

情报工程

TECHNOLOGY INTELLIGENCE ENGINEERING



中国科学技术情报学会 | 主办
中国科学技术信息研究所

2024年 第10卷 第6期

二〇二四年 第十卷 第六期

中国科学技术情报学会
中国科学技术信息研究所 主办

情报工程

QINGBAO GONGCHENG

2024年 第10卷 第6期
双月刊



主 管：中国科学技术协会
主 办：中国科学技术情报学会
中国科学技术信息研究所

编辑部

主 编：刘琦岩
执行主编：姚长青
副 主 编：王莉军 刘志辉 桂 婕 董 诚
学术编辑：张均胜
责任编辑：孙 瑶 杨 岩 李孟秋 韩颖霄
美术编辑：翟芒芒

本刊为中国科技核心期刊（中国科技论文统计源期刊），数字化期刊全文数据库（万方）、中国学术期刊全文数据库（CNKI）、中文科技期刊数据库（维普）、《中国人文社会科学期刊评价报告（AMI）》引文数据库、国家哲学社会科学学术期刊数据库、超星学术期刊“域出版”等收录。

国内统一连续出版物号：CN 10-1263/G3
国际标准连续出版物号：ISSN 2095-915X
地 址：北京市海淀区复兴路 15 号（100038）
电 话：（010）58882458
E-MAIL：qbgc@istic.ac.cn
网 址：http://tie.istic.ac.cn/
印刷单位：北京科信印刷有限公司
发行范围：公开发行
编辑出版：科学技术文献出版社有限公司
广告发布登记：京海工商广登字 20170148 号

如发现印刷、装订等质量问题，请与本刊发行部联系调换。

ISSN 2095-915X



定 价：55.00 元

CN 10-1263/G3
ISSN 2095-915X





全国科普日



科普中国
CHINA SCIENCE COMMUNICATION

9月15日 - 9月25日

2024年全国科普日活动

提升全民科学素质 协力建设科技强国

主办单位

中国科协、中央宣传部、中央网信办、教育部、科技部、国家原子能机构、自然资源部、生态环境部、水利部、农业农村部、国家卫生健康委、应急管理部、国务院国资委、中国科学院、中国工程院、国家林草局、全国总工会、共青团中央、全国妇联、中国作协、全国工商联



情报技术 | INFORMATION TECHNOLOGY

003

美国兰德公司遴选颠覆性技术的新型方法及借鉴启示

The Method of Selecting Disruptive Technologies by RAND Corporation and Its Inspiration and Suggestions

崔怡雯 吴毅华 刘巧虹 赵筱媛 苏成

014

出版产业链技术关系抽取研究

Research on Extracting the Relationship between Technology and Publishing Industry Chain

韦向峰 张全 袁毅

028

基于主题模型的政策工具自动标注研究

Automatic Labeling Research of Policy Tools Based on Topic Model

程佳雪 于卫红 齐勇凯 马一鸣

信息分析 | INFORMATION ANALYSIS

039

高校图书馆用户阅读影响因素分析

Analysis of Factors Influencing the Reading Activity of University Library Users

陆桃妹 张芹 刘天琳 景民昌

049

超级数字阅读探索研究

Research on Super Digital Reading

刘利 王东雨

059

大学生网络社会情绪表达策略：基于事件文本挖掘研究

Strategies for College Students' Online Social Emotional Expression: A Study Based on Event Text Mining

刘杰

071

大学生网络信息焦虑的影响机制研究

Research on the Influence Mechanism of College Students' Network Information Anxiety

方增泉 肖可心 元英 祁雪晶 张思康

科技评价 | SCIENCE AND TECHNOLOGY EVALUATION

082

科技管理数据安全治理实践路径研究

Research on the Practice Path of Data Security Governance in Science and Technology Management

徐晨阳 寇亚东 李子伦 王飘 王东

093

污水处理行业产学研专利产出时空演变研究

Study on the Spatial and Temporal Evolution of Industry-University-Research Institute Patent Output in Sewage Treatment

封丽 郭泊池 石亮 李伟

106

钱学森情报思想研究的可视化计量分析

Visualized Bibliometric Analysis of Qian Xuesen's Information Science Thoughts Research

王琳 陈一谕

119

“新动能·新青年·新作为”：2024年科技情报新青年学术会议综述

"New Momentum, New Youth, New Achievements": A Review of the 2024 Academic Conference on New Youth in the Field of Scientific and Technical Information

杨卓莹 赵晖 王彦妍

编委会 • EDITORIAL BOARD

主任委员：

刘琦岩

副主任委员：

乔晓东 潘云涛 孙建军 陆 伟 姚长青 朱礼军

编委委员（按姓氏汉语拼音为序）：

Ume, Anton (罗马尼亚)	安小米	Coh, Byoung-Youl (韩国)	陈仕吉
陈文广	程永红	迟东训	杜永萍
顾进广	胡春明	Jung, Hanmin (韩国)	黄智生
李广建	刘桂锋	李 贺	Meng, Lin (日本)
刘伟丽	刘玉琴	刘忠宝	卢小宾
罗准辰	孟 遥	Rana, F. Omer (英国)	漆桂林
钱 庆	孙 卫	孙 新	Mandl, Thomas (德国)
汪雪锋	吴晨生	吴广印	吴小兰
夏昊翔	杨 波	杨思洛	杨秀丹
俞立平	翟 云	张德政	张家俊
张 军	张 全	张学福	张 巍 (澳大利亚)
张智雄	章成志	真 溱	



开放科学
(资源服务)
标识码
(OSID)

美国兰德公司遴选颠覆性技术的新型方法及借鉴启示

崔怡雯 吴毅华 刘巧虹 赵筱媛 苏成

中国科学技术信息研究所 北京 100038

摘要: [目的/意义] 探索兰德公司为支撑美国政府而研制的新型颠覆性技术评估遴选方法, 以期为我国在目标约束下的颠覆性技术遴选评估工作提供理论与实践参考。[方法/过程] 以兰德公司为美国政府制定的新型遴选评估方法为研究对象, 对其遴选评估颠覆性技术的研究过程进行分析, 总结归纳该新型方法的步骤、流程与创新之处。[结果/结论] 兰德公司对于颠覆性技术遴选的研究相对较为成熟, 分析其相关做法与创新之处, 对于优化我国颠覆性技术遴选评估方案, 完善我国颠覆性技术的评选、资助方式以及实现国家长期科技创新战略布局具有重要意义。

关键词: 兰德公司; 颠覆性技术; 颠覆性技术遴选

中图分类号: G350

The Method of Selecting Disruptive Technologies by RAND Corporation and Its Inspiration and Suggestions

CUI Yiwen WU Yihua LIU Qiaohong ZHAO Xiaoyuan SU Cheng

Institute of Scientific and Technical Information of China, Beijing 100038, China

Abstract: [Objective/Significance] This paper aims to explore the methodology of the new disruptive technology assessment and selection developed by the RAND Corporation to support the US government, with a view to providing theoretical and practical references for China's selection and assessment of disruptive technologies within the constraints of the target. [Methods/Processes] This paper takes the new selection and evaluation methodology developed by RAND for the U.S. government as

基金项目 国家社科基金一般项目“面向颠覆性技术早期识别的弱信号监测体系设计与应用研究”(22BTQ067); 中国科学技术信息研究所创新研究基金青年项目“面向中国式现代化建设的颠覆性技术监测体系研究——基于行动者网络理论”(QN2024-19)。

作者简介 崔怡雯(1991-), 博士, 助理研究员, 主要研究方向为颠覆性技术创新研究、技术创新预测、科技管理; 吴毅华(1994-), 博士, 助理研究员, 主要研究方向为颠覆性技术创新研究、技术创新预测; 刘巧虹(1991-), 博士, 助理研究员, 主要研究方向为颠覆性技术创新研究、技术监测与识别; 赵筱媛(1978-), 博士, 研究员, 主要研究方向为颠覆性创新研究、竞争情报与竞争战略、公共管理与科技政策, E-mail: zhaoxy@istic.ac.cn; 苏成(1973-), 博士, 研究员, 主要研究方向为颠覆性技术、技术预测、科学计量学。

引用格式 崔怡雯, 吴毅华, 刘巧虹, 等. 美国兰德公司遴选颠覆性技术的新型方法及借鉴启示 [J]. 情报工程, 2024, 10(6): 3-13.

the research object, analyses its research process of selecting and evaluating disruptive technologies, and summarizes the steps, processes and innovations of the new methodology. [Results/Conclusions] RAND's research on the selection of disruptive technologies is relatively mature, and the analysis of its relevant practices and innovations is of great significance for optimising China's selection and assessment of disruptive technologies, as well as improving China's disruptive technology selection and funding methods, and realising the layout of the country's long-term scientific and technological innovation strategy.

Keywords: RAND; Disruptive Technologies; Prioritizing Disruptive Technologies

引言

颠覆性技术被视为“改变竞争规则”“重塑未来格局”的革命性力量^[1-2],在新一轮科技革命的新机遇与新挑战下,为抢占科技竞争战略高地与主动权^[3-4],各国政府积极开展谋划布局^[5-6]。在党的十九大报告中,习近平总书记首次提出要“突出颠覆性技术创新”^[7],随后习近平总书记在各种重要场合讲话中不断强调要发展颠覆性创新,特别是在2023年底召开的中央经济工作会议上更是强调要将颠覆性技术创新作为催生未来产业,培育发展新质生产力的关键力量^[8-9]。颠覆性技术的遴选评估是颠覆性技术发展培育资助的关键环节,直接决定了技术创新的前瞻布局方向,决定了产业升级与高质量发展的着力点。因此,加强优先选取具有颠覆影响力技术方法流程的探索,重点资助培育,对充分发挥财政资金使用效率、增强自主创新能力,实现科技跨越式发展至关重要^[10]。

兰德公司是全球著名的高端智库,也是美国最重要的综合性战略研究机构之一。其成立初期便因对全球科技革命、苏联解体、两德统一、朝鲜问题等重大国际问题做出过准确预测而名声大振,多年来开创了诸如系统分析法、德尔非法、博弈论模型等一系列研究方法 with 理论模

型^[11-12],在情报理论方法研究和科学决策支撑中发挥着重要作用,并深度参与影响美国政府决策过程。为进一步巩固加强美国在科技发展中的领导地位,兰德公司与美国政府在综合考虑当前面临的发展环境及竞争格局的基础上,合作开发了用于遴选颠覆性技术的系统评估和比较流程(Evaluation and Comparison Process, ECP)框架^[13],以助力美国空军在预算约束下优先选取能改变竞争规则的技术进行自主培育,实现资源效率最大化。本研究重点分析该新型方法的步骤、流程与创新之处,为我国相关部门开展颠覆性技术遴选评估工作提供借鉴。

1 相关研究进展

目前,美国、俄罗斯、德国、日本以及欧盟等国家地区均设立有专门的颠覆性技术项目研发管理机构^[14],并对颠覆性技术创新建立形成相对较为成熟的评估方法与遴选机制^[15-16]。美国作为在颠覆性技术创新发展方面的表率,其最为成功且著名的当数美国国防部高级计划研究署(DARPA)^[17-18]。由于DARPA资助并创造出众多包括互联网、全球定位系统(GPS)、隐身技术、自动语音识别等在内的广为人知的颠覆性创新,对经济社会发展影响深远,因此

其成功经验与创新组织管理模式为越来越多的国家、政府机构和大型企业所追随模仿^[19]。DARPA 作为美国国防部下属机构,旨在探索国防技术新概念以感知军方未来的潜在需求,因此其资助方向更多侧重于军用领域,如脑机接口、新材料、超声加速等;从 DARPA 的实践经验上来看,项目经理在项目的遴选评估方面享有绝对话语权,主要依赖其通过自上而下的方式发现并进行决策^[3,20]。俄罗斯于 2012 年筹备成立国家先期研究基金会来负责颠覆性技术创新的研发与组织实施,其运行模式部分借鉴 DARPA 模式,拟支持项目主要依靠各领域专家开展遴选评估^[21-22],支持方向涵盖物理技术(包括高速装备、数字化生产、先进水下技术、智能武器)、化学生物与医学、信息技术等。同样,在美国 DARPA 的影响与启示下,日本于 2019 年启动了促进颠覆性技术创新发展的“登月型”研发制度^[23-24],资助方向涵盖虚拟替身、增强现实、疾病预测预防、人工智能、量子计算以及资源循环利用等领域^[25],主要由项目经理负责组织各类构想与挑战的收集,并被赋予充分的决策权,鼓励大胆试错。德国于 2019 年、2020 年分别成立颠覆性创新局 SPRIN-D 和网络安全创新局 Cyberagentur,作为德国识别、促进和发展颠覆性技术创新生态系统的重要组成部分^[26]。与 DARPA 不同,SPRIN-D 将关注重点限于民用技术的创新,申请人项目建议书的审核筛选主要依靠技术专家进行。欧洲创新理事会(EIC)计划由欧盟委员会于 2017 年依据“欧洲地平线”(Horizon Europe)启动,2021 年正式设立实施,旨在识别、开发和推广突破性技术和颠覆性技术创新^[10]。在项目遴选环节,

EIC 针对不同类别的项目分别制定了具有针对性的评估标准与流程准则,评估方式主要基于同行评议与专家打分^[27]。

我国在颠覆性技术遴选评估方面的现有研究主要从国家层面对评估结果进行解读,或重点关注管理模式与组织实施,缺少对方法、工作流程等环节的系统梳理与分析,也尚未形成标准的方法体系与细化方案^[15-16]。邱俊等^[17]通过总结美国新型类 DARPA 成立背景、运行机制等方面内容,对我国科技创新发展提出政策建议。智强等^[28]基于项目经理人的专业化机构管理制度经验提出我国科技计划改革的建议启示。部分学者尝试基于当前理论与实践经验提出技术项目遴选评估的思路框架,如陈之遥等^[5]结合当前政策导向及颠覆性技术的特征属性,从实践角度提出适用于针对颠覆性技术类项目遴选的指标体系与流程设计;张玉磊等^[29]基于国内外颠覆性技术遴选现状,从理论出发提出颠覆性技术创新项目遴选角度和评审标准;马利斌^[30]对颠覆性技术的筛选评估开展研究并指出存在问题。此外,对颠覆性技术识别方法的探索也是近年来我国学者的研究热点与重点方向^[31-32],比如王康等^[33-34]、邢晓昭等^[35-36]、窦永香等^[37-38]、徐硕等^[39]、刘俊婉等^[40]都对颠覆性技术的识别发现进行了系统梳理,虽形成了相对较为全面的颠覆性技术识别发现体系,但上述方法基本仅限于对某一领域特定技术的实证研究,对多领域多学科交叉的颠覆性技术创新识别评估机制研究较少。此外,还有部分学者尝试对原创性项目、非共识性项目等与颠覆性技术项目具有类似属性项目的识别评估方法进行研究^[41-42],如刘志辉等^[43]基于隐性知识,提出汇

聚专家和大众智慧的潜在颠覆性技术评估系统设计方案；鲍锦涛等^[44]研究总结国外主要科学基金资助机构关于原创性基础研究项目的资助情况，并对其申请与评审管理过程提出具体意见建议；董高峰^[45]以美国国家科学基金会和国立卫生研究院为研究对象，分析创新性研究的非共识性与同行评议共识性之间的矛盾关系，并总结该类创新性研究资助政策的主要特点与模式；刘钊等^[46]研究非共识项目的遴选资助问题，提出非共识项目资助管理的风险管理机理与措施；殷嘉珩等^[47]提出传统同行评议模式在对创新性强、不确定性大的原创科学进行评议时会出现偏差，需要探索完善适用于原创性科学的评审方式与体系。

整体来看，探索科学的评估方法与决策机制是布局发展颠覆性技术的关键和前提，有助于降低该类科技创新的投入风险，助力政府部门、科研机构、企业遴选重点潜在颠覆性技术创新进行精准培育。但现有的研究中，对于如何衡量该类技术的颠覆性效果和影响尚未有统一的量化标准，缺乏系统性的遴选评估体系与流程设计，且可操作性不强。兰德公司作为全球顶尖智库之一，在颠覆性技术、前沿技术以及关键核心技术的研究与决策支撑方面颇负盛名^[12]，且擅长将问题转变形成体系化、流程化研究方法。因此，本文重点分析兰德公司受军方委托所制定的“颠覆性技术系统评估和比较流程框架”，总结其在预算约束下制定的颠覆性技术方向优先遴选方法与流程，以期为我国建立科学、常态化的颠覆性技术创新评估遴选机制，实现颠覆性技术创新发展精准决策提供参考。

2 兰德公司“颠覆性技术系统评估和比较流程框架”(ECP)概况

颠覆性技术的评估和遴选是一项具有战略意义的、动态的、复杂系统工程。兰德公司在该新型颠覆性技术系统评估和比较流程框架（见图1）的设计思路中，一方面坚持目标需求导向，通过在同一问题框架下对不同技术的比较，避免了上述提到的针对颠覆性技术影响缺少统一量化标准的问题；另一方面从实际问题出发，制定了流程化、体系化的评估标准与遴选方案，很大程度上解决了当前颠覆性技术评估过程中过于依赖专家智慧/项目经理决策、可操作性不强等问题。

总结来看，该方法与流程具有以下特点：一是强调高层决策领导与领域技术专家的共同参与，既能准确把握目标需求，又可保证遴选评估结果的科学性与准确性；二是以任务需求为导向，通过将待评估技术与目标场景相结合来完成不同技术的多维度对比分析，消除传统从技术角度来计算多方影响力的困难与不确定性，更具有可比较性与透明度；三是围绕目标需求制定不同层面的指标评价标准，将定性问题转化为可量化、可比较的分析指标，为技术评估遴选提供可落地、操作性强的指导性方案。

对此，针对颠覆性技术的评估遴选问题，兰德公司构建了“候选技术确定—技术能力评定与可行性分析—最终技术确定与结果展示”的流程框架，主要分为6个步骤：一是颠覆性技术潜力的评估，即确定评估和比较哪些潜在的颠覆性技术；二是颠覆性技术应用场景的详尽阐述；三是颠覆性技术应对持久挑战的能力

评估，即对颠覆性技术应用的重要性、优先级进行评价；四是颠覆性技术在现实任务中的优势验证；五是颠覆性技术优势置信度评估，也

就是从技术与实施两方面对技术可行性展开评估；六是潜在颠覆性技术对比结果展示，为决策提供支持支撑。

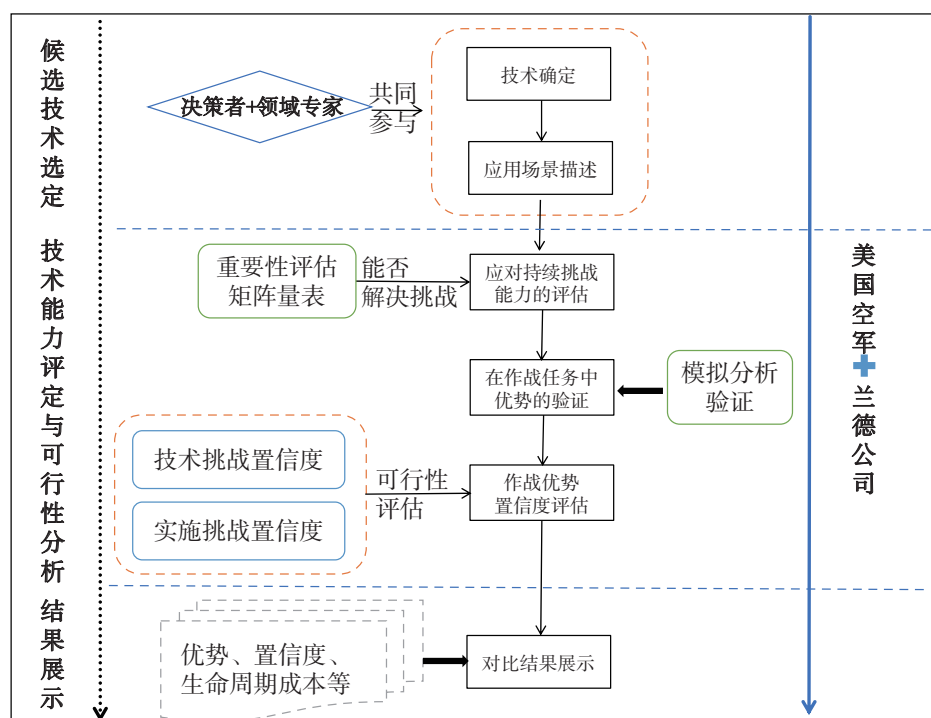


图1 颠覆性技术评估遴选框架流程

3 ECP 框架分析过程与结果

3.1 阶段1:明确目标需求,选定候选技术

颠覆性技术的选取确定要符合国家政府/决策机构的战略部署，这对于有效把握目标需求具有重要意义。对于颠覆性技术而言，因为其往往具有超前性、非共识性和高度不确定性等特点，因此技术风险要远大于常规渐进型技术，失败率高。为解决该问题、进一步降低投资风险，该方法在候选技术确定环节就特别强调高层决策者与领域专家的共同参与，以尽可能借助决策者前瞻部署能力准确把握目标需求，并充分发挥专家学者的领域知识来保证结果的

科学性与专业性。为明确目标需求、选定候选技术，此环节共分为两个步骤：

首先，由高层决策者与领域专家共同参与，并提出在未来新情境下可能产生的具有颠覆性影响的新技术或新概念。而该项工作的基础则是需要参与者对当前科技文献资料、决策机构内部计划和正在进行的研发工作尽可能全面详尽地梳理，进而辅助新概念/新技术的提出。

其次，由决策者和技术专家从技术能力、实施概念以及可以解决的重要挑战等角度对上一步骤中提出的潜在颠覆性技术的应用场景展开详尽描述，为下一环节潜在颠覆性技术的遴选评估研究奠定基础。具体来讲，每一项待评估的候选颠覆性技术或因其固有能力，或是随

着该技术的实施能够为当前带来跨越式、颠覆式的发展,或是前所未有的变革,因此要求参与者能对其应用场景、潜在价值等有一定的把握与理解。

3.2 阶段2:技术能力评定与可行性分析

颠覆性技术本身具有高度的不确定性、高风险性和非共识性等特点,在利用传统基于“头脑风暴”的专家评议方式对该类技术评判时,因主要依赖专家个人的智慧意见,主观性强且没有统一的评判标准,因此容易扼杀“小众思想”的创造力,出现“屈服于权威或大多数”的局面。针对该问题,该方法提出了“技术能力评定—优势验证—可行性评估”的评估遴选流程,通过构建创新型评价指标,以定性定量相结合、多轮迭代、相互印证的方式共同提升结论的可信度与可靠性,科学有效地支撑决策。该环节可分为三个步骤:

3.2.1 目标/问题需求导向,从应对持久挑战能力的角度对待评估技术的优先级展开研判

目标/问题需求导向,指的是从决策者目标需求角度来确定待评估技术解决方案的实际成效。考虑因素主要有两个方面:一是该技术是否能够解决当前面临的问题与挑战;二是该技术对解决目标需求贡献程度的大小。具体包含三个流程。

首先,要明确应对挑战,并坚持问题导向。高层领导者在决策过程中面临的难点是该步骤首要考虑的问题。由决策者提出目标需求以及可能需要应对的持久挑战,其目的在于以问题导向,更好地集聚力量来牵引、推动技术发展,并确定技术的优先级。也就是说,针对不同的

任务问题需求,其相关技术的重要性与优先级也不尽相同,我们需要针对具体任务情境来评判确定相关技术的重要程度。

其次,提供统一的遴选评估标准。虽然面对不同的决策需求可能会导致待评估技术的内容与边界不同,但仍需以统一的方式和评估标准开展遴选,以保证该方法的可复用性与可移植性。对此,兰德公司首创开发了重要性评估矩阵量表,使用者可依据目标需求与相关技术之间的匹配程度对其重要性 with 优先等级进行审查研判。具体来看,每项技术的重要性可分为低、中、高三个等级。其中,1~3分代表该技术对目标需求的影响很小;4~6分代表该技术能带来一定的贡献;7~9分则代表该技术具有明显的优势,能够为目标需求带来关键性的改进。分数越高代表着该技术与目标需求之间的匹配度越高,越能解决当前面临的问题挑战,也越为重要。具体评判标准与对照关系见表1。

表1 重要性评估矩阵量表

等级	重要性评分	技术能力与理念 (与以往相比)	对目标需求的影响程度
高	9	完全不同	能够引起颠覆性变化
	8	提供一些新东西	具有显著优势
	7	重要改进	具有关键贡献
中	6	重要改进	具有重要贡献
	5	重要改进	具有有益贡献
	4	轻微改进	贡献极小
低	3	轻微改进	可通过其他方式获得
	2	轻微改进	渐进式改进
	1	无改进	不显著

最后,依据技术成熟度与技术重要性评估矩阵量表来共同确定技术重要性 with 优先级。具

体来讲，在资源约束下，一项技术从被投资、到开展研发活动、再到快速走向实际应用以解决问题，需考虑以下两方面因素。一是技术需要具有足够的重要程度与被认可的潜在价值，即需要参与者依据上述评估量表的评判标准对候选技术的能力、理念以及对目标结果的影响程度进行研判。为避免意见不统一可能会带来的不确定性，具体操作时可基于评分等级的分布作出统计估计。二是技术的成熟度属性也需达到一定程度（根据美国国防部颁布的技术成熟度评价手册，技术成熟度共分为9个等级）。也就是说，如若一项技术仍处于基本原理、技术概念的研究验证阶段，其应用场景、潜在价值尚不明确、无法评判时，则无法通过评估以进行投资。

3.2.2 模拟仿真结果分析,在现实任务中开展优势验证

模拟仿真是一种基于实验或训练目的的技术，其核心在于将系统或流程模型化，以此表征其关键特性或行为/功能，常被用于评价各种应对措施或替代方法对解决实际问题的有效性。因此，为进一步验证候选技术解决目标挑战的优势能力，该步骤要求可以通过使用多种模拟仿真手段来测试候选技术在解决现实挑战时的优势能力及有效性，并根据结果对上一步骤中的技术重要性评分和优先等级进行调整。以空军基地防御这一挑战为例，可通过仿真技术来模拟多轴巡航导弹对基地的攻击，进而利用时间线分析法确定不同候选技术应对该攻击过程的实际作用与有效性。

3.2.3 技术可行性评估,从技术发展与实施风险两方面确保成功

一项技术能否实现其重要价值往往受多方

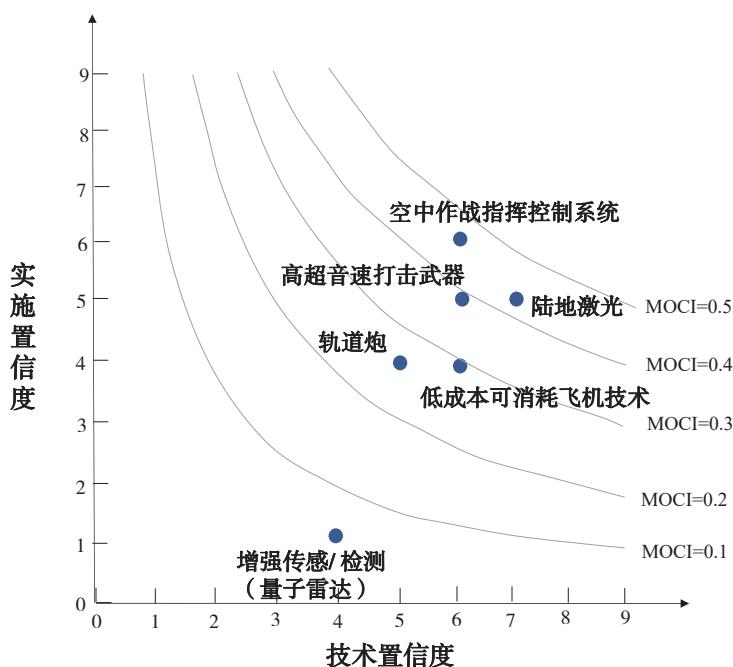
因素的影响。因此，遴选潜在颠覆性技术时，不能只是简单地采用定性或定量排序的方法，直接选取得分最高、最有共识的技术，否则可能会忽视技术对创新生态环境的依赖性，忽略技术在实施过程中可能面临的失败问题。而实施风险又包括如组织管理、基础设施、资源配置、政策制度以及人才合作等在内的方方面面。对此，兰德公司认为，必须将技术风险与实施风险一起纳入评估流程框架中，共同评判一项技术是否能解决目标需求并取得领先优势。

基于上述思想，该方法从技术挑战和实施挑战两方面制定了技术可行性评估标准（见表2），用以刻画衡量某项技术达到优势状态时所需克服困难与障碍的难易程度。上述两个量表均基于九分制评分系统，根据技术所处阶段以及对应问题是否存在解决方案，可将其评为1~9分，划分为低、中、高三个等级。具体来讲，如果某项技术相对较为成熟、所遇到的技术问题可以很好地解决，则技术置信度越高；如果某项技术在实施过程中面临的困难与障碍越少或者面临的挑战均能够得以快速解决，则实施置信度越高。

最后，基于上述两项评分结果，取其乘积做归一化处理后作为最终的技术优势置信度（Measure of Confidence in Implementation, MOCI）结果。MOCI数值范围在0~1之间，得分越高，表明待评估技术可行性越大。为更好理解、操作该步骤，下图以空军基地防御为例，对相关颠覆性技术的评估结果在同一坐标中进行展示，可行性大小则以等高线形式进行比对分析（见图2）。其中，技术点所在位置越靠近右上方说明技术和实施置信度越大，可行性越高。

表 2 技术可行性评估标准量表

等级评估	评分	技术挑战（技术开发状态）	实施挑战（实施风险状态）
高	9	所有技术问题均已解决	所有实施挑战均已解决
	8	开发中，所有技术问题均有望解决	大多数实施挑战已得到解决；剩余几个实施挑战也有望解决
	7	开发中，大多数技术问题已得到解决	解决所有实施挑战的途径是明确的，但仍需要努力
中	6	开发中，已经解决部分技术问题，但仍然存在实质性的技术问题尚待解决	正在努力解决所有挑战，并且没有挑战处于早期探索阶段
	5	开发中，并已发现正在解决的技术问题	正在努力解决所有挑战，但仍有少数挑战处于早期探索阶段
	4	开发中，但尚有未完全理解的技术问题	正在努力解决所有挑战，但许多挑战仍处于早期探索阶段
低	3	建模与模拟阶段	所有挑战均有明确的解决方案
	2	应用规划阶段	有较少尚未有明确解决方案的挑战
	1	初始概念阶段	有很多尚未有明确解决方案的挑战

图 2 以空军基地防御为例，相关颠覆性技术的置信度展示^[13]

3.3 阶段3:最终颠覆性技术确定与结果展示

遴选确定初步的潜在颠覆性技术之后，需要综合考虑当前的技术研发能力、投入成本以及对应的风险，以确定最终技术清单的优先级排序。为此，该框架提出以目标任务为导向的对比分析方法，即针对某一特定任务目标，在同一坐标轴下，对候选技术的技术优势、技术

优势置信度以及技术生命周期成本进行可视化展示和对比分析，为决策者优化备选方案的收益、风险和成本提供一致性评比标准。其中，生命周期成本主要指技术研发、维护过程中所需支付的相关成本。

图 3 以空军基地防御为例，对相关颠覆性技术的评估结果进行可视化展示与分析。图中

的作战优势和 MOCI 分别由上一环节的步骤 2、3 得出，技术的生命周期成本范围则可通过新闻文章、项目预算以及先前研究中所积累的相关信息等估计确定。其中，点越大表示培育布

局该技术的成本越高，点的相对位置越靠近右上方则说明技术的优势及置信度越高，决策者可通过图中点的位置以及大小对相关目标约束条件下的技术优先级做出快速研判。

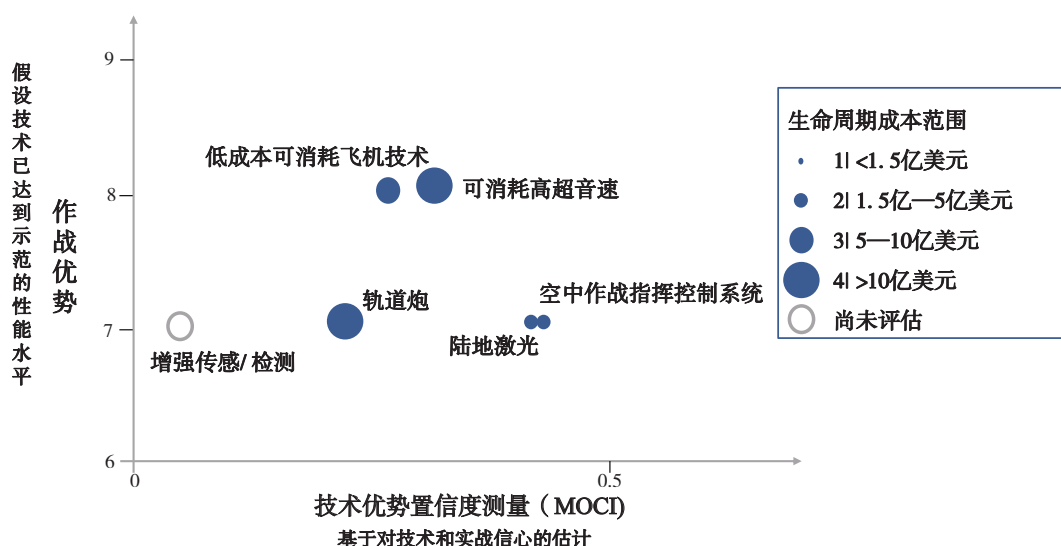


图3 以空军基地防御为例，相关颠覆性技术评比结果演示图^[13]

4 启示

理论层面看，该方法开创性地构建了 ECP 框架，为预算约束下早期技术遴选提供了动态决策方法。该评估框架聚焦目标任务，决策者、领域专家等多主体共同参与，通过建立优势评估以及技术优势置信度测量等指标来共同辅助完成对候选技术的评估比较与多轮迭代，提高技术遴选精准度。在此基础上，开发了可视化演示案例系统，全面化、直观化分析问题以服务决策需求，践行其职责使命。该框架综合了技术风险评估和实施风险评估，能更全面地洞察发展某项颠覆性技术可能会带来的机遇以及潜在风险，在当今科技创新引领驱动社会快速发展的时代背景下，这对于颠覆性技术管理机构来讲尤为重要。

同时，该框架比传统方法具有更明确的目标与步骤、更完善的反馈与调整机制，有助于更精准、高效地完成颠覆性技术遴选评估工作。传统基于专家领域知识或基于多维指标体系的颠覆性技术遴选评估缺少明确的需求目标，在发挥理论指导作用时会存在一定困难。如若仅依靠专家决策，可能会产生“一言堂”局面；非共识评价、负责人的自由度量、交叉评审、扶优评审、人本评审等非常规评价体系^[30]虽然能从一定程度上避免共识性评价带来的弊端，但该方法对评审专家的科研能力、学术水平以及战略眼光等要求都较高，且在可操作性与风险把控上存在难度；多维指标定量分析则缺少统一的标准去衡量评比具有社会、经济、生态等不同层面价值意义的技术^[5, 29, 48]。

从实践应用层面看，与 DARPA、EIC 等长

期从事颠覆性技术识别、培育、资助的研发机构做法相比,该方法聚焦目标任务,决策者全流程参与,更适用于国家主导下目标明确的技术遴选评估,且具有更强的适用性、可操作性和可复用性。如 DARPA 经过多年的实践已建立起基于项目经理主导的遴选评估机制,但该机制要求领导者/决策者或项目经理具有很强的领导能力和成就,对于新技术的发展也要具有敏锐的眼光与嗅觉,因此最终技术方向的选定与项目经理的把控息息相关且无法固化该操作流程。EIC 计划主要围绕欧盟战略规划展开,以加速相关成果与发明的技术转化,其项目的遴选评估主要由专员基于设定的各类遴选指标(如卓越程度、影响力、实施质量和效率以及对相关支持需求等)来实现,更适用于对企业类项目的评估投资。

未来,我国也可充分学习借鉴兰德公司在目标约束下开展颠覆性技术遴选评估的研究思路与方法,为国家目标导向下进行颠覆性技术方向选择、遴选评估形成有益参考。

参考文献

- [1] DAN Y, CHANG C H. Creating technology candidates for disruptive innovation: Generally applicable R&D strategies[J]. *Technovation*, 2011, 31(8): 401-410.
- [2] 克莱顿·克里斯坦森. 创新者的窘境[M]. 北京: 中信出版社, 2014: 8.
- [3] 曹晓阳, 魏永静, 李莉, 等. DARPA 的颠覆性技术创新及其启示[J]. *中国工程科学*, 2018, 20(6): 122-128.
- [4] 苏帆. 国外颠覆性技术创新资助管理经验与启示[J]. *科技中国*, 2022(11): 59-61.
- [5] 陈之瑶, 罗军, 莎薇, 等. 基于颠覆性技术创新项目评价及立项决策模型的研究[J]. *科技管理研究*, 2024, 44(1): 152-159.
- [6] 孙永福, 王礼恒, 孙棕檀, 等. 引发产业变革的颠覆性技术内涵与遴选研究[J]. *中国工程科学*, 2017, 19(5): 9-16.
- [7] 习近平. 决胜全面建成小康社会夺取新时代中国特色社会主义伟大胜利——在中国共产党第十九次全国代表大会上的报告[J]. *党建*, 2017(11): 15-34.
- [8] 习近平. 高举中国特色社会主义伟大旗帜为全面建设社会主义现代化国家而团结奋斗——在中国共产党第二十次全国代表大会上的报告[J]. *青海党的生活*, 2022(11): 4-23.
- [9] 中央经济工作会议在北京举行 习近平发表重要讲话 李强作总结讲话 赵乐际王沪宁蔡奇丁薛祥李希出席会议[J]. *旗帜*, 2023(12): 6-8.
- [10] 崔怡雯, 刘巧虹, 吴毅华, 等. 基于多国实践研究国家目标导向的颠覆性技术类科技项目选题机制[J]. *中国科技资源导刊*, 2024, 56(2): 63-72, 83.
- [11] 王伟, 张艳霞, 王锐华, 等. 兰德战略评估经验做法及启示[J]. *情报理论与实践*, 2023, 46(4): 204-209.
- [12] 叶尔达纳·肯吉别克, 莎塔娜提·热合木别克, 任真. 兰德公司遴选国家关键技术的方法及其启示[J]. *智库理论与实践*, 2023, 8(5): 149-157.
- [13] SILBERGLITT R, COOK C R, POPPER S, et al. Systematic Method for Prioritizing Investments in Game-Changing Technologies: The Evaluation and Comparison Process Framework[R/OL]. (2022-07-18)[2023-03-27]. https://www.rand.org/pubs/research_reports/RRA632-1.html.
- [14] 黄雁宁, 赵筱媛, 赵志远, 等. 颠覆性技术项目管理国际经验及启示: 基于多案例分析视角[J]. *全球科技经济瞭望*, 2023, 38(11): 19-28, 37.
- [15] 王开阳. 美国主要颠覆性技术研发机构的管理机制比较研究——以“类 DARPA”机构为例[J]. *全球科技经济瞭望*, 2022, 37(12): 35-40.
- [16] 赵志远, 赵筱媛, 苏成, 等. 颠覆性技术研发与管理资助体系: 分析模型与国外实践[J]. *科技管理研究*, 2022, 42(17): 111-117.
- [17] 邱俊, 梁正, 顾心怡, 等. 美国新型类 DARPA 项目管理创新机构的若干进展及启示[J]. *中国科学院院刊*, 2023, 38(6): 907-916.

- [18] 李曼迪, 赵筱媛, 苏成. 项目生命周期视角下美国 DARPA 先进技术开发项目管理实践与经验 [J]. 全球科技经济瞭望, 2024, 39(1): 1-8, 24.
- [19] 曹方, 何颖, 张鹏, 等. 国内外发展颠覆性技术的实践及其典型做法 [J]. 科技中国, 2021(10): 9-11.
- [20] DUGAN R E, GABRIEL K J. 'Special forces' innovation: how DARPA attacks problems [J]. Harvard Business Review, 2013, 91(10): 74-84.
- [21] 刘澈, 唐波. 超越 DARPA: 俄罗斯先期基金会创新突破及其启示 [J]. 军民两用技术与产品, 2021(3): 37-44.
- [22] 孟光. 俄先期研究基金会发展情况研究 [J]. 现代军事, 2015(8): 91-94.
- [23] 朱相丽, 王溯, 董瑜. 浅析日本“登月型”研发制度 [J]. 世界科技研究与发展, 2021, 43(1): 111-118.
- [24] 孙婕, 沈恒超. 美日宏观科技统筹协调机制及启示 [J]. 世界科技研究与发展, 2022, 44(4): 504-517.
- [25] MOONSHOT. Moonshot R&D TOP[EB/OL]. (2022-12-06)[2024-05-06]. <https://www.jst.go.jp/moonshot/en/program/millennia/about.html>.
- [26] 李万. 无尽的前沿: 诱人的未知: 新科技革命正引致一系列深刻变革 [J]. 中国科技论坛, 2021(9): 3.
- [27] European Innovation Council. EIC 2023 work programme[EB/OL]. (2021-09-17)[2024-05-06]. https://eic.ec.europa.eu/eic-2023-work-programme_en.
- [28] 智强, 林梦柔. 美国国防部 DARPA 创新项目管理方式研究 [J]. 科学学与科学技术管理, 2015, 36(10): 12-22.
- [29] 张玉磊, 戴海闻, 许泽浩, 等. 颠覆性技术遴选的基本原则与运行流程研究 [J]. 科技管理研究, 2020, 40(13): 209-216.
- [30] 马利彬. 颠覆性技术筛选评估重点问题研究 [D]. 北京: 军事科学院, 2018.
- [31] 刘雨农, 石静, 梁琴琴. 基于 STM 的颠覆性技术主题识别研究 [J]. 情报工程, 2023, 9(3): 81-91.
- [32] 石慧, 潘云涛, 苏成. 颠覆性技术及其识别预测方法研究综述 [J]. 情报工程, 2019, 5(3): 33-48.
- [33] 王康, 陈悦. 基于异质性专利的颠覆性技术早期识别研究 [J]. 科学学研究, 2023, 41(8): 1364-1375.
- [34] 王康, 陈悦, 王玉奇, 等. 基于专利引用变化的颠覆性技术识别研究 [J]. 情报杂志, 2022, 41(1): 74-80, 169.
- [35] 邢晓昭, 袁芳, 牟琳, 等. 基于动态网络位置的颠覆性技术识别方法 [J]. 高技术通讯, 2023, 33(2): 198-207.
- [36] 邢晓昭, 任亮, 雷孝平, 等. 基于专利主题演化的颠覆性技术识别研究——以类脑智能领域为例 [J]. 情报科学, 2023, 41(3): 81-88.
- [37] 窦永香, 开庆, 王佳敏. 一种基于图表示学习的潜在颠覆性技术识别方法 [J]. 情报学报, 2023, 42(6): 637-648.
- [38] 开庆, 窦永香. 颠覆性技术识别研究综述 [J]. 情报杂志, 2021, 40(11): 31-38.
- [39] 徐硕, 李静鸿, 安欣. 基于专利术语的颠覆性技术识别及实证研究 [J]. 图书情报工作, 2024, 68(2): 62-72.
- [40] 刘俊婉, 庞博, 徐硕. 基于弱信号的颠覆性技术早期识别研究 [J]. 情报学报, 2023, 42(12): 1395-1411.
- [41] 李晓萌, 赵英弘, 郝红全, 等. 基础研究原始创新的政策探索: 从非共识创新到原创探索计划 [J]. 中国科学基金, 2024, 38(2): 320-327.
- [42] 赵英弘, 郝红全, 李晓萌, 等. 国家自然科学基金资助原创性基础研究的启示 [J]. 中国软科学, 2023(10): 9-20.
- [43] 刘志辉, 张均胜, 林毅, 等. 基于隐性知识的潜在颠覆性技术评估方法研究 [J]. 情报学报, 2021, 40(12): 1271-1278.
- [44] 鲍锦涛, 郑毅, 彭一杰, 等. 原创性基础研究的内涵分析及对原创探索计划项目的启示 [J]. 中国科学院院刊, 2022, 37(3): 384-394.
- [45] 董高峰. 美国国家科学基金创新性项目资助政策研究 [D]. 北京: 中国科学院大学, 2008.
- [46] 刘钊, 李小敏. 基础研究中非共识项目的资助问题研究 [J]. 自然辩证法通讯, 2022, 44(3): 70-81.
- [47] 殷嘉珺, 罗惠文, 庄建辉. 原创性科学研究同行评议的国际经验及对原创探索计划项目的启示 [J]. 中国科学基金, 2021, 35(4): 567-572.
- [48] 冯倩倩, 张光宇, 戴海闻, 等. 颠覆性技术遴选的指标体系与流程设计——基于扎根理论的多案例研究 [J]. 科技管理研究, 2021, 41(24): 50-59.



开放科学
(资源服务)
标识码
(OSID)

出版产业链技术关系抽取研究

韦向峰^{1,2} 张全¹ 袁毅¹

1. 中国科学院声学研究所 北京 100190;
2. 富媒体数字出版内容组织与知识服务重点实验室 北京 100038

摘要: [目的/意义] 出版产业链中技术与产业链环节的关系对于出版产业技术谱系的构建和出版产业的监测具有重要意义。[方法/过程] 设计了传统出版和数字出版的产业链环节,并从业务环节、产业术语、技术术语、参与主体、产品服务维度进行了产业技术谱系设计。在获取出版产业技术谱系实体后,利用句法依存分析工具获取实体之间的关系模板,使用 Mean Teacher 深度学习训练框架和 BiGRU+Attention 神经网络编码器实现了基于关系模板质量的关系抽取模型;然后使用部分人工标注的半监督深度学习方法对关系模板进行了分类标注和关系分类的模型训练。[局限] 未来仍需研究如何提高关系模板中关系类型的识别准确率,通过改进深度学习模型框架来提高模型的性能。[结果/结论] 实验表明该关系抽取模型在实际语料库文本中可获得 66% 的准确率,消融实验表明模板质量等级划分能带来 1% 的准确率提升。

关键词: 出版产业; 产业链; 关系抽取; 关系模板; 半监督深度学习

中图分类号: G35; G23; TP391

Research on Extracting the Relationship between Technology and Publishing Industry Chain

WEI Xiangfeng^{1,2} ZHANG Quan¹ YUAN Yi¹

1. Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China;
2. The Key Laboratory of Rich-Media Knowledge Organization and Service of Digital Publishing Content, Institute of Scientific and Technical Information of China, Beijing 100038, China

Abstract: [Objective/Significance] The relationship between technology and the nodes of the publishing industry chain is of

基金项目 2023 年富媒体数字出版内容组织与知识服务重点实验室开放基金“基于预训练模型的产业技术谱系构建研究”(ZD2023-11/03)。

作者简介 韦向峰(1976-), 博士, 副研究员, 主要研究方向为人工智能、语音信号与信息处理、大数据与知识组织, E-mail: wxf@mail.ioa.ac.cn; 张全(1968-), 博士, 研究员, 主要研究方向为人工智能、语义分析与处理; 袁毅(1967-), 学士, 高级工程师, 主要研究方向为计算机软件、计算机应用。

引用格式 韦向峰, 张全, 袁毅. 出版产业链技术关系抽取研究[J]. 情报工程, 2024, 10(6): 14-27.

significant importance for constructing the technological spectrum of the publishing industry and monitoring its development. [Methods/Processes] This article designs the industrial chain and the technological spectrum for both traditional publishing and digital publishing. The design of the industrial technological spectrum includes six dimensions, industrial segments, industrial terms, technical terms, participating entities, and product services. After obtaining the entities of the technological spectrum in publishing industry, relationship templates between entities are acquired using syntactic dependency analysis tools. Then, a relationship extraction model based on the quality of relationship templates is implemented using the Mean Teacher deep learning framework and BiGRU+Attention neural network encoder. Furthermore, a semi-supervised deep learning method with partially manually annotated data is employed for relationship classification model training based on relationship template classification. [Limitations] The future research work is still needed on how to improve the accuracy of identifying relationship types in relationship templates and enhance the performance of models by improving deep learning model frameworks. [Results/Conclusions] Experimental results indicate that this model achieves 66% accuracy in actual corpus texts, and categorizing templates can lead to a 1% increase in accuracy.

