**专利技术功效矩阵构建与应用的研究进展**

马昕远

(南京工业大学经济与管理学院)

摘 要：【目的】 梳理了近年来技术功效矩阵的构建和应用的发展，综合分析目前存在的问题，提出针对性建议。 【结果/结论】 技术功效图智能构建目前水平是以人工辅助下的半自动化构建;技术功效分类精度不高;可视化功能简单。需要通过综合运用规则和统计相结合的自然语言处理技术提高抽取分类精度;加强信息可视化交互功能和跨语言技术功效图构建研究;探索跨领域技术借鉴与转移应用方案;利用SAOx结构提取技术功效词可能更好地构建技术功效矩阵;应用大语言模型会更好地解决现存问题。

关键词：技术功效矩阵；专利分析；数据挖掘；文本挖掘；研究进展

专利文献是一种特种文献，是记载专利申请、审査、批准过程中所产生的各种有关文件的文件资料；具有内容新颖、文字严谨、实用性强、形式统一等特点。因此，通过专利文献可以了解某一领域内技术发展趋势、当前竞争情况，发掘技术的热点、空白点，为企业指明未来发展方向。

专利地图是专利分析的方法，专利地图是通过对专利文献进行收集、整理、加工、预处理、分析、挖掘之后以可视化的方式方便直观地展示分析的结果，决策者可以在此基础上做出判断与决策。专利地图是当前进行专利分析的主要手段，特点是直观、简洁明了。

专利技术功效矩阵是专利地图的一种，是常见的专利定量分析方法；通过识别专利文献中的技术词、功效词，绘制成可视化图表进行分析。专利技术功效矩阵可以直观地反馈当前领域内技术和功效的类别与分布情况，帮助决策者识别“专利雷区”、“专利空白区”，识别领域内技术的热点、空白点[1]，可以为企业专利布局、技术研发、抢占市场等工作提供参考。

早期的专利技术功效矩阵构建过程比较复杂，需要首先由人工判断专利文献中的技术词、功效词，成本高、研究周期长，判断结果的好坏依赖于领域专家的知识储备，主观性强。随着大数据技术、知识图谱、机器学习、深度学习、自然语言处理、数据可视化等技术的快速发展，半自动或自动化地构建技术功效矩阵已成为可能。技术功效矩阵的元素也不只仅限于技术词和功效词，逐渐形成了IPC-技术功效短语、技术-功效-时间三维矩阵等多类型的技术功效矩阵，并将通过这些方法构建的矩阵应用于领域研究中。

