

基于 CiteSpace 的人工智能文献大数据可视化分析^①

任利强, 郭强, 王海鹏, 张立民

(海军航空大学 信息融合技术研究所, 烟台 264001)
通讯作者: 任利强, E-mail: 1194153993@qq.com

摘要: 随着数据量的上涨、计算机运算力的提升和深度学习算法的出现, 人工智能受到越来越多的关注. 文中以美国核心期刊数据库 Web of Science 收录的与人工智能相关的 6879 篇期刊论文为研究对象, 以时空知识图谱及内容知识图谱分析为主要研究方法, 基于信息可视化软件 CiteSpace 从合作国家、研究机构、引用文献、关键词和突现词五个方面对文献大数据进行可视化比较和分析, 明确了人工智能领域的研究现状及重要文献, 揭示了人工智能领域的研究热点和前沿. 最后, 通过对五个可视化分析方面的总结, 给出了在人工智能领域中选择科研方向、探测学科前沿、辅助科技决策等方面的重要参考.

关键词: 人工智能; 可视化分析; 文献大数据; CiteSpace; 学科前沿

引用格式: 任利强, 郭强, 王海鹏, 张立民. 基于 CiteSpace 的人工智能文献大数据可视化分析. 计算机系统应用, 2018, 27(6): 18-26. <http://www.c-s-a.org.cn/1003-3254/6390.html>

CiteSpace-Based Visualization Analysis of Literature Big Data on Artificial Intelligence

REN Li-Qiang, GUO Qiang, WANG Hai-Peng, ZHANG Li-Min

(Institute of Information Fusion, Naval Aviation University, Yantai 264001, China)

Abstract: With the increase of data volume, the improvement of computer computing power, and the emergence of deep learning algorithm, artificial intelligence has been paid more and more attention. Taking American Core Journal Database Web of Science as research object, in which 6879 journal papers about artificial intelligence were included, adopting knowledge map of time and space and content knowledge map as main research methods, and applying the information visualization software CiteSpace, visualize and analyze literature big data from five aspects: cooperative countries, research institutions, references, keywords, and burst terms. It clarifies the research status and important literature in the field of artificial intelligence, and reveals the research hotspots and frontiers in the field of artificial intelligence. Finally, through the summary of the five visual analyses, the article gives an important reference in the field of artificial intelligence to select the direction of scientific research, to explore the forefront of science, to help science and technology decision, and so on.

Key words: artificial intelligence; visual analysis; literature big data; CiteSpace; subject frontier

1 引言

近年来, 从人机围棋大战到无人驾驶汽车, 从智能游戏、智能检索到智能机器人, 从语音识别、图像识别到机器翻译和自然语言理解等. 人工智能技术对社

会各领域产生了广泛而深远的影响, IBM、谷歌、Facebook、微软、Amazon、百度、阿里巴巴等各国高科技企业普遍将人工智能视为下一代产业革命和互联网革命的技术引爆点, 并且投入大量资金以加快其产业

^① 基金项目: 国家自然科学基金重大研究计划 (91538201); 泰山学者工程专项 (Ts201511020)

收稿时间: 2017-10-10; 修改时间: 2017-11-03; 采用时间: 2017-11-08; csa 在线出版时间: 2018-05-28

化进程^[1]. 人工智能相关产业与技术也受到了各国政府的高度关注与积极投入. 美国白宫科技政策办公室 2016 年 10 月发布了题为《为人工智能的未来做好准备》和《国家人工智能研究与发展战略规划》^[2]的重要报告, 前者概述了人工智能的发展现状以及应用领域, 后者则提出了美国应该将人工智能提升到优先发展的战略方向; 英国政府 2016 年 12 月发布了题为《人工智能: 未来决策的机遇与影响》^[3]的报告, 以关注人工智能对社会创新与生产力的促进作用, 论述如何将英国人工智能的独特优势转化为增强英国国力的优势; 2017 年 3 月我国首次将人工智能写进 2017 年《政府工作报告》^[4]中, 2017 年 8 月国务院下发的《新一代人工智能发展规划》通知中甚至宣布举全国之力抢占人工智能全球制高点. 这些足以表明人工智能已经被多国提升到国家战略层面.

基于人工智能技术被各国广泛应用, 成为近年来舆论焦点和炙手可热的科技热搜词. 为此, 本文对有关人工智能 2007–2016 年的文献梳理、分析和比较, 全面了解和追踪近年来人工智能研究过程中方法与技术的更新. 以美国核心期刊数据库 Web of Science 收录的与人工智能相关的 6879 篇期刊论文为依托, 利用 CiteSpace 软件对人工智能领域合作国家、研究机构、引用文献、关键词和突现词五个方面进行可视化分析, 全面了解人工智能领域的研究进展. 文献大数据可视化分析结果可以为人工智能的深入研究、实践探索和产业推进提供重要参考.

2 CiteSpace 及大数据可视化

2.1 大数据可视化

人类的创造性不仅取决于逻辑思维, 还与形象思维密切相关. 利用形象思维获取视觉符号中所蕴含的信息有助于人类发现和洞察规律, 进而获得科学发现. 期间, 可视化技术对重大科学发现起到了重要作用^[5]. 数据可视化于二十世纪五十年代出现, 指运用计算机图形学和图像处理技术, 以图表、地图、标签云、动画或任何使内容更容易理解的图形方式来呈现数据^[6], 使数据所表达的内容更容易被理解.

