



科学学研究  
*Studies in Science of Science*  
ISSN 1003-2053, CN 11-1805/G3

## 《科学学研究》网络首发论文

题目： 面向有组织产学研协同创新的合作主题挖掘  
作者： 黄璐，任航，曹晓丽，陈翔  
DOI： 10.16192/j.cnki.1003-2053.20240801.001  
收稿日期： 2024-01-26  
网络首发日期： 2024-08-02  
引用格式： 黄璐，任航，曹晓丽，陈翔. 面向有组织产学研协同创新的合作主题挖掘 [J/OL]. 科学学研究. <https://doi.org/10.16192/j.cnki.1003-2053.20240801.001>



**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。

# 面向有组织产学研协同创新的合作主题挖掘

黄璐<sup>1,2</sup> 任航<sup>1</sup> 曹晓丽<sup>3,4</sup> 陈翔<sup>5</sup>

- (1. 北京理工大学经济学院, 北京 100081;  
2. 数字经济与政策智能工业和信息化部重点实验室, 北京 100081;  
3. 中国科学院文献情报中心, 北京 100190;  
4. 中国科学院大学经济与管理学院信息资源管理系, 北京 100190;  
5. 北京理工大学管理学院, 北京 100081)

**摘要:** 开展有组织的产学研协同创新, 是发挥我国新型举国体制优势、实现产学研深度融合的重要内容。其中, 对产学研“合作主题”的有效识别是实现高质量高效率产学研协同创新的首要问题。本文提出了一套基于复杂网络分析和深度学习算法的产学研协同创新合作主题挖掘方法, 首先, 围绕“有组织的产学研协同创新”概念和主题特征进行深度剖析, 提出产学研合作主题应具有高价值性和强相关性两大特征; 其次, 基于论文数据和专利数据分别构建“科学主题词-学研机构”双层网络和“技术主题词-企业”双层网络, 其中, SciBERT 模型被用来构建科学和技术主题词语义网络, 基于 Node2Vec 的链路预测模型被用来预测未来的科学和技术主题词语义网络; 之后, 应用复杂网络拓扑结构分析、社区发现、机器学习等方法对主题的新颖性、基础性、广泛性、成长性、前瞻性五大指标进行测度, 识别高价值的科学主题和技术主题; 最后, 对语义相似度指标 SimDoc 进行改进, 计算科学主题和技术主题之间的相关性, 遴选产学研协同潜力大的“科学主题-技术主题对”。本文选取人工智能领域开展实证研究, 对提出的研究方法进行验证。本研究能为国家、区域和行业组织高层级产学研协同创新提供重要的量化决策参考。

**关键词:** 产学研协同创新; 产学研合作主题; SciBERT 模型; 复杂网络分析; 有组织

**中图分类号:** G353.12 **文献标识码:** A

二十大报告提出“产学研深度融合”“推动创新链产业链资金链人才链深度融合”, 这是我国在实现高水平科技自立自强、建设科技强国的新阶段对产学研协同创新提出的新要求。近年来, “有组织”相关理论研究受到广泛关注, 如“有组织科研”“有组织的基础研究”和“有组织创新”等<sup>[1]</sup>, 在“有组织”理论指导下开展创新活动, 有利于发挥我国新型举国体制优势<sup>[2]</sup>, 打破阻碍不同组织、学科、地域间人才、资本、信息与技术等要素自由流动的传统藩篱和壁垒<sup>[3]</sup>。

开展有组织的产学研协同创新, 强调以国家战略需求为导向, 从顶层视角谋划全局, 整合产学研优势资源开展有导向性、目标性、定向性的高层级创新合作, 避免创新主体各自为政、无序发展<sup>[4]</sup>, 实现产学研深度融合<sup>[5]</sup>。由于“选题”是科学研究和创新活动的起点<sup>[6]</sup>, 因此, 在开展有组织产学研协同创新的过程中, 针对协同创新的合作目标(主题)进行顶层规划, 是“有组织”的首要体现, 更是实现高质量产学研协同创新的关键环节<sup>[7]</sup>。

现有研究多停留在概念和理论解析层面, 如围绕“有组织基础研究”内涵特征、实践经验、推进策略进行探讨<sup>[8]</sup>, 针对高校“有组织科研”组织模式、实施路径、响应机制展开分析<sup>[9]</sup>, 鲜少从量化方法的角度为政策制定和实践指导提供支撑。基于此, 本文面向有组织产学研协同创新的合作主题识别, 提出一套量化分析方法体系, 用论文表征学研机构开展的科

收稿日期: 2024-01-26; 修回日期: 2024-07-06

基金项目: 国家自然科学基金面上项目(72274013; 72371026)。

作者简介: 黄璐(1984-), 女, 教授, 博士。任航(2001-), 女, 硕士研究生。曹晓丽(1999-), 女, 博士研究生。陈翔(1976-), 男, 教授, 博士, 通讯作者, E-mail: bjchenxiang@hotmail.com。

学研究，用专利表征企业开展的技术研究，一方面，构建与有组织产学研协同创新内涵和特征相适应的高价值科学和技术主题评价指标体系，并综合复杂网络分析和深度学习算法构建两类“主题词-机构”双层网络，实现对指标的科学度量；另一方面，引入语义分析方法测算高价值科学主题与技术主题的相关性，提升产学研合作主题协同潜力分析的准确性。本文以人工智能领域为例开展实证分析，对产学研合作主题挖掘方法及结果进行验证。

## 1 有组织的产学研协同创新内涵及主题特征分析

有组织的产学研协同创新核心内涵包括三方面：一是强调对创新链和产业链上各类创新主体有组织的整合共创<sup>[10]</sup>，推动产学研合作由“自由无序”迈向“组织有序”<sup>[11]</sup>；二是强调有目的地推进一系列跨机构、跨维度、跨层次产学研协同创新活动的有组织开展，主要表现为以国家战略需求为导向<sup>[4]</sup>，保证产学研协同创新的高层级；三是强调对产学研不同创新主体间创新诉求的有组织管理，实现多元诉求协调互促<sup>[10]</sup>，避免“强买强卖”和“拉郎配”，保证产学研协同创新的高效率。

因此，在开展有组织的产学研合作主题顶层规划过程中，首先要选择出高价值的科学主题和技术主题；同时，要充分考量企业和学研机构间创新诉求的相关度，遴选关联性较高的“科学主题-技术主题对”（简称“科技主题对”）。

### （1）科学主题和技术主题的高价值性

有组织的产学研协同创新在选题伊始要紧扣国家战略需求<sup>[1]</sup>。党的二十大报告提出“集聚力量进行**原创性引领性**科技攻关，坚决打赢关键核心技术攻坚战，加快实施一批具有**战略性全局性前瞻性**的国家重大科技项目，增强自主创新能力”，为主题识别提供了方向指引和基本遵循。基于此，本研究提出**新颖性、基础性、广泛性、成长性和前瞻性**五项指标，其中，前三项指标主要反映主题当前的价值，后两项指标侧重主题未来的发展潜力。

