



科技情报系列丛书

智慧科技 与情报服务

ZHIHUI KEJI
YU QINGBAO FUWU

主 编 谢 威
执行主编 吴晨生 代根兴 高淑萍



北京邮电大学出版社
www.buptpress.com

科技情报系列丛书

智慧科技与情报服务

主 编 谢 威

执行主编 吴晨生 代根兴 高淑萍



北京邮电大学出版社
www.buptpress.com

内 容 简 介

本书以“智慧科技发展,情报服务先行”为主题,入选的论文主要是从北京科学技术情报学会 2018 年学术年会论文征文中挑选出来的,主要包括以下几个方面:在新时代背景下人工智能等颠覆性技术与实体经济融合研究;大数据环境下人工智能发展趋势研究;智慧科技发展与当今科技领域最前沿的课题探讨;科技情报服务对人工智能科技热点问题的深入分析等。这些论文反映了一年來北京科学技术情报学会专家、学者对科技情报工作的理论探讨,也反映了科技情报工作人员在实际工作中的实践创新、应用技术的研发、公共服务平台的构建、产业创新体系的强化、科技情报工作理论的探索和经验总结。

图书在版编目(CIP)数据

智慧科技与情报服务 / 谢威主编. -- 北京: 北京邮电大学出版社, 2018. 11

ISBN 978-7-5635-5643-4

I. ①智… II. ①谢… III. ①科技情报工作—学术会议—文集 IV. ①G255.51-53

中国版本图书馆 CIP 数据核字 (2018) 第 273030 号

书 名: 智慧科技与情报服务

责 任 编 辑: 徐振华 孙宏颖

出 版 发 行: 北京邮电大学出版社

社 址: 北京市海淀区西土城路 10 号 (邮编: 100876)

发 行 部: 电话: 010-62282185 传真: 010-62283578

E-mail: publish@bupt.edu.cn

经 销: 各地新华书店

印 刷: 北京九州迅驰传媒文化有限公司

开 本: 787 mm×1 092 mm 1/16

印 张: 26.25

字 数: 688 千字

版 次: 2018 年 11 月第 1 版 2018 年 11 月第 1 次印刷

ISBN 978-7-5635-5643-4

定 价: 66.00 元

• 如有印装质量问题, 请与北京邮电大学出版社发行部联系 •

编委会

主 编：谢 威

执行主编：吴晨生 代根兴 高淑萍

主任委员：谢 威

副主任委员：吴晨生 郑彦宁 刘素清 夏勇其 袁汝兵

委 员：王 强 张丰智 钟新春 付 宏 苗润莲

肖 雯 邵 颖 韩 军 侯 丹 杨 萍

刘彦君 李 辉 李 荣 杜慰纯 王婵媛

黎晓东 廖 涛

前言

2018年是贯彻党的十九大精神的开局之年，是改革开放40周年，也是站在新的历史起点上接力探索、接续奋进的关键之年。2017年，李克强总理的政府工作报告中第一次写入了人工智能，强调要加快培育壮大包括人工智能在内的新兴产业。2018年3月，全国两会在北京隆重召开，以人工智能为代表的科技创新成了此次两会的热点和亮点。

作为新一轮科技革命的重要代表之一，智慧科技发展是当今科技领域最前沿的课题；作为新一轮产业变革的核心驱动力，智慧科技正在成为全球经济发展的新动力；而作为社会风尚变化的一种最新潮流，人工智能又是离百姓生活最近的科技话题。创新是引领发展的第一动力。只有从具体的科研单位做起，从推动具体的课题研究入手，推动科研体制的改革和创新，才能为智慧科技的加速落地和产业化提供最有力的支撑。在此基础之上，北京科学技术情报学会在北京举办了主题为“智慧科技发展，情报服务先行”的论坛。

关注科技情报创新与智慧科技的关系，是基于人工智能在社会经济发展中越来越重要的作用。人工智能也是北京市发展的重要产业之一。北京人工智能研究基础深厚，聚集了一批人工智能骨干研究单位，拥有十余个国家重点实验室，神经网络芯片、计算机视觉等人工智能关键技术国际领先。目前，北京的人工智能产业成为全国发展高地，拥有全国过半数研究机构，研发企业数量居全国首位，其中5家入选全球人工智能创新100强企业。