随着大数据时代的到来, 人类社会产生的数据呈爆炸式增长, 数据的规模越来越大^[7]. 这些冰冷枯燥的海量数据使人望而生畏. 在这种背景下大数据可视化技术作为一种可以有效地简化与提炼数据流, 将海量

复杂的数据直观可视化呈现的工具逐步发展起来. 知识图谱便是其中之一, 它是以科学知识为研究对象, 描述科学知识的发展进程与结构关系的一种图形^[8]. 科学知识图谱涉及数理统计学、计算机科学、社会学、信息科学、图像学等多学科的理论, 并且与科学计量学的共词分析、共引分析等方法结合, 通过对科学知识的挖掘和处理, 绘制一系列可视化的图形, 将学科知识发展进程和结构关系直观形象地展示出来^[9]. 这种将数据通过可视化技术变成直观图形的方法不仅让冰冷枯燥的数据变得亲切和易于理解, 更是激发了人的形象思维与想象力, 从而为科学新发现创造新的手段和条件^[10].

2.2 CiteSpace

在科学探索中, 学者最关注的莫过于从自己所研究领域海量文献中发现自己最感兴趣的主题和科学文献, 对自己研究领域发展历程有清晰的认识, 以及发现该领域最为活跃的研究热点、前沿和未来发展趋势. 由美国费城德雷克塞尔大学教授陈超美开发的 CiteSpace 软件对上述一系列问题提供了解决方法. CiteSpace 是一种基于引文分析理论的科技文本挖掘及可视化分析软件^[11]. 该软件可以用来绘制科学领域发展的知识图谱, 直观地展现科学知识领域的信息全景、宏观结构以及一个学科或知识领域在一定时期发展的趋势与动向, 探测并发现若干研究前沿领域的演进历程及其重要的奠基性理论文献^[12]. 同时, CiteSpace 所独特具有的时区视图和突变词检测功能可以发现科学领域研究的热点及前沿. 陈超美教授在 2004 年发表的《Searching for intellectual turning points: Progressive knowledge domain visualization》^[13]文章中首次介绍了 CiteSpace 软件. 2005 年陈悦、刘则渊发表的《悄然兴起的科学知识图谱》^[14]文章在国内首次应用了 CiteSpace 可视化软件. 2008 年 CiteSpace 实现了对中文数据的可视化分析, 加速了该软件在国内应用的传播进程. 2016 年李杰、陈超美出版的《CiteSpace: 科技文本挖掘及可视化》专著对 CiteSpace 的设计理念与操作流程做了详细的介绍, 标志 CiteSpace 已从初级应用到高级应用. 如今 CiteSpace 软件已进入“5.1”版本, 其功能已日渐完善.

利用 CiteSpace 构建知识图谱步骤如图 1 所示. CiteSpace 在国内文献中的应用主要表现在三个方面: 一是作为科学知识图谱的引证资料; 二是作为其他可

可视化软件比较对象;三是作为科技文献可视化分析的工具.本文应用 CiteSpace 软件的第三种形式,通过该软件对国际人工智能领域的科技文献大数据进行分析,从而得到分时动态的可视化图谱.对图谱的进一步分析以识别并发现该领域科学发展的现状、研究热点、研究前沿及发展趋势^[15].

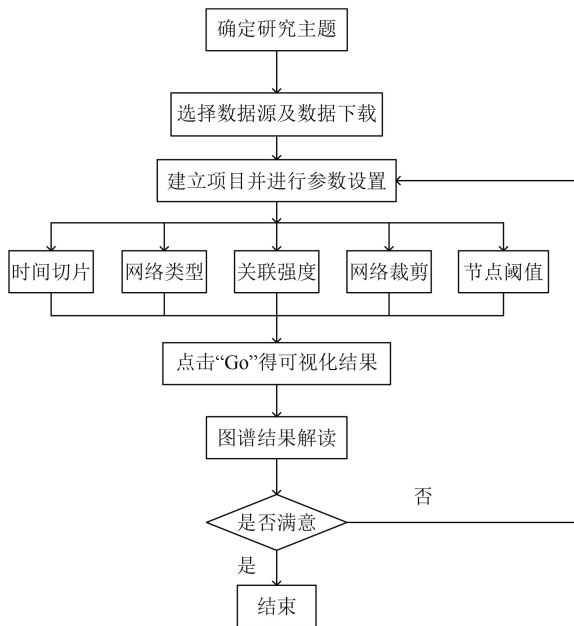


图1 CiteSpace 操作流程

3 基于 CiteSpace 的人工智能文献大数据可视化分析框架

基于 CiteSpace 的人工智能文献大数据可视化分析框架如图 2 所示.首先,在了解人工智能文献大数据可视化分析需求的基础上,提出了掌握人工智能研究领域的发展现状、重要文献并探究该领域的研究热点和前沿的研究目标;其次,根据研究目标选择大数据可视化工具 CiteSpace 作为合适的研究工具对数据进行分析处理;再次,根据研究目标选取 Web of Science 数据库中与人工智能有关的文献数据作为研究内容;接着,通过建立时间分布图谱、国家合作图谱、机构合作图谱以及地理可视化网络实现时空知识图谱分析,通过建立文献共被引图谱、关键词聚类图谱和突现词检测图谱实现内容知识图谱分析,从时空知识图谱和内容知识图谱分析两方面对人工智能研究现状、热点以及前沿进行分析;最后,根据图谱分析的结果给出结论.