Keywords: Publishing Industry; Industry Chain; Relationship Extraction; Relationship Template; Semi-Supervised Deep Learning

引言

随着技术的不断发展和进步，出版产业也在向数字化、智能化、创新化发展，并引进和应用了许多新技术，例如人工智能、大数据、云计算、增强现实和虚拟现实等。从出版产业链角度看，传统出版产业的“编—印—发”三个环节^[1]（即“编辑、印刷、发行”），可分别对应出版产业链的“上游、中游、下游”，产业链涉及的技术则包括信息采集、印刷工艺、物流技术等；数字出版的产业链可分为“内容提供—数字化加工与出版—内容销售与终端”三个环节，涉及的技术更新更广，例如智能写稿、AI创作、OCR、XML、语音合成、计算机信息检索、中文信息处理技术、数字加密技术、电子支付、虚拟现实、增强现实等。为了对出版产业上游、中游、下游等链条环节上的技术术语进行组织并构建相应的知识图谱，得到技术在出版产业链中的环节位置、技术关联

的专家人才、技术关联的企业信息等知识，需要对出版产业链涉及的业务环节、技术术语、参与主体、产品服务等进行术语抽取和关系抽取，对出版产业涉及的相关技术、产业链环节等构建得到相应的技术谱系，从而在产业技术谱系的基础上实现技术推荐、产品推荐、企业推荐等应用。

本文以《编辑与出版学名词》^[1]为基础，从文本数据中提取了产业链环节、产业术语、技术术语、参与主体、产品服务等行业技术谱系中的术语或实体，并着重研究如何从实际文本数据中获取和完善产业技术谱系中实体之间的关系，具体来说就是要确定技术术语与产业术语之间的关系类型，以及技术术语在产业链中的产业链环节（上游、中游或下游）。这里涉及文本数据的关系抽取，也称为实体关系抽取，是指从给定文本提取出实体对中实体之间语义关系^[2]。关系抽取的方法按照是否需要事先抽取出实体，可以分为整体式抽取和级联式

抽取^[3]。整体式抽取采用统一模型,从非结构化文本中提取得到“实体—关系—实体”三元组。级联式抽取需要先识别实体或已知文本中的实体对,在此基础上从文本中提取得到实体之间的关系。按照是否需要标注语料,关系抽取的方法可分为有监督方法、半监督方法、远程监督方法和迁移学习方法^[4-5]。有监督方法需要大量高质量关系标注语料,远程监督和迁移学习方法需要目标领域知识库或者关联领域标注语料,一般用于相似领域的关系抽取,很难应用于特定领域内部的关系抽取。

在本文的研究中,出版产业技术谱系的实体已使用自然语言处理方法提前获取,且获得的出版产业文本语料库没有大量标注好的高质量语料,因此在已有关系抽取研究基础上采用级联式的半监督关系抽取方法^[6-8]。本文将重点探讨如何使用半监督的深度学习模型方法,自动或半自动地进行出版产业的技术术语与产业链环节的关系抽取,从而提高出版产业技术谱系的构建效率,为出版产业的技术链监测、产业链监测、创新链监测等提供相关的技术支撑。

1 相关研究

早期的关系抽取采用规则方法来抽取文本中的实体关系,Chinatsu等^[9]利用领域知识专家编写的规则,包括特征词、词性、句法信息和语义模式等,在文本语料库中抽取与规则匹配的关系实例;Fukumoto等^[10]提出了OKI信息抽取系统,其中的关系抽取利用实体之间的谓词来判定实体之间的关系。基于规则的方法主要依赖于专家的领域知识,主观性较大、

成本较高,且无法移植到其他领域,对于跨领域的关系抽取构成难以克服的困难,带来高昂的成本。在统计机器学习时期,研究者采用基于特征向量的抽取方法或者基于核函数的方法。特征向量的特征一般包括词语特征、句法特征和语义特征,核函数是隐式地计算特征向量的内积,核函数主要有序列核函数、卷积核函数和树核函数,然后应用于最大熵模型、支持向量机(SVM)、朴素贝叶斯和条件随机场等机器学习模型。2003年Zelenko等^[11]首次将核函数应用到关系抽取任务上,使用动态规划的算法,结合两个样本的浅层句法解析树来分析两者间的相似性,采用SVM统计机器学习模型在新闻语料文本数据上取得了不错的效果。在中文关系抽取中,车万翔等^[12]设计了实体类别、实体位置、前后词信息等特征,采用Winnow和SVM两种统计机器学习算法,对中文关系抽取标注语料库进行了训练和关系识别,实验表明SVM相对于Winnow算法可以提高召回率和F1值。统计机器学习模型方法在关系抽取中虽然取得了较好的效果,但是该方法依赖于大量标注好的语料库,需要进行大量的预处理工作,耗费的人力成本也很高,且无法自动进行关系类型的扩展。

随着深度学习技术的发展和进步,许多研究者把人工神经网络模型应用到文本数据的关系抽取任务中。2014年Zeng等^[13]利用深度的卷积神经网络(CNN)进行特征提取,使用词向量和卷积提取文本中词语和语句的特征,然后通过分类器可以预测两个标记实体之间的关系。Socher等^[14]最先将循环神经网络(RNN)模型用于关系抽取中,文本中每个单词均由向

量和矩阵表示单词本身语义和对其他单词的修饰作用，这种表示可以自动学习到较长短语的深层语义，但模型需要学习的参数过多。长短期记忆网络(LSTM)、门控循环单元网络(GRU)和图卷积神经网络(GCN)等人工神经网络模型也纷纷被引入到文本的关系抽取中，促进了关系抽取的研究，成为目前的研究热点之一。

Transformer模型提出了注意力机制(Attention)，融合了语义的分布式表示和长距离语义关系发现，从而克服了RNN计算成本高、没有长距离文本语义处理的缺陷。2018年BERT预训练语言模型在Transformer的基础上应用掩码语言模型、预训练和微调技术，在自然语言处理相关任务中取得了很大进步。预训练语言模型BERT自然也被引入到实体抽取和关系抽取的任务中，例如马江微等^[15]使用BERT模型作为输入文本的编码器，采用分层强化学习方法分别进行关系与其对应实体的解码，在公开的NYT10、NYT10-sub数据集上F1值分别达到71.8%和69.0%。近年来，以GPT(生成式预训练Transformer)为基石的大语言模型获得了空前发展，其特点是基于海量数据进行了模型的预训练，结合指令学习、人类反馈强化学习、思维链和模型微调等技术，可以根据输入的自然语言指令文本生成指令相关的结果文本，也可以作为人工神经网络模型编码器实现文本关系抽取任务。但是，人工神经网络模型和预训练语言模型用于关系抽取都需要大规模标注的带关系标签的语料库。大语言模型虽然不需要标注语料库，但是其效果依赖于预先训练的文本数据。对于没有标注过的行业领域文本数据的关系抽取，以及大语言模型未进行微调或特

殊预处理的专业领域文本中的信息抽取，无法直接应用人工神经网络模型、预训练语言模型或大语言模型来进行文本中关系的抽取。因此，一些研究者转向研究如何利用半监督的方法，只需要少量的人工参与和标注语料，实现自动地从文本中提取出实体之间的关系。

半监督关系抽取方法只利用少量标注数据，结合相关算法学习训练大量未标记的文本语料，从而抽取得到文本语料中的关系。这不仅可以有效减少人工参与以及对大规模标注语料的依赖，而且关系抽取的性能也能得到提升，还可以扩展应用到其他领域的大规模文本语料的关系抽取中。半监督关系抽取方法的难点主要是初始种子关系的质量问题和如何降低迭代过程中的噪声问题，且少量的人工标注数据容易产生语义漂移。解决问题的办法之一是提高种子关系的质量，或者通过人工总结全面的种子关系。Mean Teacher模型^[16]是一种被广泛地应用于半监督的图像分类和语义分割任务的半监督分类模型，解决了时间集成(Time Ensembling)模型在大数据集下更新缓慢和无法实现在线训练的问题。Mean Teacher半监督分类模型的核心思想是：模型既可以作为教师模型，也可以作为学生模型。当作为教师模型时，可以用来产生学生模型学习时的目标，教师模型的参数由历史上前几个学生模型的参数经过加权平均得到；当作为学生模型时，则利用教师模型产生的目标来进行学习训练。与时间集成模型相比，无标签数据的目标标签来自教师模型的预测结果，且教师模型参数无需等到一个epoch训练结束后才更新。模型在CIFAR-10数据集上的实验把当时的SOTA结果错误率从

10.55% 降到 6.28%，在 ImageNet2012 数据集上把错误率从当时 SOTA 结果的 35.24% 降到了 9.11%。

本文研究的出版产业链领域的实体以及实体之间的关系并没有大量标注好的语料库，因此需要在产业链技术谱系设计的基础上抽取出出版产业链中的各种实体，同时通过语料库总结归纳出常见的关系类型。关系抽取的实质就是对实体之间的关系类型进行自动分类，本文的具体方法是首先通过依存句法分析得到实体关系模板，然后通过人工标注部分少量的高质量关系模板，对得到的剩下的关系模板再按照与高质量关系模板的相似度分为两种模板（即中等质量关系模板和低质量关系模板）。最后，利用 Mean Teacher 半监督分类模型，把高质量关系模板训练得到的分类模型作为教师模型和学生模型，把中等质量关系模板和低质量关系模板在教师模型指导下训练得到学生模型，最后综合教师模型和学生模型得到一个新的分类模型，用于语料库文本中的关系抽取，也就是关系类型分类。

2 出版产业技术谱系设计

产业技术谱系的设计离不开产业链的分析及其相关概念。产业链是产业经济学中的一个概念，其概念起源于社会分工，产业链的内涵包含了技术链、供需链、价值链、产品链和空间分布组织等方面的链条，不同视角研究会形成不同的产业链环节。例如，郁义鸿^[17]将其定义为从原材料或矿产资源出发，经过加工、生产直至最终产品达到消费者手中的整个纵向链

条；芮明杰等^[18]将其理解为生产最终产品和服务所经历的从原材料到最终消费品的各个阶段，既涵盖厂商内部过程，也包括厂商间交易的各环节；魏后凯^[19]则将其视为基于分工经济的一种产业组织形式，强调了供应商、制造商、分销商以及零售商等加盟节点企业之间的分工合作关系；郑大庆等^[20]将其定义为产业内部企业或部门之间基于技术经济关联和逻辑、空间分布关系形成的一种关联模式。根据距离产业最终产品的远近程度，产业链可以分为上游、中游和下游三个产业链环节。针对传统出版产业，普遍观点认为传统出版产业链是“编—印—发”的链式结构（分别对应上游、中游、下游）；针对数字出版产业，本文借鉴邓佳佳^[21]的观点（上游为内容提供商、中游为内容出版商和技术服务商、下游为内容销售商和终端设备商），将数字出版产业链分为“内容提供—数字化加工与出版—内容销售与终端”。本文以产业链环节（上游、中游、下游）为纵轴，以业务环节、产业术语、技术术语、参与主体、产品服务为横轴，设计构建了出版产业的技术谱系知识，如表 1、表 2 所示。

在获得类似于表 1 和表 2 的出版产业的相关知识谱系实体之后，进一步可以挖掘实体之间的关系，得到诸如技术术语与产业术语、参与主体、产品服务的关系之后，可构建得到出版产业的技术谱系知识。为了获得出版产业技术谱系中的实体（主要是产业术语、技术术语、参与主体、产品服务），本文主要以《编辑与出版学名词》^[1]、出版技术相关论文、相关企业信息为基础，使用自然语言处理的关键词分析技术作为工具，从非结构化文本中获取实体候

表 1 传统出版产业技术谱系设计

产业链环节	业务环节	产业术语	技术术语	参与主体	产品服务
上游	编辑	例如：选题、组稿、审稿、校对、装帧、排版	例如：信息采集、信息处理	例如：责任编辑、总编辑、出版社、杂志社、报社	例如：书稿、文稿、校对稿、送审稿、完成稿
中游	印刷	例如：模拟打样、图文合一、投影制版、印刷材料、喷墨印刷机、数字印刷机	例如：激光成像、激光照排、静电喷墨、离子成像、直接热成像、油墨乳化	例如：印刷厂、印刷复制单位	例如：样书、印刷品
下游	发行与经营	例如：本埠发行、外埠发行、总发行、编发合一、集中征订、电话征订、订阅代理	例如：电子支付、移动支付	例如：新华书店、发行人、总发行单位、报刊亭、读者、书商、发行代理商	例如：图书、报纸、期刊、音像制品、电子出版物

表 2 数字出版产业技术谱系设计

产业链环节	业务环节	产业术语	技术术语	参与主体	产品服务
上游	内容提供	例如：数字内容、数字资源、数字信息	例如：智能写稿、智能校对、自动化排版、语音识别、中文信息处理、人工智能	例如：出版社、杂志社、报社、内容原创网站、游戏开发商、手机内容提供商	例如：原创网文、动漫、游戏、彩铃、彩信、手机报、传统出版物的纸质内容
中游	数字化加工与出版	例如：内容加工、数字化加工、知识化加工、数据清洗、数字内容重组、移动出版、数字出版、云出版、网络出版、跨媒体出版、按需出版	例如：语音识别、语音合成、移动互联网、云计算、大数据、数字签名、虚拟现实、增强现实、数字版权保护技术、中文信息处理	例如：数字音频编辑、视频剪辑师、互联网服务提供者、互联网内容提供者、数字技术提供者、新媒体制作中心、增值服务提供商、信息服务商	例如：数字音像、数字视频、电子书、电子出版系统、内容管理系统、信息检索系统、数字出版平台、数字出版产品、网络出版支撑系统
下游	内容销售与终端	例如：数字产品分销、数字营销、电子阅报栏、报网互动、移动终端、多媒体终端、智能手机	例如：电子支付、移动支付、图形用户界面、多媒体、全媒体、富媒体	例如：订阅服务提供者、数字内容集成商、数字内容分销商、信息服务商、数字内容零售商、终端设备制造商	例如：电子出版物、网络出版物、电子期刊、数字动漫、科学网络数据库、全文数据库、手机游戏、Kindle、iPad、电子书阅读器、阅读设备

选，然后获取出版产业技术谱系中的实体。《编辑与出版学名词》的内容包括综论、编辑、印刷、音像复制、发行与经营、数字出版、出版物、著作权等 8 部分，共 3380 条名词术语，每条名词术语均给出了定义或注释，这些术语和定义注释文本可以作为实体抽取的语料库。以“出版”“数字出版”“产业”“技术”为关键词组合，检索了万方数据知识服务平台，经过筛选后得到了发表年份为 2013—2023 年的论

文共计 1528 篇，把这些论文的标题、摘要、关键词也加入到语料库中。此外，从顺企网搜索“出版”相关的企业，得到出版相关的共 2000 家企业的“主要产品”和“经营范围”等描述文本，也作为非结构化的文本加入到语料库中。在非结构化语料库文本的基础上，使用无监督术语抽取的方法获得术语候选。无监督术语抽取方法的步骤如下：（1）按语句标点、回车换行等特殊字符将文本切分为语句文本；（2）对

语句文本使用 Python 中的 Jieba 分词模块进行分词处理,得到语句文本中的词语;(3)计算词语的词频、N-gram、TF-IDF、NC-Value^[22]、PMI^[23]等统计特征,对统计特征进行加权平均后得到一个总分数;(4)将总分数降序排列得到所需数目的术语候选集合。对于术语候选集合,经过人工筛选审核后得到出版产业技术谱系的术语实体,以《编辑与出版学名词》为准,补充和完善所得到的术语实体,并按照表1和表2所示分为产业术语、技术术语、参与主体、产品服务等类别。本文对非结构化文本进行处理后得到术语实体候选集合约2500个,经过人工筛选和补充完善后得到术语实体997个,其中产业术语491个,技术术语193个,参与主体105个,产品服务208个。

3 关系抽取方法

本文中出版产业技术谱系的实体已使用自然语言处理方法提前获取,且本文获得的出版产业的文本语料库没有大量标注好的高质量语料,因此在已有关系抽取研究基础上采用级联式的半监督关系抽取方法。具体的抽取方法是:第一,定义关系模板,将产业技术谱系中的实体对(例如:技术术语—关系—产业术语)在文本语句中所对应的关键词序列定义为关系模板,利用自然语言处理的依存句法分析等工具,可以自动提取得到出现实体对的语句文本所对应的实体对之间的关系模板;第二,对关系模

板划分等级,分为高质量、中等质量和低质量三类,高质量关系模板为根据出现频率人工筛选并标注后得到的关系模板,中等质量关系模板为与高质量关系模板语义相似度高于一定阈值的模板,低质量关系模板为与高质量关系模板语义相似度低于一定阈值的模板;第三,对关系模板中的文本词语,利用词向量工具得到词语的词向量表示,将高质量关系模板的词向量表示和关系模板的关系类型(即分类结果)输入到编码器(分类器)中进行深度学习模型训练,得到一个教师模型分类器;第四,使用教师模型分类器对中等质量和低质量关系模板进行预测得到伪标签(分类结果),然后利用伪标签训练分别训练得到中等质量关系模板和低质量关系模板对应的分类器模型;第五,将高质量、中等质量、低质量关系模板对应的分类器模型参数进行加权平均,得到一个总的模板关系分类模型。使用这个训练好的分类模型就可以预测技术术语与产业术语之间的关系类别。

3.1 关系模板和等级

本文的关系模板定义为文本语句中包含实体对及其关系描述的关键词序列,借助自然语言处理的依存句法分析工具(如HanLP^[24])可以自动从语句中提取得到关系模板。如语句“人工智能技术可以帮助编辑人员更高效地生成和编辑内容”经过依存句法分析后得到如图1所示的结果。

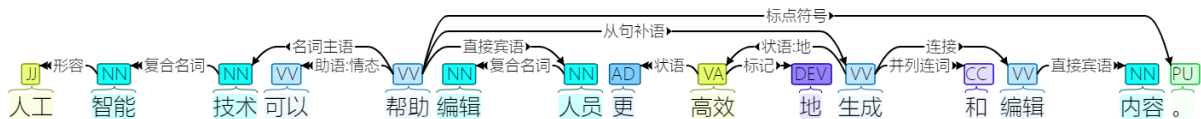


图1 依存句法分析结果例子

其中得到的包含头实体和尾实体的最短依存路径 (Shortest Dependency Path, SDP) 为:

“头实体 帮助 生成 尾实体 内容”, 该 SDP 就是包含头实体“人工智能技术”和尾实体“编辑”的关系模板, 其关系类型为“提升”即“技术术语—有助于—产业术语”, 而产业术语“编辑”所属的业务环节为传统出版产业链的上游, 因此也得到了人工智能技术与产业链上游环节的关联关系。

虽然可以自动得到包含头实体和尾实体的语句所对应的关系模板, 但无法自动获得头实体和尾实体之间的关系类型 (如上述例子中的关系类型“提升”)。为此先人工标注一部分关系模板的关系类型, 获得高质量的关系模板。由于标注成本限制且数量较少, 这部分关系模板不能直接用于神经网络分类器的训练和学习。对于未经人工标注的关系模板, 根据该关系模板与高质量关系模板的质量相似度分为两类不同的模板, 以使用不同的训练策略进行分类器模型训练, 同时可避免低质量模板带来的噪声和错误。质量相似度包括语义相似度、置信度和语义丰富度。

为了计算未标注模板 p_u 与已知关系类别的高质量模板的语义相似度, 可以利用词向量工具 (如 Word2Vec^[25]), 计算两个模板 (由词语及其向量构成) 的向量欧式距离。每一个关系类别的高质量模板可能会有很多个, 为保证计算结果的可靠性同时减少计算耗时, 取前 K 个最相似的语义相似度得分的平均值作为未标注模板与关系类别 m 的语义相似度。即首先获得构成模板的词语组合向量, 计算未标注模板 p_u 与某个关系类别 m 下所有高质量关系

模板的语义相似度, 然后得到 K 个最相似的语义相似度得分的平均值作为语义相似度 $\text{ClassSim}(p_u, P_{H,m})$ 。其中, $P_{H,m}$ 表示高质量关系模板中属于关系类别 m 的模板集合。

假设总共有 M 个关系类别, 为了突出未标注模板与最相似的关系类别的相似程度, 设计了一个置信度指标 conf_u (见式 (1))。因为语义相似度的取值为 0 到 1 之间, 语义相似度越大, 那么 1 减去语义相似度的值就越小, 取 $\log$ (自然对数) 之后再取负值则越大, 因此语义相似度越大则置信度指标的值也越大。同时需要对不是最相似的关系类别做消除相似的计算, 置信度通过式 (1) 计算得到。

$$\text{conf}_u = -\log(1 - \text{ClassSim}(p_u, P_{H,1st})) - \sum_{k \in \{2, \dots, M\}} \log(1 - \text{ClassSim}(p_u, P_{H,kst})) + \text{ClassSim}(p_u, P_{H,kst}) \quad (1)$$

在式 (1) 中, conf_u 为计算得到的置信度指标, $\text{ClassSim}(p_u, P_{H,kst})$ 则是未标注模板 p_u 与所有类别中排在第 k 位的关系类别 (高质量模板) 的语义相似度, 计算方法如前文所述。

语义丰富度主要考虑关系模板的长度, 可以取长度的三角正切函数的倒数, 使得其数值处于 0 到 1 之间。因为三角正切函数为单调递增函数, 所以取倒数和负值后仍然是单调递增函数, 即模板长度越长, 其语义丰富度越大。语义丰富度根据式 (2) 计算得到。

$$\text{dive}_u = -\tan^{-1} \text{Len}(p_u) \quad (2)$$

在式 (2) 中, $\text{Len}(p_u)$ 是关系模板 p_u 的长度, $\tan$ 是三角正切函数, dive_u 是语义丰富度。

最后, 综合语义相似度和语义丰富度, 得到未标注关系模板 p_u 的质量相似度得分值, 其值也在 (0, 1) 区间。质量相似度得分 qual_u 的

计算方法如式(3)所示。

$$qual_u = dive_u \times \log(1 - conf_u) \quad (3)$$

在式(3)中, $conf_u$ 是关系模板 p_u 根据式(1)计算得到的语义相似度, $dive_u$ 是根据式(2)计算得到的语义丰富度。 $qual_u$ 则是未标注关系模板 p_u 的质量相似度得分。

计算得到所有未标注关系模板的质量相似度得分后, 依据数据和经验设定阈值 τ , 质量相似度得分高于阈值的未标注关系模板归入中等质量关系模板集合, 质量相似度得分低于阈值的未标注关系模板归入低质量关系模板集合。由此, 所有关系模板被划分为三个等级: 高质量关系模板、中等质量关系模板、低质量关系模板, 在训练神经网络分类器模型时对不同质量关系模板采取不同的训练策略(损失函数), 不同程度地利用不同等级关系模板的类别标签及其语义信息, 实现从自然语言文本语

句到关系模板, 再从关系模板到关系类型标注的关系分类或抽取。

3.2 分类器模型和训练

在人工神经网络分类器模型中, 分类器的作用是接收输入的向量数据, 输出数据的分类类别。常见的人工神经网络分类器模型有循环神经网络(Recurrent Neural Network, RNN)、长短期记忆网络(Long Short-Term Memory, LSTM)和门控循环单元(Gate Recurrent Unit, GRU)模型。深度学习的注意力机制(Attention)借鉴了人类视觉注意力机制, 其本质是从众多信息中挑选出对于当前任务更加有用的信息, 这不仅提升了模型处理数据的效率, 同时保证了预测结果的准确性。因此本文采用双向门控循环单元加注意力机制(BiGRU+Attention)的神经网络架构作为深度学习模型中的分类器模型。

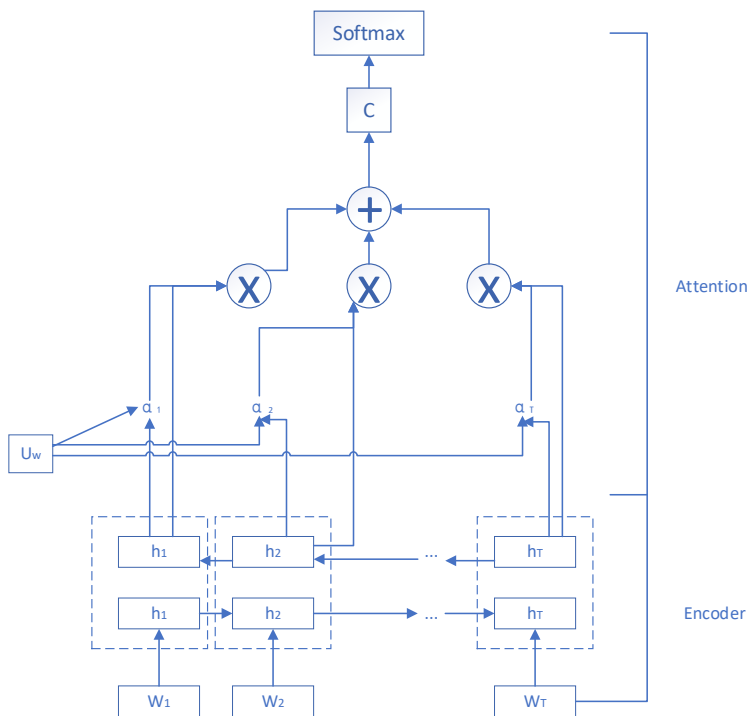


图2 分类器模型的网络结构

如图2所示,分类器模型的网络结构包括 Encoder 和 Attention 两个主体部分,其中 Encoder 部分分为嵌入层(向量序列)、双向 GRU 层; Attention 层的输出经过 Softmax 层后即分为分类编码器的分类输出。对于一个关系模板,假设其包含词语序列 $\{w_1, \dots, w_k, \dots, w_n\}$, 词语对应的词向量为 $\{c_1, \dots, c_k, \dots, c_n\}$, 词语的位置向量为 $\{v_1, \dots, v_k, \dots, v_n\}$, 通过向量拼接的方式得到输入数据的嵌入层向量序列 $\{x_1, \dots, x_k, \dots, x_n\}$, 其中 $x_i = [c_i; v_i]$ 。再把向量序列 $\{x_1, \dots, x_k, \dots, x_n\}$ 输入 BiGRU 模块(由前向 GRU 和反向 GRU 构成), 来获得隐层状态序列 $\{h_1, \dots, h_k, \dots, h_n\}$ 。即前向 GRU 在第 t 个时刻整合 $\{x_1, \dots, x_k, \dots, x_n\}$ 向量序列后,可以生成前向隐向量 $\vec{h}_t$, 而反向 GRU 可以生成反向隐向量 $\tilde{h}_t$, 这两个隐向量相加则可以得到在第 t 个时刻的隐向量: $h_t = \vec{h}_t + \tilde{h}_t$ 。

对于分类器模型的训练,本文采用一种常见的 Mean Teacher 半监督学习框架^[16],它通过在标记和未标记样本之间建立一种强大的交互关系,来提升模型的泛化能力。如图3所示,在整个分类器模型的训练框架中,关系模板的

分类模型主要由学生模型和教师模型组成,它们使用相同的 BiGRU+Attention 深度人工神经网络结构,其中教师模型的参数可以通过学生模型的参数平移得到。高质量样本为人工标注的关系模板及其对应的关系分类标签,可用于学生模型和教师模型的训练;中等质量样本为通过质量相似度得分得到的关系模板及其对应的关系分类伪标签;低质量样本也是通过质量相似度得分得到的关系模板,其伪标签质量较低,因此为无标签数据,但使用了低质量样本的语义信息。三类样本使用不同的损失函数来训练编码器分类模型,其中高质量样本用于训练教师模型(分类器),损失函数采用交叉熵;中等质量样本可同时用于训练学生模型,然后在教师模型指导下获得分类器模型参数,损失函数采用向量的欧式距离;低质量样本可用于训练学生模型,然后在教师模型的指导下获得分类器模型参数,损失函数采用 KL 散度。对高中低三类样本的模型参数进行加权平均,得到一个总的用于关系模板分类的模型,用于预测关系模板对应的关系类型。

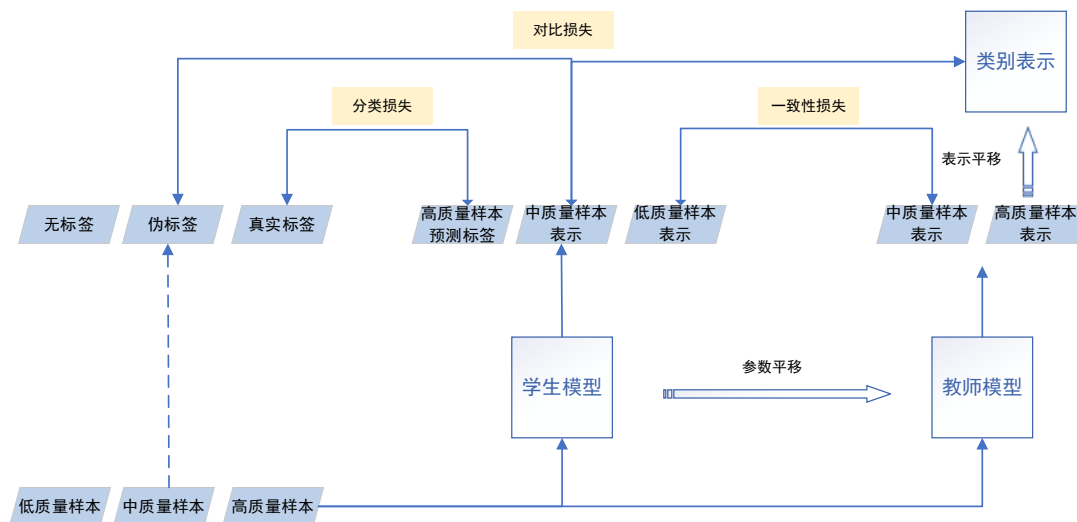


图3 基于关系模板等级的 Mean Teacher 训练框架

4 实验

为了验证基于模板质量划分的半监督 Bi-GRU+Attention 模型在抽取出版产业领域的产业术语和技术术语之间关系的性能和效果,本文以大模型检索的文本结果为基础,总结和设计了产业术语和技术术语之间的关系,并使用基于半监督的深度神经网络模型对关系模板进行学习和训练,通过实验以验证模型的关系抽取能力。由于技术术语的定义或者百科词条往往不会涉及出版行业,而出版产业术语的定义或者百科词条涉及的具体技术术语也不多,因此要寻找

同时出现技术术语和出版产业术语的大量文本极为困难。本文选择大模型搜索平台天工 AI 搜索 (<https://www.tiangong.cn/>) 作为工具,使用其“研究”搜索模型,并设计了提问问句,然后手工下载其搜索结果的大模型总结回答文本作为关系抽取的文本语料库。如表 3 所示,对于语料库中同时包含技术术语和产业术语的关系文本,通过依存句法分析可以自动地得到关系模板,同时可以人工标注其关系类型,并且通过产业术语实体的产业链环节得到技术术语所在的产业链的位置(上游、中游还是下游)。

表 3 AI 搜索提问和结果部分文本示例

提问问句	关系文本	关系模板	关系类型	产业链环节
人工智能涉及出版产业的哪些环节?	在印刷和发行环节,人工智能可以用于优化印刷流程,提高印刷效率和质量。	人工智能 用于 优化 印刷 流程 提高 印刷 质量	技术术语实体—应用于—产业术语实体 (人工智能—应用于—印刷)	传统出版产业的中游(印刷)
大数据技术在出版产业中有什么具体应用?	出版企业可以利用大数据技术探索新的出版模式,如按需出版、数字出版等,以适应数字化和个性化阅读的趋势。	出版企业 利用 大数据 技术 探索 出版 模式 按需出版	技术术语实体—应用于—产业术语实体 (大数据—应用于—按需出版)	数字出版产业的中游(按需出版)
语音识别技术和出版产业有什么关系?	语音识别技术可以用于编辑检索,提高编辑人员的工作效率。	语音识别 技术 用于 编辑 检索 编辑 人员 效率	技术术语实体—应用于—产业术语实体 (语音识别—应用于—编辑)	传统出版产业的上游(编辑)

本文设计了 3 类提问问句,使用本文第 2 节得到的部分技术术语和产业术语进行了提问检索,共得到大模型检索的回答文本 1768 篇,从中提取得到同时包含技术术语和产业术语的语句 9591 句。把 9591 句语句文本按照 3:1:1 的比例分为:训练集、验证集和测试集。对每个数据集中的语句文本使用依存句法分析工具(HanLP^[24])自动获取语句中的关系模板,同时根据关系模板的出现频率,手工挑选出少量模板(每一关系类型至少 10 个,至多 50 个)进行关系模板对应的关系类型标

注,形成高质量关系模板集合,剩余的关系模板作为未标注关系模板集合。对于未标注的关系模板集合中的关系模板,按照本文 3.1 小节的方法计算关系模板与关系类型(高质量模板)的质量相似度得分,区分为中等质量关系模板和低质量关系模板两类,K 取值为 3。这样得到了高中低质量的三类关系模板,以及这些模板对应的关系类型和词语序列样本,用于后续的半监督深度学习模型的训练和关系抽取。按照本文 3.2 小节的基于关系模板等级的 Mean Teacher 训练框架,对不同等

级质量的关系模板，使用不同的损失函数进行训练和关系抽取。训练集和验证集的语句文本用于模型的学习训练，测试集的语句文本用于进行关系抽取的正确率测试。在模型训练和关系抽取两个阶段，一些超参数的设置有所不同，主要是隐藏层的维度分别设置为 16 和 200，Batch 值分别设置为 16 和 80，关系抽取时中等质量和低质量对应的损失权重分别为 1 和 0.05（训练时权重相同），词向量维度、位置向量维度的设置在训练时和关系抽取时一样，都设置为 200 和 5。

在实验中设置了实体对的关系类型为 5 类，分别是：应用于、使用、集成、提升、其他。关系抽取的目标就是用训练好的半监督深度学习模型来预测实体间的关系类型，得到知识三元组（头实体，关系，尾实体）。但是实际的数据当中还包含大量未定义的关系类型，为了降低模型的训练难度，因此将未能识别的关系类型标记为其他。此外，部分关系类型（如使用和被使用关系）存在方向性，需要将“应用于”对应的反向关系“使用”加入模型。由于本文使用的语料库及其中的关系标注并没有标准答案和完全人工的标注关系，因此采用基于抽样的人工评估。该方法以抽样的方式从所有三元组中抽取部分样本人工标注，然后计算准确性指标作为关系类型分类的无偏估计。具体来说，按照关系类型分别抽样 200 条知识三元组，然后统计头实体、尾实体、关系类型抽取分类的正确与否。其中，关系抽取的结果需要结合样本进行判断。如果整个三元组（头实体，关系，尾实体），即标注的头实体、尾实体以及关系类型都正确，则认为知识三元组是正确的。通

过抽样标注的方法对 5 类关系和模型预测结果进行比对，其准确率结果平均为 66%，如表 4 所示。

表 4 半监督深度学习模型的预测结果准确率

关系类型	样本数	正确数量	错误数量	准确率
应用于	200	136	64	68%
使用	200	153	47	76.5%
集成	200	99	101	49.5%
提升	200	107	93	53.5%
其他	200	165	35	82.5%
平均值	200	132	68	66%

此外，为了验证本文的模型结构是否合理有效，设计了两个消融实验进行验证。一个是在模板质量划分时，仅仅考虑语义相似度而不考虑置信度和语义丰富度，即根据仅依靠最相似的高质量关系模板来确定未标注模板的关系类别，这个消融实验发现置信度和语义丰富度可以带来 1% 的正确率提升。因此，从语义丰富度、置信度等多个角度考虑未标注关系模板的划分，可以提高中等质量模板的健壮性和划分质量，从而提高了整个模型的分类正确率。另一个是消融实验考虑训练集和对比学习对于模型的影响，把中等质量模板集和高质量模板集合起来作为一个训练数据集，并且不考虑不同质量之间的损失函数对比学习。结果发现模型的预测性能在相同的测试集上比原来模型的正确率下降了 1%。这说明虽然直接使用伪标签（中等质量模板的关系类别）来训练分类器也能获得较高的分类正确率，但是加入对比学习后可以减少噪声样本对模型性能的影响，进一步提高模型分类器的关系分类效果。

5 结语

本文针对出版产业链,从产业链环节、业务环节、产业术语、技术术语、参与主体、产品服务六个维度对传统出版产业链、数字出版产业链进行了实体类型划分,并使用无监督的自然语言处理方法,以《编辑与出版学名词》为基准,获得了术语实体 997 个。为了获取产业术语和技术术语之间的关系,本文采用了半监督的深度学习模型方法。首先使用依存句法分析工具从包含术语实体的语句中获得最短依存路径的关键词语,作为关系模板并使用这些词语的词向量表示作为模板的语义表示。然后,利用模板的语义表示和长度等信息计算未标注模板与人工标注模板之间的语义距离及相似度,将关系模板划分为高中低三种类型的模板。借鉴 Mean Teacher 深度学习模型训练框架,使用基于对比学习的损失函数来训练编码器模型。不同质量的关系模板使用不同的损失函数,即高中低三类质量的关系模板分类器参数进行加权平均,得到一个总的分类器模型,可以预测包含实体对的文本语句对应的模板关系类型。通过出版产业的产业术语和技术术语之间的关系抽取实验,发现本文提出的基于模板质量的关系抽取半监督深度学习模型对于 5 种关系类型的预测准确率达到 66%。展望未来的工作,仍需要研究如何通过标注少量的关系模板获得更高准确率的关系类型,或者通过改进深度学习模型框架如加入语义知识等来提高模型的性能。

参考文献

- [1] 全国科学技术名词审定委员会. 编辑与出版学名词 [M]. 北京: 科学出版社, 2022: 16, 46, 87.
- [2] 薛平, 李影, 吴中海. 基于语言模型增强的中文关系抽取方法 [J]. 中文信息学报, 2023, 37(7): 32-41.
- [3] 任乐, 张仰森, 刘帅康. 基于深度学习的实体关系抽取研究综述 [J]. 北京信息科技大学学报 (自然科学版), 2023(6): 70-79, 87.
- [4] 武文雅, 陈钰枫, 徐金安, 等. 中文实体关系抽取研究综述 [J]. 计算机与现代化, 2018(8): 21-27, 34.
- [5] 郎春雨, 侯霞. 基于迁移学习的实体关系抽取技术综述 [J]. 北京信息科技大学学报 (自然科学版), 2022, 37(1): 65-70.
- [6] 袁清波, 杜晓明, 杨帆. 限定域关系抽取研究综述 [J]. 计算机系统应用, 2021, 30(9): 24-40.
- [7] 付瑞, 李剑宇, 王笏辉, 等. 面向领域知识图谱的实体关系联合抽取 [J]. 华东师范大学学报 (自然科学版), 2021(5): 24-36.
- [8] 刘彤, 魏静, 倪维健, 等. 基于半监督学习与 CRF 的应急预案命名实体识别 [J]. 软件导刊, 2020, 19(3): 35-38.
- [9] CHINATSU A, LAUREN H, et al. SRA: Description of the IE² system used for MUC-7 [C]// In Seventh Message Understanding Conference (MUC-7): Proceedings of a Conference Held in Fairfax, Virginia, April 29 - May 1, 1998: 1-14.
- [10] FUKUMOTO J, MASUI F, SHIMOHATA M, et al. Description of the Oki system as used for MUC-7 [C]// In Seventh Message Understanding Conference (MUC-7): Proceedings of a Conference Held in Fairfax, Virginia, April 29-May 1, 1998: 1-7.
- [11] ZELENKO D, AONE C, RICHARDELLA A. Kernel methods for relation extraction [J]. Journal of Machine Learning Research, 2003(3): 1083-1106.
- [12] 车万翔, 刘挺, 李生. 实体关系自动抽取 [J]. 中文信息学报, 2005, 19(2): 2-7.
- [13] ZENG D, LIU K, LAI S, et al. Relation classification via convolutional deep neural network [C]// International Conference on Computational Linguistics. The 25th International Conference on Computational Linguistics, 2014: 2335-2344.
- [14] SOCHER R, HUVAL B, MANNING C D, et al. Semantic compositionality through recursive matrix-vector spaces [C]// Proceedings of the 2012

- Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, 2012: 1201-1211.
- [15] 马江微, 吕学强, 游新冬, 等. 融合 BERT 与关系位置特征的军事领域关系抽取方法 [J]. 数据分析与知识发现, 2021, 5(8): 1-12.
- [16] TARVAINEN A, HARRI V. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results [C]// Advances in Neural Information Processing Systems 30, 2017: 1195-1214.
- [17] 郁义鸿. 产业链类型与产业链效率基准 [J]. 经济与管理研究, 2005(11): 25-30.
- [18] 芮明杰, 刘明宇. 产业链整合理论述评 [J]. 工业经济研究, 2006(3): 60-64.
- [19] 魏后凯. 产业特征、空间竞争与制造业地理集中 [J]. 管理世界, 2007(4): 68-77.
- [20] 郑大庆, 张赞, 于俊府. 产业链整合理论探讨 [J]. 科技进步与对策, 2011, 28(2): 64-68.
- [21] 邓佳佳. 产业链视角下的数字出版产业发展 [J]. 南昌大学学报 (人文社会科学版), 2014, 45(6): 73-76.
- [22] FRANTZI K T, MIMA H, et al. Automatic recognition of multi-word terms: the C-value/NC-value method [J]. International Journal on Digital Libraries, 2000(3): 115-130.
- [23] BOUMA G. Normalized (pointwise) mutual information in collocation extraction [J]. Proceedings of GSCL, 2009(30): 31-40.
- [24] HE H, CHOI J D. The Stem Cell Hypothesis: Dilemma behind Multi-Task Learning with Transformer Encoders [C]// Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, 2021: 5555-5577.
- [25] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality [C]// Advances in Neural Information Processing Systems 26, 2013: 3111-3119.



开放科学
(资源服务)
标识码
(OSID)

基于主题模型的政策工具自动标注研究

程佳雪 于卫红 齐勇凯 马一鸣

大连海事大学航运经济与管理学院 大连 116026

摘要: [目的/意义] 政策工具的标注工作是政策研究与政策工具识别模型建立的重要基础, 针对人工编码等传统的政策工具标注手段具有主观性且成本高的问题, 提出一种基于主题模型的政策工具自动标注方法, 以期提高政策工具标注的效率。[方法/过程] 首先, 通过政策工具理论与候选政策文本的关键词生成政策工具标注标签。随后, 将获取的政策文本进行细分处理, 对细分后的文本进行 BTM 主题挖掘, 生成政策工具主题标签。最后, 利用 Word2Vec 词嵌入模型将标注标签与主题标签转换为向量进行匹配度计算, 得到政策工具的标注结果。[结果/结论] 以“双一流”政策文本为例进行实证研究, 结果表明, 该方法能够较为准确地标注出政策文本中隐含的政策工具, 可以应用到今后的政策分析工作中。

关键词: 政策工具; BTM 主题模型; Word2Vec; 自动标注; 双一流政策

中图分类号: G35

Automatic Labeling Research of Policy Tools Based on Topic Model

CHENG Jiaxue YU Weihong QI Yongkai MA Yiming

School of Maritime Economics and Management, Dalian Maritime University, Dalian 116026, China

Abstract: [Objective/Significance] The annotation of policy tools is an important basis for policy research and the establishment of the identification model of policy tools. Aiming at the problems that the traditional annotation methods of policy tools such as manual coding are subjective and high cost, an automatic annotation method of policy tools based on topic model is proposed to improve the efficiency of policy tool annotation. [Methods/Processes] Firstly, the policy tool annotation labels are generated by policy tool theory and keywords of candidate policy texts. Then, the obtained policy texts are subdivided and subjected to BTM topic mining to generate topic labels for policy tools. Finally, Word2Vec word embedding model is used to convert annotation labels and topic labels into vectors to calculate the matching degree, and the annotation results of policy tools are obtained.

基金项目 辽宁省教育科学规划课题“电子商务专业课程供给侧改革研究”(JG21DB069)。

作者简介 程佳雪(1998-), 硕士研究生, 主要研究方向为深度学习、文本挖掘、政策量化, E-mail: chengjiaxue1998@163.com; 于卫红(1972-), 博士, 副教授, 主要研究方向为智能信息处理、文本挖掘、政策量化分析等; 齐勇凯(2000-), 硕士研究生, 主要研究方向为文本挖掘、图像处理与图像识别; 马一鸣(1998-), 硕士研究生, 主要研究方向为文本分析。

引用格式 程佳雪, 于卫红, 齐勇凯, 等. 基于主题模型的政策工具自动标注研究[J]. 情报工程, 2024, 10(6): 28-38.

[Results/Conclusions] The “double first-class” policy text is taken as an example for empirical research, and the results show that the method can more accurately label the implied policy tools in the policy text, and can be applied to the future policy analysis.

Keywords: Policy Tools; BTM Topic Model; Word2Vec; Automatic Labeling; Double First-class Policy

引言

政策是一个国家发展的总体纲领，指明了国家的总体发展战略。政策文本是政府活动存在的物质载体，是政府政策活动的具体体现形式，也是描述政府意图与施政过程的客观证据，所以政策文本研究是跟踪观察政策过程的重要方法^[1]。在政策文本研究过程中，政策工具被提到了重要的位置，政策从制定到执行的全过程都离不开政策工具的支持，基于政策工具来探究政策文本，可以更好地解析政策及其实施的内涵，进而更加深入地把握政策的核心含义，因此，识别政策文件中隐含的政策工具对于政策文本研究至关重要。从某种意义上说，基于深度学习的政策工具识别模型的建立是政策工具识别研究的未来，但要实现这一目标，需要构建海量的数据集，这就要求对政策工具进行大量的标注工作。

而在目前政策文本研究中，政策工具的标注工作多是由人工编码方式进行的，其基本流程为：政策文本采集—制定分析单元与编码标准—政策工具标注—政策工具统计^[2]。如刘春艳等^[3]对205份“新型智库”政策文本进行了统一编码，分析我国“新型智库”政策工具制定中存在的问题，并提出相应的对策建议。陈雅芝等^[4]对46个重点城市的生活垃圾分类政策按照“政策编号—（章节或部分）—具体条款”的顺序进行分析单元的编码，以此构建“目标—

工具”适配框架，探究了地方生活垃圾分类政策目标和工具的内容特征及二者间的适配程度。李宁等^[5]在乡村教师生活待遇政策制度和具体内容的基础上，将政策文本的每一分句作为基本单元对政策工具进行标注，以此来对政策工具的选择及其运用逻辑进行分析。还有一些学者运用编码工具与人工相结合的方式对政策工具进行标注，如兰娅菲等^[6]运用NVivo12软件对2002—2021年国家层面涉及民族医药的10项典型政策、地方层面42份代表性政策进行编码，基于构建的二维框架分析不同层级政策工具的选择及内部结构。何依恒等^[7]运用NVivo11软件分析了我国34篇社区教育政策文本，从而描述出21世纪以来我国社区教育政策工具的类型。Zhang等^[8]选择供给型政策工具、环境型政策工具、需求型政策工具、政策强度和医疗服务活动作为树节点，在每个树节点下构建子节点。根据节点和参考点的选择，使用Nvivo11软件对每个节点的参考点编码，以此对医疗服务政策工具进行分析。

从上述分析可以看出，目前政策工具标注方法以人工编码和机器辅助编码方法为主。人工编码依靠人工理解文本，具有很强的主观性和价值偏好，并且随着政策文本数量的迅速增加，人工方式的代价也越来越大，这极大地降低了政策工具研究的效率，严重阻碍了政策工具在政策分析中的广泛应用。政策样本中分析

单元的编码制定标准和标注结果是否合理可靠，也直接关系到后续研究中政策工具识别模型建立的准确性。因此，为了丰富现有研究，优化当前政策量化研究中政策工具的标注手段，为政策工具识别模型建立数据量足够大的并较为准确的训练库，本文结合主题模型和 Word2Vec 词嵌入模型，引入一种新的政策工具自动标注方法，替代传统的政策工具标注工作，提高政策工具标注的效率与效果，使政策工具的研究

更加深入。

1 基于主题模型的政策工具自动标注方法

基于主题模型的政策工具自动标注方法的流程主要可以分为五个步骤，分别为数据收集和预处理、政策工具标注标签的生成、政策工具主题标签的建立、标签筛选和匹配度计算，如图 1 所示。

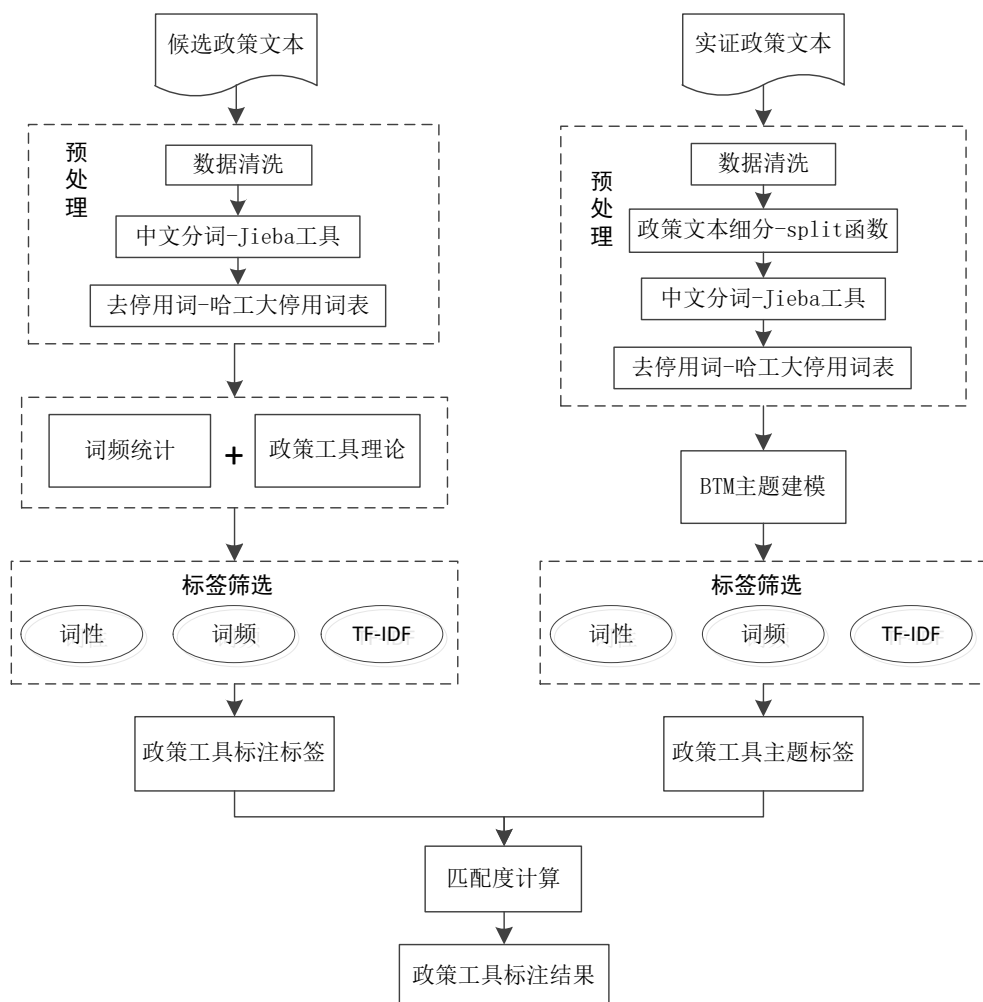


图 1 基于主题模型的政策工具自动标注方法流程

首先提取候选政策文本的关键词并结合政策工具理论生成政策工具标注标签；其次

对细分后的实证政策文本进行 BTM 主题挖掘，生成主题标签；然后运用 Word2Vec 词

嵌入模型，将筛选后的标注标签与主题标签转换为向量；最后，利用余弦相似度计算主题标签向量与标注标签向量的匹配度得到政策工具的标注结果。

1.1 数据来源

候选政策文本是为提取高频关键词建立政策工具标注标签，其要选取大量与实证文本相关的政策文本。由于公共政策内容以及作用的目标与对象的不同，可分为政治、经济、社会、教育、科技、文化等基本政策类型^[9]。如本文选取的实证政策为“双一流”政策，隶属于教育政策，则要选取教育类的政策作为候选政策文本。实证政策文本是为得到政策工具标注结果，本文以“双一流”政策文本为例进行实证研究，“双一流”建设作为我国新时代高等教育发展的重大战略工程，赋予了高等教育机构建设世界一流大学和世界一流学科的关键使命，“双一流”政策的实施为今后高等教育的发展指明了方向^[10]。基于政策工具的视角对其进行实证分析，有助于发现我国“双一流”建设政策中政策工具的结构与特征，为未来“双一流”建设政策的制定和优化提供理论支持。

1.2 政策工具理论

21 世纪初，政策工具理论被引入我国后，受到了国内学者的高度重视，他们对政策工具的概念和分类进行了深入探讨，政策工具在政策分析方面的应用迅速发展^[11]。从 1983 年英国学者胡德最早给出政策工具这个概念开始，国内外研究者从不同方面提出了不同的政策工具概念，其中，被认可程度最高的是欧文·休斯^[12]在其所撰写的《公共管理导论》中对政策工具的解释：公共政策工具是政府机构制定公共政策时为了保障政策目标的顺利实现所采取的一系列措施，以及对这种措施进行调整的机制。政策工具，简单地说，就是达到政府目的的手段。对政策工具的研究建立在政策工具分类的基础上，不同的分类标准会有不同政策理解，不同代表人物及其分类体系如表 1 所示。在众多分类中，公共政策学家麦克唐纳尔和埃尔莫尔的分类体系相对更为细化^[5]。其中，命令性工具主要是指通过强制手段对个体或组织机构的规则进行规范，如法律、标准和监督等；能力建设工具主要是指对未来产生影响的资源投入，如人才培养，设施建设等；系统变化工具是指个体或组织内部通过权力转移的方式达到

表 1 代表人物及分类体系

代表人物	分类依据	政策工具分类体系				
胡德 ^[13]	政策工具自身活动功能	权威工具	财力工具	组织工具	信息工具	
迈克尔·豪利特、M·拉米什 ^[14]	政府权力直接介入程度	强制性工具	自愿性工具	混合性工具		
罗斯韦尔、泽格维尔德 ^[15]	政策工具影响作用对象	环境型工具	需求型工具	供给型工具		
麦克唐纳尔、埃尔莫尔 ^[16]	政策工具所要达成目标	命令性工具	激励性工具	能力建设工具	系统变化工具	劝告工具
陈振明 ^[17]	现代西方国家政府改革与治理的实践趋势	市场化工具	工商管理技术	社会化工具		

政策目的,如权力重组与制度变革等;激励性工具主要是对差异性行为进行鼓励,如奖金、补贴等激励措施;劝告工具是指政府通过号召、树立榜样等手段对目标团体进行引导和启迪,使其采取符合政策目的的行动。与其他政策工具分类相比,麦克唐纳尔和埃尔莫尔提出的政策工具分类体系具有更显著的特点,针对性与操作性也更强,并且针对每种政策工具都提出了最适用的应用场景、所需成本和使用中需关注的问题。由于该分类更为细化,且为每种政策工具都设定了明确的边界和清晰的分类标准,因此,本文选择这种政策工具分类体系来结合候选政策文本的关键词生成政策工具标注标签。

1.3 主要流程

1.3.1 政策文本预处理

预处理是数据分析的基础,文本预处理的结果直接关系到实验结果的好坏。本文的预处理流程包括:

(1) 政策文本内容过滤:在爬取所有政策文献后,需要对政策文献进行人工筛选,所留文献要满足以下条件:一是政策出台单位为中央和地方(不包括港澳地区)及有关部门;二是保留通知、意见、方案等能直接反映政府态度和措施的政策文件,剔除公告、办事服务、政民互动等无效文件;三是政策发文时间要在有效时间内。

(2) 政策文本细分:为实现不同的政策目标,一篇政策文本不可避免地包含一系列相互作用的工具,这表现在其单句或相邻多句的文本中。因此,在进行政策工具标注之前,

本文首先运用 `split` 函数以符号“!”“。”“?”为分隔符,将政策文本细分成多个相对独立的片段或段落,以确保更精确地标注出每一段政策文本中所隐含的政策工具。

(3) 中文分词和去停用词:以提高文本语料的准确性和有效性为目的,需要对获取的政策文本进行分词并对其中无用的标点符号如“[]”“《》”“、”等和过渡词如“和”“或”“但”等进行删除,保证后续的量化分析工作更加高效。

