

·前沿与热点·

人工智能 2.0 时代的知识分析变革研究*

田倩飞^{1,2} 张志强¹

(1.中国科学院成都文献情报中心 四川成都 610041)

(2.中国科学院大学 北京 100049)

摘要:“人工智能”一词于 1956 年提出后经历几番起落,近年来,随着云计算、大数据等技术快速发展,人工智能的研究和应用得以新生,人工智能 2.0 时代已然到来。文章首先梳理国际政产学研界针对新一轮人工智能研发的战略布局、相关成果与应用,概述我国《新一代人工智能发展规划》及其部署进展,引出人工智能 2.0 的定义、技术特征及其颠覆性影响,然后,基于文献调研与项目跟踪,在研读全球各大智库人工智能研究报告、国内外相关研究论文、梳理学者研究原型与美国智能情报项目的基础上,从信息知识采集、信息知识搜索、信息处理与知识挖掘、知识预见预警以及知识决策服务等诸多环节阐述人工智能关键技术知识分析中的应用与影响,提出人工智能 2.0 时代知识分析的总体趋势将朝向全面、精准、自动、高效、智能和深度等方向发展。

关键词:人工智能 2.0 知识分析 DARPA IARPA

中图分类号:G250.2;TP18

文献标识码:A

DOI:10.11968/tsyqb.1003-6938.2018023

Research on Knowledge Analysis Transformation in the Era of Artificial Intelligence 2.0

Abstract Artificial intelligence is to make machines smarter and competent to accomplish the work that only human beings could do. After the introduction of "Artificial Intelligence" in 1956, it has undergone ups and downs. Recently, with the development of cloud computing, big data, etc., the research and application of artificial intelligence is entering the 2.0 Era. This paper firstly introduces the strategic plan, research achievements and application of industry, academia and government on the new wave of artificial intelligence research and development, China's Development Plan on New Generation of Artificial Intelligence, as well as the definition of Artificial Intelligence 2.0, its tech features and disruption. Based on literature research and projects tracking, this paper then analyses the key technologies of artificial intelligence and the application in knowledge analysis. It is concluded that information analysis in the artificial intelligence 2.0 Era would bound toward the direction of comprehensive, targeted, automated, effective, intelligent and profound. Finally, this paper concludes the development and effect of artificial intelligence, and gives suggestions on AI-based knowledge analysis and scientific decision making.

Key words artificial intelligence2.0; knowledge analysis; DARPA; IARPA

1 人工智能进入 2.0 时代

近年来,伴随着移动互联网、云计算、大数据等信息前沿技术的快速发展,计算机硬件性能和计算技术的突破以及机器学习算法的优化,诞生六十余年的人工智能研发再次迎来了新的机遇,相关产业规模不断增长,企业数量大幅增加。2000 至 2016 年

间,美国累积新增人工智能企业 3033 家,占全球累积总数的 37.41%;同期中国人工智能企业数累积增长 1477 家,占全球总数的 12.91%^[1]。2016 年 10 月 26 日,在北京举行的瑞银创新科技行业 CEO 高峰论坛上,瑞银发布《未来值得关注的九项颠覆性技术趋势》报告预测,认为到 2020 年,全球人工智能市场规模有望达到 1200 亿到 1800 亿美元之巨^[2]。以谷歌围

* 本文系国家社会科学重点基金项目“面向领域知识发现的学科信息学理论与应用研究”(项目编号:17ATQ008)与中科院成都文献情报中心主任基金项目“科研投入视角下的学科领域创新前沿和热点挖掘方法研究”(项目编号:Y7Z0581001)研究成果之一。
收稿日期:2018-03-11;责任编辑:魏志鹏

棋系统 AlphaGo^[3]及新一版 AlphaGo Zero^[4]、IBM 认知技术平台 Watson、微软聊天机器人“小冰”等为代表的产品成果标志着人工智能从学术课题研究全面步入产业经济爆发阶段。2016 年,谷歌首席执行官桑达尔·皮查伊宣布谷歌战略从“移动先行”(Mobile First) 转向“人工智能先行”(AI First)^[5]; 同年, Facebook 创始人兼首席执行官扎克伯格在 F8 开发者大会上,将人工智能明确列为 Facebook 未来十年发展规划战略中的重要领域,具体包括自然语言、视觉识别、逻辑规划等^[6]。除学术界、产业界聚焦人工智能研发和应用外,各国政界也开始关注和支持人工智能研发,出台发展规划和政策措施,力图在新一轮国际人工智能科技竞争中掌握主导权。2016 年 10 月,美国政府发布的《国家人工智能研发战略规划》^[7],制定出美国人工智能研发的整体框架以及七项优先战略,以期充分利用人工智能技术来增强国家经济实力并改善社会安全。同年 12 月,英国发布的《人工智能:未来决策制定的机遇与影响》^[8],阐述了人工智能的未来发展对英国社会和政府的影响,论述了如何利用英国独特的人工智能优势来增强英国国力。2017 年 4 月,法国制定的《国家人工智能战略》^[9],旨在谋划法国未来人工智能的发展,使法国成为欧洲人工智能的领军者。

在人工智能发展上,我国也不甘落后,积极动员和部署。2017 年 7 月,国务院印发的《新一代人工智能发展规划》^[10],提出了面向 2030 年我国新一代人工智能发展的指导思想、战略目标、重点任务和保障措施,部署构筑我国人工智能发展的先发优势,加快建设创新型国家和世界科技强国。其中,三步走战略目标为:到 2020 年人工智能总体技术和应用与世界先进水平同步,人工智能核心产业规模超过 1500 亿元,带动相关产业规模超过 1 万亿元;到 2025 年人工智能基础理论实现重大突破,部分技术与应用达到世界领先水平,人工智能核心产业规模超过 4000 亿元,带动相关产业规模超过 5 万亿元;到 2030 年人工智能理论、技术与应用总体达到世界领先水平,成为世界主要人工智能创新中心,人工智能核心产业规模超过 1 万亿元,带动相关产业规模超过 10 万亿元。