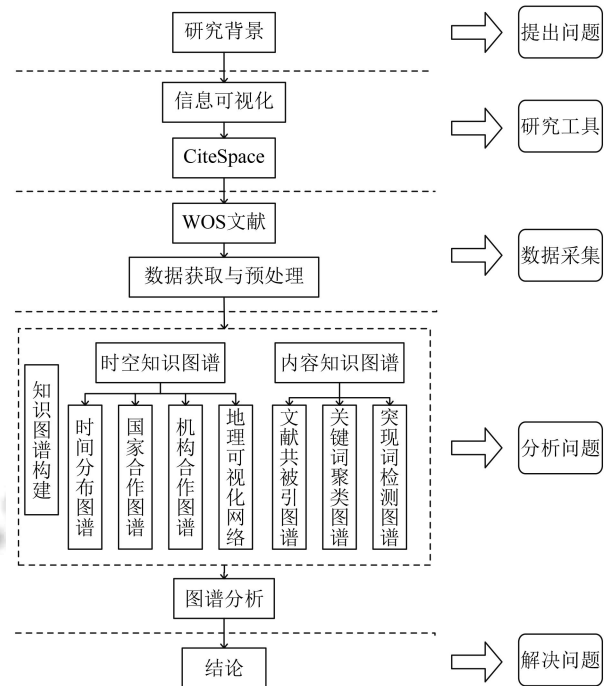


图2 基于 CiteSpace 的人工智能文献大数据可视化分析框架图

4 人工智能文献大数据可视化数据来源及数据预处理

科技文本数据的采集是 CiteSpace 可视化分析的基础.对于科技文本数据而言,索引型数据库收录了除正文外所有文献的信息.美国 Web of Science 数据库的数据结构较为完整,包括文献类型、作者、期刊、关键词、摘要、机构和参考文献.基于 CiteSpace 分析的数据便是以 Web of Science 下载的数据格式为基础.

4.1 数据采集

研究数据来源于美国 Web of Science 中的 Web of Science TM 核心数据库,检索方式为主题=“artificial intelligence” or “Artificial Intelligence”,时间跨度=2007-2016,文献类型=ARTICLE,索引=SCI-EXPANDED,检索时间为 2017 年 8 月 3 日,查询到的 7000 篇文献数据库界面如图 3 所示.剔除“书评”、“会议摘要”、“书信”以及其他“新闻报道”,最后得到有效记录 6879 条,这些文献 2007-2016 年期间被引用频次总计 70 519 次,每项平均引用次数为 10.25, h-index^[16] 为 90,表示有 90 篇论文至少被引用 90 次.

4.2 CiteSpace 软件的数据预处理

在数据分析之前需要对 Web of Science 格式的数据进行数据过滤与除重.数据过滤可以提取不同学科

领域的的数据,由于人工智能涉及多个学科领域,数据过滤操作可以略过.数据除重可以去除冗余的数据,CiteSpace 软件的数据预处理模块可以完成此步骤.如图 4 所示,进入数据预处理界面,首先选择标签“WoS”,然后加载原始数据到“Input Directory”,将保存处理后数据文件夹加载到“Output Directory”,最后点击“Remove Duplicates (WoS)”完成除重过程.



图 3 文献检索结果界面

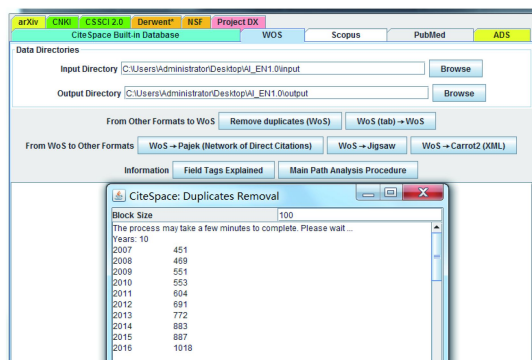


图 4 CiteSpace 软件数据除重界面

5 基于 CiteSpace 的人工智能文献大数据可视化参数设置及结果分析

5.1 参数设置

软件相关选项参数的设置如图 5 所示,时间跨度 (Time Span) 设置为 2007-2016,选择每 1 年为一个时间切片 (Time Slice),从而得到 10 个时间分段.网络节点关联强度选择 Cosine 算法,设定时间片阈值为 50,即数据抽取对象为前 50 (Top 50).最后选择最小生成树精简算法 (Minimum Spanning Tree)^[17]对网络修正裁剪,并且根据研究需要选择相应的节点类型,绘制相应的科学知识图谱.