本文介绍了专利技术功效矩阵的基本知识，梳理当前主要的技术功效矩阵的构建方法和应用过程，提出下一步专利技术功效矩阵的构建与应用的建议。

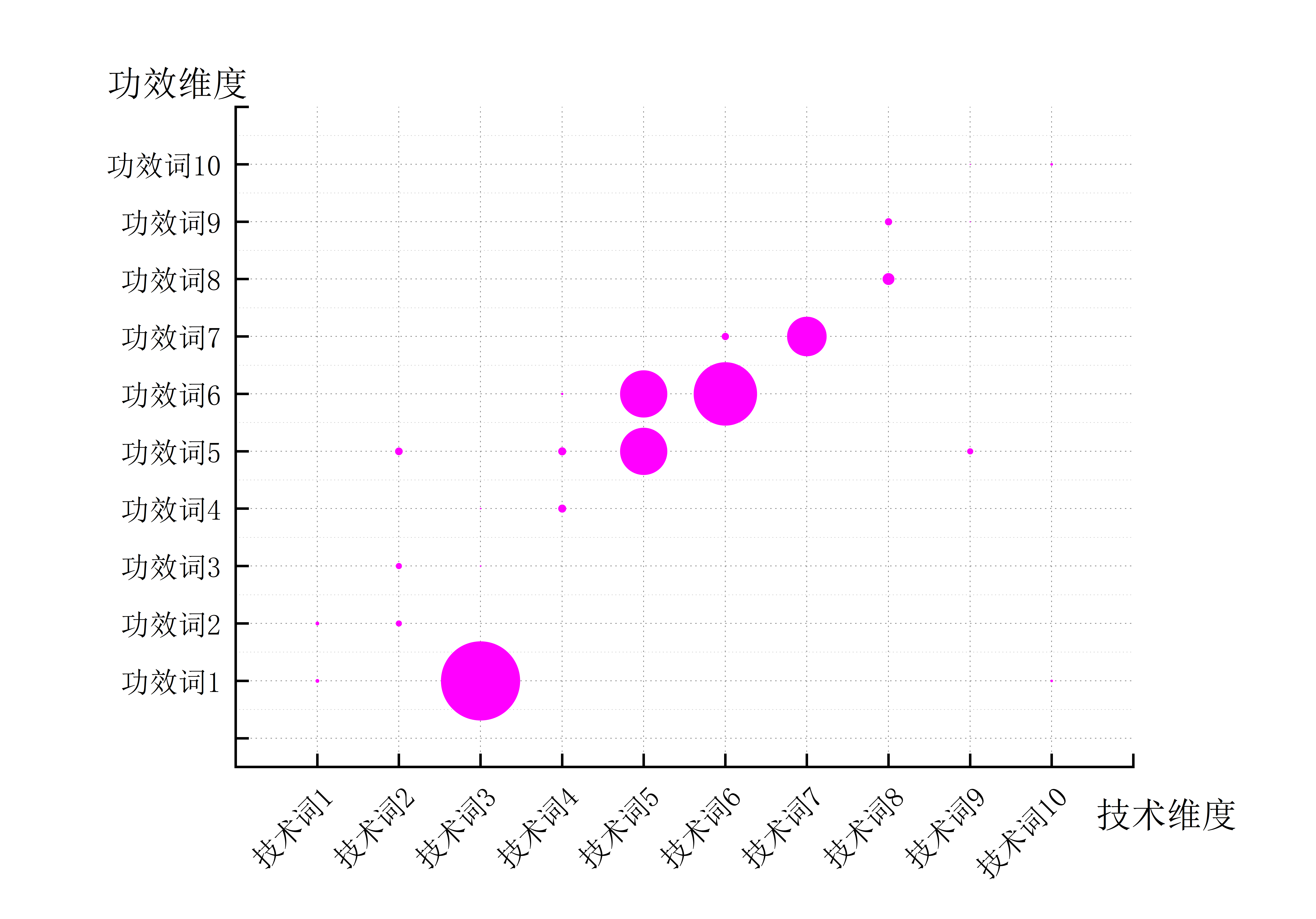
1. **专利技术功效矩的定义**

专利技术功效矩阵一般是同时包括“技术”和“功效”两种元素而构成的二维专利地图。目前对技术功效矩阵还没有明确的定义；专利技术功效矩阵也被称为技术功效矩阵（文中以下内容均称为技术功效矩阵），分析对象一般为专利文献，其他类型的文献大多没有明确描述功效，所以不适合用技术功效矩阵做相关工作。

技术词和功效词是技术功效矩阵最基本的元素，对专利文献的内容进行分析，统计技术词和功效词以及专利数量，由此构建的包含技术维度和功效维度的专利地图被称为技术功效矩阵。有观点认为只要包含技术词、功效词二者之一就可以称为技术功效矩阵[2]；也有观点认为只要包括技术功效短语就可以称为技术功效矩阵。

技术功效矩阵包括技术维度和功效维度，两个坐标轴的交点通常用专利的数量表示，可以用圆形代表专利数量，圆形越大专利数量越多；据此可以绘制出简单的技术功效矩阵图，如图1所示。

技术功效图是指对专利文献进行技术类别、功效类别标引，统计任意技术类功效类交点处的专利数量，并用可视化技术呈现统计结果，至少同时包含技术功效两种元素的专利地图。



**图1 简单的技术功效矩阵**

技术功效矩阵的技术维度可以是技术词、方法、专利分类号、IPC、材料、产品、原理等内容，并且可以在多个维度上扩展。功效维度可以是功效词、结果、效果、技术方案能解决的问题等内容进行分类。交点处可以用圆形代表专利数量；还可以增加第三维度进行详细分析。

1. **技术功效矩阵构建与应用的发展**

近年来学者们广泛采用新的方法构建技术功效矩阵，如采用新兴自然语言处理技术构建、增加维度构建三维技术功效矩阵等，并结合领域实例来论证。技术功效矩阵的构建与应用是相辅相成的过程，为了实现新的应用促进新的构建方式的产生，新的构建方式的成功实现开辟新领域的应用，为之后的研究提供借鉴。

**2.1 技术功效矩阵的构建发展**

构建技术功效矩阵的方法有很多种，从传统的人工方法到自动化半自动化的构建方法都存在类似的主要的步骤：确定研究主题、检索专利文献、分析专利内容、提取技术功效维度、技术词/功效词统计、构建技术功效矩阵，基本构建过程如图2所示；在提取、分析技术词和功效词的过程中往往也需要加入领域专家的判断，该过程存在判断结果有主观性、依赖领域专家知识储备的缺陷。



**图2 技术功效矩阵基本构建过程**

随着数据挖掘、自然语言处理、机器学习、模式匹配等技术的发展，如何在机器辅助下

实现技术功效图的自动或半自动化构建从而减少对人主观判断的依赖引起了学者们的重视。

对于技术功效矩阵构建方法的改进集中在如何高效获取准确的技术词、功效词；以及如

何处理大规模数据以提高筛选词语的效率。

过度依赖专家学者来判断技术词、功效词的过程比较耗费人力，结果也不够全面；针对该情况翟东升等[3]提出以文本挖掘理论为基础，结合专利数据库与专家辅助参与，用TLA方法提取高质量的技术词、功效词，并绘制技术功效矩阵。该方法减少了构建过程中过度依赖人工判断，缺点是数据来源格式要求较严格，可能会忽略部分技术词、功效词。