1.3.2 主题建模

本文使用词对主题模型(Biterm Topic Model, BTM)来进行文本主题挖掘。与传统的主题模型如隐含狄利克雷分布模型(Latent Dirichlet Allocation, LDA)等相比,BTM模型弥补了其在较短文本上进行主题建模时性能下降的问题,更适用于细分后的政策文本。BTM模型的核心思路在于利用整个语料库中频繁共现的词对进行聚合,以此来学习并揭示短文本中的主题,这种方法有效地克服了单个文档因稀疏性所带来的问题。具体来说,整个语料库被认为是一个主题的混合体,其中每个词对都是从一个特定的主题中独立提取的,从特定主题中提取词对的概率进一步通过词对中两个词都是从主题中提取的概率来捕获^[18]。假设 θ 是语料库中的主题概率分布, θ_z 是某主题下词的概率, α 和 β 服从 Dirichlet 先验参数。可以将 BTM 中所有词对在语料库中的具体生成过程描述如下^[19]:

(1) 对于每个主题 z ,描述这个特定主题的单词分布 $\theta_z \sim \text{Dir}(\beta)$;

(2) 为整个短文本语料库描述一个主题分

布 $\theta \sim \text{Dir}(\alpha)$;

(3) 对于词对集合 B 中的每个词对 b :

a) 绘制主题分布 $z \sim \text{Multi}(\theta)$;

b) 抽样出两个词 $w_i, w_j \sim \text{Multi}(\phi_z)$, 组成词对 b 。

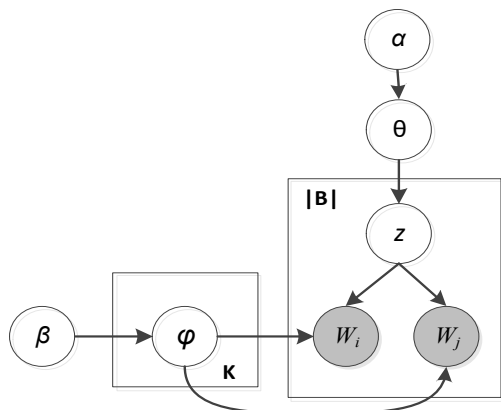


图2 BTM 主题模型

BTM 模型生成过程如图 2 所示, 假设词对独立生成, 根据上述生成过程, 一个词对 $b=(w)$ 的联合概率以及整个语料库的概率计算公式可以写成:

$$P(b) = \sum_z P(z) P(w_i|z) P(w_j|z) = \sum_z \theta_z \phi_{i|z} \phi_{j|z} \quad (1)$$

$$P(B) = \prod_{(i,j)} \sum_z \theta_z \phi_{i|z} \phi_{j|z} \quad (2)$$

1.3.3 标签词筛选

结合候选政策文本和政策工具理论生成政策工具标注词, 对细分后的政策文本进行 BTM 主题挖掘得到各个政策分句的主题词。在将标注词和主题词表转换为词向量之前, 还需对其进行针对性的筛选, 以确保匹配结果的准确性。在筛选时要考虑以下三点:

(1) 词频, 一般情况下, 低频的词语更加脱离文章的中心思想, 因此去除词频低于 30% 的单词。

(2) 词性, 从语言角度上看, 动词和名词

是更具有描述性的词语, 能更好地表达主题。

因此, 仅保留最能体现整篇文档中心思想的动词和相关名词, 过滤掉其他对于文档的重要程度较低的词汇类型, 如形容词, 介词, 专有名词等, 减少计算复杂度, 提高匹配的准确性。

(3) 语料库, 判断一个词语的重要性不仅要考虑词语本身, 还要考虑词语所在的语料库。

TF-IDF 是一种统计方法, 用于评估一个词在特定文档中相对于整个文档集合的重要性。根据 TF-IDF 权重因子, 能排除如“教育”“共和国”“国家”及“科学”等无意义的高频词。

1.3.4 相似度计算

本文使用余弦相似度进行标签向量的匹配, 该方法是计算向量相似度的有效手段。在这里本文运用 Google 公司开发的 Word2Vec 词嵌入模型, 它存在一个重要的假设是文本中相邻词语之间的相似度更高, 两个词向量的相似度可以通过它们的余弦值来衡量。本文利用 Python 中的 Word2Vec 工具包对词向量进行训练, 得到的词向量用以支持后续的实验。式 (3) 表示向量 d_1 与向量 d_2 的相似度:

$$\text{sim}(d_1, d_2) = \frac{\vec{d}_1 \cdot \vec{d}_2}{\|\vec{d}_1\| \|\vec{d}_2\|} \quad (3)$$

在公式中, 两个向量的点积作为分子, 而两个向量模的乘积则构成分母。余弦相似度值为 -1 到 1, 值越趋近 1 意味着两个词语在语义层面上的相似度越高, 值越小则相反^[20]。

2 实证研究

2.1 数据采集

本文采集的数据为需要建立标注标签的

教育政策文本和实证文本,教育政策文本依据《中华人民共和国教育法》第十一条,在北大法宝网站以七个方面的关键词进行精确检索,每一方面爬取文献二十篇左右,经预处理,保留有效文件 149 篇。实证政策文本以“双一流”“一流大学”“一流学科”及“一流大学和一流学科”为检索词,检索范围选取“标题”在中国政府网,中华人民共和国教育部和各地区人民政府网进行精确检索,共检索出 122 篇文件。将全部文件分为中央、地方两个层级,其中地方按“七五”规划划分为中部、西部、东部地区。经政策文本内

容过滤后,获得有效文件 109 篇,其中中央层级包含文献 9 篇,地方层级分别为中部 27 篇,西部 26 篇,东部 47 篇。再经政策文本细分处理后,共获 8164 条文本数据。

2.2 政策工具标注标签生成

(1) 教育政策文本高频词提取

首先预处理爬取到的教育政策文本,剔除形容词、介词、连词、专有名词及含义不明确的语词如“等于”“有些”等,仅保留动词和相关名词,用剔除后的词语建立高频词表,教育政策文本部分高频词如表 2 所示。

表 2 教育政策文本高频词表(节选)

高频词 -Noun	词频	高频词 -Noun	词频	高频词 -Verb	词频	高频词 -Verb	词频
教师	1393	技术	1025	规定	844	鼓励	270
项目	1352	改革	736	加强	725	负责	250
资金	1307	标准	736	支持	572	提供	246
职业	1305	人才	559	制定	354	投入	172
资源	1081	经费	481	培养	306	组织	171

(2) 政策工具标注标签生成

通过上述分析,根据麦克唐纳尔和艾莫尔的政策工具分类体系,结合获取的教育政策文本高频词表,通过逐一比对关键词与政策工具的定义及特征,进行了精准的分类匹配。例如,高频词“程序”或“法律”等被归类为命令性工具标签,因为其体现了政府通过直接规定等强制性措施来要求执行的行为模式;而“补贴”或“奖励”等高频词则被归类为激励性工具标签,因为其体现了通过经济手段来鼓励或促进特定行为。经过上述匹配过程,得到政策工具的标注标签如表 3 所示。

2.3 “双一流”政策文本主题标签生成

利用 python 中的 BTM 工具包对细分后的“双一流”政策文本进行模型训练,主题数设置为 $n_topics=1$,主题词设置为 5,经过 100 次迭代训练后,得到中央发布的“双一流”政策文本主题词表如表 4 所示。

2.4 相似度计算

通过 2.2 节计算得到了政策工具标注标签,将标注标签与 2.3 节获取的“双一流”政策主题标签进行相似度计算,得出各主题标签与政策工具标注标签的相似度。其中,设置相似度

表 3 政策工具标注标签表

政策工具类型	政策工具标注标签
命令性工具	程序、章程、标准、规定、规范、法规、保障、监管、监督、监察、督查、检查、审核、审批、监测、依法、法律、权益、公正、安全、公平、透明、有序、严格、合理、合格、严肃、负责、严禁、警示、约束、备案、上报、违反、追究、废止、遵守、遵循、责任、担任、权限、自主权、权责、责令、考核、评价、评议、评估
激励性工具	激励、奖励、授权、惩罚、经费、专项、资金、工资、政策、支持、补贴、补助、财政、收入、支出、经济、绩效、预算、投资、资助、奖补、效益、待遇、拨款、福利、支付、扶持、购买、购置、减免、拨付、捐资、采购、专款
能力建设工具	资源、投入、倾斜、建设、建立、基地、平台、场地、科研、文化、基础、设施、设备、装备、仪器、教师、师资、队伍、团体、项目、学生、毕业生、机构、人才、培养、培育、招收、选拔、职业、培训、区域、试验区、招生、技能、技术、创新、新建、实习、开发、试行、试点、增设
系统变化工具	权力、重组、优化、成立、组织、职能、界定、改革、改善、配置、分配、结构、战略、决策、调整、整改、晋升、任职、修订、选拔
劝告工具	鼓励、宣传、学习、引导、典型、精神、表彰、思想、示范、导向、鲜明、优秀、优良、伟大、推广、营造、振兴、树立、弘扬、理念、观念、道德、氛围、校风、教风、学风、师德、师风

表 4 “双一流”政策文本主题词表(节选)

政策文本单位	主题词
加强师德师风建设，培养和造就一支有理想信念、有道德情操、有扎实学识、有仁爱之心的优秀教师队伍	教师 学识 师德师风 情操 理想信念
坚持立德树人，突出人才培养的核心地位，着力培养具有历史使命感和社会责任感，富有创新精神和实践能力的各类创新型、应用型、复合型优秀人才	培养 创新 实践 能力 人才
要强化跟踪指导，对建设过程实施动态监测，及时发现建设中存在的问题，提出改进的意见建议	实施 强化 改进 监测 跟踪
加强组织实施，教育部加强政策措施统筹协调，成立人工智能高层次人才培养专家委员会，指导高校实施人才培养专项计划，及时总结推广可复制的经验和做法	成立 委员会 组织 指导 协调
对于建设成效特别突出、国际影响力特别显著的少数建设高校及建设学科，在资金和政策上加大支持力度	建设 成效 支持 政策 资金

值最高的为政策工具匹配结果并进行输出，相似度值保留小数点后两位，考虑到相似度计算方法及选择的语词对匹配度的影响，以及对各个标签与主题词的匹配度进行分析，发现相似度 >0.5 的主题标签能较为准确地匹

配政策工具，因此保留相似度 >0.5 的政策工具匹配结果，相似度在 0.5 以下设置为此分句不存在政策工具。以中央发布的“双一流”政策文本为例，其政策工具匹配结果如表 5 所示。

表 5 “双一流”政策工具匹配结果(节选)

政策主题标签	命令性工具标签	激励性工具标签	能力建设工具标签	系统变化工具标签	劝告工具标签	政策工具匹配结果	相似度
教师 学识 师德师风 情操 理想信念	0.67	0.62	0.68	0.51	0.88	劝告工具	0.88
培养 创新 实践 能力 人才	0.32	0.64	0.91	0.48	0.67	能力建设工具	0.91
实施 强化 改进 监测 跟踪	0.75	0.53	0.61	0.36	0.27	命令性工具	0.75
成立 委员会 组织 指导 协调	0.72	0.66	0.69	0.85	0.52	系统变化工具	0.85
建设 成效 支持 政策 资金	0.51	0.83	0.71	0.66	0.47	激励性工具	0.83

2.5 实验结果分析

通过上述分析得到“双一流”政策工具的标注结果,中央层面和地方层面的“双一流”政策工具分布情况分别如图3和图4所示。如图3所示,在中央发布的“双一流”政策文件中,命令性工具使用最为频繁,劝告工具和能力建设工具也使用得较多,激励性工具与系统变化工具的使用相对来说占比较小。而在地方发布“双一流”政策文件的政策工具分布情况中,能力建设工具使用频率最高,激励性工具和命令性工具分别排第二位和第三位,使用劝告工具和系统变化工具最少。这表明,地方政

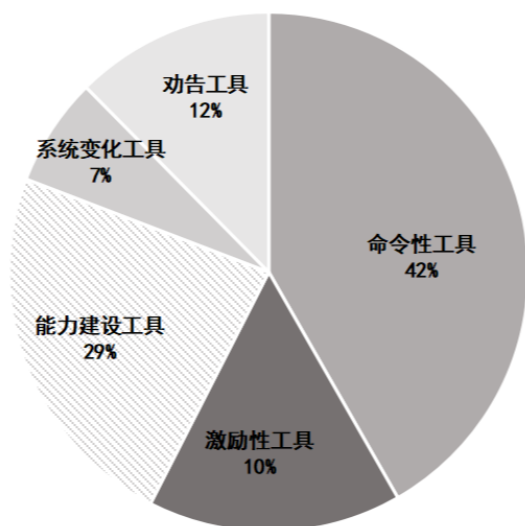


图3 中央层面“双一流”政策工具分布情况

府在地方层面,又将“双一流”政策分为东中西部,其政策工具的使用状况如图5所示,从整体情况来看,东中西部的双一流建设政策工具使用状况趋势接近,使用最多的是能力建设工具,其次是命令性和激励性工具,使用最少的是系统变化和劝告工具。从个体情况来看,在东部地区,

府推进“双一流”建设的政策工具使用情况与中央推进双一流建设的方式有一定的相似之处,存在某种程度上的模仿。中央政府在“双一流”建设中政策工具使用情况展现的特征是权威与强制,因为中国高等院校的办学主体主要是中央政府部门,而高等教育机构和政府部门有完全的依附关系,所以直接行政一直是中国政府部门使用命令性工具实施高等教育政策的最主要表现形式^[21]。地方政府在推动“双一流”工程建设时更加倾向于使用能力建设工具和激励性工具,希望借助整个省域的能力建设,推动地方一流大学和一流专业不断发展。

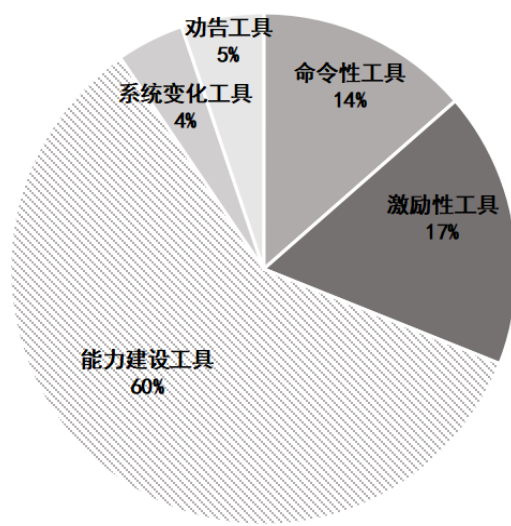


图4 地方层面“双一流”政策工具分布情况

能力建设和激励性工具运用较多,这表明地方政府在东部地区“双一流”建设中更倾向于长期性的、基础性的投入带来的政策价值。而在中部地区和西部地区,命令性工具的占比更多,尤其是在西部,说明政府在中西部地区更趋向于采用直接管制的手段达到双一流建设的目标。

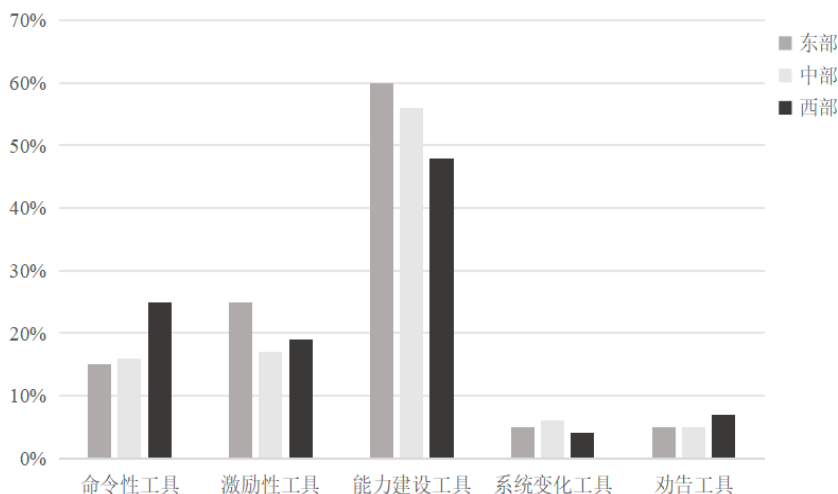


图5 东中西部“双一流”政策工具分布情况

3 模型评估

实验采取常用于评价算法质量的准确率、召回率和F值作为模型评估指标。其中，准确率用于评价政策工具标注结果表现原文档政策工具标注信息的准确度，召回率用于评价政策工具标注结果对原文档政策工具标注信息的覆盖程度，F值是衡量准确率和召回率的指标，它的大小为准确率与召回率的加权调和平均数^[22]。通常认为，F值越高，算法效果越好，即标注准确率越高。在获取的8164条文本数据中选取20%的文本数据作为测试数据，政策工具自动标注算法性能结果如表6所示。

表6 政策工具自动标注算法性能

P 值 (%)	R 值 (%)	F 值 (%)
81.28	78.29	79.76

综合以上测试数据和分析，政策工具自动标注算法实验方案合理，能够较为准确地标注出政策文本中隐含的政策工具，大量节省了人工成本，提高了政策工具标注的效率与效果，

弥补了以往研究的不足，政策工具的研究深度进一步拓展。

4 结论与展望

本文在BTM主题模型的基础上，融入了Word2Vec词嵌入模型，提出了一种基于主题模型的政策工具自动标注算法。首先通过提取候选政策文本的关键词，在政策工具理论的基础上生成政策工具标注标签；其次利用BTM主题模型对细分后的政策文本单位进行主题建模，提取主题词生成主题标签；再次运用Word2Vec词嵌入模型将标注标签和主题标签转换为可直接应用于计算的词向量；最后利用余弦相似度计算政策工具标注标签和主题标签之间的匹配度，得到政策工具的标注结果。经过分析发现，基于主题模型的政策工具自动标注算法能够较为准确地标注政策文本中隐含的政策工具。

为进一步优化和完善该研究，后续研究将从以下方面展开：

(1) 考虑主题词在政策文本单位中的比重，

将主题词占比权重加入向量转化过程中。

(2) 在“双一流”工具分类中仅设置了一级政策工具, 后续分析中会加入一级政策工具下的子政策工具的标注, 更好地探讨“双一流”政策中政策工具的运用逻辑。

(3) 结合本文的算法基础, 形成政策工具标注训练集, 尝试将深度学习技术应用于政策工具识别领域当中, 进行更精准的识别。

参考文献

- [1] 郑新曼, 董瑜. 政策文本量化研究的综述与展望[J]. 现代情报, 2021, 41(2): 168-177.
- [2] 李娜, 姜恩波, 朱一真, 等. 政策工具自动识别方法与实证研究[J]. 图书情报工作, 2021, 65(7): 115-122.
- [3] 刘春艳, 孙雨晴. 政策工具视角下中国特色新型智库政策文本分析[J]. 智库理论与实践, 2022, 7(5): 1-10.
- [4] 陈雅芝, 罗丹, 黎江平. “目标—工具”适配视角下生活垃圾分类政策量化研究——以46个重点城市为例[J]. 生态经济, 2022, 38(10): 198-205.
- [5] 李宁, 李中国. 乡村教师生活待遇政策效应研究——基于政策工具的视角[J]. 教育学术月刊, 2023(1): 98-105.
- [6] 兰娅菲, 陈颖, 韩朦, 等. 政策工具视角下我国民族医药政策分析[J]. 医学与社会, 2022, 35(10): 67-71, 85.
- [7] 何依恒, 卢德生. 21世纪以来我国社区教育政策研究——基于NVivo的政策文本分析[J]. 成人教育, 2022, 42(10): 34-40.
- [8] ZHANG H, ZHU L, ZENG C, et al. Text Mining and Quantitative Research of Medical Service Policy: Sichuan Province as an Example[J]. Frontiers in Public Health, 2021(8): 509842.
- [9] 宁骚. 公共政策学[M]. 北京: 高等教育出版社, 2011: 109.
- [10] 许杰. 试析“双一流”建设政策工具视角下的学科治理[J]. 国家教育行政学院学报, 2021(12): 48-56.
- [11] 李玲洁. 基于定性比较分析的农民工返乡创业政策效果研究[D]. 昆明: 云南师范大学, 2022.
- [12] 欧文·E. 休斯, 彭和平. 公共管理导论[M]. 北京: 中国人民大学出版社, 2001: 43.
- [13] HOOD C. The Tools of Government[M]. London: The Macmillan Press Ltd, 1983: 38.
- [14] 迈克尔·豪利特, M·拉米什. 公共政策研究: 政策循环与政策子系统[M]. 上海: 上海三联书店, 2006: 146.
- [15] ROTHWELL R. Reindustrialization and technology: towards a national policy framework[J]. Sci Public Policy, 1985, 12(3): 113-130.
- [16] MCDONNELL LORRAINE M, ELMORE RICHARD F. Getting the job done: alternative policy instruments[J]. Educational Evaluation and Policy Analysis, 1987, 9(2): 133-152.
- [17] 陈振明. 政策科学——公共政策分析导论[M]. 北京: 中国人民大学出版社, 2003: 12-18.
- [18] YAN X, GUO J, LAN Y, et al. A bitern topic model for short texts[P]. World Wide Web, 2013.
- [19] 张文伟, 赵辉. LDA与BTM概率主题模型抽取科学主题效果比较研究[J]. 情报工程, 2020, 6(2): 66-77.
- [20] 孟秋晴, 熊回香, 童兆莉, 等. 基于在线问诊文本信息的医生标签自动生成研究[J]. 情报科学, 2020, 38(5): 58-64, 72.
- [21] 徐赞. “双一流”建设中政策工具选择与运用的问题及对策[J]. 教育发展研究, 2018, 38(1): 26-32.
- [22] 李雯. 教育资源标签生成方法研究[D]. 长沙: 长沙理工大学, 2020.



开放科学
(资源服务)
标识码
(OSID)

高校图书馆用户阅读影响因素分析

陆桃妹 张芹 刘天琳 景民昌

中国石油大学(北京)图书馆 北京 102249

摘要: [目的/意义] 高校图书馆用户作为读者的重要构成,其阅读活动关系着全民阅读的推进。[方法/过程] 系统分析高校图书馆用户阅读影响因素,通过专家函询识别确定出用户因素、高校因素和社会因素3大类影响高校图书馆用户阅读的因素,构建了DEMATEL-ISM分析框架,对各影响因素的重要程度、因果属性以及层级关系进行了计算和划分。[结果/结论] 研究结果表明,就业关联性是影响阅读的本质因素,阅读需求和阅读动机是直接因素,各因素由下至上逐层产生影响。研究结果可为高校图书馆及相关阅读推广人员提升图书馆用户阅读积极性提供新的思路和决策依据。

关键词: 高校图书馆; 阅读; 决策实验室分析法; 解释结构模型

中图分类号: G35

Analysis of Factors Influencing the Reading Activity of University Library Users

LU Taomei ZHANG Qin LIU Tianlin JING Minchang

Library, China University of Petroleum, Beijing 102249, China

Abstract: [Objective/Significance] As an important constituent of readers, the reading activities of university library users are related to the advancement of nationwide reading. [Methods/Processes] This paper analyzes the factors influencing the reading activity of university library users, identifies three major categories of factors affecting the reading activity of university library users, including user factors, university factors, and social factors through expert consultation, and constructs the DEMATEL-ISM analysis framework. The importance, causal attributes, and hierarchical relationships of various influencing factors have been calculated and divided. [Results/Conclusions] The research results show that the correlation with employment is the essential factor affecting reading activity, while reading needs and reading motivation are the direct factors. Various factors produce effects from bottom to top. The research results can provide new ideas and decision-making basis for university libraries and related reading promotion personnel to improve the reading activity of library users.

Keywords: University Library; Reading; Decision-Making Trial And Evaluation Laboratory; Interpretative Structure Modeling

基金项目 北京高校图书馆研究基金项目“创新需求下基于专利文献的颠覆性技术识别--以石墨烯技术为例”(BGT2024021); 中国石油大学(北京)研究生教育质量与创新工程一般项目“基于高影响论文和专利文献的学科竞争力评价探究”(yjs2023029)。

作者简介 陆桃妹(1992-), 硕士, 馆员, 主要研究方向为图书情报分析、石油科技信息挖掘、专利信息服务, E-mail: lutm2016@cup.edu.cn; 张芹(1981-), 博士, 副研究馆员, 主要研究方向为图书情报分析及专利信息服务; 刘天琳(1992-), 硕士, 馆员, 主要研究方向为图书情报分析及专利信息服务; 景民昌(1971-), 博士, 研究馆员, 主要研究方向为大数据处理、文本挖掘。

引用格式 陆桃妹, 张芹, 刘天琳, 等. 高校图书馆用户阅读影响因素分析[J]. 情报工程, 2024, 10(6): 39-48.

引言

2014年“倡导全民阅读”第一次写入国务院政府工作报告,《“十三五”规划纲要》要求“推动全民阅读”,将全民阅读提升到了国家战略高度,《“十四五”规划纲要》则更明确提出,要“深入推进全民阅读,建设‘书香中国’”。可见阅读是一件与所有人密切相关的大事。有学者指出,阅读是衡量个人竞争力的关键要素,也是创造健康心智、促进幸福感、建立终身学习所需能力不可或缺的条件^[1-2]。

高校图书馆作为人才培养和文化引领的重要阵地,既是海量知识的提供者,也是阅读行为的引导者和阅读风气的营造者,是推进全民阅读的重要分支力量。其中,阅读推广是图书馆最具活力、最能充分体现图书馆核心价值的服务^[3]。高校图书馆用户不仅有学生,还包括相关科研工作者。近年来,伴随社会多元化和互联网信息技术飞速发展,高校图书馆用户的阅读心理和行为模式发生较大改变,阅读活动趋向主动化、社交化和移动化,阅读需求呈现个性化、多元化特征^[4-7]。

笔者基于中国知网(CNKI)检索平台,以“高校图书馆阅读”“大学生阅读”为主题进行专业检索,获得近五年(2019.1.1—2023.12.31)我国学者发表的相关基金文献共计603篇。在发表年度上呈现稳中略降的趋势,对研究主题进行统计发现,出现频次最多的是“阅读推广活动”“阅读推广模式”“阅读推广策略”“阅读行为”“图书馆阅读”。从发文机构上看,主要集中在高校图书馆及信息管理学院等相关部门。高校图书馆从图书馆角度

出发,以提升读者阅读频次、阅读范围等为目标,持续关注阅读推广活动模式^[8-9]、策略及评价体系建设^[10-11]等方面的研究,旨在推动高校图书馆用户进行阅读活动,同时也能看到,关于阅读影响因素的研究也日益增多。如韩丽^[12]将自我决定理论引入阅读动机研究,基于阅读动机构建了高校读者阅读意愿影响因素理论模型,针对受控动机、自主动机和基本心理需要分别进行了量表分析;巫芯宇^[8]指出阅读主体、阅读文本、阅读认知、终端技术和阅读环境是大学生深阅读行为的影响因素,并分析了其形成机制;刘晓莉^[13]发现阅读推广信息的相关性、及时性、个性化、互动性和趣味性是影响读者阅读意愿的重要因素;姚晓彤^[14]归纳总结了高校信息素养教育、家庭因素以及外向因素对娱乐型阅读、理性专业型阅读以及无目的型阅读的影响作用;关磊^[15]分析了技术因素、心理因素和社交因素对高校有声阅读推广有效性的影响。这一方面是因为图书馆用户阅读的研究分支比较多,包括阅读心理、阅读动机、阅读需求等多个方面,另一方面是因为阅读的指标也具有多维性,包括阅读频率、阅读时间、阅读深度等多个维度。正因如此,对影响高校图书馆用户阅读的因素进行深度分析,厘清各影响因素的重要性和层次关系,才能对症下药,更好地进行阅读推广。

本文使用DEMATEL-ISM分析法对影响高校图书馆用户阅读的因素进行分析,构建阅读影响因素分析框架,为高校图书馆及阅读推广工作者提供对策建议,进一步推动高校图书馆阅读推广工作。

1 整体研究方法思路

本文采用三步法开展研究，如图 1 所示。
第一步：通过文献检索结合专家函询，识别影响因素集，设计影响因素相关性评分表，建立影响因素矩阵；第二步：采用决策实验室分析法（DEMATEL），对得到的影响矩阵进行计算，

得到综合影响矩阵，并逐步计算各影响因素的影响度、被影响度、中心度和原因度，对各影响因素的因果属性进行归类；第三步采用解释结构模型法（ISM）厘清各影响因素的层次结构和逻辑关系，分析影响高校图书馆用户阅读的本质因素、过渡因素和直接因素。

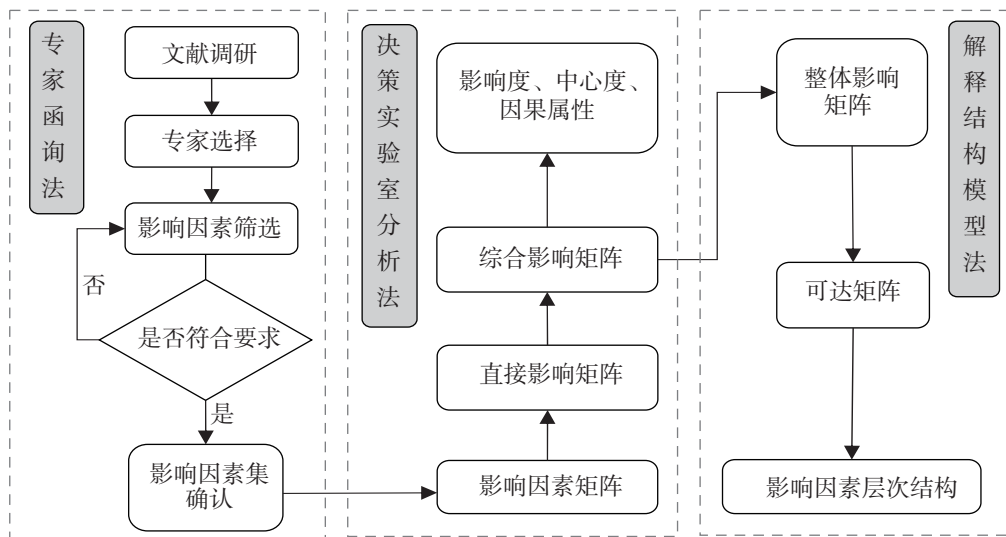


图 1 整体研究思路

2 高校图书馆用户阅读影响因素的识别

本文以中国知网（CNKI）和 SCI 科学引文索引（Web of Science 平台）为检索平台，检索出阅读影响因素的相关研究文献，剔除非核心文献后逐一阅读，对影响因素进行粗提取，共提取出 120 多个影响因素关键词。删除明显不相关因素后，对高频因素进行排序，并初步合并近义词及同义词，获得基础影响因素 38 个。通过反复函询专家，识别出影响阅读的因素。具体步骤为：首先，组织专家

对基础影响因素进行甄别，以问卷调查形式请各专家针对各影响因素进行“是”与“否”的单项选择，在回收的问卷中，将选项完全一致或选项不全的问卷作为无效问卷舍弃，基于剩余有效问卷统计各因素的“是”“否”占比情况，以“是”占比超过 70% 为选择标准，获得了 15 个被初步认可的影响因素。其次，针对 15 个初步认可的影响因素，设计第二轮函询问卷，请各专家按重要性程度由低到高对各因素分别赋 1~5 分范围内的整数分值。按前述同样方法去除无效问卷，对回收的有效问卷采用去除一个最高分和一个最低

分的平均值法对各因素的打分结果进行计算,以平均得分不小于3为筛选标准,获得了13个影响因素。这里专家对结果起决定性作用,为了提高因素识别的合理性,本文邀请了来自不同高校图书馆及相关学科的代表性专家

作为函询对象。专家的专业性、代表性和广泛性保证了多视角识别和结果的可信度。通过两轮函询最终识别出包括个人层面、高校层面、社会层面三个大类共计13个影响高校图书馆用户阅读的因素。如图2所示。

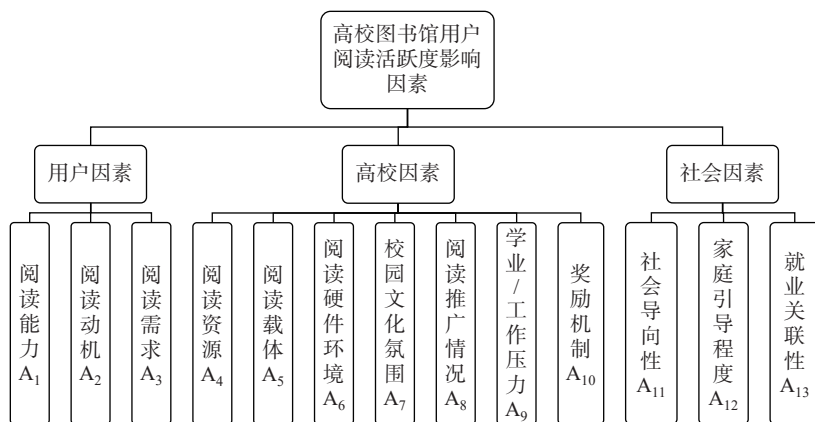


图2 高校图书馆用户阅读影响因素

用户因素包括个人阅读能力、阅读动机、阅读需求等三个因素。其中,阅读能力是指图书馆用户个体理解和吸收文字信息的能力,对阅读深度有直接影响。阅读动机是指包含阅读心理、阅读习惯、阅读兴趣等在内能够促进个体进行阅读活动的心理因素,这里主要指内在驱动因素,如为了提升自我知识面而开展的主动阅读。阅读需求是为满足外在需要而开展的阅读活动,如为了完成课程学习或开展科研工作而进行的阅读。

高校因素包含阅读资源、阅读载体、阅读硬件环境、校园文化氛围、阅读推广情况、学业/工作压力和奖励机制等7个影响因素。其中,阅读资源指高校图书馆能够给用户提供的资源种类、资源数量和资源质量等情况。阅读载体是指用户开展阅读活动需要借助及图书馆能够

为用户提供的媒介种类情况。阅读硬件环境是指图书馆阅览室等设计是否合理,用户体验感是否良好。校园文化氛围是指高校文化底蕴和传承情况。阅读推广情况是指为了促进用户阅读,阅读推广的丰富程度。学业/工作压力是指是否存在因完成学业及工作占据过多时间和精力,导致无暇进行阅读的情况。奖励机制是指对于阅读积极性高的图书馆用户,高校是否有开展相应的活动及奖励政策等。

社会因素包括社会导向性、家庭引导程度以及就业关联性这三个方面的影响因素。社会导向性是指当前社会环境及他人阅读行为对用户阅读产生的作用,家庭引导程度是指用户家庭其他人员阅读行为对用户阅读产生的影响,就业关联性是指在当下及未来就业环境中是否有对阅读的要求。

3 基于 DEMATEL 方法的高校图书馆用户阅读综合影响因素矩阵构建及计算

DEMATEL 分析法是一种借助矩阵论和图论原理分析复杂性因素间影响关系的方法^[16]。由美国布法罗纽约州立大学 (Buffalo State University) 的教授 A. Chakraborty 在 1970 年代提出。本研究采用 DEMATEL 分析法对高校图书馆用户阅读影响因素进行分析, 包括以下几个步骤:

(1) 影响因素集确定: 对前述 13 个影响因素按照 $A_1 \sim A_{13}$ 的顺序进行命名标注, 得到高校图书馆用户阅读的影响因素集 $A = \{A_1, A_2, A_3, A_4, A_5, A_6, A_7, A_8, A_9, A_{10}, A_{11}, A_{12}, A_{13}\}$ 。

(2) 影响因素相关性评分: 设计各阅读影响因素相关性评级表, 按照相关程度由低到高分别赋 1、2、3、4、5 五个等级的分数。邀请 m 位专家进行评级打分, 从而得到每位专家的阅读影响矩阵 $B^k = [b_{ij}^k]_{n \times n}$, 其中, n 是指矩阵维度,

也是因素个数, b_{ij}^k 是指第 k 位专家给出的因素 A_i 对因素 A_j 的影响程度评分, $i=j$ 时赋 0 分。

(3) 直接影响矩阵计算: 采用平均法对 m 位专家的阅读矩阵按照式 (1) 处理, 得到阅读直接影响矩阵 $C = [c_{ij}]_{n \times n}$, 计算结果如表 1 所示。

$$c_{ij} = \frac{1}{m} \sum_{k=1}^m b_{ij}^k \quad (1)$$

(4) 统一量纲处理: 采用“行和最大值”法对直接影响矩阵 C 进行规范化, 即直接影响矩阵 C 的每一个元素除以各行之和中的最大值, 得到规范化矩阵 D 。

$$D = \frac{1}{\max \sum_{j=1}^n c_{ij}} \cdot C \quad (2)$$

其中, n 为矩阵维度。

(5) 综合影响矩阵计算: 为了直观显示各因素间直接影响和间接影响关系对阅读的综合作用大小, 采用式 (3) 进行处理, 得到综合影响矩阵 E , 计算结果如表 2 所示。

$$E = D(I - D)^{-1} \quad (3)$$

其中, I 为单位矩阵。

表 1 阅读直接影响矩阵

因素	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}	A_{11}	A_{12}	A_{13}
A_1	0	3.3	3.8	1.8	1.8	1.9	1.5	1.7	3.1	1.5	2.3	3.3	2.9
A_2	3.2	0	3.6	1.9	1.5	1.8	1.7	1.5	1.3	1.6	2.1	2.3	2.8
A_3	3.3	4.2	0	2	1.9	1.8	2.1	3.2	1.2	1.1	2.1	1.8	1.9
A_4	3.6	3.9	3.1	0	3.5	1.7	2.9	3.8	1.1	1.2	2.9	3.7	1.9
A_5	3.1	3.1	3.5	3.9	0	2.1	3.3	3.1	1.3	1.2	3.1	2.5	1
A_6	2.6	3.2	3	1.5	1.2	0	3.5	2.6	1.2	1.2	2.1	1.5	1.3
A_7	3.6	3.8	3.4	2.1	3	2.9	0	4.3	1.5	1.1	3.2	1.6	1.3
A_8	4	3.9	3.1	3.6	3.5	3.1	3.8	0	1.3	1.1	1.9	1.3	1.2
A_9	1.3	3.5	4.3	1	1.2	1	3.8	3.5	0	1.2	1.3	1.2	2.8
A_{10}	2.5	4.1	3.4	2.1	1.5	3.1	3.7	2.1	3.4	0	2.1	3.4	1.9
A_{11}	3.3	3.4	2.4	2.1	2.6	2.8	2.1	1.2	1.8	2.8	0	3.1	3.2
A_{12}	4	3.5	2.9	1.2	3.1	1.2	1.3	1.3	1.2	1.1	2.8	0	2.6
A_{13}	3.9	3.8	3.8	3.1	3.8	2.3	3.7	2.6	3.2	2.7	3.2	2.8	0

表 2 阅读综合影响矩阵

因素	A ₁	A ₂	A ₃	A ₄	A ₅	A ₆	A ₇	A ₈	A ₉	A ₁₀	A ₁₁	A ₁₂	A ₁₃
A ₁	0.239	0.343	0.336	0.208	0.222	0.205	0.234	0.231	0.208	0.151	0.237	0.255	0.23
A ₂	0.29	0.237	0.305	0.194	0.197	0.187	0.217	0.206	0.153	0.141	0.214	0.215	0.21
A ₃	0.302	0.344	0.229	0.204	0.212	0.194	0.233	0.251	0.154	0.133	0.22	0.209	0.194
A ₄	0.363	0.396	0.356	0.192	0.29	0.227	0.294	0.306	0.18	0.16	0.28	0.292	0.229
A ₅	0.334	0.36	0.347	0.272	0.194	0.225	0.29	0.28	0.174	0.151	0.271	0.253	0.197
A ₆	0.271	0.307	0.285	0.181	0.185	0.141	0.254	0.228	0.145	0.128	0.209	0.19	0.17
A ₇	0.348	0.379	0.349	0.234	0.266	0.246	0.215	0.308	0.181	0.151	0.275	0.233	0.205
A ₈	0.357	0.382	0.344	0.268	0.278	0.25	0.305	0.21	0.177	0.15	0.247	0.228	0.203
A ₉	0.254	0.328	0.328	0.179	0.195	0.174	0.272	0.26	0.122	0.134	0.2	0.19	0.212
A ₁₀	0.331	0.397	0.361	0.236	0.238	0.255	0.311	0.267	0.232	0.127	0.257	0.28	0.227
A ₁₁	0.335	0.364	0.322	0.228	0.253	0.239	0.262	0.232	0.189	0.19	0.196	0.266	0.248
A ₁₂	0.315	0.325	0.296	0.183	0.237	0.177	0.211	0.204	0.154	0.133	0.235	0.165	0.21
A ₁₃	0.411	0.443	0.419	0.295	0.327	0.27	0.351	0.317	0.255	0.216	0.32	0.304	0.212

(6) 影响度 (F)、被影响度 (G)、中心度 (H)、原因度 (J) 计算：影响度 F_i 是指第 i 个因素对其他因素的综合影响程度；被影响度 G_i 是指第 i 个因素受其他因素综合影响的程度；中心度 H_i 用来衡量第 i 个因素的重要程度；原因度 J_i 用来衡量第 i 个因素的因果属性。计算结果汇总如表 3 所示。

$$F_i = \sum_{j=1}^n E_{ij}$$

(4)

$$G_i = \sum_{i=1}^n E_{ij}$$

(5)

$$H_i = F_i + G_i$$

(6)

$$J_i = F_i - G_i$$

(7)

其中， E_{ij} 是综合影响矩阵 E 中的元素， J_i 值为正数，说明第 i 个因素影响其他因素的程度大于被其他因素影响的程度，为原因因素，反之，为结果因素。

表 3 影响度、被影响度、中心度及原因度计算结果

因素	影响度	影响度排名	被影响度	被影响度排名	中心度	中心度排名	原因度	原因度排名	因果属性
A ₁	3.099	8	4.15	3	7.249	2	-1.052	11	结果因素
A ₂	2.764	12	4.607	1	7.371	1	-1.842	13	结果因素
A ₃	2.88	9	4.276	2	7.156	3	-1.396	12	结果因素
A ₄	3.563	2	2.873	9	6.436	9	0.69	3	原因因素
A ₅	3.348	6	3.094	7	6.441	8	0.254	5	原因因素
A ₆	2.694	13	2.788	10	5.482	12	-0.095	9	结果因素
A ₇	3.391	5	3.45	4	6.841	5	-0.058	8	结果因素
A ₈	3.4	4	3.299	5	6.699	6	0.1	7	原因因素
A ₉	2.848	10	2.325	12	5.173	13	0.523	4	原因因素
A ₁₀	3.52	3	1.966	13	5.486	11	1.554	1	原因因素
A ₁₁	3.324	7	3.161	6	6.485	7	0.164	6	原因因素
A ₁₂	2.846	11	3.08	8	5.925	10	-0.234	10	结果因素
A ₁₃	4.139	1	2.747	11	6.886	4	1.392	2	原因因素

从影响度指标来看,排名前三位的分别是就业关联性、阅读资源和奖励机制,这说明这三个因素对其他因素有较大影响。从被影响度指标来看,排名前三位的分别是阅读动机、阅读需求、阅读能力,说明这三个因素更容易受其他因素影响。从中心度指标来看,排名前三位的分别是阅读动机、阅读能力和阅读需求,反映了这三个因素对阅读的影响作用较大,而学业/工作压力中

心度最小,对阅读的影响力较小。在原因度方面,奖励机制、就业关联性、阅读资源等取值为正,表明这些属于原因因素,阅读动机、阅读需求等取值为负,表明这些因素属于结果因素。

(7) 因素因果关系图绘制:以中心度为横坐标,原因度为纵坐标,绘制高校图书馆用户阅读影响因素因果关系图,添加辅助线将图形四等分,如图3所示。

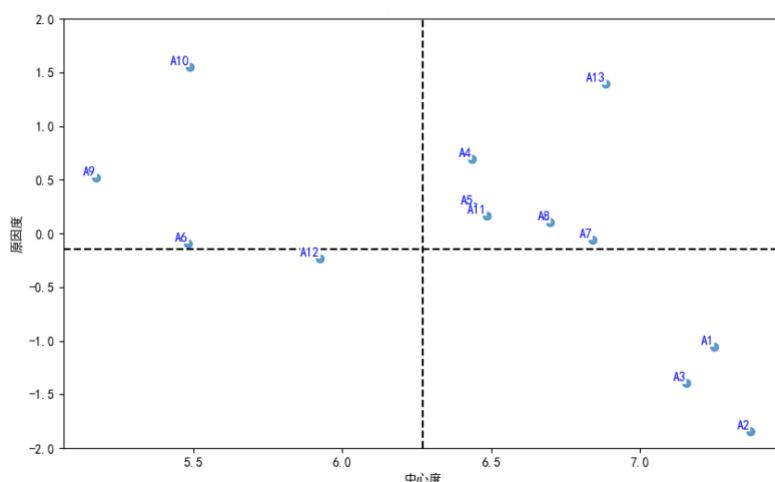


图3 因素因果关系图

图中可以看到13个影响因素在四个象限中的分布情况。其中,第一象限为原因因素,同时具有较高的中心度,本研究中有6个因素分布在此;第二象限的因素为原因因素,但中心度较低,本研究中有3个因素分布;第三象限的因素原因度及中心度均较小,相对来说不重要,本研究中家庭引导程度属于此类;第四象限的因素中心度较高,同时属于结果因素,本研究中有3个因素归于此类。

4 基于ISM方法的高校图书馆用户阅读影响因素层级模型构建

ISM (Interpretive Structural Modeling) 是一

种系统分析技术,由美国心理学家 J. Warfield 于 1971 年提出。ISM 是一种定性分析方法,主要用于理解和描述复杂系统中各元素之间的相互关系,并通过构建概念模型来揭示这些关系。本研究借助 ISM 方法,可以对影响高校图书馆用户阅读的因素进行层次划分,使得各影响因素的层级关系清晰明了。具体过程如下:

(1) 计算整体矩阵。将综合影响矩阵 E 与同阶单位矩阵 I 相加,得到整体影响矩阵 K 。

(2) 计算可达矩阵。计算综合影响矩阵 E 的均值 α 和标准差 β 及两者之和 γ 。将整体矩阵 K 中各元素值与 γ 进行比较,元素值小于 γ 的重新赋值为 0,元素值大于 γ 的赋值为 1,得

到可达矩阵 L 。本研究中, $\alpha=0.247$, 标准差为 $\beta=0.068$, $\gamma=0.315$, 可达矩阵 L 计算结果如表 4 所示。

(3) 计算可达集 M 、先行集 N 和其交集 O 。其中, $M(A_i)$ 表示可达矩阵 K 中第 i 行中所有元素值为 1 的列对应的因素集合; $N(A_i)$ 表示可达矩阵 K 中第 i 列中所有元素值为 1 的行对应的因素集合; $O(A_i)=M(A_i)\cap N(A_i)$ 。

(4) 层级机构划分。对于每一个因素 A_i , 若满足可达集 $M(A_i)=O(A_i)$, 则表示该因素是高层因素, 找出当前所有高层因素后归为同一层级。

(5) 在可达矩阵中将已找出的高层因素所在行列进行删除, 重复步骤 (3) (4), 直到将所有因素划分出层级。

(6) 根据层级划分结果绘制影响因素层级关系图, 如图 4 所示。

表 4 可达矩阵

因素	A ₁	A ₂	A ₃	A ₄	A ₅	A ₆	A ₇	A ₈	A ₉	A ₁₀	A ₁₁	A ₁₂	A ₁₃
A ₁	1	1	1	0	0	0	0	0	0	0	0	0	0
A ₂	0	1	0	0	0	0	0	0	0	0	0	0	0
A ₃	0	1	1	0	0	0	0	0	0	0	0	0	0
A ₄	1	1	1	1	0	0	0	0	0	0	0	0	0
A ₅	1	1	1	0	1	0	0	0	0	0	0	0	0
A ₆	0	0	0	0	0	1	0	0	0	0	0	0	0
A ₇	1	1	1	0	0	0	1	0	0	0	0	0	0
A ₈	1	1	1	0	0	0	0	1	0	0	0	0	0
A ₉	0	1	1	0	0	0	0	0	1	0	0	0	0
A ₁₀	1	1	1	0	0	0	0	0	0	1	0	0	0
A ₁₁	1	1	1	0	0	0	0	0	0	0	1	0	0
A ₁₂	0	1	0	0	0	0	0	0	0	0	0	1	0
A ₁₃	1	1	1	0	1	0	1	1	0	0	1	0	1

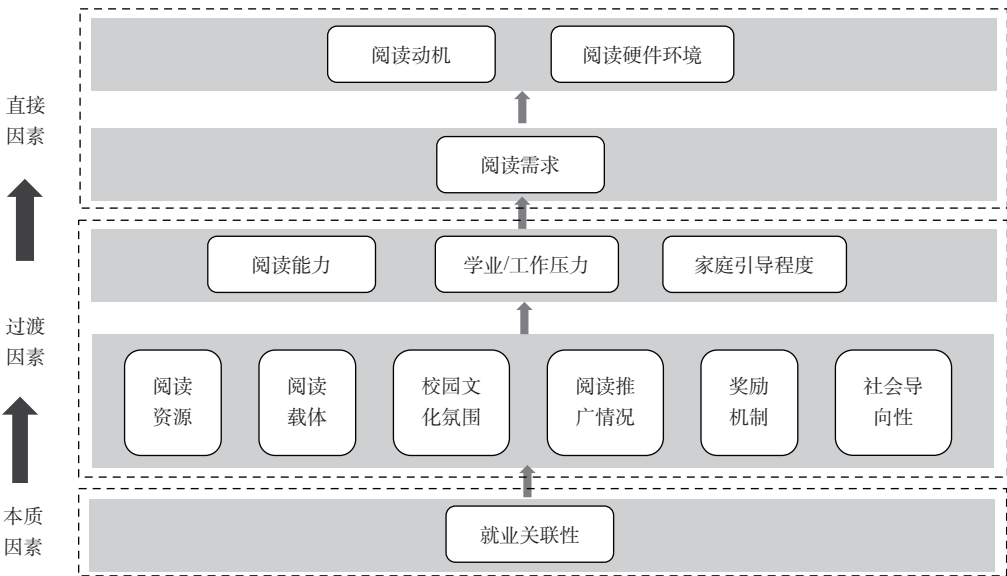


图 4 影响因素层级关系图

由图4可见,影响高校图书馆用户阅读的因素分为3层,其中,最底层的因素称为本质因素,本研究中体现为就业关联性;中间层级的因素为过渡因素,包括阅读资源、阅读载体、校园文化氛围、阅读推广情况等9项;最高层级的因素是直接因素,包括阅读需求、阅读动机、阅读硬件环境3项。

5 研究结果分析

通过利用DEMATEL分析方法得到了高校图书馆用户阅读影响因素的影响度、中心度、原因度以及中心度与原因度关系四分图,如图3所示。其中第一象限中的因素对阅读影响作用较大,且均为原因因素,应该给予重点关注,具体包括就业关联性、阅读资源、阅读载体、校园文化氛围、阅读推广情况、社会导向性。

通过ISM分析方法划分了高校图书馆用户阅读影响因素的层级关系,其中本质因素是就业关联性,过渡因素是阅读资源、阅读载体、奖励机制等,直接因素是阅读动机、阅读硬件环境和阅读需求。

针对以上计算结果,笔者分析认为,在高校图书馆用户群体中,以学生群体(包括本科生、研究生)为主体,教职人员占比较小,因此,努力完成学业以便有更好的就业机会和更高的就业平台是大多数高校图书馆用户的目标。就业将学生群体、高校和社会三者进行了联结,在社会方面,学生群体毕业后需要融入社会,必然会参与社会导向性的发展变化;在高校方面,高校在人才培养过程中会及时接收学生就业情况的反馈,从而调整培养策略,这会在阅

读资源丰富程度和偏好性、阅读载体的多样性、校园文化氛围、阅读推广力度、高校奖励机制等方面进行体现。校园阅读资源丰富度、阅读推广力度等又会影响用户阅读能力的培养,进一步影响到用户的阅读需求和阅读动机的产生。

因此,要提升高校图书馆用户阅读积极性,就要把握高校图书馆用户就业趋势,顺势而为,为其提供丰富的阅读资源、开展相关主题的阅读推广活动,设置相关的奖励机制。同时,要进一步调研图书馆用户的阅读需求和挖掘其阅读动机,从而针对性地采取刺激方法,激发图书馆用户的阅读热情。具体可尝试从以下几个方面入手:(1)高校领导者持续加大对图书馆及阅读相关配套软硬件设施的更新和投入力度,为阅读者提供更加舒适和便捷的阅读环境,激发读者的阅读需求;(2)持续开展阅读推广活动,创新活动形式、丰富活动内容,通过设置游戏闯关、打卡霸榜等多种形式的有奖竞赛提高阅读推广的覆盖面和参与度,打造充满活力的校园文化氛围;(3)高校图书馆可联合就业指导中心,在新生入学教育课程中增加阅读与就业相关部分介绍,帮助学生在其未来的职业规划中提前加入阅读计划;(4)校企联动,寻找校外知名企业开展特色主题阅读宣传活动,在吸引高校图书馆用户提升阅读积极性的同时,辐射企业文化发展,增强企业员工阅读热情。

6 结语

针对高校图书馆用户阅读影响因素繁杂的问题,本研究基于专家反馈法、DEMATEL分析和ISM分析法,识别了用户层面、高校层面

和社会层面的 13 个影响因素, 分析了 13 个因素间的影响度、原因度等指标, 并梳理了影响因素间的层次关系, 研究结果表明, 就业关联性是最本质因素, 阅读需求、阅读动机则属于直接因素。本研究可以为高校提升图书馆用户阅读积极性提供新的思路和决策支持。

参考文献

- [1] 杨军. 媒介形态变迁与阅读行为的嬗变——以印刷媒介与网络媒介为例的考察[J]. 图书馆工作与研究, 2012(2): 90-92.
- [2] 王姝, 马家伟. 全民阅读背景下高校阅读推广人培育策略研究[J]. 图书馆工作与研究, 2020(10): 115-122.
- [3] 范并思. 阅读推广与图书馆学: 基础理论问题分析[J]. 中国图书馆学报, 2014(5): 4-13.
- [4] 李倩, 郭倩倩. 基于用户类型划分的高校图书馆阅读推广策略选择研究[J]. 图书馆, 2020(6): 72-79, 100.
- [5] 王蕾. 高校图书馆阅读推广有效性的影响因素研究[J]. 大学图书馆学报, 2020, 38(2): 76-83.
- [6] 曾文. 经典文本的阅读偏好与阅读行为研究: 以图情档类本科生为例[J]. 图书情报工作, 2020, 64(22): 118-125.
- [7] 李明. 高校学生社团阅读推广工作调查与探析——以全国高校“阅读推广”十佳学生社团为例[J]. 图书馆学研究, 2020(7): 72-78.
- [8] 巫芯宇. 基于移动终端的大学生深阅读行为影响因素研究——基于扎根理论的质性分析[J]. 图书情报工作, 2021, 65(24): 80-88.
- [9] 王姝, 马家伟, 蒋振喜, 等. 高校图书馆阅读推广影响力及评价研究[J]. 图书馆, 2021(12): 74-80.
- [10] 苏文成, 于莹莹, 卢章平, 等. 高校图书馆用户学习与阅读行为的空间嗅觉影响机理研究[J]. 图书馆建设, 2024(2): 81-93.
- [11] 高凡, 欧阳娟, 吴秀明. 高校图书馆阅读推广对读者个体影响力的评价研究[J]. 大学图书馆学报, 2021, 39(6): 62-70.
- [12] 韩丽. 自我决定理论视角下高校读者阅读意愿影响因素探究[J]. 图书情报工作, 2018, 62(14): 22-28.
- [13] 刘晓莉. 高校图书馆数字阅读推广用户阅读意愿影响因素研究[J]. 图书馆工作与研究, 2022(9): 19-28.
- [14] 姚晓彤. 高校学生阅读行为类型及影响因素研究[J]. 图书馆杂志, 2021, 40(3): 49-55.
- [15] 关磊. 高校图书馆有声阅读推广效果影响因素研究[J]. 图书馆工作与研究, 2023(11): 20-31.
- [16] 白如江, 秦明艳, 张玉洁. 科研数据学术不端影响因素研究[J]. 科技进步与对策, 2023, 40(20): 110-121.



