

DOI:10.13718/j.cnki.xsxb.2023.08.010

养猪研究知识图谱可视化分析^①

周先存¹, 余德勇²

1. 皖西学院 电子与信息工程学院, 安徽 六安 237012; 2. 皖西学院 生物与制药工程学院, 安徽 六安 237012

摘要: 对养猪研究进行知识图谱可视化分析与构建, 挖掘潜在价值信息并预测未来发展方向, 为养猪业后续发展提供重要理论依据。选取了中国知网数据库近 20 年与养猪研究有关的文献, 通过 Citespace 软件对作者、研究机构和关键词进行了大数据知识图谱可视化分析。结果表明: ①研究机构主体之间关联性较弱, 合作关系有待加强; ②研究重点集中于疫情防控、规模化养殖、生态化发展和产业智能化等方面, 体现了研究的广度和深度; ③研究热点始终与社会发展需求一致; ④未来养猪产业的发展方向是规模化、生态化和智能化。

关键词: 养猪业; 知识图谱; 可视化分析

中图分类号: TP391; F302

文献标志码: A

文章编号: 1000-5471(2023)08-0083-09

Visualization Analysis of Dendrobium Study Based on Knowledge Graph

ZHOU Xiancun¹, SHE Deyong²

1. College of Electronics and Information Engineering, West Anhui University, Lu'an Anhui 237012, China;

2. College of Biological and Pharmaceutical Engineering, West Anhui University, Lu'an Anhui 237012, China

Abstract: As an important industry of animal husbandry, pig industry plays an important role in the development of national economy. This paper mainly conducts the visualization analysis and construction of knowledge graph for pig industry research. And it can provide an important reference for the follow-up development of the pig industry, which by mining potential value information and predicting future development directions. The relevant literature on pig industry research in the past 20 years was selected from CNKI database. Knowledge graph visualization analysis about authors, research institutions and keywords can be realized by Citespace software. The following four results can be obtained through knowledge graph analysis. Initially, the correlation between research institutions is weak and the cooperative relationship needs to be strengthened. Furthermore, the research mainly focuses on epidemic prevention, large-scale farming, ecological development, industrial intelligence, etc. It can reflect the breadth and depth of research. Additionally, research hotspots are always in line with the needs of social development. Finally, the future development direction of the pig industry is large-scale, ecological and intelligent.

Key words: pig industry; knowledge graph; visualization analysis

① 收稿日期: 2022-06-14

基金项目: 国家自然科学基金项目(61572366); 高校学科(专业)拔尖人才学术资助项目(gxbjZD2020084); 安徽高校自然科学研究重大项目(KJ2021ZD0116); 安徽省重点研发项目(202004h07020032); 安徽省高校自然科学基金重点项目(KJ2019A0631); 安徽省中央引导地方科技发展专项资金项目(202107d08050032)。

作者简介: 周先存, 教授, 硕士, 主要从事智能信息处理、大数据分析等方向研究。

作为国家发展的重要产业之一,畜牧业一直以来受到国家和地方政府的资金扶持.2020年,国务院相关政策提出畜牧业是关系国家经济和人民生活的重要产业,对保障国家的食物安全、促进农村经济发展和提高农民收入等方面具有重要的作用^[1].2020年,农业农村部印发的通知中包含了加快生猪生产的恢复发展、加强动物疾病防控能力等畜牧业工作要点主要目标,也提出了严禁“炒猪”、加强工作机制管理等相关政策^[2-3].

当前对畜牧业研究的成果较多^[4-6].养猪业作为畜牧业的一个重要组成部分,其研究成果被广泛应用于企业的管理和工程应用中^[7-10],但在信息化建设方面仍相对欠缺.近年来,基于知识图谱的大数据可视化分析成为了一大热点,知识图谱被成功应用于医学、农业、教育和体育等领域^[11-22].本文利用知网数据库的文献摘要部分数据可以实现基于知识图谱的养猪业研究可视化分析,构建面向养猪产业研究的“一张图”,得到的结果可为部分管理者和研究者提供参考依据,有助于政府、企业和养殖户优化资源,促进畜牧业朝向智能化、数字化方向发展.

