

基于科技大数据的 医药健康产业情报服务系统框架设计

陈晓玲^{1,2} 张世彤¹

(1. 吉林省科技创新研究院, 吉林长春 130000; 2. 吉林省科学技术信息研究所, 吉林长春 130033)

摘要: 科技大数据为产业应用研究与服务带来了重大的机遇和挑战, 迫切需要科技大数据管理驱动产业情报研究和战略决策服务的质量。在综述科技大数据与医药健康产业应用现状的基础上, 分析存在的问题和相互作用机理, 提出“科技大数据支撑、产业科技创新协同、信息精准服务、前沿技术监测”的产业情报服务模式, 构建涵盖政策、产业、科研成果、设备、科学数据、机构、人才 7 类科技大数据的医药健康产业情报服务体系, 开展智能检索与数据可视化、动态跟踪与智能推送、产业科技创新地图展示、前沿技术监测与预警等应用研究。在此基础上, 构建科技大数据中心, 提出基于科技大数据的医药健康产业情报服务体系模型及功能设计, 以提高医药健康产业科技大数据的应用研究和决策支撑服务, 并运用科技大数据管理与技术提出医药健康产业科技创新发展的实施步骤, 以及科技大数据情报分析和创新地图, 促进产业升级转型和高质量发展。

关键词: 科技大数据; 医药健康; 情报服务; 情报服务模式; 服务体系

DOI: 10.3772/j.issn.1674-1544.2023.05.007

CSTR: 15994.14.issn.1674.1544.2023.05.007

中图分类号: R-1; TP391.1; G350

文献标识码: A

Design of Medical and Health Industry Information Service System Framework Based on Science and Technology Big Data

CHEN Xiaoling^{1,2}, ZHANG Shitong¹

(1. Jilin Science and Technology Innovation Research Institute, Changchun 130000; 2. Jilin Provincial Information Institute of Science and Technology, Changchun 130033)

Abstract: Big data in science and technology has brought significant opportunities and challenges for industrial application research and services, and there is an urgent need for big data management in science and technology to drive the quality of industrial intelligence research and strategic decision-making services. This paper reviews the current situation of science and technology big data and medical and health industry applications, analyzes the existing problems and interaction mechanisms; proposes an industrial intelligence service model of “science and technology big data support, industrial science and technology innovation collaboration, precise information service, and frontier technology monitoring”; and designs and implements a system that covers policy, industry, research results, equipment, scientific data, institutions, and talents.

作者简介: 陈晓玲 (1984—), 女, 硕士, 吉林省科技创新研究院副研究员, 研究方向为信息资源服务、科学数据管理和情报研究 (通信作者); 张世彤 (1975—), 男, 吉林省科技创新研究院研究员, 研究方向为科技成果转化、科技大数据管理和情报研究。

基金项目: 2022 年基本科研经费专项资金项目“吉林省科技成果服务平台建设研究经费” (2022001); 2021 年“数字吉林”建设专项资金项目“吉林省科技大数据融合应用平台建设项目” (2021038); 2022 年吉林省科技发展计划项目“基于知识图谱技术的产业关键技术大数据分析系统的研发与应用” (20220201143GX)。

收稿时间: 2023 年 3 月 14 日。

It provides intelligent retrieval and data visualization, dynamic tracking and intelligent pushing, industrial science and technology innovation map display, frontier technology monitoring and early warning, and other application research. A big data center of science and technology is constructed, the model and functional design of the intelligence service system of pharmaceutical and health industry based on big data of science and technology are proposed, the application research and decision support service of big data of science and technology of pharmaceutical and health industry are improved, the implementation steps of science and technology innovation development of pharmaceutical and health industry are proposed by using management and technology of big data of science and technology, and the intelligence analysis and innovation map of big data of science and technology promote industrial upgrading Transformation and high quality development.

Keywords: big data of science and technology, medical health, information service, intelligence service model, service system

0 引言

大数据产业的数字产业化、产业数字化,也为医药健康产业高质量发展增加新动能。2015年,国务院^[1]印发《促进大数据发展行动纲要》,提出“公共服务大数据工程中构建医疗健康服务大数据”。为围绕医药健康领域开展产业数字化建设与应用研究,出台了一系列政策措施。2016年颁布的《关于促进医药产业健康发展的指导意见》提出“推动政产学研用深度融合”^[2],2021年颁布的《全国一体化大数据中心协同创新体系算力枢纽实施方案》强调统筹和建设行业领域数据中心^[3]。当前,科技大数据已在医药大数据、医疗健康大数据、医疗数据、大数据可视化、基因数据库等领域得到了应用研究,为医药健康发展提供了数据支撑与信息服务。随着大数据、物联网、人工智能等技术的发展,医药健康的科研人员对科技大数据的应用需求从文献信息向自动化、智能化的情报服务转变。如何利用医药健康领域科技大数据集中管理,实现对产业热点、前沿技术、关键技术的跟踪监测和对风险的分析管理;如何构建医药健康产业情报服务体系,实现产业科技服务的精准对接,解决产业发展需求甄选和关键技术“瓶颈”问题,建立有效的合作创新体系,促进科技成果转化^[4]。这些问题已是当今业界关注的重点。因此,本文试图解决以下问题:①梳理国内外科技大数据与医药健康产业大

数据应用现状,总结医药健康产业科技大数据应用研究面临的问题,并阐述科技大数据与医药健康产业科技创新的作用机理;②建立基于科技大数据的医药健康产业情报服务体系设计,构建科技大数据中心,提出情报服务模式,设计体系模型、组成及功能;③提出科技大数据为医药健康产业情报服务体系应用。本文将以此为出发点,就“科技大数据”“医药健康”和“情报服务”3个关键问题构建体系设计 and 应用,为医药健康产业科技创新生态服务中的科技大数据管理和情报服务提供应用研究,为进一步开展全产业链的情报服务奠定基础。据此,医药健康产业科技大数据情报服务体系需要设计一种统筹的系统架构,以医药健康产业科学研究应用场景模式为导向,将专题的科技大数据组织、关联、服务进行整体规划,以较好地驱动医药健康产业科研工作的知识化、智慧化、协同化的精准服务。

1 科技大数据及其应用现状

1.1 科技大数据内涵及特征

Williams等^[5]将科研活动中涉及的文献、专利、图书、报告等科技信息称为学术大数据。曾文等^[6]认为,科技大数据是一种特殊类型的大数据,指与科技信息相关的非数值型数据,最常见的一种科技大数据即科技文献数据,还有政策数据、成果数据、产业数据、人才数据等。科技大数据具有数据量大、多样化、数据价值密度化、