开放科学
(资源服务)
标识码
(OSID)

超级数字阅读探索研究

刘利¹ 王东雨²

1. 北京当代软科学研究所 北京 100006;
2. 临沂大学 临沂 276000

摘要: [目的/意义] 随着虚拟现实、区块链和大数据等技术的飞速发展,阅读方式也在发生着深刻变革。超级数字阅读场景作为新型的阅读方式,有望促进知识生产力的跃迁和发展。[方法/过程] 引入波普尔的三个世界理论,阐述超级数字阅读场景如何促进知识生产力跃迁和发展,并构建超级数字阅读场景模型:数字场景阅读模块、知识参考咨询模块、阅读社群交流模块和开放创作模块。用户可以更加深入地理解图书内容,增强阅读兴趣,和提高效率。[结果/结论] 在超级数字阅读场景中,读者不仅能够享受到更加直观、生动的视觉盛宴,还能通过沉浸式体验深入故事核心,与角色建立情感共鸣,从而获得更加深刻、难忘的阅读体验。这种全新的阅读方式不仅极大地提升了阅读的趣味性和吸引力,还构建了一个开放、共享的知识创作与传播平台,提升了知识的生产、整合与流通效率。

关键词: 超级数字场景; 数字阅读; 卡尔·波普尔; 三个世界理论

中图分类号: G252; G35

Research on Super Digital Reading

LIU Li¹ WANG Dongyu²

1. Beijing Institute of Contemporary Soft Science, Beijing 100006, China;
2. Linyi University, Linyi 276000, China

Abstract: [Objective/Significance] With the rapid development of technologies such as virtual reality, block chain and big data, the way of reading is also undergoing profound changes. As a new reading method, super digital reading scene is expected to promote the leap and development of knowledge productivity. [Methods/Processes] By introducing Popper's three worlds theory, this paper expounds how the super digital reading scene promotes the transition and development of knowledge productivity, and constructs the super digital reading scene model: digital scene reading module, knowledge reference consulting module, reading community communication module and open creation module. Users can understand the contents of books more deeply and improve their reading interest and efficiency. [Results/Conclusions] In the super digital reading scene, readers can not only enjoy

作者简介 刘利(1989-), 硕士, 副研究员, 主要研究方向为数字人文、图书馆、科技情报分析, E-mail: 389011848@qq.com; 王东雨(1991-), 博士, 馆员, 主要研究方向数字图书馆、数字可视化。

引用格式 刘利, 王东雨. 超级数字阅读探索研究 [J]. 情报工程, 2024, 10(6): 49-58.

a more intuitive and vivid visual feast, but also go deep into the core of the story through immersive experience and establish emotional resonance with the characters, thus obtaining a more profound and unforgettable reading experience. This brand-new reading method not only greatly enhances the interest and attraction of reading, but also builds an open and shared knowledge creation and dissemination platform, which accelerates the efficiency of knowledge production, integration and circulation.

Keywords: Super Digital Scene; Digital Reading; Karl Popper; Theory of The Three Worlds

引言

随着新一轮科技革命和产业变革的兴起和演化,以人工智能、云计算、区块链、大数据等为代表的数字技术迅猛发展,对图书行业产生了深远影响。然而,尽管数字化技术提高了图书接触率和利用率,但调查显示我国成年国民每天用于阅读的时间仍然有限,远不及使用网络消遣的时间^[1]。这种情况下,提高图书利用率、阅读知识转化率和创新率成为推动数字阅读发展的关键。因此,超级数字阅读场景作为一种新型的阅读方式应运而生,其通过深度融合知识再生产的各个环节,打破时空限制,推动知识要素的快速流动和高效匹配,有望促进知识生产力的跃迁。

1 超级数字阅读场景概述

当前,数字阅读的场景化发展正迎来一个多元化和智能化的新时代。超级数字阅读场景在未来将是一个融合了个性化、智能化和沉浸体验的全方位阅读空间。在这个空间中,用户不仅能够获得丰富多样的阅读内容,增强沉浸体验,还能够通过社交功能与他人进行深入的交流和讨论,深化阅读社交。随着技术的不断进步,数字阅读场景将更加生动和真实,不但能为用户提供更加高效、有趣和个性化的阅读

体验,还能提供知识流通和知识创作的便捷环境^[2]。

1.1 超级数字场景内涵

根据2023年11月发布的《超级数字场景源动力——游戏科技与创新应用研究》^[3]报告,超级数字场景是一个经由技术整合后对其作用对象能够形成多维场力,特别是高效服务于海量数据处理和结果生成的数字环境,是万物互联、数实融合社会新形态的交互和生产新工具。超级数字场景具有海量用户参与的鲁棒稳定、海量多维交互的即时协同、海量个体决策的复杂拟合、海量现实元素的极致仿真、海量数据处理的超级算力^[4]。

1.2 超级数字阅读场景概念和特点

超级数字阅读场景是在超级数字场景概念下结合虚拟现实、增强现实和人工智能等技术的数字阅读方式。它通过将图书内容数字化和场景化,为用户提供沉浸式、交互式、高效性和个性化的阅读体验。在数字化时代,超级数字阅读场景作为新兴的阅读形式,正逐渐改变着人们的阅读习惯^[5]。其特点主要体现在以下方面:

(1) 沉浸式:超级数字阅读场景通过虚拟现实和增强现实等场景技术体系,为用户带来

身临其境的阅读体验。这种体验使读者仿佛置身于图书所描述的世界中，有助于加深对知识内容的理解和掌握^[6]。较于传统纸质图书，超级数字阅读场景更能吸引读者的注意力，激发阅读兴趣。

（2）交互式：超级数字阅读场景不仅包含文字、图片等传统媒体形式，还可以融入音频、视频、动画等多媒体元素。这些丰富多样的内容形式有助于提升图书的表现力，使知识传递更加生动有趣。用户可以与虚拟场景中的角色、物体进行互动，参与故事情节的发展^[7]。这种交互式阅读方式提高了阅读的趣味性和参与度，使读者不再是被动的接受者，而是主动的参与者，从而更好地投入图书所描绘的世界中。

（3）高效性：超级数字阅读场景借助人工智能技术，可以辅助读者进行深入研究。AI会为读者的想法提供数据和条件支持，帮助读者快速论证新的思想和技术的可行性^[8]。这不仅

提高了知识生产的效率，也提升了知识应用的效率。

（4）个性化：超级数字阅读场景系统可以根据用户的偏好和需求，提供个性化的阅读体验。例如，用户可以选择不同的视角、调整场景亮度、添加背景音乐等。这种个性化定制使得每一段阅读旅程都独一无二，满足读者在视觉、听觉等多方面的需求^[9]。

1.3 超级数字阅读场景解析

由上述趋势和超级数字场景概念总结，超级数字阅读场景是将图书内容进行数字化、场景化和立体化呈现的阅读。超级数字阅读场景利用超级计算技术、AI技术、计算机图形学技术、虚拟现实、增强现实等技术，将书籍的内容进行数字化处理，呈现出动态、立体的场景和互动效果，让读者获得更加真实和生动的阅读体验。



图1 超级数字阅读场景效果图

如图1所示，读者直接穿戴数字阅读场景设备，进入一个由超级数字场景模拟构建出的全新世界。这个世界结合了图书内容的拆分设

计，并利用场景技术进行演绎，将一本书的知识世界呈现得栩栩如生。在这个阅读过程中，知识不再是简单的文字，而是变成了立体的、

有温度的、可交流的、可触摸的存在。在这个世界里,读者可以根据自己的喜好和兴趣选择不同的角色和场景,亲身感受到书中所描述的场景、人物、情感和思想,与书进行对话,并在书中体验另一个世界的奇妙之旅。读者在超级数字场景中进行阅读仿佛置身于游戏化的书中世界,可以尽情探索和发现、学习和体验、交流和互动,从而更好地理解 and 掌握书中的知识和思想,让阅读变得更加有趣、生动和有意义,也让知识变得更加易于理解和掌握。同时它可以激发读者的兴趣和好奇心,帮助他们更好地探索和发现知识的世界,从而更好地提高他们的阅读和学习效果。

2 打造超级数字阅读场景的理论基础和知识创新模型

2.1 卡尔·波普尔三个世界理论

三个世界理论是由英国科学哲学家卡尔·波普尔提出的,这一理论将宇宙划分为三个世界^[10]:物理世界(世界一)、主观知识世界(世界二)和客观知识世界(世界三)。世界一包括物质和能量,以及生物有机体。世界二是指人的精神和心理状态,如意识、观念和情感。世界三涉及人类精神产物,如科学理论、文学作品和技术装备等。波普尔的世界三理论指出,物理世界(世界一)、主观知识世界(世界二)和客观知识世界(世界三)都是实在且相互作用的。世界一和世界二相互影响,例如,物质需求满足精神需求,精神意志能克服物理困难。世界二和世界三也相互作用,如情感影响艺术作品

的创作。同时,世界三能反作用于世界一、世界二,科技进步改变物理世界和人们的精神世界。波普尔认为世界三是思想内容的客观世界,包括科学、诗歌和艺术作品。

国内相关学者也对波普尔的三个世界理论开展了研究,如王延屹^[11]提出客观知识是一种虚拟存在,虚拟经济源自实体经济,具有自身发展规律,并与实体经济互动。刘伟^[12]认为“世界三”的发展推动了“世界一”(物质世界)和“世界二”(精神世界)的进步,计算机等技术提高了知识组织和传播效率。赵涛^[13]指出网络作为一种低成本的知识发布系统,加速了主观知识向客观知识的转化,并促进了知识的个人和社会建构,网络的协作特性将加快客观知识的进化,例如通过互联网协作完成小说创作等。在三个世界理论下,这三位专家都强调了客观知识的重要性,以及前沿科技在促进知识进化、提高社会文明程度和认识世界能力方面的重要作用。

超级数字阅读场景系统作为世界三(客观知识的世界)的一部分,为知识的进化提供了共享、交流的平台和媒介。在这个世界中,知识通过虚拟现实技术和网络技术被转化为虚拟的形式,如视频、文字、图像、符号等,更加逼真地还原现实,可以被其他人所理解和使用。通过超级数字场景技术群建构世界三,使得知识要素在三个世界之间更加高效地传播、交流和应用,从而在不同的个体和群体之间实现共享和积累。这种共享和积累的过程加速了科学知识的开发和创新,使得科学知识得以不断进化和发展^[14]。

2.2 卡尔·波普尔三个世界知识创新模型

知识创新模型通常用于描述在特定环境或系统中知识如何被创造、传播和应用的过程。超级数字阅读场景模型是基于波普尔的三个世

界理论，构建的一个知识创新模型，该模型强调了物理世界（世界一）、主观世界（世界二）和客观世界（世界三）之间的相互作用和转化，如图2所示。

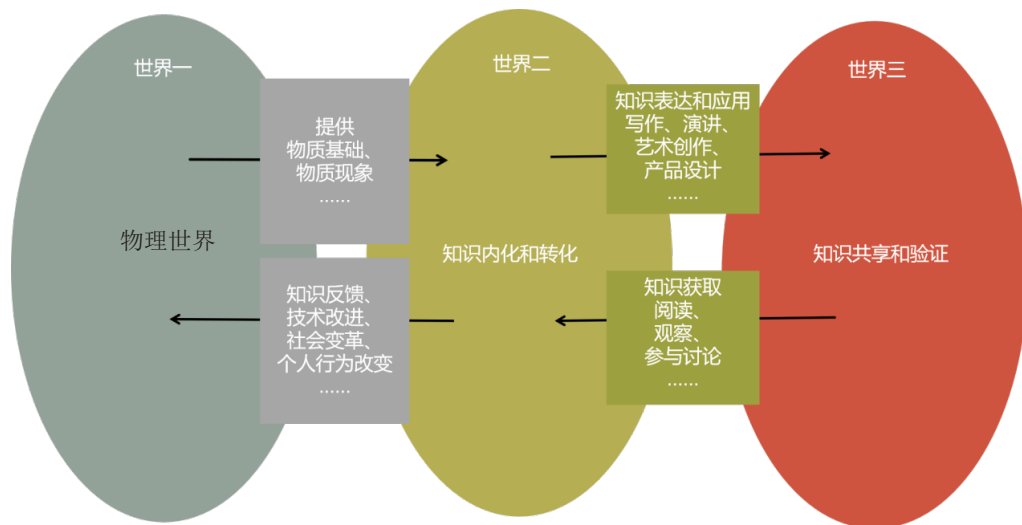


图2 波普尔三个世界知识创新模型

（1）知识获取（从世界三到世界二）

第一步涉及从世界三（客观知识世界）中获取和接触知识。这可能包括阅读书籍、文章、观察数据、参与讨论等。在这个过程中，个体通过感觉器官和思维活动，将这些客观知识内化到自己的主观世界中，形成个人的理解和认识。

（2）知识内化和转化（在世界二内部）

在世界二（主观知识世界）中，个体对获取的知识进行内化和转化。这包括理解、解释、关联、批判性思考和创新性的重新组合。这一步骤涉及个人的认知过程，如记忆、想象、直觉和逻辑推理，它们共同作用以产生新的知识见解。

（3）知识表达和应用（从世界二到世界三）

当个体在主观世界中形成新的理解或知识时，他们可以通过各种方式将这些知识表达出来，如写作、演讲、艺术创作、产品设计等。这些表达形式构成了世界三的新增内容，即新的客观知识。它们可以被其他人获取、理解和应用，从而进一步推动知识的传播和创新。

（4）知识共享和验证（在世界三内部）

在世界三中，新的知识与其他知识相互作用，通过共享、讨论、验证和批判性评估来增强其有效性和可靠性。这个过程可能涉及科学共同体、学术界、行业专家或普通公众的互动和合作，以确保知识的质量和适用性。

（5）知识反馈和循环（从世界三回到世界一和世界二）

最终，经过验证和优化的新知识会通过世

界二反馈到世界一，影响个体的主观世界和物理环境。这可能涉及技术改进、社会变革、个人行为改变等方面，从而形成一个持续的知识创新循环。

这个模型强调了知识创新是一个跨越三个世界的动态过程，涉及知识的获取、内化、表达、共享和反馈。在每个步骤中，世界一、世界二和世界三都相互关联和相互作用，共同推动知识的不断发展和创新。

3 超级数字阅读场景模型架构设计

根据卡尔·波普尔知识创新模型，超级数字阅读场景系统应该具有沉浸式阅读功能、知识交互功能、知识交流功能、知识创作功能，对应的超级数字阅读场景系统应该包含数字场景阅读模块、知识参考咨询模块、阅读社群交流模块、开放创作模块、如图3所示。

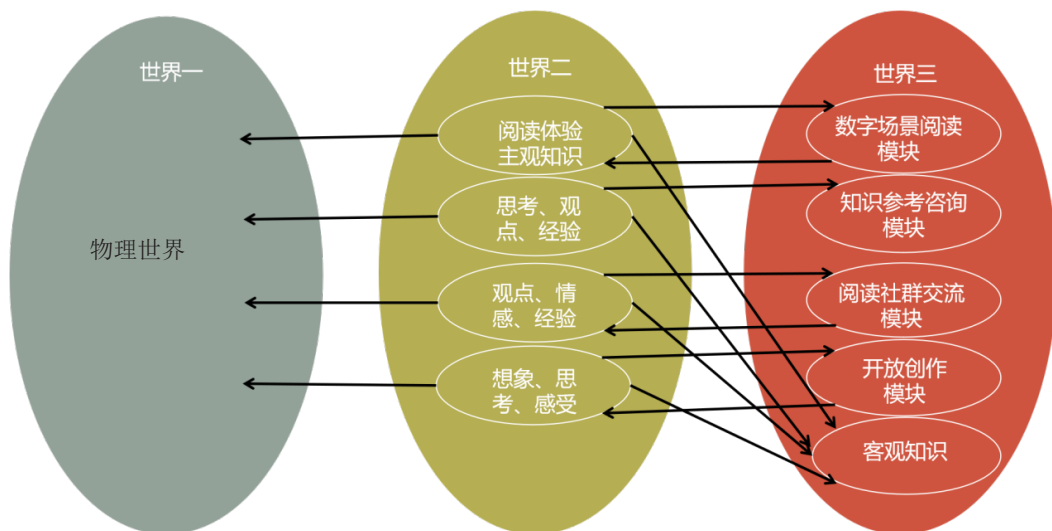


图3 超级数字阅读场景架构设计

3.1 数字场景阅读模块

数字场景阅读模块中的数字化内容本身则属于第三世界，即客观知识世界。这些数字化内容是人类精神活动的产物，包括文字、图像、音频等各种形式的信息。它们以数字化的形式存储和传输，可以在不同的设备和平台上进行访问和共享。这些数字化内容不仅具有客观性和可传承性，还能够通过用户的互动和反馈不断得到更新和完善。

用户的阅读行为和体验则属于世界二，即主观知识世界。用户在阅读过程中会产生个人的理解、感受和思考，这些主观经验构成了他们独特的阅读体验。同时，用户的个人背景、兴趣爱好和认知结构等因素也会影响他们的阅读选择和理解方式。

用户通过物理世界的数字阅读设备访问和获取世界三的数字化内容，这些内容在用户的世界二中引发主观的思考和体验。用户可以根据自己的需求和兴趣对这些内容进行筛选、整

合和创新性的解读,从而形成新的知识和理解。这些新的知识和理解又可以以数字化的形式反馈到系统中,成为世界三新的客观知识,供其他用户访问和共享。

3.2 知识参考咨询模块

知识参考咨询模块的基础技术架构是一个大型的语言模型,它是基于大量的文本数据训练而成的。这些数据来自世界一中的物理对象——书籍、文章、网页等。然而,关于这本书的语言模型不仅仅是对这些数据的简单复制。它通过对这些数据的模式和关系进行学习和理解,生成新的、有意义的文本。这种生成的文本是主观的,因为它反映了模型对给定问题的理解和解释。但同时,它也是客观的,因为它基于大量的数据和一定的算法生成的,而不是基于某个特定个体的主观经验或意识状态。因此,知识参考系统可以被视为世界三的一部分。它不仅仅是一个工具或技术,更是思想内容的生成器。它可以帮助我们理解复杂的问题、生成新的观点,甚至在某些情况下,可以作为我们的对话伙伴,帮助我们思考和表达。

3.3 阅读社群交流模块

社群交流模块在世界三中表现为共享的知识、文化和规范。在社群交流模块中,成员们通过文字、语音、图像等方式表达自己的内心世界,分享观点、情感和经验。这种精神层面的交流是社群互动的核心(世界二),它促进了成员之间的相互理解和认同。社群成员在交流过程中不断创造和积累新的知识,形成共同的文化和价值观。这些客观知识反过来又影响

成员们的精神世界和交流行为,塑造着社群的独特氛围和特征。

3.4 开放创作模块

世界三在开放创作模块中指创作者们通过共享和交流所形成的共同的知识和文化遗产。在这个世界里,创作者通过想象、思考、感受等精神活动(世界二),将这些物理素材转化为具有个性和创意的作品。这些作品可以是文字、图像、音频、视频等多种形式,它们构成了开放创作系统的主要内容。在这个世界中,创作者们相互学习、相互启发,共同推动创作水平的提高和文化的繁荣。

4 超级数字阅读场景实现路径分析

根据超级数字阅读场景架构设计,超级数字阅读系统共分为四个部分:数字场景阅读系统、知识参考咨询系统、阅读社群交流系统、开放创作系统。而超级数字阅读场景的实现需要首先对图书内容进行分析设计,然后利用场景技术,根据图书内容和场景系统功能进行设计和渲染,最后制作完成一部部超级数字阅读场景,如图4所示。

4.1 图书内容策划与准备

内容是场景的核心,它决定了场景的主题、情节和角色等核心要素。所以首先需要确定图书的主题和内容,对图书内容进行拆分,确定好人物、情节、线索以及图书风格,分析图书内容以便构建数字场景。创建故事板或概念图,明确每个数字场景的目标和互动点。然后根据场景设计制作多媒体素材,如音频、视频、图像等。

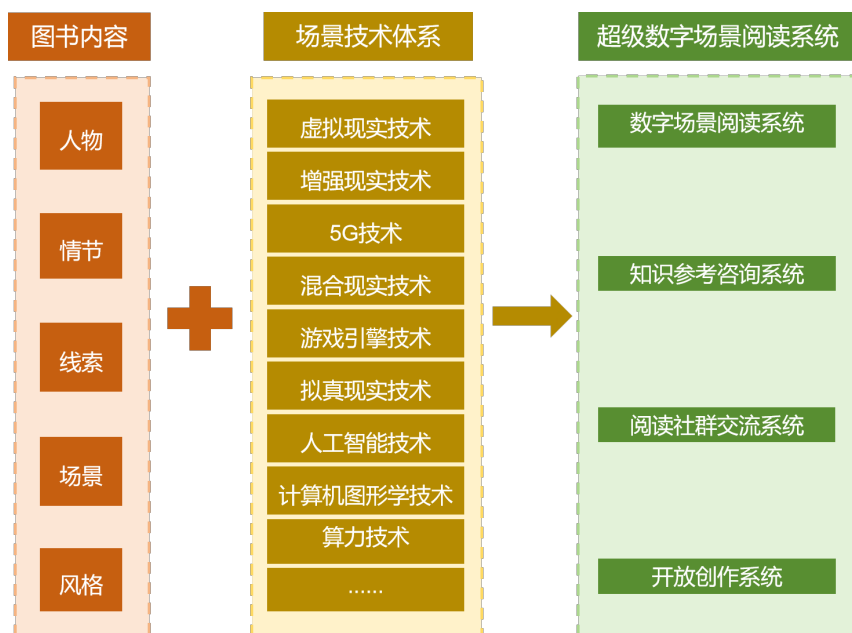


图4 超级数字阅读场景架构

超级数字场景所要实现的目标是打造一种高度沉浸式的阅读体验，使读者能够更深入地参与和体验知识。图书内容作为场景的核心，通过故事情节、角色塑造等手段吸引读者的情感投入；而数字场景则通过技术手段营造出逼真的环境和氛围，使读者仿佛置身于故事的世界中。内容与场景的有机结合共同构建了这种沉浸式的阅读体验。

4.2 场景技术体系集成

构建超级数字阅读场景系统需要综合运用多项技术，以确保用户的沉浸式、交互式、高效性阅读体验和系统的稳定性和安全性运行。这些技术包括大数据分析、人工智能、云计算、用户界面与交互式设计技术、移动设备与跨平台技术、安全性与隐私保护技术，以及阅读内容管理与版权保护技术等。其中最核心的场景技术包括虚拟现实、增强现实、混合现实、拟

真现实技术、人工智能技术、计算机图形学、游戏引擎技术、算力技术等。

4.3 超级数字阅读场景系统

超级数字阅读场景系统是一个将超级数字场景技术应用于图书领域的系统。该系统通过利用先进的图形渲染、物理模拟、人工智能等技术，将传统的图书内容转化为高度逼真的三维虚拟场景，为用户提供沉浸式的阅读体验和开放式的创作环境。

在超级数字阅读场景系统中，用户可以通过虚拟现实（VR）设备或增强现实（AR）设备进入虚拟图书世界，在数字场景阅读系统内与图书中的角色、场景和物体进行互动，在虚拟场景中自由探索、阅读、互动，甚至参与故事情节的发展。在知识参考咨询系统内可以选择图书的任意知识点向AI大模型进行咨询学习相关纵深知识^[15]。在阅读社群交流系统内可以

和相关兴趣读者交流，通过讨论交换观点激发灵感。在开放创作系统内根据图书内容自动生成三维场景，或者通过扫描实体图书来创建虚拟场景，或者进行论文的写作和标引。

超级数字阅读场景系统的发展潜力巨大，它可以为出版业、图书馆、教育机构等带来创新性的变革。通过该系统，用户可以更加深入地理解图书内容，提高阅读兴趣和效率。同时，该系统也为图书创作者提供了更广阔的创作空间，可以创作出更加丰富、多样的知识作品。

5 超级数字阅读场景研究结论与展望

先进数字技术作为时代进步的强大引擎，正以前所未有的力量推动着图书出版行业的深刻变革，催生了超级数字场景图书这一创新形态的出现与蓬勃发展。这类图书通过深度融合虚拟现实(VR)、增强现实(AR)、混合现实(MR)乃至计算机图形学等尖端技术，彻底颠覆了传统图书的静态呈现模式，实现了信息的动态化、多维度传递，以及用户与内容之间前所未有的深度互动。

在超级数字阅读场景中，读者不仅能够享受到更加直观、生动的视觉盛宴，还能通过沉浸式体验深入故事核心，与角色建立情感共鸣，从而获得更加深刻、难忘的阅读体验。这种全新的阅读方式不仅极大地提升了阅读的趣味性和吸引力，还构建了一个开放、共享的知识创作与传播平台，提升了知识的生产、整合与流通效率。知识创作者能够利用这些先进技术，以更加直观、形象的方式表达思想，激发灵感，

促进知识的创新与发展。

超级数字阅读场景的出现，不仅提升了知识生产的效率，还促进了数字场景技术集群的广泛应用与深入研发。它推动了人工智能、大数据、云计算等智能技术的交叉融合与迭代升级，为科技产业的持续进步注入了新的活力。然而，尽管前景广阔，超级数字阅读场景的研发与应用目前仍处于探索阶段，面临着诸多挑战与难题。

首先，资金短缺是制约其快速发展的关键因素之一。超级数字场景图书的制作成本高昂，涉及复杂的技术研发、内容创作与市场推广等多个环节，需要大量资金投入。以《黑神话：悟空》等3A级游戏项目为例，其巨额投入显示了技术实现的难度与成本。因此，如何降低成本、提高性价比，成为当前亟待解决的问题。其次，人才短缺也是制约超级数字阅读场景发展的瓶颈之一。随着技术的不断进步，对具备跨学科知识背景与创新能力的人才需求日益迫切。然而，目前我国在沉浸式技术人才的培养方面还存在数量不足、素质参差不齐等问题，人才培养体系尚需进一步完善。再者，开发机制的不完善也限制了超级数字场景图书的规模化开发与商业化应用。当前，国内尚未形成一套成熟的市场合作机制与商业模式，导致市场发展空间受限，难以吸引更多资本与资源的投入。因此，相关部门与领域需加强合作，共同探索适合我国国情的开发模式与市场机制。最后，法律法规的滞后也是制约超级数字阅读场景发展的重要因素之一。随着技术的快速发展与应用，知识产权、数据安全、隐私保护等问题日益凸显。然而，现有的法律法规体系尚不能完全适应新

技术的发展需求,需要进一步健全与完善。特别是针对超级数字场景图书及开放创作系统的知识产权问题,需要明确保护范围与责任归属,保障知识开发、创作与流通的有序进行。

超级数字阅读场景作为未来图书出版行业的重要发展方向之一,其研发与应用具有广阔的前景与潜力。然而,要实现其规模化、商业化发展仍需克服诸多挑战与难题。通过加强技术研发、人才培养、机制创新与法规完善等多方面的努力与探索,我们有理由相信超级数字阅读场景将在不久的将来迎来更加辉煌的篇章。

参考文献

- [1] 韩婧,于准,马春林.全媒体背景下国民阅读趋势及对策研究——基于第十五次至第十九次全国国民阅读调查报告分析[J].图书馆学刊,2023,45(8): 59-64.
- [2] 张建友,卜彦芳.可供性视角下元宇宙阅读场景的建构[J].出版广角,2022(18): 44-50.
- [3] 游戏科技创新研究项目组.游戏科技与超级数字场景的跨行业溢出效应研究——基于《超级数字场景源动力——游戏科技与创新应用研究》报告数据[J].中国数字出版,2024,2(1): 49-59.
- [4] 赵红勋,侯珮桦.重新部落化:数字时代社会化阅读场景建构及路径重塑[J].中国出版,2023(24): 3-9.
- [5] 张建友,卜彦芳.场景可供性视角下数字阅读发展研究[J].编辑之友,2023(5): 39-45.
- [6] 郑弘,丰树琪,谭言.科普节目《共上一堂课》全场景阅读模式的实践探索[J].新阅读,2023(7): 61-63.
- [7] 林凌.场景式阅读的结构特征及其发展前景研究[D].广州:广东财经大学,2021.
- [8] 许哲,徐和祥.全景式阅读场景下数据驱动的学生阅读素养评价的内涵、方法及实现路径[J].电化教育研究,2022,43(10): 79-85.
- [9] 宫蛟飞,宫雅琦.数字图书馆沉浸式阅读项目建设探讨——以国家图书馆“沉浸式阅读体验”项目为例[J].河南图书馆学刊,2024,44(1): 78-80.
- [10] 卡尔·波普尔.《客观知识:一个进化论的研究》[M].上海:上海译文出版社,1987: 98-102.
- [11] 王延屹.波普尔“三个世界”理论在中国虚拟经济中的作用研究[D].大连:东北财经大学,2017.
- [12] 刘伟.波普尔的“三个世界”理论对构建图书馆哲学的思考[J].内蒙古科技与经济,2012(21): 138-139.
- [13] 赵涛.电子网络与知识生产——基于波普尔“三个世界”理论视角的考察[J].学术界,2013(10): 74-84, 308.
- [14] 李荣坤.数字技术为文化破圈注入新活力[N].中国文化报,2023-06-30(002).
- [15] 李鹏,宋西贵.AIGC技术赋能图书馆阅读推广工作的创新应用[J].农业图书情报学报,2023,35(12): 84-93.



开放科学
(资源服务)
标识码
(OSID)

大学生网络社会情绪表达策略： 基于事件文本挖掘研究

刘杰

中国石油大学（华东）海洋与空间信息学院 青岛 266580

摘要： [目的/意义] 针对热点事件中大学生网民在线评论具有的主题概念模糊、内容信息分散、情绪表达多样等特点，利用文本挖掘技术手段，开展事件刺激下大学生网络社会情绪影响因素及情绪表达策略研究，对高校育人工作者提升网络舆情识别精度、增强网络舆论引导能力具有重要意义。 [方法/过程] 搭建了大学生网络社会情绪表达“语义特征-情感识别-生成机理-表达策略”研究框架。以大学生权益诉求事件为例，通过 BERTopic 主题聚类、深层语义分析以及 BiLSTM 情感识别，实现对热点事件中大学生网络社会情绪表达策略的精细刻画。 [局限] 热点事件的范畴应包括负向、中性、正向事件性质，研究仅选取了一例权益维护的负向事件进行了框架分析，数据样本量较小，情绪表达的影响因素及内在路径仍具有较大探讨空间。 [结果/结论] 基于 SOR 理论模型等，揭示了热点事件中大学生网络社会情绪表达生成机理，并进一步提出了“事件刺激→个体感知、事件归责与权益权衡→权益诉求”网络社会情绪表达策略具体路径，为高校舆论引领及信息化管理工作提供一种分析研判的工作框架、研究范式及对策启示。

关键词： 网络社会情绪；语义特征；情感分析；表达策略；大学生

中图分类号： G35; TP391

Strategies for College Students' Online Social Emotional Expression: A Study Based on Event Text Mining

LIU Jie

College of Oceanography and Space Informatics, China University of Petroleum (East China), Qingdao 266580, China

Abstract: [Purpose/Significance] In response to the characteristics of vague thematic concepts, scattered content information, and diverse emotional expressions in online comments by college students during hot events, text mining technology is used

基金项目 2024 年度教育部人文社会科学研究专项任务项目（高校辅导员研究）“经济社会热点事件情境下大学生积极网络社会心态培育研究”（24JDSZ3087）；2023 年山东省社会科学规划学校思想政治教育（全环境立德树人）研究专项“热点事件中大学生网络社会心态影响因素及其引导策略研究”（23CSZJ39）。

作者简介 刘杰（1987-），博士，讲师，主要研究方向为用户行为、网络社会心态，E-mail: liujie_upc@163.com。

引用格式 刘杰. 大学生网络社会情绪表达策略：基于事件文本挖掘研究 [J]. 情报工程, 2024, 10(5): 59-70.

to conduct research on the influencing factors and emotional expression strategies of college students' online social emotions under event stimulation. This is of great significance for university education workers to improve the accuracy of online public opinion recognition and enhance their ability to guide online public opinion. [Methods/Processes] We have established a research framework for the expression of college students' online social emotions, which includes semantic features, emotion recognition, generation mechanisms, and expression strategies. Taking the issue of college students' rights demands as an example, this study uses BERTopic topic clustering, deep semantic analysis, and BiLSTM emotion recognition to achieve a detailed characterization of college students' online social emotion expression strategies in hot events. [Limitations] The category of hot events should include negative, neutral, and positive event properties. This study only selected one negative event related to rights protection for framework analysis, with a small sample size. There is still a lot of room for exploration on the influencing factors and internal paths of emotional expression. [Results/Conclusions] Based on the SOR theory model, the generation mechanism of college students' online social emotional expression in hot events has been revealed, namely "event stimulation → individual perception, event attribution and rights balance → rights and interests demands". All these provided an analytical framework, research paradigm, and policy recommendations for public opinion guidance and information management in universities.

Keywords: Online Social Emotions; Semantic Features; Sentiment Analysis; Expression Strategy; College Student

引言

网络媒体时代,网络社会情绪是网络社会心态的重要组成部分^[1-2],也是网络社会舆情产生的基础,同时网络社会情绪的爆发会造成网络社会心态的失衡^[3]。2000年后的大学生群体,被称为网络原住民,热点事件发生作为外部刺激,会引起大学生网络社会情绪波动,甚至对网络社会心态产生显著影响^[4]。前人采用质性分析与问卷调查等方式针对网络社会情绪全貌进行了大量卓有成效的研究^[5-8],但是,当前热点事件一旦出现,其时间周期短但网络评论数据量巨大,传统的人工处理方式已不能满足当前大规模数据的处理需求^[9],即无法明确某段时间大学生网络社会情绪表达全貌。因此,有必要针对大学生围绕某热点话题的评论,采用计算机建模定量方式进行文本数据采集和挖掘,以便及时、准确、高效地为政府、高校、社区、家庭等针对大学生网络社会情绪和行为倾向提供政策制定和情绪疏导的数据支撑和理论依据。

本研究选取了高校日常教育管理中涉及的大学生权益维护事件,作为热点事件的典型案例进行分析,通过文本挖掘重点开展热点事件中大学生网络社会情绪表达策略研究。

1 相关工作

情绪研究起源于心理学科,后波及人文学科和社会科学的全部领域^[10]。个体是情绪的基本载体,由于社会网络关系产生了人物、地点、时间、空间、事件所集成而来的各种情绪体验与情绪感知^[11],因此产生了社会情绪的概念。21世纪互联网被大范围普及后,网民在网上的情感交流日益增多,并引起学界的关注,“网络情绪符号”一词被首次提出。此后,网络情绪多用来指个体、群体或者社会大众在网络空间中体现出的情绪体验与反应。

突发事件的刺激唤醒了个体情绪,社交媒体的智能化使个体情绪演变为社会情绪。缺乏安全感、公平感的社会情绪,会联动触发对人

身安全、公民权利、生存环境、经济地位甚至国家信心的负面态度^[12]；同时，其传播速度异常迅速，具有“社会放大效应”，继而引发具有高度复杂性的政治、社会和文化问题^[13]。显然，情绪成为穿插于事件之中的重要效能因素——谁能有效沟通情感，谁就能更大程度地化解风险^[14]。由此，探究突发事件中网络社会情绪的影响因素及生成机理，分析其背后隐藏的社会问题，有助于疏导反向社会情绪积聚^[15]、强化正向社会情绪共鸣^[16]。前人已运用多个案定性比较分析等方式^[15, 17-20]，对突发公共事件中网民社会情绪的影响因素及其机理进行了探讨，为网络社会情绪疏导提供了有力的理论支撑；此外，大学生作为网络社会的重要群体^[21]，其情绪性集群行为^[22]及个体权益表达与管理模式冲突下的负向网络情绪表达^[21]等现实问题仍值得高度关注。

综上所述，前期虽然已有研究关注网络社会情绪的影响因素及生成机理，但是研究方法以主观测量或模糊定性等截面分析方式为主，缺乏大规模、实时动态、客观数据支撑的精准研究。因此，本文着重利用大数据采集、文本评论挖掘分析等方式开展网络社会情绪的影响因素、生成机理及表达策略分析，特别是为把握事件情境下网络社会情绪波动特点并进行有效疏导提供理论依据。

2 研究设计

2.1 研究框架

本次研究以某高校宿舍调整引发大学生维权事件为例，构建网络社会情绪表达策略分析

模型。研究框架主要包括语义特征、情感识别、生成机理、表达策略4个部分。首先，利用BERTopic模型提取评论主题特征，据此判断促发大学生产生网络社会情绪的要素和关注着力点；利用深层语义分析，进一步明确关注点之间的内在联系，明确事件刺激下大学生网络社会情绪的影响因素及情绪发展路径；其次，利用深度学习模型BiLSTM进行不同主题网络社会情绪特征识别提取和正反向语义特征探究；最后，基于SOR等理论模型，分析研判高校权益维护事件中大学生网络社会情绪生成机理及表达策略。结合上述研究，提出针对性的政策建议。

2.2 理论模型

(1) SOR 理论模型

SOR模型也就是刺激—机体—反应模型^[23]，其前身为S-R模型，早先用来探究外部环境因素对个体行为的影响^[24]，但是忽视了人的心理活动与主观能动性；直到1974年Mehrabian等^[25]在S-R的基础上添加O（机体心理状态）构建了S-O-R模型，该模型认为外部的环境刺激个体后会激发其内部的情绪认知变化，从而产生外在的行为反应^[26]，这表明个体的认知情绪等心理状态在从刺激到反应的过程中起着中介作用^[27]。

本文以“S-O-R模型”为基础，“S”（stimulus）指刺激大学生产生网络社会情绪的外部事件，“O”（organism）指大学生在事件刺激下机体产生的认知、感知等心理反应过程，继而引发情绪情感波动变化，受到多重因素持续作用，情绪变化形成一定的情感倾向和表达

意愿,利用意愿外化为网络社会情绪的表达策略 R (response)。

(2) 第一人效应和第三人效应

1983 年,菲利普·戴维森 (Phillips Davison) 在《传播中第三人效应的作用》一文中提出了一个理论假说——第三人效应理论,即人们认为媒介讯息对他人在态度、行为层面产生的影响小于这种讯息对自己产生的影响^[28]。诸如这事轮不到我或别人更严重的这类想法,正是第三人效应理论的直接表现之一^[29],具体理解可以分为认知和行为两个层面。研究发现在负面或争议性的信息中大多数人都会出现“第三人效果”,但当信息转向为符合社会期望的、符合需要的或正能量性质的时候,大多数的人们又会出现一种“第一人效果”(或者称之为“反第三人效果”),即认为自己所受到媒介信息的影响要远大于他人受到的影响。

本文探讨事件刺激下,大学生网民网络社会情绪传播中会出现大量有亲身经历或围观群众的“感同身受”,并以此为契机,产生大量评论、分享以宣泄情绪,因此初步推断网络社会情绪的弥散性传播、权益诉求表达均与第三人效应有关;另外,在实际事件情境中,部分网民会代入第一人角色,认为这些事情会发生在自己身上,因而表达出更加激烈的情绪,这应与第一人效应有关。

2.3 研究方法

(1) BERTopic 模型主题聚类

BERTopic 模型是基于 BERT 模型提出的一种完整的文档主题聚类方法^[30],是一种主题聚类无监督深度学习模型。该模型的优点在于其

为各种语言模型提供使用方案的同时,不需要预先设置大部分的模型超参数,模型在训练过程中可以自行动态调整,同时给出最优聚类结果。该模型在其他相似模型中具有更为出色的表现^[31]。

本文收集了微博平台一定时间范围内围绕事件进行的评论数据,引入 BERTopic 模型,且样本以短文本为主,覆盖事件讨论全周期,代表性较强;研究希望通过对评论数据进行聚类分析,明确大学生网络社会情绪产生的着力点。

(2) 深层语义分析

深层语义分析恰能实现评论文本的深层次分析和解读,该方法在词频分析的基础之上,将焦点落于词频与词频间的关联和意义,能够实现词条语义的深度解构^[32]。可以通过节点、关系链对人、物、动作、关系、属性等要素加以模型化,用以形式化表达物理量、心理量、条件、目标等要素的关系和逻辑表述^[33]。

本文将大学生网络社会情绪评论文本进行分词后,将内容聚合到语义网络中,如果两个关键词同时出现在一段文本中,则视为存在一次共现关系。共现次数越多,则关键词之间的联系更为紧密。以此将各个词语之间的关联规则以语义网络的形式进行可视化展现,明确词语之间的联系以及文本的深层次结构。

(3) BiLSTM 模型情感识别

BiLSTM 情感识别研究。考虑到作为自然语言处理的常用模型 LSTM,其运算是单向的,但在实际语句编码中一个词的含义通常需要结合上下文信息进行判断,因而,组合形成双向的长短期记忆网络 (BiLSTM) 模型^[34],其包含更丰富的信息,能取得更好的网络情绪评价

结果。

本文利用 BiLSTM 情感识别模型，对大学生网络社会情绪评论文本各主题进行情感倾向分析，明确随时间演化的情感发展特征，以及对不同主题下的正负情绪进行识别分析。

2.4 数据收集与预处理

本文将某高校毕业生宿舍调整至 10 人间引

发大学生大规模网络评论作为高校大学生权益维护领域代表案例进行研究，研究数据来源于微博社交平台，以“** 高校毕业生宿舍调整”作为搜索关键字，采集 8 月 17 日—9 月 18 日的热门评论数据，具体包括数据 2233 条。经过数据预处理，删除重复和无效评论后，最终共得到 2053 条微博数据作为实验的语料库。部分数据样本如表 1 所示。

表 1 数据样本示例

情感极性	微博评论
积极（1）	96届老学姐告诉你，***还真是个读书的好地方，当年化学系英语系的快乐时光
消极（2）	凭什么逼着在考研考公的大四搬四人寝，大四不是你们决策失误的牺牲品，19级大一能住差宿舍为什么22级不行，凭什么又让我们牺牲

数据预处理具体过程包括对收集来的语料文本按时间进行归类，进行数据清洗筛选掉无关文本，去除特殊符号、空格和无用微博表情等重复值和缺失值；进行中文分词、去停用词等，并利用 Word2vec 工具训练词向量模型；同时针对高校宿舍调整权益维护语料进行情感词性标注。

3 结果与讨论

3.1 事件评论文本的特征分析

（1）事件评论文本的 BERTopic 主题分布

对事件评论文本进行 BERTopic 模型聚类分析后，提取出 3 个相关性较高主题，主题内容如表 2 所示。从表 2 可以看出相关词汇均分布在同一主题，说明评论文本 BERTopic 模型聚类效果较好，并以此作为文档分类的支撑依据。同时，基于“社会心态”和“网络社会心态”的文献调研及理论模型的建构，经过整理，3 个主题最终确定划分成大学生网络社会情绪表达的三个维度，即事件性质与权益权衡、救援响应诉求、媒体报道诉求等 3 个主题，如图 1 所示。通过 BERTopic 模型聚类分析，初步获知了大学生网络社会情绪的关注点。

表 2 事件评论文本的 BERTopic 主题分布

主题编码	前10个高频词	文档数
T1	学校，学生，宿舍，大学，大四，没有，十人间，不是，真的，扩招	1904
T2	中国 消防，消防，中国，消防 中国，消防 消防，消防 ***，*** 消防，消防 在线，消防 反应，在线 中国	72
T3	人民日报，共青团中央，观察者，新闻周刊，中国 新闻周刊，新闻，东方网，观察者 新华网，新华网，央视	18

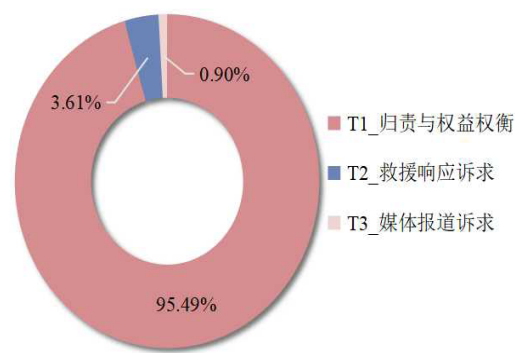


图1 事件评论文本的BERTopic主题占比分布图

(2) 事件评论文本的深层语义分析

在BERTopic模型聚类分析后，初步获知了大学生网络社会情绪的关注点，虽然避免了主

观臆断对于大学生情绪关注点和行为的分析，但是在主题内部，语义之间的关联性上仍然没有较高表现。而深层语义分析恰能实现评论文本的深层次分析和解读，该方法在词频分析的基础之上，将焦点落于词频与词频间的关联和意义，能够实现词条语义的深度解构^[32]。故本研究引入学术点滴软件的社会网络和语义网络分析功能对汇总评论进行分析，得共词矩阵（见表3），之后又将构建完成的共词矩阵导入拥有可视化功能的Gephi 0.1.0，进而生成了深层语义分析网络图，如图2所示。

表3 共词矩阵(基于评论数据)

频次	学校	学生	大学	宿舍	**地区	消防	没有	**省	扩招	寝室	...
学校		68	19	36	1	2	26	4	24	13	...
学生	68		23	26	14	4	25	5	18	1	...
大学	19	23		23	56	2	17	6	9	7	...
宿舍	36	26	23		8	2	9	4	19	6	...
**地区	1	14	56	8		4	9	26	5	2	...
消防	2	4	2	2	4		2	5	1		...
没有	26	25	17	9	9	2		2	9	7	...
**省	4	5	6	4	26	5	2		2		...
扩招	24	18	9	19	5	1	9	2		5	...
寝室	13	1	7	6	2		7		5		...
...	...	...	...	...	...	...	...	...	...	...	...

