**主题模型应用在产业链视角下专利分析的报告**

**1.主题模型**

主题模型在文献的分析中广泛应用，经过分词、去除停用词、加载用户词典之后再根据困惑度、主题一致性确定主题数量，经过反复调参最终确定主题个数。

专家学者经常利用LDA主题建模，传统的主题模型在文本分析和信息检索领域得到了广泛应用，它们通过概率模型、矩阵分解等方法来发现文本数据中的主题结构，为文本理解和挖掘提供了有力工具。最近的基于深度学习的方法BERTopic能提供更好的性能和灵活性。

**2.BERTopic主题模型**

BERT是一种基于Transformer的深度学习模型，通过双向的上下文建模来生成词语的向量表示。它在大规模的无标签文本数据上进行预训练，学习了丰富的语义和上下文信息。通过BERTopic内置的方法可以实现可视化地查看主题聚类、文档分布、每个主题下的主要关键词等操作，这是LDA主题模型不具备的。

**3.主题模型应用于某领域的产业链专利分析中**

产业链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。

首先要阅读相关领域综述性文献、查阅资料确定某领域产业链的结构，产业链各层次蕴含着很多活动，通过对检索出来的专利文献应用主题模型，可以实现专利文献——产业链活动的匹配，也可以找出在识别产业链过程中未发现的活动，是个双向互补的过程。

构建产业链上游、中游、下游的技术功效矩阵，识别产业链各层级的技术热点、空白点、可能存在的技术机会。从技术角度结合专利数量识别专利布局的难易程度。

**4.主题模型应用在产业链中的缺点**

现有的主题模型都是用高频关键词确定主题，这个过程主观性强、实际效果一般。对于产业链初步构建没有发现主题的而主题模型发现的，也是需要人工判断的过程，依赖领域专家的知识。

**5.有关技术功效词的提取**

最近阅读文献发现了有关提取英文技术功效词的方法，SAOx分析。对于英文专利的SAO结构之后存在的修饰成分如“for”“to”等结构也可以作为技术功效词语，这样可以更全面地识别技术功效短语。这篇文献用到的技术包括R语言的包OpenNLP和CoreNLP，整个过程需要较多的人工工作。

**基于技术功效矩阵和主题模型的膜技术产业链专利分析**

马昕远

南京工业大学经济与管理学院

**摘 要**：【目的】基于技术功效矩阵和主题模型探究膜技术产业链中各部分技术机会，为企业技术机会识别、专利布局提供参考。【方法】构建膜技术领域产业链模型, 基于BERTopic主题模型实现产业链环节–技术主题–专利文本的映射;分析申请公开趋势、技术发展趋势、构建产业链上中下各部分的技术功效矩阵。【结果】在算法上BERTopic主题模型能更好地识别领域主题，无需反复调参; 在分析结果上，得出膜技术领域产业链技术发展趋势,当前领域内的技术热点、空白点。【局限】本文主要针对膜技术产业链进行研究，该建模方法若推广应用于其他产业,需要具有一定的相关产业背景知识，本文在产业链环节到专利文本的匹配过程中依赖领域专家的知识丰富程度和和判断，具有一定的主观性。【结论】有效识别了膜技术领域的产业链重点、热点环节和技术空白点，该产业链下游有关膜技术的应用方面是当前应重点研发的方向。

**关键词**：膜技术；产业链；技术功效矩阵；BERTopic主题模型；专利分析；数据挖掘

**1 引 言**

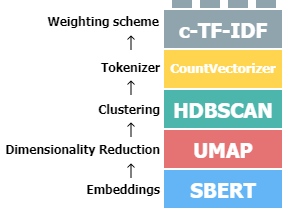
2020年，碳达峰与碳中和被正式列入我国“十四五”发展纲要。膜技术在实现“双碳”目标，服务我国节能减排、生态环境保护、提高资源利用率等方面迎来新的机遇，也面临新的挑战。膜是具有选择性分离功能的材料，当膜两侧存在推动力时如压力差、浓度差、电位差等，原料组分能选择性地透过膜，可以实现不同液体或气体组分的分离、分级、浓缩与提纯。反渗透膜的研发、海水淡化、污水处理等膜技术方向都在助力我国节能减排环境保护事业的发展；近年来我国产业结构不断升级，膜技术领域快速进步，因此本文以膜技术领域作为研究对象，在产业链视角下结合BERTopic主题模型，识别产业链上游、中游、下游领域内的热点、空白点，帮助我国企业做好专利布局，并为膜技术领域产业发展提供建议。

**2 理论综述与研究思路**

**2.1 基于主题模型的信息分析方法**

目前专家学者常用的主题模型是LSA、pLSA、NFM、LDA等传统的主题建模方法，广泛使用的是LDA主题模型，LDA用于挖掘和识别目标文档主题，一些学者通过对专利文档进行LDA主题建模分析专利文本的技术特征。Kim等[1]使用LDA主题模型对美国减少温室气体排放技术专利进行分析,挖掘不同时期技术发展趋势及特征。Chen等[2]提出一种基于潜在Dirichlet分布的话题变化识别方法,以最少的人为干预模拟、分析主题的变化和趋势,使得隐藏在专利权力要求下的潜在语义主题自动显示,并且将该方法在2009年–2013年间,以澳大利亚为受让国的案例研究中进行验证。Venugopalan等[3]基于LDA主题模型算法识别专利关键主题,将主题分布作为专利特征,构建分类器进行学习,将专利划分到不同技术子领域以进行技术融合以及创新地域研究。宋凯等[4]从专利法律状态视角计算技术发展态势，结合LDA主题模型和word2vec进行相似度计算用于挖掘潜在竞争对手，使用pyLDAvis进行主题可视化判断主题个数是否合适，最后以“云计算”领域为例进行实证研究。张杰等[5]利用LDA主题模型结合产业链的方法分析微藻生物燃料领域，有效识别了微藻生物燃料产业链重点及热点环节。

BERTopic是一种主题建模技术是自然语言处理领域最新的主题模型，它利用huggingface的Transformer和c-TF-IDF创建聚类簇，使主题易于理解，同时在主题描述中保留重要的单词，BERTopic基本工作流程如图1所示。



**图1 BERTopic工作流程**

Okazaki等[6]利用BERTopic主题模型从新闻数据中提取每个公司部门所属的产品和与每个部门相关的特征主题，并分析了这些主题随时间的变化，该方法能有效地获取关于公司部门的及时信息，帮助投资者和利益相关者更好地了解公司的活动和趋势。曹树金等[7]利用BERTopic探究信息资源管理领域学科发展方向和拓展空间，建议学科发展在强基固本、不断促进跨学科交叉融合的基础上，新增“行业信息资源管理”、“数据资源管理”等二级学科。沈棋琦等[8] 应用BERToopic主题建模方法对移动图书馆APP在线评论数据进行主题挖掘,揭示了影响用户情感体验的关键因素并建立了完整的移动图书馆APP用户情感体验评价体系， 为在线评论情感分析和用户情感体验度量提供了重要依据。张敏等[9]基于政务话语框架和BERTopic模型对政务短视频标题进行主题挖掘，构建突发公共卫生事件期间政务短视频的话语框架,为突发公共卫生事件中政务短视频运营管理提供了一种全新的发布策略。