1 知识图谱可视化分析与构建原理

本文主要选用CNKI数据库中2002—2022年文献摘要内容作为基础数据.可视化分析与构建流程图如图1所示,主要分为3个过程:数据准备、参数设置、图谱分析.以下分别对这3个过程进行详细的介绍.

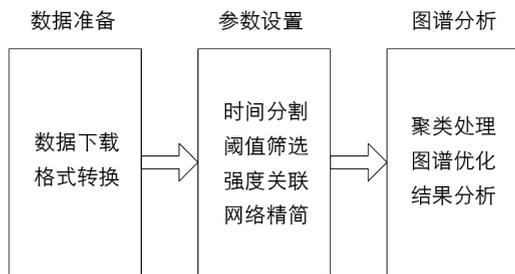


图1 可视化分析与构建流程图

1.1 数据准备

CNKI导出的数据并不能直接用于数据分析,可通过Citespace软件自带的转换器对数据进行转换处理后再进行数据分析.由于CNKI数据库中尚未对外界开放引文数据选项,因此引文处理功能在转换器中并未被提供.

1.2 参数设置

在图谱可视化分析之前需要对多个参数进行设置,如:时间分割、阈值筛选、强度关联和网络精简等.

1.2.1 时间分割

软件中的熵值曲线可以反映各时间区间段的变化情况,从而衡量养猪业研究图谱网络的有序性.

1.2.2 阈值筛选

知识图谱的构建并不需要所有的数据,过多的数据量会导致较高的图谱冗余度从而削弱可视化效果.图谱数据筛选可以达到去除冗余的作用,使得结果变得更加清晰.citespace软件给出了数据筛选的4种方式:

- 1) 选择每一个时间切片内出现频次或被引用频次最高的若干个节点数据;
- 2) 选择每一个时间切片内出现频次或被引用频次比例最高的若干个节点数据;
- 3) 设置最低出现(或被引用)频次、时间切片内出现(或被引)频次和、共现率(或共被引率);
- 4) 选择被引频次在单一时间切片的节点数据.

1.2.3 强度关联

图谱节点关联强度可反映网络节点之间的关联性.软件提供了3种计算节点关联强度的方法,本文选用Cosine方法.Cosine算法的计算公式如下:

$$\text{Cosine}(c_{ij}, s_i, s_j) = \frac{c_{ij}}{\sqrt{s_i s_j}} \quad (1)$$

其中 c_{ij} 为节点 i 和节点 j 共同出现的次数, s_i 为节点 i 出现的频率次数, s_j 为节点 j 出现的频率次数. 标准化后的连接强度值在 0 到 1 之间. 连接强度值越大, 节点之间的关联性越强; 反之, 节点之间的关联性越弱.

1.2.4 网络精简

Citespace 软件提供了网络精简的两种方法, 即最小生成树算法(MST, minimum spanning tree)和寻径网络算法(Pathfinder). MST 算法具有运算简洁高效的优点; Pathfinder 算法具有完备性, 且可以简化图谱网络的结构, 重点显示主要的结构信息. 两种算法处理后的图谱网络均会丢失一些节点信息. 与 MST 算法相比 Pathfinder 算法处理后的图谱会保留更多的网络结构信息, 数据量较大且网络连线较多时 Pathfinder 算法效果更好.

1.3 图谱分析

图谱分析主要包括聚类分析、结果优化和结果分析.

1.3.1 聚类分析

最大期望聚类算法(EM)可用于图谱网络节点的处理. EM 聚类算法的原理如下: 首先, 将网络节点随机分配到制定的若干个聚类簇中; 然后再统计网络节点在聚类簇中的分布比例; 最后, 通过比例结果得到所需模型参数.

对于 CNKI 数据库而言, Citespace 软件提供了基于标题、作者和关键词等部分的数据聚类分析, 主要通过潜在语义索引算法、假设检验算法和互信息算法提取文本数据中的关键信息.

1.3.2 结果优化

得到的图谱可能会出现网络节点过多导致网络混乱而难以获取有用信息的情况, 在获得的图谱结果中需要对关键词进行合并和排除, 使得结果得到优化, 观察者可以直观获取有用信息.

1.3.3 图谱分析

对获得的图谱进行可视化分析, 如文献基础分析、作者机构分析、关键词与热点分析内容. 其中文献基础分析主要观察养猪业领域的研究热度变化情况, 作者机构分析主要是分析作者和机构之间的关联性情况, 关键词和热点分析主要是通过关键词共现分析、聚类分析和热点分析观察养猪业领域的研究方向和动态变化趋势.