关注科技情报发展与智慧科技的关系，是基于科技创新在社会经济发展中越来越重要的作用。科技创新活动是不断循环往复、螺旋式上升的知识创造或应用过程，在科技创新活动中，从科技规划到科技成果评估，几乎其中的每一个环节都或多或少地需要不同程度的科技情报服务支撑。然而，当今科技情报服务仍以传统的服务方式为主，并且存在若干需要思考的问题。因此，北京科学技术情报学会围绕“新时代背景下，人工智能等颠覆性技术与实体经济相融合”“科技创新中心建设的新动能研究”“面向2035年的北京科技前沿基础研究保障的体制机制研究”以及“科技情报服务在科技创新领域中发挥的作用”等主题，集中开展研讨，为大家提供一个专业化的学习、交流、讨论与分享的机会与平台。该论坛将借助北京科学技术情报学会作为北京市科技情报服务平台的多年服务经验与优势，吸引全国优秀的情

报专家、企业相关人士、情报从业者、产业与市场研究人士以及学者。这场“智慧科技发展与科技情报创新”的交流盛宴将为各政府部门、企业人士及创业者提供学习、提升和拓宽视野的机会，让与会者真正享受科技情报业界的思想大餐。本次论坛共收到论文 60 余篇，这些论文涵盖了科技情报理论与实践的多个方面，包括情报信息、计算机网络、人工智能智库、综合服务等，具有较高的学术和实践价值，体现了北京市科技情报工作在这一年来的最新进展。

北京以建设全国科技创新中心为目标，面向世界科技前沿，面向经济主战场，面向国家重大需求，已经取得了举世瞩目的成就，在若干领域开始成为全球创新的并跑者，甚至成为领跑者。北京科学技术情报学会在服务北京市政府的同时，为政府与企业之间、科技成果与市场之间搭起了一座良好的沟通桥梁，积极开展多种形式的科技活动，加速以情报服务为支撑的科技创新，推动以科技创新为核心的全面创新。我们将协同其他情报机构，认真贯彻党中央、国务院的决策部署，坚持创新、协调、绿色、开放、共享的发展理念，加快推进以科技创新中心建设为核心的情报支撑进程，以科技情报的方法与手段，将最前沿的科技动态、科技发展趋势、科技战略布局等有价值的情报信息服务落到实处，牢牢把握高质量发展的根本要求，坚持把科技创新作为引领发展的第一动力，为北京高质量发展和创新型国家建设再添新贡献。

北京科学技术情报学会理事长

谢 威

2018 年 10 月

目 录

情报信息类

基于多源数据的前沿科学领域与新兴研究方向识别和遴选方法研究	曾 文 李 辉 李 荣(3)
大数据环境下企业科技信息服务模式研究	付芳杰(10)
基于文献计量的领域基础研究竞争态势分析框架研究——以智能机器技术为例	李 荣 李梦辉 刘 静(16)
基于论文与专利视角下世界人工智能发展研究	罗梓超 范漪萍 刘彦君(28)
英国智能机器人技术创新研究及发展趋势分析	王晓迪 刘彦君 张 炜(36)
大数据环境下人工智能发展趋势研究	邱克锦(45)
英、美城市开放政府数据平台调查与分析	许玥姮 刘光宇 王冠宇 丛 琳(51)
基于优秀青年科学基金资助人才国际合作数据分析跨领域科研发展趋势	王晓迪 张 媛 范漪萍(55)
大数据时代信息资源服务的机遇、挑战与对策	燕 娜 郭建伟 陈佳宇(62)
第二代开源情报及其分析工具及方法	侯 丹 杜元清 李薇濛(66)
区块链技术国防应用潜力简析	陈 菲 姚保寅 侯 丹(70)
基于文献计量分析的国内脑科学研究	毛维娜 李 梅 孟 潇 董 洁(76)
基于大数据的北京留学回国人才就业分析研究——以中关村国家自主创新示范区为例	陈佳宇 张素娟 燕 娜 郭建伟(84)
基于文献计量的国内信息资源领域研究现状与对策分析	凡庆涛 周 雷 杜 贇 谢海涛(89)