BERTopic基于BERT词向量，它利用BERT嵌入、统一面域逼近和投影（UMAP）、HDB-SCAN聚类算法以及C-TF-IDF来创建密集的集群，轻松解释主题并在主题中保留重要词[10],相比于LDA、CTM等主题模型，BERTopic的优势在于弥合了基于密度聚类和基于中心采样之间的不兼容问题，使用时可以生成较优的主题个数，不需要LDA主题模型预先设定主题个数后反复调参，本文研究中将采用BERTopic主题模型。

**2.2 技术功效矩阵**

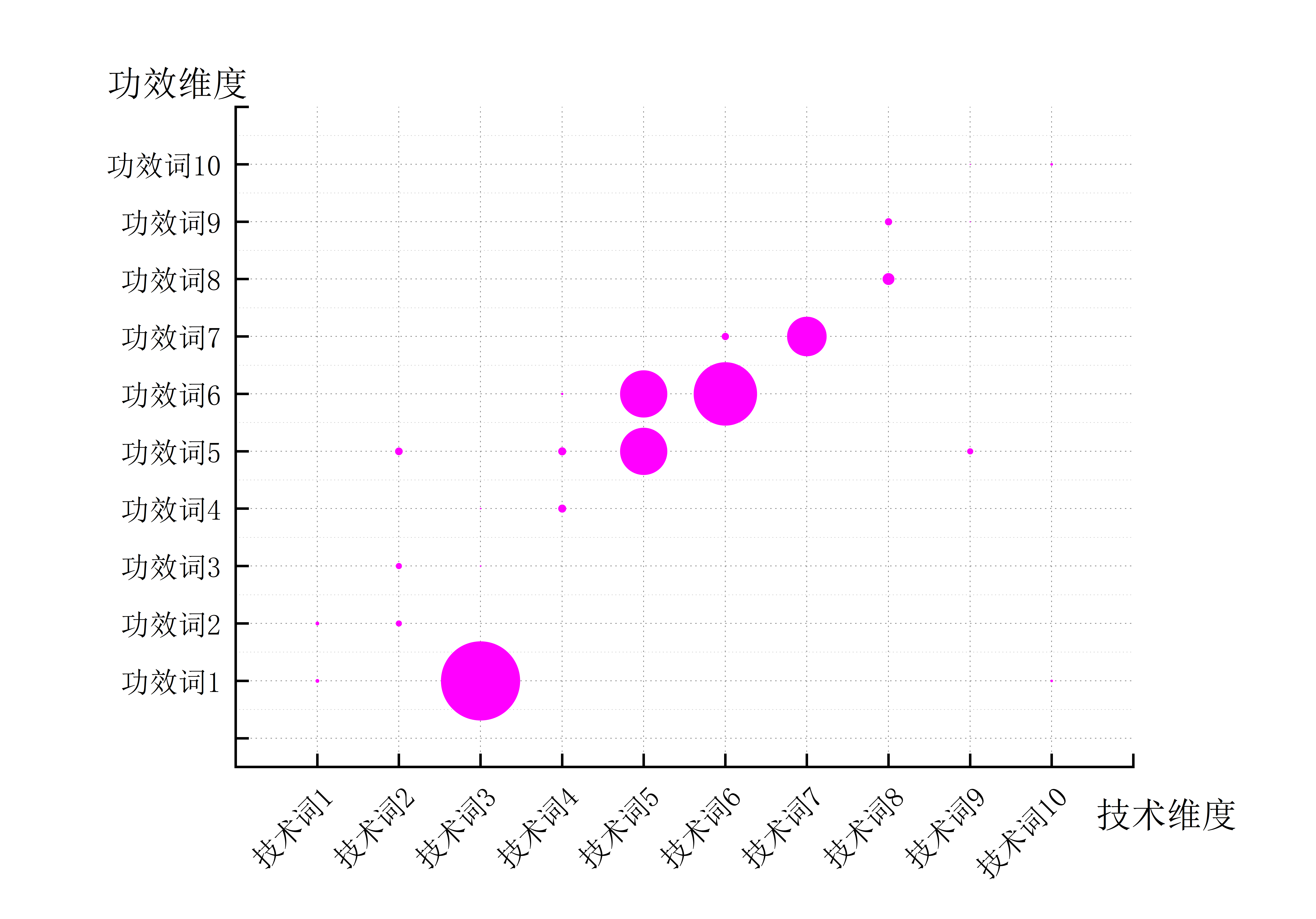
技术功效矩阵是专利地图的一种，是常见的专利定量分析方法；通过识别专利文献中的技术词、功效词，绘制成可视化图表进行分析。技术功效矩阵可以直观地反馈当前领域内技术和功效的类别与分布情况，帮助决策者识别“专利雷区”、“专利空白区”，识别领域内技术的热点、空白点[11]，可以为企业专利布局、技术研发、抢占市场等工作提供参考。专利技术功效矩阵一般是同时包括“技术”和“功效”两种元素而构成的二维专利地图。目前对技术功效矩阵还没有明确的定义；专利技术功效矩阵也被称为技术功效矩阵（文中以下内容均称为技术功效矩阵），分析对象一般为专利文献，其他类型的文献大多没有明确描述功效，所以不适合用技术功效矩阵做相关工作。

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[12]；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2所示。

技术功效矩阵能反映出领域内技术热点、空白点，刘化然等[13]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[14]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[15]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。



**图2 技术功效矩阵**

**2.3 产业链分析技术**

产业链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[16],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，基本的产业链结构如图3所示，当前围绕产业链技术进行专利分析或者技术路径分析已成为非常多学者研究的重点热点。



**图3 产业链基本结构**

基于以上文献分析可知当前主题模型已经广泛应用于专利数据分析，国内外学者都对技术热点、难点、空白点进行识别分析，本文通过产业链进行分析是可行的。

**2.4 本文研究思路**

本文研究思路如图4所示。



**图4 本文研究思路**

本文研究主要分为三个部分：

1. 查阅领域内相关文献结合领域专家的判断分析并确定产业链模型，识别产业链各环节上的所有活动。
2. 运用BERTopic模型对该领域专利摘要进行主题建模,挖掘潜在技术主题。综合产业链环节以及主题标注结果,将专利分布于产业链不同环节,实现产业链环节–产业链活动–专利的匹配。
3. 根据产业链环节–产业链活动–专利的匹配结果构建上游中游下游产业链的技术功效矩阵，开展各层次的技术机会识别以此发现专利雷区、技术空白点，分析当前申请人情况等其他分析过程。

**3 产业链视角下的膜技术领域专利技术主题分析**

**3.1 膜技术领域产业链构建**

本文通过查阅文献[17–19]浏览膜技术技术公司网站信息,收集产业资料,构建膜技术产业链组成模型,并结合专家意见对产业链环节进行修正,最终确定膜技术上中下游产业链环节。其中膜产业链上游主要包括膜材料、膜设备和泵阀等相关专业配套装备的生产制造;膜产业链中游主要是膜组件和装置的装备制造，膜产业链下游主要系膜系统的集成与膜分离技术应用，包括污水处理、市政供水、海水淡化等，膜技术产业链模型如图5所示。