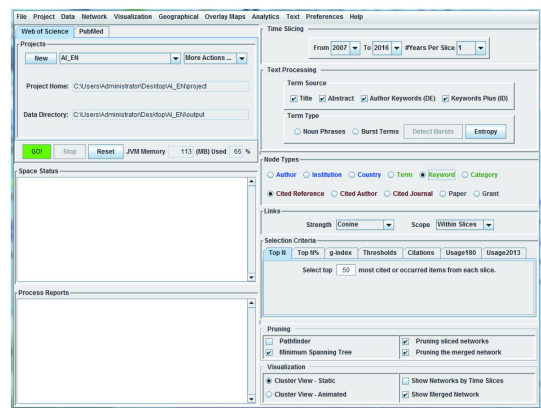


图 5 CiteSpace 软件相关选项参数设置界面

5.2 时空知识图谱及其分析

通过建立时间分布图谱、国家合作图谱、机构合作图谱以及地理可视化网络实现时空知识图谱分析,从时空知识图谱对人工智能的发展现状进行不同层面的分析与探究.

5.2.1 人工智能研究时间分布图谱

某一领域的发展状况可以从该领域发文量历时性变化趋势中得到反映^[18].为了解人工智能近十年的研究成果,统计了 2007-2016 年人工智能领域各年发表的文献量如图 6 所示.图 6 显示,自 2007 年以来围绕人工智能的研究文献量呈现持续增长的趋势,尤其是 2016 年恰逢“人工智能 60 周年”,围绕人工智能的研究文献量出现急增.人工智能正受到广泛的关注,迎来发展热潮.明显看出,全球人工智能发展持续火热,研究正进入高速发展阶段.



图 6 人工智能各年发文量统计图

5.2.2 国家空间分布图谱

为了解世界各国对人工智能领域的研究现状,对开展人工智能相关研究的国家空间特征进行分析.选择网络节点为“Country”,运行 CiteSpace 软件,得到人工智能研究国家合作图谱,如图 7 所示.在图 7 中圆

圈的大小代表文献作者所在国家出现的频次, 圆圈越大, 表明该国作者出现的频次越多, 反之越少. 圆圈的颜色代表相应国家作者发表文献时间, 单圈年轮环越宽表示该单位时间段内该国作者出现的频次越多^[19]. 可以明显看出在人工智能研究领域, 美国 (USA) 的辐射面积最大, 以频次 1168 居于世界首位, 具有绝对的优势. 这说明美国在人工智能领域具有较强的研究实力和影响力. 中国 (PEOPLES R CHINA) 以 784 频次位居第 2, 具有一定的研究实力; 西班牙 (SPAIN) 频次 581 位列第 3; 英国 (ENGLAND) 频次 526 位列第 4; 伊朗 (IRAN) 频次 497 位列第 5; 第 6 和第 9 分别是印度 (INDIA) 和土耳其 (TURKEY)、意大利 (ITALY) 和法国 (FRANCE). 由此可见, 欧洲和亚洲各国都高度重视人工智能技术的发展, 积极开展人工智能领域相关研究.

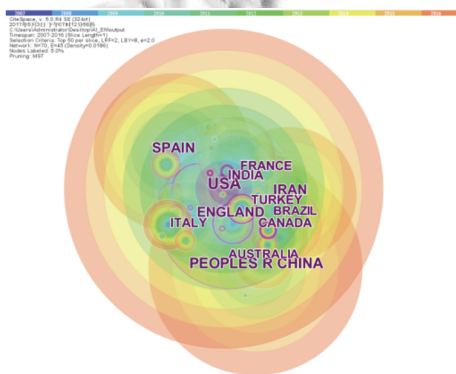


图 7 人工智能研究国家合作图谱

5.2.3 研究机构空间分布图谱

构建人工智能的研究机构分布图谱时, 设定网络节点类型为“Institution”, 生成人工智能研究机构合作图谱, 如图 8 所示. 从图 8 可以发现, 伊斯兰自由大学 (Islamic Azad Univ) 频次最高, 在人工智能领域发表论文最多, 具有绝对高的研究实力. 其次高产的机构依次为中国科学院 (Chinese Acad Sci)、德黑兰大学 (Univ Tehran)、台湾科学技术大学 (Taiwan Univ Sci & Technol)、马来亚大学 (Univ Malaya)、南洋理工大学 (Nanyang Technol Univ), 表明这四个机构在人工智能方向具有较强的研究潜力. 通过对这些机构的所在地进行分析, 发现发文量排在前十的研究机构都位于亚洲, 其中包括 4 所伊朗名校, 3 所中国研究机构, 仅从发文量上可见, 近年来人工智能在亚洲受到广泛关注与

积极研究. 同时表明, 伊朗和中国对人工智能的研究相当重视且具有一定的基础实力.

为了研究不同机构间的合作情况, 如图 8 所示. 节点为研究机构名称, 节点之间连线描述机构的合作, 连线的颜色表示节点首次合作的时间, 连线的粗细表示机构合作次数, 线条越细表示合作次数越少^[20]. 机构合作网络中共有节点 429 个, 连线 137 条, 整体网络密度仅为 0.0015, 这说明国际人工智能研究团队分散, 不同机构的作者之间存在较少合作, 极具凝聚力的科研群体尚未能形成. 未来需要国内外高校及相关研究团队之间加强合作交流, 打破消极界限, 创建跨机构、跨学科的合作队伍, 以促进人工智能研究的快速发展.