①**新颖性**：指主题出现时间较短，且内容与以往或当前的研究主题相比展现出显著的创新性<sup>[12]</sup>，能有效体现**原创性**。

②**基础性**：指主题能为其他研究提供基础性的研发手段和技术支持，在科学研究和产业应用中占据重要地位<sup>[13]</sup>，能有效体现**战略性**。

③**广泛性**：指主题具有在多种产品和技术工艺中广泛应用的潜力，所包含的知识和技术能实现跨机构、跨领域的扩散传播<sup>[14]</sup>，能有效体现**全局性**。

④**成长性**：指主题处于其生命周期的成长阶段，所包含的知识和技术展现出快速增长趋势<sup>[15]</sup>，能有效体现**引领性**。

⑤**前瞻性**：指主题代表了领域未来的发展趋势和方向，具有引导未来科学研究和产业发展的潜力<sup>[16]</sup>。

### （2）科技主题对的强相关性

研究发现，企业和学研机构在社会职能和目标上存在差异，二者在合作过程中易产生分歧和摩擦，会影响产学研协同创新的效率与成效<sup>[17]</sup>。因此，本研究通过语义分析挖掘具有内在强相关性的科技主题对，有组织地推动多元诉求协调共治。

## 2 研究思路与方法

本文用论文表征学研机构开展的科学研究、专利表征企业开展的技术研究，引入复杂网络分析、深度学习和语义分析方法开展合作主题挖掘研究，研究框架见图 1。

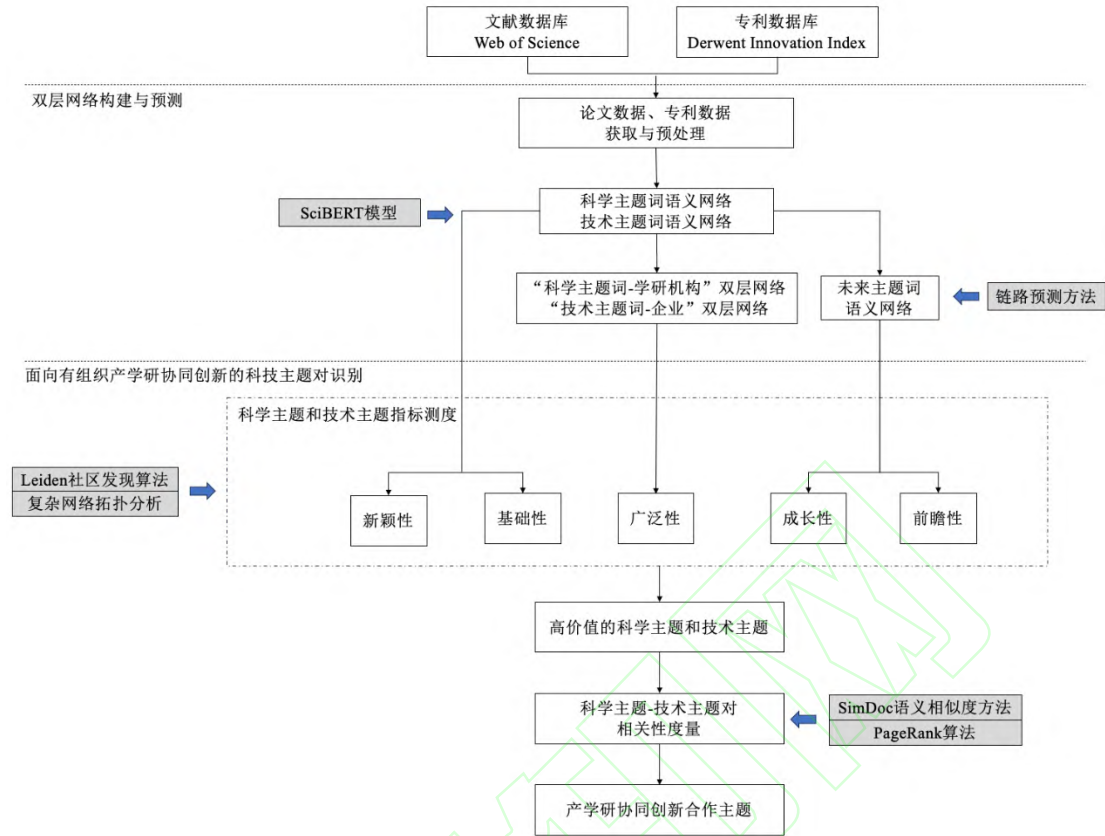


图 1 研究框架

Figure 1 Research framework

## 2.1 双层网络构建与预测

由于五项指标中的“广泛性”表征主题词在不同机构间的扩散程度，使用传统主题词单层网络无法实现有效测度。因此，本文构建了主题词-机构双层网络，不但可以实现对主题词多元特征的测度，更能有效反映主题词和机构间的耦合关系。

首先，基于 WoS 文献数据库和 DII 专利数据库，围绕特定研究领域下载论文和专利数据，并利用文本挖掘软件 ITGInsight<sup>1</sup>和自然语言处理技术<sup>[18]</sup>对数据进行处理和清洗，得到论文和专利主题词集合，以及对应的学研机构和企业集合。然后，分别构建“科学主题词-学研机构”双层网络（图 2a）和“技术主题词-企业”双层网络（图 2b）。以图 2a 为例，第一层是科学主题词语义网络，第二层是学研机构合著网络，两层网络都是无向图，层与层之间连线表示论文主题词与学研机构的共现关系。构建过程为：

### （1）科学主题词语义网络构建

SciBERT 是基于 BERT 的预训练语言模型，优化了对科技文献文本的处理能力<sup>[19]</sup>，因此本文运用 SciBERT 模型提取论文主题词的语义向量表征。然后，基于论文主题词间的语义相似度构建科学主题词语义网络。其中，网络节点为论文主题词，节点间的连边代表主题词间的语义关联，连边权重为主题词间的语义相似度。使用相同方法构建技术主题词语义网络，为后续“新颖性”和“基础性”指标测度提供网络分析基础。

### （2）学研机构合作网络构建

<sup>1</sup> ITGInsight 是面向文献计量数据的文本挖掘与可视化软件，详情见官方网站：[cn.itginsight.com](http://cn.itginsight.com)