速度快、时效高和数据在线等特点,也具有敏感性和积累性,科技大数据的处理和分析具有复杂性。

1.2 国外科技大数据应用现状

国外对大数据建设的投入较早且取得了显著的成效,以科技大数据管理平台或工具为主。

①在国外科技大数据管理平台方面:美国、英国、日本等均建立了大数据创新中心,促进了对数据管理、公开、检索的研究,且增强了大数据生态系统的内部联系^[7]。美国哈佛大学定量社会科学研究中心的Dataverse平台,提供了广泛的社会学、人文学、地球与环境学、医学等数据库^[8]。Baumann等^[9]提出构建地球大数据分析Earth Server系统,提供了齐全的、生产级的、灵活的、可互操作的数据服务。美国政府数据网data.gov向公众提供政府数据的访问、下载和使用等服务,加强政府数据的开放性^[10]。世界上最大的社会科学数据中心是美国校际社会科学数据共享联盟ICPSR (Inter-University Consortium for Political and Social Research),为全球提供社会科学数据访问、管理和分析方法等服务^[11]。美国Dryad数据仓储覆盖生态学、医学、环境科学、自然科学等领域的研究数据,实现论文与科学数据的关联,提供开放获取、重用与存储^[12]。英国最大的社会科学和人文科学数据库是数据存档库UKDA (The National Archives of United Kingdom),它包括国家统计局、国家社会研究中心和人口普查数据等资源^[13]。②在国外科技信息分析平台方面:Web of Science、wizdomAI、Dimensions、Incopat、SciVal、Incites等文献、专利、项目多类型的综合性数据评价与分析、自动生成报告的智能平台^[14]。VOSviewer、CiteSpace、Bibexcel、Sci2、Ucinet等科技信息知识图谱工具,提供论文、专利、文件等多类型多维度的可视化分析与数据挖掘^[15]。

1.3 国内科技大数据应用现状

国内科技大数据主要以数据管理平台和数据应用系统为主。①在研究数据或科研数据管理平台或工具方面:国家科技基础条件平台中心的中

国科技资源共享网,包括科学数据、生物种质与实验材料、重大科研基础设施和大型科研仪器等资源,涵盖国家的20个科学数据中心和31个资源库,具有资源检索、数据汇交、数据管理等功能^[16]。中国科学院科学数据中心提供数据集、软件、论文、著作、标准、关联发现等多源数据服务,涵盖16个所级中心和18个学科中心,建立标识服务、数据汇交、数据出版、支撑工具、数据安全、元数据术语等数据挖掘分析与学科应用平台,促进学科领域科学数据的深度应用^[17]。北京大学开放研究数据平台面向机构内部或社会的学者开展研究数据的保存、分享与再利用活动的信息网络平台,提供创建、编辑、发布、查询等数据空间和数据集服务^[18]。复旦大学社会科学数据平台汇聚中国社会经济发展数据,提供数据服务、定量研究方法教学、跨学科研究、社会调查等服务^[19]。②利用科技大数据知识发现、情报分析进行科技信息监测与知识获取方面:清华大学科技情报分析与挖掘平台(AMiner系统),包括论文、学者、引用关系、知识实体等科研数据,利用数据挖掘和社会网络分析与挖掘技术整合数据源,提供学者画像、AI报告、人才智库、趋势分析等服务^[20]。中国科学院图书文献情报中心科技文献知识人工智能引擎(SciAIEngine系统),利用科技文献大数据和深度学习方法,提供文献摘要、语义识别、文献分类、关键词识别、命名实体识别、概念定义句识别等服务^[21]。中国农业科学院科技信息研究所农业专业知识服务系统,基于农业领域的文献、项目、科学数据、行业报告、专利、专家和学者、机构等知识资源,采用集成知识组织、知识计算、深度搜索、协同众包和可视化分析等关键技术构建农业智库,面向农业战略咨询、科技创新和农业生产提供专业化知识服务^[22]。国家科技图书文献中心专题情报数据管理与智能分析平台,集成论文、专利、基金项目等多来源多类型数据,面向成员单位和服务站的情报分析人员,提供专题数据汇聚、清洗规范功能,从计量分析、内容分析、模型分析等方面提供专题情报分析功能^[23]。广西科技大数据平台

利用“互联网+”和大数据技术,以共享机制为核心,以资源整合为主线,为管理者及决策者进行全局性层面的管理及决策提供支持^[24]。

综上所述,国内外都高度重视科技大数据信息化工作,研制了相关分析工具和数据库,基于科技资源和科学数据,研发数据监测模型、监测算法和工具,在数据资源系统和工具取得了重大突破,但利用科技大数据为产业科技创新服务的科研匮乏。鉴于此,本文提出利用科技大数据管理促进产业科技创新与应用的设想,支撑产业科技管理和科学决策。

2 医药健康产业科技大数据及其应用

2.1 科技大数据的组成

科技大数据作为医药健康产业大数据的一个分支,与医药健康产业科技活动、科学研究紧密相关,是医药健康产业各类科学数据资源的总和。目前,医药健康产业的科技大数据主要包括科技投入、企业培育、项目建设、品牌培育、品种培育、创新平台、科技成果转移转化、金融支持、人才培养、优化环境等政府措施;医药健康产业分类包括医疗与健康服务、医药流通、制药设备与检测仪器、医疗器械与医用健康材料、化学药业、生物医药产业、中药产业等;医药健康产业链包括原料、研发、生产、流通、消费;产业数据包括产业目录、医疗器械、药品信息、制药企业、新品种新产品等。

2.2 面临的问题

翟运开等^[25]提出基于“互联网+医疗健康”的医药电商平台,提供的在线挂号、在线问诊、专家会诊、线上医药销售等功能和服务。李亚凡等^[26]提出构建生物医药和大健康产业可视化平台设想,直观展示云南省生物医药产业成果数据。孟彦霖等^[27]提出依托药品销售数据的可视化展示探究,对比购药患者的行为、性别、年龄段,对比各类药物销量、各省份的药物销量。刘倩^[28]提出了大数据产业链的“T”型价值链结构,区域科技创新发展要素由知识集聚技术创新、产业结构转型升级、大众创新万众创业、科技成果转移

转化、科技创新管理构成,大数据产业推动区域科技创新的驱动效应、集聚效应、协同效应和规范效应的作用机制和对策建议。李梅等^[29]从空间视角研究京津冀协同创新,提出以资源要素协同、主体协同、产业协同、布局协同和政策协同的创新发展地图系统,为区域协同创新评价指标提供数据,为政府决策提供科学依据。龚轶等^[30]从地方空间、流动空间维度,结合产业和创新层面构建城市群协同创新的系统,包括环境层、资源和要素层、产业分工/联系层、协同创新网络层。

目前,医药健康产业大数据主要侧重药品数据与销售数据的应用研究。科技大数据管理主要存在以下问题:①数据分散性强导致应用效率低。科技管理部门和科研单位长期形成的数据库和数据平台的建设、使用和管理分散,造成跨部门数据统计、汇总、分析等困难,导致应用效率低下。②数据统一性弱导致数据流动性差。数据管理缺乏统一的标准。同一个服务对象需要填充多个数据。每次根据不同的填充需求生成不同的数据版本,因此数据一致性较差。③数据标准的不一导致数据适用性差。数据更新缺乏统一的标准。数据更新和验证不及时,导致数据不准确、不完整,随着时间的推移数据可用性较差。④数据碎片化严重导致管理效率低下。对于全方位、多结构化的实时数据和历史数据,特别是隐性数据的在线采集和实时分析,各种科技计划存在重复、分散、封闭、低效等问题。有些申报项目和资源配置还存在“碎片化”等问题,不能完全适应实施创新驱动发展战略的要求。⑤数据参数多,口径不同。各项指标统计口径不同,科技指标项目较少,不能准确、详细、深入地了解科技产业、项目、成果等。

