀有金属、制备、成本
T3	铂合金催化剂浆料的制备和涂布	涂层、薄膜结构、油墨、催化剂、Pt、浆料
T4	碳纳米管的电催化活性提升	碳纳米管、催化剂、碳纸、活性、提升
T5	蒸汽重整技术对催化剂比表面积	重整、碳氢化合物、表面积、催化剂、水蒸气
T6	粘结剂复合催化剂制备	催化剂、复合材料、活性、制备、粘结
T7	催化剂用稀有金属替代	纳米粒子、铂、钌、碳载体
T8	碳纳米管长线复合结构的阳极	膜电极组件、长线复合、碳纳米管、阳极
T9	耐腐蚀涂层保护的金属基板	耐腐蚀、变质、金属、双极板、涂层
T10	低成本碳载体材料	石墨、双极板、碳载体、贵金属、成本、替代
T11	控制电池内部的湿度和温度系统	含水量、内燃发动机、控制、电源
T12	具有可控孔结构的炭气凝胶	孔碳、多孔、分布结构、炭气凝胶
T13	水热管理和控制系统的序贯膜电极技术	热管理、控制系统、膜电极、序贯、氧化钛、电解水
T14	电解水制氢中 CO 的去除	电解水、CO、扩散层、还原反应、制氢

3.3 基于 ORCLUS 聚类识别 PEMFC 的科学知识主题

在 ORCLUS 聚类的实施中,四个参数需要被预先确定:最终聚类数(k),最终聚类集中的子空间维数(l),初始簇数(k₀)和每次迭代中聚类数减少的算法因子(a和a<1)。Shen 等^[16]研究发现当维数在 2 到 8 之间时,ORCLUS 表现良好,维数的最佳值为 6。应用 ORCLUS 提供的聚类稀疏系数初步确定合适的参数

设置,本研究中指定的子空间维数 l 不小于(k-2),其中指定的最终集群 k 的范围在 5 到 15 之间,指定的集群初始数量 k₀是接近计算机可以处理的最大值^[16]。通过审查科学集群内的论文标题,本研究将阈值设置为 0.001,并剔除了聚类稀疏系数高于阈值的参数设置,共聚类出 7 项 PEMFC 技术的科学知识主题,聚类结果如下表 2 所示。

表 2 PEMFC 技术的科学知识主题聚类

主题	数量	均年	主题内容	代表主题词
S1	3451	2018.1	新型质子交换膜制备	质子传输、聚合物电解质膜、复合膜、高温、质子交换膜、段共聚物
S2	2152	2015.5	气体扩散层材料	多孔材料、孔碳、甲醇渗透率、孔径分布、低成本、制备
S3	2863	2017.4	催化剂性能提升	碳纳米、Pt 催化剂、制备、表面积、碳纳米管、活性
S4	2365	2014.6	电极组件的结构与材料	甲醇穿越、电极结构、催化层、气体扩散层、金属双极板、多孔材料、膜电极
S5	1762	2013.7	双极板的制造	金属、石墨、流场、反应堆、阴极和阳极、膜结构
S6	897	2016.10	控制与辅助系统	控制器、电压控制、冷却水、热交换、加湿器、蒸汽重整
S7	562	2016.2	燃料电池应用	手机、电脑、便携电源、燃料电池车、发电厂

由表 2 可知,围绕质子交换膜燃料电池的科学研究主要聚类为新型质子交换膜的材料研发与制备、气体扩散层的多孔材料制备、催化剂中的贵金属替代与提高催化性能、膜电极组件的结构与新型材料、双极板的结构与制造、氧化还原反应中控制温度与湿度的系统的优化、燃料电池技术的新型应用场景 7 大科学知识主题。

3.4 基于科技差距对 PEMFC 技术创新机会的评估与筛选

对上文识别出的潜在技术创新机会与科学知识主题进行余弦相似度分析,以评估这些潜在技术创新机会的科学可行性。科学知识与技术机会的主题相似度过高表明该技术创新机会越易实现;相反,科技之间的余弦相似度越低,该技术越难以突破。为了有效筛选出具备科学可行性的技术机会,有必要建立一个阈

