



开放科学
(资源服务)
标识码
(OSID)

国际人工智能研究主体结构和知识焦点分析

郑丽霞¹ 林卓^{1,2}

1. 福建省科学技术信息研究所 福州 350003;
2. 福建省信息网络重点实验室 福州 350003

摘要: 为展现国际人工智能领域研究主体结构和知识焦点,以 Web of Science 核心集为数据源,利用 CiteSpace 可视化软件,从研究区域和机构合作的视角来展现当前国际人工智能领域研究群体结构。利用主题词分析来挖掘当前国际人工智能领域研究热点和前沿,并借助文献共被引分析来挖掘重要研究成果。研究结果较为完整地展现国际人工智能研究主体和知识焦点。

关键词: 人工智能; 热点; 前沿; 主体结构; CiteSpace

中图分类号: G35

Analysis of Research Subjects Structure and Knowledge Focus in the Field of International Artificial Intelligence

ZHENG Lixia¹ LIN Zhuo^{1,2}

1. Fujian Institute of Science & Technology Information, Fuzhou 350003, China;
2. Fujian Key Laboratory of Information and Network, Fuzhou 350003, China

Abstract: The paper is to reveal the research subjects structure and knowledge structure in the field of international artificial intelligence. The paper takes Web of Science Core Collection as data source, uses CiteSpace visualization software, reveals research regional structure and subjects structure of the international artificial intelligence from the perspective of research institute collaboration. The paper uses subject words analysis to mine the research hotspots and research front; and uses document co-citation analysis to mine the important research results. The results reveal the research subject structure and knowledge focus in the field of international artificial intelligence completely.

Keywords: Artificial intelligence; research hotspots; research front; subject structure; CiteSpace

基金项目: 中国科学技术信息研究所情报工程实验室 2018 年开放基金项目“中国人工智能国际合作现状分析与对策研究(2019)”；2019 年福建省属公益类科研院所基本科研专项“产业集群知识产权风险识别及预警研究”(2019R1008-4)。

作者简介: 郑丽霞(1983-), 硕士, 助理研究员, 研究方向为科技情报、情报服务、专利情报等; 林卓(1986-), 通讯作者, 博士, 助理研究员, 研究方向为数据挖掘与可视化、信息管理与信息系统、知识图谱等。

作为引领未来前沿性、战略性技术,人工智能正重塑传统产业发展模式,给教育、交通、物流、医疗、制造等多个行业带来新的变革。国际上,不少国家和地区将人工智能列为国家层面战略技术,积极布局其研发及在产业中的应用实践。近年来,不少国家都在人工智能领域做了进一步的布局^[1]。欧盟2018年签署《人工智能合作宣言》旨在共同推进人工智能发展;荷兰2019年公布《人工智能战略行动计划》;美国2019年发布《国家人工智能研究与发展战略规划》更新版;韩国2019年制定《推动数据、人工智能、氢经济发展规划》。中国2018年推出《新一代人工智能产业创新重点任务揭榜工作方案》;2019年又公布《关于促进人工智能和实体经济深度融合的指导意见》;2020年,新基建多次进入国家层面的布局,并明确将人工智能纳入“新基建”范畴。可见,人工智能在世界经济、社会发展中的重要意义,厘清人工智能的研究活跃主体和知识焦点显得十分必要。

从国内近年对人工智能领域研究概貌的文献来看,较多集中在对热点、前沿和总体概貌的研究。①人工智能领域研究热点、前沿的探测。姜宇星^[2]、王曰芬^[3]、范丽鹏等^[4]基于论文视角从主要国家(地区)对比、领域全局演化、高产机构对比等多个维度识别国际人工智能研究的前沿。陶于祥等^[5]亦以论文为样本,利用科学计量法梳理了国内外人工智能领域研究的热点、发展脉络等。邓启平等^[6]利用复杂网络社区探测算法和文献计量指标等探测国际人工智能的研究前沿。②人工智能领域研究总体格局的分析。牛奉高等^[7]从加权创新系数出发,

分析了人工智能研究的网络特征与演化。张春博等^[8]从 AAAI2002~2011 的年会论文入手,分析和总结国际上人工智能领域研究的国别、机构和作者,以及研究热点和前沿等。此外,还有部分学者从专利视角来挖掘人工智能领域的研究热点。如,王友发等^[9]利用专利地图展现人工智能领域研究总概貌、技术热点等。

在此现实和理论研究背景下,本文以基于 Web of Science 核心集(Web of Science Core Collection)为数据源,利用 CiteSpace 分析工具,分析人工智能领域研究的群体结构和研究热点前沿及重要成果,厘清国际人工智能领域的研究主体结构 and 知识焦点。