科技大数据管理存在的分散性强、流通性差、标准不一、应用低效等问题,导致利用科技大数据进行医药健康产业情报服务存在一定的局限性,阻碍了医药健康产业科技创新发展。主要表现在以下方面:①传统的情报分析方法研究缺乏全面和充分的产业情报分析研究。②普遍的

产业情报分析主要关注于结构性科技文献、专利等数据资源，缺乏对科技政策、专题报告、成果等半结构化或非结构化数据的产业情报融合分析研究。③现有情报分析工具侧重于特定的数据源或数据类型的统计与分析，由于情报分析工具主要是国外产品，有些分析数据源只能分析外文文献，对中文数据资源分析存在局限性，如TDA分析工具支撑外文数据库的专利和论文文献，CiteSpace、VOSviewer知识图谱工具主要侧重论文、专利文献^[31]。

2.3 科技大数据与医药健康产业科技创新

医药健康产业为了突破科学创新与技术创新的界限，必须将创新数据资源汇聚在一起，基于一定的理论观念和作用机理，充分发挥创新数据资源作用，产生新知识和新技术，由此应用到新工艺、新产品和新产业，最终实现新价值的动态复杂过程，也是促进医药健康产业科技进步和经济发展的动力源泉。

(1) 科技大数据与医药健康产业科技创新的构成。科技大数据由数据资源、数据应用和数据产品组成，即前期的数据资源集成、中期数据资源的挖掘分析及应用、后期知识服务与情报研究的数据产品。医药健康产业科技创新要素构成包括知识汇聚、技术创新、产业结构升级转型、创新主体、科技成果转化、科技创新环境6个方面。知识汇聚和技术创新是出发点和落脚点，科技数据资源的汇聚推进科学创新，形成信息和知识转换为智慧服务，推动产业技术创新。利用新

知识和新技术促进产业创新发展，突破“卡脖子”技术和关键技术，从而带动区域产业结构升级转型^[32]。创新主体要提升创新活力和动能，培育产业的核心企业、重点实验室、示范区等创新载体的创新精神，激发创新载体内部的人才创新活力和动能，提升创新生产实践。科技成果转化有效落地，能够提升产业的商业价值，也是产业技术创新推动经济发展的重点环节。科技创新环境管理要提升产业技术创新的体制机制、政府支撑产业创新政策、人才激励、项目资金支持等，构建新型的产业科技创新生态系统。

(2) 科技大数据与医药健康产业科技创新的作用机理。科技大数据与医药健康产业科技创新两者的关系：一是医药健康产业创新发展离不开大数据资源，科技大数据是医药健康产业科技创新发展的数据基础和保障；二是科技大数据为产业科技创新发展提供了新的平台和技术，推动医药健康产业发展新模式。两者密不可分。

科技大数据与医药健康产业科技创新的作用机理，如图1所示。科技大数据驱动了知识和技术创新、科技成果转化模式、资源优化配置，为医药健康产业科技创新发展提供驱动力；科技大数据聚集了前沿知识和技术、高层次人才，为医药健康产业科技创新发展提供了智库决策；科技大数据带来了管理系统、创新资金的规模效应，科技大数据协同了创新主体、资源共享、知识传递，激活了医药健康产业科技创新发展活力。



图1 科技大数据与医药健康产业科技创新的作用机理

3 基于科技大数据的医药健康产业情报服务体系构建

鉴于医药健康产业科技大数据应用所面临的问题,本文明确了医药健康产业情报服务体系设计需求与建设目标,提出了体系的科技大数据中心构建和情报服务模式,设计了基于科技大数据的医药健康产业情报服务体系的模型、组成、功能和安全保障。

3.1 设计需求与建设目标

根据科技管理在科技大数据的特点,管理部门的情报分析需求侧重于医药健康产业的宏观和中观管理,如决策、预测、制定和分析指标以及制定政策、目标和战略规划等,偏向于宏观和整体数据使用。在科技活动中,在科研阶段为科研人员提供项目选项思路、技术路线图、竞争情报分析、合作交流;在管理阶段,辅助科研人员申报立项、资金支持、团队合作、项目验收、成果发表;在科技成果转化阶段,提供资金投向、市场分析、技术趋势、产业监测与预警等决策支撑。因此,本文的医药健康产业情报服务体系主要面向从事科技管理的部门以及从事产业研究的科研机构、高校、企业、创新主体等用户,目标用户为科技管理决策者,从事产业研究和产业服务、产业转化服务的科研人员和企业管理人员。

针对上述医药健康产业大数据应用需求,本文提出的基于科技大数据的医药健康情报服务体系,涵盖医药健康产业的科技政策、科技资讯、科研成果、科学数据、设备、机构、人才等数据。从政策协同、数据协同、主体协同、产业协同、情报服务5个方面进行应用框架设计^[33]。其目标是建立医药健康产业的政产学研用一体化的情报服务平台,系统地展示近年来医药健康产业协同创新发展现状和趋势,满足多用户的智能化、专业化知识需求和情报服务,为医药健康产业协同创新发展水平和成效提供科学评价和决策支撑。

3.2 科技大数据中心建设

3.2.1 数据的选取及特征分析

本文以科技管理中科技大数据为基础,划分了政策数据、产业数据、科研成果、设备数据、科学数据、机构数据、人才数据7个类型^[34],构建医药健康产业的科技大数据中心,并以科技政策、科研成果、科学数据、科技机构、科技人才为核心(图2),对选取的19种数据进行元数据描述及分析,寻找数据间的特征及关联关系。为了实现与医药健康产业数据的关联,部分数据需要补充其学科分类、产业领域、技术领域等属性。

3.2.2 数据关联研究

科技大数据作为战略性、基础性的大数据资源,其关联研究和利用将有助于医药健康产业创新生态视角下的全域数据融合应用^[35]。依据科技大数据的元数据类型及结构描述,建立医药健康产业的科技大数据关联关系。目前科技大数据关联关系主要有以下4类。

(1) 元数据关联。依据数据密集型科学发现,构建科技大数据关联的产业应用场景,将原始科技大数据元数据的外部 and 内部特征进行描述与分析,建立各数据之间的关联关系和节点,形成逻辑与物理关联,如机构关联、作者关联、项目关联等。柴苗岭等^[36]提出面向农业产业情报分析模型、科学数据与科技文献关联研究,建立基于产业情报分析的科学数据和科技文献关联平台。

(2) 语义关联。对于半结构化或非结构化数据进行语义关联,从数据实体语义层面出发寻找关联关系,增强数据之间的共性关系,如研究热点、共词聚类。苏晓娟等^[37]提出科技大数据的中英语种对应语言关联研究。

(3) 模型关联。为了深入挖掘数据之间的相关性及其距离的方法,需要通过数学方法、统计方法进行算法关联,如共现频次、合作频次等。钱力等^[38]提出科技文献资源的智慧服务的模型设计。