为了能够大幅度提高传统功效矩阵分析的工作效率，实现专利微观分析工作的自动化与工具化。王丽等[4]学者构建了标引功效图的自动化工具，通过词库构建、主题标引、文本提取等步骤对专利文本进行挖掘和分析，实现技术功效矩阵的自动生成最后结合CMOS工艺技术进行实例验证。

由于技术功效矩阵多由人工完成的现状,陈颖、张晓林等[5]提出一种基于特征度指标和矩阵构建词汇模型的矩阵结构生成方法。特征度指标的作用是增加候选技术词和功效词之间的相关性，以改进构建矩阵结构的过程。而矩阵构建词汇模型则用于优化技术词和功效词的聚类，并生成更合理的矩阵结构。

随着专利数量的激增技术功效矩阵的构建面临着处理海量数据低效的问题，翟东升等[6] 引入MapＲeduce分布式计算的概念，通过集群系统解决了传统方法处理海量数据低效的问题，选择触屏技术作为研究对象，演示了整个技术功效矩阵的构建过程。该方法解决了构建技术功效矩阵在海量数据条件下的低效问题，但抽取的特征词的精度不高仍需要较多的人力参与后续的筛选工作。

对于传统人工构建技术效率低下以及过度依赖专家主观性判断的问题，张兆锋等[7]提出了基于知识图谱的自动构建解决方案，并以新能源汽车领域为实例论证，该方案实现技术功效图的自动构建，减少了构建时间和成本。虽然该方案没有人工判断的精确程度高，但已经基本达到工程化的应用标准；该方案的缺点是自动构建结果的好坏依赖于汉语科技词系统术语及术语关系的丰富程度，因此下一步的研究将集中于如何更多地从专利文献本身抽取词间关系构建知识图谱进而绘制技术功效矩阵。

专利的摘要和说明说中都蕴含着特定语言结构，主语(S)谓语(A)宾语(O)结构全称为Subject-Action-Object结构，该结构可以清晰地描述专利组成部分之间的关系[8]。据此原理段庆锋、蒋保健[9]提出了基于SAO结构的技术功效矩阵构建方法并以石墨烯传感领域为例进行实例分析，该方法相较于全文，SAO结构有更细粒度的结构单元，对于技术维度、功效维度的构建结果更加细致，扩展后可应用于其它领域创新分析应用，通过该方法进行SAO结构语义分析存在技术语义表征模糊、语义相似度计算准确性不高的问题。基于此问题，翟东升等[10]提出基于SAO结构和词向量的方法，能够从海量的专利数据中智能化凝练出技术主题和功效主题并识别技术词和功效词，提高了技术功效矩阵的构建效率；并采用无人驾驶汽车领域毫米波雷达技术作为实证进行研究。

以上均方法是通过研究新的构建方法再辅以实例论证，开辟了技术功效矩阵在新领域的研究，为领域决策者提供参考依据。

* 1. **技术功效矩阵的应用发展**

技术功效矩阵能反映出领域内技术热点、空白点，刘化然等[11]分析构建出的二维技术功效矩阵提出了专利布局难易程度的计算方法，并结合技术种类、功效种类、专利申请的数量对特定领域做技术机会分析，将领域内分为专利壁垒区、专利潜力区、专利活跃区、专利沉寂区，并建议专利布局应优先考虑专利潜力区和专利活跃区。

掌握领域内重点研发方向的形势对于下一步的研发工作至关重要，余梦霞等[12]通过构建细粒度技术功效矩阵揭示了第三代功率半导体的重点研发方向为“提高击穿电压、提高器件可靠性、降低成本、降低 导通电阻、降低损耗”等内容，并全方位地对改进方向、技术布局、国内外现状做了细致分析。刘月等[13]依靠技术功效矩阵探究肿瘤免疫细胞领域新药研发的技术现状以及识别技术机会，得出技术热点集中在 A61P35，C12N5等领域的结论，并依靠技术功效矩阵对肿瘤免疫细胞的发展提出建议。