人工智能技术对计算机网络技术的应用及存在的问题	何 葳 陈 茜 李世玉(98)
语义检索——专利检索技巧与方法	李时玉 何 葳 孟 莹(102)
区块链技术在专利保护上的应用研究	魏 晨 龚 龔 鲁 啸 马 燃(105)
基于图像识别的数据新闻自动生成方法	谢海涛 肖 倩 凡庆涛(111)
基于文献计量的北京市国际科研合作伙伴城市特征的研究	周 雷 杜 贇 凡庆涛(117)
大数据时代科技查新服务的转型升级	张云倩 燕 娜 朱诚岩(126)
大企业开放创新平台的特征和模式研究	孟 潇 邵 颖 杨海丽 毛维娜(131)

综合类

美国《拜杜法案》对科技成果转化的促进及其启示	付 宏 张一博 夏梦虎(139)
国内省级科技情报机构科技查新现状调查与分析	凡庆涛 张云倩 刘平平 陈 数(147)
京津冀地区科技创新能力比较研究	童爱香 孙艳艳 张 红(157)
中文电子信息资源的使用现状分析与评估——以北京市科学技术研究院 CNKI、万方和 维普数据库为例	凡庆涛 杨 萍 杜 贇 郭建伟(168)
数据库资源-期刊竞品与用户检索行为分析——以超星期刊为例	郭建伟 燕 娜 陈佳宇 李时玉(178)
产业技术联盟推动京津冀创新共同体发展与建设综述浅析	刘平平 谢海涛 凡庆涛 邵 颖 张云倩(182)
面向 2035 年的我国电力企业发展环境及路径研究——以华电重工为例	薛 菲 翟 杨(188)
大数据时代年鉴工作的创新启示	申峥峥(193)
城市科技创新能力比较研究——2016 年北、上、广、深科技创新能力实证研究	童爱香 张 敏 张 红(197)
北京市研发经费投入的思考	西桂权 王冠宇 杨濛濛(206)

美国国家实验室的运营与技术转移机制特点——以劳伦斯·伯克利国家实验室为例	许玥姮 刘光宇 丛 琳(212)
人工智能视角下面向科研院所的知识服务新走向	燕 娜 张云倩 郭建伟 翟 杨(217)
自由基聚合理论视域下北京新经济发展路径研究	杨海丽 孟 潇 李成龙 杨 柳(222)
基于配位聚合理论的北京怀柔科学城建设路径分析	杨海丽 孙若丹 毛维娜(231)
美国科学促进会科研发展浅析及对我国科技发展的经验借鉴	詹同叙 张 媛 高子涵 张晓静(241)
京津冀科技资源协同与创新发展研究	张 红 孙艳艳 童爱香(247)
国外科技服务业集群形成和发展的政府策动模式和机制评述——基于“钻石模型”的分析	张 敏 卢凤君 张 红 苗润莲(255)
北京科技社团智库建设初探——以北京科学技术情报学会为例	邵 颖 刘彦君 刘 如 郝 琦 刘平平(262)
基于 Innography 平台的黄芪提取物全球专利情报分析	张素娟 陈佳宇 董 洁 曹 元(271)
用于治疗 2 型糖尿病的中药及其代谢组学研究	张素娟 申利娜(280)
发达国家科技人才发展战略浅析及经验借鉴	张晓静(289)
大数据时代发达国家科技创新产业联盟政府作用研究	张云倩 谢海涛 刘平平 陈 树(294)
网络环境下科技传播研究	张子成(298)
数据资源采购与用户调研分析	郭建伟 燕 娜 陈佳宇(301)
知识服务“场景”化研究初探	赵 龙(305)
石墨烯材料技术发展及产业化影响简析	孙棕檀(310)
基于政府行为的科技创新中心建设发展研究	张玉娟 杨海丽 孙若丹(318)
浅析重大科技基础设施在基础研究中的关键作用	李梦茹 孙若丹(323)
科技声像情报在成果转化中的应用研究	黄 宇 廖 涛 陈 茜(337)
区块链技术发展态势及应用前景	刘丹枫(340)