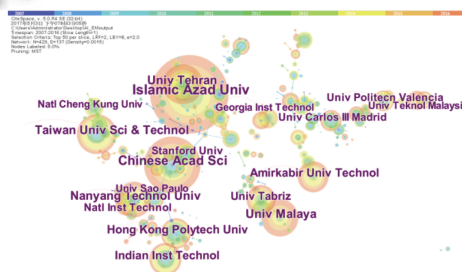
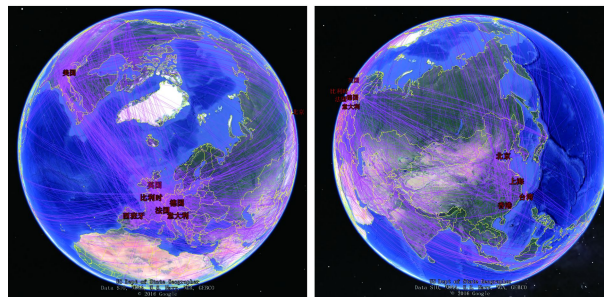


图 8 人工智能研究机构合作局部图谱

5.2.4 合作网络地理可视化分析

对国际人工智能研究作者合作网络进行地理可视化分析, 如图 9 所示. 图 (a) 为国际人工智能研究的合作网络; 图 (b) 为我国在参与人工智能研究中的合作网络. 如图 9 所示, 美国和欧洲在人工智能领域具有密切的合作, 同时国际人工智能研究中心位于美国和以西班牙、英国、意大利、法国和比利时为代表的欧洲国家. 中国和美国、欧洲都有合作, 但合作密度较小, 中国人工智能研究中心主要位于北京、上海、香港和台湾.



(a) 国际合作网络 (b) 中国合作网络

图 9 人工智能研究国际合作网络地理可视化

5.3 内容知识图谱及其分析

通过建立文献共被引图谱、关键词聚类图谱和突现词检测图谱实现内容知识图谱分析,从内容知识图谱对人工智能研究重要文献、研究热点和前沿进行分析与探究。

5.3.1 人工智能研究引用文献分析

当两篇文献同时被第三篇文献引用时,这两篇文献之间就存在共被引关系,文献被引频次可以较客观地反映该文献被学术界或知识域所认可的程度,以及在网络中的地位^[21]。持续高被引的文献被认为是一个领域的经典文献,一个领域的早期奠基性文献和经典文献则反映该领域研究的发展脉络和研究基础。本文利用 CiteSpace 软件进行文献共被引分析,设定网络节点为“CitedReference”,生成人工智能研究的文献共被引图谱,如图 10 所示。图谱中节点大小、颜色以及连线的粗细和颜色代表的含义和前文图谱的解读类似。

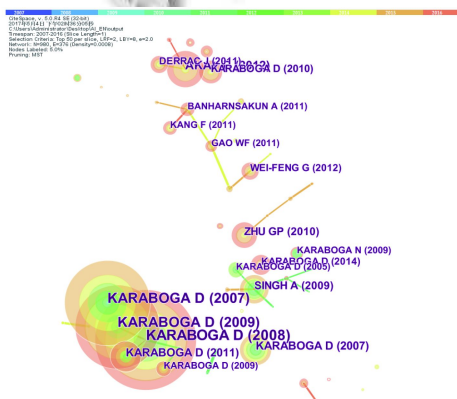


图 10 人工智能研究的文献共被引图谱

图 10 显示了各共被引文献的发表年份以及共被引用情况。从中可以看出 Karaboga D 和 Basturk B 于 2008 年所著的《On the performance of Artificial Bee Colony (ABC) algorithm》^[22]、Karaboga D 和 Akay B 2009 年所著的《A comparative study of Artificial Bee Colony algorithm》^[23]、Karaboga D 和 Basturk B 2007 年所著的《A powerful and efficient algorithm for numerical function optimization: Artificial Bee Colony (ABC) algorithm》^[24]被引频次分列前三。他们提出了一种较新的群体智能优化算法——人工蜂群算法。而被引频次排在前八的文献都是研究关于人工蜂群算法。这足以看出人工蜂群算法在人工智能领域被广泛应用,是人工神经网络训练、组合优化问题、系统和工程设

计等多个领域研究的基础理论,也是经典理论^[25]。

5.3.2 人工智能研究热点可视化分析

5.3.2.1 关键词研究

关键词是对文献研究内容的高度概括,在一定程度上可以代表一篇文献的研究内容。网络节点的中心性是反映节点核心程度的重要指标之一^[26],可以衡量某个节点在网络中的重要性,以及与其他节点之间联系的密切程度。从而,一段时间内科学研究者共同关注的核心问题即研究热点可以从中心性和频次高的关键词中得到体现^[27]。

分析人工智能研究热点时,设定网络节点为“Keyword”,运行软件后将同义关键词合并后得到研究文献中关键词共现网络,如图 11 所示。图谱中每个关键词用一个圆形节点表示,圆圈越大表明该关键词在人工智能研究领域内出现的频次越高,即高频关键词。其中需要关注的是带有紫红色光圈的节点,该节点代表的关键词具有较高的中心性,表示与其它关键词联系密切程度较高。其中关键词共现网络中有节点 104 个,连线 54 条,网络密度为 0.0101。相对于机构合作网络,关键词共现网络图谱的结构形态有了较大的优化和提升。但总体上关键词共现网络结构仍比较疏松、密度不高。这提示我们,未来研究中不仅要求科研机构之间加强合作,而且还需要科学学者专注于自己所研究的主题,选择合适的方向展开深入研究、精准研究,以避免对所研究的方向浅尝辄止。