2017 年 11 月,科技部召开新一代人工智能发展规划暨重大科技项目启动会,宣布成立新一代人工智能发展规划推进办公室,宣布成立新一代人工智能战略咨询委员会(潘云鹤院士任组长)。潘云鹤(2016)针对人工智能 2.0 给出初步定义,即:基于重大变化的信息新环境和发展新目标的新一代人工智能。他指出人工智能 2.0 的技术特征表现在:一是从传统知识表达技术到大数据驱动知识学习,转向大数据驱动和知识指导相结合的方式,其中机器学习不但可自动,还可解释,更广泛;二是从分类型处理多媒体数据(如视觉、听觉、文字等),迈向跨媒体认知、学习和推理的新水平;三是从追求智能机器到高水平的人机协同融合,走向混合型增强智能的新计算形态;四是从聚焦研究个体智能到基于互联网络的群体智能,形成在网上激发组织群体智能的技术与平台;五是将研究的理念从机器人转向更加广阔的智能自主系统,从而促进改造各种机械、装备和产品,走上智能化之路^[11]。

人工智能的深入发展和深度应用,对各行各业的影响都是颠覆性的,最为突出的影响包括:开启互联网新纪元、推动产业转型与变革、打造未来国防军事的“杀手锏”、掀起科研创新模式新革命、颠覆生活方式等^[12]。在人工智能 2.0 时代到来之际,总结人工智能关键技术对知识分析的影响,预测其未来变革,深思应对措施,提升基于人工智能的知识发现能力和水平,是值得关注的研究热点之一。本文在调研和分析国际智库有关人工智能的研究报告、梳理国内外相关研究与项目部署的基础上,探讨人工智能 2.0 时代知识分析各环节的变化,最后提出启示建议。

2 人工智能 2.0 时代:无限算力与大数据背景下的机器感知学习时代

2.1 人工智能关键技术

自 1956 年达特茅斯会议诞生“人工智能”一词以来,距今已六十余年。美国麻省理工学院人工智能实验室原主任(1972 至 1997 年)帕特里克·温斯顿(Patrick Winston)提出“人工智能就是研究如何使计算机完成曾经只有人类才能完成的工作” 斯坦福大学人工智能研究中心名誉教授尼尔斯·尼尔森(Nils

Nilsson)(1982年)将人工智能定义为“关于知识的学科——怎样表示知识、怎样获得知识、怎样使用知识的科学”^[13]。北京大学信息管理系王延飞(2016年)提出,人工智能是试图挖掘人类智能的实质,从而对人类智能进行模拟和扩展的一门新兴技术科学^[14]。

埃森哲2016年9月发布了《人工智能:经济发展新动力》报告,其指出:人工智能由多种技术组成,能以不同方式组合起来,教会机器感知、理解、行动和学习(埃森哲将新兴人工智能技术及其功能和解决方案示例见表1)^[15]。

表1 新兴人工智能技术及其功能

新兴人工智能技术	功能	解决方案示例
计算机视觉 音频处理	感知	虚拟代理、身份分析、认知机器人、语音分析、推荐系统、数据可视化
自然语言处理 知识表示	理解	
机器学习 专家系统	行动	

阿里云研究中心、阿里巴巴创投公司及波士顿咨询公司于2016年10月合作发布的研究报告《人工智能:未来制胜之道》指出:人工智能是一系列技术的集合,其下包括机器学习、推理、感知和自然语言处理等。随着人工智能被更广阔地应用到人类社会和经济生活各个方面,新的机遇和挑战随之而生。其巨大的潜在影响让人类不得不谨慎思考人工智能技术的发展与应用^[16]。

信息科技咨询公司高德纳(Gartner)于2017年7月发布了《2017年新兴技术成熟度曲线图》,强调了“无处不在的人工智能”这一技术趋势。由于计算能力快速提升、数据量不断增加、深度神经网络向前推进,人工智能技术将在未来十年展现出强劲的颠覆力。值得重点关注的人工智能重点技术包括:深度学习、深度增强学习、强人工智能、自动驾驶汽车、认知技术、商用无人机、对话用户接口、企业分类与本体管理、机器学习、智能尘埃、智能机器人和智能工作空间^[17]。

2.2 人工智能与知识分析

科学研究的范式已进入数据密集型科学范式的大数据时代,正在推动和牵引科技情报研究工作进入知识分析与知识发现服务的新时代^[18]。情报研究范式不断演进,从事实型情报搜集、综述型情报分析、计算型情报研究发展到数据驱动型知识发现。由于

机器拥有远超人类的强大计算能力,它们能对海量数据进行收集和分类,进而进行专业化计算与分析,以做出更好的决策。人工智能就是利用机器为人类完成繁杂的工作,具体包括:寻找模式、预测趋势和发现相关性,学习并改善,执行相关计划,基于历史趋势预测未来结果,基于事实给出决策建议等。机器学习是人工智能的技术之一,能在面对新数据时自动调整计算机程序,得出客观分析的新信息和见解^[19]。

大数据情报与知识分析需要人工智能方法技术。国际数据公司(IDC)的《数据时代2025》白皮书指出,认知/人工智能系统将改变格局。大量涌现的数据催生出了一系列全新的技术,如机器学习、自然语言处理和人工智能(统称为认知系统),它们将数据分析从不常见的、追溯式的实践转变成为战略决策和行动的前摄式推动因素。认知系统可以大大提高各种行业、环境和应用数据分析的频率、灵活性和即时性。IDC预计,到2025年,用于数据分析的全球数据总量将增长至原来的50倍,达到5.2ZB,而认知系统“触及”的分析数据总量将增长至原来的100倍,达到1.4ZB^[20]。2017年7月28日,Gartner发布了《2017年分析方法与商业智能成熟度曲线图》,其指出分析方法和商业智能领域将从可视化数据发现时代转变为增强分析时代,即运用机器学习、自然语言接口等技术,缩短数据准备时间、自动发现数据模式,将分析结果分享给更多用户^[21]。