区块链时代与机遇伴生的安全威胁·····	隆苏妍 靳晓宏 樊艳芳 郝琦 李佳娛	(343)
展会的情报工作效用以及展会情报工作的开展·····	鲁啸 魏晨 王冰琪 卢絮	(347)
科技创新在构建和谐宜居城市中的重要作用·····	罗婧璇	(352)
信息时代下电子期刊与纸质期刊之间的互补和利用·····	马燃	(356)
浅析北京“三城一区”如何联动发展构建科技创新中心·····	李梦茹 孙若丹	(359)
大数据下科普传播途径分析·····	陈茜 廖涛 黄宇 何葳 孙天垚	(362)
基于文献调查分析中医药治疗 2 型糖尿病的研究进展·····	董洁 张素娟 邓奕	(365)
科技情报产品的二次开发新方向初探·····	侯丹 陈菲 邢哲 孙棕檀	(372)
市属科研机构内外部竞争环境分析·····	王楠	(376)

图书馆类

“双一流”背景下图书馆对科研新手快速成长的支撑作用·····	冯菁 王昕	(383)
国内外科技创新中心建设与发展对比分析研究·····	盛穆彤	(388)
高校图书馆文献资源建设探索·····	秦疏影 关静霞 孙晶岩 归燕	(393)
互联网环境下高校信息服务与阅读心理的互适性研究·····	高淑芳 赵世华	(398)
高校图书馆向社会公众开放的原则和对策分析·····	王昕	(403)
基于大数据的科技资源分析与服务平台建设研究·····	王爽	(407)



情报信息类

基于多源数据的前沿科学领域与 新兴研究方向识别和遴选方法研究^①

曾 文¹ 李 辉² 李 荣²

(1. 中国科学技术信息研究所, 北京, 100038; 2. 北京市科学技术情报研究所, 北京, 100044)

【摘 要】新一轮的科技革命正在兴起,若要在激烈的国际竞争中快速提升科技创新能力,占领前沿技术的制高点,则必须牢牢把握科技进步的方向,掌握科技发展的最新动态,制定战略性的科技发展规划。从浩瀚的科技信息数据中识别和遴选前沿科学领域和新兴研究方向是实现科技创新的基础,也是国家制定科技发展战略的重要依据,但国内研究与国外研究相比仍有较大距离。大数据的涌现和技术的日益成熟需要智能化和科学化地识别和遴选前沿科学领域和新兴研究方向。本文研究和确定前沿科学领域和新兴研究方向的定义,分析并总结国内外已有研究方法,提出将研究对象从以传统的单一数据为主转向以多源数据对象为主,运用情报学和计算机科学的研究方法和工具,实现定性 with 定量方法的科学有效融合,提出前沿科学领域和新兴研究方向识别和遴选的方法论。

【关键词】 前沿科学领域;新兴研究方向;识别;遴选;多源数据

1 引言

过去数十年具有重大变革性影响的前沿科学和新技术,正在影响全球的科技环境和经济环境。中国科学院院长白春礼认为:“世界历史上已发生的五次科技革命,中国错失了前四次科技革命的机遇。第五次科技革命中,中国也只是一个跟踪者,而且是一个没有取得优良成绩的跟踪者,即将到来的第六次新科技革命涉及科学和技术的深刻变革,为中国科技发展提供了难得机遇。中国再不能与其失之交臂,中国必须要抢抓机遇,前瞻布局,以期在新一轮科技革命中赢得主动。”纵观几次科技革命,每次都深刻地影响了世界力量的格局,西方发达国家依靠领先世界的前沿科学技术而称雄世界。因此,只有掌握前沿科学技术,拥有强大的科技创新能力,才能掌握国际竞争和发展的主动权,从根本上保证国家安全。科技创新是历史进步的动力、时代发展的关键,创新发展居于我国发展全局的核心位置。2016年5月,习近平总书记在全国科技创新大会上发表重要讲话,从战略和全局的高度分析了我国科技创新的历史方位、时代定位和国际地位,围绕建设世界科技强国的目标进一步明确了科技事业发展的总体布局。李克强总理强调要发挥科技创新在全面创新中的引领作用,塑造更多依靠创新驱动的引领型