1 数据来源及研究方法

1.1 数据来源

对于数据源,选用 Web of Science 核心集作为国际论文来源的检索库。基于清华大学研发的 Aminer 科技情报大数据平台^[10]和领域专家咨询调研结果,本文对人工智能的子领域进行重构,确定 21 个子领域及二级主题词共 778 个作为检索关键词,确定的子领域包括“artificial intelligence(经典人工智能)”“data mining(数据挖掘)”“machine learning(机器学习)”“social network(社交网络)”“visualization(可视化)”“computer graphics(计算机图形学)”“computer vision(计算机视觉)”“virtual reality(虚拟现实)”“multimedia(多媒体技术)”“human computer interaction(人机交互)”“knowledge engineering(知识工程)”“speech recognition(语音识别)”“internet of things

(物联网)”“information system (信息系统)”“security and privacy (计算机安全与隐私)”“information retrieval(信息检索)”“database (数据库)”“natural language processing (自然语言处理)”“robotics(机器人学)”“operating system (操作系统)”“recommender system (推荐系统)”为检索词,数据检索截止时间为2020年3月31日。剔除重复文献后,收集的总文献量为313597篇。

1.2 研究方法

本文以 Web of Science 的论文文献数据为研究对象,利用 CiteSpace 知识图谱科学文献分析工具,从区域合作和研究机构合作网络把握当前国际人工智能研究知识网络中的主体结构。而对于研究焦点的分析,一是利用主题词分析把握国际人工智能的研究热点和前沿;二是基于文献共被引分析,挖掘重要研究成果。研究思路和过程如图1所示。

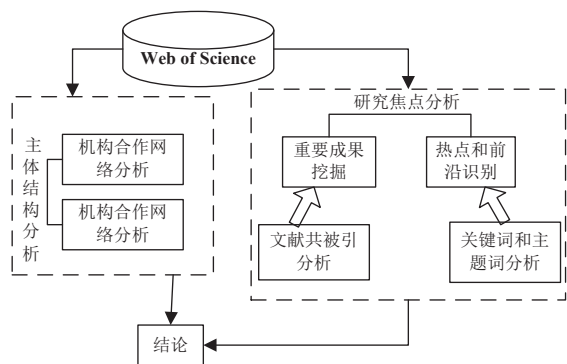


图1 研究思路

2 主体结构：研究主体合作网络分析

知识主体之间构成的知识网络^[11],也被认

为是知识主体之间的知识合作网络^[12]。本文试图从论文合作的视角来探讨国际人工智能领域知识群体之间构成的网络,即主体结构。

2.1 国家(地区)合作网络分析

图2为国家(地区)间人工智能研究合作图谱。图谱中的国家为合作频次排名前26的国家(地区),其中排名21-26的国家或也是突现度较高的国家或地区。

如图2所示,合作网络中节点大小反映国家(地区)中心度,主要呈现两大合作模式。一是以英国、美国为中心的辐射式合作方式。这两个国家在人工智能领域论文的合作方式是以他们为中心,向外辐射,其中英国的主要合作对象是德国、荷兰、希腊、瑞典、意大利等;美国的主要合作对象是新加坡、印度、澳大利亚等国,其中澳大利亚还跟伊朗开展了深入合作。二是两两合作模式,如亚洲的中国与日本,欧洲的葡萄牙与西班牙、法国与瑞士。

表1列出中心度前10及合作频次前10的人工智能领域研究论文合作国家(地区)。从各国家(地区)对应的中心度值来看,人工智能领域论文合作有较大影响的国家(地区)可以大致分为三个梯队。第一梯队是美国和英国,其中英国节点最大,其次美国,这两个国家的节点远大于其他国家或地区。结合图谱显示(图2),以英、美两国为中心向外发散的辐射式的网络构成整个图谱的两大架构。第二梯队是意大利、荷兰、中国、葡萄牙、澳大利亚、比利时等国,节点较大,这些节点在图2中可见,这些国家的中心度大于0.2、小于0.5,仅次于第一梯队的美国和英国。第三梯队的国家(地区)