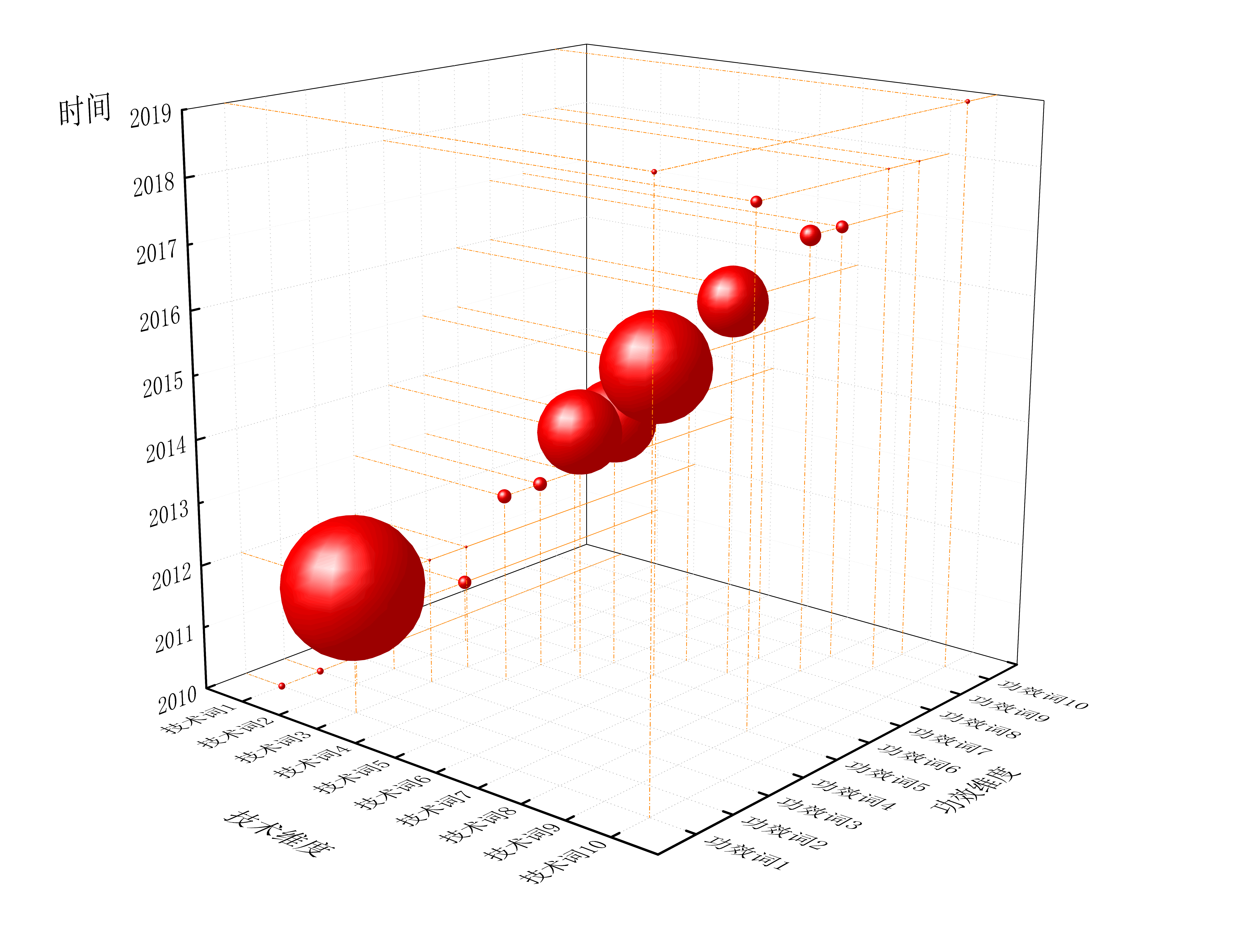
现代社会技术、经济高速发展的背景下，造成产品更新迭代速度加快，因此精准把握用户需求以引导企业高效开展产品创新，对企业实现可持续发展至关重要，刘鹏等[14] 将技术功效矩阵与多维技术创新地图融合并以重点用户需求导航构建的需求导向下的三级技术功效矩阵识别企业创新的机会。

在记录技术创新信息的各类文献中，专利有着数量可观，记录技术信息真实详细，新颖可靠以及内容格式较为标准化等优点，记载着80%以上的技术相关知识[15]。专利文献分为发明专利、实用新型专利、外观设计专利，保护期限分别是20年、10年、10年；因此通过构建技术维度、功效维度和时间维度相结合的三维技术功效矩阵进一步分析领域内的情况是可行的，简单的增加时间维度的三维技术功效矩阵如图3所示；加入时间的三维技术功效矩阵可以拆解成技术-时间、技术-功效、时间-技术的二维技术功效矩阵，弥补了单纯二维技术功效矩阵的不能从时间角度全面考察领域内技术发展、功效演变的缺点。