(4) 引用关联。数据之间的引用关系是最基

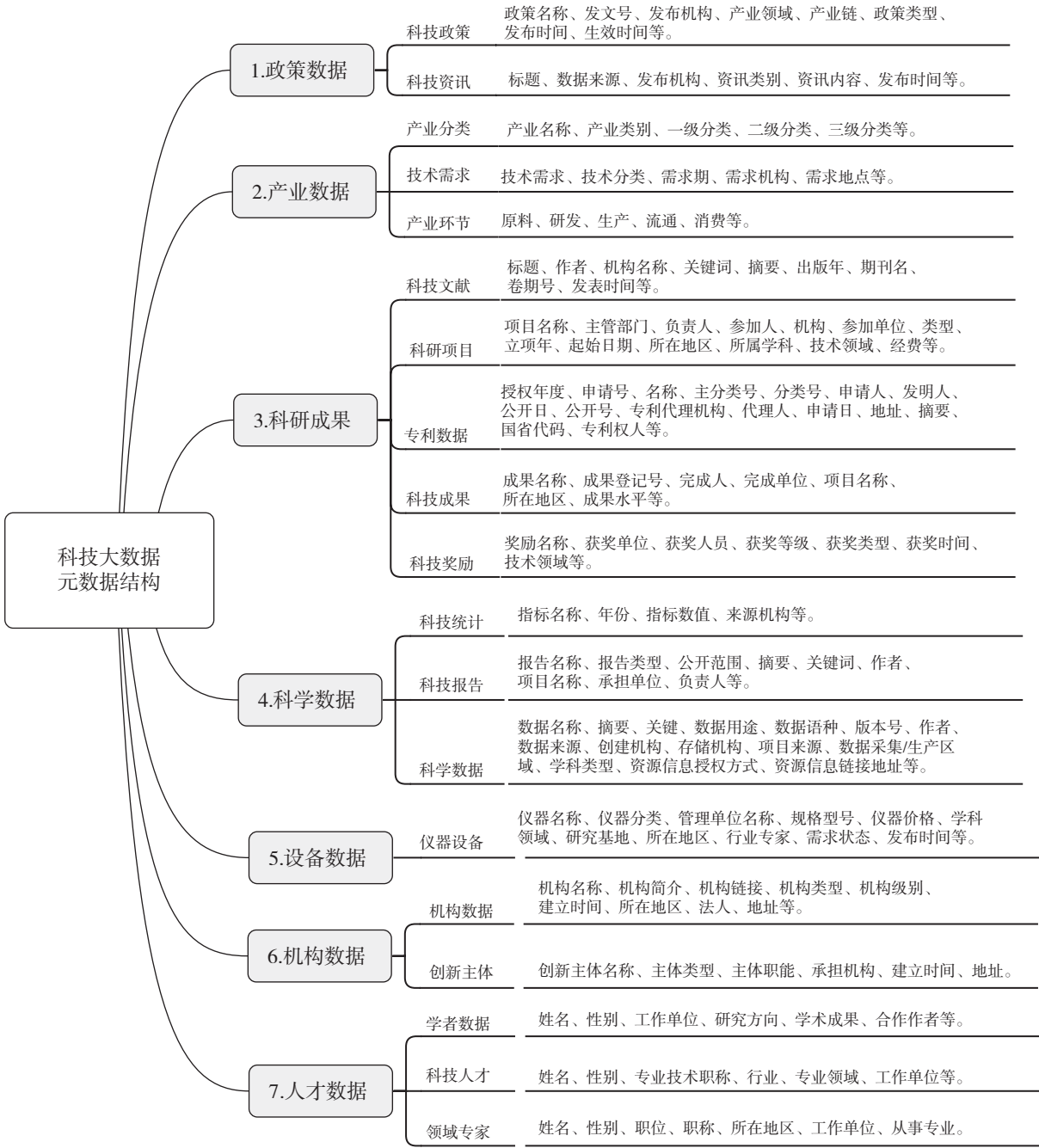


图2 科技大数据元数据结构

本的也是最直接的关联方式，构建引文网络，如科学数据、文献之间的引用关系。周春园等^[39]提出科技大数据的文献引用关联的知识图谱研究。

医药健康产业情报服务需求贯穿于全产业链服务，可以推断出各类用户主体对科技大数据的需求，涵盖多源、多类型、多样化的数据。作为医药健康产业科技大数据应用研究，在产业管理特征上采用多种关联方法有一定的研究意义和利

用价值^[40]。

3.2.3 数据集成设计

根据《科学数据管理办法》，科学数据生命周期包括数据采集、数据存储、数据传输与交换、数据分析等4个方面。本文提出了建立科技大数据的全流程全周期的情报分析，主要有以下4个数据集成阶段。

（1）数据获取。科技大数据采集对多源异构

的结构化、半结构化和非结构化的海量数据进行整合,将分散的、标准不一的数据进行集成,利用ETL工具进行数据抽取、清洗转换,并存储到数据仓库。针对跨部门的数据交换,主要采用协商数据交换、开放数据获取2种方式。网络资源借助数据采集或爬虫工具,利用数据平台或系统的数据进行接口接入,转化为一致的结构化和规范化数据。

(2) 数据存储。科技大数据结构具有复杂性,数据存储要采用善于处理半结构化、非结构化、复杂抽取、数据交换、计算模型等技术,且具有扩展、封装的大数据集成技术来实现,采用云存储、分布式存储、主从数据库等基础架构来适应科技大数据的海量存储、高效计算、灵活的扩展性,并保证数据安全^[41]。

(3) 数据处理。基于数据的特征和情报分析方法构建产业本体特征,形成非结构化数据的语义关联,对数据采集的原始数据进行清洗、补充、合并、规范、一致性检验等预处理,提高科技大数据质量,为决策分析打下基础。

(4) 数据分析。利用可视化分析、数据挖掘算法、预测性分析、语义引擎等技术和方法,对复杂无序的数据进行萃取、提炼和分析,进行科技大数据资源的信息挖掘和利用,利用图形化工具解决分散异构的科技大数据,实现关联、分析并给出完整的图表,达到直观、清晰的效果。还可以利用深度学习、统计方法进行最佳参数模型,建立趋势预测、用户画像、建模等高级知识图谱分析功能,利用科技大数据进行预测性分析、科技研究热点、主题热度趋势等各种前沿分析,为政府、企业调整科技战略和研发投入提供数据参考依据^[42]。

汇聚科技大数据可以对海量异构多源数据进行采集、计算、存储、加工,统一数据标准,并存储形成科技大数据资产,进而为用户提供高效服务;数据与数据间强关联,服务与业务强相关,是医药健康产业独有且能复用的,也是产业数据的沉淀与积累;从成本角度看,也避免了重复建设,减少了信息孤岛式的协作成本。

3.3 医药健康产业情报服务体系建设

鉴于底层科技大数据中心的构建,本文提出的医药健康产业情报服务体系是“科技大数据支撑、产业科技创新协同、信息精准服务、前沿技术监测”的情报服务平台,面向机构或企业的高层决策用户、中层管理用户、一般业务用户的科技大数据服务,提供集成化、智慧化、可视化、个性化的情报服务^[43]。

3.3.1 情报服务体系模型设计

医药健康产业通过设立园区和示范区、构建创新平台、优化产业政策等科技创新布局,促进创新主体和创新要素的流通,加强区域内的协作与发展,取得显著成效。由于协同创新相关的科技大数据分散在线上线下乃至不同机构、不同部门、不同地域,离散化严重,难以融合和利用^[44]。利用情报采集加工方法和大数据技术,对反映产业创新发展的基础数据进行定期跟踪、采集和归类储存,以图谱可视化方式展示,对产业协同创新现状评估和趋势研判具有重要意义^[45]。