人工智能方法技术推动情报与知识分析跨越新阶段。王志宏等^[22]指出,人工智能、认知计算、决策支持技术的概念及研究已经跨出实验室,进入真正的实用阶段,其现实价值超过了技术研究者最初的认识,其本质是对信息、数据进行加工生产的技术,而这种生产加工不但可以给人们提供更加便捷的生活,还将对未来人们的生产、生活方式产生根本性的影响,对未来国家经济生活的运行方式、未来社会的组织结构产生深远的影响,郭璇等^[23]探讨了在反恐情报信息工作中,利用当前人工智能领域的研究热点——深度学习技术开展反恐情报挖掘和分类,有效减少情报人员的工作量,提高反恐情报信息工作效率,Zhuang等^[24]指出,针对大数据的下一代人工智能系统将是可解释的、强健的、通用人工智能,它不

只是暴力地完成浅显计算,而是能执行深度神经推理;它能基于结构化逻辑规则发挥数据驱动模型的作用;它能从已有经验中学习;Morris 等^[25]认为,机器学习、大数据、图像处理单元(GPU)等技术的迅猛发展与融合,引发新一轮的人工智能研发与应用热潮;Dhar^[26]指出,此轮人工智能热潮与以往的根本性不同在于机器学习已经具备感知和处理能力。从前,机器无法阅读、听到或看到信息,只能获取和处理经组织后的信息。现有技术进展使得机器能直接从外界获取输入信息,而无需人类参与,继而创建机器内部表示用于进一步处理。

综上,人工智能关键技术及其对知识分析、信息服务、决策支撑等的变革作用已引起国内外学者的研究和重视。本文将结合学者研究成果、企业研发动态及美国相关政府部门的项目部署,具体探讨人工智能对知识分析各环节的变革影响。

3 人工智能对知识分析各环节的变革影响

人类社会的主流,已从利用地表资源的农业社会历史、挖掘地下资源的工业社会现实、正迈向开发数据与智力的智能社会未来^[27]。随着互联网、高性能计算、云计算、大数据、人工智能等信息技术的持续发展与应用,针对海量数据的检索存储与情报知识的分析利用均已发生翻天覆地的变化。从传统的本地数据库到数据中心再到云端,从目录检索到关键词

再到词语标注,从自然语言搜索到语义搜索再到推理,技术推动着人类情报分析的能力发展与效率提升。人工智能 2.0 时代知识分析的总体趋势将朝向全面、精准、自动、高效、智能和深度发展。人工智能技术和系统将对信息知识采集、信息知识搜索、信息处理与知识挖掘、知识预见预警以及决策服务等诸多环节,产生深刻影响。在本文的分析中,主要遴选了美国安全和情报部门——美国国防高级研究计划局(DARPA)以及美国情报高级研究计划局(IARPA)的案例,两者通过开展各种高风险、高回报的情报研究项目,以期大幅提高美国的情报分析能力。作为国际上引领性的情报信息分析机构,其项目案例可显现人工智能技术与信息知识分析应用的最新发展。

3.1 信息知识采集

在信息知识采集方面,智能化采集技术能实现信息资源(特别是网络资源)的自动搜索、甄别、过滤、监测、跟踪,使多种数据源情境下的数据采集能力大大提高。美国 DARPA 的洞见(Insight)项目和中国工程科技知识中心(CKCEST)项目均强调多源数据的收集、索引和存储,并在此基础上,构建集成系统,让计算机高效地完成从非结构化数据到知识这一过程,支撑协作与决策(这两项典型研发项目的具体信息见表 2)。

3.2 信息知识搜索

针对已采集的情报成果,结合元数据、本体、语

表 2 多源数据采集的典型研发项目案例

项目名称	主要内容	技术特征	主研机构或人员
DARPA-洞见(Insight) ^[28]	针对情报、监视和侦查(ISR),打造一款可调整的集成系统,增强情报分析人员对战场紧迫性操作的支撑。系统能收集、索引和存储来自多种源头的的数据,然后分析并关联这些信息,在分析师和数据提供商之间建立请求和共享链接。自动化后台处理能力包括行为学习和预测算法等,帮助分析师发现和确定潜在威胁。项目成果包括跨情报系统运作的集成系统,针对低层级分析任务的优化自动化操作,使分析师能更好地聚焦于其他复杂信息收集与分析;针对动态变化任务与威胁的传感诊断能力;开放、基于标准、即插即用的架构,支撑现有 ISR 技术的快速集成;工具与自动化功能,提升分析效用和效率;多用户界面,以强化理解并支撑协作和快速决策。	多源数据处理、数据驱动的机器学习、系统集成、知识图谱、知识可视化	Steve Jameson
中国工程院与浙江大学-KS Studio知识计算引擎的研发计划 ^[29]	2012年,中国工程院启动建设“中国工程科技知识中心(CKCEST)”项目,旨在打通和汇聚各类工程科技数据资源,通过技术分析处理形成知识库,并开发各种应用提供知识服务,推动国家工程科技战略思想库的建设,服务于国家的战略决策。如何将无序繁杂的文本、图像、视频等原始的非结构化数据加工转化为有序、可用、标准的结构化知识,是知识中心建设的核心问题。这个问题的解决,需要数据汇聚、知识加工、图谱构建、数据可视化等诸多关键技术的支撑。2015年,作为中国工程科技知识中心的关键技术研发中心,浙江大学提出 KS-Studio(Knowledge Service-Studio)知识计算引擎的研发计划,旨在综合上述关键技术,探索有效解决这一问题的途径。通过技术手段,让计算机高效地完成从非结构化数据到知识这一过程 ^[29] 。	数据汇聚、知识加工、图谱构建、数据可视化	中国工程院、浙江大学