为德国、法国、伊朗和葡萄牙，中心度小于0.2、大于0.1，也有一定的合作影响力。最后，其他国家（地区）处于第四梯队，中心度为0，合

作影响力弱。简而言之，从中心度这一指标来看，前三大梯队构成了国际人工智能领域研究国家层面的主体结构。

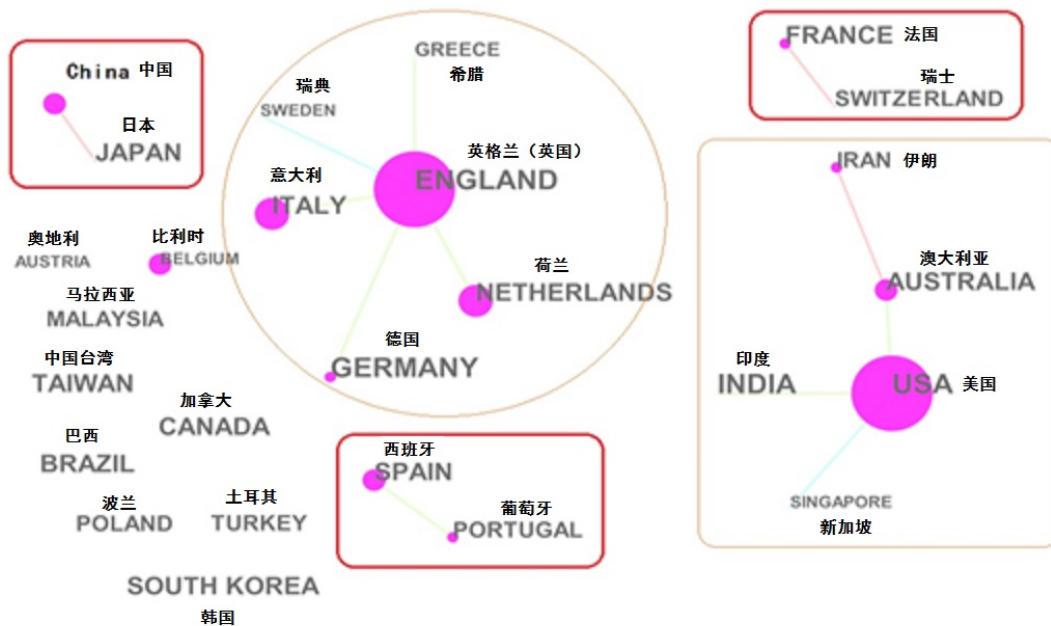


图2 国家（地区）间人工智能研究合作图谱

表1 人工智能领域研究论文合作国家（地区）的中心度和合作频次排名

排名	中心度	合作频次	国家或地区	排名	合作频次	中心度	国家或地区
1	1.31	16533	英国（欧洲）	1	67400	1.28	美国（北美洲）
2	1.28	67400	美国（北美洲）	2	53758	0.3	中国（亚洲）
3	0.42	11773	意大利（欧洲）	3	16534	1.31	英格兰（欧洲）
3	0.42	5255	荷兰（欧洲）	4	16389	0.15	德国（欧洲）
5	0.3	53758	中国（亚洲）	5	16319	0	印度（亚洲）
6	0.29	11846	西班牙（欧洲）	6	12472	0	日本（亚洲）
6	0.29	9283	澳大利亚（大洋洲）	7	11971	0	加拿大（北美洲）
6	0.29	363	比利时（欧洲）	8	11846	0.29	西班牙（欧洲）
9	0.15	16388	德国（欧洲）	9	11786	0.15	法国（欧洲）
9	0.15	11785	法国（欧洲）	10	11773	0.42	意大利（欧洲）
9	0.15	3784	伊朗（亚洲）				
9	0.15	3443	葡萄牙（欧洲）				

从合作频次来看，美国、中国是合作发文量较多的国家，这两个国家总的合作频次超过

之后的8个国家合作频次之和。可见，中、美两国十分重视学术研究过程中的合作模式，遥

遥领先其他国家（地区）。美国人工智能领域文章的合作频次位居世界之首，为 67400 次，其中心度达到 1.28，仅次于英国的 1.31，中心度排名位居第 2，合作影响力大，可见美国在人工智能国家合作网络上起到重要作用。其次是中国，合作频次为 53758 次，中心度为 0.3，中心度排名位居第 5，合作影响力也较大。换言之，从合作频次这一指标来看，美国和中国成为国际人工智能的主体结构。

结合合作频次和中心度来分析国家合作的区域分布，欧洲国家的人工智能国际合作水平较高。英国、德国、西班牙、法国和意大利等 5 个国家在合作频次排行中进入前 10。在中心度排行中，欧洲的国家也表现甚好，中心度排名前 10 的国家中，欧洲占了 8 个。第二个值得关注的区域是亚洲，除了中国外，印度、日本在人工智能研究领域也占有重要的贡献。但这些国家和地区中，只有中国的中心度大于 0.1，其他的均为 0。可见合作频次进入前 10 的亚洲国家或地区，大部分的影响力弱。

总体而言，国家层面上看，美国、中国、意大利、荷兰、葡萄牙、澳大利亚、比利时等构成国际人工智能领域主体结构。区域上，值得关注的是欧洲和亚洲。

2.2 机构合作网络分析

图 3 展示的是全球中心度大于 0 的研究机构，以及合作频次位于前 50 的所有的中国大陆研究机构的合作情况。根据各主要研究机构中心度的大小，在学术论文视角下人工智能领域研究机构学术合作网络呈现以下 3 种结构：