构建学研机构合作网络，网络中的节点集合是学研机构集合，节点间的连边代表学研机构间的合著关系，连边权重为学研机构间合著论文的次数。

### (3) “科学主题词-学研机构” 双层网络构建

基于论文主题词与学研机构的共现关系构建双层网络，两层网络节点间的连边权重为论文主题词与学研机构的共现次数。

同理，使用相同方法构建“技术主题词-企业” 双层网络，为后续“广泛性” 指标测度提供网络分析基础。

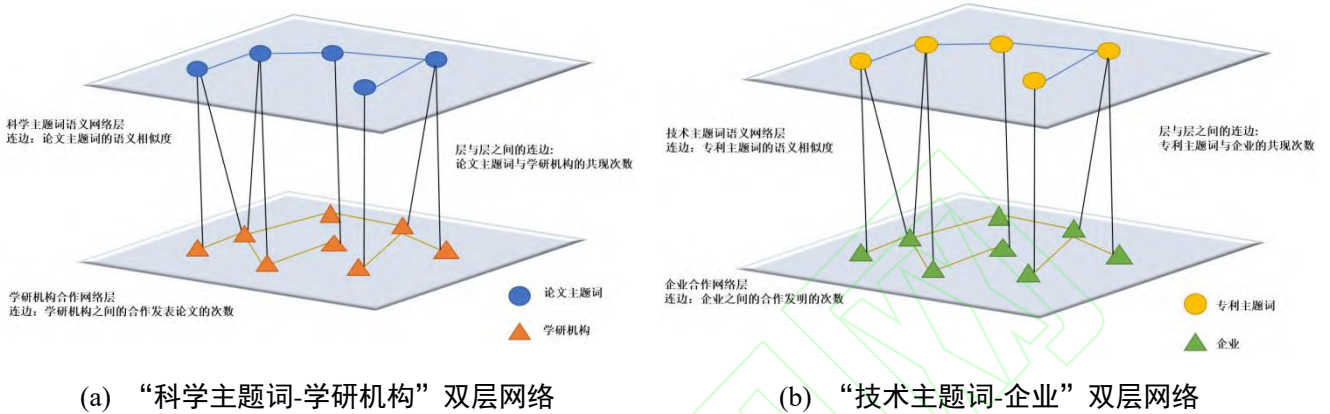


图 2 “主题词-机构” 双层网络

Figure 2 “Keyword - institution” bi-layer network

最后，由于“成长性”和“前瞻性”两项指标更关注主题未来的发展趋势和潜力，本文引入基于 Node2Vec 的链路预测方法，得到未来的科学主题词语义网络和技术主题词语义网络。

## 2.2 面向有组织产学研协同创新的科技主题对识别

### 2.2.1 科学主题和技术主题指标测度

本小节首先利用 Leiden 社区发现算法分别在科学主题词语义网络和技术主题词语义网络中识别科学社区（主题）和技术社区（主题），为后续的高价值产学研合作主题遴选提供分析基础。这里，Leiden 是一种基于模块度的社区发现算法，能够在快速运行的同时保证社区内部的紧密连接，更好地捕捉网络中的社区结构<sup>[20]</sup>。

使用 5 项指标进行测度，识别出服务国家战略需求的高价值科学主题和技术主题。其中，“新颖性”和“基础性”基于主题词语义网络进行测度，“广泛性”指标基于双层网络进行测度，“成长性”和“前瞻性”指标基于预测网络进行测度。

#### (1) 新颖性

当前关于主题新颖性的测度方法大多考虑主题的出现时间，忽略了内容上的创新性<sup>[12]</sup>，本文综合考虑主题在时间<sup>[21]</sup>和内容<sup>[22]</sup>维度上的新颖程度，并使用熵权法<sup>[23]</sup>赋权，测算主题新颖性。主题 $t_c$ 在时间上的新旧程度 $Time\_Novelty$ 和内容上的新颖程度 $Content\_Novelty$ 计算公式为

$$Time\_Novelty = \frac{1}{n} \sum_{i=1}^n y_i - \frac{1}{h} \sum_{k=1}^h Y_k \quad (1)$$

$$Content\_Novelty = 1 / \left( \frac{1}{m-1} \sum_{j=1}^{m-1} Cos(t_c, t_j) \right) \quad (2)$$

$$Cos(t_c, t_j) = \sum_{p=1}^q \sum_{i=1}^n Cos(i, p) \quad (3)$$

其中,  $n$ 为该主题 $t_c$ 中的主题词总数,  $y_i$ 为该主题中主题词 $i$ 的出现时间;  $h$ 为主题词语义网络中的主题词总数,  $Y_k$ 为网络层中主题词 $k$ 的出现时间;  $m$ 为主题词语义网络中的主题总数,  $Cos(t_c, t_j)$ 为主题 $t_c$ 与主题 $t_j$ 之间的余弦相似度;  $q$ 为主题 $t_j$ 中的主题词总数,  $Cos(i, p)$ 为主题 $t_c$ 中的主题词 $i$ 和主题 $t_j$ 中的主题词 $p$ 之间的余弦相似度。

#### (2) 基础性

参考 Zhang 等的研究<sup>[13]</sup>, 本文综合考虑主题的度中心性、接近中心性和中介中心性, 分析主题在网络中的全局结构特征和重要性。其中, 度中心性能反映节点的局部连接程度, 接近中心性侧重节点在网络中的全局连接程度, 中介中心性关注节点在网络中的信息传递作用, 故综合三者能更全面地表征节点在网络中的重要性。然后, 使用熵权法计算主题的基础性。

#### (3) 广泛性

参考 Zhang 等的研究<sup>[13]</sup>, 本文基于双层网络度量主题在不同机构间的扩散程度。当一个主题隶属多个机构, 说明其更容易扩散到多个机构, 被多个机构所应用。主题的广泛性 $Width$ 计算公式为

$$Width = \frac{1}{n} \sum_{i=1}^n D_i \quad (4)$$

$$D_i = \sum_{m=1}^{|V^a|} w_{i, v_m^a} \times L(v_m^a) \quad (5)$$

$$L(v_m^a) = \sum_{n=1}^{|V^a|} w_{v_m^a, v_n^a} \quad (6)$$

其中,  $D_i$ 代表主题词 $i$ 的广泛性,  $n$ 是该主题中主题词的总数;  $w_{i, v_m^a}$ 表示主题词语义网络的主题词 $i$ 和机构合作网络的机构 $v_m^a$ 之间的连边权重,  $L(v_m^a)$ 表示机构合作网络中与机构 $v_m^a$ 有链接的连边权重之和;  $w_{v_m^a, v_n^a}$ 表示机构合作网络中机构 $v_m^a$ 和 $v_n^a$ 之间的连边权重,  $V^a$ 代表机构合作网络的节点总数。

#### (4) 成长性

参考 Wang 的研究<sup>[24]</sup>, 本文使用未来主题词语义网络中主题词节点连边的增长率, 评估主题的成长性 $Growth$ , 计算公式为

$$Growth = \frac{1}{n} \sum_{i=1}^n r_{i, t, t'} \quad (7)$$

$$r_{i, t, t'} = \frac{Edge_{i, t'} - Edge_{i, t}}{t - t'} \quad (8)$$

其中,  $r_{i, t, t'}$ 为主题词 $i$ 从 $t$ 到 $t'$ 时间连边的增长率,  $n$ 为该主题中主题词的总数;  $Edge_{i, t'}$ 表示该主题中的主题词 $i$ 在 $t'$ 时刻(未来网络)的连边数,  $Edge_{i, t}$ 表示主题词 $i$ 在 $t$ 时刻(当前网络)的连边数。

#### (5) 前瞻性

参考 Huang 等的研究<sup>[25]</sup>, 本文基于 PageRank 算法测度主题的前瞻性 *Foresight*, 以综合考量节点的局部拓扑结构以及连边节点的重要性等多重因素, 其计算公式为

$$Foresight = \frac{1}{n} \sum_{i=1}^n PR_i \quad (9)$$

$$PR_i = (1 - d) + d \times \sum_{j \in L(i)} \frac{1}{|Count(j)|} PR_j \quad (10)$$

其中,  $PR_i$  是该主题中主题词  $i$  在未来主题词语义网络中的 PageRank 值,  $n$  表示该主题中的主题词总数;  $d$  为阻尼因子,  $1 - d$  表示随机跳转到其他主题词的概率, 一般取  $d = 0.85$ <sup>[26]</sup>;  $L(i)$  表示与主题词  $i$  存在链接关系的所有主题词的集合,  $Count(j)$  表示与主题词  $j$  存在链接关系的主题词数量。