义网等方式对其进行描述与组织,并开发新的搜索方法,提升搜索效率与质量,为情报用户提供满足需求的情报资源。

Martín A 等^[30]研究了利用智能技术增强数字图书馆的语义互操作性。文章提出一种语义与智能搜索引擎概念架构,利用本体和人工智能查询在线知识库的元数据,利用智能代理(Agent)实现更高效的知识获取方法。文章还引用欧盟互操作框架来介绍其关注重点之一——语义互操作分析(见图1)。

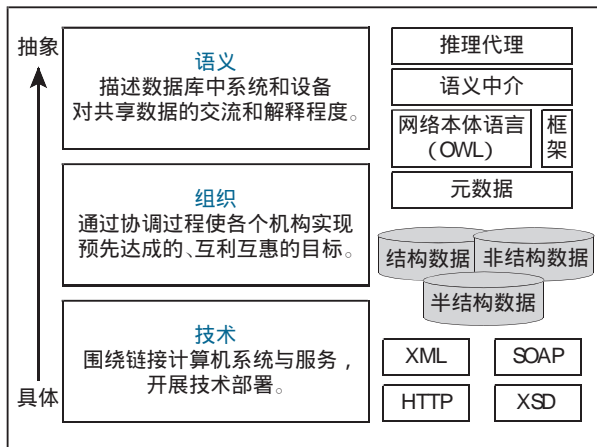


图1 欧盟互操作框架-语义与智能搜索引擎概念架构(Martín A, 2015)

美国艾伦人工智能研究所(Allen Institute for Artificial Intelligence)于2015年11月推出基于人工智能的语义搜索引擎 Semantic Scholar,通过更加深入地理解学术论文的内容和背景,对学术论文进行排序^[31]。最初推出时 Semantic Scholar 的搜索范围仅为计算机科学领域的300万篇论文,后期通过与艾伦脑科学研究所(Allen Institute for Brain Science)合作,该网站新增了数百万篇神经学和医学领域论文,并针对该领域量身定制了新的筛选器,用户可以根据各种内容(大脑组成部分或细胞类型、模式生物、研究方法等)进行搜索。

德国马普信息学研究所正开发一款名为 DeepLife、专注健康和生命科学领域的搜索引擎。Microsoft 也于2016年5月公开发布其人工智能学术搜索工具 Microsoft Academic。该工具由 Microsoft 网络搜索引擎 Bing 提供语义搜索功能支持,覆盖约1.6亿份文献^[32]。

除上述针对出版文献的信息搜索外,因情报数据源的扩展,美国 DARPA 还开展了针对多源信息筛选的引擎项目 AIDA 等, IARPA 则专门设置针对语音和视频搜索的情报能力提升项目(见表3)。

表3 美国 DARPA 与 IARPA 的搜索引擎项目简表

项目名称	主要内容	技术特征	研发机构
DARPA- 不同来源主动诠释 (AIDA) ^[33]	该项目将开展模糊性多源信息流的重要数据筛选研究,开发一种多假设语义引擎,根据从各种来源获得的数据,产生对现实世界事件、形势和趋势的显性化释义;建立自动将多媒体来源信息聚合和映射成共性表述或故事情节的技术能力,并生成和探索与关系利益事件、形势和趋势的确切性质和影响的多种假设,解决当今数据环境下的数据繁杂、矛盾和潜在的欺骗问题,确定每个信息的置信水平,以及由语义引擎生成的每个假设,理解并掌握其原始形式的信息或数据,通过调整或转移变量和概率来产生替代上下文,以提高准确性并解决歧义问题。	语义引擎、自动映射、置信水平、消歧技术	2017年4月发布项目指南,获资机构尚未公布。
DARPA- 扩展存储器 (Memex) ^[34]	该项目启动于2014年,不同于谷歌,必应这样的商业搜索引擎,它专为政府组织设计使用,可针对特定领域进行复杂的信息图表搜索并对大量信息进行关联分析形成关系网络。Memex的设计有两个重要的特点,第一是能够从极小的像素中推断出图形的含义。第二是能够搜索图像和视频的元数据。元数据非常重要,能够描述图片和视频,从而发现它们之间的联系。	元数据、信息抽取、内容发现、信息关联	卡内基·梅隆大学,MIT林肯实验室,纽约大学,斯坦福大学,美国航空航天局;SRI 国际公司等。
IARPA- 巴别塔 (Babel) ^[35]	该项目启动于2011年,旨在针对多语种,开发灵活、强健的语音识别技术和搜索能力,使分析师能有效地处理真实世界录制的海量语音。	多语言语音识别、关键词搜索算法、噪音环境语音识别、机器学习	卡内基·梅隆大学,IBM Watson 研究中心,雷神技术公司;加州大学伯克利国际计算机科学研究所。
IARPA- 阿拉丁 (Aladdin) ^[36]	Aladdin 视频项目启动于2010年,旨在以全新的方式集成先进的视频提取、音频提取、知识表达和搜索技术等,开发一种快速、准确、强健和可扩展的技术来支持未来的多媒体分析需求。	计算机视觉、自然语言处理、图像处理、大数据、视频分析、机器学习、语音处理	卡内基·梅隆大学;查尔斯·斯塔克·德拉普尔实验室,IBM Watson 研究中心,Kitware 公司;雷神技术公司,Sarnoff 公司,SRI 国际公司等。

3.3 信息处理与知识挖掘

在信息处理与知识挖掘方面,数据挖掘、文本挖掘、知识发现、智能计算、专家系统等智能化分析处理技术,可以大幅提升海量文档调查的工作效率,帮助情报知识人员分析发展趋势,并从中发现一些表达隐晦的缄默信息、未知的事实和潜在的情报。DARPA 与 IARPA 在自然语言处理、机器学习和信息理解等智能信息技术项目研发上呈现加力和发力现象^[14](详细项目信息见表 4。图 2 至图 4 还分别展示了 DARPA 大机制(Big Mechanism)、数据驱动的模式发现(D3M)以及文本深度探索及过滤(DEFTE)项目的架构、模型与数据流等信息)。