医药健康产业情报分析应用服务是面向产业协同创新而研制的,能够动态地、可视化地展现产业科技创新的协同进展、成效,实时更新最新的产业发展动态、前沿技术、科技成果和成果转化等,提供多维度、多样化、交互性强的信息展示应用服务。本文提出的基于科技大数据的医药健康产业情报服务体系模型如图3所示。

体系模型分为基础层、数据层、服务层、应用层4层结构,为机构和企业的用户群体提供医药健康产业创新成果展示、知识展示、情报服务。

(1) 基础层支撑平台运行的服务器、高速网络环境、海量存储与可视化设备等^[46],包括平台的网络设备、储存设备、计算设备、可视化设备等基础设施,提供平台运维的计算服务、网络服务和存储服务。

(2) 数据层对基础空间数据、机构、创新主体和载体数据、科技政策、产业数据、科研成果数据等进行存储和转换,建立统一的标准规范数据库结构,实现不同类型、不同结构、不同系统

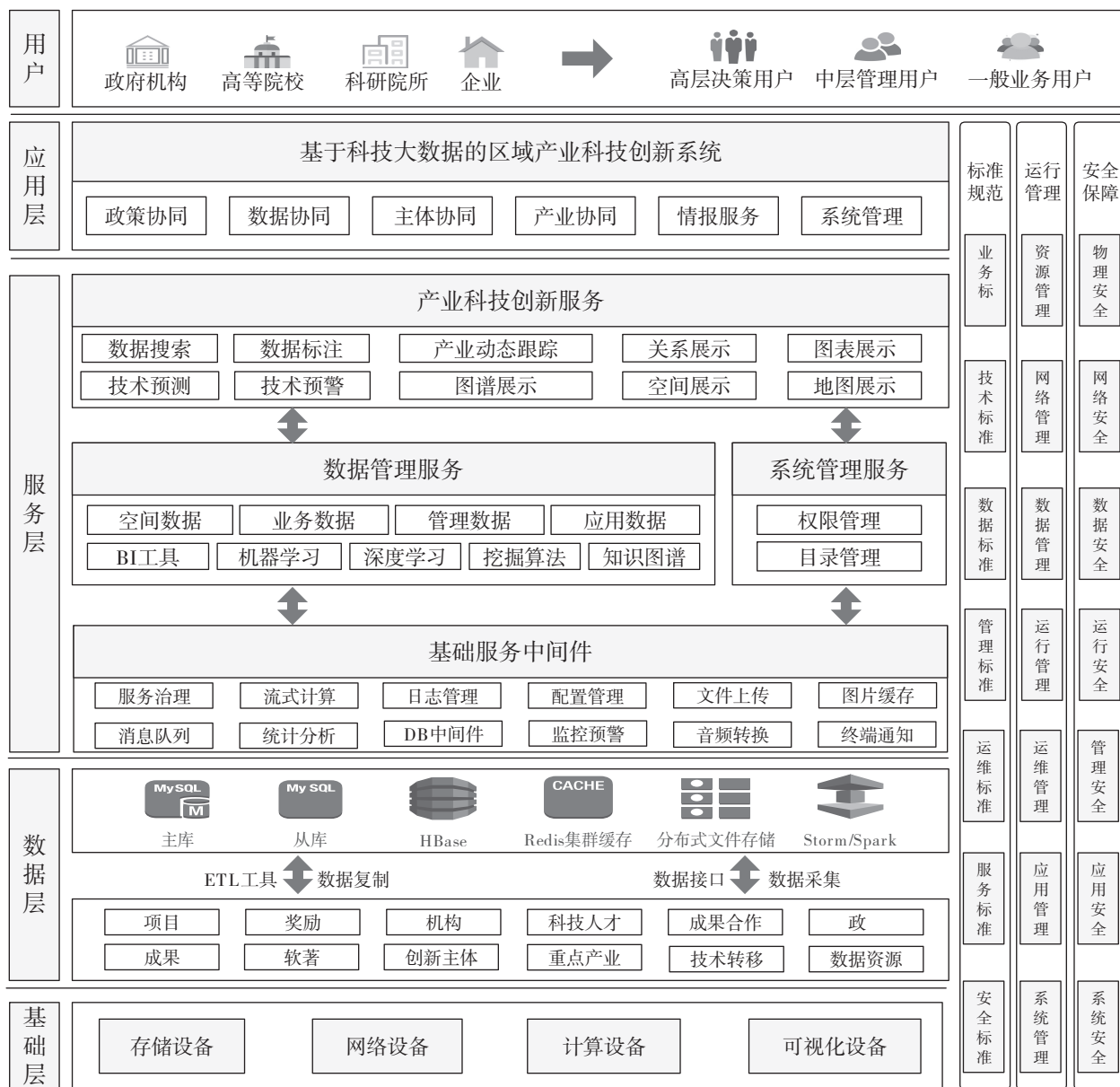


图3 基于科技大数据的医药健康产业情报服务体系模型

间的数据关联与集成。

(3) 服务层采用基础服务中间件进行数据加工和采集、产业地图展示、知识图谱展示和数据管理等服务，建立统一的产业地图展示进行数据集成与共享服务，提供业务服务资源、数据服务资源进行数据治理，用来支撑包括知识图谱、机器学习、大数据集成、网络安全等数据分析服务。

(4) 应用层依托数据资源、创新主体、产业数据、政策数据等进行协同内容的展示与应用，为企业科技创新、政府决策提供知识服务和情报

支撑。提供产业动态跟踪可视化、技术监测与预警、科技情报、科技成果转化、科技金融等服务功能模块，开展专题订阅、技术风险预警、科技中介等科技服务。

3.3.2 情报服务体系功能设计

医药健康产业情报服务体系的数据支撑是科技大数据资源，实现科技大数据资源的协同创新和应用，为医药健康产业提供决策咨询、情报分析、技术服务和应用研究^[47]。因此，根据科技管理中医药健康产业创新和技术服务的需求，利用空间、时间和地理信息来可视化展示医药健康产

业科技创新生态发展现状,揭示医药健康产业的创新资源数据、创新主体和载体的合作数据、重点产业数据、创新政策数据等,数据来源有自有数据、采购数据、合作数据、官网数据、调研数据等,根据来源数据的结构和系统模型,设计数据库表、数据结构、字段、属性和标准制定^[48]。本文提出了科技大数据中心子系统、可视化展示子系统、科技情报服务子系统和系统管理子系统,建立政策协同展示、数据协同展示、主体协同展示、产业协同展示、科技情报分析和系统管理6大功能模块(图4)。

在系统功能设计中,系统的地理位置用来实现企业、高校、科研院所等创新主体的定位;各数据库关联用来实现以机构为载体的相关科技人才、成果、项目、基础设施等数据资源要素的关联,以应用点分布图、热力图、统计图表、词云图、图谱等多种类型方式进行可视化展示。

(1) 创新政策协同是系统的关键环节,主要提供政策分类、查看、统计、政策热点等功能,便于用户查看相关产业领域的政策信息,及时把握区域的战略规划和指导方向。

(2) 数据资源协同是系统的基础,主要提供各类创新数据资源分类展示、详情查看、统计、关联等^[49]。

(3) 创新主体协同展示,主要对创新主体(企业、高校、科研院所等机构)、创新载体(实验室、园区、示范区、技术联盟等)、创新平台、成果合作、技术转移等进行可视化展示^[50]。