2 实验结果与分析

CNKI 数据库由于开放权限的问题, Citespace 软件对数据分析的类别较少. 本实验针对养猪业研究的发展、研究机构与人员、关键词等部分内容做主要分析, 最后利用 Citespace 软件实现针对性的图谱构建.

2.1 CNKI 数据预处理

本文选取了近 20 年主题为“养猪业”的 CNKI 数据库中的参考文献, 共统计参考文献有 1.38 万篇, 删除包含的新闻文稿、会议通知等部分无用数据, 通过软件预处理后共获得参考文献 14135 篇. 获取的数据包含研究领域的机构、作者和关键词等主要内容, 通过对这些数据的图谱分析可以展示不同阶段的发展情况, 揭示未来养猪业的热点研究趋势.

2.2 文献热度分析

年度文献统计可以反映该领域的研究热度变化情况, 图 2 反映了近 20 年养猪业研究的热度变化情况, 从中可以看出养猪业研究的领域经历 3 个阶段: 快速增长、平稳波动和缓慢下降.

快速增长阶段: 2002—2008 年, 有关养猪业的研究文献从一百多篇增长到近千篇, 解决了养猪业领域有关养殖和销售等方面存在的很多问题, 但养猪业的规模化发展仍处于初步阶段.

平稳波动阶段: 2009—2015 年, 有关养猪业的研究文献在这个阶段处于平稳波动时期, 养猪业的规模化和商业化在此阶段快速发展, 期间养殖模式的转型、疾病的危害和质量的品控等方面取得了较多成果.

缓慢下降阶段: 2016—2021 年, 有关养猪业的研究文献数量逐年下降至三百篇左右, 有关养猪业的研究处于成熟期, 养猪业的研究方向主要为大规模化、多元化和智能化.

文献统计的三个阶段也反映了养猪业行业的发展动态. 快速增长阶段反映出该时间段正处于养猪业的初步发展阶段, 即传统养殖向规模化养殖转型阶段, 该阶段面临的问题比较多, 因此这与研究文献数量快

速增长相吻合；平稳波动阶段可以反映出此时养猪业研究文献已趋于稳定，该阶段主要面临养殖的方式、疾病的防控和品质的把控等问题；缓慢下降阶段可以反映出对传统问题的研究已趋于饱和，此时研究方向已转型到养殖业的现代化、多元化和智能化。