3.4 知识预见预警

数据挖掘、智能算法等新一代技术在知识预见预警方面的成功案例已不少见。如北美在线影片租赁提供商 Netflix 在并未看到《纸牌屋》电视剧一个画面的情况下,出价 1 亿多美元获得这部剧集的首播权。它通过推荐引擎、数据算法等方式,提前获知观

众们喜欢看的内容,从而进行准确的内容订购授权。路透社开发一款“路透新闻追踪”软件,通过监视 Twitter 并寻找具有群体影响力的新报道,然后将相关的推文汇总为事件,最终生成与事件相关的信息和元数据。该软件可以识别哪些事件具有新闻价值、从相关观点结论中过滤出事实,并验证报道的真实性。该软件比其他媒体抢先 8 分钟报道 2016 年初的布鲁塞尔爆炸案事件,抢先 15 分钟报道 2016 年 10 月纽约发生的切尔西爆炸案(美国 IARPA 部署的两项科学与技术预测项目见表 5)。

3.5 知识决策服务

随着数据库技术和搜索引擎技术的成熟而逐渐“下移”为一种社会能力,传统信息和资料提供的职能已在科技和其他领域的决策中被边缘化。在大数据与人工智能蓬勃发展的新时代,决策者对情报机构的要求已远远不能满足于信息分析与知识服务,情报机构唯有提供更加智能化的决策方案,不仅要能回答个体“是什么”,而且还必须快速地告诉决策

表 4 美国 DARPA 与 IARPA 信息处理与知识挖掘典型项目

项目名称	主要内容	技术特征	研发机构
IARPA- 隐喻 (Metaphor) ^[37]	该项目启动于 2011 年,以概念隐喻理论为基础,研究英语、波斯语、俄语和西班牙语中的隐喻使用,以期洞悉语言中所隐含的文化观念,并开发隐喻语言自动分析软件。	自然语言处理、认知科学、语义技术。	卡内基·梅隆大学;伊利诺伊理工大学;雷神技术公司;纽约州立大学研究基础中心;加州大学伯克利分校。
IARPA- 深度联合视频分析 (DIVA) ^[38]	DIVA 项目于 2016 年发布申请指南,旨在针对多摄像头流视频环境,开发强健的自动化活动探测技术,用于取证及实时警告等。	机器学习、多层建模、深度学习、物体/人体检测、人类行为识别、视频活动检测、视频三维重构。	尚未公布
DARPA- 探索数据 (XDATA) ^[39]	该项目启动于 2012 年,XDATA 是创造和利用开源技术与架构、寻求在大数据处理和计算的计算技术与软件工具方面的创新性研究。旨在开发计算技术和软件工具,用于分析大量数据,包括半结构化数据和非结构化数据。	分布式数据存储、人机交互、视觉推理、可视化技术。	佐治亚理工大学,Continuum Analytics 公司等。
DARPA- 大机制 (Big Mechanism) ^[40]	该项目启动于 2014 年,旨在开发自动搜索技术阅读研究论文与摘要,通过先进大数据分析来发现因果关系模型,如研究导致细胞癌变的复杂分子之间的相互作用。	基于统计和知识的自然语言处理、管理与本体、表达与推理;大规模随机网络的可视化、模拟等。	卡内基·梅隆大学;哈佛医学院;英国合成与系统生物中心等。
DARPA- 数据驱动的模式发现 (D3M) ^[41]	该项目于 2016 年启动,为没有数据科学专业知识背景的用户研发机器学习建模方法,从而弥补设计人员在数据科学方面的专业素养不足。如果成功,使用 D3M 工具的研究人员将成为一支“虚拟数据科学家”部队,从而改进军事规划、后勤和情报工作。	自动建模、原型库、复杂模型自动化合成、人与模型的交互。	密歇根大学;查尔斯·里弗分析公司等。
DARPA- 文本深度探索及过滤 (DEFTE) ^[42]	该项目于 2012 年中期启动,目的是提升自然语言的处理能力。自然语言处理技术具备人工智能属性,并且已有一定实用经验,可以帮助国防分析人员大幅提升海量文档调查的工作效率,并从中发现一些表达隐晦、可能付诸行动的信息。DARPA 希望这种技术能够帮助情报分析人员扫描更多的文本文档和语音文件,并从中发现一些关键信息,如人物、事件、地点、时间等。	自然语言、深度学习、计算语言、对话分析。	伊利诺伊大学厄巴纳-香槟分校;华盛顿大学;卡内基·梅隆大学;斯坦福大学;康奈尔大学等。

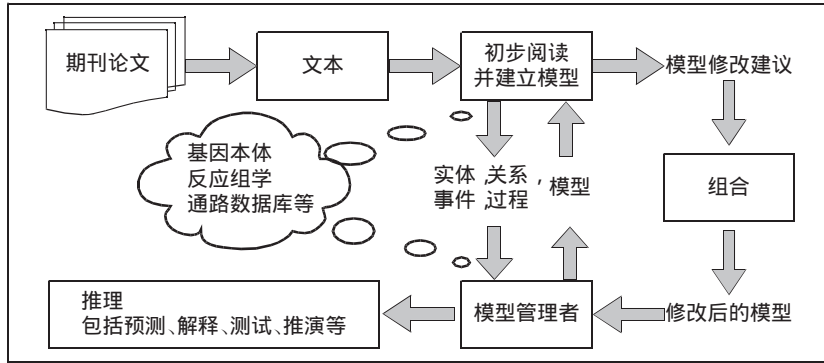


图2 美国 DARPA Big Mechanism 架构图^[43]