第一种是“辐射式”合作网络。该合作网络主要是以中心度较大的美国的杜克大学、斯坦福大学和中国的中科院三大研究机构为中心向外辐射。这三大研究机构偏向于国内合作。其中，中国科学院的合作对象主要为中国的北京航空航天大学、清华大学、北京大学、上海交通大学、武汉大学等，且这些机构间没有十分明显的合作关系，主要是围绕着中科院相关研究所展开合作。

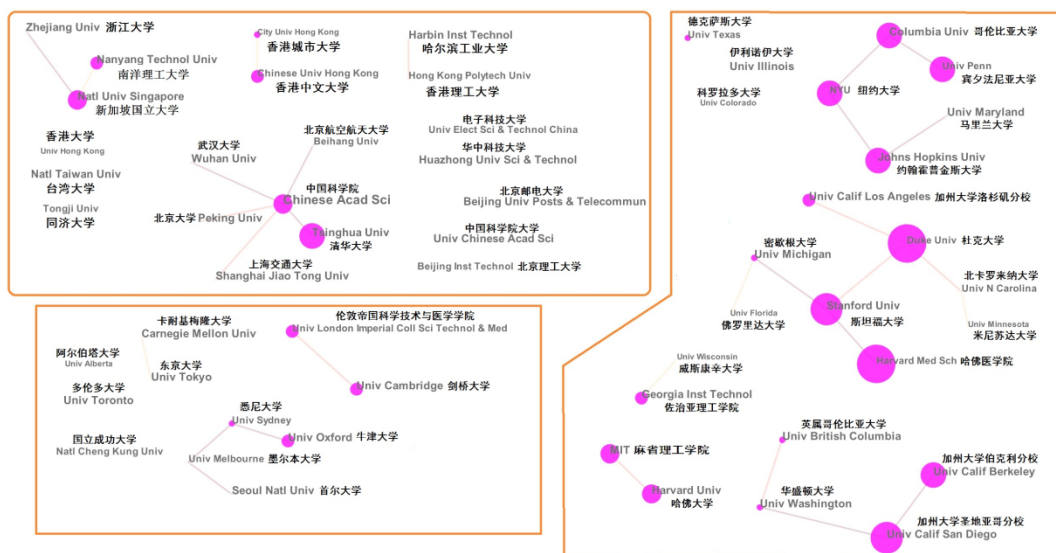


图 3 人工智能领域研究机构论文合作网络图谱

第二种是“串联式”合作网络。“串联式”合作模式，形成链条式的合作网络。图3中显示，“串联式”合作网络主要有4组研究机构。第一组是由美国的加州大学伯克利分校、加州大学圣地亚哥分校、华盛顿大学和加拿大的英属哥伦比亚大学等机构串联而成的国内、国际合作网络；第二组是由美国的宾夕法尼亚大学、哥伦比亚大学、纽约大学、约翰霍普金斯大学、马里兰大学等研究机构串联而成的美国国内合作网络；第三组是中国的浙江大学、新加坡国立大学和南洋理工大学等串联而成的国内、国际合作网络；第四组是由英国的牛津大学、澳大利亚的悉尼大学和墨尔本大学、韩国的首尔大学等机构形成国际合作网络。

第三种是“两两式”合作网络。美国形成麻省理工学院和哈佛大学、威斯康星大学与佐治亚理工学院的国内合作网络；中国形成哈尔滨工业大学和香港理工大学、香港城市大学与香港中文大学的国内合作网络；英国形成伦敦帝国科学技术与医学学院和剑桥大学的国内合作网络；美国的卡内基梅隆大学和日本的东京大学形成国际合作网络。

从机构间的合作频次来看，如表2所示，排名前3的分别为中国科学院4075次、清华大学1719次、美国的伊利诺伊大学1417次。从合作机构的性质来看，合作频次和中心度排名前20的机构均为高校或科研院所。从合作机构所属区域分布来看，合作频次排名前20的机构主要分布在美国、中国、新加坡、加拿大、日本、巴西和中国台湾等国家（地区）。其中，数量最多的是美国（9个）；其次是中国大陆（5个），

美国和中国两国占了70%，美国占了45%。美国的这9家研究机构合作频次之和为11337，约占美国合作总频次的17%，对美国人工智能领域论文合作的贡献较大。中国大陆的人工智能领域的论文合作频次上，中国大陆这5家研究机构对中国大陆人工智能领域论文的合作贡献较大，合作频次为9624，约占中国大陆合作总频次的18%。这也反映出，在研发机构的论文合作能力分布上，中国大陆的研发机构相比美国而言较为集中。