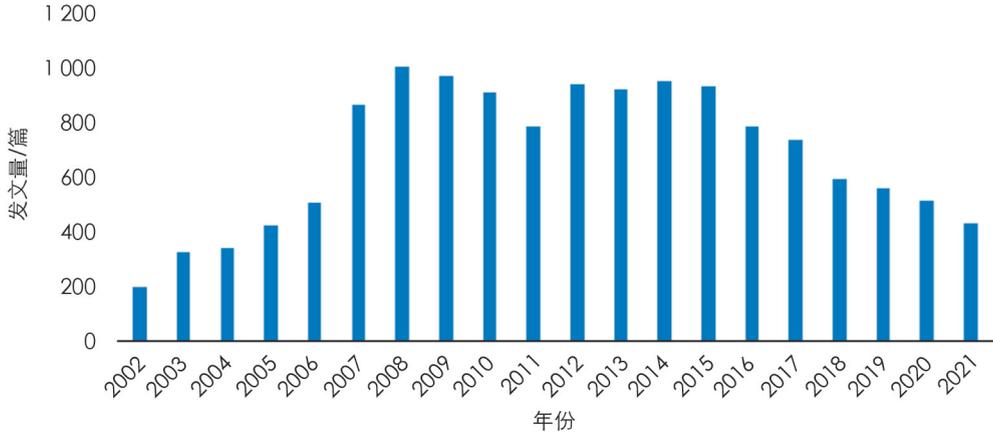


图 2 年度文献统计变化图

如图 3 所示，养猪业文献主要发表于畜牧业和兽医学相关期刊，文献的来源也反映了养猪业研究的热点方向。

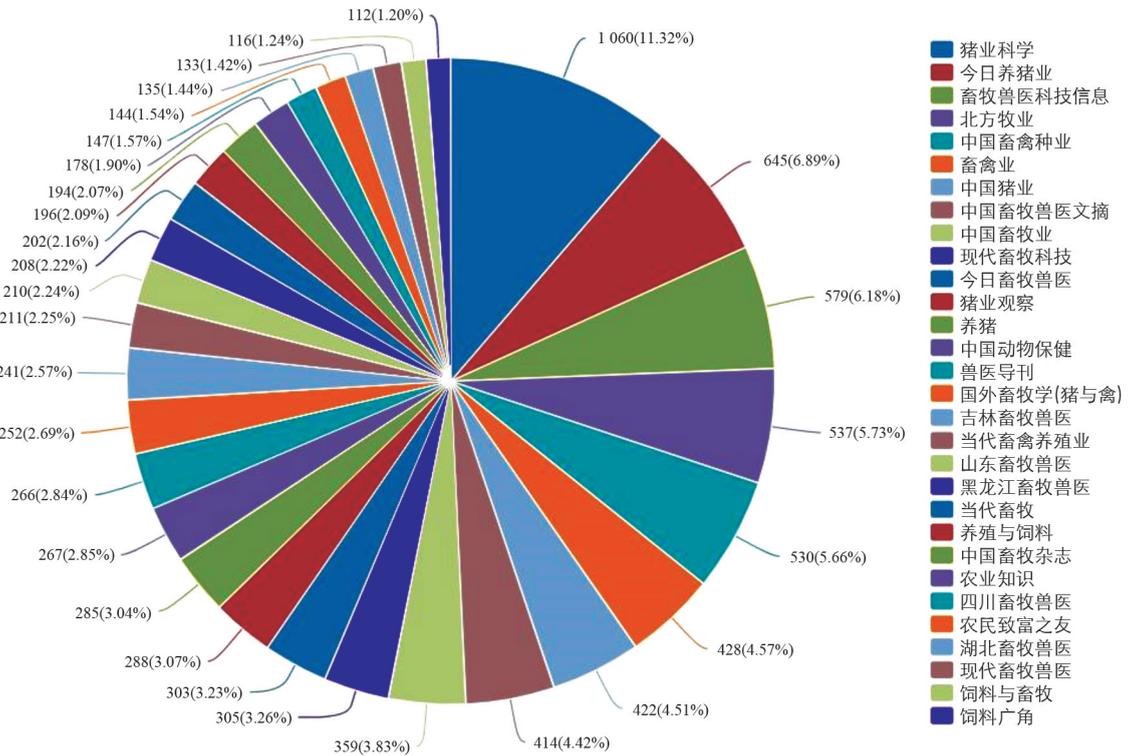


图 3 文献来源分布图

2.3 作者与机构分析

作者和研究机构作为研究养猪业发展动向的主体，通过对它们之间的关联性分析可以获取养猪业研究作者和机构的合作情况，通过 citespace 软件可以获得基于 Pathfinder 算法的知识图谱，结果如图 4 所示。

统计发文量排名靠前的作者，与图谱进行比对分析出养猪业研究领域作者的关联性。从图 4 中可以看出养猪业文献作者之间的聚类关联性不强，呈现的关系网较弱。与其他作者关联性较强的作者有孙世民、张园园等。然而，从表 1 中可以看出虞华、刘国信等为高发文量作者，但发文量靠前的作者在图谱网络中呈现的关联性较弱，体现出养猪业研究领域作者的合作性较弱。