^① 本文为北京市科技计划项目“重点领域前沿技术识别及跟踪研究”(项目编号:Z181100006018004)的成果。通讯作者:李辉。

发展。我国近三十多年来,在载人航天、载人深潜、超级计算等科技领域已经取得了世界瞩目的成就,但在基础研究科学领域和科技自主创新能力方面还存在不足。当前,第六次科技革命正在兴起,为了在激烈的国际科技竞争中,加速提升自主创新能力,占领前沿科学技术的制高点,把握前沿科学领域和新兴研究方向,超前科学预测和规划科技研究发展战略是极其必要的。在 20 世纪四五十年代,美国等发达国家已把前沿科学技术预测作为国家实施科技前瞻布局和创新发展的“武器”,对其进行全球产业技术分布发挥重要作用。日本自 20 世纪 60 年代末和 70 年代初已开展系统性的科技预见研究,这为其科技和经济发展起到巨大作用。目前,我国已有部分技术领域处于“领、并、跑”状态,未来如何实现引领性的发展,对世界前沿科学技术的预判尤为重要,而实现前沿科学领域和新兴研究方向的识别与遴选是前提和基础。

2 国内外研究现状及存在的问题

前沿科学领域与新兴研究方向是近年来国际组织、政府机构、科技管理机构,以及各国、各领域科研人员关注的热点研究问题之一。前沿科学领域与新兴研究方向并无非常明确的定义,是两个相互联系又存在差异的概念。共同之处是:两者的研究目的均是为国家和学术界揭示具有前瞻性、先进性和创新性的前沿研究与重大进展。不同之处是:①概念范围不同,前沿科学领域是较宏观的概念,是指一个领域的前沿,而新兴研究方向是较微观的概念,一个领域可以包括若干个研究方向;②数据来源不同,前沿科学领域常见于各种科技规划文本、政策文档、基金项目指南、政府“优先领域”或“重点领域”的报告中,而新兴研究方向则常见于各种科技论文、研究报告、专利等科技文献中;③识别和遴选方法不同,前沿科学领域多采用“专家智慧+定性分析”的方法,而新兴研究方向多采用“科技文献+定量分析”的方法。

(1) 研究现状分析

经 Scopus 数据库文献统计(1900 年 1 月至 2018 年 8 月)可以发现:自 20 世纪 90 年代起,有关前沿科学领域和新兴研究方向的研究呈急剧增长态势。2016 年达到顶峰,年发表论文数量为 2 452 篇。从学科类别分析来看,Scopus 数据库中包含的 27 个学科领域中的研究人员都在关注前沿科学领域的发展,说明前沿科学领域是所有学科领域共同关注的热门。从国家分析来看,美国、中国和英国是从事前沿科学领域和新兴研究方向研究的主要国家。

1) 前沿科学领域的识别和遴选方法

前沿科学领域的识别和遴选的主要途径是搜集和研究世界科技强国的战略规划、路线图、战略目标、愿景,国际大型研究计划,权威机构或部门发布的目标领域重要研究报告,重要资助机构最新部署的项目,项目申请书中的研究内容,重要的国际组织、学会、协会、科研团体撰写的有关科技发展前沿的研究报告或战略文件^[1]。

目前前沿科学领域的识别和遴选大多由国际组织、政府、权威机构或部门、科研管理机构、科技资助机构组织相关领域的专家,以现场研讨、在线讨论、通信调查、访谈等多种方式,汇集不同领域专家的知识 and 智慧,形成最终的前沿科学领域预见。例如:世界顶级科学期刊 *Science* 于 2005 年通过征询本刊高级编辑委员会、审稿编委会以及编辑和作者的建议,公布了 125 个未来 25 年最具挑战性的科学前沿问题。美国国家卫生基金会(NSF)2006 年 8 月 1 日发布了题为“进化生物学前沿(Frontiers in Evolutionary Biology)”的研究报告,其对未来进化研究领域前沿进展的识别是通过专家组对新兴研究主题的判定而作出的。联合国政府间气候