**图5 膜技术产业链模型**

**3.2 基于BERTopic的专利文本主题分析**

（1）文献数据来源和预处理

本文采用Incopat数据库进行专利检索，通过查阅相关资料、浏览相关报告后确定检索领域的表达式，检索膜技术领域的发明专利，发明专利有效期是20年，检索时间选择在2004年起始；经过同族合并、人工去重筛选共3498篇专利文献，选取专利文献的摘要部分用于建立膜技术语料库。

对收集到的进行使用Python语言的jieba库对语料库分词、加载领域内术语辞典保留重要的关键词[20]、加载哈工大停用词表去除停用词、统一大小写、去除数字等相关降噪处理，再经人工判断去重删除无意义的词，确定最终的语料库。

（2）BERTopic主题模型分析

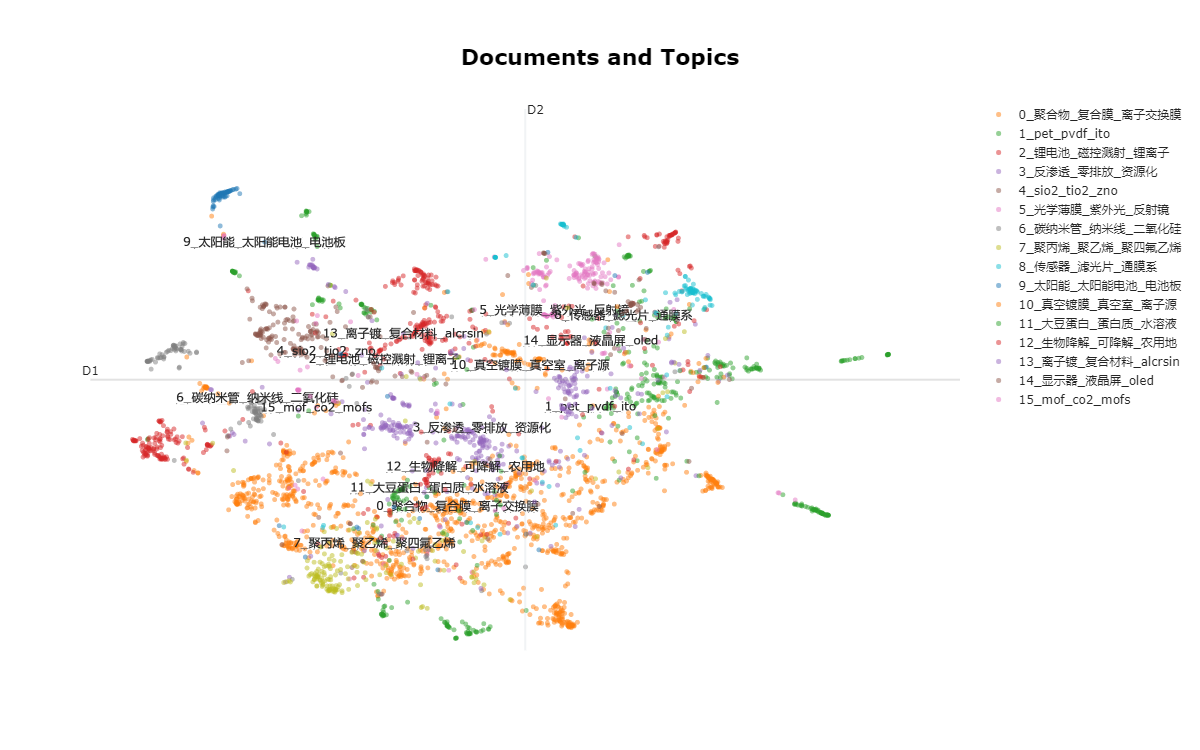
BERTopic主题模型的优点是是通过将文本表示为 BERT 向量，然后使用聚类算法（如 DBSCAN）来识别主题。BERTopic能自动地生成比较合适的主题数量，相比于LDA模型不需要反复调参，比较好的结果是每一个主题之间尽可能不交叉，主题之间的距离尽可能大。经过分析降维处理后的聚类结果如图6所示，将主题数量确定为16个，导出每个主题下权值较大的关键词如图7所示，为了更细粒度地展示每个主题下的文档调用BERTopic自带的model.visualize\_documents()函数得到图8，每个颜色的簇代表同一主题下的文档，可以看出膜技术领域中有大量的专利存在于产业链上游膜材料以及莫配套设备的制备相关活动中。



**图6 BERTopic聚类图**



**图7 每个主题下权值排名前五主题词**



**图8 文档-主题分布**

**3.3 文档主题与产业链各活动的匹配**

通过充分了解膜技术领域技术知识,对膜技术领域专利进行技术主题标注,确定主题名称。根据BERTopic主题模型挖掘到的主题,如表1所示。

（1）主题1中有权值较大的聚合物、超滤膜、离子交换膜、膜技术等词特征直接体现超滤膜的制备方法，因此将主题1定义为“超滤膜的制备方法”，与产业链上有活动中的“超滤膜”相匹配。

（2）主题2中有“保护膜组件”、“保护膜”体现膜组件的词特征，所以将主题2定义为“膜组件制备”，与产业链中游“膜组件”活动匹配。

（3）主题3中存在“膜装置”、“制备方法”、“锂电池”等体现膜装置的词语特征，故将主题3定位为“膜装置制备与应用”，匹配产业链中游的“膜装置”活动。

（4）主题4中有“反渗透”、“海水淡化”、“电渗析”等体现反渗透膜应用的词语特征，因此将主题4定义为“反渗透膜的海水淡化与废水处理应用”，匹配产业链下游的“海水淡化”活动。