值以消除不太相似的关系,本文选择相似值(0.500)作为阈值,计算结果绘制成图 6。图中虚线连接的是实现技术创新所需的科学知识主题集。

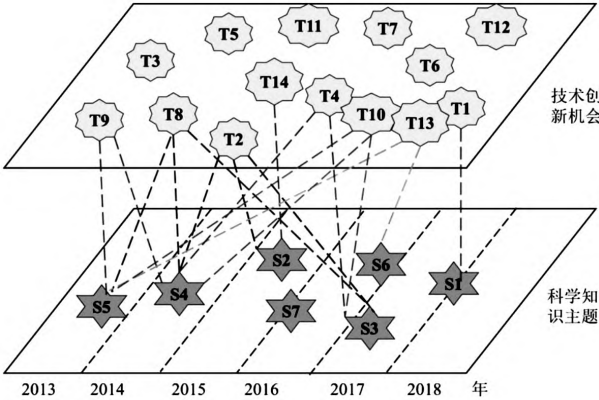


图 6 技术创新机会与科学知识主题的语义关联筛选图
由上述对潜在技术创新机会的科学可行性评估结

果可知,技术机会 T1、T2、T4、T8、T9、T10、T13 和 T14 具有较高的科学知识基础,是相对容易突破的技术空白,因此筛选出这 8 项科学可行的技术创新机会。最

后,针对这 8 项技术创新机会按照主题相似度均值和科学主题平均年限排序,可知 T1 和 T2 技术不仅较容易研发突破而且在科学研究中也具备较高新颖性。

表 3 PEMFC 领域技术创新机会的评估与筛选结果

技术机会	技术机会内容	相似度 均值	创新 难度	科学主题 年限均值	技术 新颖性
T1	具有磺化侧链的多嵌段共聚物的电解质膜	0.728	容易	2018.1	很新
T4	碳纳米管的电催化活性提升	0.657	容易	2015.11	较新
T13	水热管理和控制系统的序贯膜电极技术	0.619	容易	2014.10	一般
T9	耐腐蚀涂层保护的金属基板	0.617	容易	2014.1	一般
T8	碳纳米管长线复合结构的阳极	0.613	容易	2015.2	较新
T2	多孔材料的低成本制备	0.608	容易	2015.9	较新
T10	低成本碳载体材料	0.573	较易	2015.2	较新
T14	电解水制氢中 CO 的去除	0.502	较易	2015.5	较新

3.5 技术创新机会的消融实验检验

为验证上述融合方法的有效性,本文收集 PEMFC 技术 2021–2022 年的发明专利申请数据,由 PatSnap 专利分析系统生成 PEMFC 的专利地图,对上述识别结果进行后验检验。由图 7 可知,在 GTM 识别出的 14 项技术机会中仅有 T1、T2、T3、T4、T8、T9、T11、T13 八得到了验证,所以仅 GTM 方法识别结果的有效率为 57.1%。而融合技术差距筛选后的技术机会中有 75%的结果得到了验证。可见,本文的提出的融合 GTM 专利地图、技术差距评估与筛选的方法能有效提高技术机会识别的准确性,且有助于企业降低创新失败的风险。

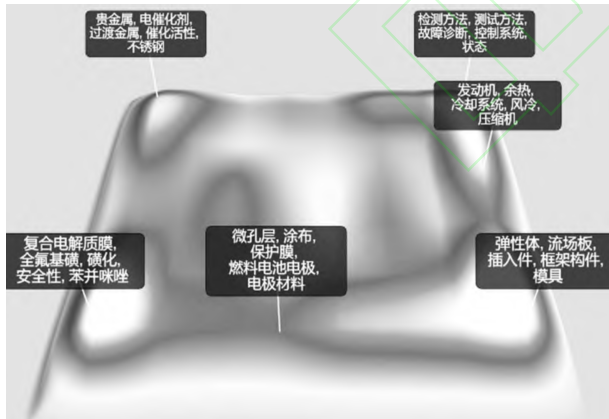


图 7 PEMFC 技术验证数据集的专利地图

4 结论与不足

围绕以往技术机会识别研究中存在的结果呈现不够直观、解读不够客观以及未对识别结果的可行性进行评估等不足,本文构建了一套融合文本挖掘、GTM 拓扑映射、ORCLUS 聚类 and 语义相似度的方法优势来识别科学可行的技术创新机会的方法。首先,对专利数据进行文本挖掘生成了 GTM 专利地图,并逆向解释了地图中的空白,获得了潜在技术创新机会;其次,

对已有论文数据进行 ORCLUS 主题聚类;然后,通过 TF-IDF 向量的余弦值测度科学知识与技术机会的语义相似度,评估了识别结果的可行性,筛选出了 8 项科学可行的 PEMFC 技术创新机会。

本文提出的融合方法减少了从专利数据中识别技术创新机会的主观性,增强了识别结果的科学可行性,能有效降低企业创新过程中的失败风险。本研究的创新性贡献主要有:①通过文本挖掘与 GTM 拓扑映射技术构建了更加直观和易于解读的专利地图,地图中的空白区域可被视为新的技术机会,该方法更加客观高效。②通过 ORCLUS 聚类和语义相似度分析提出了一套识别科技差距来评估技术创新机会可行性的量化方法。③基于组合式创新提出融合 GTM 专利地图和科学技术间语义相似度优势的组合格架,并将该方法新应用于质子交换膜燃料电池技术。