变化专门委员会(IPCC)集各领域数以百计首席专家的技术专长来撰写在气候领域具有重要影响力的 IPCC 评估报告。我国国家自然科学基金委员会设立的科学部专家咨询委员会主要由相关领域的战略科学家组成,负责为科学部的优先领域和资助格局、重大研究计划和重大项目立项、学科发展战略等具有战略性的资助决策与管理工作提供咨询建议和指导性意见。此外,许多权威学术网站或者团体公布的当年的研究热点或科技进展等都是经过领域专家们的评选而得出的。

综上,已有前沿科学领域的识别和遴选方法基本是以专家经验为基础的定性分析方法,其缺点是:①研究方法问题,以专家知识和认识为支撑,具有较强的主观性;②研究对象问题,以宏观性科技政策类数据为主,数据源较单一,若数据量不够,则会影响识别和遴选的结果;③成本问题,识别和遴选过程花费的成本较高,时间长,适合以国家为主体的前沿科学领域的识别和遴选活动。

2) 新兴研究方向的识别和遴选方法

新兴研究方向的识别和遴选多数以科技文献为核心,利用科学计量学和统计学方法,对各类学者发表的科技论文、专利、专著等进行挖掘,探测某一科学领域内在的规律和特点,进而揭示领域内新兴的研究方向。

新兴研究方向的识别和遴选方法主要包括以下几种。①基于引文的方法:该方法主要有基于共被引的方法^[2-5]、基于文献耦合的方法^[6-7]、基于直接引文的方法^[8-10]和基于引文的主路径分析方法^[11]。该方法存在的问题是引用时滞问题:论文达到一定的引用频次或高被引需要一定的时间,所以无法对新兴的研究热点和新出现的研究趋势进行识别和评价。②基于主题词的方法:该方法主要有爆发监测算法^[12]、共词方法^[13-14]和概率主题模型方法^[15-18]3类。该方法不存在时滞问题,利用自然语言处理技术和主题统计模型从自由文本数据中识别有重要意义的研究主题(词语),但由于词语在不同的语境下会表达不同的含义,单个词表达的意思并不具体,需要由用户依据自己的领域知识来判断是否是前沿研究或新兴研究方向。因此,对研究方向识别结果的解读是因人而异的,这种方法存在较多的主观因素影响。③基于引文和主题词的复合方法:该方法主要有结合共词和引文的方法^[19]、结合主题词和引文的方法^[20]以及结合全文本和引文的方法^[21]。该方法虽然弥补了基于引文与基于主题词的方法各自的缺点,但在大量数据应用时,该方法变得复杂、烦琐,实用性差。④其他方法:基于文献计量学指标的方法^[22]、直接聚类方法^[23]、基于 ESI(Essential Science Indicators)的定性和定量相结合的方法^[24]、使用神经网络模型探测研究前沿的方法^[25]、基于知识地图探测研究前沿的方法以及使用补充计量学探测研究前沿的方法^[26-27]等。这些方法针对上述方法的不足,进行数据和方法的侧重性研究,研究成果缺乏应用性。

综上,已有的新兴研究方向识别与遴选方法基本是定量分析方法,其缺点是:①研究方法问题,以数据内容分析为主,具有较强的客观性,会影响识别和遴选的准确性;②研究对象问题,数据源以某一类型科技文献研究为主,缺乏数据的融合;③研究视角问题,依赖国际上普遍使用的 Web of Science、Scopus 等数据库,虽然该数据库中包括论文、专利、图书、述评等多种科技文献类型,但其所关注的重点更倾向于微观的个体研究行为,而忽略了研究机构、国际组织、科研管理机构等发布的科技规划文本与正在研发的项目数据中蕴含的近期和远期的前瞻科学研究前沿思想信息。