（5）主题5中含有“氧化物”、“污水处理”、“膜应用”等体现膜技术应用的词语，所以将主题5定义为“膜应用—污水处理”匹配产业链下游的“污水处理”活动。

（6）主题6蕴含“钢化玻璃膜装置”、“反射镜”、“光学薄膜”等代表膜装置制备的特征词，因此将主题6定义为“膜装置制备”，对应产业链中游的“膜装置”活动。

**表1 主题词-主题对应表**

|  |  |  |
| --- | --- | --- |
| 编号 | 主题名称 | 主题词 |
| 1 | 超滤膜制备方法 | 聚合物,复合膜,离子交换膜,中空纤维膜,超滤膜,膜技术,纳滤膜,聚酰亚胺，膜材料,保护膜 |
| 2 | 膜组件制备 | pet,pvdf,ito,pvc,gan,led,制作方法,二极管,保护膜组件,保护膜 |
| 3 | 膜装置制备与应用 | 锂电池,磁控溅射,锂离子,金刚石,半导体,半导体器件,电解质,氧化铝,制备方法,膜装置 |
| 4 | 反渗透膜的海水淡化与废水处理应用 | 反渗透,零排放,资源化,电渗析,海水淡化,热熔胶,浓缩液,分离器,废水处理,二氧化碳 |
| 5 | 膜应用——污水处理 | sio2,tio2,zno,al2o3,氧化物,钙钛矿,光催化,vo2,电致发光,污水处理,膜应用 |
| 6 | 膜装置制备 | 光学薄膜,紫外光,反射镜,折射率,可见光,oca,钢化玻璃膜装置,微结构,反射层,紫外线 |
| 7 | 膜制备 | 碳纳米管,纳米线,二氧化硅,纳米银,硅橡胶,碳化硅,纳米管,硅氧烷,膜制备,二氧化钛 |
| 8 | 泵阀装置制备方法 | 聚丙烯,聚乙烯,聚四氟乙烯,收缩膜,低密度聚乙烯,中间层,聚乳酸,泵阀装置,装置制备,高密度聚乙烯 |
| 9 | 市政供水中的膜应用 | 传感器,滤光片,通膜系,过滤器,热电偶,温度传感器,滤波器,敏感膜,细胞培养,城市供水 |
| 10 | 提高太阳能电池中膜耐受性 | 太阳能,太阳能电池,电池板,pbt,太阳电池,提高膜耐候性,中间层,玻璃纤维,膜装置,碲化镉 |
| 11 | 真空镀膜方法 | 真空镀膜,真空室,源腔室,塑胶制品,真空管,镀膜方法,中间轴,plc,主动轴 |
| 12 | 可食用膜制备方法 | 大豆蛋白,蛋白质,水溶液,25,鹰嘴豆,铁蛋白,喷雾干燥,可食用膜,微滤膜,制备方法 |
| 13 | 生物降解膜在海水淡化应用 | 生物降解,可降解,农用地,聚乳酸,数据库,海水淡化,生物膜,提取物,生态系统 |
| 14 | 超滤膜制备方法 | 离子镀,复合材料,alcrsin，过渡层，crn，自润滑，aln，alcrn，超滤膜，膜制备 |
| 15 | 膜装置应用 | 显示器,液晶屏,oled,显示装置,显示屏,pani,传送带,液晶显示,电流表,膜装置 |
| 16 | 气体分离膜组件制备 | mof,co2, mofs,mos2,psu,气体分离膜,cu2,重金属,nh2,pebax |

（7）主题7中包括“膜制备”、“二氧化硅”、“纳米银”等膜制备原料的词语特征，因此将主题7定义为“膜制备”匹配膜技术领域产业链的上游活动。

（8）主题8中有“泵阀装置”、“装置制备”、“聚乙烯”、“聚丙烯”等涉及泵阀装置制备的词语，所以将主题8定义为泵阀装置制备方法，对应膜技术产业链的上游“膜制泵阀等”活动。

（9）主题9中包括“城市供水”、“传感器”、“敏感膜”等涉及膜供水应用的词语，故将主题9定义为“城市供水中的膜应用”，对应膜技术产业链下游的“市政供水”活动。

（10）主题10中含有“太阳能”、“提高膜耐受性”、“膜装置”等囊括膜装置改进的词语特征，所以将主题10定义为提高太阳能电池中膜耐受性，匹配膜技术产业链中游“膜装置”活动。