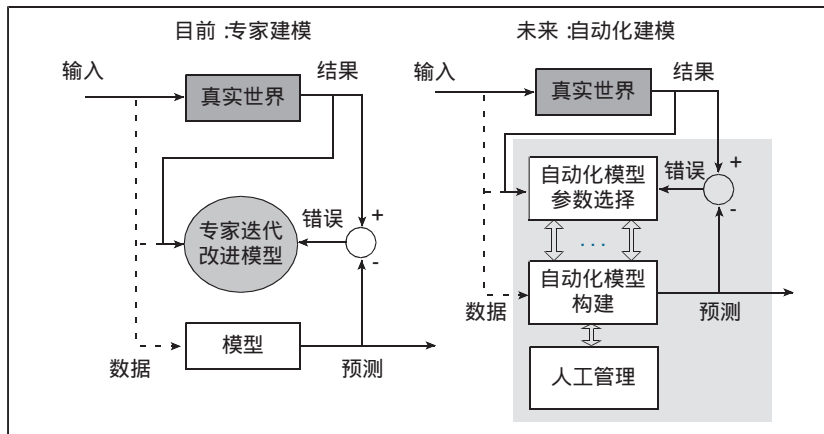


图3 美国 DARPA D3M 模型自动化改进示意图^[44]

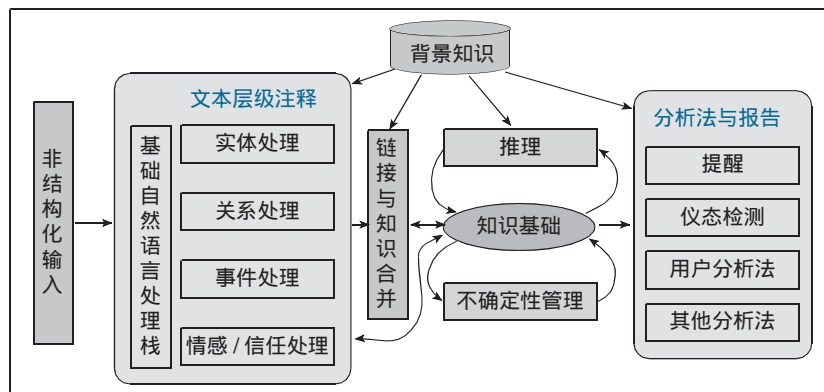


图4 DARPA DEFT 项目数据流^[45]

表5 IARPA 科技预测项目部署

项目名称	主要内容	技术特征	研发机构
IARPA- 预测科学与技术 (ForeST) ^[46]	该项目启动于 2013 年,旨在开发和测试方法,通过组合专家意见,精确地预测重大的科技里程碑事件。相关技术创新包括:(1)利用组合预测市场方法,获得有效的启发与汇集意见;(2)基于科学与专利文献指标,生成科技预测问题;(3)确定众包问题,并通过全球 1 万多名科学家与工程家确定解决方案。	预测技术、机器学习、文本分析、大数据、自然语言处理	BAE 系统公司;乔治梅森大学,SRI 国际公司
IARPA- 科研论述中的预见与理解 (FUSE) ^[47]	该项目旨在开发自动化方法,利用公开出版的科学、技术与专利文献中的信息,帮助完成针对技术出现的系统性、持续性和复杂性评估。该项目的假设是技术出现会在公开出版的科学与专利文献中留下可寻的痕迹,由此开发一个系统:(1)处理大量的、跨学科的、多语言的科学与专利文献;(2)在新兴技术领域自动生成技术术语并确定优先度。	文本分析、知识发现、大数据、社会网络分析、自然语言处理、机器学习、预测技术	哥伦比亚大学;雷神 BBN 技术公司,SRI 国际公司

者整体和整体中的任意部分“是什么”“为什么”准确地表明“怎么办”,方能发挥情报服务应有的效能。整个情报服务业正在信息技术推动下发生巨大的变化,正在迎来一个崭新的时代^[48]。

Tweedale 等^[49]指出人工智能与计算智能技术被成功用于决策制定,同时能改善信息可获性,解决数据密集问题。决策支持系统(DSS)融合了一系列技术,旨在将决策者知识与来自特定信息源的相关数据组合起来,应用数学与统计方法和模型,最终改善个体或团体的决策能力。智能决策支持系统(IDSS)利用计算机科学和新兴领域的先进智能方法,扩展 DSS 能力来解决复杂问题(IDSS 及其组件见图 5)。

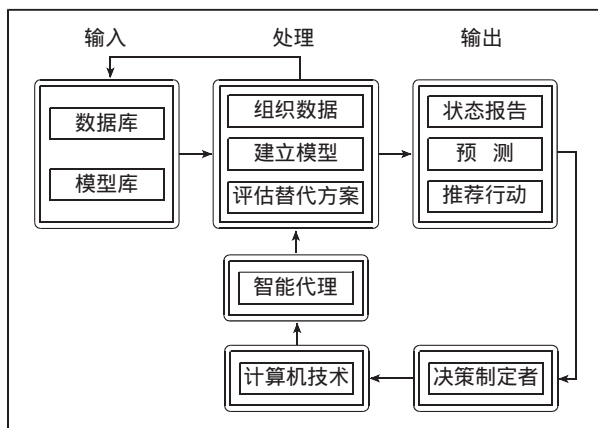


图 5 IDSS 及其组件

IARPA 指出,现有情报分析与汇报工具难以帮助用户评估竞争性假设、提供清晰的/有支撑的论点、发现或克服偏见。一些结构化技术(如论点制图、竞争假设分析、贝叶斯推理网络等)已被用于改善推理。但这些技术软件较难被使用,既费力又耗时,且所需输入的知识常常超越了个体用户的特有知识领域。IARPA 于 2016 年发起“众包证据、议论、思维和评估”(CREATE)项目,旨在开发技术工具来帮助情报分析人员更好地评估数据,或检测出可能产生曲解的假设,辅助相关决策(项目信息见表 6)。

表 6 IARPA CREATE 项目助力情报推理与决策

项目名称	主要内容	技术特征	研发机构
众包证据、议论、思维和评估 (CREATE) ^[50]	旨在开发利用众包和结构化分析方法的系统并加以实验测试,以改善分析推理能力,辅助决策。	社会与行为科学;形式推理;计算机科学;结构化分析技术;众包	乔治梅森大学;莫纳什大学;墨尔本大学;雪城大学