(4) 产业协同是系统的重要环节,主要对重点产业分布、产业资源、产业布局、科技成果、科技机构、科技人才等产业资源信息进行可视化

展示。

(5) 科技情报分析是系统的核心内容,提供产业技术监测与预警。

3.3.3 医药健康产业科技大数据的安全性设计

医药健康产业科技大数据资源展示了行业的发展现状与未来趋势,数据资源的完整性与安全性非常重要。本文提出以下几点系统与数据安全措施,以确保医药健康产业科技大数据的安全性。

(1) 对数据进行分类统一管理。按数据资源类型、学科分类、技术领域、研究价值等建立数据分类指标体系和数据标准规范,确定数据的公开范围和使用权限。

(2) 对数据资源实行审查、审核与校验机制。建立面向系统数据管理的提供者和管理者的审核与监管服务,保证医药健康产业科技大数据的完整性、规范性与正确性。

(3) 对平台的用户与数据权限进行设置。一是建立用户授权与实名认证,对用户的基本信息进行审核,针对用户类型与用户需求建立多级访问权限设置,实现不同类型用户的不同权限范围与功能;二是对科技数据资源的权限设置,针对不同应用场景提供不同的数据资源。

4 医药健康产业情报服务系统的应用研究

本文以医药健康领域科技大数据为基础,采集数据集成、创新地图展示、知识图谱等技术,实现产业科技大数据的智能检索、产业动态跟踪与智能推送、产业科技创新地图展示、前沿技术监测与预警分析等应用,促进医药健康产业链的

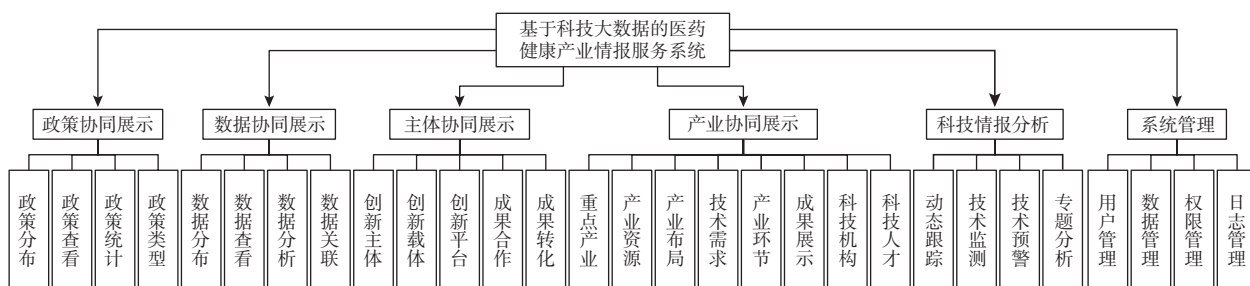


图4 基于科技大数据的医药健康产业情报服务平台功能设计

稳链、强链、补链、延链。

4.1 构建科技大数据中心服务引擎提供智能检索与数据可视化

为解决医药健康产业科技大数据分散、多源异构、应用价值低及数据获取难等问题，本文设计提出医药健康产业大数据智能搜索服务，解决不同用户快速、精准推送产业相关知识资源。以科技大数据为基础，进行数据整合、加工处理、数据关联，搭建产业科技大数据应用服务平台，为用户直观展示区域内的科技大数据资源分布动态、发现问题、挖掘资源和评价资源分布趋势提供基础数据参考，也为用户提供专业的、智能的“一站式”情报分析，形成产业链与数据链的结合，深挖科技大数据之间的关系，有助于医药健康全产业链数据集成和应用，有利于制定产业技术创新发展战略规划^[51]，有利于扩展科技大数据的管理、运营、交易等服务，较好地发挥产业科技创新发展与应用^[52]。针对产业大数据的现状，提供数字化赋能服务，提升优化产业布局的前瞻性，从根本上解决大数据集成为产业关键技术情报感知、预测提供协同服务，为进一步优化产业布局的合理性，为实施科技创新驱动发展战略、打造一流科技创新生态提供有力保障。

4.2 科技大数据知识发现提供医药健康产业动态跟踪与智能推送

利用科技大数据和机器学习数据挖掘技术，通过跟踪产业前沿动态，对风险进行分析管理，为产业创新发展提供产业态势监控、产业规划设计等应用服务，为政府合理配置创新资源和引导科技创新、产业创新提供服务支持^[53]。将基于科技大数据体量及高效的数据分析技术，紧跟前沿科技和产业发展趋势，提升科技成果就地转化效率，增强企业综合竞争实力。为产业创新发展提供产业发展态势、技术前沿、研究热点、规划设计等服务，为政府合理配置科技创新资源、引导产业科技创新、产业发展提供服务支持，推动产业链、创新链和资金链融合。医药健康产业科技创新发展和决策支撑中引用多源异构科技大数据进行情报分析和创新成果展示，提供科技大数

据治理机制下的产业决策咨询服务。只有科技大数据得到有效应用才能产生巨大的价值，才能推进我国医药健康产业技术创新和科技发展，成为推动医药健康产业科学研究、产业发展、科技创新的重要节点^[54]。跟踪各领域的前沿动态、研究报告、行业动态、产业信息、科研活动等，为用户提供内容浏览、专题定制和邮件自动推送等服务。为用户提供整合相关产业的人、财、物等信息，提供快速搜索相关产业和技术资源，便于用户查看前沿技术和研究热点。汇集并实时更新产业链中各环节企业、人才、技术等科技资源信息，了解产业科技主体和资源分布情况，洞察产业发展态势，实现产业发展的动态跟踪。

4.3 科技大数据可视化提供医药健康产业科技创新地图展示

面向政府决策、企业科技创新、科研活动，本文建立医药健康产业科技大数据互联互通，通过定义数据标签，建立各种类型数据的关联关系，可多维度全方位地了解产业投入、规模、科研产出等情况，以具有图形可视化视觉方式，建立时间、空间与地理位置的医药健康产业科技大数据地图展示，采用数据可视技术、GIS技术，将所有数据在线化、可视化、可交互，按时间、区域、产业、成果等多维度监测产业相关数据，全面立体透视产业关键技术发展，可视化展示产业的资源配置、产业链、企业布局。实现人才、项目、成果、机构、政策等多视角展示数据关联和快速定位，有效展示产业的相关数据分布、知识聚合，同时利用关联关系图、等级划分、数值区间颜色图等展示手段，为用户提供创新主体的发展趋势、创新成效，进一步揭示创新主体间的协同效率。

4.4 科技大数据情报分析提供医药健康产业前沿技术监测与预警分析

围绕医药健康产业情报服务需求，利用科技大数据集成管理对产业关键技术进行技术监测预警，为科学决策提供及时高效的预警支持。采用科技大数据、机器学习、知识图谱等先进技术，搭建产业链的关键技术知识图谱，使平台向便捷