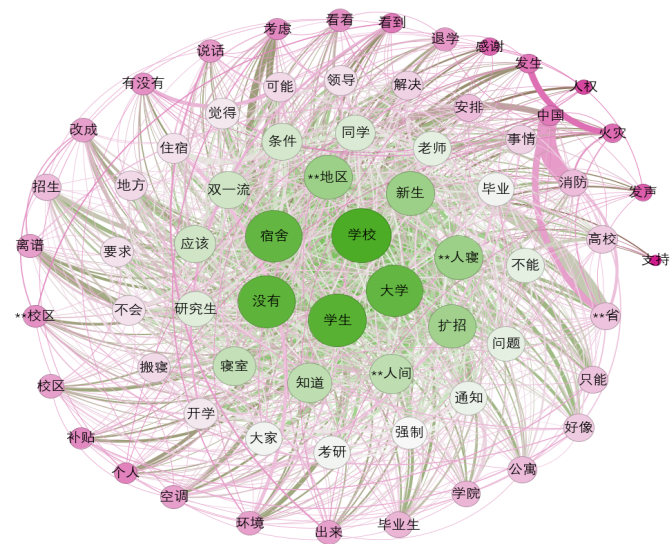


图2 事件评论文本的深层语义分析网络图

从词语关联来看，基于点度中心性，节点与其他节点相连越多，节点的圆圈就越大。颜色表示聚类，同一颜色是同一类。线条越粗，代表两点之间的共现频次越高。从层次划分来看，根据“核心—边缘”的层次结构，可将深层语义分析网络图划分为“核心—次核心—过渡—边缘”四层结构^[33]。由此划分出四层，即第一层由“学校”“大学”“学生”“宿舍”“没有”构成，主要是事件性质的描述，这也构成了情境化的刺激源头，是学生网络社会情绪展开的源头和核心利益关注点。第二层包括“**地区”“新生”“**人寝”“扩招”“**人间”“知道”“寝室”“研究生”“应该”“双一流”“条件”“同学”“老师”“毕业”“问题”“不能”“通知”等核心词汇，围绕着个体感知、权益权衡、事件归责等进一步展开，是对事件刺激认知体验的进一步展现，如对于扩招需求和住宿条件供给不匹配的政策方面的关注和思考。第三层由“强制”“考研”“大家”“开学”“搬寝”“不会”“要求”“地方”“住宿”“觉得”“可能”“领导”“解决”“安排”“事情”等核心词汇构成，进一步展现权益权衡和事件归责后的具体关注

点，如考研搬寝面临的实际困难。第四层包含“中国”“消防”“高校”“空调”“补贴”“招生”“考虑”“人权”“退学”等词汇，表达了学生对于毕业前搬寝的政策不满以及面对实际情况的无能为力，提出了利益诉求和政策建议。

综合上述研究，结合聚类分析、深层语义分析，以及SOR（刺激—机体—反应）理论的基本结构进行综合分析。事件性质作为大学生网络社会情绪的刺激源，引起了大学生群体的机体意识和集体感知；在网络社会情绪的表达中透露出个体感知、权益权衡和事件归责的指向；情绪表达引起的反应或行为倾向为向外进行权益诉求表达，这包括对于媒体媒介求助的诉求、消防安全的诉求，以及对于政策调整的期盼和建议。

3.2 事件评论文本的情感分析

（1）事件评论文本的情感演化特征

通过对时间进行切片划分，利用BiLSTM模型能够捕捉网民关于大学生宿舍调整权益维护事件评论文本的情感变化，相关趋势如图3所示。

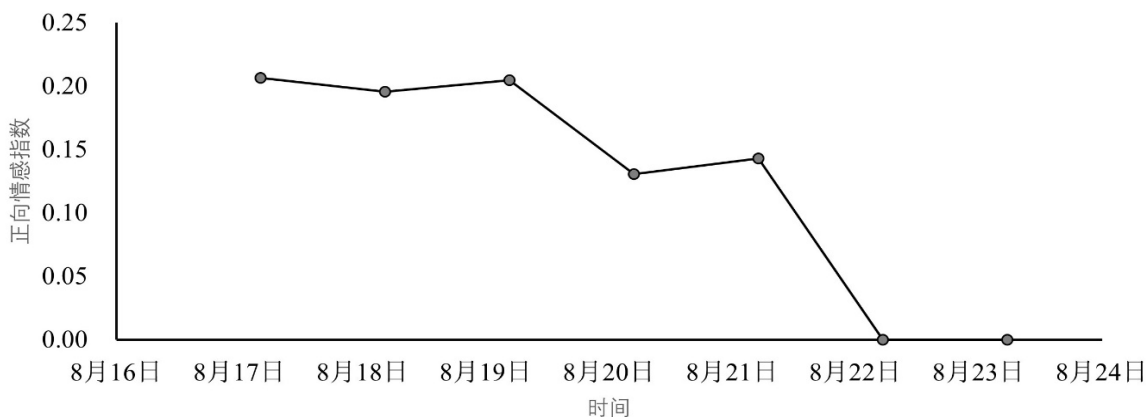


图3 大学生宿舍调整权益维护事件正向情感指数演化图

从大学生宿舍调整权益维护事件正向情感指数演化图中可以了解到大学生网民们对大学生宿舍调整权益维护事件的情感变化。以8月17日作为事件情感演化分析的分水岭,8月暑期学校发布了毕业生宿舍调整信息,部分学生出现疑问继而进行网络维权,网络空间仅有少量评论,因评论数据量较少不具有统计意义。8月17日**大学官方回应“此事还在研究当中”,微博用户的负向情感倾向值为0.79,正向情感倾向值为0.21,这意味着自官方未做正面回复起大学生网民情感呈现爆发性负面增长。8月18日,正向情感指数为0.19,负向情感指数为0.81,正向情感指数与17日相比呈现了小幅下降,整体变化不显著,大学生网民情绪以消极为主。8月19日,正向情感指数为0.20,负向情感指数为0.80。17—19日,正向情感指数基本在0.20上下波动,负向情感指数为0.80上下波动,整体以消极情绪为主。经过三天网络舆情的持续发展,大学生网民未得到官方正面回复,这也被大学生网民认为“高校意在拖延时间、等热度降温、消极不作为”等,这使得同地域其他高校和不同地域经历过同类事件的高校大学生网民不断卷入,负向评论占比持续增多。因此,8月20日,正向情感指数较之前出现较大幅度下降,为0.13,负向情感指数进一步上升为0.87,这意味着大学生网民消极情绪进一步增强。经历了8月21日短暂的正向情感指数小幅上升后,8月22日、23日,正向情感指数已趋近0,负向情绪达到顶点。8月23日之后,虽有部分大学生网民进行零星评论,基本以负向评论为主,但日均数据量不大,无统计意义。

综上可知,本研究提出的基于BiLSTM情

感识别模型得到的情感变化趋势与实际发展情况较为吻合,一定程度上反映了大学生网民在权益维护事件中的情感变化。即充分表明线上大学生网民情感演化与线下**高校应对举措进展是紧密相关的。特别启示是,在权益维护事件中,利益诉求的指向组织或个人应及时、正面地做出针对性和建设性的反馈,否则,网络评论挖掘以及情感演化分析均表明,拖沓、回避、消极应对只会让网民情感持续负向恶化。因此,热点事件中官方情感沟通反馈机制、方式和态度在网络社会情绪疏导中的功能发挥不容忽视。

(2) 不同主题情感倾向差异化特征

基于此,运用BiLSTM模型分别对各个主题的评论文本进行正、负向情感倾向分析,具体分布情况如图4所示。T1主题以事件归责与权益权衡为主,评论文本内容总文档占比为95.49%,其中,正向情感比例为21.00%、负向情感比例为79.00%;这是由于该事件为典型的大学生权益维护事件,因此采集的评论数据多聚焦于事件本身与权益维护,由事件结果及解释所引发的负向情绪整体偏高,总体来看,呈现出“强情绪弱信息”特点,与前人研究的突发公共事件的特点一致。T2主题中以救援响应诉求为主,评论文本内容总文档占比为3.61%,其中,正向情感比例为97.22%、负向情感比例为2.78%;在这一主题中,大学生主要表达对于事件结果涉及的消防安全的担忧以及向国家机构的求助,整体正向评价或正向情绪居多。T3主题中,主要以媒体报道诉求为主,评论文本内容总文档占比为0.90%,其中,正向情感比例为100.00%、负向情感比例为0.00%;在这

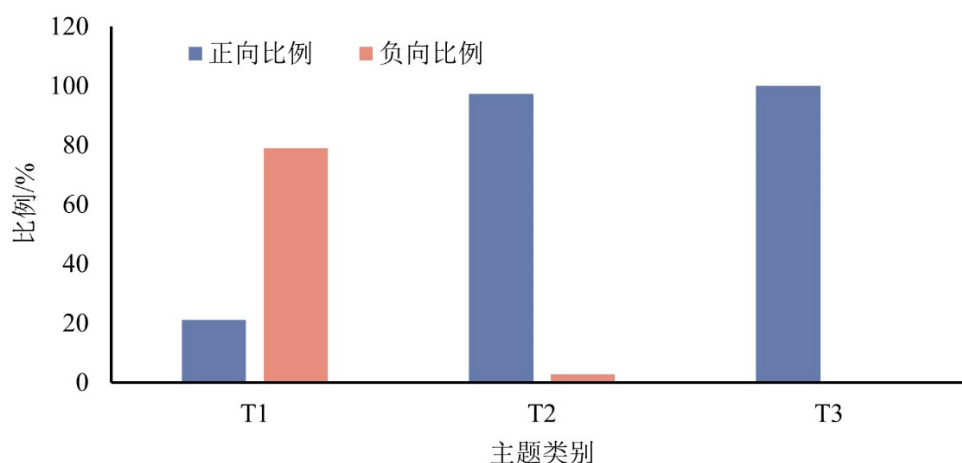


图4 不同主题下情感倾向比例差异化分布特征

一主题中，主要是大学生寻求媒体报道等社会性渠道在舆论方面的支持，均为正向情绪。

不同主题下情感倾向比例差异化分布，整体呈现出“负向表达情绪、正向表达诉求”的特征。事件性质、处置态度是决定网络社会情绪发展趋势的决定性因素；在个体感知、权益权衡、事件归责等机体反应部分，大多数不能保证理性评论，以消极态度吐槽、埋怨、不信任等负面情绪居多，这与校方没有积极回应有很大关系。对于媒体和消防安全诉求的正向表达，表明了大学生在消极网络情绪中的理性态度，能够正确表达情绪关注点的解决途径。

3.3 大学生网络社会情绪生成机理及表达策略分析

(1) 大学生网络社会情绪生成机理阐释

综合上文事件文本特征和情感分析，可以探索运用SOR理论模型进行网络社会情绪生成机理阐释，即把“热点事件（性质）”（S）作为外部刺激；大学生网民对于权益维护事件的“个体感知、事件归责与权益权衡”（O）作

为大学生网民对于权益维护事件的个体认知；情绪的动允性^[35]使得“救援响应诉求、媒体报道诉求”（R）可以作为行为倾向，因而，这一模型机制可以揭示权益维护热点事件中大学生网络社会情绪的生成机理。具体从网络社会情绪生成条件来看，热点事件作为外在刺激并不直接产生情绪，由于认知行动者在信息行为过程中处于特定环境和状态下^[36]，评估和判断构成的感知认知则成为情绪体验产生的必要和充分条件^[35]。由此，对于事件的认知和评判是网络社会情绪生发的基本条件。从网络社会情绪过程机制来看，外在热点事件刺激触发了机体响应，亲身参与学生个体或网络围观集体开始了觉察感知，并以信息表达和分享行为，将个人或集体的情绪通过“社会放大效应”，将这种或正向或负向的社会情绪在网络空间进一步渲染和发酵。从网络社会情绪演化机制来看，网络社会情绪会随时间演化出现波动性，就本研究开展的权益维护热点事件中的网络社会情绪波动特点而言，前期正向情感指标较高，代表大学生网民有所期待；随着权益诉求被指向

群体回应滞怠、大学生群体负向情绪极化，后期正向权益诉求表达急剧减少，几乎被负向情绪湮灭；正向情绪演化整体趋势呈波动性下降。

此外，在“个体感知、事件归责与权益权衡”（O），“救援响应诉求、媒体报道诉求”（R）这两部分，网络社会情绪表达策略内在具有次要驱动机制，具体可由“第一人和第三人效应”理论进行阐释。如，大学生网民认为“宿舍调整后，校园周边租房价格必定会涨”，认为周围房东会受到比本人更多的影响，同时，大学生网民通过@消防机构或@国家媒体的方式，基于“第三人效应”的感知定势，认为事件的扩散对于他人或组织会产生更大影响，通过这种影响以获取国家机构或媒体的支持，权益也会被诉求方妥善处理，从而达到个人或集体权益维护目的。“在**这样的高校宿舍还有很多”，可见舆论的扩散并不都是实际参与人，也有其他地区大学生，认为这样的处置也会出现在自己身边，因而在网络上进行抵制和声援，这是基于“第一人效应”，认为本人比其他人受到更多影响。因此，该类事件的扩散以及网络社会情绪表达策略生成的驱动机制中实际上不能排除“第一人和第三人效应”同时存在。

（2）大学生网络社会情绪表达策略分析

网络社会情绪生成机理解释了网络社会情绪何以形成，并构成了网络社会情绪表达策略的理论基础。同时，因网络社会情绪受到的影响因素组合形式（单一因素或复合因素）不同，网络社会情绪表达策略会呈现具体实践路径。

综合前文对于事件评论文本特征分析，明确了大学生网络社会情绪产生的影响因素可分

为事件性质、个体感知、事件归责与权益权衡、权益诉求四个方面。在本研究中大学生网民则围绕“学校”“大学”“学生”“宿舍”等事件核心要素进行文本评论和情绪表达，其中，事件性质作为影响大学生网络社会情绪的重要因素，其决定了网络社会情绪表达的基本走向；大学生网络社会情绪表达的关注点集中在“个体感知”“事件归责与权益权衡”“权益诉求”等方面，这一部分是情绪表达的核心，由于受到关键节点校方回应被动及内容态度消极的影响，大学生网络社会情绪表达的发展趋势主要表现为弥散性负向情绪；权益诉求总体的表达体量较少，虽然是正向情绪表达，但是从诉求表达的方式和求助内容看，是不得已而为之，甚至是被动、无可奈何的表现。进一步结合SOR等理论模型构建的大学生网络社会情绪生成机理，探究事件刺激下大学生网络情绪表达策略路径为“事件刺激（S）→个体感知、事件归责与权益权衡（O）→救援响应诉求、媒体报道诉求（R）”。

4 结论与启示

4.1 研究结论

本研究搭建了大学生网络社会情绪表达“语义特征—情感识别—生成机理—表达策略”研究框架，为高校思政工作者、信息化建设工作提供—种分析、研判网络社会情绪的工作框架及研究范式。在选取的权益维护热点事件中，基于事件评论文本的特征和情感分析、结合SOR理论模型等，探析了大学生网络社会情

绪生成机理，并提出了大学生网络情绪表达策略路径为“事件刺激（S）→个体感知、事件归责与权益权衡（O）→救援响应诉求、媒体报道诉求（R）”。总体而言，负向事件刺激下，大学生网络社会情绪表达呈现“负向表达情绪、正向表达诉求”的特征，多以负向情绪宣泄代替理性事件阐释；事件归责与权益权衡仍然是大学生关心的主体，行为倾向则是以通过向国家机构、社会渠道等寻求舆论支持等方式进行权益诉求表达；同时，利益诉求的指向组织或个人情感沟通反馈机制、方式和态度在网络社会情绪疏导中的功能发挥不容忽视。

4.2 研究启示

基于事件文本挖掘，进行了大学生网络社会情绪表达策略研究，形成如下研究启示：其一，运用大数据进行事件文本挖掘作为传统主观测量的有效补充，可以更加及时、高效、真实地为高校呈现事件情境下大学生网络社会情绪全貌。同时也为高校育人工作提供了新技术新方法新思路，为构建数字时代背景下思政教育、舆论引领工作框架提供了理论支撑。其二，深入把握事件情境下大学生网络社会情绪生成机理及表达策略，有助于高校正确把握大学生网络社会情绪发展规律、规避舆论风险点、科学制定网络突发公共事件情感沟通反馈机制。高校既要及时、正面、高效回应学生关切、避免情绪过分渲染发酵，又要引导大学生围绕事实客观阐释、正确表达合理诉求。其三，高校应将微观层面大学生网络社会情绪波动置于宏观时代背景下进行观察、解读、疏导和调适，消解圈层冲突对抗思维、增强价值共识凝聚，

引导大学生贡献主流价值观，以推动形成自尊自信、理性平和、积极向上的大学生网络社会心态秩序。

4.3 不足与展望

由于热点事件的范畴应包括负向、中性、正向事件性质，本研究仅选取了一例权益维护的负向事件进行了框架分析，数据样本量较小，研究仍具有局限性。因此，下一步工作将采集不同性质热点事件、结合群体异质性进行文本挖掘分析，以丰富大学生网络社会情绪表达策略的理论框架。

参考文献

- [1] 刘博,董倩倩.情境结构与动力机制:青年群体社会心态的网络表达[J].中国青年研究,2021(10):93-102.
- [2] 王俊秀,张衍.风险认知、社会情绪和未来预期:疫情不同阶段社会心态的变化[J].社会科学战线,2022(10):220-237.
- [3] 郑雯.中国网络社会心态演进:规律及其分析框架[J/OL].西安交通大学学报(社会科学版),1-11[2024-12-17].<http://kns.cnki.net/kcms/detail/61.1329.C.20241023.1932.002.html>.
- [4] 郑雯,乐音,桂勇.网络新生代与网络社会心态:代际更替、心态变迁与引导路径[J].青年探索,2022(2):37-45.
- [5] 代亚君,郝艳华,吴群红,等.突发公共卫生事件公众风险认知量表编制及信效度检验[J].中国公共卫生,2020,36(2):227-231.
- [6] 胡维芳,刘将.我国青年社会心态影响因素分析——基于CGSS2015年调查数据的分析[J].青海社会科学,2021(6):128-134.
- [7] 郭未,沈晖.重大突发公共卫生事件中的网络社会心态:一个整合分析框架[J].西南民族大学学报(人文社会科学版),2020(12):157-164.
- [8] 侯静.社会转型中社会心态的理论内涵、逻辑建构及变迁[J].北京社会科学,2022(4):117-128.

- [9] 王道康, 张吴波. 基于 MacBERT-BiLSTM 和注意力机制的短文本分类研究 [J]. 现代电子技术, 2023, 46(21): 123-128.
- [10] 陆扬. “情感转向”的理论资源 [J]. 上海大学学报 (社会科学版), 2017(1): 30-38.
- [11] KEMPER T D. Sociological Models in the Explanation of Emotions[J]. Handbook of Emotions, 1993, 41: 51.
- [12] 郑雯, 黄荣贵. 把握网络社会心态演进趋势 [N]. 中国社会科学报, 2017-06-01(003).
- [13] KASPERSON R E, RENN O, SLOVIC P, et al. The Social Amplification of Risk: A Conceptual Framework[J]. Risk Analysis, 1988, 8(2): 177-187.
- [14] 雷少杰. 重大突发性公共事件后风险系统的“再”生成及其防范——基于媒介化治理视角的思考 [J]. 广州大学学报 (社会科学版), 2024(6): 17-29.
- [15] 王中邮. 突发事件中网络反向社会情绪的影响因素及机理研究——基于多案例的组态效应分析 [J]. 信息资源管理学报, 2024, 14 (5): 116-131.
- [16] 李娜 姚曦. 情绪引导与情感再造: 突发公共事件中公益广告情感动员机制探析 [J]. 青年记者, 2022(22): 53-55.
- [17] 张彦, 李汉林. 国家治理视角下的总体性社会情绪 [J]. 社会, 44(2): 1-35.
- [18] 朱代琼, 王国华. 突发事件中网民社会情绪产生的影响因素及机理——基于三元交互决定论的多个案定性比较分析 (QCA)[J]. 情报杂志, 2020, 39(3): 95-104.
- [19] 朱代琼, 王国华. 热点公共事件中网络社会情绪演进的影响因素研究——基于“高铁霸座”事件的探讨 [J]. 情报杂志, 2020, 39(8): 94-116.
- [20] 朱丽萍, 陈旭. 重大公共卫生安全事件的消极社会情绪及其治理 [J]. 重庆大学学报 (社会科学版), 2023, 29(6): 295-306.
- [21] 刘晶, 纳欣悦. 高校危机管理: 观念冲突、行为差异与制度缺失 [J]. 高等教育研究, 2022, 43(10): 49-55.
- [22] 杨钧期. 大学生参与集群行为的风险及管控 [J]. 教育理论与实践, 2021, 41(9): 8-11.
- [23] LAZAEUA R. Emotion and Adaptation[M]. New York: Oxford University Press, 1991: 87-214.
- [24] WATSON J B. Psychology as the behaviorist views it[J]. Psychological review, 1913, 20(2): 158-177.
- [25] MEHRABIAN A, RUSSELL J A. An approach to environmental psychology[M]. Cambridge: MIT Press, 1974: 176-200.
- [26] GUO J, LIU Z, LIU Y. Key Success Factors for the Launch of Government Social Media Platform: Identifying the Formation Mechanism of Continuance Intention[J]. Computers in Human Behavior, 2016, 55: 750-763.
- [27] EROGLU S A, MACHLEIT K A, DAVIS L M. Empirical testing of a model of online store atmospherics and shopper responses[J]. Psychology & Marketing, 2003, 20(2): 139-150.
- [28] 江凌, 姜博伦, 颜叶清. 当代大学生对影视剧中吸烟镜头的认知态度与情感偏向研究——基于涵化效果与“第三人效应”现象的实证分析 [J]. 新闻与传播研究, 2014(4): 89-103, 127.
- [29] 司徒凌云, 李益婷, 石进. 基于第三人效应的微博隐私悖论产生路径研究 [J]. 情报杂志, 2022, 41(12): 89-97.
- [30] GROOTENDORST M. BERTopic: Neural topic modeling with a class based TF-IDF procedure[J]. arXiv preprint arXiv:220305794, 2022.
- [31] ABUZAYED A, AL-KHALIFA H. BERT for Arabic topic modeling: an experimental study on BERTopic technique[J]. Procedia Computer Science, 2021, 189: 191-194.
- [32] 刘桂海, 崔福龙, 卢彩菡, 等. 公众对假房源的关注点和态度: 基于微博评论的文本挖掘研究 [J]. 管理评论, 2023, 35(11): 153-165.
- [33] 张维亚, 汤澍, 陈玲玲, 等. 基于网络文本分析的文化遗产生态环境游客感知及其对遗产旅游形象影响研究——以南京明城墙为例 [J]. 生态经济, 2024, 40(6): 135-140, 171.
- [34] 张治鹏, 毛煜升, 张李义. 基于领域 ERNIE 和 BiLSTM 模型的酒店评论观点原因分类研究 [J]. 数据分析与知识发现, 2022, 6(9): 65-76.
- [35] 王继瑛, 陈欣. 意义建构: 生成论视角下的情绪理论 [J]. 心理研究, 2024, 17(5): 387-394.
- [36] 张爱霞, 张新民, 罗卫东. 信息查寻与信息检索的整合研究——对 ISR 集成研究框架的评述 [J]. 图书情报工作, 2007(10): 10-12, 55.



开放科学
(资源服务)
标识码
(OSID)

大学生网络信息焦虑的影响机制研究

方增泉 肖可心 元英 祁雪晶 张思康

北京师范大学新闻传播学院 北京 100875

摘要: [目的/意义] 网络信息焦虑会对大学生的健康成长产生诸多负面影响,因此,根据文献将网络素养划分为六大维度,明确了网络素养不同维度与信息焦虑的关系及上网时长的中介作用,并提出了缓解我国大学生网络信息焦虑的对策建议,从而更好地促进大学生健康成长。[方法/过程] 采取问卷调查法,以全国东中西部 62 所大学的在校生为研究对象,将上网时长作为中介变量,研究大学生网络素养六大维度与信息焦虑之间的具体影响关系。[结果/结论] 研究发现,我国大学生普遍存在中高程度的信息焦虑问题,其中信息安全焦虑尤为严重。在网络素养的六大维度中,网络信息搜索与利用能力、网络信息分析与评价能力、网络印象管理能力、网络安全与隐私保护能力对大学生信息焦虑具有显著的正向影响。上网注意力管理能力、网络价值认知与行为能力对大学生信息焦虑具有显著的负向影响。上网时长在上述的影响路径中均起到部分中介作用。

关键词: 信息焦虑;网络素养;大学生;认知行为理论

中图分类号: G35; TP391

Research on the Influence Mechanism of College Students' Network Information Anxiety

FANG Zengquan XIAO Kexin YUAN Ying QI Xuejing ZHANG Sikang

School of Journalism and Communication, Beijing Normal University, Beijing 100875, China

Abstract: [Objective/Significance] Information anxiety can have many negative impacts on the healthy growth of college students, in view of this, this paper, based on the literature on the six dimensions of network literacy, clarifies the relationship between different dimensions of network literacy and information anxiety and the mediating role of the online time, and proposes countermeasures and suggestions to alleviate the information anxiety of college students in China, so as to better promote the healthy growth of college students. [Methods/Processes] This paper adopts the questionnaire survey method to study the specific

作者简介 方增泉 (1972-), 博士, 教授, 主要研究方向为青少年网络素养、教育政策, E-mail: 15201376625@163.com; 肖可心 (2000-), 硕士研究生, 主要研究方向为网络素养、国际传播; 元英 (1989-), 博士, 讲师, 主要研究方向为网络素养、教育传播; 祁雪晶 (1986-), 博士, 助理研究员, 主要研究方向为青少年网络素养、网络思想政治教育; 张思康 (1995-), 博士, 主要研究方向为 UGC、定量研究、社交平台。

引用格式 方增泉, 肖可心, 元英, 等. 大学生网络信息焦虑的影响机制研究 [J]. 情报工程, 2024, 10(6): 71-81.

influence relationship between the six dimensions of college students' network literacy and information anxiety by using the online time as the mediating variable with students enrolled in 62 colleges and universities in the eastern, middle and western regions of China as the research object. [Results/Conclusions] It was found that Chinese college students generally have medium to high levels of information anxiety, with information security anxiety being particularly serious. Among the six dimensions of network literacy, the ability to search for and utilize online information, the ability to analyze and evaluate online information, the ability to manage online impressions, and the ability to protect online security and privacy had a significant positive effect on the level of college students' information anxiety, whereas the ability to manage online attention and the ability to perceive and act on online values had a significant negative effect on the level of college students' information anxiety. Online time partially mediates each of these pathways of influence.

Keywords: Information Anxiety; Network Literacy; College Student; Cognition and Behavior Theory

引言

根据中国互联网络信息中心（CNNIC）第54次报告，截至2024年6月，我国网民规模近11亿人，互联网普及率达78%^[1]。随着媒介技术与网络系统的高度发展，信息成为与物质和能源同等重要甚至比之更加重要的资源，获取更多信息成为人们对抗不确定性的重要依据。然而，当前迅速传播的互联网信息呈现爆炸式增长的态势，这在一定程度上会导致各类信息无序地堆积在一起。当用户接触到的信息量远远超出其自身的处理能力时，会导致其产生紧张、焦虑、烦躁等消极情绪。

大学生作为数字原住民，对于网络信息的接触、搜索、保存、利用等活动比其他群体更多，因而更容易受到信息焦虑的影响。信息焦虑不仅会影响大学生的心理健康，还可能对其学习和社会适应能力产生负面影响。因此，深入探究大学生信息焦虑的影响机制，对于促进其全面发展具有重要的理论和实践意义。本研究的创新之处在于，第一，突破了现有文献的局限，将传统的信息素养概念更新为网络素养，并结合认知行为理论，将网络素养细分为六大维度

加以研究；第二，通过在全国东中西部62所大学开展大样本问卷调查，详细分析了大学生网络素养的六大维度对信息焦虑的不同影响；第三，引入上网时长作为中介变量，将影响大学生信息焦虑的各种因素进行综合考虑。这种多维度、多层次的研究方法，不仅丰富了网络素养和信息焦虑的理论框架，也为制定缓解大学生信息焦虑的对策提供了科学依据。

1 文献综述

1.1 信息焦虑

信息焦虑（Information Anxiety）一词最早由Wurman于1989年所提出。他认为，信息焦虑产生于人们实际能够获取的信息与以为能够获取或者希望获取的信息之间的“鸿沟”^[4]。Naveed^[5]进一步将信息焦虑定义为包括任何人、任何时间、任何地点都会经历的信息搜索焦虑，同时也包括信息相关的其他焦虑。此后，诸多学者从心理学、信息技术等不同角度针对这一主题进行了相关研究。其中，信息焦虑的影响机制是一大重点。总结已有研究发现，信息焦

虑的产生受到内外部因素的共同影响。内部因素方面,李玉玲等^[6]指出,信息搜集阶段的信息素养、信息饥渴等问题以及信息处理阶段的信息悬念、信息消化不良等原因均会引发用户的信息焦虑。外部因素方面,若出现信息同质化、信息缺失、信息过载等问题,可能会使用户产生信息焦虑^[7]。

1.2 网络素养与信息焦虑

信息素养 (Information Literacy) 是信息焦虑的重要影响因素。曹锦丹等^[8]指出,若在信息检索技能、信息质量的辨别能力等信息素养方面出现欠缺可能会引发信息焦虑。随着互联网的发展,信息的工具性逐渐被网络的普遍性所掩盖,因而也应与时俱进地使用网络素养 (Network Literacy) 代替传统的信息素养进行研究。1994年,McClure^[9]首先使用网络素养这一概念来描述个人识别、访问并使用网络中的电子信息的能力。卜卫^[10]则认为,网络素养是指了解计算机和网络的基础知识及对其管理的能力、创造和传播信息的能力以及具备网络安全保护的能力。莱茵戈德^[11]将网络素养划分为注意力、识别力、参与力、协作力与联网智慧这五个层面。胡余波等^[12]则将网络素养划分为网络安全与道德、网络行为管理、网络认知与评价、网络批判意识、网络自我发展五个方面。

1.3 上网时长的中介作用

既有研究表明,青少年的媒介素养尚未成熟,这会导致其花费更多时间上网^[13]。武文颖^[14]也讨论了网络素养对网络沉迷的影响。而上网时长又会对信息焦虑产生显著影响。刘根勤等^[15]

指出,网络接触时间越长,信息焦虑越严重。吴贤华^[16]也发现,手机依赖会正向预测信息焦虑。而对于上网时长、网络素养、信息焦虑之间的关系,姚丝缘^[17]表示,由于部分大学生的信息素养较低,缺乏对于网络的掌控,因而会出现频繁查看手机APP、没有目的地浏览网页新闻等行为。随着接触网络时间的延长,这些大学生更容易产生信息焦虑。

综上所述,当前对于信息焦虑的研究尚存以下三点不足。第一,当前的研究未能充分揭示网络素养各细分维度如何独立作用于信息焦虑;第二,目前对于网络素养、上网时长与信息焦虑三者关系的探讨仍不充分,没有综合考虑信息焦虑的不同影响因素。即便有少数研究涉及了三者关系,也往往只是简单的理论思辨,缺乏实证数据的支持;第三,我国幅员辽阔,但以往针对国内大学生网络素养与信息焦虑的研究样本的覆盖范围较小,有待更大范围的实证研究。

2 研究设计

2.1 研究方法设计

2.1.1 假设提出与模型构建

既有研究表明,大学生的信息素养显著影响其信息焦虑程度^[18]。而网络素养作为当今网络时代信息素养的发展,也会对大学生的信息焦虑程度产生影响^[19]。本研究以认知行为理论为基础,将网络素养定义为人们对网络世界的信息、事件和情境的认知和行为能力,并将其划分为上网注意力管理能力、网络信息搜索与

利用能力、网络信息分析与评价能力、网络印象管理能力、网络安全与隐私保护能力、网络价值认知与行为能力六大维度,进一步研究各细分维度如何具体作用于信息焦虑。

本研究选择上述六大细分维度的原因如下:学者们普遍认为上网注意力管理能力是网络素养的重要组成部分。荣姗姗^[20]认为,用自我管理的方法增强行为自律能力,是人们必须具备的网络素养;掌握良好的网络信息搜索与利用能力至关重要,也是个人网络素养的重要体现。早在1974年,美国信息产业协会主席Zurkowski^[21]就指出,信息素养就是利用大量的信息工具和资源使问题得到解答的技术和能力;网络信息分析与评价能力是对网络信息进行思考、理解、质疑和分析的能力。张艳秋^[22]将媒介素养划分为三个相互联系的层面,其中一个层面即掌握具体的、批判性的媒介使用能力;网络印象管理能力的本质在于社会交往和互动。由于网络成瘾的程度可以反映个人网络素养的高低,而印象管理是辨别网络成瘾的重要变量^[23],因此本研究也将印象管理作为网络素养测量的一个重要维度;当前,网络空间出现了信息诈骗、隐私泄露、网络暴力、网络谣言等问题。本研究将网络安全与隐私保护能力、网络价值认知与行为能力作为网络素养的重要组成部分,旨在使人们合法、合规地约束自己的网络使用行为,尽可能降低网络使用过程中的各种风险。

综上,本研究提出如下研究假设:

H1: 大学生的上网注意力管理能力会对其信息焦虑产生显著影响。

H2: 大学生的网络信息搜索与利用能力会对其信息焦虑产生显著影响。

H3: 大学生的网络信息分析与评价能力会对其信息焦虑产生显著影响。

H4: 大学生的网络印象管理能力会对其信息焦虑产生显著影响。

H5: 大学生的网络安全与隐私保护能力会对其信息焦虑产生显著影响。

H6: 大学生的网络价值认知与行为能力会对其信息焦虑产生显著影响。

网络素养较低的大学生无法控制自己的上网时长,因此可能会导致严重的信息焦虑。但随着社会媒介化程度的加深,具备较高网络素养的用户群体善于从网络中获得自己所需的信息资源^[24],其上网时长也会有所增加,而随着上网时长延长,其信息焦虑程度可能也会加深。因此,本研究认为大学生的网络素养可以通过上网时长影响信息焦虑,但具体的影响方式有待进一步研究,故提出如下研究假设:

H7: 上网时长在上网注意力管理能力对信息焦虑的影响中存在中介作用。

H8: 上网时长在网络信息搜索与利用能力对信息焦虑的影响中存在中介作用。

H9: 上网时长在网络信息分析与评价能力对信息焦虑的影响中存在中介作用。

H10: 上网时长在网络印象管理能力对信息焦虑的影响中存在中介作用。

H11: 上网时长在网络安全与隐私保护能力对信息焦虑的影响中存在中介作用。

H12: 上网时长在网络价值认知与行为能力对信息焦虑的影响中存在中介作用。

本研究构建的研究模型如图1所示,具体阐述了网络素养的六个细分维度对信息焦虑的影响以及上网时长的中介作用。

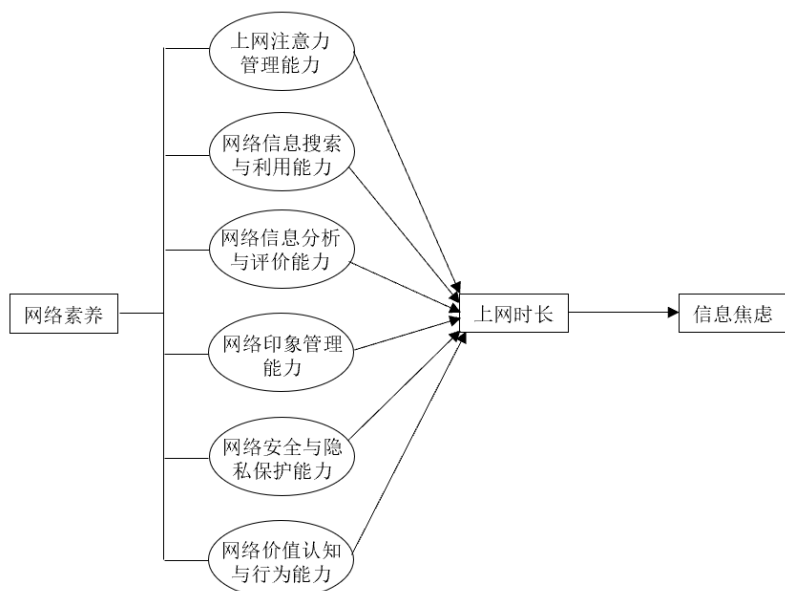


图1 大学生信息焦虑影响机制研究模型

2.1.2 问卷设计

本研究采用问卷调查法，对上述研究假设进行检验。调查问卷共分为三部分，第一部分是对全国大学生个人基本信息的收集，共9个题项（见表1）；第二部分为大学生信息焦虑量表（见表2）。该量表参考曹锦丹、吴贤华、刘根勤等的研究编制，设置4个维度，共25道题项，均采用李克特五分量表进行测量；第三部分是对大学生信息焦虑影响因素的测量（见表3），网络素养部分使用本研究团队在《中国青少年网络素养绿皮书（2022）》中构建的网络素养量表，该量表已经过实证检验，适合进行大学生网络素养的系统测量^[25]。

表1 问卷测量题项(第一部分)

大学生个人信息收集	
个人信息	性别、学校、专业、年级、户口类型、常用媒介类型、上网设备拥有类型、获取信息常用网络平台、上网获取信息类型

2.2 数据搜集与处理

本研究采取整群抽样的方式在全国东中西

部62所大学发放问卷8213份。样本覆盖全国25个省市自治区，包括北京、上海、黑龙江、吉林、广东、浙江、河南、河北、江西、陕西、山西、山东、福建、南京、湖北、湖南、安徽、四川、重庆、云南、内蒙古、贵州、青海、甘肃、澳门特别行政区；学生年级为大一到博三的10个年级；专业分布为理工类、文史类、艺术类与其他类专业。剔除无效样本后，最终确定有效样本量7904份，问卷有效率为96.24%。

2.3 量表信效度检验

结果显示，信息焦虑量表的Cronbach系数值为0.936>0.9，信度质量较高；KMO值为0.942>0.8，显著度为0.000，数据可以被有效提取信息；旋转后4个维度变量的特征根值均大于1，信息焦虑相关统计数据可以分成4个维度进行完全有效的解释分析；旋转后4个维度的方差解释率值分别是21.46%，15.78%，14.73%，11.63%，累积方差解释率为63.62%>50%，研究项的信息量可以被有效提取。

表 2 问卷测量题项(第二部分)

大学生信息焦虑测量		
维度	具体题项	
信息搜索焦虑	T1	我常常因为不知道该从哪里搜索信息感到着急
	T2	根据不同信息（图片、视频、色彩、数字、字符等）使用不同的搜索技巧让我感到困惑
	T3	在搜索信息的过程中，我常常忘记我要搜索的是什么
	T4	大多数时候我对人工智能个性化推荐的结果感到不满意
	T5	由于网上干扰信息繁多，我常常担心自己找不到想要的信息
	T6	在网上搜索信息的时候经常让我感觉烦躁甚至失去耐心
	T7	网上接收到的大量信息让我感觉压力很大
	T8	上网时需要处理太多消息让我感到疲惫
	T9	当面对大量网络信息必须做出选择时，我感到有压力
	T10	我觉得我跟不上网上信息更新的速度，这让我感到焦虑
信息错失焦虑	T11	当我不看手机时，我会很担心错过了重要的信息
	T12	我会每隔几分钟就检查我的手机是否有新消息或者通知
	T13	当我的手机没电或者没法上网时，我会很焦虑
	T14	我需要随时把手机放在身边，否则就会感到焦虑不安
	T15	我担心别人比我知道更多更新更有价值的信息
	T16	我总觉得有一种提示（手机震动声或屏幕闪烁），在提醒我还需要关注一些信息
信息利用焦虑	T17	我会因为网上的信息专业术语难以理解感到沮丧
	T18	在归纳和整理网上搜索到的信息时，我常常因为没有头绪而感到烦恼
	T19	对信息进行分类并提炼出共同点或核心观点时，我会感到有压力
	T20	我常常因为看完信息或新闻很快就遗忘而感到沮丧烦闷
信息安全焦虑	T21	当网络平台要求获取我的个人信息时，我会感到烦躁
	T22	我担心提供给平台的信息会被别人获取
	T23	我担心网络平台收集了太多我的个人信息
	T24	我总有一种我的个人信息会被泄露的预感
	T25	网络平台过度搜集我的信息让我感到很焦虑

表 3 问卷测量题项(第三部分)

大学生信息焦虑影响因素测量	
人口学因素	性别、学历、专业、户口类型、地区、上网时长、网络技能熟练度
网络素养因素	上网注意力管理能力、网络信息搜索与利用能力、网络信息分析与评价能力、网络印象管理能力、网络安全与隐私保护能力、网络价值认知和行为能力

综上所述，本文用于测量全国大学生信息焦虑的问卷量表整体信效度均相对良好，可以据此问卷所收集的数据展开进一步的实证研究与分析。

2.4 描述性统计分析

利用 SPSS26.0 进行统计，结果显示，调查样本总体为 7904 人，从地区分布来看，东部地区 3625 人，中部地区 2224 人，西部地区 2055 人，符合我国的人口统计情况以及教育资源分布东部多，中西部较少的特点；从年级构成来看，大一占比 33.7%，大二占比 20.4%，大三占比 13.5%，大四占比 9.8%，研一占比 7.7%，研二占比 5.9%，研三占比 4.5%，博士生占比 3.6%；从专业分布来看，文史类专业样本 3530 人，

占比 44.7%，理工科专业样本 2371 人，占比 30%，艺术类专业样本 1248 人，占比 15.8%，其他类 755 人，占比 9.6%。

3 结果与讨论

3.1 我国大学生信息焦虑整体情况

研究发现，我国大学生普遍存在中高程度的信息焦虑，信息安全焦虑尤为突出。我国大学生的信息焦虑总分由上述信息焦虑问卷各项目得分相加得出。该问卷共设 25 个题项，每 1 道题项由低到高设置为 1~5 分，问卷满分为 125 分。问卷总得分超过 100 分则表示大学生的信息焦虑较高，总得分低于 50 分则表示其信息焦虑较低。结果显示（见图 2），我国大学生的信息焦虑总分平均值为 79.92，88.9% 的大学生得分介于 50 分到 100 分之间，属中度信息

焦虑。

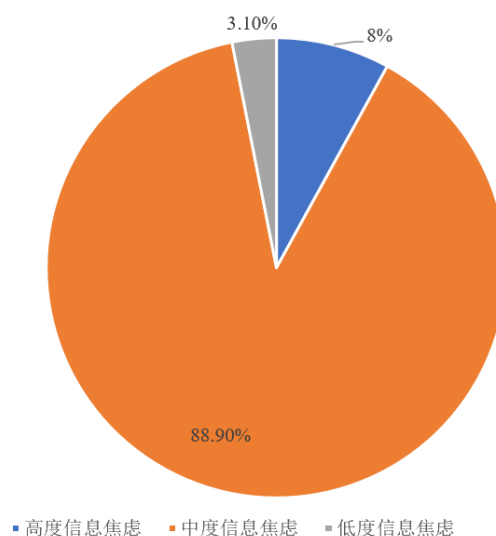


图2 我国大学生低中高度信息焦虑比例

如图 3 所示，我国大学生信息焦虑整体得分平均分为 3.20 分。其中，信息安全焦虑的平均分最高，为 3.65 分，其次为信息错失焦虑和信息搜索焦虑，分别为 3.15 分和 3.07 分，信息利用焦虑的平均分最低，为 3.01 分。

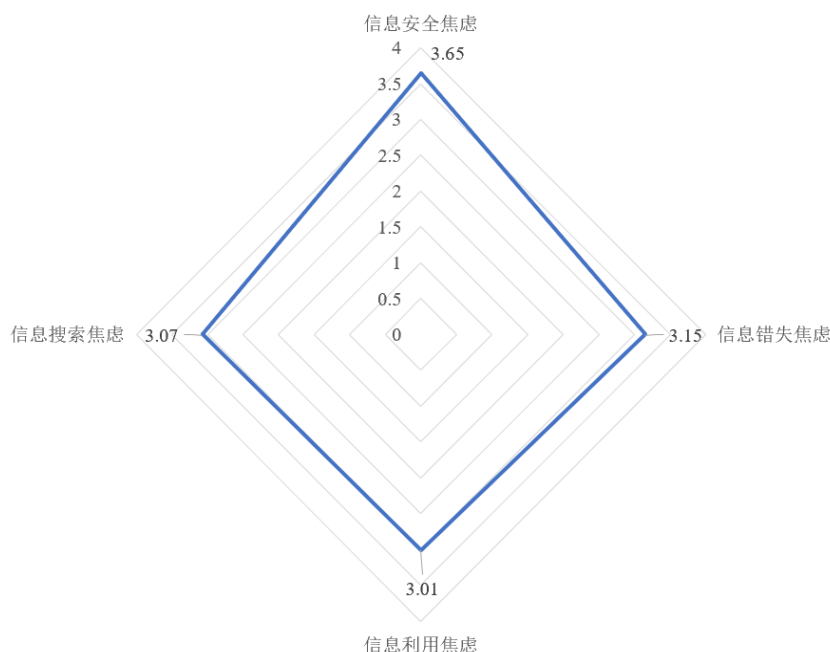


图3 我国大学生信息焦虑各维度得分

3.2 多元线性回归结果

SPSS26.0 结果分析显示,网络素养六大维度对信息焦虑的影响方式各不相同。上网时长在上述影响路径中起部分中介作用,即网络素养六大维度不仅会直接影响信息焦虑,还会部分通过上网时长来影响。

3.2.1 上网注意力管理能力维度

研究发现,上网注意力管理能力负向预测信息焦虑,上网时长起部分中介作用,H1 和 H7 成立。如表 4 和表 5 所示,上网注意力管理能力 ($\beta = -4.37, t = -14.10, p < .001$) 和上网时长 ($\beta = 1.53, t = 8.54, p < .001$) 显著负向 / 正向预测信息焦虑。

表 4 中介模型各变量关系的回归分析 (N = 7904)						
变量	模型 1		模型 2		模型 3	
	β	t	β	t	β	t
上网注意力管理能力	-4.56	-14.69***	-0.12	-6.37***	-4.37	-14.10***
上网时长					1.53	8.54***
R ²	0.03		0.01		0.04	
F	215.69***		40.62***		145.26***	

注: 模型 1: 上网注意力管理能力预测信息焦虑; 模型 2: 上网注意力管理能力预测上网时长; 模型 3: 上网注意力管理能力和上网时长共同预测信息焦虑。

表 5 预测信息焦虑的中介效应分析					
	效应值	标准误	Bootstrap 95%CI		占总效率比率
			下限	上限	
总效应	-4.56	0.31	-5.17	-3.95	
直接效应	-4.37	0.31	-4.98	-3.76	
间接效应	-0.19	0.04	-0.27	-0.11	4%

3.2.2 网络信息搜索与利用能力维度

研究发现,网络信息搜索与利用能力正向预测信息焦虑,上网时长起部分中介作用,H2 和 H8 成立。如表 6 和表 7 所示,网络信息搜

索与利用能力 ($\beta = 2.81, t = 10.83, p < .001$) 和上网时长 ($\beta = 1.62, t = 9.01, p < .001$) 都能显著正向预测信息焦虑。

表 6 中介模型各变量关系的回归分析 (N = 7904)						
变量	模型 1		模型 2		模型 3	
	β	t	β	t	β	t
网络信息搜索与利用能力	2.92	11.20***	0.07	4.10***	2.81	10.83***
上网时长					1.62	9.01***
R ²	0.02		0.00		0.03	
F	125.47***		16.83***		103.97***	

注: 模型 1: 网络信息搜索与利用能力预测信息焦虑; 模型 2: 网络信息搜索与利用能力预测上网时长; 模型 3: 网络信息搜索与利用能力和上网时长共同预测信息焦虑。

表 7 预测信息焦虑的中介效应分析					
	效应值	标准误	Bootstrap 95%CI		占总效率比率
			下限	上限	
总效应	2.92	0.26	2.41	3.43	
直接效应	2.81	0.26	2.30	3.32	
间接效应	0.11	0.03	0.05	0.18	3%

3.2.3 网络信息分析与评价能力维度

研究发现,网络信息分析与评价能力正向预测信息焦虑,上网时长起部分中介作用,H3 和 H9 成立。如表 8 和表 9 所示,网络信息分

表 8 中介模型各变量关系的回归分析 (N = 7904)						
变量	模型 1		模型 2		模型 3	
	β	t	β	t	β	t
网络信息分析与评价能力	1.35	4.13***	0.10	5.01***	1.18	3.62***
上网时长					1.67	9.23***
R ²	0.00		0.00		0.01	
F	17.02***		25.15***		51.29***	

注: 模型 1: 网络信息分析与评价能力预测信息焦虑; 模型 2: 网络信息分析与评价能力预测上网时长; 模型 3: 网络信息分析与评价能力和上网时长共同预测信息焦虑。

表 9 预测信息焦虑的中介效应分析

	效应值	标准误	Bootstrap 95%CI		占总效率比率
			下限	上限	
总效应	1.35	0.33	0.71	1.99	
直接效应	1.18	0.33	0.54	1.82	
间接效应	0.17	0.04	0.09	0.26	12.6%

析与评价能力 ($\beta=1.18, t=3.62, p<.001$) 和上网时长 ($\beta=1.67, t=9.23, p<.001$) 都能显著正向预测信息焦虑。

3.2.4 网络印象管理能力维度

研究发现,网络印象管理能力正向预测信息焦虑,上网时长起部分中介作用,H4和H10成立。如表10和表11所示,网络印象管理能力($\beta=9.77, t=40.74, p<.001$)和上网时长($\beta=0.92, t=5.55, p<.001$)都能显著正向预测信息焦虑。

表 10 中介模型各变量关系的回归分析 (N = 7904)

变量	模型 1		模型 2		模型 3	
	β	t	β	t	β	t
网络印象管理能力	9.93	41.60***	0.17	10.49***	9.77	40.74***
上网时长					0.92	5.55***
R ²	0.18		0.01		0.18	
F	1730.32***		110.14***		883.82***	

注:模型1:网络印象管理能力预测信息焦虑;模型2:网络印象管理能力预测上网时长;模型3:网络印象管理能力和上网时长共同预测信息焦虑。

表 11 预测信息焦虑的中介效应分析

	效应值	标准误	Bootstrap 95%CI		占总效率比率
			下限	上限	
总效应	9.93	0.24	9.46	10.39	
直接效应	9.76	0.24	9.29	10.24	
间接效应	0.16	0.04	0.08	0.23	1.7%

3.2.5 网络安全与隐私保护能力维度

研究发现,网络安全与隐私保护能力正向预测信息焦虑,上网时长起部分中介作用,H5和H11成立。如表12和表13所示,网络安全与隐私保护能力($\beta=5.34, t=22.06, p<.001$)和上网时长($\beta=1.51, t=8.61, p<.001$)都能显著正向预测信息焦虑。

表 12 中介模型各变量关系的回归分析 (N = 7904)

变量	模型 1		模型 2		模型 3	
	β	t	β	t	β	t
网络安全与隐私保护能力	5.45	22.42***	0.07	4.52***	5.34	22.06***
上网时长					1.51	8.61***
R ²	0.06		0.00		0.07	
F	502.78***		20.38***		290.75***	

注:模型1:网络安全与隐私保护能力预测信息焦虑;模型2:网络安全与隐私保护能力预测上网时长;模型3:网络安全与隐私保护能力和上网时长共同预测信息焦虑。

表 13 预测信息焦虑的中介效应分析

	效应值	标准误	Bootstrap 95%CI		占总效率比率
			下限	上限	
总效应	5.45	0.24	4.97	5.92	
直接效应	5.34	0.24	4.87	5.82	
间接效应	0.11	0.03	0.05	0.17	2%

3.2.6 网络价值认知与行为能力维度

研究发现,网络价值认知与行为能力负向预测信息焦虑,上网时长起部分中介作用,H6和H12成立。如表14和表15所示,网络价值认知与行为能力($\beta=-4.14, t=-18.57, p<.001$)和上网时长($\beta=1.79, t=10.12, p<.001$)显著负向/正向预测信息焦虑。

表 14 中介模型各变量关系的回归分析 (N = 7904)

变量	模型 1		模型 2		模型 3	
	β	t	β	t	β	t
网络价值认知与行为能力	-4.09	-18.20***	0.03	2.23*	-4.14	-18.57***
上网时长					1.79	10.12***
R ²	0.04		0.00		0.05	
F	331.38***		4.96*		218.99***	

注：模型 1：网络价值认知与行为能力预测信息焦虑；模型 2：网络价值认知与行为能力预测上网时长；模型 3：网络价值认知与行为能力和上网时长共同预测信息焦虑。

表 15 预测信息焦虑的中介效应分析

	效应值	标准误	Bootstrap 95%CI		占总效率比率
			下限	上限	
总效应	-4.08	0.22	4.97	5.92	
直接效应	-4.14	0.22	4.87	5.82	
间接效应	0.06	0.03	0.05	0.17	2%

4 对策与建议

4.1 个人：针对性提升网络素养，提高网络信息利用率

研究结果显示，网络素养六大维度对信息焦虑的影响方式各不相同。因此，第一，大学生要加强信息检索知识技能的学习，如检索表达式的合理运用等，并使用辅助记录工具将网络信息分门别类地保存；第二，养成良好的网络使用习惯，在上网前设置阶段性上网目标，在上网时利用 APP 监督媒介使用行为，在上网后制作媒介接触日志，对上网行为总结反思；第三，邓巴认为，个体的精力是有限的，人类的“平均社群规模”为 148^[19]。将所维持的关系数保持在邓巴数字左右，对网络社交关系进

行筛选，可以减少对于无效信息的接触；第四，大学生不仅要做到文明上网，自觉遵守社会公德与网络道德，也要对网络信息持有审慎态度，在良莠不齐的网络信息面前保持冷静和理性。

4.2 高校：开设网络素养课程，培养正确网络价值观

高校是大学生观念与习惯养成的重要场所。因此，第一，除了拓展图书馆电子信息资源库外，可以开设网络素养通识课程，提高网络素养；第二，举办多样网络素养活动，树立健康价值观念。可以借助学工办、共青团等的力量，举办健康网络价值观系列讲座活动、网络道德楷模表彰会、网络伦理知多少知识竞赛等；第三，由于大学生的上网时长与信息焦虑具有显著的相关性，因此高校可以通过健全网络管理规范，引导学生减少上网时长，并通过心理辅导的方式帮助其摒弃对网络的过度追求，加强对线下生活的关注。

4.3 社会：共筑信息安全保护网，减轻网络信息安全焦虑

研究发现，在信息焦虑的四个维度中，大学生的信息安全焦虑最为严重，亟需社会各界共同参与解决。对于国家相关部门而言，应根据时代发展变化，不断更新法律法规，继续做出细节规范，给予信息安全更强的法律保障，加大网络违法行为的打击力度；对于各大网络平台而言，应积极承担主体责任，加强技术研发保护。不仅要畅通反馈渠道，做到定时检查、全时在线，在遇到网络安全问题时能够第一时间修补漏洞，也要严格遵守国家法律法规，明

确用户数据库的使用范围,不得过度获取网民的个人信息,也不能制定霸王条例,强迫用户接受信息授权。

5 总结

本研究结合认知行为理论,对网络素养做出了全新定义,并将其细分为六大维度,揭示其对信息焦虑的不同影响,还引入上网时长作为中介变量,综合考虑了影响信息焦虑的不同因素。

本研究的不足之处在于,第一,外部信息环境也会影响大学生的信息焦虑程度,未来要考虑多方面的影响因素,进一步做好变量控制;第二,本研究将不同年级的学生合并在一起加以分析,未来也可以专门针对本、硕、博的某一群体展开探究,或比较不同层次之间存在的差异。

参考文献

- [1] 中国互联网络信息中心.第54次《中国互联网络发展状况统计报告》[EB/OL].(2024-08-28)[2024-08-29].<https://www.cnnic.cn/n4/2024/0828/c208-11063.html>.
- [2] 张帆.显与隐之间:985学子焦虑的传播机制与媒介调适策略——兼论数字青年与新媒介技术的复杂关系[J].中国图书评论,2024(5):29-41.
- [3] 周茂春,高钰媛.人工智能缓释用户信息焦虑——以ChatGPT为例[J].大学图书情报学刊,2024,42(3):80-89.
- [4] 理查德·索·乌曼.信息饥渴——信息选取、表达与透析[M].北京:电子工业出版社,2001:29-33.
- [5] NAVEED M A, ANWAR M A.Towards information anxiety and beyond[J]. Webology, 2020, 17(1): 65-80.
- [6] 李玉玲,曹锦丹.信息焦虑的概念界定[J].图书馆学研究,2011(3):2-4.
- [7] BAWDEN D, ROBINSON L. The dark side of information: overload, anxiety and other paradoxes and pathologies[J]. Journal of information science, 2009, 35(2): 180-191.
- [8] 曹锦丹,王畅,刘鑫,等.用户信息焦虑影响因素及其干预模式研究[J].情报科学,2010,28(10):1461-1463,1468.
- [9] MCCLURE C R. Network literacy: A role for libraries?[J]. Information Technology & Libraries, 1994, 13(2): 115-125.
- [10] 卜卫.媒介教育与网络素养教育[J].家庭教育,2002(11):16-17.
- [11] 霍华德·莱茵戈德.网络素养:数字公民、集体智慧和联网的力量[M].北京:电子工业出版社,2013:279-286.
- [12] 胡余波,潘中祥,范俊强.新时期大学生网络素养存在的问题与对策——基于浙江省部分高校的调查研究[J].高等教育研究,2018,39(5):96-100.
- [13] 庞博.青少年媒介依赖应对策略问题研究[D].黑龙江:黑龙江大学,2019.
- [14] 武文颖.大学生网络素养对网络沉迷的影响研究[D].大连:大连理工大学,2017.
- [15] 刘根勤,曹博林.高校学生网络接触与信息焦虑实证研究[J].中国青年研究,2012(9):53-57.
- [16] 吴贤华.大学生手机依赖、睡眠质量与信息焦虑的关系[J].内江师范学院学报,2020,35(12):7-11.
- [17] 姚丝绦.信息过载背景下当代青年信息焦虑及对策研究[J].新西部,2024(2):107-111.
- [18] 龚花萍,吴灏驰,梅娇.大学生信息焦虑行为影响因素及应对策略[J].情报理论与实践,2020,43(2):55-60.
- [19] 何琪琪,钟起,徐雨洁,等.本科护生网络素养的潜在剖面分析及其在线学习效果比较[J].护理学报,2022,29(2):62-67.
- [20] 荣珊珊.安徽高校学生网络素养现状及其教育实践探究[D].安徽:安徽师范大学,2007.
- [21] ZURKOWSKI P. The information service environment relationships and priorities[M]. Washington D.C.: National Commission on Libraries and Information Sciences, 1974: 6.
- [22] 张艳秋.理解媒介素养:起源、范式与路径[M].北京:人民出版社,2012:167.
- [23] 赵球霞,邓银.青少年网络成瘾的影响因素及应对策略[J].社会科学前沿,2023,12(6):2996-3002.
- [24] 喻国明,耿晓梦.“深度媒介化”:媒介业的生态格局、价值重心与核心资源[J].新闻与传播研究,2021,28(12):76-91,127-128.
- [25] 方增泉,祁雪晶.中国青少年网络素养绿皮书(2022)[M].北京:人民日报出版社,2022:29-46.