4 建议

(1)人工智能发展迅速并产生颠覆性影响。近年来,随着移动互联网、大数据和云计算的快速发展,诞生 60 多年的人工智能再获新生并取得重大突破性进展。从学术界的课题研究到产业界的成果与经济突破,再到各国政府在人工智能领域紧锣密鼓的战略布局,均彰显人工智能是未来发展的重大前沿领域,有着广阔的应用前景并将对各行各业的发展产生颠覆性影响。瑞银《未来值得关注的九项颠覆性技术趋势》报告指出,人工智能技术发展迅速,正处于弱人工智能向通用人工智能发展的关键阶段,未来农业、零售、制造业、金融、交通等众多行业将受到人工智能技术的巨大影响,并因之改变。预计,未来全球因人工智能而受到影响的工作岗位将多达 5000 万至 7500 万^[51],但人工智能的应用并不会在全球范围内导致大规模失业,反而会因此创造出更多的工作岗位。

人工智能的发展,最根本的推动技术包括不断增加的无限算力和不断膨胀的大数据技术;人工智能发展的核心是人类知识组织和再生系统的重大变革,即由人类自身的知识创造,延伸到机器感知学习的机器化知识再生系统。毫无疑问,人工智能的发展将深刻改变一切依赖人类知识运行的系统规则,也将深刻改变人类自身的发展。

(2)加强基于人工智能的知识分析与科学决策。在大数据时代,知识分析密切依赖于大数据分析 with 智能挖掘技术,特别是一些以科学大数据为鲜明特征的重要科技领域(如生命科学、空间科学、天体物理、地球科学等)科学大数据呈爆发式增长。借助人工智能技术,改造提升知识分析环节和知识分析预测,能极大地改善知识分析效率和更好服务于科学决策。开发基于人工智能的知识分析流程和平台,是开展知识分析、知识发现的必由之路。

5 结语

本文综合分析了国际上代表性机构 DARPA/IARPA 应用人工智能技术提升知识分析的案例。人工智能技术能从广度、深度、精度等多方面提高知识分析工作的水平,知识分析将越来越依赖于先进的

智能技术。建议重点建设领域知识分析数据库并促进开放共享；开发知识分析智能化工具并打造系统平台；推进智能化知识分析应用示范。通过大力支持、加强人工智能技术在知识分析系统中的应用研发,提升知识分析效率和知识挖掘深度,提高知识预见预警能力,最终辅助人类知识决策。

参考文献:

- [1] 乌镇智库.全球人工智能发展报告产业篇[EB/OL].[2018-01-10].<http://h5.iwuzhen.org/pdf/AI-Report20170803.pdf>.
- [2] 侯云龙.人工智能市场规模将达1800亿美元[EB/OL].[2018-01-10].<http://finance.people.com.cn/h1/2016/1027/c1004-28811231.html>.
- [3] Silver D,Huang A,Maddison C J,et al.Mastering the game of Go with deep neural networks and tree search[J].Nature,2016,529(7587):484.
- [4] David S,Julian S,Karen S,et al.Mastering the game of Go without human knowledge[J].Nature,2017,550(7676):354-359.
- [5] 沈怡然.谷歌AI中国中心建立AI First战略亚洲落地[EB/OL].[2018-01-10].<http://www.eeo.com.cn/2017/1214/318814.shtml>.
- [6] 纪振宇.透视三巨头开发者大会:人工智能成战略重点[EB/OL].[2018-01-10].<https://tutorials.hostucan.cn/microsoft-facebook-google-intelligence>.
- [7] 田倩飞.编译.美国发布人工智能领域国家级研究报告与战略规划[EB/OL].[2017-08-29].http://www.casid.cn/zkcg/ydkb/kjqykb/2016/201612/201707/t20170703_4821935.html.
- [8] 钟新龙,韩健.美英人工智能报告带来的启示和思考[EB/OL].[2017-09-10].<http://www.ccidgroup.com/sdgc/9569.htm>.
- [9] 陈晓怡.编译.法国发布人工智能战略[EB/OL].[2017-08-25].http://www.casid.cn/zkcg/ydkb/kjzcyzkb/2017/201705/201707/t20170703_4821779.html.
- [10] 国新网.《新一代人工智能发展规划》政策解读[EB/OL].[2017-07-30].<http://www.scio.gov.cn/34473/34515/Document/1559231/1559231.htm>.
- [11] Pan Y.Heading toward Artificial Intelligence 2.0[J].Engineering,2016,2(4):409-413.
- [12] 中国科学院颠覆性技术创新研究组.颠覆性技术创新研究——信息科技领域[M].北京:科学出版社,2018.
- [13] Nilsson N J.Artificial intelligence:engineering,science or slogan?[M].Readings from the AI magazine.American Association for Artificial Intelligence,1989:2-9.
- [14] 王延飞,赵柯然,何芳.重视智能技术凝练情报智慧——情报、智能、智慧关系辨析[J].情报理论与实践,2016,39(2):1-4.
- [15] 埃森哲.人工智能:经济发展新动力[EB/OL].[2017-09-01].<https://www.accenture.com/cn-zh/insight-ai-artificial-intelligence-future-growth>.
- [16] 云栖智库.人工智能:未来制胜之道[EB/OL].[2017-08-20].http://image-src.bcg.com/Images/BCG_Artificial-Intelligence_CN_HN_Oct2016_tcm55-156025.pdf.
- [17] 田倩飞.编译.Gartner发布《2017年新兴技术成熟度曲线图》[EB/OL].[2017-09-20].http://www.das.cas.cn/xwzx/2016/kxw2016/xxjst/2016/201709/t20170918_4861886.html.
- [18] 张志强.论科技情报研究新范式[J].情报学报,2012,31(8):788-797.
- [19] SAP Analytics Cloud.How Artificial Intelligence Can Solve Industry Challenges?[EB/OL].[2017-09-20].<https://www.sapanalytics.cloud/how-ai-solves-industry-challenges/>.
- [20] IDC.Data Age 2025: The Evolution of Data to Life-Critical [EB/OL].[2017-09-21].<http://www.seagate.com/www-content/our-story/trends/files/Seagate-WP-DataAge2025-March-2017.pdf>.
- [21] Gartner.Hype Cycle for Analytics and Business Intelligence 2017 [EB/OL].[2017-09-25].<https://www.gartner.com/document/3772080>.
- [22] 王志宏,杨震.人工智能技术研究及未来智能化信息服务体系的思考[J].电信科学,2017,33(5):1-11.
- [23] 郭璇,吴文辉,肖治庭,等.基于深度学习和公开来源信息的反恐情报挖掘[J].情报理论与实践,2017,40(9):135-139.
- [24] Zhuang Y T,Wu F,Chen C,et al.Challenges and opportunities from big data to knowledge in AI 2.0 [J].Frontiers of Informa-