化、智能化、规范化、协同化方向发展,形成面向产业关键技术提供产业动态、技术监测、科技成果转化和科技情报等服务的一体化平台。利用科技大数据集中管理达到对产业技术监测与预警,为科研管理的提供高效的决策支撑。一是产业技术监测。面向科技管理和服务需求,主要包括对科技动态、科技项目、科技人才的监测预警模块以及对科技前沿的专题分析、前沿分析、简报推送等功能模块。二是产业技术预警。对地区重点产业的技术状态进行监测,对产业技术进行横纵向对比分析和跨地区对比分析,能够满足指标查询、状态监测、区域对比和异常警示等需求。对可能产生威胁和侵害的情况提前感知、预判、预测和预警。在智能报告方面,按简报、专题报告等线上生成报告。在科研管理方面,构建前沿的产业关键技术情报态势感知和预警平台,能够追踪、跟踪产业发展趋势,形成科技决策、跨学科、跨领域的产业关键技术科技情报服务体系,围绕科技计划管理的全流程管理与服务,推动产业创新生态发展的科技管理效率,与科技计划管理对接,提供精准化、专业化、智能化的产业科技情报服务,突破基于关键的核心共性的的大数据、人工智能、知识图谱技术。在产业创新生态的技术监测和预警系统对所收集的数据进行模型训练和预测推理,从而在产业动态跟踪功能模块中实现对用户的精准推荐,并实现智能风险预测和预警。

5 结语

本文旨在解决医药健康产业的科技大数据信息不对称和精准服务带来的局限性问题。在全面分析当前医药健康产业的科技大数据元数据类型及结构的基础上,设计了医药健康产业科技大数据情报服务体系模型,并进行了精准知识服务功能设计,从需求、数据集成与关联、服务模式等层面详细阐述了精准情报服务的逻辑与方法,为应用系统开发提供理论支撑。大数据技术为知识经济带来了新的机遇与挑战,依托科技大数据资源支撑医药健康产业科技创新发展,有利于发挥

医药健康产业的科技大数据基础资源作用,更有能力和经验设计出符合应用场景的医药健康产业情报服务产品应用,利用专业的智能化的大数据处理技术和情报分析工具,开拓更广泛的产业科技创新新路径,形成新的机制。而本文设计的精准情报服务架构涉及面广,未来还需要与传统的产业情报方法进行对比分析,应用先进技术实现科技大数据融合应用与互联互通,并结合具体应用进一步实现与验证;针对复杂的应用场景和用户需求还不能有效解决;还有待进一步激发医药健康产业科技大数据的共享交流和知识创新服务。

参考文献

- [1] 国务院.关于印发促进大数据发展行动纲要的通知[EB/OL].[2022-04-25].http://www.gov.cn/zhengce/content/2015-09/05/content_10137.htm.
- [2] 国务院办公厅.国务院办公厅关于促进医药产业健康发展的指导意见[EB/OL].[2022-04-25].http://www.gov.cn/zhengce/content/2016-03/11/content_5052267.htm.
- [3] 国家发展改革委,中央网信办,工业和信息化部,等.关于印发《全国一体化大数据中心协同创新体系算力枢纽实施方案》的通知[EB/OL].[2022-04-25].http://www.gov.cn/zhengce/zhengceku/2021-05/26/content_5612405.htm.
- [4] 李辉,曾文,谭晓,等.科技大数据资源平台建设研究[J].科技情报研究,2022,4(1): 71-77.
- [5] WILLIAMS K, WU J, CHOUDHURY S R, et al. Scholarly big data information extraction and integration in the CiteSeer X digital library[C]//IEEE International Conference on Data Engineering Workshops. IEEE, 2017.
- [6] 曾文,车尧.科技大数据的情报分析技术研究[J].情报科学,2019,37(3): 93-96.
- [7] 刘明鹏,王忠明,马文君.基于科技大数据的我国林业知识服务体系研究设计[J].世界林业研究,2022,35(1): 94-99.
- [8] Harvard Dataverse[EB/OL].[2022-04-14].<https://dataverse.harvard.edu/>.
- [9] BAUMANN, PETER, MAZZETTI, et al. Big data analytics for Earth sciences: the Earth server approach[J]. International journal of digital Earth, 2016, 9(1/2/3): 3-29.
- [10] Data[EB/OL].[2022-04-14].<http://data.gov/>.

- [11] Inter-university Consortium for Political and Social Research. About the ICPSR[EB/OL].[2022-03-03].
<http://www.icpsr.umich.edu/web/pages/about/>.
- [12] Dryad Digital Repository[EB/OL].[2022-03-05].<http://datadryad.org/>.
- [13] UK Data Archive. The process[EB/OL].[2022-01-23].
<http://www.data-archive.ac.uk/curate/process>.
- [14] 王颖, 钱力, 谢靖, 等. 科技大数据知识图谱构建模型与方法研究[J]. 数据分析和知识发现, 2019, 3(1): 15-26.
- [15] 胡吉颖, 谢靖, 钱力, 等. 基于知识图谱的科技大数据知识发现平台建设[J]. 数据分析和知识发现, 2019, 3(1): 55-62.
- [16] 中国科技资源共享网[EB/OL].[2022-04-14].<https://escience.org.cn/>.
- [17] 中国科学院科学数据中心[EB/OL].[2022-04-14].
<http://www.casd.cn/>.
- [18] 北京大学开放研究数据平台[EB/OL].[2022-04-04].
<https://opendata.pku.edu.cn/>.
- [19] 复旦大学社会科学数据平台[EB/OL].[2022-04-04].
<https://dvn.fudan.edu.cn/>.
- [20] AYYAPPAN G, NALINI C, KUMARAVEL A. a novel k-nn classification approach using topic modelling in aminer dataset[J]. Indian journal of computer science and Engineering, 2019, 10(2):40-44.
- [21] 张智雄, 刘欢, 于改红. 构建基于科技文献知识的人工智能引擎[J]. 农业图书情报学报, 2021, 33(1): 17-31.
- [22] 赵瑞雪, 李娇, 张洁, 等. 多场景农业专业知识服务系统构建研究[J]. 农业图书情报学报, 2020, 32(1): 4-11.
- [23] 于倩倩, 钱力, 程冰, 等. 专题情报数据管理与智能分析平台的构建[J]. 图书情报工作, 2020, 64(24): 92-105.
- [24] 李堃怡, 莫锦华, 谢卫. 浅谈建设广西科技大数据平台的意义[J]. 大众科技, 2018, 20(7): 145-146, 165.
- [25] 翟运开, 李思琪, 孙东旭, 等. 面向“互联网+医疗健康”的医药电商平台构建[J]. 中国卫生事业管理, 2021, 38(3): 161-163, 202.
- [26] 李亚凡, 徐磊, 王莉, 等. 生物医药和大健康产业数据可视化应用[J]. 科技创新与应用, 2020(28): 167-168.
- [27] 孟彦霖, 王祖维, 叶力铭, 等. 基于医药健康的大数据可视化探究[J]. 电子商务, 2019(2): 30-31.
- [28] 刘倩. 大数据产业对区域科技创新发展的影响研究[J]. 科技管理研究, 2019, 39(2): 217-223.
- [29] 李梅, 张红, 孙艳艳, 等. 京津冀协同创新发展地图系统设计及应用[J]. 知识管理论坛, 2020, 5(3): 191-199.
- [30] 龚轶, 王峥, 高菲. 城市群协同创新系统: 内涵、框架与模式[J]. 改革与战略, 2019, 35(9): 61-70.
- [31] 孔贝贝, 谢靖, 钱力, 等. 科技大数据增值丰富化方法研究与工具研发[J]. 数据分析和知识发现, 2019, 3(7): 113-122.
- [32] 陈满, 向秦, 魏香梅, 等. 高新技术领域大数据中心平台的构建与应用[J]. 自动化与仪表, 2017, 32(5): 5-8, 13.
- [33] 冯立杰, 王亚星, 冯奕程, 等. 基于多维技术创新地图的技术预测研究[J]. 情报杂志, 2017, 36(8): 58-63, 26.
- [34] 刘召栋, 周亿城. 科技大数据资源及分类分级研究[J]. 科技与创新, 2021(18): 123-126.
- [35] 吴振新, 钱力, 谢靖, 等. 面向智慧知识服务的科技文献大数据体系建设[J]. 图书情报工作, 2020, 64(24): 63-72.
- [36] 柴苗岭, 邹弈星, 谭荣志, 等. 面向农业产业知识服务的科学数据与科技文献关联研究与实践[J]. 农业图书情报学报, 2022(3): 37-50.
- [37] 苏晓娟, 张英杰, 白晨, 等. 科技大数据背景下的中英双语语料库的构建及其特点研究[J]. 中国科技资源导刊, 2019, 51(6): 87-92.
- [38] 钱力, 刘细文, 张智雄, 等. AI+智慧知识服务生态体系研究设计与应用实践: 以中国科学院文献情报中心智慧服务平台建设为例[J]. 图书情报工作, 2021, 65(15): 78-90.
- [39] 周园春, 王卫军, 乔子越, 等. 科技大数据知识图谱构建方法及应用研究综述[J]. 中国科学: 信息科学, 2020, 50(7): 957-987.
- [40] 彭茂祥, 韩子丹. 基于专利数据的国防科技大数据构建及管理应用探讨[J]. 科学管理研究, 2016, 34(4): 29-32.
- [41] 赵华, 赵瑞雪, 金慧敏, 等. 农业科技大数据仓储建设与服务[J]. 数字图书馆论坛, 2020(8): 48-55.
- [42] 于阳. 科技资源信息相关集成方法研究[J]. 江苏科技信息, 2020, 37(11): 25-28.
- [43] 赵剑冬. 基于区域科技大数据平台的科技服务模式研究[J]. 产业与科技论坛, 2019, 18(14): 61-62.
- [44] 孙惠娟, 陈红梅, 沈杰, 等. 以科技创新地图推动高新技术企业培育加速发展的研究与思考: 以苏州市吴江区高新技术企业为例[J]. 江苏科技信息, 2019, 36(23): 5-9.
- [45] 陈伟, 杨锐, 何涛, 等. 大数据环境下科技情报研究的新模式[J]. 科技导报, 2018, 36(16): 78-85.