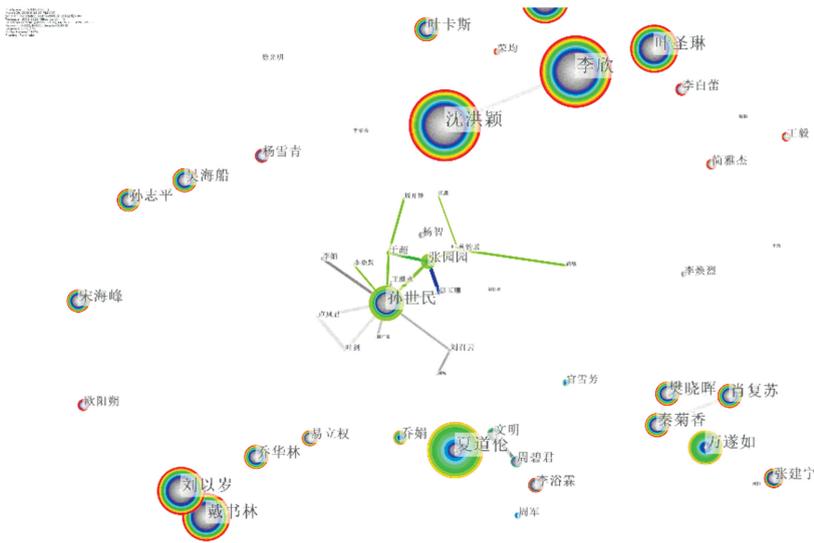


图 4 Pathfinder 作者图谱

表 1 养猪业研究部分高发文量作者统计表

作者	虞华	刘国信	虞丽娜	葛翔	刘诺	齐振宏	孙德林	石有龙	沈志强
发文量/篇	45	34	43	30	30	17	29	29	19

机构之间的合作关联性分析亦可体现出养猪业领域的关联性. 由于单个机构分析图谱网络关联性较差, 可通过对作者与机构图谱联合分析观察合作情况. 图 5 为基于 MST 算法的作者与研究机构联合图谱网络, 可以发现养猪业研究的机构主体为高校和科研院所. 从表 1 发文量统计中可以看出葛翔教授所在的中国农业大学、齐振宏教授所在的华中农业大学和王林云教授所在南京农业大学等高校和研究所为主要的研究机构, 但从图 5 图谱中可以发现研究机构之间关联较为紧密的有孙世民所在的山东农业大学和张媛媛所在的山东财经大学, 而这些作者所在的机构相对发文量并非靠前, 充分反映出当前养猪业研究的机构之间合作性不强.

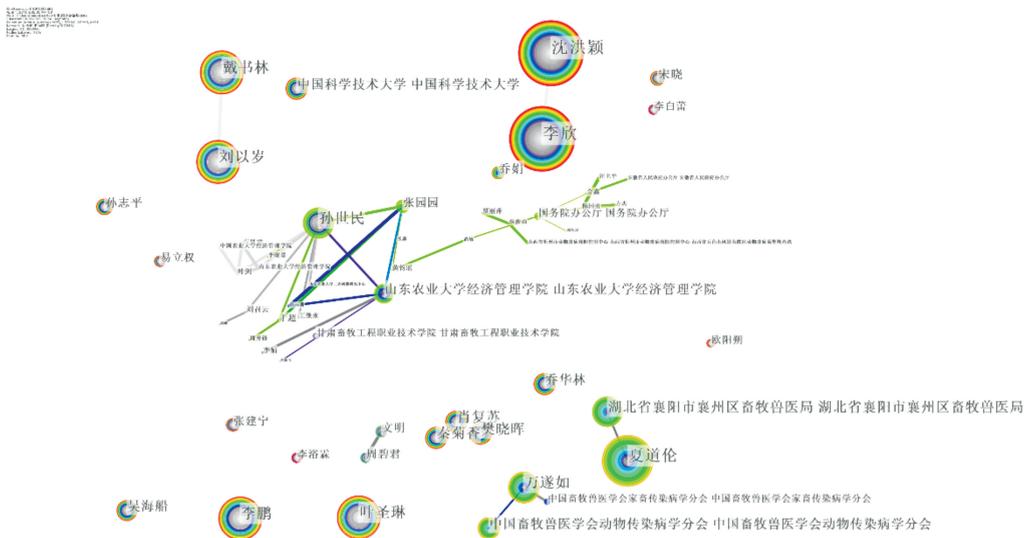


图 5 MST 作者-机构联合图谱

作者的时序图可以反映研究机构在养猪业领域的研究变化情况. 图 6 给出了近 20 年养猪业研究相关作者的 Top-k 关联强度显示图, 反映了作者机构的研究时间段、作者与机构在不同时间段的变化情况等内容, 还可从侧面反映出不同时间段养猪业研究的热点变化情况.