比较国家（地区）的合作频次排名情况，英国、德国、西班牙、法国、意大利、荷兰、瑞士、葡萄牙等欧洲国家的研究机构均未出现在研究机构合作频次的排名前20中，表明这些国家在人工智能领域论文合作机构的分布较为分散，换言之在人工智能领域，这些国家的研究机构的合作能力较为不够突出。新加坡在国家层面的合作频次和排名中，未进入前20，但在合作频次的机构排名中有出现新加坡的研究机构。新加坡南洋理工大学的合作频次为1379，位居第6；新加坡国立大学的合作频次为1135，位居第16。可见新加坡这两个机构具有较强的合作研究能力，且侧面也反映出在学术论文视角下新加坡合作能力较强的研究机构仅集中在这2家。另外，新加坡国立大学和南洋理工大学的中心度均较高，分别为0.49、0.39，说明新加坡的这2个机构在国际合作方面也具有一定的影响力。

从合作中心度来看，中心度排名前20的研究机构均多数为高校，个别为科研院所。从合作机构所属区域分布来看，中心度排名前20的研究机构主要分布在美国、中国、新

加坡、英国和中国香港等地。其中，美国拥有最多合作效果较好的研究机构。中心度排名前 10 的研究机构中，美国的研究机构占了 9 个。结合图 2 人工智能领域研究机构论文合作网络图谱来看，美国的这些机构主要是在

美国国内范围内开展人工智能论文合作。值得关注的，哈佛医学院虽然中心度较高，但合作频次仅为 459，且突变率较高，说明其在近年才出现与人工智能相关的重要合作文献且突增现象显著。

表 2 人工智能领域研究机构论文合作频次排名

排名	合作频次	中心度	研究机构	排名	中心度	合作频次	研究机构
1	4075	0.58	中国·中国科学院	1	1.03	459	美国·哈佛医学院
2	1719	0.7	中国·清华大学	1	1.03	280	美国·杜克大学
3	1417	0	美国·伊利诺伊大学	3	0.95	1274	美国·斯坦福大学
4	1385	0.04	中国·上海交通大学	4	0.86	914	美国·加州大学圣地亚哥分校
5	1380	0.09	美国·卡耐基梅隆大学	5	0.71	916	美国·加州大学伯克利分校
6	1379	0.39	新加坡·南洋理工大学	6	0.7	1719	中国·清华大学
7	1374	0.17	美国·华盛顿大学	7	0.66	467	美国·宾夕法尼亚大学
8	1349	0.04	加拿大·多伦多大学	8	0.64	917	美国·哥伦比亚大学
9	1340	0	中国·浙江大学	9	0.62	985	美国·约翰霍普金斯大学
10	1287	0.49	美国·麻省理工学院	9	0.62	480	美国·纽约大学
11	1274	0.95	美国·斯坦福大学	11	0.58	4075	中国·中国科学院
12	1233	0.04	日本·京东大学	12	0.49	1287	美国·麻省理工学院 MIT
13	1197	0.13	美国·密歇根大学	12	0.49	1135	新加坡·新加坡国立大学
14	1184	0.42	美国·哈佛大学	14	0.42	1184	美国·哈佛大学
15	1160	0.25	美国·佐治亚理工学院	15	0.39	1379	新加坡·新加坡南洋理工大学
16	1135	0.49	新加坡·新加坡国立大学	15	0.39	967	英国·剑桥大学
17	1105	0	中国·北京邮电大学	17	0.33	424	中国香港·香港中文大学
18	1064	0	美国·马里兰大学	18	0.32	280	英国·伦敦帝国科学技术与医学学院
19	1059	0.09	中国·台湾大学	19	0.25	1160	美国·佐治亚理工学院
20	1036	0	巴西·圣保罗大学	19	0.25	817	英国·牛津大学

其次是中国。中国大陆的清华大学和中国科学院中心度分别以 0.7 和 0.58 位居第 6 和第 11。中国大陆的这两家研究机构在人工智能领

域的论文合作目前主要也是在国内范围内开展。中国（香港）的香港中文大学的合作中心度为 0.33，位居第 17，其合作对象主要为香港城市

大学。

再次是新加坡的国立大学和南洋理工大学，中心度分别以 0.49 和 0.39 位居第 12（并列）和第 15 位（并列）。其中，新加坡国立大学与中国的浙江大学开展较多的合作。例如 2019 年 6 月，新加坡国立大学联合浙江大学、网易开发分布式大规模深度学习云项目；2019 年 8 月，由新加坡国立大学与中国的浙江大学、北京理工大学联合发表的论文在 LDB 会议（即 Very Large Data Bases）上获得最佳论文奖。

最后是英国。在合作频次排名中，英国的研究机构表现一般，均未进入前 20；但在中心度排名中，剑桥大学、伦敦帝国科学技术与医学学院，以及牛津大学，这三所高校的中心度分别为 0.39、0.32 和 0.25，位居第 15（并列）、第 18 和第 19（并列）。根据图谱显示（图 3），剑桥大学主要与伦敦帝国科学技术与医学学院开展两两合作；牛津大学主要与澳大利亚的悉尼大学等开展“串联式”的合作。