（11）主题11涉及“真空镀膜”、“镀膜方法”等改进膜装置的词特征，故将主题11定义为真空镀膜方法，与产液量中游膜装置相对应。

（12）主题12有“微滤膜”、“可食用膜”、“大豆膜蛋白”等涉及微滤膜的词语，所以将主题12与产业链上游“微滤膜”活动对应。

（13）主题13包括“海水淡化”、“生物降解”、“可降解”等词语，所以将主题13定义为“生物降解膜在海水淡化中的应用”与产业链下游海水淡化相匹配。

（14）主题14有“离子镀”、“超滤膜”、“膜制备”等涉及膜制备的主题词，所以将主题14定义为超滤膜制备方法，匹配产业链上游“超滤膜”活动。

（15）主题15囊括“显示器”、“液晶屏”、“膜装置”等代表膜装置应用的词语特征，因此将主题15定义为“膜装置应用”，对应产业链中游“膜装置”活动。

（16）主题16有“气体分离膜”、“重金属”、“CU2”等涉及膜装组件制备的词特征。所以将主题16定义为“气体分离膜组件制备”，对应“膜组件”活动。

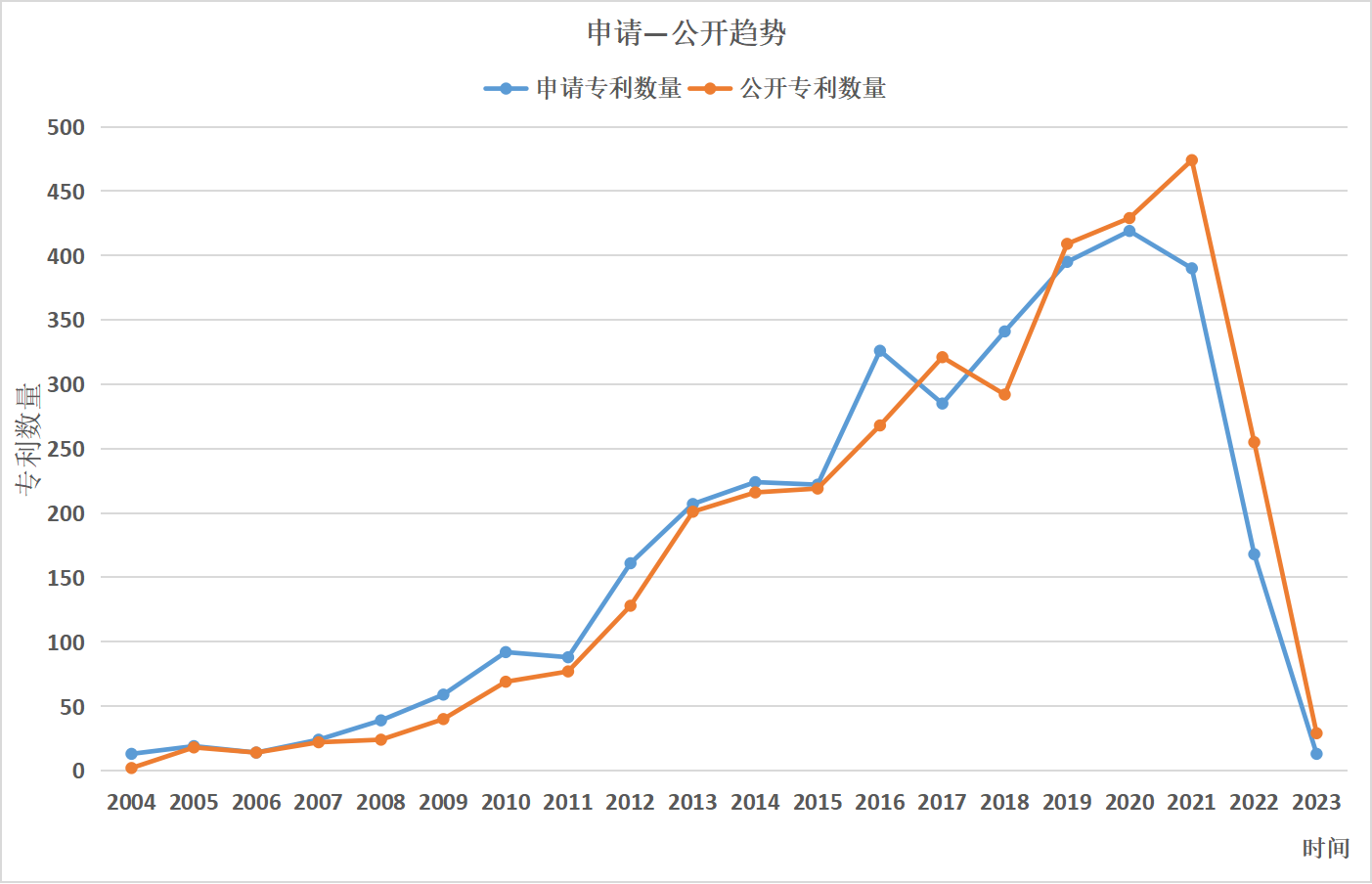
**3.4构建IPC-技术功效矩阵**

通过国际专利分类（IPC）技术功效矩阵分析可以了解技术热点和潜在技术热点，对侧重点进行划分；旨在为膜产业链开发利用、产业链专利布局及相关产业竞争力的提升提供参考依据。

**4 膜技术领域专利分析**

**4.1 申请——公开趋势分析**

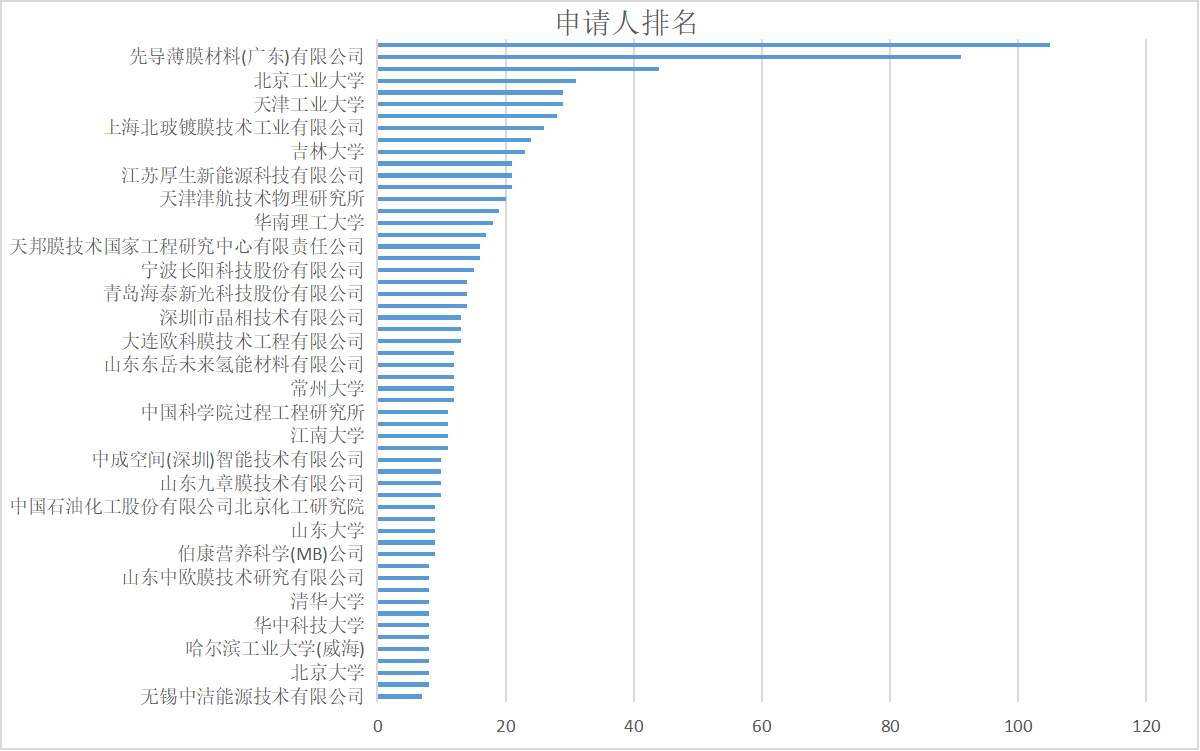
申请公开趋势分析展示的是专利申请量和公开量的发展趋势。通过趋势可以从宏观层面把握分析对象在各时期的专利布局变化。专利公开和专利申请相比有一定滞后，一般发明专利在申请后3~18个月公开，实用新型专利和外观设计专利在申请后1~15个月公开，因此2022~2023年的申请公开趋势有所下降。从图9可以看出膜技术领域的专利申请公开趋势稳步上升，我国企事业单位、发明人都对于膜技术领域给与很高重视并做出投入。



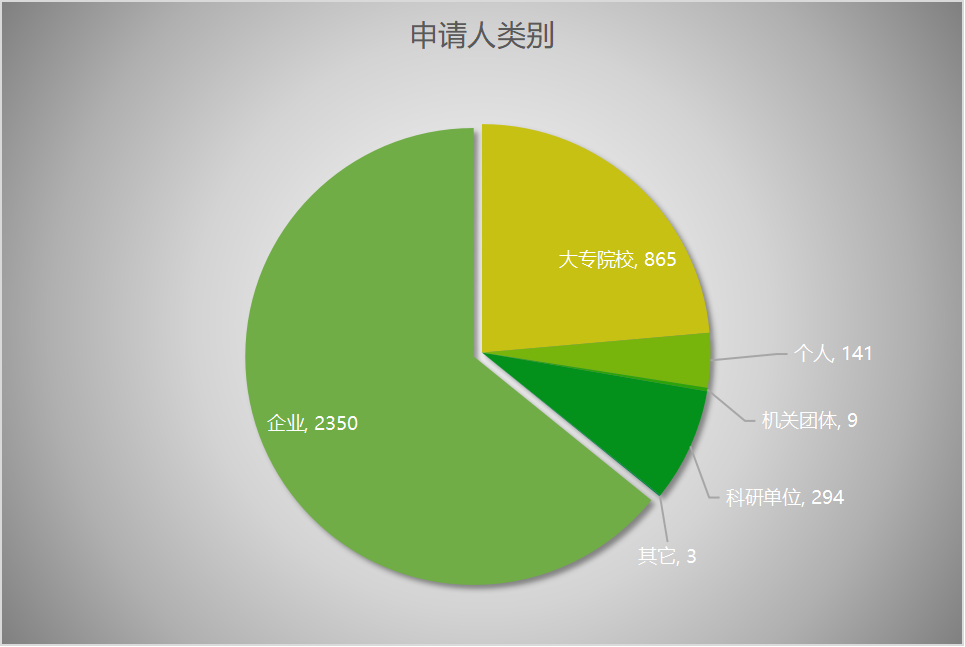
**图9 申请——公开趋势分析图**

**4.2 研发主体分析**

对膜技术产业领域专利的主要研发主体进行分析，绘制统计图, 申请人排名如图10所示，申请人类型如图11所示。可以看出企业、科研院校、大专院校专利申请量占总申请量的绝大部分，对膜技术领域给予足够重视。