尽管本研究提出的方法是有效的,但仍存在部分不足:首先,GTM 方法对参数设置非常敏感,如果设置不正确,可能会识别出不恰当的空白技术。然后,ORCLUS 聚类结果的解读仍受到专家知识主观性的影响。未来可以尝试其他更加优化的方法,以在没有专家参与的情况下客观的解读技术创新机会。

参 考 文 献

[1] Yoon B, Magee C L. Exploring technology opportunities by visualizing patent information based on generative topographic mapping and link prediction[J]. Technological Forecasting and Social Change, 2018, 132: 105–117.

[2] Cho C, Yoon B, Coh B Y, et al. An empirical analysis on purposes, drivers and activities of technology opportunity discovery: the case of Korean SMEs in the manufacturing sector[J]. R&D Management, 2016, 46(1): 13–35.

[3] Lee M, Lee S. Identifying new business opportunities from competitor intelligence: An integrated use of patent and trademark databases[J]. Technological Forecasting and Social Change,

- 2017, 119: 170–183.
- [4] Choi Y, Hong S. Qualitative and quantitative analysis of patent data in nanomedicine for bridging the gap between research activities and practical applications[J]. *World Patent Information*, 2020, 60: 101943.
- [5] Jones R J, Barnir A. Properties of opportunity creation and discovery: Comparing variation in contexts of innovativeness[J]. *Technovation*, 2019, 79: 1–10.
- [6] Teng F, Sun Y, Chen F, et al. Technology opportunity discovery of proton exchange membrane fuel cells based on generative topographic mapping[J]. *Technological Forecasting and Social Change*, 2021, 169: 120859.
- [7] Park Y, Yoon J. Application technology opportunity discovery from technology portfolios: Use of patent classification and collaborative filtering[J]. *Technological Forecasting & Social Change*, 2017, 118(5): 170–183.
- [8] 许学国, 桂美增. 基于 GTM 逆向映射的技术创新机会识别—以新能源汽车为例[J]. *情报理论与实践*, 2021, 44(6): 146–153, 198.
- [9] Aaldering L J, Leker J, Song C H. Analysis of technological knowledge stock and prediction of its future development potential: The case of lithium-ion batteries[J]. *Journal of Cleaner Production*, 2019, 223(20): 301–311.
- [10] 王雪原, 金秀丽. “技术-市场”三级技术创新逻辑图谱设计: 以智能制造为例[J]. *情报杂志*, 2022, 41(7): 197–207.
- [11] 吴菲菲, 米兰, 黄鲁成. 以技术标准为导向的企业研发方向识别与评估[J]. *科学学研究*, 2018, 36(10): 1837–1847.
- [12] Son C, Suh Y, Jeon J, et al. Development of a GTM-based patent map for identifying patent vacuums[J]. *Expert Systems with Applications*, 2012, 39(3): 2489–2500.
- [13] 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(3): 182–195.
- [14] Ogawa T, Kajikawa Y. Assessing the industrial opportunity of academic research with patent relatedness: A case study on polymer electrolyte fuel cells[J]. *Technological Forecasting and Social Change*, 2015, 90: 469–475.
- [15] Mejia C, Kajikawa Y. Emerging topics in energy storage based on a large-scale analysis of academic articles and patents[J]. *Applied Energy*, 2020, 263: 114625.
- [16] Shen Y-C, Wang M-Y, Yang Y-C. Discovering the potential opportunities of scientific advancement and technological innovation: A case study of smart health monitoring technology[J]. *Technological Forecasting and Social Change*, 2020, 160: 120225.
- [17] Shibata N, Kajikawa R, Sakata R. Extracting the commercialization gap between science and technology — Case study of a solar cell[J]. *Technological Forecasting and Social Change*, 2010, 77(7): 1147–1155.
- [18] Müller E, Günnemann S, Assent I, et al. Evaluating clustering in subspace projections of high dimensional data[J]. *Proceedings of the VLDB endowment*, 2009, 2(1): 1270 – 1281.
- [19] Wang X, Zhang S, Liu Y. ITGInsight - discovering and visualizing research fronts in the scientific literature[J]. *Scientometrics*, 2021, 12(1): 6096.
- [20] 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(3): 182–195.
- [21] Shibata N, Kajikawa Y, Sakata I. Detecting potential technological fronts by comparing scientific papers and patents[J]. *Forecasting*, 2011, 13(5): 51–60.