(下转第110页)

- <https://www.chinanews.com.cn/gj/2019/08-29/8940893.shtml>.
- [14] 韩国产业通商资源部. 24시간누구나반도체집 설계가 능! [EB/OL]. (2020-06-29)[2023-01-14]. http://www.motie.go.kr/motie/ne/motienewse/Motienews/bbs/bbsView.do?bbs_seq_n=155117359&bbs_cd_n=2¤tPage=781&search_key_n=title_v&cate_n=&dept_v=&search_val_v=.
- [15] 韩国产业通商资源部. 세계 최초·최고기술향확보 하 라 [EB/OL]. (2020-09-10)[2023-01-14]. http://www.motie.go.kr/motie/ne/motienewse/Motienews/bbs/bbsView.do?bbs_seq_n=155117300&bbs_cd_n=2¤tPage=841&search_key_n=title_v&cate_n=&dept_v=&search_val_v=.
- [16] 韩国产业通商资源部. 한국산업통상자원부. 소부장 2.0 전 략 [EB/OL]. (2020-07-09)[2023-01-14]. http://www.motie.go.kr/motie/ne/presse/press2/bbs/bbsView.do?bbs_cd_n=81&bbs_seq_n=163128.
- [17] 韩国科技信息通信部. 제2차위성정보활용종합계획 (2019—2023) [EB/OL]. (2022-07-17)[2023-01-14]. <https://www.msit.go.kr/bbs/view.do?sCode=user&mId=84&mPid=83&pageIndex=5&bbsSeqNo=65&nttSeqNo=2085457&searchOpt=ALL&searchTxt=>.
- [18] 韩国产业通商资源部. 차세대 지능형반도체 ‘드림팀’ 출범 [EB/OL]. (2020-09-10)[2023-01-14]. http://www.motie.go.kr/motie/n/motienewse/Motienews/bbs/bbsView.do?bbs_seq_n=155117446&bbs_cd_n=2¤tPage=691&search_key_n=title_v&cate_n=&dept_v=&search_val_v=.
- [19] 财联社. 韩国: 上半年韩国材料零件设备进口对日本依存度降至15.4%的纪录新低 [EB/OL]. (2022-07-20)[2023-01-14]. <https://baijiahao.baidu.com/s?id=1738832288891587301&wfr=spider&for=pc>.
- [20] 韩国产业通商资源部. 일본의존도90% 반도체소재 연 내 양 산! [EB/OL]. (2020-09-10)[2023-01-14]. http://www.motie.go.kr/motie/ne/motienewse/Motienews/bbs/bbsView.do?bbs_seq_n=155117253&bbs_cd_n=2¤tPage=891&search_key_n=title_v&cate_n=&dept_v=&search_val_v=.

(上接第77页)

- [46] 王磊, 谭清美. 智能生产与服务网络条件下产业创新平台战略地图研究 [J]. 科技进步与对策, 2017, 34(1): 53-58.
- [47] 李亚莉. 云南化工产业大脑建设之思考 [J]. 云南化工, 2020, 47(8): 184-186.
- [48] 钱力, 谢靖, 常志军, 等. 基于科技大数据的智能知识服务体系研究设计 [J]. 数据分析与知识发现, 2019, 3(1): 4-14.
- [49] 陈宫晴, 勾鑫峰, 郭婷, 等. 基于科技资源共享的创新地图探究与思考 [J]. 中国新通信, 2019, 21(21): 155-156.
- [50] 冯立杰, 曹健, 王金凤, 等. 基于FBS和多维技术创新地图的技术创新机会识别方法及其应用 [J]. 情报理论与实践, 2020, 43(12): 89-95.
- [51] 李亚丹, 任艳君, 刘洪麟, 等. 科技情报机构建设基于事实型数据库的科技创新智库初探 [J]. 技术与市场, 2020, 27(7): 42-43, 46.
- [52] 张勇, 苏学, 谢振峰. 面向科技大数据的元数据仓储建设实践探索 [J]. 情报工程, 2020, 6(6): 84-96.
- [53] 徐源, 胡正银, 宋亦兵, 等. 面向学科知识服务的微信小程序研究与实践 [J]. 图书情报工作, 2020, 64(14): 54-62.
- [54] 王俊, 王修来, 庞威, 等. 面向科技前瞻预测的大数据治理研究 [J]. 计算机科学, 2021, 48(9): 36-42.