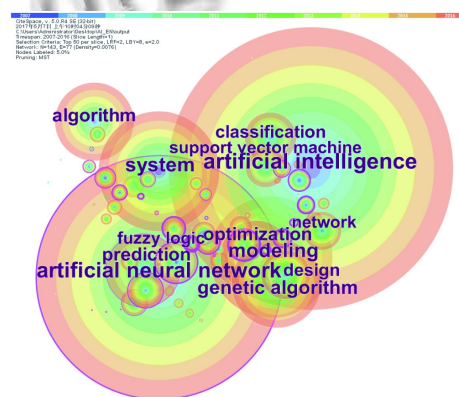


图 11 国内人工智能关键词共现图谱

近十年人工智能研究文献中出现的高频次和高中心性关键词如表 1 所示。研究文献中出现频次前十的关键词有“artificial intelligence”,“artificial neural network”,“modeling”,“system”,“genetic algorithm”,

“prediction”, “optimization”, “support vector machine”, “classification”, “design”. 结合高中心性关键词, 可以发现近十年国际人工智能研究的热点领域包括机器学习、数据挖掘、人工神经网络、群智能、计算智能、知识库系统; 人工智能研究的热点算法包括遗传算法、蚁群优化、决策树; 人工智能研究的热点方法及应用包括建模、预测、优化、分类、设计.

表1 人工智能研究文献高频次

序号	高频关键词		高中心性关键词	
	关键词名称	频次	关键词名称	中心性
1	artificial intelligence	1931	computational intelligence	0.93
2	artificial neural network	1660	swarm intelligence	0.91
3	modeling	838	artificial immune system	0.89
4	system	821	ant colony optimization	0.77
5	genetic algorithm	612	machine learning	0.71
6	algorithm	547	environment	0.71
7	prediction	505	data mining	0.68
8	optimization	498	decision tree	0.68
9	classification	357	knowledge based system	0.67
10	design	344	decision support system	0.65

5.3.2.2 关键词聚类分析

数据挖掘中聚类是对复杂网络中的数据根据相似性进行聚合与分类的过程^[28]. 在关键词聚类分析中, 聚类反映了网络中节点的相似性, 有助于识别和探测某一研究领域代表性知识子群, 即该研究领域内的热点主题. 设定网络节点为“Keyword”, 在生成关键词共现网络的基础上对其进行聚类处理, 最终得到如图12所示的人工智能研究领域的关键词聚类图谱.

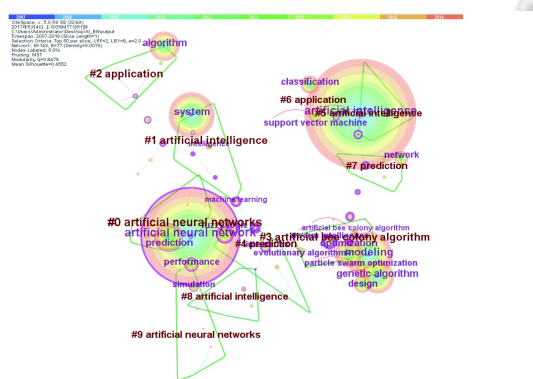


图12 关键词聚类图谱

人工智能的研究热点知识子群可总结为五大主题: 以预测 (prediction)、回归 (regression)、模糊推理系统 (fuzzy inference system) 为代表的人工神经网络方向的

研究热点; 以智能 (intelligence)、框架 (framework)、系统 (system)、环境 (environment) 为代表的机器学习方向的研究热点; 以算法 (algorithm)、智能体 (intelligent agent)、架构 (architecture)、管理 (management) 为代表的语义 Web 方向的研究热点; 以建模 (modeling)、优化 (optimization)、群智能 (swarm intelligence)、进化算法 (evolutionary algorithm) 为代表的人工蜂群算法方向的研究热点; 以模糊逻辑 (fuzzy logic)、计算智能 (computational intelligence)、识别 (identification)、时间序列 (time series) 为代表的软计算方向的研究热点. 上述研究热点主题与先前的研究结果部分相重复, 以人工神经网络为首的研究热点主题值得科研学者们高度关注. 近年来, 携着“深度学习”之势, 人工神经网络重新成为最强大的机器学习算法之一, 成为人工智能领域的研究热点主题.

5.3.3 结果分析中展示的人工智能领域研究前沿

研究热点和研究前沿不能混淆. 陈美超指出, 研究前沿是一组凸起的动态概念和潜在的研究问题, 研究前沿术语即出现频次快速增加的专业术语^[29]. CiteSpace 的一大突破便是提供一种基于词频增长算法 (Burst Detection) 的突现词分析方法, 该方法通过对引文关键词的统计, 将短时间内快速增长的专业词汇检测出来^[30]. 凭借这些突现词的时间分布和动态变化性的特点可以较好地反映知识域内的研究前沿和发展趋势. 设定网络节点为“Keyword”, 词语类型选择突现词 (Burst Terms). 如表2所示, 利用 CiteSpace 软件提供的词频增长算法, 将词频变化率高的词从大量关键词中探测出来, 并选取部分突现词按突现值大小排列, 通过考察突现词的历史共现频次分布总结其在一段时期内的趋势, 来确定人工智能研究领域的前沿.