**图10 申请人排名**



**图11 申请人类别**

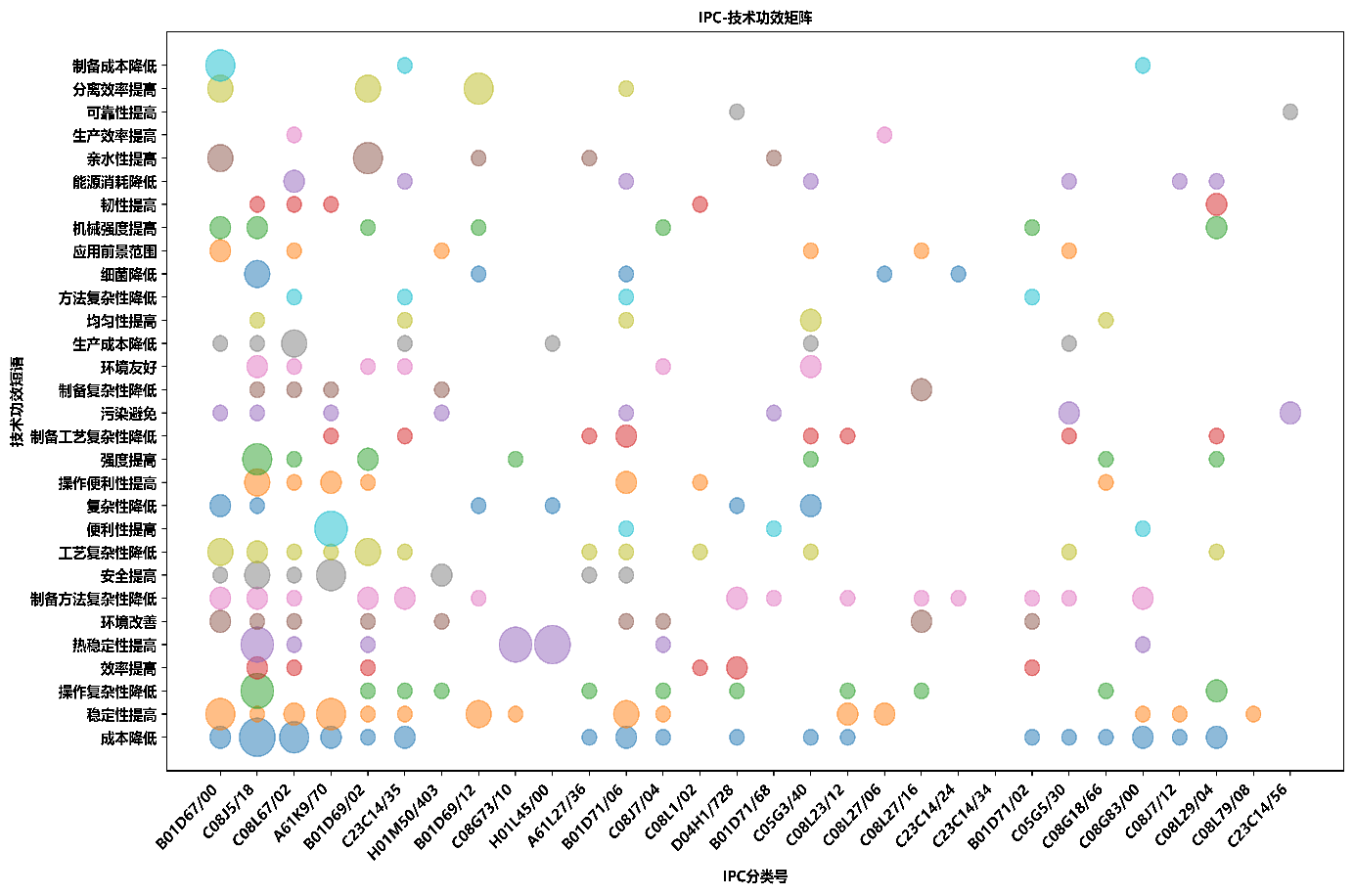
**4.3 产业链视角下的技术机会识别**

（1）产业链上游的技术技术机会识别

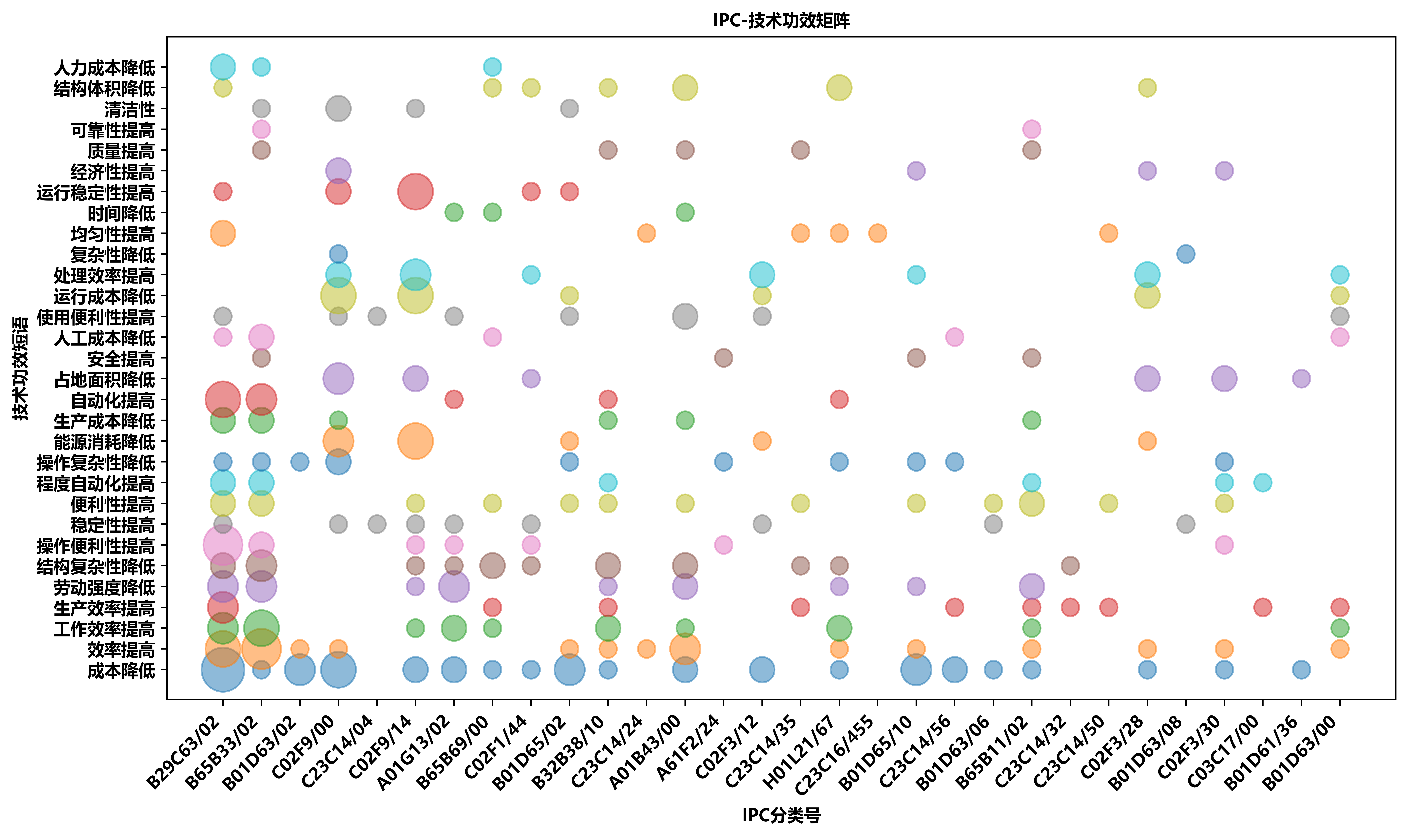
构建产业链上游的IPC-技术功效矩阵如图12所示，可以看出上游企业对于成本降低、操作复杂性降低、稳定性提高、工艺复杂性降低等方面投入较大，可以实现这些效果的IPC类别专利也很多，因此这类方向属于专利壁垒区，不适合后来申请人在这些方向上投入太多。相反的制备成本降低、方法复杂性降低、均匀性提高、生产效率提高、韧性提高这类效果中申请专利数量非常少，可实现这类功效的专利类别也很少，属于专利沉寂区不建议开展专利布局。可能存在技术机会的功效是环境友好、强度提高、安全性提高、操作便利性提高、复杂性降低、热稳定性提高、污染避免等技术功效，这类功效上的专利数量适中，有很多类别的专利可以达成此类技术功效，因而有较大的技术机会。