总体而言，中、美两国研究机构的合作能力较大，且中国国内的研发机构合作能力较美国集中；英国、德国、西班牙、法国、意大利、荷兰、瑞士、葡萄牙等欧洲国家的研究机构的合作能力则较为分散；美国、中国、新加坡、英国和中国香港等地的研究机构的合作影响力较大。

3 知识焦点：研究热点前沿、及重要成果挖掘

在分析国际人工智能研究主体结构的基础上，我们从研究热点前沿和重要成果的视角来

探讨国际人工智能领域研究的聚集点。

3.1 研究热点前沿识别——基于主题词分析

词频分析中的高频词在一定程度上代表研究热点^[13]，而主题词突变检测可呈现领域研究的前沿^[14-15]。且基于主题词检测（term burst）的主题分析可以更为深入、全面把握研究热点与识别研究前沿。为此，笔者从主题词的词频监测和突变检测来识别人工智能领域研究热点前沿。剔除其中重复的、专指度较低的系统（system）、模型（model）、管理（management）；选择频次及突变率排名靠前的关键词，对 CiteSpace 所导出的主题相关数据进行抽取，进而对国际人工智能领域研究热点进行识别。结果见表 3。

基于主题词的词频分析和突变率分析，剔除已经退出热点的主题词（如信息检索），检测结果发现，国际人工智能领域研究近年研究焦点中，聚焦理论层面的研究主要围绕“算法”和“分类”；理论研究和应用研究兼顾的研究涵盖了“数据挖掘”“机器学习”“物联网”“数据库”“语音识别”“深度学习”和“大数据”等相关领域。

3.2 重要成果挖掘——基于文献共被引分析

图 4 为 CiteSpace 绘制的人工智能领域被引频次大于 100、中心度大于 0.05 的文献共被引知识图谱。共现网络谱图显示，以斯洛文尼亚的学者 Demšar J 在 2006 年发表的论文 *Statistical Comparisons of Classifiers Over Multiple Data Sets*、美国斯坦福大学统计学、生物医学数据科学教授 Hastie T 在 2001 年出版的书 *The Ele-*

ments of Statistical Learning: Data Mining (2009年, 出版新的版本) 和伊利诺伊大学计算机科学教授 Han J 出的书 Data Mining: Concepts and

Techniques 为中心的三大共现网络组成主体结构。这三大中心文献都是数据挖掘领域的研究成果。

表 3 人工智能研究热点前沿

研究热点前沿关键词	相关内容
分类 (classification)	出现频次和中心度均最高, 突变权重较大。近5年依然备受关注的焦点有理论层面的分类算法、模糊分类、数据流分类、标签分类等; 应用层面的情感分类、恶意软件分类、图像分类、证型分类、交通分类等。
数据挖掘 (data mining)	至今还在延续的研究焦点有: 大数据挖掘、教育领域数据挖掘、数据挖掘技术等
机器学习 (machine learning)	至今还在延续的研究焦点有: 利用机器学习预测、机器学习算法、方法、机器学习模型、机器学习运用等。
物联网 (internet of thing)	中心度和突变度排名均靠前。物联网领域的研究前沿热点主要涉及物联网应用、物联网设备、物联网服务等。
算法 (algorithm)	中心度显示其影响力较大。至今还备受关注的的前沿热点有学习算法、萤火虫算法、人工蜂群算法等。
数据库 (database)	大约于2001年成为一大研究焦点。目前的研究焦点主要为数据库健康医疗、保险行业的应用研究。
语音识别 (speech recognition)	研究焦点主要有理论层面的分布式语音识别、连续语音识别、语音识别技术、汉语语音识别、噪音语音识别等; 应用层面上的主要有试听教学语音识别、孩子语音识别、电话语音识别、新闻方面的语音识别等。
深度学习 (deep learning)	在本研究范围内, “深度学习” 出现的频次没有进入前20, 但其突变率位居第一, 为机器学习中一个新的研究领域。
大数据 (big data)	突变率较高, 大数据是2012年开始兴起的一大研究领域, 至今依然备受关注, 主要领域为大数据应用、大数据分析、大数据挖掘、大数据管理、大数据技术等。