表2 人工智能研究突现词

突现词	突现值	起始年份	终止年份	趋势
logic	15.85	2007	2011	上升到平稳
differential evolution	11.97	2014	2016	突然出现
global optimization	11.64	2014	2016	上升到下降
learning (artificial intelligence)	10.66	2014	2016	一直上升
architecture	10.64	2007	2010	下降到平稳
case based reasoning	9.69	2008	2011	上升到下降
complexity	9.53	2007	2009	一直上升
expert system	9.40	2007	2009	一直下降
artificial immune system	9.27	2009	2010	突然出现
Intelligent agent	8.60	2007	2009	下降到上升

表2 显示人工智能研究近十年词频变化率较高的关键词有逻辑 (logic)、差分进化 (differential

evolution)、全局优化 (global optimization)、人工智能学习 (learning (artificial intelligence))、架构 (architecture)、案例推理 (case based reasoning)、复杂性 (complexity)、专家系统 (expert system)、人工免疫系统 (artificial immune system) 和智能体 (intelligent agent), 这些突现词反应了近十年人工智能领域的研究前沿。

突现词一段时期内出现频次的趋势可以体现一个领域前沿发展趋势。如表 2 所示, 呈突然出现型的突现词有差分进化算法 (differential evolution) 和人工免疫系统 (artificial immune system), 这些词在近年突然出现, 可以视为未来发展的趋势; 呈上升型趋势的突现词有人工智能学习 (learning (artificial intelligence))、复杂性 (complexity) 和智能体 (intelligent agent), 这类关键词的研究时间并不长但仍然是领域内的研究热点, 研究者对其会继续关注; 呈平稳型趋势的突现词有逻辑 (logic) 和架构 (architecture), 这类词虽然在一段时间有上升或者下降的波动, 但总体趋于平稳的状态, 说明一直处于研究的热点之中。

6 研究结论

通过 CiteSpace 软件, 对 Web of Science 数据库中有关人工智能 2007–2016 年的文献数据进行不同层面的分析以及直观的可视化展示, 构建了时间分布图谱、国家合作图谱、机构合作图谱、地理可视化网络、文献共被引图谱、关键词聚类图谱, 突现词检测图谱, 研究得出以下结论。

1) 时间分布图谱表明: 人工智能作为一种前沿学科在整个计算机科学领域正处于被极度关注的情况, 近十年国际人工智能研究领域文献量持续增长, 全球人工智能发展持续火热, 研究正进入高速发展阶段。

2) 空间分布图谱表明: 在国家层面上, 美国、中国、西班牙、伊朗、英国等国发文量较多, 具有一定的人工智能研究基础; 在机构层面上, 跨机构之间的合作明显较少, 尚未形成明显的凝聚子群, 缺少具有足够影响力的科研机构; 在作者合作地理分布层面上, 美国和欧洲在人工智能领域有密切的合作, 是国际人工智能的研究中心, 中国和美国、欧洲都有合作, 但合作密度较小, 中国人工智能研究中心主要位于北京、上海、香港和台湾。

3) 共被引文献图谱表明: 人工智能领域的重要研

究文献有 Karaboga D 和 Basturk B 于 2008 年所著的《On the performance of Artificial Bee Colony (ABC) algorithm》、Karaboga D、Akay B 在 2009 年所著的《A comparative study of artificial bee colony algorithm》以及 Karaboga D 和 Basturk B 于 2007 年所著的《A powerful and efficient algorithm for numerical function optimization: Artificial Bee Colony (ABC) algorithm》, 这些文献是人工智能研究领域的基础理论, 也是经典理论。

4) 关键词共现图谱表明: 人工智能研究的热点领域包括机器学习、数据挖掘、人工神经网络、群智能、计算智能和知识库系统; 人工智能研究的热点算法包括遗传算法、蚁群优化和决策树; 人工智能研究的热点方法及应用包括建模、预测、优化、分类和设计。人工智能研究的热点主题有人工神经网络、机器学习、语义 Web、人工蜂群算法和软计算。

5) 突现词检测图谱表明: 人工智能领域近十年研究前沿包括逻辑 (logic)、差分进化 (differential evolution)、全局优化 (global optimization)、人工智能学习 (learning (artificial intelligence))、架构 (architecture)、案例推理 (case based reasoning)、复杂性 (complexity)、专家系统 (expert system)、人工免疫系统 (artificial immune system) 和智能体 (intelligent agent)。

人工智能领域已经成为世界各国科技革命和产业变革的新竞技场。积极研究和利用人工智能技术, 对我国科学研究发展和社会经济增长具有重要促进作用, 也是我国把握未来发展先机并提升自身国际竞争力的重要途径。

参考文献

- 1 朱巍, 陈慧慧, 田思媛, 等. 人工智能: 从科学梦到新蓝海——人工智能产业发展分析及对策. 科技进步与对策, 2016, 33(21): 66–70. [doi: 10.6049/kjbydc.2016040495]
- 2 Government Office for Science. Artificial intelligence: Opportunities and implications for the future of decision making. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/566075/gs-16-19-artificial-intelligence-ai-report.pdf. [2016-12].
- 3 National Science and Technology Council. The national artificial intelligence research and development strategic plan. http://www.360doc.com/content/16/1015/20/37334461_