- tion Technology & Electronic Engineering ,2017 ,18(1):3- 14.
- [25] Morris K C ,Schlenoff C ,Srinivasan V.Guest Editorial A Remarkable Resurgence of Artificial Intelligence and Its Impact on Automation and Autonomy[J].IEEE Transactions on Automation Science & Engineering ,2017 ,14(2):407- 409.
- [26] Dhar V.The future of artificial intelligence[J].Big Data ,DOI :10.1089/big.2016.29004.vda.
- [27] 王飞跃.知识产生方式和科技决策支撑的重大变革——面向大数据和开源信息的科技态势解析与决策服务[J].中国科学院院刊 ,2012 27(05):527- 537.
- [28] DARPA.Insight[EB/OL].[2017- 10- 13].<https://www.darpa.mil/program/insight>.
- [29] 中国人工智能学会通讯.KS- Studio:一个知识计算引擎[EB/OL].[2017- 10- 12].<https://yq.aliyun.com/articles/216744?spm=5176.100239.blogcont216742.14.JWfETU>.
- [30] Martí n A ,Leó n C ,Ló pez A.Enhancing semantic interoperability in digital library by applying intelligent techniques[C].Sai Intelligent Systems Conference.IEEE ,2015:904- 911.
- [31] Luther J.Discovery in an age of artificial intelligence[J].Learned Publishing ,2016 ,29(2):75- 76.
- [32] Nicola Jones.AI science search engines expand their reach[EB/OL].[2017- 09- 01].<http://www.nature.com/news/ai-science-search-engines-expand-their-reach-1.20964>.
- [33] DARPA.AIDA[EB/OL].[2017- 10- 01].<https://www.darpa.mil/program/active-interpretation-of-disparate-alternatives>.
- [34] DARPA.MEMEX[EB/OL].[2017- 10- 01].<https://www.darpa.mil/program/memex>.
- [35] IARPA.BABEL[EB/OL].[2017- 09- 10].<https://www.iarpa.gov/index.php/research-programs/babel>.
- [36] IARPA.ALADDIN[EB/OL].[2017- 09- 10].<https://www.iarpa.gov/index.php/research-programs/aladdin-video>.
- [37] IARPA.METAPHOR[EB/OL].[2017- 09- 18].<https://www.iarpa.gov/index.php/research-programs/metaphor/baa>.
- [38] IARPA.DIVA[EB/OL].[2017- 09- 18].<https://www.iarpa.gov/index.php/research-programs/diva>.
- [39] DARPA.XDATA[EB/OL].[2017- 10- 01].<https://www.darpa.mil/program/xdata>.
- [40] DARPA.BIG MECHANISM[EB/OL].[2017- 09- 25].<https://www.darpa.mil/program/big-mechanism>.
- [41] DARPA.Data- Driven Discovery of Models (D3M)[EB/OL].[2017- 09- 25].<https://www.darpa.mil/program/data-driven-discovery-of-models>.
- [42] DARPA.Deep Exploration and Filtering of Text(DEFT)[EB/OL].[2017- 09- 25].<https://www.darpa.mil/program/deep-exploration-and-filtering-of-text>.
- [43] William C.Regji ,Brian Pierce.DARPA and Data :A Portfolio Overview[EB/OL].[2018- 02- 05].<https://www.nitrd.gov/nitrd-groups/images/3/31/DARPA-and-DATA.pdf>.
- [44] Wade Shen ,Brian Sandberg.DARPA Overview and Data Driven Discovery of Models (D3M)[EB/OL].[2017- 09- 25].<http://www.amstat.org/asa/files/pdfs/POL-JSM-DARPA.pdf>.
- [45] Boyan Onyshkevych.KB Representation of Text ,Audio ,Images ,and Video[EB/OL].[2017- 10- 01].<http://www.akbc.ws/2014/slides/onyshkevych-nips-akbc.pdf>.
- [46] IARPA.FOREST[EB/OL].[2017- 10- 01].<https://www.iarpa.gov/index.php/research-programs/forest>.
- [47] IARPA.FUSE[EB/OL].[2017- 10- 01].<https://www.iarpa.gov/index.php/research-programs/fuse>.
- [48] 吴晨生 ,李辉 ,付宏,等.情报服务迈向 3.0 时代[J].情报理论与实践 ,2015 ,38(9):1- 7.
- [49] Tweedale J W ,Phillips- Wren G ,Jain L C.Advances in Intelligent Decision- Making Technology Support[M].Intelligent Decision Technology Support in Practice.Springer International Publishing ,2016:3- 28.
- [50] IARPA.CREATE[EB/OL].[2017- 09- 10].<http://www.iarpa.gov/index.php/research-programs/create>.
- [51] 人工智能市场规模将达 1800 亿美元[N/OL].[2018- 01- 21].http://dzjjckb.cn/www/pages/webpage2009/html/2016- 10/27/content_24758.htm.

作者简介 田倩飞,女,中国科学院成都文献情报中心助理研究员,中国科学院大学情报学博士研究生,研究方向:情报研究方法与技术、信息科技战略情报研究;张志强,男,中国科学院成都文献情报中心研究员,博士生导师,研究方向:情报研究方法与技术、学科信息学与领域知识发现、科学学与科学评估。