开放科学
(资源服务)
标识码
(OSID)

科技管理数据安全治理实践路径研究

徐晨阳 寇亚东 李子伦 王飘 王东

中国科学技术信息研究所 北京 100038

摘要: [目的/意义] 科技管理数据作为国家科技创新与经济社会发展的关键资源,在大数据时代,其安全治理显得尤为重要。[方法/过程] 随着科技投入增加和创新能力提升,科技管理数据量激增,质量提高,但同时也面临着数据安全的新挑战,包括合规性不足、数据孤岛、数据质量参差不齐、数据流转复杂化以及安全风险加剧等问题。为应对这些挑战和问题,结合科技安全战略背景,规划了科技管理数据安全治理体系,并对摸清现状、安全建设规划、分级分类、全生命周期管理、运营监管等安全治理实践路径做了详细介绍。[结果/结论] 科技管理数据安全治理是一个系统工程,需要从战略高度出发,综合运用管理与技术手段,构建动态适应、全面覆盖的安全防护体系,以保障数据安全,促进科技管理工作的高效、健康发展,支撑国家科技创新战略的顺利实施。

关键词: 科技安全; 科技管理; 数据治理; 数据安全

中图分类号: G35; TP391

Research on the Practice Path of Data Security Governance in Science and Technology Management

XU Chenyang KOU Yadong LI Zilun WANG Piao WANG Dong

Institute of Scientific and Technical Information of China, Beijing 100038, China

Abstract: [Objective/Significance] Science and technology management data is a key resource for national science and technology innovation and economic and social development. In the big data era, its security governance is particularly important. [Methods/Processes] With the increase of science and technology investment and the improvement of innovation capacity, the volume and quality of science and technology management data have soared, but at the same time, they are faced with new challenges of data security, including insufficient compliance, data silos, uneven data quality, complicated data flow, and increased security risks. In order to cope with these challenges and problems, this paper, combined with the background of science and

作者简介 徐晨阳(1992-), 硕士, 助理研究员, 主要研究方向为科技安全、数据治理、安全管理; 寇亚东(1992-), 硕士, 助理研究员, 主要研究方向为信息安全、信息系统运维, E-mail: kouyd@istic.ac.cn; 李子伦(1994-), 硕士, 助理研究员, 主要研究方向为计算机应用、信息系统建设; 王飘(1992-), 硕士, 助理研究员, 主要研究方向为计算机应用、信息系统建设; 王东(1987-), 博士, 副研究员, 主要研究方向为系统建设、自然语言处理、知识图谱。

引用格式 徐晨阳, 寇亚东, 李子伦, 等. 科技管理数据安全治理实践路径研究[J]. 情报工程, 2024, 10(6): 82-92.

technology security strategy, plans the science and technology management data security governance system, and makes a detailed introduction to the security governance practice paths such as finding out the current situation, security construction planning, classification, whole life cycle management, and operation supervision. [Results/Conclusions] Science and technology management data security governance is a systematic project, which needs to start from a strategic height, comprehensively apply management and technology means, and build a dynamic adaptive and comprehensive security protection system to ensure data security, promote the efficient and healthy development of science and technology management, and support the smooth implementation of national science and technology innovation strategy.

Keywords: Scientific and Technological Security; Science and Technology Management; Data Governance; Data Security

引言

科技管理数据是国家科技创新发展的重要资源^[1],是信息时代影响面最宽、开发利用潜力最大的科技资源。在大数据时代,科技创新、科技战略决策越来越依赖于大量、系统、高可信度的科技管理数据。近年来,我国高度重视科技创新工作,科技投入和创新能力不断提升,科技管理数据也大幅增长。随着信息化技术的快速发展,科技管理业务数字化转型迎来新机遇,数据成为科技管理业务发展的重要资源要素,但新技术、新需求、新场景也给数据安全带来新的压力与挑战,泄漏、窃取、滥用风险与日俱增。当前我国正处于实施创新驱动发展战略和建设科技强国的关键时期^[2],开展科技管理数据安全治理是加强我国科技创新能力建设和保障国家安全、科技安全的重要方式和手段。

1 数据安全治理研究现状

科技管理数据质量和应用效果是各国开展科技统筹^[3]、科技竞争的内容,其在全球范围内关注度与重要性的日益提高引发了各界对如

何高效进行数据治理,最大限度地发掘其内在价值的考虑。加快科技管理数据开发、共享与保护的步伐成为世界各国竞相提升数据治理水平、实现科技强国的重要战略目标^[4]。

面对数字化浪潮下数据的井喷式增长引发的乱象,许多国家都从国家安全、社会稳定、企业和个人信息安全层面出台相应的法律法规和管理要求,加强对信息安全,特别是数据安全的风险防范。梅傲等^[5]从顶层设计、监管机制、国际合作3个方面出发,梳理日本数据安全治理制度,并阐释日本数据安全治理制度现状。陈毅等^[6]探讨了国家安全体系和能力现代化视域下数据安全治理的困境及突围路径。吕明元等^[7]从政策、法律、制度、行业等方面分析我国数据安全治理进展,剖析目前在数字基础设施建设、数据保护法律体系、数据安全监管、国家数据主权方面存在的问题,分析了美国、德国、日本、韩国的数据安全治理措施。李雪莹等^[8]和王庆德等^[9]分别从数据安全治理实践和数据安全治理的行业实践等方面开展研究,提出数据安全治理理念,并从民航、企业、金融、能源和零售5个行业实践展开分析,提出数据安全治理的政策建议。

在不同行业的数据安全治理研究中,王玉等^[10]对政务数据安全治理体系开展研究探索,提出了适用于我国政务数据安全治理的体系框架,并介绍了相关技术以及项目实践案例。朱洪斌等^[11]研究了电力大数据安全治理体系,提出了电力大数据安全治理要求。杨超等^[12]对工业互联网数据安全治理实践进行了研究,论述了工业互联网数据安全治理所采取的各种策略、技术和活动,从组织建设、业务流程、规章制度、技术工具等方面介绍了提升数据安全风险应对能力的过程。侯鹏等^[13]对金融数据安全治理智能化技术与实践开展了研究。原磊^[14]研究了平台企业数据安全治理。在数据治理相关技术工具上,朱佳妮等^[15]、程伟等^[16]、许杰等^[17]和胡剑等^[18],分别研究了区块链、大数据等技术在数据安全治理体系中的应用和创新。

在国家层面发布加强和规范数据安全保护的政策制度,推动数据开放共享^[19],对于服务科技创新、支撑科技决策等方面具有重要意义。盛小平等^[20]研究了科学数据开放共享中的数据安全治理,运用规范分析法,梳理与界定科学数据开放共享中的数据安全问题,并从多个维度探究科学数据安全治理措施。面对当前科技创新对科技管理数据的应用和服务需求,尤其是与欧美发达国家相比,我国科技管理数据的治理、管理与应用仍然存在明显不足,对科技管理数据的安全治理研究也较少。随着科技管理业务的发展,在利用数据资产创造价值、辅助决策的同时,对数据质量和稳定性、安全性要求也有所提升。因此在科技安全视角下,为了有效应对严峻的国际信息安全和数据泄漏风

险等巨大挑战,需开展数据安全治理体系^[21]的探索研究、建设。

2 科技管理数据安全治理的痛点和问题

结合数据安全治理研究现状,以往数据安全建设中沿用自下而上的“堆产品”思维、烟囱式建设,导致数据安全普遍存在管理制度体系不健全、数据资产权责不清晰、数据业务流程不明确等问题挑战。科技管理数据安全治理是科技安全的重要组成部分也是难点之一,对于科技管理数据的利用与安全的平衡,仍然存在“对数据识别不完整”“数据流转无法溯源”“安全与业务难以兼顾”等一系列现实问题。

数据梳理与分类分级实施难度大。数据资产量级不清,在科技管理业务过程中,因其内部业务模块多、数据类别多、分布广的特点,所引发的数据标准不统一、数据质量有差别等一系列问题亟须解决。在数据分类分级方面^[22],一旦数据梳理不清,很难依据规范开展落地实践的分类分级。

数据安全实施细则尚不完善。科技管理数据会在不同载体和场景环境下流转,数据价值、量级、周期性都会发生动态变化。数据安全建设要求、数据安全风险评估等^[23]是基于某个标准,进行相对静态、固化的评估,无法适应数据在不同科技业务场景、不同数据应用目标下的安全要求。

数据流转场景下的高安全需求。随着科技管理业务的发展和科技管理信息系统的不断完善,数据量大、数据调用常态化、数据处理活

动复杂。国产化建设背景下,数据流转在终端、应用、组件等场景发生巨大改变,数据安全性要求也随之增多。数据流转频繁下的数据安全风险^[24]已打破传统的网络安全区域划分,传统的边界、主机、系统、终端、数据库的单点防护已无法满足数据安全防护的需求。

数据安全风险态势持续加剧。近年来,数据安全风险事件频发,常规状态下管理、技术层面的安全防护属于静态防护,难以及时根据科技管理业务和合规性的变化动态调整,导致数据安全策略滞后,无法有效^[25]。

3 科技管理数据安全治理框架

结合上述痛点和问题,科技管理数据安全治理亟需与时俱进、体系化开展,以数据为中心、数据安全合规为底线,驱动科技管理数据开展分级分类,落实全生命周期安全防护。通过数据安全运营,持续开展治理效果评估与优化。保障数据在安全可控的情况下应用并发挥价值,实现安全与业务发展兼顾、从管理到技术全方位体系融合,持续提升数据安全保障能力。框架如图1所示。

科技管理数据安全治理以数据为中心,一方面是站在数据的视角看数据,当科技管理数据被访问时,从内外层次去区分,分别是数据载体、数据逻辑、数据访问拓扑和数据身份等。另一方面是从流动的视角来看,科技管理数据本身是为科技管理业务服务的,基于业务的需求需要看到数据、互相调用数据,最终数据流动至各个科技管理业务系统,因此相比较网络安全,数据安全风险暴露面无限大,任何数

据会从有序、无序最终到失控,因此数据流动到任何位置,安全防护就应该覆盖到该位置。

在科技管理数据安全治理过程中,需以安全合规为前提。在满足合规性要求的同时,兼顾业务实际发展状况,开展科技管理数据分级分类,并结合数据级别和类型,制定数据全生命周期安全防护策略,一方面动态支撑数据治理安全运营,另一方面结合运营效果,反向监管、动态更新防护策略和技术手段。该框架突出环形治理思路,是互相支撑、互相驱动、互相补充完善的过程。

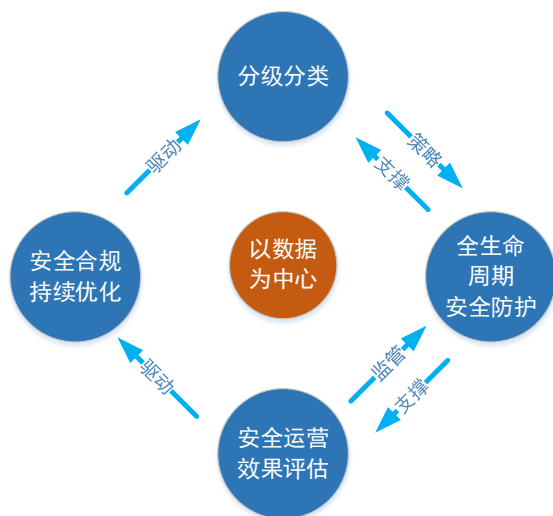


图1 科技管理数据安全治理框架

4 科技管理数据安全治理实践路径

科技管理数据安全治理的目标是开展治理工作的前进方向,根据上述框架和思路,本文重点开展科技管理数据安全治理实践路径研究,通过厘清数据现状、摸清安全风险现状、开展数据分级分类、规划数据安全建设,进行常态化数据安全防护以及持续有效的监管,逐步实现数据安全全域可管、风险全局可视,以及数据安全可信的目标。实践路径如图2所示。

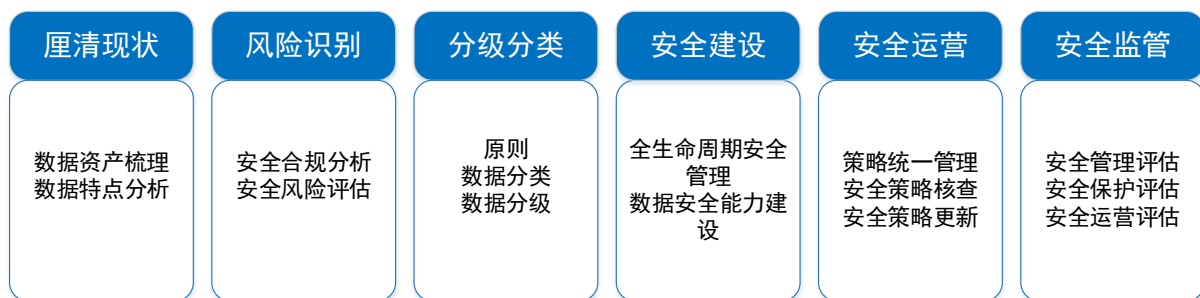


图2 科技管理数据安全治理实践路径

4.1 厘清现状

4.1.1 数据资产梳理

数据资产梳理是数据安全治理的基础,通过对数据资产的梳理,可以确定数据的类型和量级、数据范围、数据业务场景、数据流转形式、敏感性数据的分布等。通过开展科技管理数据资产梳理,可以进一步明确防护重点,改善决策制定,保证数据准确性,加强数据安全和合规性,提高数据管理效率,以便能最大程度地利用这些数据,提升科技管理服务水平。

4.1.2 科技管理数据特点分析

科技管理数据是实施国家科技创新战略的重要基础性资源。大数据时代,科技创新越来越依赖于科技管理数据的治理与应用。科技管理数据不同于其他数据,具有体量大、影响面广等特点,在进行数据治理过程中,需结合特点针对性开展。科技管理数据是支撑科技发展的基础,因此对科技管理数据进行有效安全治理,本身就是辅助科技创新发展的一种方式。

4.2 风险识别

4.2.1 数据安全合规性分析

科技管理数据安全合规是开展科技管理业务工作的底线要求。近年来,陆续出台的法律

持续健全我国数据安全法律法规矩阵^[26],构建了我国网络空间治理和数据保护的基本法。因此对照国家法律法规、政策文件、标准规范,通过数据安全合规性分析,尽可能发现数据安全方面存在的问题,能够更有针对性建设全方位的数据安全治理体系。

4.2.2 数据安全风险评估

科技管理数据安全风险评估,主要围绕数据和数据处理活动,聚焦可能影响数据的保密性、完整性、可用性和数据处理合理性的安全风险,掌握数据安全总体状况,发现安全隐患,为后续数据安全治理建设规划提出数据安全管理和技术防护措施建议,提升数据安全防护攻击、防破坏、防窃取、防泄露、防滥用能力。

4.3 数据分级分类

4.3.1 原则

结合科技管理数据特点,科技管理数据在开展数据分级分类时,需遵循相关原则。

合法合规原则,遵循有关法律法规和标准规范规定,满足数据安全要求^[27]。

科学系统化原则,综合考虑科技管理数据敏感性高、关联性强、涉及主体多等特点进行科学系统化的分类。

就高从严原则，数据分级时采用就高不就低的原则进行统一定级，例如某一数据类别包含多个级别的数据项，按照数据项的最高级别对数据进行定级。

适用客观性原则，确保分级结果能够为数据安全提供可落地的管控措施，避免过于复杂的分级规划，保证执行的可行性。

自主灵活性原则，根据科技管理业务特性和数据安全需要，依据相关分类分级实施

指南，确立数据的审核机制，落实数据逐级报批流程。

动态调整原则，数据的类别级别可能因时间变化、管理要求变化而发生改变，因此需要对数据分类分级进行定期审核并及时调整。

4.3.2 数据分类

结合业务场景，科技管理数据按照数据特性可分为科研项目、科研主体、过程管理、科技监督及基础运行五个类别。如表 1 所示。

表 1 数据分类表

数据类型	数据描述
科研项目类数据	各类科技计划（专项、基金等）项目（课题）基本信息、各阶段的申报材料，以及科研成果等。
科研主体类数据	开展科技计划、科技专项项目申报的科研单位、申报人员及科技专家数据，以及开展科技管理的专业机构、管理部门等单位及人员数据。
过程管理类数据	各类科技计划从需求征集、实施方案、管理通知、评审结果、预算安排、监督检查、结题验收结论等过程中产生的管理数据。
科技监督类数据	科技管理工作中科研诚信数据。
基础运行类数据	支撑科技管理工作的系统运行维护及安全保障过程中所产生的数据，包括日志、流量等。

4.3.3 数据分级

结合科技管理数据类别，根据数据在遭受泄露、破坏或非法利用后，对国家安全、科技

安全、社会利益、个人及组织的合法权益带来的危害程度和影响，将数据从低到高分成公开、一般、重要、核心四个级别。如表 2 所示。

表 2 数据分级表

数据级别	数据描述
公开级别	数据完全公开，可被公众获知、使用。数据被泄露、破坏或非法利用后，对国家安全、科技安全、社会利益、个人及组织的合法权益带来轻微影响。
一般级别	数据有条件公开，可被认证后的对象获知、使用。数据被泄露、破坏或非法利用后，对国家安全、科技安全、社会利益、个人及组织的合法权益带来轻微危害。
重要级别	数据不对外公开，可被科技管理单位获知、使用。数据被泄露、破坏或非法利用后，对个人及组织的合法权益带来严重危害，对国家安全、科技安全、社会利益带来一定危害。
核心级别	数据不对外公开，可被科技管理政府机构获知、使用。数据被泄露、破坏或非法利用后，对个人及组织的合法权益带来特别严重危害，对国家安全、科技安全、社会利益带来严重危害。

4.4 数据安全建设

数据作为资产，具备完整生命周期，在科技管理数据全生命周期各个环节均存在数据安全风险。为了更好地应对数据安全风险，需从

管理和技术两方面入手，实现对数据安全使用和全生命周期管理^[28]，数据安全将一方面服务科技管理业务，保障数据的安全使用与共享，另一方面满足数据安全合规性要求，达成“进

不来、拿不来、看不懂、改不了、走不脱”的
安全管理目标。科技管理业务涵盖项目申报、
立项、过程管理、验收管理等多个阶段，每个
阶段都有其独有的业务服务场景，科技管理数

据就是这些不同的业务服务过程中产生的全部
数据，既包括科技管理和服务业务数据，也包
括系统运行维护数据。科技管理数据安全建设
技术框架如图3所示。

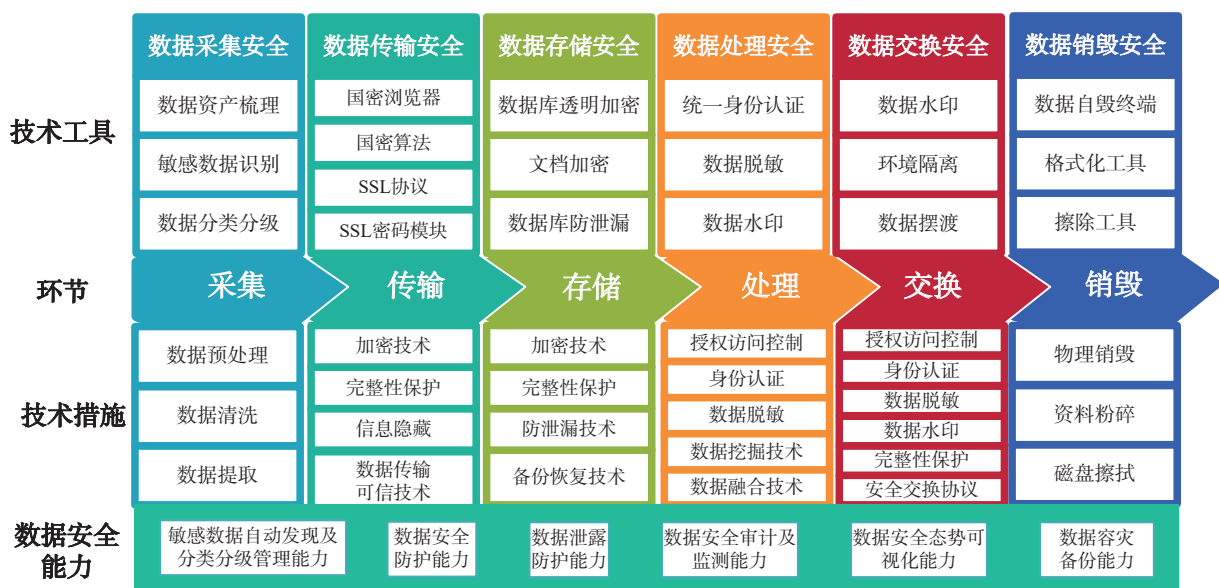


图3 科技管理数据安全建设框架

4.4.1 数据全生命周期管理

按照数据产生的环节，科技管理数据全生命周期包括收集、传输、存储、处理、交换、销毁六个阶段，以数据保密性、完整性、可用性为核心目标，秉持用户授权、最小够用、专事专用、全程防护原则，结合科技管理数据分级分类结果，对各个环节落实相关的管理措施和技术防护措施。

(1) 采集阶段。采集指获取数据的过程，包括但不限于科研单位在线填报、外部应用对接共享、公开数据获取等方式。针对科技管理数据采集阶段可能面临的非授权采集、数据分类分级不清、敏感数据识别不清和缺乏细粒度的访问控制等风险，技术上应采取传统的数据安全防护措施，包括数据传输加密、文档加密

技术、数据访问控制、安全认证等。在管理层面，应遵循合法合规、正当必要的原则，在规定的职责和范围内开展数据的采集活动，不收集与科技管理业务无关的信息。

(2) 传输阶段。传输指按照一定的规程，实现点对点之间数据的安全传输和交换的通信过程。科技管理数据传输阶段主要指数据在科研用户和科技管理部门之间，或科技管理业务平台系统间传输的过程。数据传输阶段需采用符合国密算法的密码技术构建安全的传输通道，确保传输过程中的保密性和完整性。通过身份认证、权限限制等技术措施实现传输过程中的访问控制，保证传输的节点安全；使用成熟的安全传输协议，保证传输的通道安全。

(3) 存储阶段。存储是指通过一定的计算

机技术将数据存储在服务、存储等特定的介质中，包括结构化、非结构化存储形式。科技管理数据存储面临数据分类分级不清、重要敏感数据的安全管理不足、缺乏细颗粒度访问控制等问题。主要技术措施包括对数据进行识别和分类分级、对重要数据存储提供加密手段、对存储在平台中的敏感数据提供更细粒度的访问控制能力。结合数据分级分类结果，对不同类别和级别的数据采取相应的安全存储措施。必要时，建立科技管理数据异地备份和容灾管理机制，同时采用必要的冗余策略和管理措施，确保数据备份与恢复操作过程规范，保障备份数据的有效性、可用性。

（4）处理阶段。处理指对科技管理数据进行清洗、转换、整合、服务的过程。在处理过程中，严格遵循最小必要原则，对于进行数据处理的操作人员，应当授予且仅授予必须的数据访问和操作权限，降低数据泄露的风险。数据处理时面临的安全风险包括数据缺乏访问控制、数据结果的访问接口缺乏控制、数据处理结果缺乏敏感数据保护措施、缺乏安全审计和数据溯源的能力，应加强数据细颗粒度访问控制、敏感数据标记访问控制、数据脱敏、处理过程日志记录以及数据追踪溯源技术的应用。在开展敏感数据处理时，建立安全隔离的数据加工处理环境，确保在数据安全可控的基础上配合特殊科技管理业务的顺利开展。同时，针对数据展示场景或需求，应建立敏感数据去标识化规范，使用数据脱敏、数据加密等技术对数据进行去标识化处理，降低数据泄露安全风险。

（5）交换阶段。数据交换主要是指数据在系统平台中各节点之间、各组件之间以及从数

据平台和其他外部系统之间交换的过程。针对可能存在泄露问题，应提供数据加密能力，保证数据交换时的安全，并提供相关的安全审计能力。当进行线下交换时，应使用符合数据安全要求的介质对数据进行保存，并做好介质管理，确保专事专用、及时回收。

（6）销毁阶段。销毁指删除科技管理数据及其相关备份副本的过程。当进行销毁时，应通过有效技术手段，保证数据被完整、有效删除，且不可恢复。如涉及存储介质，应采用不可恢复的方式（如消磁、焚烧、粉碎等）对介质进行销毁处理。销毁过程应进行全程监督与控制，对销毁介质的登记、审批、交接、销毁执行等过程进行监督，并保留销毁过程有关记录。

4.4.2 数据安全能力建设

考虑到不同数据安全技术可能复用于科技管理数据全生命周期各阶段，依照整体规划，统一部署的原则，科技管理数据全生命周期安全管理将从安全措施可实现的角度，规划五大数据安全能力建设。

（1）敏感数据自动发现及分类分级管理能力建设。规划建立科技管理业务敏感数据特征库，依据法律法规及组织规范，建立敏感特征与级别类别的关联，形成敏感数据自动识别能力，并对接数据自动分类分级。

（2）数据泄露防护能力建设。规划建设敏感数据泄露防护能力，实现对科技管理业务网络、终端敏感数据的监测、识别、分析、审计和阻断控制，通过在重要网络节点处（如内网与外网节点、上联区域中心节点）部署数据泄露防护工具，通过预制策略与外发的敏感数据匹配，实现敏感数据外发监测。

(3) 数据安全审计及监测能力建设。维护敏感数据全生命周期各环节日志,并实现数据操作审计,规划建立全周期化的数据流动审计能力,监督监测用户行为,防范敏感数据的违规使用,该能力将覆盖数据的采集、使用、共享、使用等阶段^[29]。

(4) 数据安全态势可视化能力建设。从科技管理业务、平台、用户和数据资源等角度,规划建设科技管理数据可视化能力,呈现数据全生命周期的安全风险态势。通过敏感数据日志信息审计能力,综合记录用户行为的情况,形成数据安全感知全景,评估数据安全风险。

(5) 数据容灾备份能力建设。科技管理数据应具备数据备份容灾功能^[30],在本地备份的基础上,需具备跨数据中心、建立异地容灾机制。

4.5 科技管理数据安全运营

开展科技管理数据安全运营^[31],全局管控数据资产、身份、权限等,结合数据分类分级结果,实现安全策略的统一管理及下发、安全风险监测、安全威胁分析、安全事件处置,达到对科技管理数据全生命周期的统一、全局的管理和掌握、数据可视化、数据风险可控的目标。例如在数据安全策略统一管理层面,统一管理数据安全策略,集中展示各个数据安全组件状态,对各个数据安全组件涉及的识别类及防护类安全策略进行统一配置,实现安全策略集中管理。在数据安全策略核查方面,制定核查策略模板,基于模板,检验、核查数据安全策略设定的完备性及执行效果。此外,基于对策略、事件、数据综合分析,全局化分析安全态势,提升科技管理数据安全运营能力。

4.6 科技管理数据安全治理监管

数据安全治理监管^[32]主要是对科技管理数据安全保障状况定期开展合规性评估,从而对数据安全保护的管理制度、安全防护措施、数据运营等方面进行有效监管。数据安全治理监管主要从安全管理、数据安全保护、数据安全运营三个方面开展评估工作。在安全管理评估层面,包括组织架构评估和管理制度评估两个方面,其中组织架构评估是对组织架构的设定完备性、岗位工作职责与分工的明确性方面进行评估;管理制度评估是指从技术管理、人员管理、流程管理多个维度对管理制度进行评估。在数据安全保护评估层面,包括数据分类分级管理与数据全生命周期管理评估两个方面,其中数据分类分级管理评估是对数据分类分级目标和原则、涉及的数据内容、分级方法和具体要求、分级结果等内容进行评估;数据全生命周期管理评估是围绕数据处理活动涉及的采集、传输、存储、处理、交换、销毁等环节的数据安全保护措施合规性、完备性进行评估。在数据安全运营评估层面,包括从事前的安全访问控制,事中的安全监测、应急响应与事件处置,事后的安全审计等维度对数据安全运营的效果进行评估。

5 总结与展望

在科技竞争背景下,科技管理数据治理是数据价值发挥、支撑科技决策、维护科技安全的基础性工作。在数据要素时代,随着科技管理数据价值的不断提升,科技管理数据治理就是以释放数据价值为核心目标,旨在提升数据

质量,保障数据安全,进一步维护科技安全。科技管理数据安全治理作为保障国家科技创新能力和科技安全的重要组成部分,面临着前所未有的机遇与挑战。随着大数据时代的到来,科技管理数据不仅成为推动科技进步的关键资源,同时也伴随着潜在的安全隐患。当前,虽然我国已出台多项法律法规强化数据安全保护,不少学者也在研究、实践中探索了一系列治理措施,但数据安全治理依然存在诸多瓶颈,如管理制度不完善、数据安全建设不到位、数据资产权责不明晰、数据流转缺乏有效监管等。新形势下,科技管理数据治理应充分认识科技管理改革新形势和科技管理数据安全合规治理的新变化新需求,坚持规范引领,从多维度、多角度、多层次作出科技管理数据安全战略规划,坚持数据安全保障与业务服务并重的基本立场,秉承合规优先、逐层推进原则,构建科学、合理、有效的数据安全治理体系,从而积极应对数据安全治理面临的困难挑战,实现平稳、有序推进科技管理数据安全治理与业务发展。

展望未来,在科技管理数据安全治理实践路径下,结合框架各部分内容,需进一步提高智能化与自动化水平,借助人工智能和机器学习等先进技术,实现数据的自动发现、分类分级及异常检测,从而提升治理效能。同时,建立统一的标准与规范,促进不同组织间的数据互信,减少安全风险。在数据隐私保护方面,平衡数据利用和个人隐私的重要性将日益凸显,促使研发更多专注于隐私保护的技术方案。此外,数据安全治理不仅是技术层面的问题,更是管理层面的任务。因此,需制定科技管理数据安全保障的战略规划和顶层设计,不断完善

相关法律法规体系;加强科技管理数据安全治理宣传教育,培养专业的人才队伍,不断提升全民的数据安全意识。只有通过政府、企业和社会界的共同努力,才能构建起一个既安全又高效的科技管理数据安全生态系统,为我国的科技创新提供坚实的数据保障。

参考文献

- [1] 糟玉庆,赵捧未,尹丽英,等.面向政府宏观科技决策的科技管理数据服务模式构建[J].科技管理研究,2023,43(2):167-176.
- [2] 李品.统筹科技发展与科技安全的情报体系框架研究[J].情报学报,2024,43(3):357-368.
- [3] 孙婕,沈恒超.美日宏观科技统筹协调机制及启示[J].世界科技研究与发展,2022,44(4):504-517.
- [4] CHAOHUI M A, RUIHUA N, HAOXIANG T, et al. Research on data schema and security in data governance[J]. Big Data Research, 2016(3): 89-101.
- [5] 梅傲,李坤佳.日本数据安全治理制度述评及其启示[J].情报理论与实践,2023,46(7):195-200.
- [6] 陈毅,华蕊.国家安全体系和能力现代化视域下数据安全治理的困境及突围路径[J].学习与探索,2023(12):41-47.
- [7] 吕明元,弓亚男.我国数据安全治理发展趋势、问题与国外数据安全治理经验借鉴[J].科技管理研究,2023,43(2):21-27.
- [8] 李雪莹,张锐卿,杨波,等.数据安全治理实践[J].信息安全研究,2022,8(11):1069-1078.
- [9] 王庆德,吕欣,王慧钧,等.数据安全治理的行业实践研究[J].信息安全研究,2022,8(4):333-339.
- [10] 王玉,安鹏,栗文科,等.政务数据安全治理体系研究与实践[J].信息安全研究,2023,9(9):900-907.
- [11] 朱洪斌,安龙,杨铭辰.电力大数据安全治理体系研究[J].电信科学,2019,35(11):140-145.
- [12] 杨超,郭刚,叶林佳,等.工业互联网数据安全治理实践[J].信息安全与通信保密,2022(9):18-27.
- [13] 侯鹏,李智鑫,张飞,等.金融数据安全治理智能化技术与实践[J].网络与信息安全学报,2023,9(3):174-187.

- [14] 原磊. 平台企业数据安全治理研究 [J]. 世界社会科学, 2024(1): 103-118.
- [15] 朱佳妮, 赵丽梅. 基于主权区块链的科学数据安全治理研究 [J]. 科技管理研究, 2023, 43(9): 171-176.
- [16] 程伟, 马成, 凌捷. 大数据技术在数据安全治理中的应用 [J]. 大数据, 2023, 9(6): 3-14.
- [17] 许杰, 张锋军, 陈捷, 等. 面向大数据环境下的数据安全治理技术 [J]. 通信技术, 2021, 54(12): 2659-2665.
- [18] 胡剑, 戚湧. 基于区块链跨链机制的政务数据安全治理体系研究 [J]. 现代情报, 2023, 43(9): 85-97.
- [19] 陈媛媛, 赵晴. 数据利他: 全球治理观下基于公共利益的数据共享机制 [J]. 图书馆论坛, 2023, 43(5): 71-80.
- [20] 盛小平, 郭道胜. 科学数据开放共享中的数据安全治理研究 [J]. 图书情报工作, 2020, 64(22): 25-36.
- [21] 李晓东, 董少平, 吴菁. 总体国家安全观视域下数据安全治理的路径选择 [J]. 中国科技论坛, 2024(2): 147-157, 167.
- [22] 徐婧欣, 郭丰, 苏鹏. 数据分类分级政策演化研究 [J]. 图书馆, 2023(2): 48-55.
- [23] 陈永刚, 赵增振, 陈岚. 政务数据安全合规评估要点及实践 [J]. 信息安全研究, 2022, 8(11): 1050-1054.
- [24] ZHOU F, HUANG J. Cybersecurity data breaches and internal control[J]. International Review of Financial Analysis, 2024(93): 67-75.
- [25] 程啸. 论个人信息权益 [J]. 华东政法大学学报, 2023, 26(1): 6-21.
- [26] 徐玉梅, 王欣宇. 我国重要数据安全法律规制的现实路径——基于国家安全视角 [J]. 学术交流, 2022(5): 37-48.
- [27] 马费成, 熊思玥, 孙玉姣, 等. 数据分类分级确权对数据要素价值实现的影响 [J]. 信息资源管理学报, 2024, 14(1): 4-12.
- [28] 段尧清, 姜慧, 汤弘昊. 政府开放数据全生命周期: 概念、模型与结构——系统论视角 [J]. 情报理论与实践, 2019, 42(5): 35-40.
- [29] 戴荣峰, 陶晓英, 于萌, 等. 大数据驱动下的数据全生命周期安全监测方法 [J]. 信息安全研究, 2023, 9(12): 1226-1232.
- [30] CHAUDHARY J, VYAS V, SAXENA M. Backup and Restore Strategies for Medical Image Database Using NoSQL[J]. Communication, Software and Networks, 2023(10): 161-171.
- [31] 刘志勇, 何忠江, 阮宜龙, 等. 大数据安全特征与运营实践 [J]. 电信科学, 2021, 37(5): 160-169.
- [32] 梅傲, 陈子文. 总体国家安全观视域下我国数据安全监管的制度构建 [J]. 电子政务, 2023(11): 104-115.



开放科学
(资源服务)
标识码
(OSID)

污水处理行业产学研专利产出时空演变研究

封丽¹ 郭泊池² 石亮³ 李伟⁴

1. 河海大学图书馆 南京 210098;
2. 南京奥凯知识产权服务有限公司 南京 210000;
3. 上海海事大学图书馆 上海 200135;
4. 新疆水利水电科学研究院 乌鲁木齐 830049

摘要: [目的/意义] 我国污水处理行业地区创新发展和产学研创新产出不均衡现象较为突出,解构专利产出驱动因素,揭示其时空演变特征,为促进行业创新区域均衡发展提供决策依据。[方法/过程] 基于专利文本分析、LMDI 分解模型和空间自相关分析方法对我国 2000—2023 年污水处理行业产学研专利授权量的空间动态演变及驱动因素进行分析,并探讨主要驱动效应在省区尺度的空间差异特征。[局限] 受数据所限,没有开展研发经费投入、区域水平等因素影响下的驱动效应分析。[结果/结论] 研究表明 2015 年开始专利数量增幅明显增大,2020 年达到峰值,C02F1/00、C02F1/28 分类下的专利申请量增幅显著。节能环保产业规模效应和污水处理行业份额效应是驱动产学研专利授权量增加的主要因素。产学研内驱效应有明显差别,企业内驱效应在研究期内正负交替,院校和科研单位在大部分时期以负效应为主,研究期末表现出明显的正效应。企业和院校的内驱效应在省区空间尺度上呈现显著聚集特征。

关键词: 驱动效应; 空间计量; 污水处理; 产学研; 专利

中图分类号: G306

Study on the Spatial and Temporal Evolution of Industry-University-Research Institute Patent Output in Sewage Treatment

FENG Li¹ GUO Bochi² SHI Liang³ LI Wei⁴

1. Library of Hohai University, Nanjing 210098, China;
2. Nanjing Aokai Intellectual Property Services Co., Ltd., Nanjing 210000, China;
3. Library of Shanghai Maritime University, Shanghai 200135, China;
4. Xinjiang Institute of Water Resources and Hydropower Science, Wulumuqi 830049, China

基金项目 中国博士后科学基金面上项目“环境污染突发事件的时空数据挖掘及协同治理机制研究”(2019M650108);江苏省高校哲学社会科学研究一般项目“科技创新团队视角下技术多元化对专利颠覆性及其创新绩效的影响研究”(2024SJYB0038);江苏省高校图工委教改研究课题一般项目“AIGC 驱动下工科院校大学生知识产权素养教育实施路径研究与探索”(2024JTYB07)。

作者简介 封丽(1982-), 硕士, 馆员, 主要研究方向为数据分析与挖掘;郭泊池(1988-), 学士, 工程师, 主要研究方向为知识产权信息服务;石亮(1984-), 博士, 馆员, 主要研究方向为知识组织, E-mail: liangshi@shmtu.edu.cn;李伟(1980-), 博士, 正高级工程师, 主要研究方向为水资源节约利用。

引用格式 封丽, 郭泊池, 石亮, 等. 污水处理行业产学研专利产出时空演变研究[J]. 情报工程, 2024, 10(6): 93-105.

Abstract: [Objective/Significance] The imbalance between regional innovation development and Industry-University-Research Institute innovation output in China's sewage treatment industry is quite prominent. Deconstructing the driving factors of patent output, revealing its spatiotemporal evolution characteristics, and providing decision-making basis for promoting regional balanced development of industry innovation. [Methods/Processes] Based on the patent text analysis, LMDI decomposition model, and spatial autocorrelation analysis method, this paper analyzes the spatial dynamic evolution and driving factors of Industry-University-Research Institute patent authorization in China's sewage treatment industry from 2000 to 2023, and explores the spatial differences of the main driving effects at the provincial scales. [Limitations] Due to data limitations, no analysis was conducted on the driving effects of factors such as research and development funding investment and regional economic level. [Results/Conclusions] Research has shown that since 2015, the number of patents has increased significantly, reaching its peak in 2020. The number of patent applications under categories C02F1/00 and C02F1/28 has also increased significantly. The scale effect of the energy-saving and environmental protection industry and the share effect of the sewage treatment industry are the main factors driving the increase in the number of patent authorizations. There should be a significant difference in the internal driving effect of Industry-University-Research Institute. The internal driving effect of enterprises alternates between positive and negative during the research period, while the negative effect is mainly observed in universities and research institutions during most of the period. At the end of the research period, there is a clear positive effect. The internal driving effects of enterprises and universities exhibit significant clustering characteristics at the provincial scale.

Keywords: Driving Effect; Spatial Econometric Analysis; Sewage Treatment; Industry-University-Research Institute; Patent

引言

污水处理技术的创新与应用是深入打好水污染防治攻坚战的重要保障,对于全面落实“绿水青山就是金山银山”的发展理念,改善城镇人居环境,促进城市治理体系和治理能力的现代化,推动经济高质量发展具有重要的作用。

《“十四五”城镇污水处理及资源化利用发展规划》中明确要求,到2025年,全国城市生活污水的集中收集率应力争超过70%,县城的污水处理率达到95%以上^[1]。随着“十四五”规划的推进实施,各地多措并举,积极推进污水处理行业创新发展^[2-4]。专利是评估一个地区或机构科技研发能力及技术创新程度的关键性指标^[2]。从地区分布看,我国污水处理领域的专利正处于迅猛发展的阶段,但主要集中在东部地区,中西部地区相对落后^[5]。从创新主体看,产学研(企业、院校、科研单位)三大类型中,

企业授权专利占比最高,根据 incoPat 数据库统计,2000—2023年,污水处理行业产学研三大创新主体的专利授权量比值为81:15:4。可见,我国污水处理行业地区创新发展不均衡、产学研创新产出不均衡现象较为突出,在此背景下,本文以我国污水处理行业专利授权量为数据信息,从产学研三大创新主体专利产出角度,分析专利授权量的空间动态轨迹,解构专利授权量增减的主要驱动因素,并阐明主要驱动因素在省级区域尺度上的空间差异特征,以期准确把握污水处理行业创新产出的时空演变趋势,为制定针对性创新扶持政策,促进污水处理行业区域均衡发展提供借鉴参考。

1 相关研究

目前在专利产出影响因素研究领域,传统回归模型(Ordinary Least Squares, OLS)、空

间滞后模型 (Spatial Lag Model, SLM) 是常用的计量分析工具^[6-8]。为避免经典回归分析方法中的“伪回归”及“内生性”问题,刘志迎等^[9]较早地构建了反映长江经济带专利产出变化的LMDI分解模型,综合考量区域经济结构、区域研发强度、区域创新效率及整体经济发展4个驱动因素对长江经济带创新产出的影响。LMDI (Logarithmic Mean Divisia Index) 驱动效应分解模型最初由Ang^[10-11]提出,因为其可以有效解决残差项和零值问题,被广泛地运用于能源消耗和环境污染的因素分解研究领域^[12-14]。在刘志迎等探索应用于专利分析领域之后,蔡中华等^[15]利用LMDI模型对我国PCT国际专利申请量的增长驱动因素进行了分解,将其归因于规模效应、结构效应和强度效应。陈爱萍等^[16]基于LMDI模型,采用中国上市公司绿色技术专利申请数据,构建包括绿色专利研发优先序、绿色专利份额和研发活动规模的因素分解模型。魏莱等^[17]基于黄河流域绿色专利申请数据,运用LMDI与M-R模型从时间与空间两方面分析经济增长、产业结构、能源结构、碳强度、研发效率对黄河流域绿色技术创新的影响。与传统影响因素回归分析方法相比,LMDI方法在辅助宏观政策制定方面体现出较强的适用性^[16]。

基于专利数据自身具有的地理空间属性,采用空间自相关分析技术对该数据进行空间计量分析,发掘揭示空间上的聚集性和相关性,是专利计量分析的常用方法。张玉明等^[18]将专利数据应用于空间自相关分析,以揭示不同省份区域间创新产出的空间相关模式。此后,空间自相关广泛应用于旅游^[19]、风电^[20]、绿色低

碳^[21]等不同产业创新水平空间分布特征研究。

从现有研究看,尽管LMDI模型和空间分析模型分别被广泛用于研究专利产出驱动因素和专利地理位置属性的空间演变特征,但将两种方法结合起来,特别是综合利用LMDI分解模型和空间分析模型来探讨驱动因素空间差异性的研究还鲜有报道。已有学者从大学视角进行专利产出的空间相关性分析^[7, 22],但从产学研三大主体创新产出视角开展的专利空间动态特征对比研究还比较少见。在污水处理行业,虽然已有部分学者运用专利分析方法研究本行业创新发展态势^[2-5, 23],但没有从产学研视角细化专利授权主体类型,也未深入开展驱动因素分解及其空间相关性研究。基于此,本文以2000—2023年污水处理行业专利授权量为研究对象,采用专利文本分析法、LMDI分解模型和空间自相关分析方法,分析产学研三类创新主体专利授权量的空间动态演变特征及驱动影响因素,并探讨主要驱动效应在省区尺度的空间差异特征。

2 研究方法和数据来源

2.1 研究方法

2.1.1 专利文本分析法

专利文本分析是专利计量分析常用的方法^[24-26],其核心在于利用数据分析工具对专利文本数据进行处理与分析,旨在提炼出其中的有用信息。研究者可采纳文本分析技术对专利文献中的关键词语、技术属性及叙述内容进行分析,以此来辨识技术特征传递的隐性信息和演变趋势。

2.1.2 驱动因素分解方法

本文研究的创新主体有三类：企业、院校、科研单位。首先依据 Kaya 恒等式将某类创新主体专利授权量 (Q_i) 分解如下：

$$Q_i = \frac{Q_i}{W} \times \frac{W}{T} \times T \quad (1)$$

式中： Q_i 代表第 i 类创新主体专利授权量； W 代表污水处理行业专利授权总量； T 代表节能环保产业专利授权总量。 $\frac{Q_i}{W}$ 表示第 i 类创新主体专利授权量占污水处理行业专利授权总量的比重，比重越大，表明此类创新主体在行业内创新产出主导性越强，该指标是反映创新主体自身主观能动性的内驱因素； $\frac{W}{T}$ 表示污水处理行业专利授权量占节能环保产业专利授权总量的比重，污水处理是节能环保产业中的核心领域^[4]，该指标从专利产出角度反映了污水处理行业的重要性，代表污水处理行业在节能环保产业中的份额； T 代表节能环保产业专利授权总量，反映了节能环保产业的创新研发活动总规模。综上，式 (1) 表示产学研各类创新主体专利授权量受到创新主体自身内驱因素、污水处理行业份额以及节能环保产业创新总体规模等三类因素的共同影响。

依据 LMDI 分解模型， t 期相对于 $t-1$ 期第 i 类创新主体专利授权量的变化量 ΔQ_i 可以表示为：

$$\Delta Q_i = \Delta Q_{i1} + \Delta Q_{i2} + \Delta Q_{i3} \quad (2)$$

$$\Delta Q_{i1} = \frac{Q'_i - Q^{t-1}_i}{\ln(Q'_i / Q^{t-1}_i)} \times \ln\left(\frac{Q'_i / W_t}{Q^{t-1}_i / W_{t-1}}\right) \quad (3)$$

$$\Delta Q_{i2} = \frac{Q'_i - Q^{t-1}_i}{\ln(Q'_i / Q^{t-1}_i)} \times \ln\left(\frac{W_t / T_t}{W_{t-1} / T_{t-1}}\right) \quad (4)$$

$$\Delta Q_{i3} = \frac{Q'_i - Q^{t-1}_i}{\ln(Q'_i / Q^{t-1}_i)} \times \ln\left(\frac{T_t}{T_{t-1}}\right) \quad (5)$$

式中： ΔQ_{i1} 表示创新主体内驱因素变化引起的专利授权量变化，定义为内驱效应； ΔQ_{i2} 表示污水处理行业份额变化引起的专利授权量变化，定义为份额效应； ΔQ_{i3} 表示节能环保产业创新活动规模变化引起的专利授权量变化，定义为规模效应。综上，创新主体专利授权量的变化可以通过内驱效应、份额效应和规模效应三方面的综合影响来分解，如果分解效应为正值，则起到正向促进作用，反之，起到负向抑制作用。

2.1.3 空间自相关分析方法

空间自相关是指同一个变量在不同空间位置上的相关性。当某个变量的相似性在空间上聚集在一起的时候，就表现为正的空间自相关性；相反，当某个变量的差异性在空间上聚集在一起的时候，就表现为负的空间自相关性。空间自相关分为全局自相关和局部自相关。全局自相关用于描述一个现象的整体分布，判断该现象在空间上是否具有聚类特征，但不能准确指出其集中的位置。局部空间自相关用于衡量聚合空间单元相对于整个研究区域的空间相关性是否足够显著。

全局莫兰指数 (Moran's I) 是用以描述区域空间特征要素全局自相关的常用指标，计算公式为：

$$I = \frac{n \sum_i \sum_j \omega_{ij} (a_i - \bar{a})(a_j - \bar{a})}{(\sum_i \sum_j \omega_{ij}) \sum_i (a_i - \bar{a})^2} \quad (6)$$

式中， n 为研究对象总数； i 和 j 代表第 i 和第 j 个研究对象； ω_{ij} 是研究对象的空间权重； a_i 和 a_j 是研究对象的空间属性值； $\bar{a}$ 是全部研究对象空间属性值的平均值。得出莫兰指数后，需要进行显著性检验，常假设变量的空间分布