（2）产业链中游的技术机会识别

构建产业链中游的IPC-技术功效矩阵如图13所示，可以得出结论：中游企业对于成本降低、效率提高、工作效率提高等方面投入较大，可以实现这些效果的IPC类别专利也很多，因此这类方向属于专利壁垒区，不适合继续在这些方向上投入太多。相反的人力成本降低、可靠性提高、清洁性、便利性提高等技术功效中申请专利数量非常少，可实现这类功效的专利类别也很少，属于专利沉寂区，所以不建议在此方向开展专利布局。可能存在技术机会的功效是人工成本降低、自动化提高、便利性提高、劳动强度降低、结构复杂性降低、稳定性提高、使用便利性提高、运行成本降低、均匀性提高、生产成本降低、操作便利性提高等技术功效，这类功效上的专利数量适中，有很多类别的专利可以达成此类技术功效，因而有较大的技术机会。



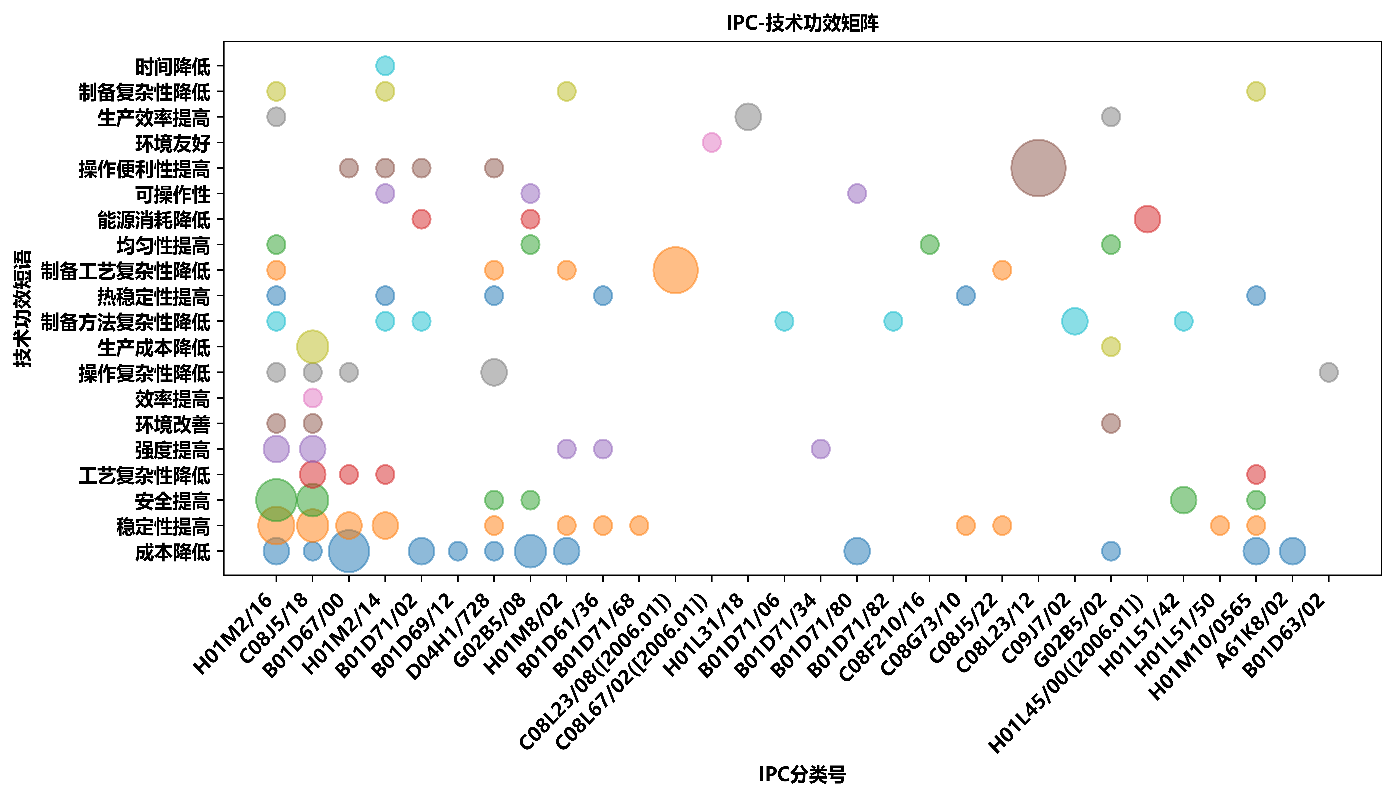
**图12 产业链上游的IPC-技术功效矩阵**



**图13 产业链中游的IPC-技术功效矩阵**

（3）产业链下游的技术技术机会识别

构建产业链下游的IPC-技术功效矩阵如图14所示，产业链下游的专利总量较少说明在膜技术应用方面国内还有很大空白，企业应抓住机会开展专利布局抢占市场，通过技术功效矩阵可以得出结论：下游企业对于成本降低、稳定性提高、工作效率提高等方面投入较大，可以实现这些效果的IPC类别专利也很多，因此这类方向属于专利壁垒区，不适合继续在这些方向上投入太多。相反的时间降低、环境改善、效率提高、能源消耗降低、等技术功效中申请专利数量非常少，可实现这类功效的专利类别也很少，属于专利沉寂区，所以不建议在此方向开展专利布局。可能存在技术机会的功效是操作便利性提高、热稳定性提高、制备方法复杂性降低、强度提高、工艺复杂性降低、制备工艺复杂性降低、操作便利性提高等技术功效，这类功效上的专利数量适中，有很多类别的专利可以达成此类技术功效，因而有较大的技术机会，适合下游企业在这类方向上进行专利布局。