(2) 结论

① 研究目标具有一致性。尽管前沿科学领域与新兴研究方向在概念上有差异,但是在本质上有着必然的联系,前沿科学领域和新兴研究方向识别和遴选结果均代表着国际科学技术的先进前沿,对我国科技事业的发展起到指南针的重要作用。

② 研究对象具有局限性。无论是前者依赖于专家的定性分析,还是后者注重文献的定量分析,面对当前信息数据的大数据量和多元化,以及科技发展的多学科、多领域交叉融合的趋势,传统单一的研究对象不利于分析。

③ 研究方法应具有互补性。精准的前沿科学领域识别可以指导新兴研究方向的精确辨识。同样,在一定程度上,新兴研究方向存在超前于前沿领域定性的现象,因此,对新兴研究方向的正确识别可以优化前沿科学领域的识别和遴选,反之亦然。

3 需要解决的关键问题和研究方法

实现前沿科学领域和新兴研究方向的识别与遴选的核心是数据源和数据分析方法,主要涉及 3 个关键问题的解决。

(1) 数据源的筛选与确定

应获取的数据源包括:部分世界重点科技情报机构的数据,Web of Science、Scopus、中国万方、中国知网、国家科技报告系统的科技文献数据,国内科技政策文本数据。待筛选与确定的数据源是:①常见学术社交网络数据(Google Scholar、ResearchGate、arXiv、CiteULike)、文章引用数据和在线关注行为数据,如阅读次数、被转发次数等;②出版商的使用数据,如Springer、Nature、ScienceDirect 的下载数据、浏览数据等。分析重要的社交网络和数据商使用的数据资源的数据特点,进行定制化的数据采集、筛选与确定等处理研究,并保证获取数据的质量。

(2) 前沿科学领域和新兴研究方向的动态挖掘与感知

从海量多源的科技信息中挖掘并分析前沿科学领域和新兴研究方向是大数据环境下科技情报预测服务的一个关键问题,也是实现非结构化知识发现的一个难题。前沿科学领域和新兴研究方向的动态挖掘与感知的研究目的是通过过滤海量多源和实时的科技信息数据,动态识别数据中表达的研究主题、研究趋势和创新内容,这些数据内容含有前沿科学领域、新兴研究方向信息,是后续数据的遴选基础。可以通过对科研产出数据内容的要素和条件进行分析,分别建立研究主题、研究趋势和创新内容要素的关键变量和语义关系,建立研究主题、研究以趋势和创新内容为主的本体模型。并基于模型构建研究主题、研究趋势和创新内容的动态挖掘模板,以短语片段形式抽取可以代表前沿科学领域、新兴研究方向的深度数据内容,并加以聚类,实现动态挖掘与感知。

(3) 前沿科学领域和新兴研究方向遴选指标体系和权重模型的构建

指标体系是实现科学化遴选的基础,权重是评判指标重要程度的依据。权重模型构建方法主要包括主观赋权法和客观赋权法。主观赋权法会由于决策者的经验或偏好,带有一定的主观随意性。客观赋权法利用了数据的客观信息,但是忽视了专家的经验信息。因此,需要针对前沿科学领域、新兴研究方向分别建立遴选指标内容,对于一致的指标内容,其权重也应各有侧重。可以根据前沿科学领域、新兴研究方向的定义和差异化特点,建立可以有效地、多维度地体现两者共性和差异性的前沿科学领域和新兴研究方向的遴选指标体系,可以利用神经