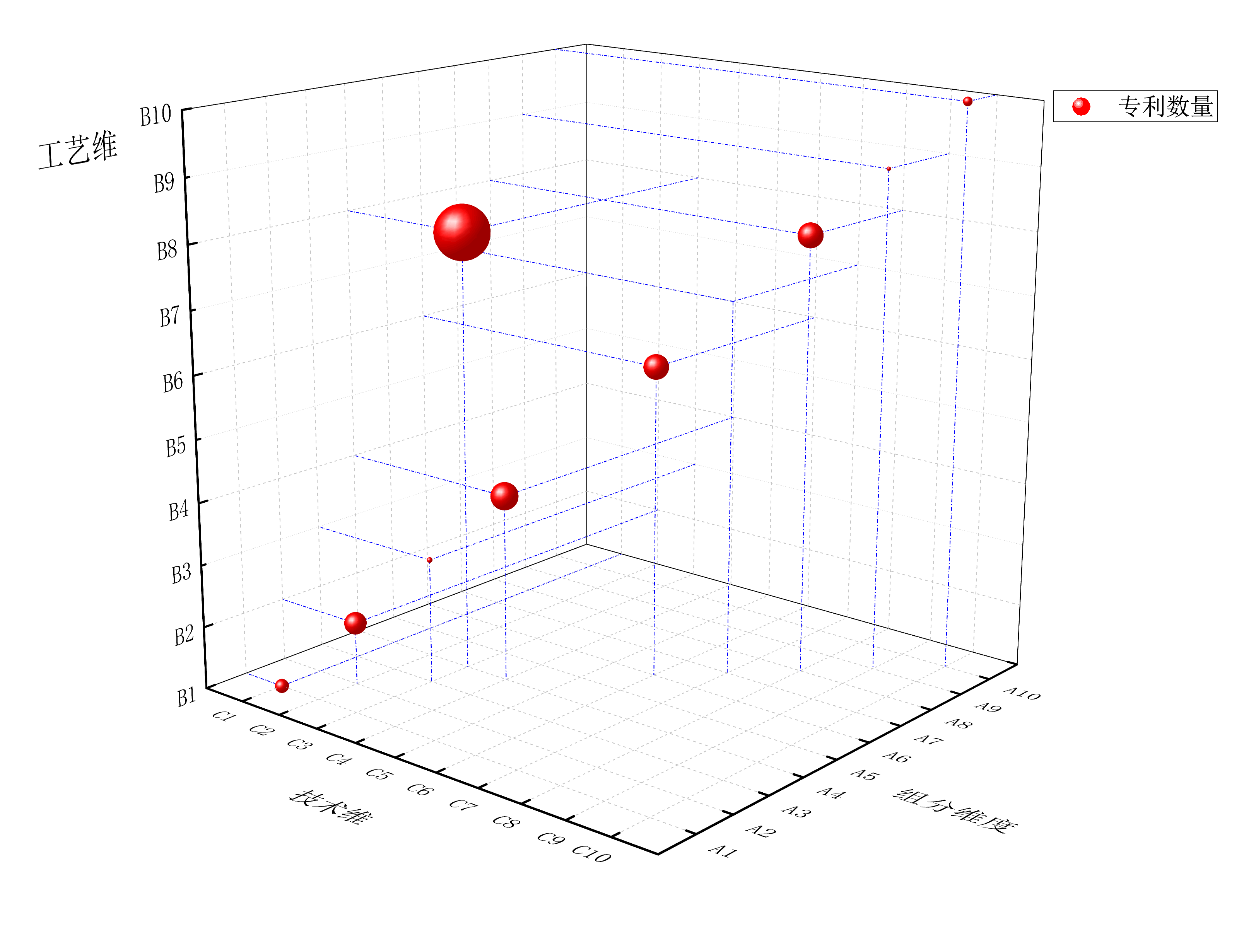
通过建立三维技术功效矩阵分析模型不仅可以体现专利技术与矩阵的映射，还能从多个维度全面展示技术的发展情况，相关工作如朱琳等[16]通过构建增加时间维度的三维技术功效矩阵全面分析豆浆机领域技术研发的技术成熟度和延续情况，更精准地识别出企业创新的机会如：通过改善电机来减震降噪、改善电机来提升粉碎效果等。张浩、张云秋[17]利用融合技术-功效-时间的三维技术功效矩阵识别中药饮片炮制领域的研究热点、重点。

技术功效矩阵还可以结合其他维度进行细致地分析，药学领域中可以融合技术、组分、工艺，分析药物研发可能存在的；相关工作如王巍洁等[18]为了分析吉林省人参产业情况构建了融合技术、组分、工艺维度的技术功效矩阵模型，如图4所示，识别出在功效上以主要应用于抗肿瘤、抗疲劳、保健与提高免疫力等临床范围为基础继续投入研发，进行进一步的研究、开发与拓展。