**图14 产业链下游的IPC-技术功效矩阵**

**5 结论**

本文通过BERTopic主题模型结合产业链与产业链上游、中游、下游的技术功效矩阵对膜技术领域进行详细的专利信息分析，利用BERTopic本身对于关键词抽取相对于其他主题模型更精准且不需要反复调参的优势，将产业链上各个活动与模型识别出的主题进行匹配，进而实现专利文本和产业链活动的精准对应；利用技术功效矩阵识别可能存在的技术机会，为相关企业专利布局提供依据。本文得出的结论主要有以下方面：

1. 产业链下游专利数量较少，通过技术功效矩阵反馈出的技术空白点很多，政府可以出台相关政策促进企业、高校、科研院所在下游膜应用上的研发，如海水淡化、污水处理、市政供水等方面。
2. 识别出了适合上游、中游、下游企业的开展专利布局技术点和不适合专利布局的壁垒区。

本文存在的不足指出在于提出的分析方法仅适合膜技术领域，对于应用于其他领域的研究需要依赖领域专家知识丰富程度；主题识别的过程存在一定主观性，因为这一过程需要分词、去除停用词再根据权值较大的词判断主题，采用GPT等大语言模型可能有更好的效果，大语言模型的部署相对困难，现阶段实现起来难度较大；对于识别出的主题缺少验证的环节，对于结果判断的准确性缺少验证；分析过程缺少对上游、中游、下游企业识别的过程以及技术继承性的考虑，未能针对每个企业提供个性化的参考建议；对于技术功效矩阵中专利壁垒区、沉寂区、活跃区的识别缺少量的判断，布局难度判断的主要问题是以下公式：

分子较大分母很小，导致布局难易程度高，但是该技术点上专利文献的申请量如果非常大，那么该在这个技术点上后来者继续申请专利是更容易或更困难，较难确定。

**参考文献**

[1] KIM G, PARK S, JANG D. Technology Analysis from Patent Data Using Latent Dirichlet Allocation[M/OL]//LEE K M, PARK S J, LEE J H. Soft Computing in Big Data Processing: 卷 271. Cham: Springer International Publishing, 2014: 71-80[2023-06-05].

[2] CHEN H, ZHANG Y, ZHANG G, 等. Modeling technological topic changes in patent claims[C/OL]//2015 Portland International Conference on Management of Engineering and Technology (PICMET). Portland, OR, USA: IEEE, 2015: 2049-2059[2023-06-05]. http://ieeexplore.ieee.org/document/7273098/.

[3] VENUGOPALAN S, RAI V. Topic based classification and pattern identification in patents[J/OL]. Technological Forecasting and Social Change, 2015, 94: 236-250.

[4] 宋凯, 冉从敬. 基于专利法律状态的企业技术竞争画像构建研究[J/OL]. 情报理论与实践, 2022, 45(12): 178-185. DOI:10.16353/j.cnki.1000-7490.2022.12.023.

[5] 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.

[6] OKAZAKI H, TAKAHASHI H. Nowcasting of Corporate Research and Development trends through news article analysis by BERTopic: the case of Japanese electric company[C/OL]//2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME). Maldives, Maldives: IEEE, 2022: 1-6[2023-06-05]. https://ieeexplore.ieee.org/document/9987867/.

[7] 曹树金, 曹茹烨. 基于研究主题和引文分析的信息资源管理学科发展探究[J/OL]. 信息资源管理学报, 2023, 13(2): 12-29.

[8] 沈棋琦, 张雪, 赵杨. 基于评论主题挖掘的移动图书馆App用户情感体验评价体系构建[J/OL]. 图书馆学研究, 2023(4): 41-51. DOI:10.15941/j.cnki.issn1001-0424.2023.04.007.

[9] 张敏, 沈嘉裕. 突发公共卫生事件中政务短视频主题与用户行为的关联演化研究[J]. 情报杂志, 2023, 42(3): 181-189.

[10]GROOTENDORST M. BERTopic: Neural topic modeling with a class-based TF-IDF procedure[M/OL]. arXiv, 2022[2023-06-05]. http://arxiv.org/abs/2203.05794.

[11]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.

[12]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.

[13]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.

[14]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.

[15]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.

[16]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.

[17]刘雨. 膜技术在工业废水处理中的应用研究进展[J]. 化工设计通讯, 2023, 49(2): 177-179.

[18]林钰青, 张以任, 邱宇隆, 等. 膜技术在盐湖提锂中的进展和展望[J/OL]. 无机盐工业, 2023, 55(1): 33-45.

[19]郑佩君, 谢威, 白菊, 等. 气体分离膜技术在天然气提氦中的研究进展[J/OL]. 膜科学与技术, 2022, 42(6): 168-177.

[20]王从厚. 膜技术术语辞典[M]. 北京. 化学工业出版社. 2008.2-31.

**Patent Analysis of the Membrane Technology Industry Chain Based on Technology Efficacy Matrix and Topic Modeling**

Ma Xinyuan

(Nanjing Tech University School of Economics & Management)

**Abstract: [Purpose]** The purpose of this study is to investigate various technological opportunities within the membrane technology industry chain using the Technology Efficacy Matrix and the BERTopic topic model. The findings aim to provide valuable insights for enterprises in identifying technological opportunities and planning their patent strategies. **[Methods]** To achieve the research objectives, an industry chain model was constructed for the membrane technology field. The BERTopic topic model was then utilized to map the different segments of the industry chain, technological themes, and patent texts. Furthermore, an analysis was conducted to examine the trends in patent applications and technological developments. A Technology Efficacy Matrix was constructed to evaluate the effectiveness of various technologies across different parts of the industry chain. **[Results]** The application of the BERTopic topic model proved effective in accurately identifying domain-specific themes without the need for extensive parameter adjustments. The analysis yielded valuable results regarding the technological development trends within the membrane technology industry chain. Additionally, it identified the current technological hotspots and areas with limited research progress. **[Limitations]** It is important to note certain limitations of this study. Firstly, the research primarily focuses on the membrane technology industry chain. Generalizing the proposed modeling method to other industries would require a substantial understanding of the relevant industry backgrounds. Moreover, the matching process from the industry chain to patent texts heavily relies on the expertise and judgment of domain experts, which introduces a certain level of subjectivity. **[Conclusion]** This study successfully identified key components, hotspots, and technological gaps within the membrane technology industry chain. The findings emphasize the importance of focusing on downstream applications of membrane technology within this industry chain for further research and development endeavors.

**Keywords:** Membrane technology Industry chain Technology efficacy matrix

BERTopic topic model Patent analysis Data mining