计算得到每个主题的 5 项指标值后, 使用熵权法进行赋权, 最终得到每个主题对应的综合得分, 本研究将得分排名靠前的主题视为高价值科学主题和技术主题。

## 2.2.2 科技主题对相关程度量

本小节主要分析高价值的科学和技术两类主题间的相关性, 遴选出有较大产学研协同潜力的科技主题对。考虑到每个主题包含多个主题词, 本文将主题内不同主题词的重要性信息纳入度量模型进行考虑, 以提高重要主题词的权重, 从而更精确地捕捉高价值科学主题和技术主题之间的潜在语义关联。具体来看, 对 Maheshwari 等提出的文本相似度指标 SimDoc<sup>[27]</sup> 进行改进, 采用主题词的 PageRank 值为主题词向量赋权, 使用加权求和方法得到主题向量, 最后基于余弦相似度计算科技主题间的相关性。科学主题  $t_i^s$  和技术主题  $t_j^t$  间的相关性  $Correlation(t_i^s, t_j^t)$  具体计算公式为

$$Correlation(t_i^s, t_j^t) = Cos(Vec(t_i^s), Vec(t_j^t)) \quad (11)$$

$$Vec(t_i^s) = \frac{1}{P_i^s} \sum_{k=1}^{P_i^s} PR(V_k^s)' \times X_k \quad (12)$$

$$Vec(t_j^t) = \frac{1}{Q_j^t} \sum_{h=1}^{Q_j^t} PR(V_h^t)' \times Y_h \quad (13)$$

其中,  $Vec(t_i^s)$  为科学主题  $t_i^s$  的向量表示,  $Vec(t_j^t)$  为技术主题  $t_j^t$  的向量表示,  $Cos(Vec(t_i^s), Vec(t_j^t))$  为  $Vec(t_i^s)$  和  $Vec(t_j^t)$  之间的余弦相似度;  $P_i^s$  表示科学主题  $t_i^s$  中论文主题词总数,  $Q_j^t$  表技术主题  $t_j^t$  中专利主题词总数;  $X_k$  表示科学主题  $t_i^s$  内论文主题词  $V_k^s$  的向量表示,  $Y_h$  表示技术主题  $t_j^t$  内专利主题词  $V_h^t$  的向量表示;  $PR(V_k^s)'$  表示论文主题词  $V_k^s$  经过极差标准化后的 PageRank 值,  $PR(V_h^t)'$  表示专利主题词  $V_h^t$  经过极差标准化后的 PageRank 值。

## 3 实证研究

随着算法复杂度和精度不断增加、数据规模指数级增长、算力持续提升, 人工智能 (AI) 在全球蓬勃发展, 该领域科学与技术相互交织、彼此推动, 是产学研协同创新的典型代表。近年来, 我国政府持续加强 AI 战略布局, 强化科研机构与企业之间的协作网络, 2022 年, 科技部、工业和信息化部等六部门发布《关于加快场景创新以人工智能高水平应用促进经济高质量发展的指导意见》, 指导企业、学研机构等创新主体协同促进 AI 场景创新, 着力解决 AI 应用和产业化难题。在实践中, AI 领域产学研协同创新质量和效率亟待进一步提升<sup>[5]</sup>, 本文选择 AI 领域开展实证分析。

### 3.1 人工智能领域面向有组织产学研协同创新的合作主题识别

本文分别从 WoS 文献数据库和 DII 专利数据库获取 AI 领域论文和专利数据。其中，论文数据检索策略主要借鉴 Liu<sup>[28]</sup>等人的研究，基于核心关键词和 AI 学科相关的学术期刊进行搜索，构建初始论文数据库，之后提取这些论文中的高频关键词形成扩展关键词，进而建立完整的检索框架；专利数据检索策略是参考 Zhou 等人<sup>[29]</sup>的研究，从 AI 的技术支持、基础设施、算法和应用三方面构建专利检索框架。最终获得 2018-2022 年 75800 条论文数据和 123551 条专利数据。

然后，参考 Wang 等<sup>[30]</sup>的研究，利用 ITGInsight 文本挖掘软件分别遴选词频及长度大于等于 2、PC-Value（术语度）最大的 6000 个论文主题词和专利主题词作为主题词集合；分别遴选发表论文和专利数量排名前 2000 的学研机构和企业构建机构集合。

按照 3.1.1 的步骤构建“科学主题词-学研机构”双层网络和“技术主题词-企业”双层网络，并预测得到未来的科学主题词语义网络和技术主题词语义网络。接着，利用 Leiden 社区发现算法分别识别出 AI 领域 64 个科学社区（主题）和 51 个技术社区（主题）。计算网络中每个主题词对应的 PageRank 值，将数值最高的主题词作为主题的名称，并计算每个主题的新颖性、基础性、广泛性、成长性、前瞻性指标。为了保持指标的一致性，本文使用 min-max 归一化方法对所有指标进行标准化，并对 5 项指标使用熵权法得到每个主题的综合得分。将得分前 20 的主题作为具有高价值的科学主题和技术主题，如表 1 和表 2 所示。



表 1 Top-20 科学主题及指标值

Table 1 Top-20 science topics and indicator values

序号	科学主题名称	新颖性	基础性	广泛性	成长性	前瞻性	综合得分
1	神经网络	0.2493	0.7500	0.6568	0.6196	1.0000	0.6517
2	社交网络分析	0.2336	0.6749	1.0000	0.4377	0.8252	0.6370
3	人机交互	0.4298	0.6376	0.1506	1.0000	0.9378	0.6216
4	面部表情识别	0.6819	0.4305	0.1878	0.8243	0.7998	0.5816
5	视觉图像分析	0.4020	0.5102	0.3734	0.6912	0.8398	0.5602
6	分类方法	0.1095	0.7855	0.6802	0.4534	0.5839	0.5182
7	对抗学习	0.5895	0.5586	0.1958	0.8998	0.3429	0.5102
8	工程设计	1.0000	0.5259	0.0166	0.7395	0.2688	0.5092
9	聚类算法	0.1330	0.6468	0.5549	0.3582	0.7888	0.4937
10	优化算法	0.5731	0.4331	0.4422	0.6441	0.3774	0.4935
11	大模型训练	0.0654	1.0000	0.8296	0.2817	0.2337	0.4767
12	金融数据分析	0.5228	0.6198	0.1818	0.9514	0.1213	0.4700
13	智能设备	0.2374	0.7556	0.1533	0.8247	0.3847	0.4587
14	统计学方法	0.2601	0.8248	0.1197	0.5340	0.5723	0.4509
15	在线学习	0.2679	0.7827	0.0838	0.6477	0.4479	0.4339
16	数值模拟	0.1452	0.8576	0.3840	0.5708	0.2225	0.4257
17	模糊数学模型	0.0567	0.8807	0.2454	0.8108	0.1639	0.4167
18	随机模型	0.1984	0.8395	0.4185	0.3268	0.2340	0.3960
19	距离函数	0.1032	0.6736	0.3709	0.3865	0.2311	0.3433
20	遗传算法	0.1474	0.4249	0.1893	0.4735	0.4137	0.3244