- 598685262.shtml. [2016-10-15].
- 4 中华人民共和国政府工作报告. http://www.gov.cn/premier/2017-03/16/content_5177940.htm. [2017-03-16].
 - 5 陈明. 大数据可视化分析. 计算机教育, 2015, (5): 94–97.
 - 6 李健, 王丽萍, 刘瑞. 美国的大数据研发计划及对我国的启示. 中国科技资源导刊, 2013, 45(1): 17–23.
 - 7 李国杰. 大数据研究的科学价值. 中国计算机学会通讯, 2012, 8(9): 8–15.
 - 8 梁秀娟. 科学知识图谱研究综述. 图书馆杂志, 2009, 28(6): 58–62.
 - 9 Xu T Y, Qu H N, Zhao S S, *et al.* The visualization analysis of research hotspot and frontier technology of the smart power distribution and utilization based on the cite space. *Energy and Power Engineering*, 2017, 9(4B): 515–524.
 - 10 刘则渊, 王贤文, 陈超美. 科学知识图谱方法及其在科技情报中的应用. 数字图书馆论坛, 2009, (10): 14–34. [doi: 10.3772/j.issn.1673-2286.2009.10.004]
 - 11 梁静, 任增元. 我国研究生教育研究进展的文献计量分析. 现代教育管理, 2015, (12): 103–108. [doi: 10.3969/j.issn.1674-5485.2015.12.020]
 - 12 侯剑华, 胡志刚. CiteSpace 软件应用研究的回顾与展望. 现代情报, 2013, 33(4): 99–103.
 - 13 Chen CM. Searching for intellectual turning points: Progressive knowledge domain visualization. *Proceedings of the National Academy of Sciences of the United States of America*, 2004, 101(S1): 5303–5310.
 - 14 陈悦, 刘则渊. 悄然兴起的科学知识图谱. 科学学研究, 2005, 23(2): 149–154.
 - 15 姜俊锋, 丁香乾, 侯瑞春, 等. 基于 Citespace III 的大数据研究的可视化分析. 计算机与数字工程, 2016, 44(2): 291–295, 299.
 - 16 郭俊仓. 个人学术水平评价指标——h 因子与 ph 因子概念介绍. 西安工业大学学报, 2015, 35(11): 909.
 - 17 Liu HL, Zhu YP, Guo YZ, *et al.* Visualization analysis of subject, region, author, and citation on crop growth model by CiteSpace II software. In: Wen ZK, Li TR, eds. *Knowledge Engineering and Management*. Berlin, Heidelberg: Springer, 2014. 243–252.
 - 18 蔡建东, 马婧, 袁媛. 国外 CSCL 理论的演进与前沿热点问题——基于 Citespace 的可视化分析. 现代教育技术, 2012, 22(5): 10–16.
 - 19 Chen C. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology*, 2006, 57(3): 359–377. [doi: 10.1002/(ISSN)1532-2890]
 - 20 李杰, 陈超美. CiteSpace: 科技文本挖掘及可视化. 北京: 首都经济贸易大学出版社, 2016. 180–185.
 - 21 金丽娜. 基于复杂网络的文献关联关系模型的建立与分析 [硕士学位论文]. 北京: 北京化工大学, 2013. 27–32.
 - 22 Karaboga D, Basturk B. On the performance of artificial bee colony (ABC) algorithm. *Applied Soft Computing*, 2008, 8(1): 687–697. [doi: 10.1016/j.asoc.2007.05.007]
 - 23 Karaboga D, Akay B. A comparative study of artificial bee colony algorithm. *Applied Mathematics and Computation*, 2009, 214(1): 108–132. [doi: 10.1016/j.amc.2009.03.090]
 - 24 Karaboga D, Basturk B. A powerful and efficient algorithm for numerical function optimization: Artificial bee colony (ABC) algorithm. *Journal of Global Optimization*, 2007, 39(3): 459–471. [doi: 10.1007/s10898-007-9149-x]
 - 25 秦全德, 程适, 李丽, 等. 人工蜂群算法研究综述. 智能系统学报, 2014, 9(2): 127–135. [doi: 10.3969/j.issn.1673-4785.201307019]
 - 26 刘军. 社会网络分析导论. 北京: 社会科学文献出版社, 2004. 170–185.
 - 27 Freeman LC. Centrality in social networks conceptual clarification. *Social Networks*, 1978-1979, 1(3): 215–239. [doi: 10.1016/0378-8733(78)90021-7]
 - 28 Chen CM. Visualising semantic spaces and author co-citation networks in digital libraries. *Information Processing & Management*, 1999, 35(3): 401–420.
 - 29 陈超美. CITESPACE II: 科学文献中新趋势与新动态的识别与可视化. 陈悦, 侯剑华, 梁永霞, 译. 情报学报, 2009, 28(3): 401–402.
 - 30 辛伟. 知识图谱在军事心理学研究中的应用 [硕士学位论文]. 西安: 第四军医大学, 2013. 52–59.