服从正态分布，通过服从标准正态分布的 z 统计量来判断地区间的空间相关性，计算公式为：

$$z = \frac{I - E(I)}{\sqrt{Var(I)}} \quad (7)$$

式中， $E(I)$ 是莫兰指数的期望， $Var(I)$ 是方差。在 $p=0.1$ 的检验水准下，当 $|z|>1.65$ 时，表明地理事物之间存在空间相关性。

本文采用局部莫兰指数分析专利产出的驱动效应在全国各省区之间的空间聚集特征。局部莫兰指数 I_i 主要用于验证研究对象是否存在相似值或相异值的局部集聚现象，计算公式如下：

$$I_i = \frac{(a_i - \bar{a}) \sum_j \omega_{ij} (a_j - \bar{a})}{\sum_j (a_i - \bar{a})^2 / n} \quad (8)$$

若计算所得 I_i 为正，表示区域 i 与其临近区域均为高值区，属于高高 (H-H) 聚集或者区域 i 与其临近区域均为低值区，属于低低 (L-L) 聚集，即区域 i 存在相似值的空间聚集。若计算所得 I_i 为负，表示区域 i 与其临近区域呈相反的值，区域 i 为高值则其临近区域为低值，属于高低 (H-L) 聚集，反之则属于低高 (L-H) 聚集，即区域 i 存在相异值的空间聚集。

2.2 数据来源与处理

本文的数据来源于 incoPat 科技创新情报平台。专利类型限定为授权发明专利和实用新型专利，时间为 2000 年 1 月至 2023 年 12 月，省区范围包括 31 个省、自治区和直辖市（不含港澳台）。污水处理行业专利按照国际专利分类 IPC 号检索，共计 372316 件，按照 incoPat 数据库申请人类型划分功能，可进一步获取企业、院校、科研单位的专利数量。在 incoPat 数据库自有的新兴产业分类中检索节能环保产业专利，

共计 2571650 件。基于上述数据，采用 EXCEL 进行专利文本分析和驱动因素分解计算，采用 ArcGIS 自带的 toolbox 工具箱进行空间自相关分析，数据分析结果图件采用 EXCEL 绘制。

3 结果分析

3.1 总体特征分析

3.1.1 总体技术分布特征

表 1 展示了我国污水处理行业的技术分布，揭示了污水处理技术研究的热点领域，可以看出，C02F1/00 分类下的水、废水或污水处理技术的研究集中度居首位，其专利申请数量显著超过其他技术领域。紧随其后的是 C02F1/28 吸附法、C02F101/30 有机化合物、C02F1/52 悬浮杂质的絮凝或沉淀，以及 C02F1/44 渗析法、渗透法或反渗透法等技术分支，这些领域的专利申请量也颇为可观，均超过 3 万件，表明这些技术分支同样是中国污水处理技术研究的重要方向。

3.1.2 省区技术分布特征

表 2 所示为我国污水处理行业主要省区的技术分布，可以看出，江苏省在污水处理行业申请的专利数量最多，相较于其他地区专利数量遥遥领先。广东省专利数量占据第二，浙江省占据第三。其次为山东省、北京市、上海市、安徽省、四川省、湖北省和天津市，分别位列第四至第十。从国内各省区的技术分布来看，各省区的专利申请仍主要集中在 C02F1/00，各省区均在 C02F1/00 布局的专利数量最多，其次是 C02F1/28，与国内整体技术分布基本一致。

表 1 我国污水处理行业技术分布

分类号	类别	数量
C02F1/00	水、废水或污水的处理	72128
C02F1/28	吸附法（离子交换法入 C02F1/42; 吸附剂的组成入 B01J）	47600
02F1/30		38210
C02F1/52	悬浮杂质的絮凝或沉淀	37506
C02F1/44	渗析法、渗透法或反渗透法	37099
C02F9/04	至少有一个化学处理步骤	25574
cC02F9/14	至少有一个生物处理步骤	25067
C02F101/20	重金属或重金属化合物	23253
C02F101/16	氮的化合物，例如氨	21154
C02F1/72	氧化法	21056

表 2 我国主要省区技术分布统计表

IPC 分类	江苏	广东	浙江	山东	北京	上海	四川	湖北	福建	安徽
C02F1/00	8103	7377	4763	5186	2319	2423	2051	1994	2032	2063
C02F9/04	6208	4655	3581	3669	2532	1775	1889	1495	1558	1314
C02F9/14	5668	4524	3578	2789	3575	2008	1926	1605	1457	1185
C02F9/02	5052	5715	3793	3067	1486	1572	1516	1164	1541	1315
C02F1/28	4211	2911	2347	2386	1956	1359	1060	1131	962	1064
C02F1/44	3927	4185	3174	2183	2202	1531	875	686	941	565
C02F1/52	3775	2125	1898	2419	1723	1194	1085	983	771	997
C02F101/30	3359	2273	1579	1546	1786	1067	752	835	600	665
C02F3/34	2469	1857	1449	1244	1852	917	560	727	506	479
C02F9/00	2173	1843	1132	1330	973	752	601	614	475	498

3.1.3 专利数量趋势

通过对污水处理行业的专利年申请量进行统计分析，国内专利总体申请趋势如图 1 所示。总体来看，2000 年后污水处理行业专利申请数量呈上涨趋势，其中 2000—2014 年专利年申请量呈稳步上升的缓慢发展趋势，2015 年国务院印发《水污染防治行动计划》，制定了水污染防治的工作目标及指标等，进一步推动了污水处理的发展，因此自 2015 年开始专利申请量涨幅明显增大，污水处理行业也步入快速发展期，

2020 年达到高峰。2021—2022 年专利年申请量小幅下降。

3.1.4 技术演进趋势

图 2 是我国污水处理行业近十年技术演进趋势图，可以看出，国内污水处理行业在 C02F1/00- 水、废水或污水的处理，C02F1/28- 吸附法，C02F101/30- 有机化合物，C02F1/52- 悬浮杂质的絮凝或沉淀，C02F1/44- 渗析法、渗透法或反渗透法，C02F101/20- 重金属或重金属化合物等技术方向的专利申请量基本呈现逐年

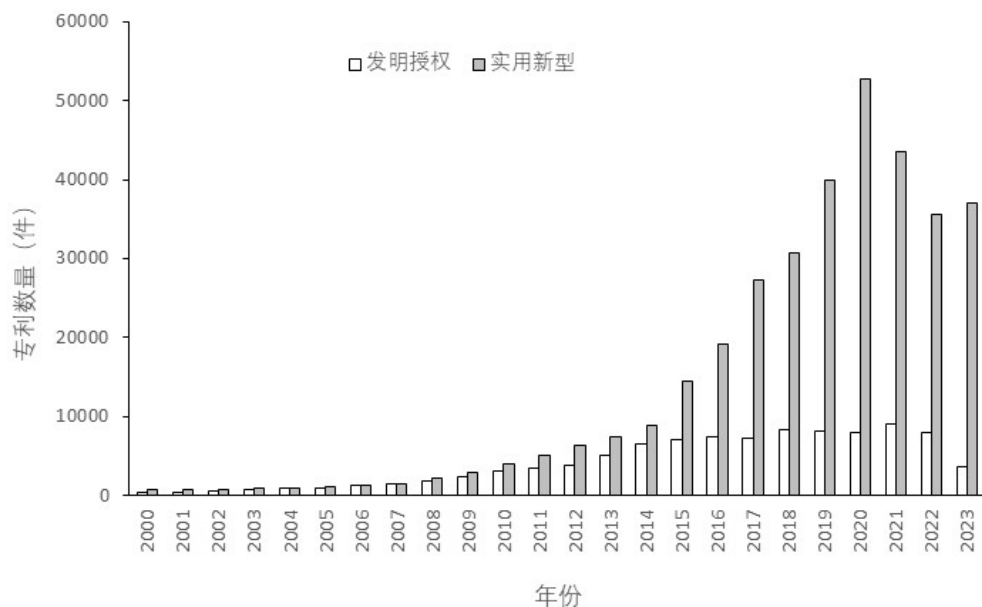


图1 我国污水处理行业专利申请趋势

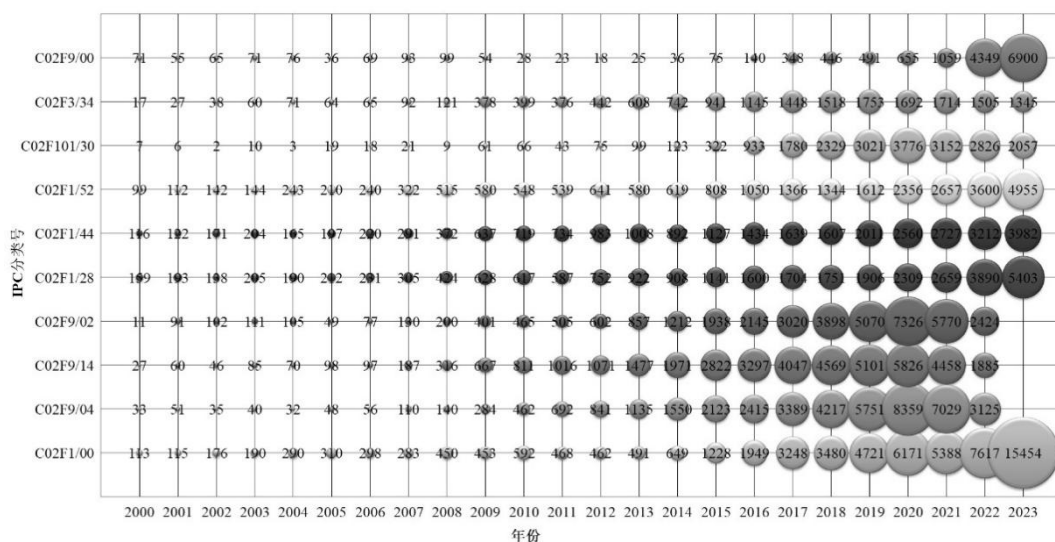


图2 我国污水处理行业技术演进趋势分析

上升的趋势，尤其是 C02F1/00、C02F1/28 近两年专利申请量涨幅显著，这反映出近两年的研究热点主要集中在这两个领域。

3.2 驱动因素分解分析

2000—2023 年企业、院校、科研单位专利授权量变化的 LMDI 驱动因素分解结果如图 3—

5 所示。

企业是污水处理行业专利授权的主体。企业授权专利数量在 2000—2021 年持续增加，特别是从 2011 年开始，专利年授权量超过 1000 件，2021 年专利授权量超过 15000 件，随后的 2022—2023 年，专利授权量下降明显。从驱动效应上看，企业授权专利数量的增加主要受到

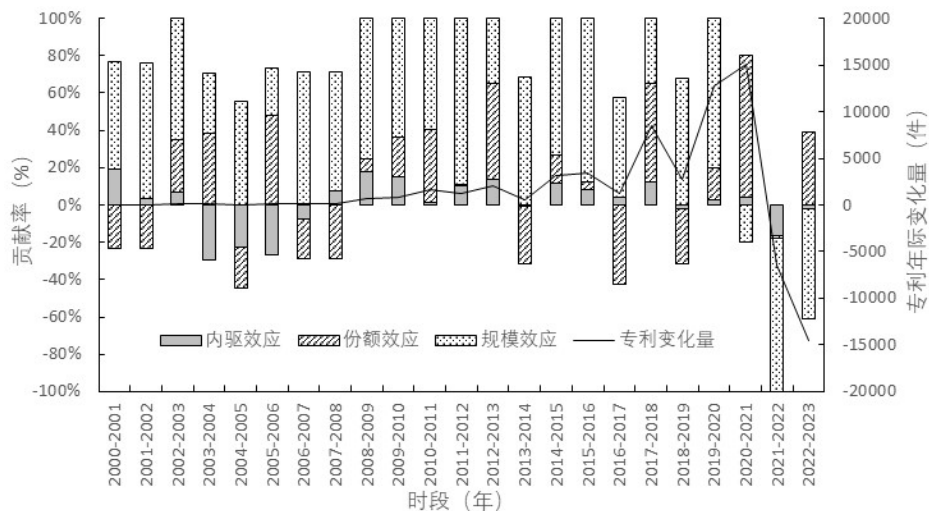


图3 企业专利授权量变化驱动效应分解

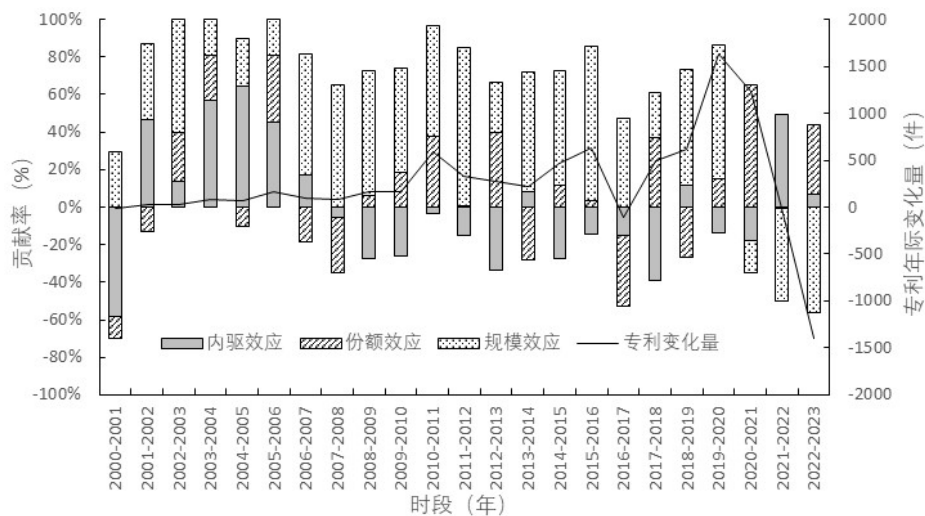


图4 院校专利授权量变化驱动效应分解

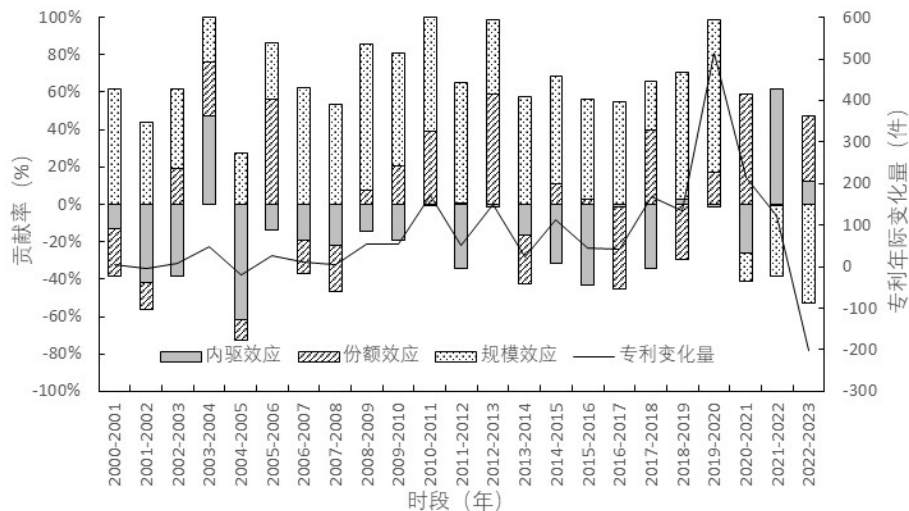


图5 科研单位专利授权量变化驱动效应分解

节能环保产业规模效应和污水处理行业份额效应的正向驱动。企业自身内驱效应在各阶段表现不一，分阶段看，在研究初期，2007 年之前，企业内驱效应不足，在一定程度上抑制了专利数量的增加，2008 至 2021 年，企业内驱效应显著增强，大幅促进了专利数量的增加，但近年来受疫情影响，企业生产活力减弱，其内驱效应由正转负，导致了专利数量的减少。

院校在污水处理行业专利授权数量变化趋势与企业类似，但在数量级上与企业有明显差距，2020 年授权数量达到最高，超过 1500 件，约占企业授权数量的十分之一。节能环保产业规模和污水处理行业份额仍然是院校专利授权量增加的主要驱动因素。院校自身内驱效应在 2002 至 2007 年表现强劲，对专利数量增加发挥了显著的正效应，但在 2008 至 2021 年的绝大多数年份驱动效应为负，发挥了一定的抑制作用，近两年院校自身内驱效应又由负转正，体现了较强的正效应。

科研单位专利授权量的变化趋势与院校较为相似，同样在 2020 年授权量达到最大，超过 500 件，但其数量是产学研三类主体中最小的。

从整个研究期看，2021 年以前，节能环保产业规模效应始终为正，是推动科研单位专利授权量增加的主要因素，相比之下，科研单位自身内驱效应在绝大多数年份为负值，是制约科研单位专利授权量增加的主要因素。近两年，在节能环保产业专利授权量大幅下降的背景下，科研单位自身内驱效应由负转正，对于维持污水处理行业专利授权量发挥了积极作用。

3.3 内驱效应空间差异性分析

从上述分析可以看出，企业、院校、科研单位三类创新主体的内驱效应具有明显差异性。为深入探究各创新主体内驱效应在各省区之间的差异性，以省区专利数据为样本深入分析各创新主体的内驱效应空间分布特征。为了降低专利数据的年际波动影响，对各省区逐年的原始数据进行了平均化处理，即使用“十五”“十一五”“十二五”“十三五”“十四五”这五个时期的年均值来进行不同时期的驱动效应分析。基于上述数据采用 LMDI 驱动因素分解方法，得到各省区各类创新主体的内驱效应，按照内驱效应的正负分类统计省区个数，结果见表 3。

表 3 省区产学研内驱效应类型统计表

		单位：个			
创新主体	效应类型	“十五”至“十一五”	“十一五”至“十二五”	“十二五”至“十三五”	“十三五”至“十四五”
企业	正效应	13	28	26	23
	负效应	18	3	5	8
院校	正效应	22	7	6	10
	负效应	9	24	25	21
科研单位	正效应	6	5	10	14
	负效应	25	26	21	17

从表 3 结果可以看出，在“十五”至“十一五”期间，院校在 22 个省区表现出了正效应，促进

了这些省区专利数量的增加。与之相比，科研单位在大部分省区表现为负效应，企业的正负

效应基本持平。在“十一五”至“十二五”期间，企业正效应的省区大幅增加至 28 个，企业内驱效应促进了这些省区专利数量的增加。相比之下，院校和科研单位在大部分省区表现为负效应，后续“十三五”“十四五”时期基本延续了这一特征。总体而言，省区企业内驱效应以正向为主，省区院校和科研单位内驱效应以负向为主，在研究期末即“十四五”时期，院校和科研单位为正效应的省区有增加趋势。

对企业、院校、科研单位三类创新主体的内驱效应分别进行全局空间自相关计算分析。针对各个时期分别计算莫兰指数值和对应的 Z 值、P 值，结果见表 4。可以看出，企业、院

校、科研单位三类创新主体在各个时期的内驱效应空间自相关性均通过了显著性检验($Z>1.65$, $P<0.1$)，表明具有显著的空间聚集特征，即内驱效应大的省区聚集在一起，低的省区聚集在一起。进一步对比分析莫兰指数绝对值的大小，可以看出，企业内驱效应和院校内驱效应的莫兰指数均呈现先增大后减少趋势，在“十二五”至“十三五”期间达到最大，说明这一时期的空间集聚特征最为显著。科研单位内驱效应的莫兰指数一直呈现增加趋势，但是数值均小于企业内驱效应和院校内驱效应的莫兰指数，说明科研单位内驱效应的空间聚集特征有所增加，但显著程度较弱。

表 4 不同时期产学研内驱效应的全局莫兰指数

创新主体	统计变量	“十五”至“十一五”	“十一五”至“十二五”	“十二五”至“十三五”	“十三五”至“十四五”
企业	Moran's I	0.2366	0.2859	0.3449	0.2366
	z 值	2.5188	3.0141	3.6203	2.7081
	p 值	0.0118	0.0025	0.0003	0.0068
院校	Moran's I	0.2193	0.2636	0.3721	0.2506
	z 值	2.4192	2.8729	3.9582	2.7306
	p 值	0.0156	0.0041	0.0001	0.0063
科研单位	Moran's I	0.0114	-0.0381	0.1228	0.1315
	z 值	1.7789	1.6641	1.6761	1.9438
	p 值	0.0863	0.0964	0.0899	0.0519

全局莫兰指数代表了整个区域的空间相关性，为了深入揭示各个省区之间的空间聚集特征，进一步采用局部莫兰指数对各类创新主体内驱效应进行空间差异性分析，结果见表 5。

从企业内驱效应空间聚集特征看，“十五”至“十一五”时期，高高聚集区位于广西，表明这一地区企业授权专利数量显著增加，到了“十二五”“十三五”时期，高高聚集区转移

到了东部的山东、江苏、浙江、上海，表明这些东部省区企业正向内驱效应表现突出。到了“十三五”至“十四五”时期，高高聚集区又转移到了山西、河南等中部省区。与此同时，低低聚集区从东部的江苏、上海、浙江、安徽等省区逐步向西部内地转移，到了“十二五”至“十三五”时期，停留在宁夏、甘肃、青海、四川四个省区，表明这些西部内地省区企业专利数量增加不明显。

表 5 不同时期产学研内驱效应的空间聚集区分布

创新主体	空间聚集类型	“十五”至“十一五”	“十一五”至“十二五”	“十二五”至“十三五”	“十三五”至“十四五”
企业	高高	广西	山东、江苏、上海、浙江、安徽	山东、江苏、上海、浙江	山东、山西、河南
	高低				
	低高		河南、江西、福建	安徽、福建	上海、安徽
	低低	江苏、上海、浙江、安徽	西藏	宁夏、甘肃、青海、四川	
院校	高高	山东、江苏、上海		甘肃、青海、西藏	
	高低	重庆	河南、安徽、江西、福建	河南、安徽、福建	山西、上海
	低高	吉林、辽宁、安徽			
	低低	贵州、广西	山东、江苏、上海、浙江	山东、江苏、上海、浙江	山东、河南、安徽
科研单位	高高			甘肃、四川	
	高低	吉林、内蒙古	内蒙古、河北、山东、江苏	河北、安徽	安徽
	低高	四川、广东			
	低低	辽宁、天津、河北		山东、江苏、上海、浙江	上海

从院校内驱效应空间聚集特征看，高高聚集区在“十五”至“十一五”时期，位于山东、江苏、上海，到了“十二五”至“十三五”时期，位于甘肃、青海、西藏。低低聚集区从西南的贵州、广西逐步转向东部的山东、江苏、上海、浙江，继而到中北部的河南、安徽。这些省区院校数量多，但授权专利数量增加不明显。“十一五”至“十三五”时期，安徽、福建两省处于相对稳定的高低聚集区，表明在这一时期，其周边省份江苏、浙江、广东等地院校专利产出明显不足。

从科研单位内驱效应空间聚集特征看，各省区科研单位内驱效应明显弱于企业和院校。高高聚集区仅出现在了“十二五”至“十三五”时期的甘肃、四川。低低聚集区在“十五”至“十一五”时期位于辽宁、天津、河北，在“十二五”至“十三五”时期位于山东、江苏、上海、浙江。这些东部省区科研单位授权专利数量增加不明显。

4 结论与建议

本文基于专利文本分析法、LMDI 分解模型和莫兰指数空间计量对 2000—2023 年污水处理行业产学研三类创新主体专利授权量的空间动态演变特征及影响因素进行分析，并探讨了主要驱动效应在省区尺度的空间差异特征。得到以下主要结论：

（1）江苏、广东、浙江占据我国污水处理行业专利数量的前三位，C02F1/00 分类下的水、废水或污水处理技术的研究集中度居首位，紧随其后的是 C02F1/28 吸附法。自 2015 年开始专利申请量涨幅明显增大，污水处理行业也步入快速发展期，2020 年达到高峰，尤其是 C02F1/00、C02F1/28 近两年专利申请量涨幅显著。

（2）产学研三类创新主体的专利授权量在 2010—2020 年均有显著增加趋势，特别是在“十三五”中后期，专利授权量大幅增加，这种趋势与国家在生态环境保护、污水处理及资源循

循环利用方面的一系列政策密不可分,特别是党的十九大以来,生态文明建设加快推进,《水污染防治行动计划》逐步落实,为行业创新发展,加快专利布局提供了政策保障。2021年以来,受新冠疫情、外部经济环境等影响,三类主体的年度专利授权量有所下降,其中企业授权量下降最为突出。相比之下,院校和科研单位授权量略有下降,客观上避免了污水处理行业专利授权量大幅减少,保障了该行业创新发展的多元化和可持续性。

(3)从驱动因素分解结果看,节能环保产业规模的发展壮大是产学研各主体专利授权量增加的主要驱动因素,在节能环保产业中,污水处理行业份额效应呈现一定的波动性,但总体上呈现正效应。三类主体内驱效应有明显差别,企业内驱效应在研究初期以负效应为主,中后期表现出明显的正效应,近期又转为负,这一变化趋势也反映了企业创新活动的周期性。院校和科研单位在大部分年份以负效应为主,但近两年表现出明显的正效应,这在一定程度上缓解了企业创新动力不足的问题,保障了污水处理技术的持续创新发展。

(4)从省区层面产学研内驱效应差异性看,企业内驱效应以正向为主,院校和科研单位内驱效应以负向为主,在研究期末,即“十四五”时期,院校和科研单位为正效应的省区有增加趋势。从企业内驱效应空间聚集特征看,“十五”至“十四五”时期,高高聚集区从广西移到了东部的山东、江苏、浙江、上海,继而又转移到了山西、河南等中部省区。从院校内驱效应空间聚集特征看,高高聚集区从东部的山东、江苏、上海,转移到甘肃、青海、西藏。从科研单位内驱效应空间聚集特征看,各省区科研

单位内驱效应明显弱于企业和院校。

上述结论反映出我国污水处理行业专利创新产出在各省区、产学研创新主体之间不平衡的问题。为促进污水处理行业区域均衡发展,推进产学研多主体创新,提出如下对策建议:

(1)我国污水处理行业创新产出已经呈现出一定的空间聚集特征,这表明我国的区域创新产出有明显的空间近邻溢出效应,相邻地区之间的创新行为相互影响。因此,应充分发挥高高聚集区的带头引领作用,利用区位优势、空间依赖等地理条件,继续加大区域间交流,扩大创新辐射范围,增大强度,推进高低聚集区向高高聚集区演化,实现区域双赢。

(2)省区之间污水处理行业专利产出不均衡现象较为明显,总体看东南、华北地区要优于西北、西南地区。应鼓励西北地区、低低聚集区通过设立人才飞地、设置创新平台分中心、地方研发基地等形式积极引进创新领先地区的技术力量,同时选派科技人员到创新领先地区进修交流,培育壮大本地区创新主体,逐步实现低低聚集区向高低聚集区演进。

(3)企业是污水处理行业创新的主要力量,与之相比院校和科研单位创新产出较为薄弱,这一方面与创新主体数量紧密相关,企业数量远大于院校和科研院所,另一方面也反映了院校、科研单位创新主观能动性不足问题。政府应出台鼓励政策,加大院校、科研单位成果转化力度,促进产学研各主体在污水处理行业的合作,将院校、科研单位的人才优势与企业自身发展需求相结合,推进污水处理行业的产学研联合创新。

本研究也存在以下三点局限:一是受数据

资料所限,驱动因素分解过程中以专利授权量作为主要数据源进行分解分析,未能考虑产学研各创新主体研发经费投入、各省区国民经济水平等其他影响创新产出的因素;二是本次研究主要聚焦省级区域,没有将地市或县区级等更精细的空间单元作为研究对象,在空间特征分析的精细化程度上有所不足;三是本研究没有考虑产学研合作网络下的专利产出特征。今后,考虑研发经费投入、区域经济水平、产学研合作网络等影响下的驱动因素分解及精细化空间特征分析研究有待拓展深化。

参考文献

- [1] 国家发展改革委,住房和城乡建设部.“十四五”城镇污水处理及资源化利用发展规划[EB/OL].(2021-06-06)[2024-2-15].https://www.ndrc.gov.cn/xxgk/zcfb/ghwb/202106/t20210611_1283168_ext.html.
- [2] 许美芳,黄晓东,魏家香,等.基于专利分析视角的浙江省污水治理技术态势研究[J].浙江农业学报,2018,30(12): 2121-2128.
- [3] 许美芳,金李杰,楼昭涵,等.长三角一体化视角下的污水治理专利计量分析[J].浙江农业科学,2021,62(7): 1428-1432, 1460.
- [4] 周运迪,赵晗,邱传虹,等.江苏省污水处理及其再生利用专利分析研究[J].江苏科技信息,2022,39(24): 30-34.
- [5] 王森薇,雷一东.我国污水处理领域专利地图分析[J].复旦学报(自然科学版),2013,52(1): 99-104.
- [6] 于伟,张鹏.我国省域专利授权分布及影响因素的空间计量分析——基于2007—2009年统计数据的实证研究[J].宏观经济研究,2012(6): 83-86.
- [7] 沈能,周晶晶.我国大学专利产出的空间分布及趋同特征——基于空间计量分析[J].科研管理,2013,34(S1): 133-139.
- [8] 张明倩,柯莉.“一带一路”跨国专利合作网络及影响因素研究[J].软科学,2018,32(6): 21-25, 29.
- [9] 刘志迎,陈侠飞.基于LMDI模型的长江经济带创新产出变化的驱动效应分析[J].研究与发展管理,2016,28(6): 61-70.
- [10] ANG B W. The LMDI approach to decomposition analysis: a practical guide[J]. Energy Policy, 2005(33): 867-871.
- [11] ANG B W. LMDI decomposition approach: a guide for implementation[J]. Energy Policy, 2015(86): 233-238.
- [12] CHO J H, SOHN S Y. A novel decomposition analysis of green patent applications for the evaluation of R&D efforts to reduce CO₂ emissions from fossil fuel energy consumption[J]. Journal of Cleaner Production, 2018(193): 290-299.
- [13] FUJII H, MANAGI S. Trends and priority shifts in artificial intelligence technology invention: A global patent analysis[J]. Economic Analysis and Policy, 2018(58): 60-69.
- [14] FUJII H. Decomposition analysis of green chemical technology inventions from 1971 to 2010 in Japan[J]. Journal of Cleaner Production, 2016(112): 4835-4843.
- [15] 蔡中华,余薇佳.我国PCT国际专利申请量增长的驱动因素分析——基于LMDI方法[J].科技管理研究,2016,36(20): 143-148.
- [16] 陈爱萍,陈会英.基于LMDI的中国上市公司绿色专利分解分析[J].情报杂志,2020,39(9): 103-109, 32.
- [17] 魏莱,彭佑元.黄河流域绿色技术创新时空演化分解分析——来自绿色专利数据的证据[J].河南科学,2023,41(11): 1687-1696.
- [18] 张玉明,李凯.中国创新产出的空间分布及空间相关性研究——基于1996-2005年省际专利统计数据的空间计量分析[J].中国软科学,2007(11): 97-103.
- [19] 王毅.我国旅游产业科技创新能力时空演变研究[D].南京:南京师范大学,2015.
- [20] 张建宇.风力发电产业专利技术研发概况、研发热点与演进路径研究[D].呼和浩特:内蒙古大学,2020.
- [21] 彭聪,吕拉昌.中国低碳技术专利的时空格局及影响因素——基于中国284个地级及以上城市数据的实证分析[J].技术与创新管理,2023,44(4): 424-434.
- [22] 桂还官尚.近十年我国高校专利产出数量的空间差异性分析[D].兰州:西北师范大学,2017.
- [23] 王曼娜,陈晨,陈晓芬,等.我国再生水领域的研究和应用趋势——基于博硕士论文与专利的分析[C]//中国环境科学学会.2014中国环境科学学会学术年会论文集.成都:四川大学,2014: 1167-1174.
- [24] 赵阳.面向企业技术创新的专利技术信息挖掘与分析研究[J].现代情报,2020,40(12): 84-89.
- [25] 朱佳林.我国高校图书馆知识产权服务研究热点及前沿可视化分析[J].图书馆工作与研究,2022(2): 77-84, 95.
- [26] 秦奋,高健,宋妙茹.双碳背景下基于区位熵方法的绿色专利文献计量研究[J].图书馆工作与研究,2024(1): 80-88.



开放科学
(资源服务)
标识码
(OSID)

钱学森情报思想研究的可视化计量分析

王琳 陈一谕

杭州电子科技大学中国科教评价研究院 浙江 310018

摘要: [目的/意义] 钱学森是我国杰出的战略科学家和航天事业的奠基人,同时他在情报学和情报事业研究领域也作出了重大贡献。其情报思想在我国学术界和业界影响极为深远。钱学森 1983 年发表《科技情报工作的科学技术》的学术论文,是其情报学理论研究进入思想化体系化阶段的重要标志。四十年来,钱学森情报思想研究日益繁荣。本文探究钱学森情报思想的研究现状,更好地深化对钱学森情报思想研究的理解和把握,并展望其未来的发展趋势,从而为今后的研究提供支持和参考。[方法/过程] 利用文献计量学方法,使用 CiteSpace 可视化工具对中国知网 1983—2023 年间有关钱学森情报思想研究的文献进行定量分析。[结果/结论] 钱学森情报思想研究总体上呈递增之势,课题热度逐渐升高,呈现出良好的发展态势。主要研究方向发生了变迁,热点词汇从“思维科学”“情报学”“情报思想”“学科发展”等转向“情报感知”“大数据”“悟性思维”,显示研究重点从情报概念辨析、学科建设指导原则等学科基础理论研究转向了对其思想在数智场景的落地应用和方案设计。钱学森情报思想研究也存在一些局限,如钱学森情报思想选题同质化严重,跨学科研究较少,有显著影响力的学者数量较少,研究机构呈现出高度集中的特点,这不利于钱老思想的传播和发展,不利于争鸣的展开。需要更加重视搭建跨学科研究的桥梁,拓宽研究视野,并加强研究机构之间的协作与交流。

关键词: 钱学森; 情报学; 情报事业; 文献计量学; 可视化

中图分类号: G35; G2

Visualized Bibliometric Analysis of Qian Xuesen's Information Science Thoughts Research

WANG Lin CHEN Yiyu

Chinese Academy of Science and Education Evaluation, Hangzhou Dianzi University, Zhejiang 310018, China

Abstract: [Objective/Significance] Qian Xuesen is an outstanding strategic scientist and founding figure of China's aerospace industry. Simultaneously, he made significant contributions to the field of intelligence studies and intelligence operations. His intelligence ideology has had a profound influence on both academic and industrial sectors in China. In 1983, Qian Xuesen published the academic paper *Scientific Technology of Science and Technology Intelligence Work*, marking an important milestone

基金项目 国家社科基金重点项目“健康中国战略下信息贫困群体的健康焦虑及疏导行动研究”(21ATQ005)。

作者简介 王琳(1979-), 博士, 教授, 主要研究方向为情报学理论与方法, 信息资源管理学理论。陈一谕(2001-), 硕士研究生, 主要研究方向为信息行为, E-mail: wanglinpk@163.com。

引用格式 王琳, 陈一谕. 钱学森情报思想研究的可视化计量分析[J]. 情报工程, 2024, 10(6): 106-118.

in the development of his intelligence theory into a systematic framework. Over the past forty years, research on Qian Xuesen's intelligence ideology has flourished. This paper explores the current research status of Qian Xuesen's intelligence ideology, aiming to deepen our understanding and grasp of his intelligence ideology research while also envisioning its future development trends. It thereby provides support and reference for future research endeavors. [Methods/Processes] This paper employs bibliometrics methodology, utilizing the CiteSpace visualization tool, to conduct a quantitative analysis of the literature related to Qian Xuesen's intelligence ideology research on the CNKI platform from 1983 to 2023. [Results/Conclusions] Research on Qian Xuesen's intelligence ideology has shown an overall increasing trend, with the topic's popularity gradually rising, indicating a promising development trajectory. There has been a shift in the main research directions, with hot topics transitioning from concepts like "cognitive science" "intelligence studies" "intelligence ideology" and "disciplinary development" to "intelligence perception" "big data" and "intuitive thinking". This shift reflects a move from fundamental theoretical studies such as conceptual analysis of intelligence and guiding principles for disciplinary construction towards the practical application and program design of his ideology in data-driven scenarios. However, research on Qian Xuesen's intelligence ideology also faces limitations, such as significant topic homogeneity, limited interdisciplinary studies, a small number of influential scholars, and a highly concentrated research institution landscape, which hinders the dissemination and development of Qian's ideology and inhibits open debates. There is a need to prioritize building interdisciplinary research bridges, broadening research perspectives, and enhancing collaboration and communication among research institutions.

Keywords: Qian Xuesen; Information Science; Information Enterprise; Bibliometrics; Visualization

引言

钱学森是我国两院资深院士，航空航天事业的开拓者，两弹一星事业的奠基人。钱老也被誉为“我国国防科技情报事业的导师”^[1]。钱老是“在我国最关心、支持和重视科技情报工作的科学家，也是发表有关科技情报工作见解和论述最多的科学家”^[2]。他归国之后为建设我国科技情报事业，尤其是国防科技情报事业作出了重大贡献，使我国科技情报工作在较短的时间内实现了体系化和现代化，并形成了以情报研究为代表的一整套具有中国特色的情报工作内容与方法。改革开放以后，钱老将情报学研究作为关注重点，1983年7月2日，钱学森在国防科技情报工作会议上作了关于《科技情报工作中的科学技术》的报告，并成文发表，提出了科技情报是一门学问而不仅是一项工作，

情报是激活、活化的知识，情报学属于思维科学等著名学术观点，对情报学基本概念和基本理论问题作出了深刻阐述，确立了情报学的学科地位，明确了情报学的目标与任务，为新时期情报学学科建设打开了新的局面。这篇论文也成为其情报学理论研究进入思想化体系化阶段的重要标志。从20世纪80年代情报学学科建设到90年代后提出的“将人的思维、思维结果、知识、智慧以及各种情报、信息集成起来”、从定性到定量的综合集成思想，无不昭示着钱老情报学思想的深邃。其情报思想在我国传播和影响之深远，尚未有人能及。

从20世纪80年代以来，钱学森情报思想研究热度一直不减，尤其是近年来随着建立中国特色、中国风格、中国气派的情报学自主知识体系的热潮，钱学森情报思想作为该知识体系重要的思想源泉，研究日益繁荣。钱学森情

报思想所蕴含的新型思维方式和思维体系,是数智赋能情报学和情报工作的重要基础^[3]。深入研究、努力传播践行钱学森情报思想,对推进数智时代情报学理论创新和科技情报工作发展具有重要的现实意义^[4]。本文对中国知网1983—2023年间有关钱学森情报思想的研究进行文献计量学分析,使用 CiteSpace 工具对40年来钱学森情报思想研究的文献进行系统性的梳理与量化探究,旨在从不同角度深入了解该研究的发展历程,把握该领域的研究现状,展望其未来的发展趋势,更好地深化对钱学森情报思想研究的理解,传承和发展钱学森情报思想。

1 文献收集与研究工具

本文利用中国知网进行文献检索。设定“网络”“图书馆”“数据”“知识”“智能”“钱学森”“情报”“信息”“思维科学”等为检索词。在知网数据库中进行主题检索(检索时间为2023年6月20日),共筛选出177条结果。文献出版的时间跨度为1983—2023年。将177篇文献以 Refworks 的形式输出并保存起来,在 CiteSpace 中进行数据分析。

本研究运用文献计量学研究方法,以

CiteSpace6.2.R4 为分析工具,从多元、分时和动态的视角对钱学森情报思想研究的文献集合进行深入研究。以 CiteSpace 所产生的关键字共现视图、聚类视图、时区视图等所获得的相关信息为依据,通过对文献相关数据的深入挖掘,将钱学森情报思想研究的现状展现出来,通过对其研究热点与研究方法的分析,发现当前的热点、重点与新兴研究领域及其动态演变。

2 数据分析

2.1 总体趋势与特点

为了解四十年来钱学森情报思想研究之整体走向,笔者将文献按照发表年份进行统计,着重探讨其中具有重大意义的转折点和关键点。本研究先以177篇论文的发表年代为基础,对其发文量进行分析,得到钱学森情报思想研究文献发表时间分布图,如图1所示。可以看出,关于钱学森情报思想的研究总体上是呈递增之势,特别是2010年以后有了显著的增长。从2010到2023年,年均发文量为9篇,尤其是2020年的发文量达到了16篇,体现了该领域良好的发展态势。

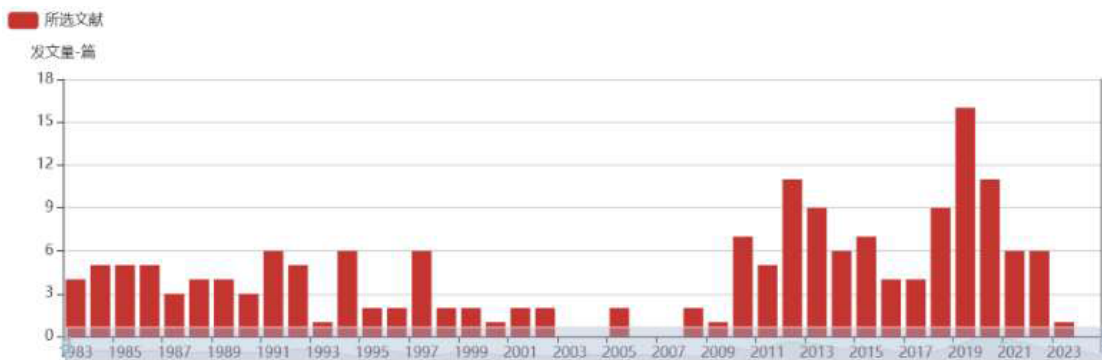


图1 钱学森情报思想研究文献发表时间分布

钱学森情报思想研究之所以长年不衰，主要原因是钱老的情报学思想具有前瞻性和预见性。集中体现出钱老作为战略科学家的情报战略眼光，以及综合集成和知识融通的特点。钱学森科技情报理论不仅将情报学作为一门独立学科来看待，对情报学进行了准确的学科定位，而且开创性地拓展了情报理论的深度和广度，为情报学的发展提供了广阔空间，这是其他情报理论所不具备的特质^[5]。不同时代的学者，总能从钱老的情报学和情报工作的相关著述中找到重要的思想源泉，从中获得启发，继而对当时的情报学研究主题做出创新性的研究或论释，不断丰富学科的内涵，拓展学科的边界，使情报学理论大厦建设得更加坚固，更加具有时代性。从情报学的决策属性，情报的知识观，综合继承方法论到灵境，钱老的这些思想观点依次照进了学科不同阶段的发展现实之中，繁荣了信息分析与预测、智库研究、知识管理、军民融合新大情报观、情报综合集成研讨厅、元宇宙等研究领域或主题，使情报学成为了数字时代的当代学科。在图书情报与档案管理一级学科更名为信息资源管理后，面对学科名和学科体制结构的重大调整，如何一方面更好地坚守情报初心，继承和发展好学科的核心知识和方法，稳固情报学在二级学科中的地位，另一方面积极创新，主动求变，构建面向总体国家安全观和新时代国家战略的现代情报学理论体系，是摆在情报学界面前的重大课题，非常有必要从钱学森情报学思想中继续汲取力量，实现数智赋能环境下情报科学的守正创新。深入研究钱学森情报

研究思想，探寻其中的方法论价值，对于提升情报研究工作的层次和质量具有重要的现实意义^[6]。

从图1中可以观察到，1983—1997年间的发文量较为平均。1997—2009年发文量较少。2009年10月，钱老逝世，情报学界开始系统地整理和回顾钱老给我们留下的思想遗产。从2010年开始，关于钱学森情报思想研究的文献数量呈现出显著的增长趋势，并在2019年达到峰值。从时间节点上看，2019年恰逢钱老去世10周年，情报学界掀起了对钱老情报思想挖掘发现和重新认识其时代价值的新一轮热潮，各种学术研讨交流活动层出不穷，学界不仅是纪念钱老这位国防科技情报导师，而且是在复杂多变的外部环境中努力寻求对钱学森情报学思想体系的继承，并以此力图消除当前情报学发展中的迷惘与徘徊^[7]。此外，近十年来，大数据、物联网、云计算、人工智能等新一代重要技术的迅猛发展，使情报学原有的物质基础—人类信息基础环境发生了重大变迁。这必然使情报学学科产生新的变革，学科需要构建适应新环境需要的学科理论体系。而这方面缺乏可供参考的理论资源，钱学森的综合集成思想和大成智慧学恰好可为新环境下情报学理论大厦建设提供理论支持与指导原则，是现成可借鉴的思想库与方法库。钱学森以人为主、人机（网）结合、从定性到定量的综合集成思想为情报学实现这种革命性转变提供了可能的方法和思路。这方面的深入研究是值得情报学界关注的，成果的涌现也是可以预期的^[8]。大成智慧思想对于未来情报工作的创新发展具有重要意义^[9]。因此基于钱学森综合集成思想的情报学研究逐

渐成为方兴未艾的研究热点，这同时也促进了情报学与系统科学、思维科学和智能科学的进一步融合，从学科交叉中生成新的知识领域和知识生长点，进一步拓展了学科的边界与范围，增强和提升了学科的核心知识结构，相关研究成果不断涌现，研究论文数量和质量双双攀升。

元宇宙的兴起改变了信息内容生态，为情报学发展带来了新的机遇与挑战。赛博空间、元宇宙以及泛在智能符合钱学森所描绘的未来信息世界，或如钱老所说的“智界”和“灵境”。大数据与钱学森情报思想具有一定的内在联系，钱学森情报思想被赋予了新的内涵和意义^[10]。

3 钱学森情报思想研究的可视化

在此基础上,本研究利用 CiteSpace 中多种类型数据来可视化地研究主题词、聚类,研究钱学森情报思想研究的主题词、演化、突现等领域知识结构特点。

3.1 关键词共现

首先,运行 CiteSpace 获得有关钱学森情报思想研究的关键词共现图,如图 2 所示。关键词共现图可以显示关键词之间的关系和共现频率,通过观察关键词共现图,可以识别出重要或热门的主题及相关概念。

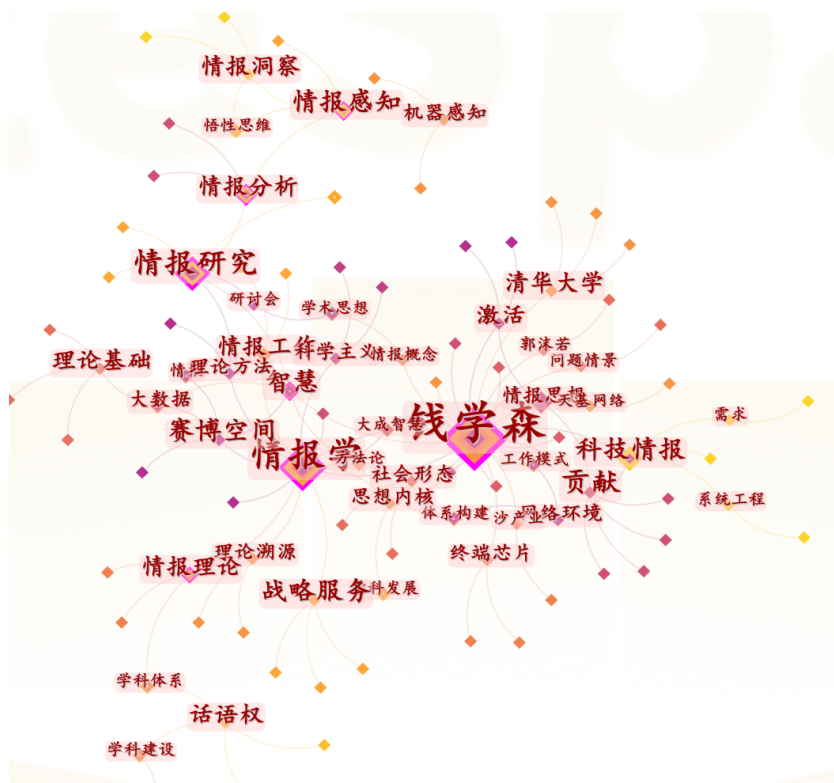


图2 涉及钱学森情报思想的研究文献的关键词共现

其中以关键词为中心的直线代表着关键词间的关联,字体愈大代表着关键字出现的频率愈高。从分析图表中不难看出,涉及钱学森情

报思想研究文献的关键词主要有：“情报研究”“情报学”“科技情报”“情报感知”和“情报理论”。其中情报研究又称信息分析，是指

根据用户的特定需求,对情报进行针对性地选择、分析、综合、预测,为用户提供系统、准确、及时、大流量的知识与情报的科学研究活动,综合集成探讨厅作为情报研究的新方法而受到重视。“情报理论”和“情报学”是钱学森情报思想研究的重头戏,“钱学森情报学思想之用”是在继承钱学森情报学思想体系的基础上,用钱学森情报主张的指导方法去解决情报工作中现存的问题,或者将钱学森情报学思想应用于情报学新兴领域的探索^[7]。体现着对学科建设的重视。而“科技情报”研究是钱学森情报思想的最初源头。情报感知将主动察觉环境要素、处理信息的“感”与清楚理解、诠释、展望的“知”相融合,使其成为应对复杂巨系统问题和探索创新的有力抓手^[10]。这些都是情报学科的重点领域。

图2中的“科技情报”“战略服务”是我国情报学科的历史悠久的研究方向。随着信息社会的建设和发展,情报学科也有了新的研究方向。由关键词共现图展现,就战略服务而言,钱学森一直非常重视科技情报的战略服务功能,提出要“能站在高处,远眺信息大洋,能观察到洋流的状况,察觉大势,做出预见”。他对情报“耳目、尖兵和参谋”的作用进行过详细阐述。近年来学术界大力倡导的“南京共识”可视作对情报战略服务功能的发展。2017年10月29日,中国科学技术情报学会、中国社会科学情报学会联合发布了《情报学与情报工作发展南京共识》。“南京共识”重新定位情报学科的发展目标:在国家创新驱动发展战略与总体国家安全观的指导下,从国家经济、社会发展与人民安全的需要出发,将科技情报、社科情报、军事情报、安全情报等联为一体,形成

大情报科学,促进各情报领域的相互融合与相互支持,实现军(军事情报、安全情报等)民(科技情报、社科情报等)情报学的融合。这与钱学森情报思想中的主导逻辑是完全吻合的:即充分发挥情报学和情报工作对国家安全和发展的预警、决策和引领作用。同时,钱学森情报思想研究的论文主题已经扩散到了“赛博空间”“大数据”等近年来的新兴热点。这反映了学术界对于钱学森思想在数字化时代的应用和探索。将钱学森思想与信息化、网络化发展趋势相结合,探讨了情报工作在数字化环境下的新思路和新方法。

由关键词的学科分布可以看出,近年来钱学森情报思想的研究热点较为集中,跨学科研究较少。研究热点虽从角度和思考方式有了转变,但与其他学科的跨学科研究动向并不明显。

钱学森情报思想研究的文献机构共现图展示了发文机构之间的联系,如图3所示可以分析出机构之间的连接情况。较大的字体表明该机构有较多的出版物。由此可见,当前对于钱学森情报学思想的研究中,“北京大学信息管理系”“中国国防科技信息中心”“北京市科学技术情报研究所”“杭州电子科技大学”等机构的发文频次较高。这些机构在对于钱学森情报思想的研究上发挥着重要作用。其中钱老曾多次前往中国国防科技信息中心视察指导,钱学森情报思想在中国国防科技信息中心被不断深化、探索。北京大学信息管理系的王延飞教授对钱老思想深入研究,提出了情报感知理论与方法,使北京大学成为了钱学森情报学术思想研究的另一个中心。由图3可以看出,发文机构的分布呈现出高度集中的特点。主要集