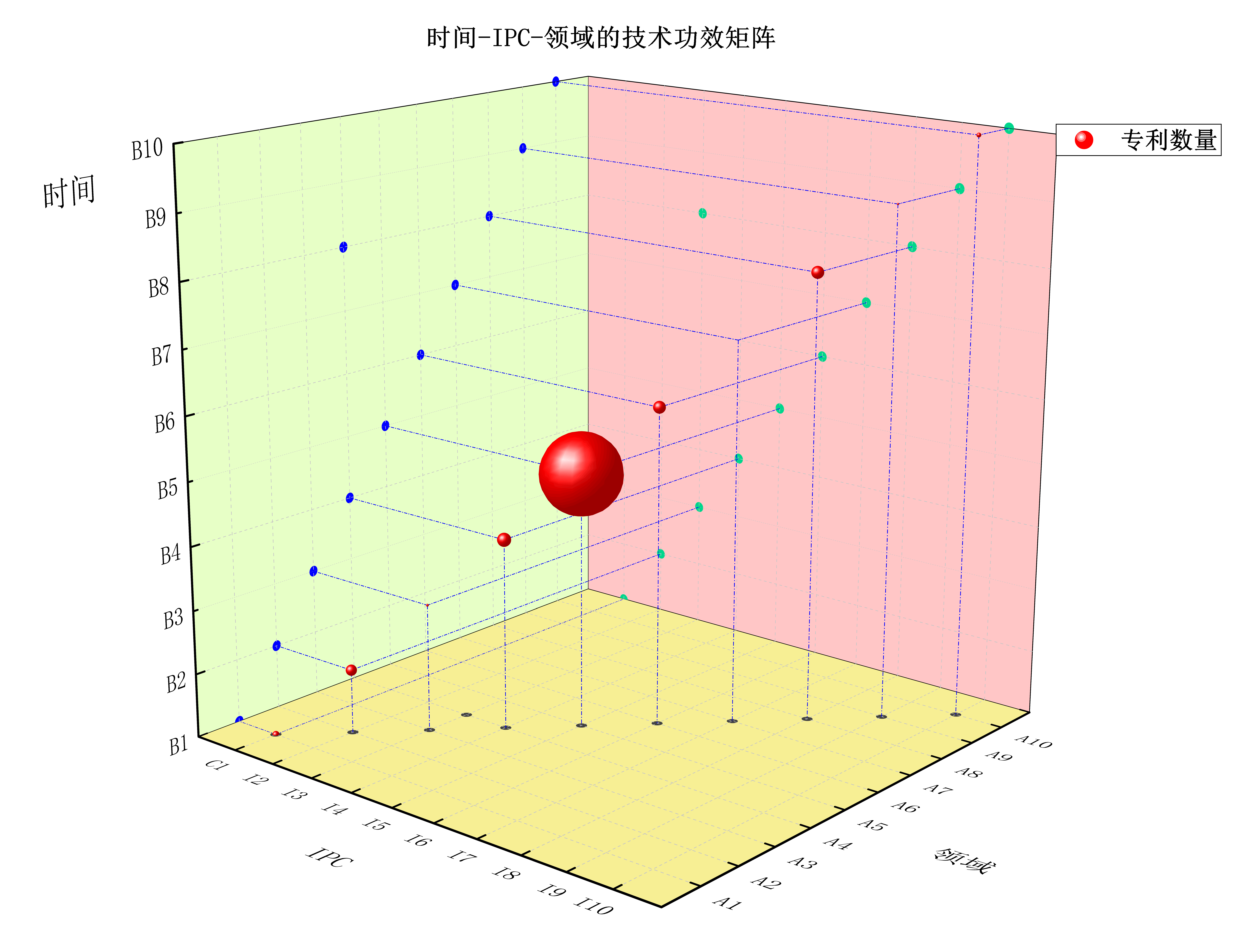
IPC分类号由粗到细分成了部、大类、小类、主组、分组，正是由于IPC的特点可以选择某一个IPC分类号由粗到细构建技术-功效-IPC三维矩阵进行分析；构建领域-IPC分类号-时间的三维矩阵模型可以判断领域内重点技术、空白技术，以及各类技术都集中在哪类专利文献中，如图5所示，交点球体表示专利数量；相关工作有房廷秀等[19]构建融合IPC-领域-时间的三维技术功效矩阵并结合人参领域的研究得出A61K36/8988组成的药物在治疗神经精神疾病方面具有潜在技术热点，如抑郁症，可以在该方向投入更多研究等结论。



**图3 三维技术功效矩阵示意图**



**图4 技术-组分-工艺的三维技术功效矩阵模型**



**图5 IPC-领域-时间的三维技术功效矩阵模型**

1. **当前研究**

**3.1 “供应链+大语言模型”的研究模式**

专利文献分析的结果最好体现就是市场环境中。供应链通俗地说,即是以受市场认可和欢迎、技术实力强、产品服务相互依赖的企业和优质产品服务为核心,以技术和资本为依托形成的价值创造活动链,从上游原材料至下游功能产品整体而言是一个价值增值的过程[20]。

围绕供应链技术进行专利分析或者技术路径分析已成为当前诸多学者研究的重点，近年来几乎所有与专利信息分析的研究都会结合供应链，这是一个研究热点。张杰等[21]构建供应链模型,基于改进的LDA方法实现供应链环节–技术主题–专利映射，有效识别了微藻生物燃料供应链重点及热点环节,该供应链技术创新需多环节协同；缺点是在使用过程中对于主题的判断需要人工完成，这个过程有主观性的缺陷。

大语言模型已经逐渐成熟，生成式语言模型ChatGPT掀起了人工智能正式取代人工的浪潮，相似的模型同样可以应用在专利信息分析的各种手段中，如果在专利信息分析的过程中加入BERT、GPT这样的大语言模型，将能更好地避免领域专家主观性判断造成的结果不准确。