网络方法进行指标的筛选和权重的设定,即综合使用主客观权重模型构建方法来确定权重。

基于多源数据的前沿科学领域与新兴研究方向识别和遴选的方法体系如图 1 所示。

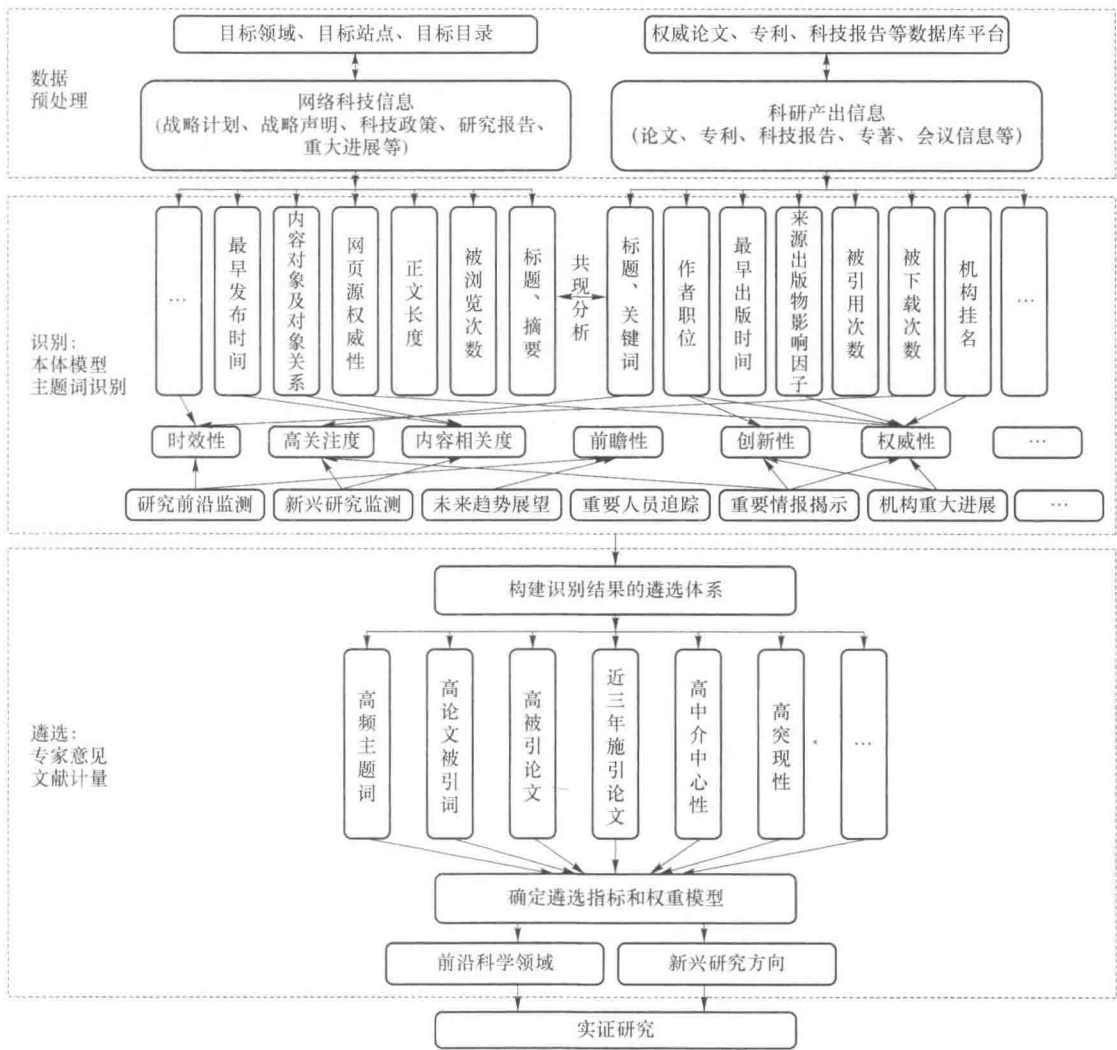


图 1 方法体系框图

4 结论

历史和现实都说明,前沿研究领域和新兴研究方向的研究是促进科学技术发展的“利器”,也是世界各国科技界的研究重点。在大数据时代,对科学前沿研究和技术进行有效预判,掌握世界科学技术发展的动态,占据科学技术研究的主动性,对于提高国家科技创新能力尤为重要。其中有效地、多维度地实现前沿研究领域和新兴研究方向的识别和遴选是科技预判研究的重点和核心内容。本文解析相关定义,提出前沿研究领域和新兴研究方向的识别和遴选的研究思路 and 关键技术等,下一步将重点开展相关研究工作,付诸实践和应用。