中在国家和北京地区科研院所以及双一流的高校。各研究机构之间的跨区域合作虽然较少，

但已有的合作较为稳定，如中国国防科技信息中心与杭州电子科技大学中国科教评价研究院。

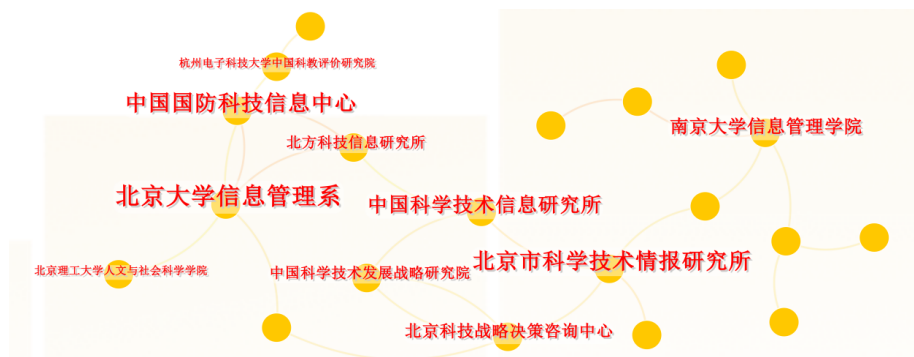


图3 钱学森情报思想研究的文献机构共现

钱学森情报思想研究的相关文献作者共现图显示了发表相关文献的主要作者，如图4所示，以作者为中心的连线表明了作者之间的联系，字体越大表示作者发表的文章越多。由此可见，当前对于钱学森情报学思想的研究中，吴晨生、卢胜军、刘如、李辉、栗琳等作者的发文频次较高，表明了这部分作者在对钱学森情报思想的研究中扮演了重要的角色。李辉、吴晨生来自北京市科学技术情报研究所。王延飞来自北京大学信息管

理系。卢胜军、栗琳、袁有雄、史秉能皆来自钱老曾多次指导的中国国防科技信息中心。

分析文献作者共现可以发现，钱学森情报思想研究中影响力显著学者数量较少，且多分布于重点科研院所和双一流高校，这不利于钱老思想的传播和发展，不利于争鸣的展开。原因可能在于对于钱学森情报学思想研究选题同质化严重，有待进一步开拓思路、进行话语创新。

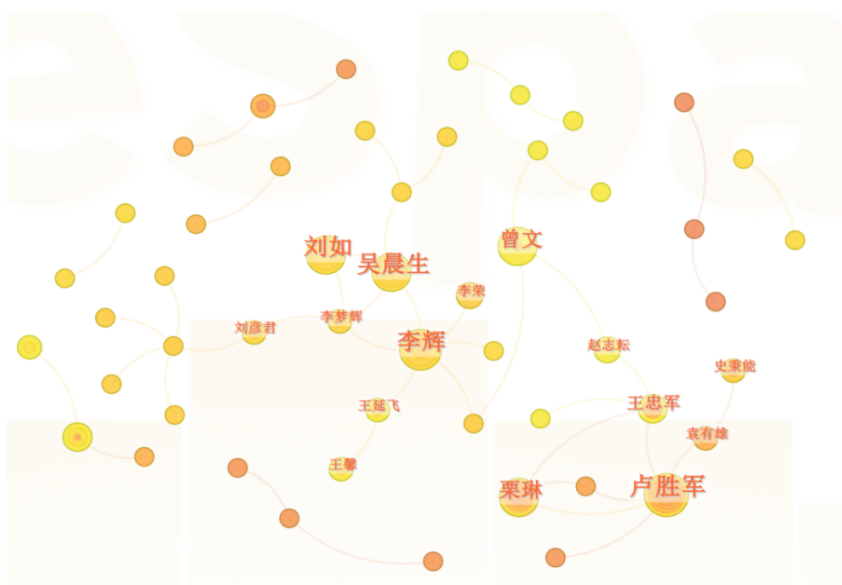


图4 钱学森情报思想研究相关文献作者共现

3.2 聚类视图

通过使用 CiteSpace 软件，我们能够快速地把钱学森情报思想的相关研究文献按照其所涉及的主题词进行聚类，生成关于钱学森情报思

想研究的主题词聚类视图，如图 5 所示，接着，我们将全部文献按照 8 个聚类用表格的方式展示，可得出聚类信息总结表（表 1），在表格中，聚类编号与聚类大小成反比，即 0 # 为最大聚类。

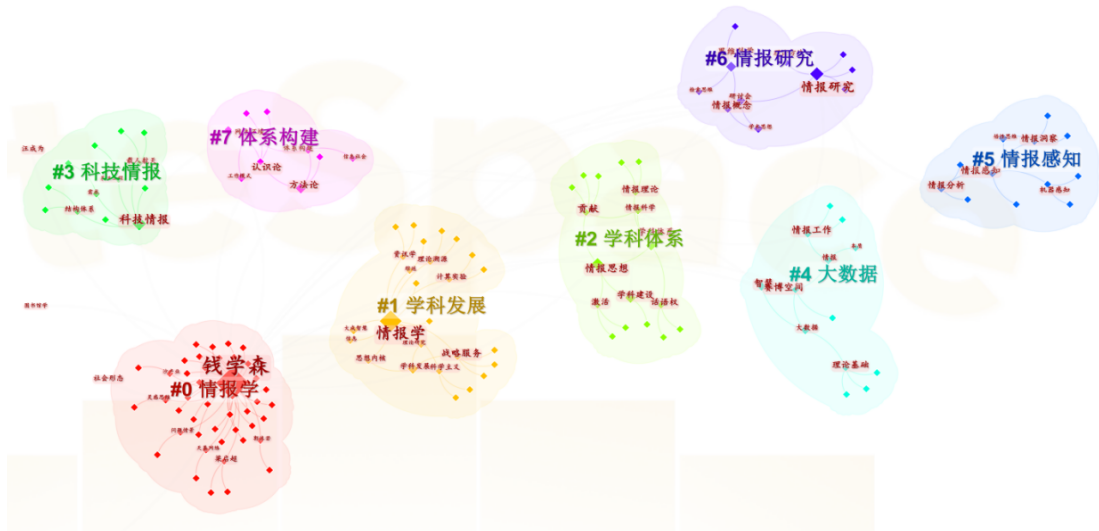


图 5 中国知网 177 篇关于钱学森情报思想研究文献关键词聚类视图

表 1 当前对钱学森情报思想研究的聚类统计

聚类编号（ClusterID）	聚类名称	标题（terms）
0	情报学	灵感思维；问题情景；天基网络
1	学科发展	情报学；战略服务；理论溯源；计算实验
2	学科体系	情报思想；情报理论；学科建设；贡献
3	科技情报	结构体系；载人航天；系统工程
4	大数据	赛博空间；理论基础；情报工作；智慧
5	情报感知	情报感知；情报分析；情报洞察；机器感知
6	情报研究	情报研究；情报概念；思维科学
7	体系构建	认识论；方法论；网络环境

根据表 1 的数据，“#0 情报学”是相关研究规模最大的聚类，该聚类中的关键词有“灵感思维”“问题情景”“天基网络”等。其中“天基网络”作为情报收集和传输的基础设施，其高速和广域覆盖特性使得情报数据可以更快速、更可靠地传输，从而加快情报分析和决策的过

程。“问题情景”指代情报学科是面向国家战略的研究领域、解决实际问题。这表明，情报学已开始重视理论要联系实际。比起理论推演，实践落地同样至关重要。

在“#1 学科发展”聚类中出现了“情报学”标题。赛博空间、大数据、元宇宙等学科热点

的出现,带动了情报学学科的热度。一级学科的更名也昭示着情报学科建设在数智赋能时代下有了新方向、新光景。在数据和信息易得而情报难求的时代,情报学的初心和使命就是要不失时机地把数据和信息转化为情报,从而服务于决策,为国家安全战略和发展战略做出自己应有的贡献^[11]。

“#2 学科体系”聚类中的标题“情报理论”,探讨了情报学的基本原理,并提供了指导情报工作的理论框架。理论溯源追溯了情报理论的发展历程和演变过程,从最初的军事情报分析到现在多领域应用的演化。钱学森情报思想在作为中国特色情报学理论体系重要组成部分的同时,钱老曾指出的“科技情报工作建资料库、检索系统,情报研究要用智能机等现代化发展路径”等思想对我国科技情报探索数智赋能时代下的发展方向有极大的指导价值。

“#3 科技情报”聚类的出现,说明了钱老对于科技情报事业、工作的重大历史贡献。科技情报领域是钱老最早投入心血的情报思想领域。在新的历史条件下,钱学森科技情报学术思想仍然是科技信息工作谋划发展、破解难题、推动建设的锐利思想武器,对于推动新形势下科技信息工作创新发展具有十分重要的指导意义^[13]。

“#4 大数据”聚类昭示着大数据带来的理论基础变迁。钱学森情报思想指明了情报研究必须与智能技术的研究相结合。钱老在过去几十年领导和指导科技情报工作的历程中,一直围绕如何提高获取情报信息的质量和速度问题,利用科学技术的发展特别是情报信息技术的发展,对科技情报工作不断地提出新的要求和发

展目标^[14]。而大数据技术提供了强大的信息采集、存储和处理能力,能够处理大规模的异构数据源,从中提取有价值的信息,为情报工作提供支持。大数据环境下情报学理论创新、科技情报服务与工作模式的转型成为研究的前沿问题,热度至今不减。与大数据相伴的云计算、物联网、人工智能、移动互联网也同样影响着情报学和科技情报工作的发展^[15]。

“#5 情报感知”聚类是由王延飞创立的体系。钱学森在国防科技情报工作会议上作了著名的“科技情报工作的科学技术”讲话,全面系统地阐述了关于情报工作中的“系统辨识”方法。钱学森所述的系统辨识过程与情报感知过程都是主动双向、反复修正调整的过程,其目的就是通过对数据反复的意义构建,使情报感知的结果无限趋近于还没有发生的未来。而情报洞察,对于情报专业人员来说,洞悉描述了获取注意和感知过程所需原始数据的方法,也描述了情报人员寻找信息从而满足需求的过程^[16]。

“#6 情报研究”作为钱老开创的中国特色领域概念界定,是当前主要研究内容和主题。在大数据环境下,利用大数据技术和工作,收集多方开源数据,运用数据分析和数据挖掘技术,对整理后的数据进行探索分析、统计分析、模式挖掘等。发现数据中的潜在模式、关联关系和趋势,通过聚焦基于赛博空间的综合集成研讨厅,提供更准确、全面的情报信息。

“#7 体系构建”体现了情报学科从“认识论”“方法论”“网络环境”等方面不断构建的重要元件。“认识论”涉及如何感知情报、识别情报来源的过程。“方法论”指探究通过

建立科学有效的方法和技术，以收集、分析和利用情报数据。在“网络环境”快速变迁的情况下，需要发展和采用新的工具和技术，例如应用大数据分析、机器学习和人工智能等技术来处理海量的信息。建设现代化情报工作体系是当前的热点之一。

通过观察八大聚类的时区视图，我们发现“情报感知”“情报洞察”等主题词的出现，反映出近几年来钱学森情报学思想的研究的重点，而且这种趋势也在关键词共现中体现，此外，时区视图还揭示出，随着时间的推移，钱学森情报学思想研究的热点也在不断发生变化，如

图6所示。前期研究热点集中在“情报学”“情报研究”以及“学科发展”等领域，其核心在于情报学的学科基础理论的构建。近年来研究热点转向“科技情报”“大数据”“情报感知”等。随着情报问题的复杂性和综合性的增加，钱学森情报学思想与新技术、新理论、新思想的融合，不断迸发出生机和活力。“科技情报”热点的复兴，大数据作为新兴技术在情报工作中的应用，“情报感知”作为创新理论的价值和指导意义，都映射出钱学森情报思想研究能够与时俱进，具有在新社会环境和新信息环境下不断创新的潜力。

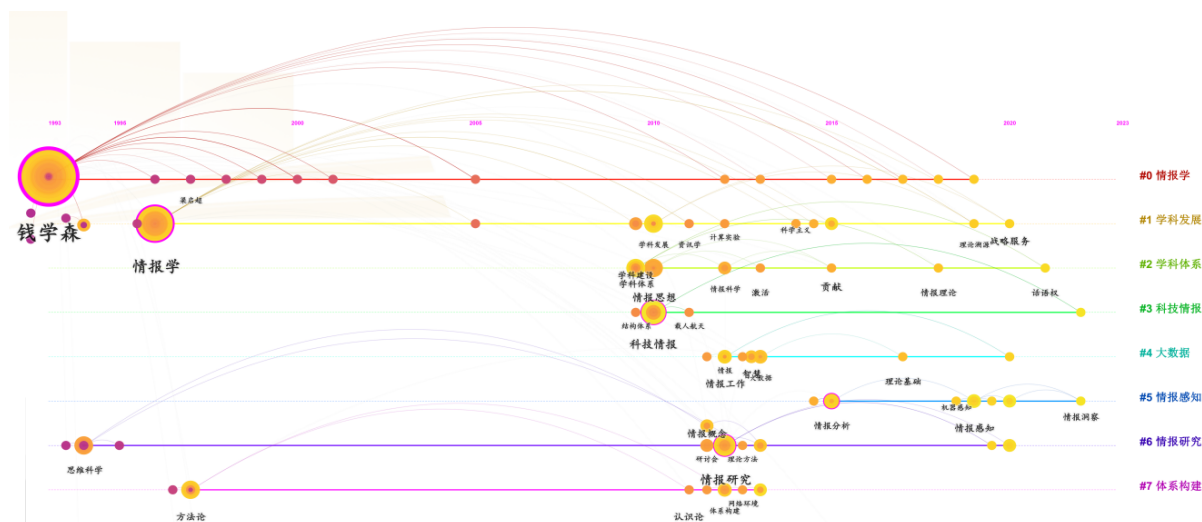


图6 钱学森情报思想研究的时区视图

钱学森情报思想研究总体上呈递增之势，课题热度逐渐升高，该领域研究呈现出良好的发展态势。钱学森的情报思想理念贯穿我国情报学学科发展的始终，并且在不同时期呈现出不同侧重的指导作用^[17]。

3.3 聚类节点信息表

结束了对 CiteSpace 图像的分析，笔者建立

了一个关键词节点信息表（表2），从而揭示出一些潜在的重要信息。

在聚类节点信息表中，突现性（Burst）是一个重要的概念，它指的是一个主题词在不同时间段内发生了显著的变化，这种变化可以帮助我们更好地理解和掌握某一学科的研究前沿^[12]。根据主题词突现性图，“情报学”一词的突现性检测值高达2.33，表明该词

在 2010—2014 年间的关注度有了显著的变化,如图 7 所示。在此期间,综合集成思想与情报学理论研究深度融合,并成为了情报学重要的理论基础,使情报学学科建设有了新的突破。作为早期钱老情报学思想研究的方向,“思维科学”的突现期在 1994—1995 年,“情报思想”的突现期在 2013—2015 年,说明当时思想体系化已经成熟。“大数据”在 2013—2018 年突现,体现了大数据技术的出现对情报学研究的重要影响。“学科发展”是近年关注的热点,其突现期为 2019—2021 年。“情报感知”“悟性思维”“算法风险”为延伸到当下的突现词,反映出目前的研究热点。“悟性思维”与钱学森情报思想中的“灵感”联系紧密。可以明显观察到,钱学森情报思想研究热点已经从“思维科学”“情报学”“情

报思想”“学科发展”等转向“情报感知”“大数据”“悟性思维”。表明现阶段钱学森情报思想研究的关注重点更偏向于思想落地,落实到具体的实践方法和方案。

表 2 有关钱学森情报思想研究的关键词节点信息

频次	中心度	年份	关键词
47	0.66	1993	钱学森
19	0.26	1996	情报学
8	0.12	2010	科技情报
7	0.15	2012	情报研究
5	0.04	2010	情报思想
4	0.01	1994	思维科学
4	0.03	2010	学科建设
4	0.01	2010	学科体系
4	0.02	1997	方法论
4	0.00	2010	学科发展

Top 14 Keywords with the Strongest Citation Bursts

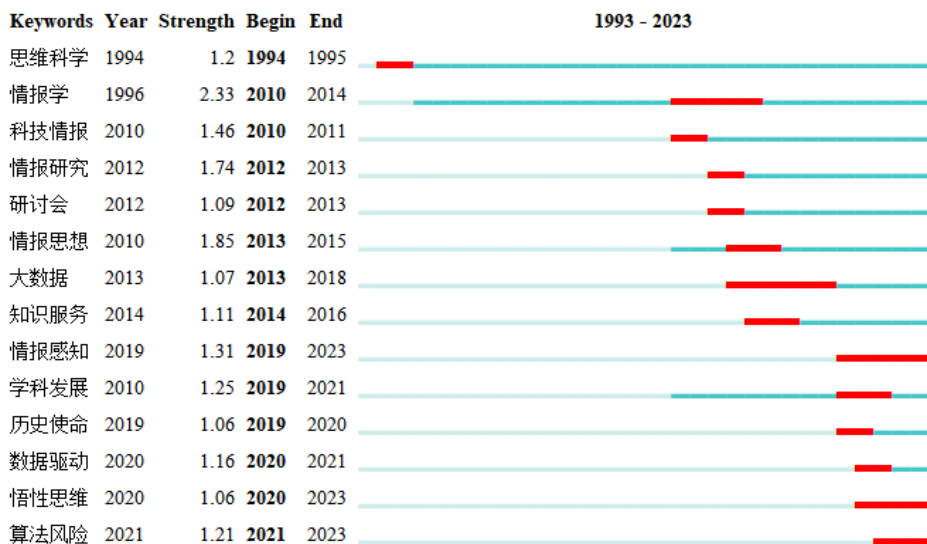


图 7 主题词突现性

4 讨论与启示

四十年来,钱学森情报思想研究日益繁荣。本文利用文献计量学方法,使用 CiteSpace 可视

化工具对中国知网 1983—2023 年间有关钱学森情报思想研究的文献进行定量分析,采用文献发表趋势图、关键词共现图谱、作者共现图谱、

机构共现图谱、关键词聚类、时区视图等计量可视化手段,梳理了钱学森情报思想研究的发展历程、研究热点转变,分析了未来的发展趋势。通过分析,本文发现钱学森情报思想研究总体上呈递增之势,课题热度逐渐升高,呈现出良好的发展态势。主要研究方向发生了变迁,热点词汇从早期的“思维科学”“情报学”“情报思想”“学科发展”等,逐步转向“情报感知”“大数据”“悟性思维”等,显示研究重点从情报概念辨析、学科建设指导原则等学科基础理论研究转向了对其思想在数智场景的落地应用和方案设计。

通过梳理钱学森情报思想研究的发展历程和研究热点的转变,并分析其未来的发展趋势,本研究列出钱老思想对情报学界的三点重要启示。

(1) 研究热度递增,表明持续关注的必要性

研究发现钱学森情报思想的研究热度逐渐升高,显示出良好的发展态势。这表明钱学森的思想在情报学领域具有持久的影响力和重要性。因此,情报学界应继续深入研究钱学森的情报思想,不断发掘其理论价值和应用潜力,以推动情报学的发展。

(2) 研究重点转向数智场景应用,突显前瞻性

研究热点从早期的“思维科学”“情报学”“情报思想”等基础理论,逐步转向“情报感知”“大数据”“悟性思维”等数智场景应用。这一转变展示了钱学森情报思想的前瞻性,表明其学术观点和思路在当今数智时代依然具有深刻的现实意义。情报学界应注重将钱学森的思想与

现代信息技术结合,探索其在大数据、人工智能、元宇宙等领域的应用,提升情报学的实践效能。

(3) 促进情报学科自主知识体系构建,彰显中国特色

当前我国正在构建信息资源管理一级学科和情报学二级学科的自主知识体系,钱学森的情报思想可以为之提供重要的思想源泉、丰富的理论方法和可行的工程指南。这有助于建立具有中国风格、中国气派和中国特色的信息资源管理和情报学学科体系、学术体系和话语体系,增强我国在情报学领域的国际影响力和话语权。

5 结论与展望

钱学森在情报学和情报事业研究领域作出了重大贡献。其情报思想贯穿于我国情报学学科发展和事业建设始终,并在不同时期呈现出不同侧重的指导作用,在我国学术界和业界影响极为深远。当前钱学森情报思想主要研究方向的转变不仅展示了钱学森情报思想前瞻性,其学术观点和思路放在现在都是非常恰当和深刻的,也表明了情报学在数智时代对思想理论落地和实践效能的关注,随着赛博空间、大数据、元宇宙和数字孪生的应用普及,钱学森情报学思想的现实意义和启示价值也日益凸显,值得我们全面地继承和发展。

同时我们也应认识到,尽管钱学森情报思想研究取得了一定的进展,但仍存在一些局限。首先,钱学森情报思想选题同质化严重,创新性不足。其次,跨学科研究较少,研究视野较为狭隘。再次,有显著影响力的学者数量较少,

研究机构呈现出高度集中的特点,不利于钱老思想的传播和发展,不利于争鸣的展开。未来钱学森情报思想研究需鼓励创新与选题多样化,学者应积极从信息资源管理和情报学新兴领域、前沿数智技术在学科中应用的视角出发,寻求钱学森情报思想的新发展与新价值;聚焦现代科学技术体系中跨学科交叉学科范式在钱学森情报思想研究中的应用,打造跨学科和交叉学科团队。支持地方高校和研究机构开展钱学森情报思想研究,通过建立协作网络促进研究机构之间的合作与交流,共享研究资源,推动研究的均衡发展。

值得一提的是,本文的数据主要来源于中国知网,一定程度上限制了研究的全面性和代表性。其次,限于篇幅,本文对于钱学森情报思想的深层次内涵和具体应用案例的分析尚有较大提升空间。再次,本文侧重于宏观的趋势分析,对具体研究成果的评价和比较较少,可能无法全面展示钱学森情报思想在不同研究领域和应用场景中的具体贡献和影响。此外,未来的研究需要持续跟踪和借鉴国际情报学、信息哲学、人工智能、复杂系统和神经认知科学等学科领域的最新发展动态和技术应用,以促使钱学森情报思想研究新知识增长点的不断涌现。

参考文献

- [1] 江和平. 钱学森:“我国国防科技情报事业的导师”[J]. 党史博览, 2022(1): 39-43.
- [2] 史秉能. 钱学森科技情报学术思想及其意义[J]. 钱学森研究, 2018(2): 74-83.
- [3] 袁有雄. 钱学森情报研究学术思想探析[J]. 情报理论与实践, 2013, 36(9): 18-20.
- [4] 刘芝玮, 周庆山. 基于钱学森大成智慧思想的国防科技情报工作模式初探[J]. 情报理论与实践, 2021, 44(9): 68-73.
- [5] 卢胜军, 赵需要, 栗琳. 钱学森科技情报理论体系及其意义[J]. 情报科学, 2012, 30(9): 1418-1423, 1435.
- [6] 赵一男, 赵超阳. 大数据时代钱学森情报研究思想的方法论意义[J]. 中华医学图书情报杂志, 2018, 27(4): 44-49.
- [7] 李竹, 曹文振. 钱学森情报学思想研究: 定名、脉络与内核——纪念钱学森院士逝世十周年[J]. 情报理论与实践, 2019, 42(10): 15-20, 14.
- [8] 王琳. 基于钱学森综合集成思想的情报学理论研究论要[J]. 情报理论与实践, 2013, 36(4): 6-11.
- [9] 卢胜军, 栗琳, 王忠军. 科技情报工作发展史观下的钱学森情报思想[J]. 情报理论与实践, 2015, 38(3): 21-25.
- [10] 马雨萌, 王延飞. 面向战略性新兴产业政策制定的情报感知研究[J]. 情报学报, 2023, 42(8): 883-892.
- [11] 王知津. 数智时代情报学学科体系构建[J]. 情报理论与实践, 2024, 47(3): 1-15.
- [12] 卢胜军, 王忠军, 栗琳. 赛博空间与大数据双重视角下的钱学森情报思想[J]. 情报理论与实践, 2013, 36(4): 1-5.
- [13] 史秉能, 袁有雄, 卢胜军. 钱学森科技情报学术思想研究及其意义[J]. 情报学进展, 2014, 10: 1-44.
- [14] 史秉能, 汤珊红, 赵柯然. 钱学森指引科技情报工作现代化发展[J]. 情报理论与实践, 2024, 47(1): 2-5, 21.
- [15] 王琳, 赖茂生. 中国科技情报事业回顾与展望: 基于情报学理论的视角[J]. 中国图书馆学报, 2021, 47(4): 28-47.
- [16] 刘如, 吴晨生, 王延飞, 等. 基于钱学森系统辨识理念的情报感知研究[J]. 情报理论与实践, 2019, 42(5): 7-13.
- [17] 马海群, 蒲攀. 钱学森情报思想影响力分析——兼评《情报理论与实践》的学术贡献[J]. 情报理论与实践, 2014, 37(9): 26-29.



开放科学
(资源服务)
标识码
(OSID)

“新动能·新青年·新作为”：2024 年科技情报 新青年学术会议综述

杨卓萱 赵晖 王彦妍

中国人民大学信息资源管理学院 北京 100872

摘要： [目的 / 意义] 2024 年 10 月 19 日，由中国科技情报学会青年人才工作委员会、中国科学技术信息研究所主办，中国人民大学信息资源管理学院、《情报工程》编辑部承办的“2024 科技情报新青年学术会议”在中国人民大学成功召开。本次会议旨在推动科技情报工作的高质量发展，聚焦面向国家战略和时代发展的情报理论研究、赋能情报用户的理论与实践探索、情报技术的创新发展路径以及紧跟时代的情报业务实践四个主题。 [方法 / 过程] 会议汇聚了科技情报领域的资深专家与中青年学者，共同探讨科技情报工作所面临的新形势、新问题、新挑战、新使命。 [结果 / 结论] 本次会议为学术交流和青年人才培养提供了良好的平台，为情报学术研究与实践应用相互融合创造了新机遇，为实现国家科技创新战略目标注入了新动能。

关键词： 科技情报；会议综述；情报理论；情报用户；情报技术；情报业务

中图分类号： G35

“New Momentum, New Youth, New Achievements”: A Review of the 2024 Academic Conference on New Youth in the Field of Scientific and Technical Information

YANG Zhuoxuan ZHAO Hui WANG Yanyan

School of Information Resource Management, Renmin University of China, Beijing 100872, China

Abstract: [Objective/Significance] On October 19, 2024, the “2024 Academic Conference on New Youth in the Field of Scientific and Technical Information” was successfully held at Renmin University of China, sponsored by the Youth Chapter, China Society for Scientific and Technical Information and the Institute of Scientific and Technical Information of China, and hosted by the School of Information Resource Management of Renmin University of China and the Editorial Office of Technology Intelligence Engineering. The conference aims to promote the high-quality development of scientific and technical information work, focusing on four topics: intelligence theory research for national strategy and development of the times, theoretical and practical exploration

基金项目 国家社科基金青年项目“媒体融合环境下用户健康信息搜索行为模式研究”（20CTQ023）。

作者简介 杨卓萱（2002-），硕士研究生，主要研究方向为信息行为；赵晖（2002-），硕士研究生，主要研究方向为信息行为、信息检索、人机交互，E-mail: zhaohui1127@ruc.edu.cn；王彦妍（1989-），博士，副教授，主要研究方向为信息行为、数字治理、数字人文。

引用格式 杨卓萱, 赵晖, 王彦妍. “新动能·新青年·新作为”：2024 年科技情报新青年学术会议综述 [J]. 情报工程, 2024, 10(6): 119-127.

of empowering intelligence users, innovative development paths of intelligence technology, and innovative practices of intelligence business that keep up with the times. [Methods/Processes] The conference brought together experts and young scholars in the field of Scientific and Technical Information to discuss the new situation, new problems, new challenges, and new missions facing scientific and technical information work. [Results/Conclusions] The conference provided a good platform for academic exchanges and young talent development, created new opportunities for the integration of intelligence academics and practices, and injected new momentum into the realization of the national science and technology innovation strategic goals.

Keywords: Scientific and Technical Information; Conference Review; Intelligence Theory; Intelligence Users; Intelligence Technology; Intelligence Business

引言

2024年1月31日,习近平总书记在中共中央政治局第十一次集体学习时强调:“发展新质生产力是推动高质量发展的内在要求和重要着力点”。从一五计划建立全国科技情报体系开始,科技情报行业始终围绕着国家科技创新发展的战略需求持续完善科技信息资源保障体系,创新情报理论和方法工具,协助制定科技战略、规划与政策^[1],助推我国的科技发展和创新。在二十届三中全会胜利召开以及新一轮科技革命和产业变革的双重背景之下,为了深入贯彻落实新质生产力的发展要求,夯实科技情报工作建设,推动科技情报工作者深化交流合作,构建高水平的科技情报人才体系,2024年10月19日,由中国科技情报学会青年人才工作委员会、中国科学技术信息研究所主办,中国人民大学信息资源管理学院、《情报工程》编辑部承办的“2024科技情报新青年学术会议”在中国人民大学成功举办。

本次会议主题为“新动能·新青年·新作为”,设置了嘉宾致辞、专家讲坛、获奖青年讲坛、专题对谈、学术盲盒等多项议程,吸引了来自高校、研究机构及企事业单位的百余位学者参会。会议开幕式由中国科学技术情报学会青年人才工作委员会主任、研究员刘志辉主持。中国科学技术情

报学会理事长戴国强和北京大学信息管理系教授赖茂生分别发表了学会领导致辞和学术致辞,指出青年学术力量是促进科技创新的重要力量,科技创新如何赋能新质生产力是情报工作者需要把握的关键,勉励青年人才“扎根成长”,并对本次会议的召开表示热烈祝贺。

1 沙龙环节

1.1 面向国家战略和时代发展的情报理论研究

“情报理论方法沙龙”聚焦情报学的目标使命、研究范式和新理论的发展方向,明晰情报理论的基础框架和研究边界,探讨情报学在大数据、人工智能等新兴技术环境下的演进路径,由杭州电子科技大学教授王琳主持。

在新时代、新环境、新技术、新战略的背景下,情报学的主要研究内容在不断变化,情报工作的重点也在不断转移,但创新情报理论方法、完善情报学科体系、振兴我国的情报事业始终是情报学工作者的坚守^[2]。南开大学商学院信息资源管理系王知津教授从国内外的情报学演变轨迹入题,强调了我国的科技情报工作要坚持“不忘初心”,遵循我国科技情报事业创建初期国家领导人提出的“打仗”思想和“耳目”“尖兵”“参谋”的角色地位,为各

级领导和决策者制定策略、规划和战略提供情报支持；也要“牢记使命”，按照我国科技情报事业初期提出的“广、快、精、准”的要求，完成情报搜集获取、情报存储检索、环境情报监测、目标情报分析、技术情报跟踪、决策情报支持、情报探测预警、情报安全保护等八大基本任务。在数智时代，情报事业和情报学的初心使命要体现在助力科技创新，促进新质生产力发展，推动高质量发展等层面上，为服务国家战略和社会需求贡献科技情报的重要力量。

情报学作为一门兼具理论探索与实践应用的学科，正在快速发展并面临深刻的范式转型。中国人民大学周晓英教授从情报观角度分析中国情报学研究范式的发展，通过构建情报定义的时序图^[3]，梳理出情报学演变的历史脉络。报告系统地阐述了情报认知观从军事情报观、科技情报观、信息情报观发展至综合情报观的演变过程，强调当前情报认知观已步入“大情报观”的阶段。“大情报观”包含情报跨领域的多元性，情报来源的多样性，情报价值的综合性和多样性三个重要维度。当前中国情报学的研究范式可概括为“情报范式+信息范式+融合范式”，强调情报源自信息，而信息中亦包含情报，二者相互关联，形成整体研究框架。

面临着科技革命与国际竞争的双重挑战，情报工作者需要不断更新研究方法 with 思维模式，其中不仅包括采用先进的分析工具和技术，还要求情报工作者培养综合的思维能力和识别潜在风险和机遇，为战略规划提供有力的支撑。浙江省科技信息研究院查新中心主任林志坚立足碳中和背景下科技情报工作者的参谋角色，介绍了“智江南”自主研发数智化智库的概况。

报告分享了智库型知识产权研究与服务的方法，即坚持问题导向进行文献调研，针对当前问题进行初步研究，遵循对标问题、补短板、明方向的原则进行深入研究；同时介绍了新兴技术对比分析、国外来华专利布局分析、重点专利申请人分析、发明人及团队分析、应用示范项目分析的研究内容，为产学研合作提供了宝贵的经验。北京大学信息管理系博士生鞠孜涵介绍了地平线扫描方法，强调使用该方法时需具备全谱思维，关注未知未闻等问题，进行全方位扫描。该方法适应多重需求的情报分析，能够帮助从业者理解和准备未来的风险。科技情报工作者在关注基于历史变化统计得到的有着强信号的前沿科技的同时，也应关注未来有可能成为强信号但现在信号较弱的技术。

1.2 赋能情报用户的理论与实践探究

情报用户是情报工作的服务对象。“获奖青年讲坛·情报用户沙龙”聚焦情报赋能用户的实现途径、安全保障、成果评价与应用聚合等多个方面，由南开大学图书馆副馆长樊振佳主持。

北京大学信息管理系情报学教研室主任、长聘副教授刘畅提出可通过情报赋能用户，提高其决策能力，以应对决策过程中的获取障碍、内容质量不佳、真伪迷雾等挑战。情报赋能的实现途径有二：一是从主体视角出发提高情报用户的能力，如简化情报用户的决策过程、增强决策体验过程、改善情报内容的呈现形式等；二是从客体视角增强情报工作的效能，如多元的线索发现、精准的分析方法、灵活的决策保障等。

南京大学信息管理学院张斌副教授聚焦数

据合规与隐私保护,指出当前监管要求的复杂性及动态变化特征,强调完善数据的全流程合规和监管体系的重要性。为应对企业在数据合规方面的成本高、质量不稳定、周期长等挑战,提出了基于数智驱动的解决方案,如大规模DLM标注和领域大模型AIGC,从而赋能整个数据产业链条。

中国科学院文献情报中心陈莉玥助理研究员从劳动分工的视角探究了科学贡献的复杂性。通过识别论文的实际贡献和劳动分工的五种类型(理论型、方法型、实验型、数据型和其他型),分析论文学科贡献的特征,发现论文的实际贡献模式在不同学科之间存在一定差异,论文的实际贡献与劳动分工投入之间并无显著相关性。

中国航天系统科学与工程研究院一所数据服务部副主任、高级工程师李金钊在报告中指出,航天科技信息动态跟踪工作面临信息来源多、数据规模大、响应时效高、成果类型多等挑战,迫切需要创新技术支撑手段和应用服务。因此,团队提出构建基于智能技术的航天科技信息基础数据库,以此减轻研究人员负担,提升产品生产线的智能化和自动化水平。该数据库具有数据密集、智能嵌入、人机协同的动态跟踪智能化生产线平台,具备基础数据检索查阅、智能问答、可视化专题分析栏目构建和产品生成能力。未来研究团队将进一步探索构建面向复杂任务的AI智能体,扩展行业知识范围和智能应用场景,以提升科技信息工作的服务质效。

1.3 多视角解读情报技术的创新发展路径

“情报技术沙龙”聚焦情报领域的学科建

设、情报技术的实践应用和产业落地,探讨大模型、实体识别等技术在情报研究中的应用,推动情报技术在科研情报分析和科技创新管理中的深度融合。沙龙由南京农业大学信息管理学院教授、中国科学技术情报学会青年人才工作委员会副主任杨波主持。

2022年,“信息资源管理”一级学科的更名为中国的信息分析学科开启了学科改革部署与战略布局的新篇章^[4]。中国人民大学信息资源管理学院教授、中国人民大学智能信息分析研究中心主任卢小宾梳理了中国自主信息分析知识体系从初始形成、快速发展到持续深化的发展脉络,展示了信息分析学科由信息分析原理、信息分析方法体系与工具以及信息分析应用构成的知识体系结构,强调方法、技术和工具共同构成的数智融合型方法体系是信息分析的关键。此外,卢小宾阐明了中国自主信息分析知识体系的特色:(1)基于中国智慧的信息分析原理自主创新;(2)具有中国特色的信息分析方法体系研究;(3)科学回答中国之问、时代之问的信息分析应用。

大模型技术的发展为科技情报的智能化提供了新的可能性,也对情报学科的理论体系构建提出了新的挑战,对技术融合和产业应用提出了更高的要求。针对这一情况,武汉大学信息管理学院程齐凯副教授从科技情报工作的角度指出,当前情报工作需要顺势而为,建立契合发展阶段的创新理论体系、突破科技封锁的支撑技术体系和完善独立自主的创新服务体系。他总结了目前大模型技术在情报领域落地应用中亟待突破的6类问题:可控生成、安全伦理、隐私安全、路线实现、模型分割和微调能力漂移。

此外，他也介绍了团队基于合成数据的可靠引用生成研究，从实践层面切实回应了大模型落地应用的可行性。科大讯飞北京研究院执行院长、正高级工程师伍大勇从产业界实践的视角，介绍了认知大模型的智能涌现和国家大模型产业的激烈竞争态势，以及大模型多模态能力上限被持续抬高的发展趋势。他指出了通用大模型在可信问题、算力集群建设方面面临的技术挑战，介绍了基于科技文献大模型的星火科研助手，该工具可提供科技情报监测、情报分析、情报研读和情报写作等智能化服务，助力科研提质增效。

凭借信息采集、分析、挖掘与管理的重要能力，情报技术在不同场景中的发展和应用愈加广泛。在社交媒体场景下，南京大学信息管理学院博士研究生刘江峰聚焦网络舆情中的群体极化和极端观点检测现象，揭示了网络舆情空间中群体极化和极端观点互相反馈和促进的内在机制，发现其背后的原因可以归纳为社交媒体泡沫效用，情绪化和点击驱动报道，以及虚假新闻泛滥。南京农业大学信息管理学院博士生刘禹彤面向生物信息学领域，介绍了研究数据引用实体识别方法，发现了目前研究人员倾向于引用数据集的初创论文，且现有数据引用识别以人工和半自动化方法为主。随后，刘禹彤介绍了基于“先分类后识别”策略的数据资源发现模型的构建流程，该模型可以为进一步探索数据复用及复用指标的评价提供支持，为数据共享、人才评价和科学发现作出贡献。

在现代化产业体系建设与产业数字化高质量发展的背景下，情报技术正逐渐成为产业转型与升级的重要推动力量。华中师范大学信息

管理学院博士研究生董焕晴在数智赋能的视角下，系统介绍了产业大脑“汇‘智’成‘脑’”的构成要素和“聚‘链’融‘智’”的运行逻辑，提出了由数字底座、智能中枢、信息服务、管理与保障体系构成的总体框架，总结了面向产业布局与管理、产业实践与升级的应用场景，包括产业监测、产业诊断、产业评价、精准施策、智能招商和企业全景评估、政策精准匹配、个性化智能服务等，揭示了情报技术在支持产业智能化管理和提升产业竞争力中的关键作用。

1.4 紧跟时代的情报业务创新实践

“情报业务沙龙”邀请六位嘉宾，共同探讨提升情报业务的不同途径，由中国科学院文献情报中心韩涛研究员主持。专题报告既探讨了专题报告撰写、提高情报能力等具体情报业务工作，也分享了从人才培养、科研资源配置、数据资源建设、技术支持等方面提高情报业务质效的途径。

北京大学信息管理系王延飞教授认为事实、观点和策论是专题报告的质量抓手。报告指出，决策分析依赖于“情报事实”，即情报工作场域中的信息。观点虽不足以称为事实，但有相当程度的数据和逻辑支持，分析过程中对观点的检验与鉴识是重要工作。策论质量系于情报智慧，厘清方法工具对象谱系是其中的关键。最后，王延飞指出报告结构、标题表达、资料分析、结论摘要呈现研究之美的重要方面。

人才培养是情报业务持续优化的不竭动力，有效评估人才是人才工作的重要抓手。中山大学信息管理学院侯剑华教授使用科学计量学的方法支撑科技管理的实践业务，分享科研人员

学术节律测度方法及其创新应用探索。学术节律即科研人员在进行科研活动时所呈现的节奏和规律。侯剑华团队基于以往研究中根据发文轨迹、学术影响力轨迹、学术合作轨迹的三种测度方法,结合知识的游离与重组理论以及知识扩散理论,提出基于知识生产能力的学术节律测度方法,用于测度对已有知识的整合能力,计算知识的扩散与影响能力,最终构建知识生产能力矩阵,对科研人员的知识生产能力进行测度。

中国科学技术信息研究所李孟秋助理研究员从科技安全的新形势与新内涵出发,分析了科技安全能力建设对科技情报工作提出的三方面新要求。总体要求是实现对国家科技安全态势更为全面系统的感知,以专家智慧融合计算智能,深度解析存在的科技风险及其表现;外部风险应对要求为,增加大国科技博弈情报分析的时效性与准确性,为有效应对国外科技遏压做好提前谋划;内部风险应对要求为面向创新能力不足等内生性问题建立体系化的动态评估系统,及时调整科技管理政策与研发方向。同时,科技情报工作也面临开源数据资源水平不够高、情报工作整体支撑框架尚未建立、分析科技安全态势的理论与方法不够成熟、科技情报工作仍未能有效支撑科技安全能力建设四方面挑战。据此,李孟秋构建了科技情报工作保障科技安全能力建设的框架:保障科技安全能力建设的底层数据支撑框架、覆盖科技安全全生命周期的情报支撑框架、涵盖科技安全各环节的全链条分析框架。

了解跨学科知识流动情况有助于科研人员

和机构合理配置资源、明确科研方向、有效布局科研战略^①。武汉大学信息管理学院的博士研究生刘可辉介绍了结合前向引文及后向引文来识别高影响力的跨学科知识流动的方法。将该方法应用于 ISLS 学科的跨学科知识流动分析,发现以下结论:(1)跨学科流入知识在目标学科中以跨学科吸纳知识及跨学科内化知识的形式存在,且跨学科内化知识的形式更容易被前向引用;(2)结合跨学科知识的流入分析及持续影响分析,能够将跨学科知识流动路径划分为四种类型,其中路径四“跨学科流入知识→跨学科吸纳知识→跨学科内化知识→前向引用知识”中的跨学科流入知识对目标学科的影响力最高;(3)对 ISLS 学科产生高影响的跨学科流入知识的特点为扩展性及实用性强,且涉及创新应用与多学科融合。

数据资源建设和技术支持是情报业务的重要基础。中国人民大学信息资源管理学院博士研究生邵琦汇报了 Europeana 的文化遗产数据资源建设路径和现状。Europeana 的数据资源建设包含数据资源规划、资源组织和资源调度三个环节,它们相互依存、相互促进,共同实现文化遗产数据资源的保护传承和创新发展。中国科学院文献情报中心硕士研究生陈昱成介绍了“智能科学家”数智驱动能力的建设框架与具体实践。智能 AI 模型,观测数据、实验数据和计算数据,以及领域内的专业科学知识共同构成智能科学家的数智驱动能力。其建设从高质量多模态数据基座、领域化模型工具、专业知识嵌入三方面着手,三者循环交互,在螺旋中互相促进优化。

① 选自武汉大学信息管理学院博士研究生刘可辉发言内容。

2 对谈环节

为促进学术交流和学科融合发展，推动科技情报工作的守正创新，会议设置对谈环节，邀请来自不同院校、不同学科背景、不同研究方向的学者围绕当前科技情报研究工作的重大需求与迫切任务等热点与前沿议题，进行面对面的交流与讨论。会议对应 4 场沙龙设置情报理论、情报用户、情报技术和情报业务 4 场专题对谈。

2.1 情报理论在变革背景下的继承与发展

情报理论主题对谈由王琳主持，邀请五位嘉宾就情报理论的创新、发展和应用进行了深入探讨。情报学理论揭示了信息社会的深层机制，是开展扎实研究的基础，其重要性不容忽视^②。王知津教授曾指出当前情报学理论的五大发展趋势：情报学哲学思潮多元化、多学科与跨学科共存、技术与人文并重、从数字化到智能化以及从信息服务到知识服务^③。随着技术的不断发展，情报分析过程中人工智能辅助的情报分析较以往的情报分析工作产生了很大的变化。人工智能辅助得到的情报可以归类为智能情报，而原有的人工分析所得的情报则归类为智慧情报。两种情报都是必需的，但人工智能技术只能解决信息分析等客观细微的问题，而无法彻底取代综合判断信息成为情报分析的主角^④。同时，情报学的研究范式不断演变以应对新的社会需求和技术挑战，“大情报观”的形成正是技术发展与社会需求变

化共同驱动的结果，是情报学科不断适应外部环境变化的必然产物^⑤。

吉林大学商学与管理学院邓君教授指出数字人文、数据管理等新兴二级学科的涌现为科研工作者提供了更丰富的学术身份认同感和归属感，并与信息资源管理一级学科形成了合作共赢的关系：新兴学科的出现扩展了信息资源管理学科的边界和应用领域。同时，情报学理论及相关方法技术为这些新兴学科提供了关键的理论支撑和方法论指导，促进了学科融合与协同发展。

情报理论的创新发展同样体现在实践层面。林志坚结合科技情报工作向科技智库转型的经验，提出了“大格局、大方法、大成果”的主旨，强调情报工作必须立足国家发展与战略需求，以创新的方法体系支撑情报工作转型升级。他认为，情报理论的发展不仅是理论体系的创新，还需要通过多方面的协同合作来践行，最终在学术攻关和服务社会方面取得实质性的大成果。

2.2 多维视角下的情报用户探讨

对情报用户及其需求的精准把握是有效开展情报研究的关键。为了实现这一目标，情报机构需要运用多样化的技术手段满足用户动态变化的需求，该过程也体现了情报用户和情报行业相互作用的关系。第二场对谈由樊振佳主持，五位嘉宾从不同视角出发，分享了各自有关情报用户的观点。

②选自北京大学信息管理系博士研究生鞠孜涵发言内容。

③选自南开大学商学院信息资源管理系王知津教授发言内容。

④选自中国人民大学信息资源管理学院周晓英教授发言内容。

⑤选自北京大学信息管理系刘畅副教授发言内容。

⑥选自南京大学信息管理学院张斌副教授发言内容。

要开展关于情报用户的研究,首先应明确情报用户是谁^⑤,即考虑谁愿意为自己的产品买单的问题^⑥。其次是针对用户需求,从任务的视角切入,理清服务对象目前要解决的问题与完成的任务^⑤;并且由于用户对于自身需求的描述不一定准确,在制作大模型产品时也需考虑如何设计 prompt 等相关问题^⑦。再次,在数字时代信息传输速度大幅提升,信息量爆炸增长的背景下,情报用户的需求也在发生变化。在以前的信息背景下,用户获取信息的途径较为单一,决策周期压力相对较小;现今新的信息获取途径大量涌现,用户对于提供产品与服务的有效性要求更高,也倒逼行业提升服务速度与质量^{⑦⑧}。与此同时,随着科技情报服务范围的扩大,一些重点行业将面临潜在的用户隐私安全问题,需要从业人员加强关注^⑨。

2.3 情报技术在数智时代的应用与挑战

大模型、人工智能和数据挖掘等技术的不断演进为情报技术带来了全新的发展契机与挑战。情报技术如何在这一背景下实现有效应用并助力学科的创新与发展,成为学界和业界共同关注的重要议题。论坛设立情报技术主题对谈,邀请了来自学界和业界的五位嘉宾,从理论与实践双重角度展开讨论,由北京师范大学政府管理学院信息管理系主任刘晓娟教授担任主持人。

程齐凯、刘禹彤、刘江峰和董焕晴探讨了

大模型在情报工作中的应用场景,包括生物信息学数据引用实体识别^⑩、数据治理^⑪等方面。研究发现,大模型技术在数据采集和处理、实体抽取和识别^⑫、内容生成等情报工作流程中能够发挥潜力,但在战略情报等需要高度精确判断的领域仍然存在局限,情报工作者在使用该项技术时需要因地制宜、审慎评估其适用性^⑬。

北京市科学技术研究院科技情报所副所长吴晨生研究员结合自身丰富的研究经验,提出了“技术是情报的工具”这一观点,技术的不断进步可以赋能情报工作,理解新技术的本质和来源可以更好地将其应用于情报工作中。伍大勇以跨界视角解读了情报技术与信息技术之间密不可分的关系,进一步深化了“技术是情报的工具”的概念,指出情报学的任务是获取和分析知识内容,这一核心任务不会因技术手段的变化而改变。情报行业应主动拥抱新兴信息技术,实现学科的拓展和深化。在此讨论基础上,吴晨生和伍大勇也分别从学界和业界的视角对年轻情报工作者在未来的技术学习和研究重点提出了建议。伍大勇指出大模型技术在应用和落地方面仍面临许多挑战,情报工作者要保持技术背景下的学科本位主义,不被技术潮流裹挟。吴晨生则指出大模型技术对于小概率和零概率事件的预测能力不足,依然需要依赖人类的经验判断,提醒年轻学者要在技术应用中保持清醒,既要利用技术的优势,也要意识到其局限性。

⑦选自中国科学院文献情报中心助理研究员陈莉玥发言内容。

⑧选自中国航天系统科学与工程研究院一所数据服务部李金钊副主任发言内容。

⑨选自南开大学商学院信息资源管理系博士研究生杨丽娟发言内容。

⑩选自南京农业大学信息管理学院博士研究生刘禹彤发言内容。

⑪选自华中师范大学信息管理学院博士研究生董焕晴发言内容。

⑫选自南京大学信息管理学院博士研究生刘江峰发言内容。

⑬选自武汉大学信息管理学院程齐凯副教授发言内容。

2.4 情报研究中的主客观平衡之论

在强调数据驱动和数智融合的时代，情报工作者不仅要关注数据和技术在情报业务中的基础性作用，也要考虑到情报研究中的主观性因素。在天津师范大学管理学院副院长、副教授翟羽佳的主持下，七位嘉宾围绕“如何把握情报研究中的主客观平衡”问题展开讨论。

首先，无论是学术研究还是业务实践，最终输出的都是人类认知分析的结果，在实际操作中不可能完全排除主观因素的存在，因此没有绝对的主客观之分，不能随意偏废其中一方，通过定量的技术得出的结果，也需人类进行定性的判断^⑭。其次，情报学的交叉属性极强，涉及大量定性方法与定量方法的结合，定量的手段会使研究更加高效、客观，但也有追求最省力法则的“趋易”嫌疑；很多情况下，定性的主观判断比定量的数据分析更难^⑮。韩涛提到自然科学研究的两个极致，一个向宇宙、向宏大去探索；一个向微观、向极小去探究。科技情报首先要从大的数据和模型出发，这也是从业者们必须面对并拥抱的行业趋势；但拥有大量数据并形成宏观认知后，也需要做精微的“小模型”^⑯以完成具体微观的情报工作。

邵琦以自身研究为例，分享了文化遗产数字化研究从定性到定性与定量相结合的转变——从最初的采纳专家意见，到后来结合专家意见与数据分析结果。由此认识到数据是用来揭示其背后所反映的客观规律的工具，借此工具可以帮助研究人员进行定性的最终判断，

体现了情报研究中的主客观平衡之思。

3 总结

教育、科技、人才是全面建设社会主义现代化国家的基础性、战略性支撑^⑰，在发展新质生产力中发挥着关键作用。本届科技情报新青年会议响应时代号召，聚焦科技情报研究的核心议题和创新动态，邀请学界与业界老中青三代专家学者分享创新思维与研究成果，通过 20 个主题报告、4 场专题对谈、30 个学术盲盒开启等环节，为广大科技情报从业者提供了宝贵的交流与合作机会。本次会议具有崇尚平等学术交流的新风格，开辟了分主题对话、分层报告的新传统。中国人民大学信息资源管理学院党委书记闫慧教授在会议总结中提出殷切希望，期待会议在形式与内容上不断创新，引领新理念，培养新人才。

参考文献

- [1] 赵志耘. “十四五”科技情报创新的思考[J]. 信息资源管理学报, 2021, 11(6): 4-9.
- [2] 苏新宁. 不忘初心、牢记使命展望情报学与情报工作的未来[J]. 科技情报研究, 2019(1): 1-12.
- [3] 周晓英, 陈燕方, 裴俊良, 等. 中国情报与情报学: 概念模型和学科演进[J]. 情报理论与实践, 2024, 47(5): 1-11.
- [4] 侯鹏娟, 陈兰杰, 李伟. 构建中国特色信息资源管理学科自主知识体系面临的机遇及实践逻辑[J]. 西华大学学报(哲学社会科学版), 2024, 43(1): 101-110.
- [5] 当代国际情报学理论进展与中国情报学的理论创新[C] // 全国哲学社会科学工作办公室. 国家社会科学基金项目优秀成果选介汇编(第1辑), 北京: 社会科学文献出版社, 2021: 850-856.
- [6] 习近平. 高举中国特色社会主义伟大旗帜为全面建设社会主义现代化国家而团结奋斗[N]. 人民日报, 2022-10-26(001).

⑭选自北京大学信息管理系王延飞教授发言内容。

⑮选自中山大学信息管理学院侯剑华教授发言内容。

⑯选自中国科学院文献情报中心研究员韩涛发言内容。