表 2 Top-20 技术主题及指标值

Table 2 Top-20 technology topics and indicator values

序号	技术主题名称	新颖性	基础性	广泛性	成长性	前瞻性	综合得分
1	深度学习	0.8587	0.8991	0.6765	0.6496	1.0000	0.8177
2	计算机视觉	1.0000	0.7201	0.5545	0.7375	0.8220	0.7713
3	神经网络	0.7385	0.8715	0.7135	0.5730	0.8571	0.7503
4	知识图谱	0.7100	0.5494	0.2978	0.6943	0.9233	0.6369
5	计算设备	0.4295	1.0000	0.1810	0.8981	0.7289	0.6347
6	医学影像	0.6523	0.6651	0.2662	0.7852	0.6201	0.5950
7	人脸识别	0.5250	0.5739	0.1103	1.0000	0.7785	0.5923
8	目标检测	0.5184	0.7892	0.2370	0.6735	0.6426	0.5659
9	大数据处理	0.1167	0.8718	1.0000	0.4728	0.3261	0.5463
10	控制系统	0.4513	0.7034	0.1096	0.7606	0.3904	0.4750
11	语音识别	0.4273	0.7306	0.4985	0.3470	0.3877	0.4741
12	信息处理系统	0.3944	0.5847	0.3421	0.6292	0.3107	0.4467
13	特征提取	0.2900	0.4498	0.6858	0.3827	0.4296	0.4462
14	预测模型	0.4278	0.3977	0.3904	0.6169	0.3708	0.4393
15	智能设备	0.6300	0.4677	0.1325	0.6844	0.2425	0.4304
16	三维模型	0.3500	0.6724	0.0708	0.7711	0.2710	0.4171
17	分类方法	0.3500	0.3728	0.6925	0.2200	0.3531	0.3996
18	图像处理	0.2917	0.4679	0.2160	0.4725	0.4927	0.3843
19	阈值设置方法	0.2970	0.6751	0.3564	0.3385	0.2856	0.3843
20	质量控制	0.4355	0.4874	0.1426	0.5005	0.3468	0.3798

表 1 显示，科学主题主要包括 AI 领域的核心算法理论（如神经网络、分类方法、聚类算法等）、关键方法研究（如人机交互、视觉图像分析、大模型训练等），以及 AI 交叉领域的核心理论方法（如社交网络分析、金融数据分析、数值模拟等）。这一结果能为政府部门聚焦关键基础科学问题提供决策支持。例如，2023 年 4 月，国家自然科学基金委员会发布的《可解释、可通用的下一代人工智能方法重大研究计划 2023 年度项目指南》指出，要围绕神经网络等基础科学问题开展研究，重点支持人机交互、多模态数据分析、数值方法等研究方向。2023 年 11 月，中国、美国、欧盟等 28 个国家共同签署《布莱切利宣言》，重点聚焦包括基础模型、大语言模型等 AI 大模型训练问题。这也体现本文高价值科学主题识别结果的可靠性。

表 2 显示，技术主题主要包括 AI 领域的关键技术（如计算机视觉、语音识别等）、应用场景（如医学影像、智能设备等），以及底层基础设施建设（如计算设备、大数据处理等）。这一结果能为相关部门制定关键技术发展规划提供重要支撑。例如，2022 年 7 月，科技部等六部门发布的《关于加快场景创新以人工智能高水平应用促进经济高质量发展的指导意见》

指出，要着力打造医学影像智能辅助诊断、智能设备应用等 AI 场景，加强 AI “数据底座”建设，提升 AI 场景创新的算力支撑。2023 年 5 月，美国发布的《国家人工智能研究与发展战略计划》提到，要对计算机视觉、医学影像等领域进行长期投资，并重点关注数据、计算等关键基础设施建设，实现对大规模数据的快速访问和处理。这体现出本文高价值技术主题识别结果的可靠性。

从表 1 和表 2 的主题分布来看，存在如“神经网络”“分类方法”“智能设备”等共性主题。同时，高价值科学主题和技术主题也呈现出差异性，其中，科学主题更加聚焦神经网络训练算法、人机交互方法、大模型训练算法等基础理论研究，而技术主题更加重点关注计算机视觉技术、计算设备、大数据处理技术等应用研究。因此，本文进一步通过语义分析，分析高价值科学和技术主题间的相关性，识别具有较大产学研协同潜力的科技主题对，相关性热力图结果如图 3 所示。

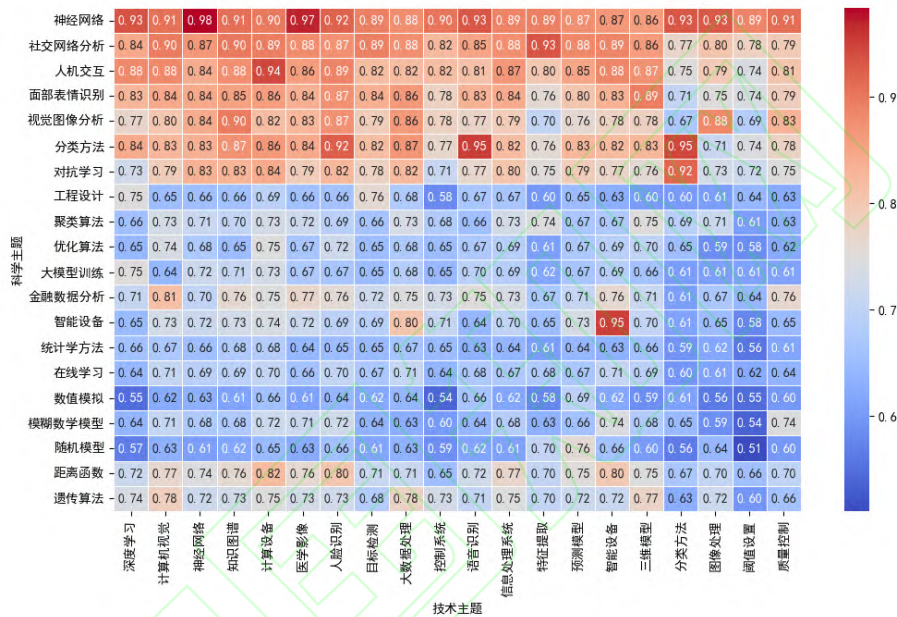


图 3 高价值科学主题和技术主题相关性热力图

Figure 3 Heat map of the correlation between high-value science and technology topics

图 3 显示，AI 领域相关性在 0.9 及以上的科技主题对有 22 对（占总数 5.5%）。尤其，科学主题“神经网络”作为 AI 领域的核心算法，与所有高价值技术主题的相关性均超过 0.85，这表明要进一步重视发挥“神经网络”在 AI 领域的基石作用，通过推进产学研协同创新，强化“神经网络”在图像处理、语音识别、计算机视觉、医学影像分析等领域的应用。

在上文提到的三大共性主题（“神经网络”“分类方法”和“智能设备”）构成的科技主题对之外，本部分选取相关性得分较高的另外三对科技主题对进行分析，探究推进有组织产学研协同创新的实践路径。