* 1. **技术思路建议**

1. 分析确定供应链：通过查阅领域内专利文献并咨询领域专家，确定供应链以及供应链各环节的各项活动。
2. 检索专利文献：构建正确的检索式，检索相关专利文献，因为需要检索领域内大量文献，不能针对某一主题构建检索式，而应该更关注查全率。
3. 划分供应链各环节和活动并根据主题模型实现专利文献和供应链各环节各活动的匹配，利用大语言模型的效果更好地实现该任务。
4. 分析供应链上游、中游、下游的技术机会。
5. 对各个层次进行技术机会识别或其他专利信息分析过程如申请人排行、申请趋势、申请——公开趋势等。

该研究点的主要问题在于主题模型的应用，无论是LDA主题模型还是BERTopic模型都是展示关键词来代替主题，从关键词到主题提取的过程依然是依靠人工判断，仍然没有解决之前的缺点，并且结果非常依赖领域内关键词典和分词工具的效果；在大数据量的情况下极度消耗计算资源，这两个模型对于各种Python包的要求也很多，因此近两年的数据挖掘、文本挖掘领域的研究者逐步放弃主题模型转而投入针对不同任务开发大语言模型去实现目的的研究中去。

主题模型在当前LLM（大语言模型）的快速发展下已经完全没有竞争力，通过LLM可以不经分词直接提取文档主题[22]实现专利信息到供应链各活动的匹配，尤其是随着ChatGPT的出现，大语言模型已经可以替代原先的主题模型实现新的主题提取过程。因此在未来的所有专利信息分析、数据挖掘、文本挖掘的研究中，依托大语言模型训练大模型向下游任务的发展提供了良好的参考技术路径，未来的多模态大模型构建以及下游任务实现过程中[23]，可以充分利用高质量的指令微调等技术来显著提升多模态大模型的下游任务性能。

1. **未来发展建议**
2. 提高可视化效果，加强人机交互能力。

当前采用各种方法构建出技术功效矩阵在交点气泡上主要通过专利的数量来体现，专利数量越多交点气泡越大。可以在交点处进行更多的变化，如用饼形图、折线图、年轮图等替换，以展示更多的信息，如该交点处可展示专利权人占比、专利数随时间趋势等，进而提升用户体验。提供交互功能，实现用户随时修改或合并技术功效分类，即时生成调整后的结果。为避免用户等待时间过长，可引入分布式计算等大数据处理技术。

1. 融合市场行情，构建紧密联系的技术功效矩阵。

专利信息分析最终结果要落实在市场、服务于决策者做出最正确的决定，要紧密结合经济、市场等其他学科的建议进行专利信息分析。

1. 智能化构建技术功效矩阵，提高构建效率，减少对专家判断的依赖。

综合运用句法、词法、语义分析方法，结合专利文本撰写特点，运用机器学习模型，加入词系统、知识图谱等知识库的支持，实现“规则+统计+知识库”综合构建方法。提高技术功效分类层次的合理性和专利分类标注的准确性。依靠大数据技术处理海量专利文献，提高构建的效率。现阶段构建技术功效矩阵的过程还存在依赖专家判断的缺点，主要集中在对命名实体的识别和相关领域技术的判断，加强知识图谱的构建能更好地解决该问题。

4) 改进构建方法，当前功效矩阵的构建是主要基于专利摘要数据，在英文专利文献原文中的专利描述性语句也蕴含着与该专利密切相关的内容，因此综合分析摘要、原文将会更详细的提取专利文献的技术功效词语；提取技术功效词常常基于原文语句的SAO结构，在SAO结构之后修饰O的句法部分也包括很多可以提取技术功效词的成分，例如“for”、“to”和动名词短语等，表示目的、因果等等相关表达，但当前的SAOx结构分析主要依赖人工，所以对于SAOx结构的自动化/半自动化分析将会是下一步技术功效词提取的研究方向。

5）大语言模型（LLM）的使用，当前GPT、BERT等模型广泛使用，在所有专利信息分析的手段中加入LLM会起到更好效果，在专利信息分析过程中加入这样先进的技术对减轻于本文提到的所有构建过程中所有缺点将是好的选择。

**参考文献**

[1] 陈颖, 张晓林. 专利技术功效矩阵构建研究进展[J]. 现代图书情报技术, 2011(11): 1-8.

[2] 张兆锋, 贺德方. 专利技术功效图智能构建研究进展[J/OL]. 情报理论与实践, 2017, 40(1): 139-144. DOI:10.16353/j.cnki.1000-7490.2017.01.026.

[3] 翟东升, 陈晨, 张杰, 等. 专利信息的技术功效与应用图挖掘研究[J]. 现代图书情报技术, 2012(Z1): 96-102.

[4] 王丽, 张冬荣, 张晓辉, 等. 利用主题自动标引生成技术功效矩阵[J]. 现代图书情报技术, 2013(5): 80-86.

[5] 陈颖, 张晓林. 专利技术功效矩阵构建词汇模型研究[J/OL]. 情报科学, 2012, 30(11): 1704-1708+1719. DOI:10.13833/j.cnki.is.2012.11.007.