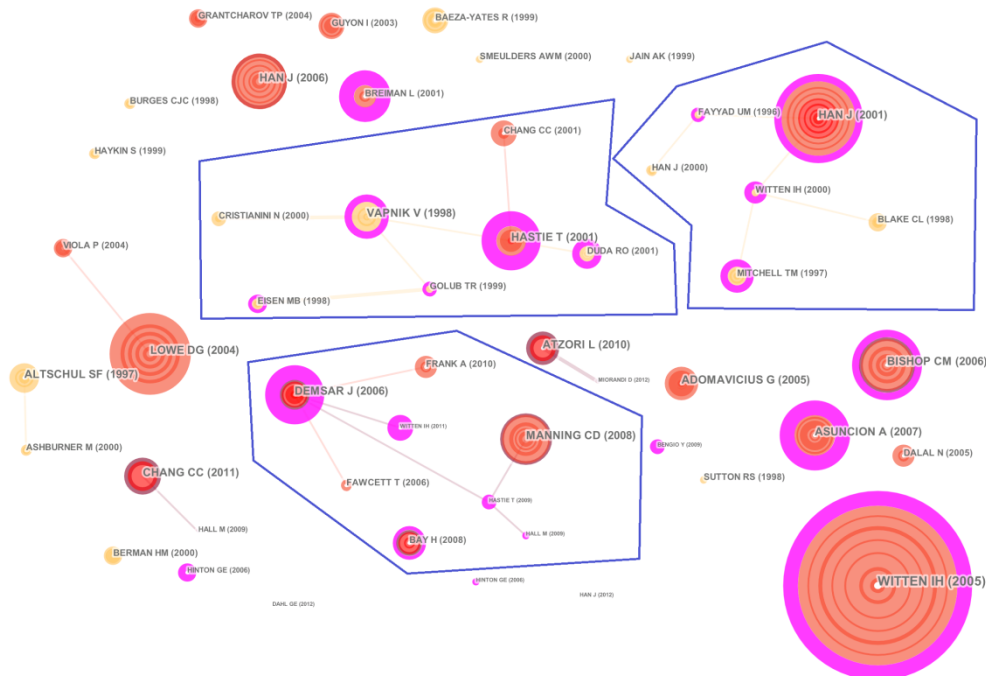


图 4 人工智能领域研究重要文献共被引网络图谱

学者 Demšar J 在 2006 年发表的论文 *Statistical Comparisons of Classifiers Over Multiple Data Sets* 的这一中心文献，中心度高达 0.76，位居第一，主要探讨的内容是可用在多个数据集的分类器，属于数据挖掘领域的研究成果。Demšar J 的这篇文章还延伸出中心度较小的、与其有这共被引关系的文献群。从文献群所属学科领域来看，多数属于数据挖掘、机器学习领域。如新西兰怀卡托大学的 Witten IH 在 2011 年出的书 *Data Mining: Practical Machine Learning Tools and Techniques* (2000 年版，中心度为 0.29，位居第 13；2005 年版，中心度也高达 0.64，位居第 3，主要探讨机器学习和数据挖掘领域的技术变革和方法，属于数据挖掘领域的经典教材。来自美国斯坦福大学统计学、生物医学数据科学教授 Hastie T 在 2009 年出版的书 *The Elements of Statistical Learning: Data Mining* (2001 年，出第一版，为第二个共现网络的中心文献)，属数据挖掘领域的研究成果。

第二大共现网络是以 Hastie T 教授的书 *The Elements of Statistical Learning: Data Mining* 为中心，向外延伸，文献群主要展现的是数据挖掘、机器学习、统计学、计算机等领域的研究成果。这本书主要探讨数据挖掘的理论与案例，属于数据挖掘领域。与其共现的文献主要有：来自中国台湾的“台湾国立大学”Chang CC 的 *LIBSVM: A Library for Support Vector Machines*，主要探讨支持向量机的 LibSVM 工具包、属于机器学习领域的研究成果；美国 NEC 实验室研究员统计学、机器学习领域 Vapnik V 教授的文章 *Statistical Learning Theory*，该文章主要探讨统计学习理论，属于统计学领域的研究成

果；美国加州圣何塞州立大学电气工程系名誉教授 Duda R 的文章 *Pattern Classification*，主要研究模式识别相关理论，研究内容涉及神经网络、机器学习、支持向量机等理论，为模式识别指明研究方向，属于计算机领域研究成果。

第三大共现网络是以 Han J 在 2001 出的书 *Data Mining: Concepts and Techniques* 为中心文献，该文献的中心度为 0.21，位居第 15，主要展现数据挖掘技术相关的内容。与其共现的文献相对没有前面两大中心文献多，主要为新西兰怀卡托大学的 Witten IH 的 2000 版 *Data Mining: Practical Machine Learning Tools and Techniques*，亦为数据挖掘领域研究成果；卡内基梅隆大学机器学习领域教授 Mitchell TM 的书 *Machine Learning*，机器学习领域的经典教材。

从中心度来看，排名前十的文献多数为数据挖掘、机器学习等基础性学科领域。从被引频次来看，被引频次较高的文献主要涉及数据挖掘、机器学习、文献检索、物联网、统计学习理论等学科领域。其中，中心度排名较高的除了前面在三大共现网络中提及到的部分文献外，还有：加州大学伯克利分校统计学领域的教授 Leo Breiman 在 2001 年发表的论文 *Using Iterated Bagging to Debias Regressions*，中心度位居第 4，探讨迭代 bagging 回归算法，属于机器学习领域的研究成果；机器学习领域的数据集“UCI Machine Learning Repository”，中心度位居第 5；英国微软剑桥研究院学者 Bishop CM 的 *Pattern Recognition and Machine Learning*，中心度位居第 7，介绍模式识别和机器学习的最新进展，属于机器学习领域的研究成果。