(1) 神经网络-医学影像（科技主题相关性 0.97）

建议政府部门有组织地引导“神经网络”科学主题下的学研机构和“医学影像”技术主题下的企业，重点围绕“卷积神经网络”“图神经网络”等科学子主题和“三维 CT 图像”“超声检查”等技术子主题推进人才、资金、数据等多类产学研资源的流动、共享与集成，并聚焦“三维图像重建”等技术难题进行联合攻关，助推医学影像技术向数字化、精准化、智慧化迈进。

(2) 分类方法-语音识别（科技主题相关性 0.95）

随着以 ChatGPT 为代表的大语言模型技术的逐步成熟,语音识别技术迎来了新的发展机遇,而分类方法在其中发挥着关键作用。建议相关部门引导学研机构和企业集中优势资源围绕“基于注意力机制模型”“长短时记忆网络”等科学子主题和“语音情感”“语音转写”等技术子主题开展协同创新,共同攻克“语音情感识别”“语音实时分析”等技术难题,实现智能语音交互和情感智能分析。

### (3) 人机交互-计算设备(科技主题相关性 0.94)

人工智能大模型的出现推动通用 AI 发展的同时也给 AI 计算设备带来了算力和能耗挑战,亟需开展跨领域、跨任务场景的人机交互方法研究。建议学研机构和企业共同围绕“三维交互”“数字孪生”等科学子主题和“云计算”“GPU”等技术子主题开展协同创新,联合攻关“多模态感知”“数智计算”等技术难题,提升计算设备的数智化交互体验。

总结来看, AI 正处于“强算法+大数据+大算力”的发展阶段,例如,学研机构开展了大量“神经网络+大模型训练”的强算法科学基础研究,企业基于“大模型”对大数据平台进行智能化升级,并不断优化算力基础设施以赋能“数智化转型”。本文开展的科技主题相关性分析结果也进一步证明 AI 领域的科学和技术紧密交织,学研机构的研究成果正不断向各行各业渗透,从单点应用向多元化应用、从通用场景向行业特定场景、跨领域场景不断深入;同时,企业的技术需求也为学研机构的基础研究提供了驱动力,推动了传统 AI 算法的优化和升级,催生了新兴的科学研究主题。基于此,建议相关管理部门打造一批产学研协同创新示范中心,有组织地推动创新链产业链各创新主体的深度协同,例如,在医学影像、语音识别、图像处理领域中加快神经网络算法的应用,在特征提取、计算机视觉、知识图谱领域中推进社交网络分析方法的应用,优化产学研创新资源配置。

## 3.2 方法验证

### 3.2.1 链路预测模型验证

为定量验证本文训练得到的基于 Node2Vec 的链路预测模型的性能优势,我们在科学主题词语义网络和技术主题词语义网络上基于 AUC、和 Precision 两个评估指标,将本文方法与 4 种传统的链路预测方法进行对比,包括基于 Adamic-Adar(AA)指标的方法、基于 Jaccard 指标的方法、基于 Preferential attachment(PA)指标的方法和基于谱聚类(Spectral Clustering, SC)的方法,结果如表 3 所示。

表 3 链路预测模型性能对比

Table 3 The comparison of link prediction model performance

方法	科学主题词语义网络		技术主题词语义网络	
	AUC	Accuracy	AUC	Accuracy
基于 Node2Vec 的链路预测模型 (本文方法)	<b>0.997</b>	<b>0.996</b>	<b>0.995</b>	<b>0.994</b>
基于 AA 指标的方法	0.914	0.896	0.917	0.886
基于 Jaccard 指标的方法	0.956	0.943	0.933	0.925
基于 PA 指标的方法	0.815	0.829	0.909	0.916
基于 SC 的方法	0.584	0.580	0.673	0.582

可以看出, 本文方法在科学主题词语义网络和技术主题词语义网络上的两个评估指标均优于 4 种传统的链路预测方法, 模型性能较好。

### 3.2.2 科学主题和技术主题识别方法验证

为定量验证本文使用的 Leiden 社区发现算法进行主题识别的有效性, 我们在科学主题词语义网络和技术主题词语义网络上基于模块度指标, 将本文方法与层次聚类算法、谱聚类算法、K-Means 算法进行对比。模块度是衡量网络社区划分效果的指标, 模块度越大, 社区内部连接越紧密, 社区之间连接越稀疏, 表示社区划分效果越好<sup>[20]</sup>, 故本文分别对 4 种方法得到的聚类结果计算模块度, 以此衡量科学主题和技术主题识别效果, 结果如图 4 所示。

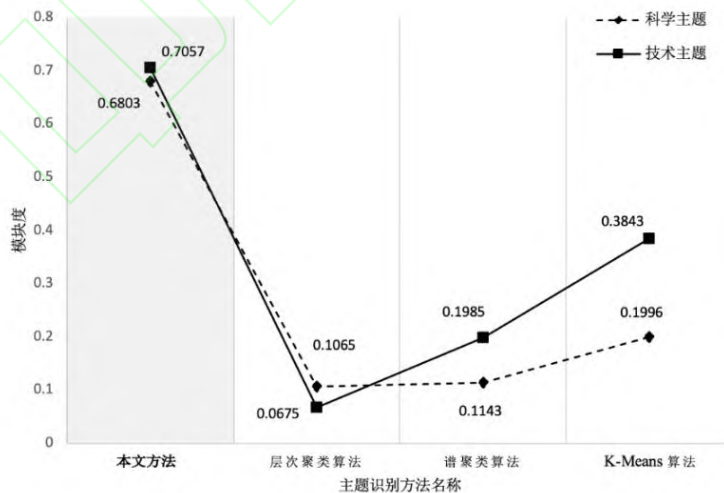


图 4 主题识别结果对比

Figure 4 Comparison of topic identification results

可以看出, 本文所用方法在科学主题识别的模块度 (0.6803) 和技术主题识别的模块度 (0.7057) 上的表现均优于其他 3 种传统的主题识别算法。

## 4 结论与讨论

### 4.1 研究结论

本文提出了一套面向有组织产学研协同创新的合作主题挖掘方法，通过深度剖析“有组织产学研协同创新”的内涵和特征，构建与之相适应的科学和技术主题评价指标体系，并综合使用复杂网络分析和深度学习算法对指标进行精准测度，识别高价值科学主题和技术主题；同时，对传统的文本语义相似度算法进行改进，重点考量了主题内部不同主题词的重要性信息和语义信息，实现对高价值科学主题和技术主题之间语义相关性的精准度量，进而遴选出具有较大协同潜力的产学研合作主题。

本研究分别识别出 AI 领域 20 个高价值科学主题和技术主题，结果显示，科学主题更加聚焦神经网络训练算法、人机交互方法、大模型训练算法等基础理论研究，而技术主题更加重点关注计算机视觉技术、计算设备、大数据处理技术等应用研究。科技主题相关性分析结果表明，AI 高价值科技主题间的整体相关性均在 0.5 以上，得分在 0.9 及以上的科技主题对共有 22 对，表明在 AI 领域开展有组织的产学研协同创新具有良好的基础。其中，科学主题“神经网络”与高价值技术主题之间的相关性分值均超过 0.85，凸显了神经网络相关理论方法在 AI 领域的基础和应用研究中扮演着关键桥梁作用。