[6] 翟东升, 蔡万江, 陈晨, 等. 基于MapReduce构建专利技术功效图的研究[J]. 情报杂志, 2013, 32(6): 29-33+46.

[7] 张兆锋, 张均胜, 姚长青. 一种基于知识图谱的技术功效图自动构建方法[J/OL]. 情报理论与实践, 2018, 41(3): 149-155. DOI:10.16353/j.cnki.1000-7490.2018.03.028.

[8] YOON J, KIM K. Identifying rapidly evolving technological trends for R&D planning using SAO-based semantic patent networks[J/OL]. Scientometrics, 2011, 88(1): 213-228. DOI:10.1007/s11192-011-0383-0.

[9] 段庆锋, 蒋保建. 基于SAO结构的专利技术功效图构建研究[J]. 现代情报, 2017, 37(06): 48-54.

[10]翟东升, 张京先, 胡等金. 基于SAO结构和词向量的专利技术功效图自动构建研究[J/OL]. 情报理论与实践, 2020, 43(03): 116-123. DOI:10.16353/j.cnki.1000-7490.2020.03.020.

[11]刘化然, 曹旭, 张晓冬, 等. 基于专利技术功效矩阵的技术机会识别方法[J]. 图书情报导刊, 2020, 5(6): 65-70.

[12]余梦霞, 张宇娥, 凌世婷, 等. 专利技术功效视域下的领域重点研发方向布局研究——以第三代功率半导体领域为例[J/OL]. 图书情报工作, 2022, 66(17): 116-128. DOI:10.13266/j.issn.0252-3116.2022.17.011.

[13]刘月, 李雪, 曾莹莹, 等. 基于专利地图的我国肿瘤免疫细胞治疗领域技术情报分析[J]. 中国新药杂志, 2023, 32(5): 448-454.

[14]刘鹏, 闫煜析, 冯立杰, 等. 用户需求导向下基于三级技术功效矩阵的产品创新机会识别[J]. 情报理论与实践, 2023: 1-13.

[15]冯立杰, 吴汉争, 王金凤, 等. 技术创新过程的基因图谱构建及其实证[J]. 情报学报, 2016, 35(8): 806-816.

[16]朱琳, 战洪飞, 余军合, 等. 基于专利数据的企业技术创新机会识别方法研究[J/OL]. 生产力研究, 2019(7): 134-141. DOI:10.19374/j.cnki.14-1145/f.2019.07.022.

[17]张浩, 张云秋. 三维技术功效分析模型构建与实证研究[J/OL]. 情报理论与实践, 2018, 41(05): 74-78. DOI:10.16353/j.cnki.1000-7490.2018.05.014.

[18]王巍洁, 穆晓敏, 王琰, 等. 多维专利技术功效分析模型构建及应用研究[J/OL]. 情报理论与实践, 2020, 43(6): 131-134+130. DOI:10.16353/j.cnki.1000-7490.2020.06.020.

[19]房廷秀, 刘柏均, 刘月, 等. 基于专利地图的人参产业态势分析[J]. 中草药, 2022, 53(12): 3881-3888.

[20]郑学益. 构筑产业链 形成核心竞争力──兼谈福建发展的定位及其战略选择[J]. 福建改革, 2000(8): 14-15.

[21]张杰, 赵君博, 翟东升, 等. 基于主题模型的微藻生物燃料产业链专利技术分析[J]. 数据分析与知识发现, 2019, 3(2): 52-64.

[22]SI Y, WANG J, XU H, 等. Enhancing Clinical Concept Extraction with Contextual Embeddings[M/OL]. (2019-08-14)[2023-06-05]. http://arxiv.org/abs/1902.08691. DOI:10.1093/jamia/ocz096.

[23]赵朝阳, 朱贵波, 王金桥. ChatGPT给语言大模型带来的启示和多模态大模型新的发展思路[J]. 数据分析与知识发现, 2023, 7(3): 26-35.

**Research Progress on Construction and Application of Patent Technology Efficacy Matrix**

Ma Xinyuan

(Nanjing Tech University School of Economics & Management)

**Abstract: [Purpose/signficance]** In recent years, there has been significant development in the construction and application of technology-performance matrices. This paper aims to provide a comprehensive review of this progress, analyze the current challenges, and propose targeted recommendations. **[Result/conclusion]** The current level of intelligent construction for technology-performance matrices is predominantly based on a semi-automated approach with human assistance. However, the accuracy of technology-performance classification is suboptimal, and the visualization capabilities are relatively simplistic. To address these limitations, a comprehensive approach that combines rule-based and statistical natural language processing techniques should be employed to improve the accuracy of extraction and classification. Additionally, research efforts should focus on enhancing the information visualization and interactive functionalities within technology-performance matrices, as well as exploring cross-lingual construction methods. Furthermore, investigating interdisciplinary technology transfer and adaptation strategies and harnessing the potential of large language models offer promising solutions to overcome the existing challenges.

**Keywords:** technical effectiveness matrix; patent analysis; data mining；text mining;

research progress