4 结论

通过 CiteSpace 分析了国际人工智能领域的主体结构和知识焦点, 研究发现: ①主体结构上, 国家层面上看, 美国、中国、意大利、荷兰、葡萄牙、澳大利亚、比利时等构成国际人工智能领域主体结构。机构网络呈现出“辐射式”“串联式”“两两式”3种合作网络, 中国的研究机构大多围绕着中国科学院呈辐射式合作, 而美国的研究机构呈现多种的国内、国际合作形式, 在机构合作中占据主导地位。②通过主题词的共现图谱得到人工智能研究领域的热点与前沿, 发现近年来备受关注的人工智能学术热点前沿领域, 包括有机器学习(machine learning)、深度学习(deep learning)等9个。文献共被引分析结果显示, 三大共现网络的中心文献为斯洛文尼亚的学者 Demšar J、美国斯坦福大学统计学、生物医学数据科学教授 Hastie, 以及伊利诺伊大学计算机科学教授 Han J 等学者的研究成果, 且三大共现网络中的中心文献都是数据挖掘领域的研究成果。

在研究过程中, 搜集的数据是基于人工智能领域的论文文献, 不可避免的会遇到部分数据缺失以及不完整的情况。虽在研究时已尽可能将这类数据的影响降低, 但由于基础数据量巨大, 难免会出现许多遗漏。但综合来看, 通过国际人工智能领域研究主体和知识焦点的研究对于未来进一步开展人工智能的研究具有重要意义。其主体结构的成员对于接下去人工智能领域合作对象的选择具有一定的参考意义; 而研究焦点对于未来人工智能研究细项领域的选择具有一定的借鉴意义。

参考文献

- [1] 中国信通院. 全球人工智能战略与政策观察(2019) [EB/OL]. (2019-08-26). <http://aiaaorg.cn/uploadfile/2019/0829/20190829102318489.pdf>.
- [2] 姜宇星, 王曰芬, 范丽鹏, 等. 人工智能研究前沿识别与分析: 基于主要国家(地区)对比研究视角[J]. 情报理论与实践, 2019, 42(9):8-15.
- [3] 王曰芬, 曹嘉君, 余厚强, 等. 人工智能研究前沿识别与分析: 基于领域全局演化研究视角[J]. 情报理论与实践, 2019, 42(9):2-7.
- [4] 范丽鹏, 余厚强, 姜宇星, 等. 人工智能研究前沿与分析: 基于高产机构对比研究视角[J]. 情报理论与实践, 2019, 42(9):16-21.
- [5] 陶于祥, 汪书悦, 袁野. 基于科学知识图谱的国内外人工智能领域研究热点、前沿与比较[J]. 重庆社会科学, 2019(11):89-100.
- [6] 邓启平, 陈卫静, 张玲玲, 等. 基于多维特征测度的人工智能领域研究前沿分析[J]. 情报杂志, 2020, 39(3):56-62.
- [7] 牛奉高, 高旭霞, 李志欣. 基于文献的人工智能领域研究的网络特征与演化分析[J]. 情报工程, 2019, 5(4):33-43.
- [8] 张春博, 丁堃, 贾龙飞. 国际人工智能领域计量与可视化研究——基于 AAAI 年会论文的分析[J]. 图书情报工作, 2012, 56(22):69-76.
- [9] 王友发, 罗建强, 周献中. 基于专利地图的人工智能研究总体格局、技术热点与未来趋势[J]. 中国科技论坛, 2019(10):80-89, 127.
- [10] Tang J, Zhang J, Yao L, et al. Arnetminer: Extraction and Mining of Academic Social Networks[C]. Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York:ACM, 2008: 990-998.
- [11] Jarvenpaa S K, Tranriverdi H. Leading Virtual Knowledge Networks[J]. Organizational Dynamics, 2003, 31(4):403-412.
- [12] 席运江, 党延忠. 基于知识网络的专家领域知识发现及表示方法[J]. 系统工程, 2005, 23(8):110-115.
- [13] 李杰, 陈超美. CiteSpace: 科技文本挖掘及可视化[M]. 北京: 首都经济贸易大学出版社, 2016:194-198.
- [14] 邱均平, 杨思洛, 宋艳辉. 知识交流研究现状可视化分析[J]. 中国图书馆学报, 2012, 38(3):78-89.
- [15] 胡静, 李璐. 基于词频突变的我国阅读推广研究前沿挖掘[J]. 情报科学, 2017, 35(10):75-78.