### 4.2 理论贡献

本研究的理论贡献如下：

(1) 深化了产学研协同创新理论研究。本研究将“有组织”理念融合到产学研高效协同体系设计中，通过深度剖析“有组织产学研协同创新”的内涵，提出了在此场景下的产学研合作主题应该具有高价值性和强相关性两大特征。

(2) 拓展了复杂网络分析和深度学习算法在产学研协同创新研究领域中的应用。本研究构建了新颖性、基础性、广泛性、成长性和前瞻性五大指标，并综合运用链路预测、拓扑结构分析等复杂网络分析方法和 SciBERT 深度学习模型精准捕获网络的拓扑结构和动态变化特征，提升了高价值科学主题和技术主题识别的科学性和准确性。

(3) 融合机器学习和文本语义分析方法为产学研合作主题内在关联的深度挖掘提供了有益补充。本研究改进了传统的文本语义相似度方法，能有效捕捉主题词的重要性信息和语义信息，提升科学主题和技术主题之间相关性度量的精准度，该指标还可以用于协同创新路径识别、知识流动效率测度等研究中。

(4) 丰富了创新资源优化配置和科技成果转化等领域的理论方法研究。本研究提出的科学主题和技术主题的关联分析方法，能够为推动创新链产业链深度融合和开展创新生态系统治理提供重要的理论支撑。

### 4.3 实践启示

本研究能为政府部门有组织地引导企业、高校和科研院所开展高层级产学研协同创新提供重要决策支持，并在促进创新链产业链深度融合的实践中发挥重要量化支撑作用。一方面，能科学高效地整合产学研优势资源，引导各方集中精力围绕高价值且有协同创新潜力的科技主题开展合作，提高产学研协同创新质量和效率；另一方面，可以有效推进重要领域的科技成果转化和产业应用布局，引导科研机构以实际需求为导向开展前沿创新研究，促进产学研协同创新成果在产业实际应用中的迅速落地和广泛推广。例如，医学影像领域企业可以通过

设立企业创新研究专项基金，联合高校共同围绕医学影像领域的“图像融合”“图像超分辨率重建”等技术难题进行攻关，并聚焦图神经网络在医学图像中的迁移学习、医学图像的数据增强和自动标注方法等方向开展基础研究；AI 领域学研机构可以主动与医学影像领域企业开展深度合作，共同推动基于神经网络的深度学习、自然语言处理和大语言模型等前沿科研成果在医学影像领域的开发与应用，例如，使用卷积神经网络进行多模态医学图像识别与分割，将残差神经网络用于超分辨率医学影像分析，提高疾病诊断的准确性和效率。

#### 4.4 研究局限与展望

本研究也存在一些不足。首先，产学研协同创新不仅涵盖了科学研究和技术开发，还包括创新政策的制定、创新环境的构建以及创新文化的培育，未来应融合政策文件、产业报告等多源异构数据开展研究；其次，有组织的产学研协同创新在不同的发展阶段具有不同的目标和特点，后续将深入探讨各个阶段的特征，并开展有针对性、差异化的研究；此外，未来将围绕未来产业和战略性新兴产业，开展更为广泛的实证分析，提升方法的可靠性和实用性。

#### 参考文献：

- [1] 潘教峰,鲁晓,王光辉.科学研究模式变迁:有组织的基础研究[J].中国科学院院刊,2021,36(12):1395-1403. Pan J F, Lu X, Wang G H. Transforming scientific research: organized basic research[J]. Bulletin of Chinese Academy of Sciences,2021,36(12):1395-1403.
- [2] 曾明彬,李玲娟.我国基础研究管理制度面临的挑战及对策建议[J].中国科学院院刊,2019,34(12):1440-1447. Zeng M B, Li L J.Challenges and countermeasures of management System of basic research in China[J]. Bulletin of Chinese Academy of Sciences,2019,34(12):1440-1447.
- [3] D'Este P, Llopis O, Rentocchini F, et al. The relationship between interdisciplinarity and distinct modes of university-industry interaction[J], Research Policy. 2019, 48(9): 1-23.
- [4] 陈劲,阳镇.融通创新视角下关键核心技术的突破:理论框架与实现路径[J], 社会科学, 2021(5):58-69. Chen J, Yang Z. The breakthrough of key core technology under the perspective of integration innovation: theoretical framework and realization path[J].Journal of Social Sciences, 2021(5):58-69.
- [5] 涂振洲, 顾新. 基于知识流动的产学研协同创新过程研究 [J]. 科学学研究,2013,31(9):1381-1390. Tu Z Z, Gu X.Study on process of industry - university - research institute collaborative innovation based on knowledge flow[J]. Studies in Science of Science, 2013,31(9):1381-1390.
- [6] 贾宝余,杨明,应验.高水平科技自立自强视野中重大科技项目选题机制研究[J].中国科学院院刊,2022,37(9):1226-1236. Jia B Y, Yang M, Ying Y.Subject selection mechanism of major science and technology programs from perspective of high-level self-reliance in science and technology[J]. Bulletin of Chinese Academy of Sciences, 2022, 37(9):1226-1236.
- [7] Zhu G, Wen M, Fan X, et al.A case study on the mechanism of university-industry collaboration to improve enterprise technological capabilities from the perspective of capability structure[J].Innovation and Development Policy.2020,2(2): 99-125.
- [8] 万劲波,张凤,潘教峰.开展“有组织的基础研究”:任务布局与战略科技力量[J].中国科学院院刊,2021,36(12):1404-1412. Wan J B, Zhang F, Pan J F. Promoting organized basic research: strategic layout and strategic capacity in science and technology[J]. Bulletin of Chinese Academy of Sciences, 2021, 36(12): 1404-1412.

- [9] 赵宏媚,赵文华,姚建建.高校有组织科研核心要素响应机制研究[J/OL].科学学研究, 2023 [2024-3-25].<https://doi.org/10.16192/j.cnki.1003-2053.20230810.001>.Zhao H M, Zhao W H, Yao J J. Research on the core elements and response mechanism of organized research in universities[J/OL]. Studies in Science of Science, 2023 [2024-3-25].<https://doi.org/10.16192/j.cnki.1003-2053.20230810.001>.
- [10] 张煜,苏竣.创新链视角下产学研合作、高校创新与技术进步[J].科学学研究,2023,41(7):1294-1305,1315. Zhang Y, Su J. Industry-university-research cooperation, universities' innovation and technological progress from the perspective of innovation chain[J]. Studies in Science of Science, 2023,41(7):1294-1305,1315.
- [11] 曾粤亮,司莉.自由探索还是有组织科研:跨学科科研合作的内涵、类型与特点[J].图书情报知识,2023,40(4):81-91,51. Zeng Y L, Si L. Free exploration or organized research: connotation, types and characteristics of interdisciplinary research collaboration[J]. Documentation, Information & Knowledge, 2023,40(4):81-91,51
- [12] Rotolo D, Hicks D, Martin B R. What is an emerging technology?[J]. Research Policy, 2015, 44(10): 1827-1843.
- [13] Zhang Y, Wu M, Miao W, et al. Bi-layer network analytics: A methodology for characterizing emerging general-purpose technologies[J]. Journal of Informetrics, 2021, 15(4): 101202.
- [14] Petralia S. Mapping general purpose technologies with patent data[J]. Research Policy, 2020, 49(7): 104013.
- [15] Xu H, Winnink J, Yue Z, et al. Multidimensional scientometric indicators for the detection of emerging research topics[J]. Technological Forecasting and Social Change, 2021, 163: 120490.
- [16] 钟荣丙.引领性技术创新: 界定、评价与治理[J].创新,2022,16(5):48-60. Zhong B R. Leading technological innovation: defining, evaluating and governing[J]. Innovation, 2022, 16(5): 48-60.
- [17] 孙成,崔维军,陈光.大学、公共研究机构及企业对产学研合作问题的认知差异——基于日本深度访谈的文本分析[J].中国科技论坛,2021(2):179-188.Sun C, Cui W J, Chen G. The cognitive differences of the industry-university-institute cooperation problem between universities, public research institutions and enterprises - a text analysis based on in-depth interviews in Japan[J]. Forum on Science and Technology in China, 2021(2):179-188.
- [18] Zhang Y, Porter A L, Hu Z, et al. "Term clumping" for technical intelligence: a case study on dye-sensitized solar cells[J]. Technological Forecasting and Social Change, 2014(85): 26-39.
- [19] Dinh T N, Pham P, Nguyen G L, et al. Enhancing local citation recommendation with recurrent highway networks and SciBERT-based embedding[J]. Expert Systems with Applications, 2024, 243: 122911.
- [20] Traag V A, Waltman L, Van Eck N J. From louvain to leiden: guaranteeing well-connected communities[J]. Scientific Reports, 2019, 9(1): 5233.
- [21] Xu S, Hao L, Yang G, et al. A topic models based framework for detecting and forecasting emerging technologies[J]. Technological Forecasting and Social Change, 2021, 162: 120366.
- [22] 徐路路,靳杨.基于 FSD 模型的政府资助项目新兴主题探测与分析[J].科学学与科学技术管理,2019,40(2):40-54.Xu L L, Jin Y. Emerging topics detection and analysis of government funded projects based on FSD model[J]. Science of Science and Management of S.& T.,2019,40(2):40-54.
- [23] 章穗,张梅,迟国泰.基于熵权法的科学技术评价模型及其实证研究[J].管理学报,2010,7(1):34-42. Zhang S, Zhang M, Chi G T. The science and technology evaluation model



- based on entropy weight and empirical research during the 10th five-year of China[J]. Chinese Journal of Management, 2010, 7(1): 34-42.
- [24] Wang Q. A bibliometric model for identifying emerging research topics[J]. Journal of the Association for Information Science and Technology, 2018, 69(2): 290-304.
- [25] Huang L, Chen X, Ni X, et al. Tracking the dynamics of co-word networks for emerging topic identification[J]. Technological Forecasting and Social Change, 2021, 170: 120944.
- [26] Gleich D F. PageRank beyond the web[J]. Siam Review, 2015, 57(3): 321-363.
- [27] Maheshwari G, Trivedi P, Sahijwani H, et al. SimDoc: topic sequence alignment based document similarity framework[C]. Proceedings of the 9th Knowledge Capture Conference, 2017:1-8.
- [28] Liu N, Shapira P, Yue X. Tracking developments in artificial intelligence research: constructing and applying a new search strategy[J]. Scientometrics, 2021, 126(4): 3153-3192.
- [29] Zhou X, Huang L, Zhang Y, et al. A hybrid approach to detecting technological recombination based on text mining and patent network analysis[J]. Scientometrics, 2019, 121: 699-737.
- [30] Wang J, Xu B, Zu Y. Deep learning for aspect-based sentiment analysis[C]. International Conference on Machine Learning and Intelligent Systems Engineering (MLISE). IEEE,2021:267-271.

# Collaboration topic mining for organized university-industry-research synergy innovation

Huang Lu<sup>1,2</sup> Ren Hang<sup>1</sup> Cao Xiaoli<sup>3,4</sup> Chen Xiang<sup>5</sup>

*(1. School of Economics, Beijing Institute of Technology, Beijing 100081, China;*

*2. Digital Economy and Policy Intelligentization Key Laboratory of Ministry of Industry and Information Technology, Beijing 100081, China;*

*3. National Science Library, Chinese Academy of Sciences, Beijing 100190, China;*

*4. Department of Information Resources Management, School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190, China;*

*5. School of Management, Beijing Institute of Technology, Beijing 100081, China)*

**Abstract:** Organized industry-university-research (IUR) collaboration is crucial for leveraging China's new national system advantages, enhancing the overall efficacy of the national innovation system, and achieving deep integration of industry, university, and research. Identifying collaboration topics is paramount for high-quality, efficient synergy in such collaborations and constitutes a primary challenge. This paper proposes a set of identification methods for IUR collaboration topic pairs based on complex network analysis and deep learning algorithms. Firstly, this paper analyzes the concept of "Organized IUR Synergy Innovation" and related topic features, proposes that the topic of IUR collaboration should be high-value and strong-relevance. Secondly, a bi-layer network of "science keywords-universities and research institutions" and a bi-layer network of "technology keywords-enterprises" are constructed based on paper data and patent data. The SciBERT model is used to construct a semantic network of science keywords and a semantic network of technology keywords. The Node2vec-based link prediction model is used to generate the future semantic network of science keywords and the semantic network of technology keywords. Then, the five indicators of novelty, fundamentality, width, growth, and foresight are measured using methods of complex network topology analysis, community discovery, and machine learning, to identify high-value scientific and technological topics. Finally, the correlation between scientific and technological topics is calculated using the improved semantic similarity index, SimDoc. This enables the selection of "science-technology topic pairs" that exhibit high potential for collaboration across IUR. In this paper, an empirical study is conducted using paper and patent data in the field of "artificial intelligence (AI)" from 2018 to 2022 to validate the research method and results.

This study identifies 20 high-value scientific and technological topics in the field of AI. The results reveal that scientific topics focused more on basic theoretical research such as neural network training algorithms, human-computer interaction methods, and large model training algorithms, while technological topics focused more on application research such as computer vision technologies, computing devices, and big data processing techniques. The correlation analysis results between scientific and technological topics show that the overall correlation is above 0.5, and there are 22 science-technology topic pairs with scores of 0.9 or above, indicating that organized IUR synergy innovation in the AI field has a good foundation.

The primary theoretical contributions are as follows. First, integrating the concept of "organization" into the design of an efficient IUR collaborative system, deepening the theoretical research on IUR synergy innovation. Second, expanding the application of complex network analysis and deep learning algorithms in the field of IUR synergy innovation research. Third, the integration of machine learning and text semantic analysis methods provides a useful supplement for deep mining of the intrinsic relationships between IUR cooperation topics. Fourth, enriching the theoretical and methodological research in areas such as optimizing the allocation of innovative resources and transforming scientific and technological achievements. This study can provide important decision-making support for government departments to guide enterprises, universities, and research institutes to carry out high-level IUR collaborative innovation in an organized manner, and play an important quantitative supporting role in promoting the deep integration of the innovation chain and industrial chain in practice.

**Key words:** Industry-University-Research Synergy Innovation; Industry-University-Research Collaboration Topic; SciBERT Model; Complex Network Analysis; Organized