关键词知识图谱时序图可以反映养猪业、养猪生产、疾病防控、环境污染和产品加工等领域在不同时间段研究重点的变化情况,同时也从侧面反映出研究重心由基础的繁殖生长、疾病防控和环境污染等问题朝着现代化、智能化和多元化方向发展。

2.4.3 热点分析

热点分析可以反映每个时间切片内养猪业领域研究热点内容的变化情况。图 9 给出了养猪业研究关键词在时间切片上的突变图。图 9 中可以看出热点关键词的关联强度范围为 7.750 3~111.221 4 之间,时间范围为 2002 年—2022 年。从图 9 中可以看出养猪业的关键词之间关联性较强,养猪业领域研究的核心问题主要集中于疫情防控、规模化养殖和安全生产等内容。如疫情防控的时间跨度为整个时间段,充分体现了该问题一直是研究的重点内容;养殖生产经历了集约化、规模化和安全化 3 个过程,充分体现了当前养猪业领域从工业化、规模化朝向产业化、智能化方向发展。

引用突现最强的前 21 个关键词



图 9 养猪业研究热点分析图

3 总结和展望

本文主要对养猪业相关参考文献的作者、机构和关键词进行了相应的知识图谱可视化构建与分析。分析结果可以发现养猪业研究的文献呈现“快速增长、平稳波动和缓慢下降”3 个阶段,有关养猪业研究内容趋于成熟化和专业化;文献分布主要与畜牧业和兽医学有关,与国家政策指导方向有关;作者与机构的图谱分析结果发现网络关联可视化较弱,体现了研究主体之间关联性不强;关键词共现分析和热点分析可以发现养猪业研究的关键词主要包括疾病防控、规模化、污粪处理、生猪产业等内容,未来养猪业朝着产业化、规模化和智能化方向发展。

养猪业的发展对于保障人民的生活水平、促进社会经济的发展等方面具有十分重要的意义。目前我国养猪业依然存在技术水平参差不齐、生产效率和经济效益差别很大、疾病防控和环境治理难度大、猪场抗

风险能力较弱等问题,下一步将利用本文的知识图谱分析技术解决养猪业领域疾病防控、污粪处理和生猪生产等单一方面的图谱生成和分析,构建面向养猪产业具有针对性研究的“知识图谱”。

参考文献:

- [1] 国务院办公厅. 国务院办公厅关于促进畜牧业高质量发展的意见 [J]. 中华人民共和国国务院公报, 2020(29): 20-24.
- [2] 农业农村部办公厅. 农业农村部办公厅关于印发《2020 年畜牧兽医工作要点》的通知 [J]. 中华人民共和国农业农村部公报, 2020(3): 101-105.
- [3] 农业农村部办公厅, 公安部办公厅. 农业农村部办公厅, 公安部办公厅关于打击“炒猪”行为保障生猪养殖业安全的通知(农办牧[2020]3 号) [J]. 中华人民共和国农业农村部公报, 2020(5): 66-67.
- [4] 宋协法, 郑书星, 董登攀, 等. 利用 CFD 技术对养殖工船养鱼水舱温度场和流场模拟及验证 [J]. 中国海洋大学学报(自然科学版), 2018, 48(6): 37-44.
- [5] 张俊华, 陈睿华, 刘吉利, 等. 宁夏养牛场粪污和周边土壤中抗生素及抗生素抗性基因分布特征 [J]. 环境科学, 2021, 42(6): 2981-2991.
- [6] 周勋章, 李广东, 孟宪华, 等. 非洲猪瘟背景下养猪户决策行为及其影响因素 [J]. 农业工程学报, 2020, 36(8): 316-324.
- [7] 王丽莎, 李希, 李裕元, 等. 亚热带丘陵区绿狐尾藻人工湿地处理养猪废水氮磷去向 [J]. 环境科学, 2021, 42(3): 1433-1442.
- [8] 高云, 陈斌, 廖慧敏, 等. 群养猪侵略性行为的深度学习识别方法 [J]. 农业工程学报, 2019, 35(23): 192-200.
- [9] 高云, 郭继亮, 黎焯, 等. 基于深度学习的群猪图像实例分割方法 [J]. 农业机械学报, 2019, 50(4): 179-187.
- [10] 刘向阳, 张千, 罗万东, 等. 菌剂挂膜 3D-RBC 联合 BCO 工艺处理养猪沼液废水 [J]. 农业工程学报, 2020, 36(20): 49-56.
- [11] 钟琳颖, 沈婕, 毛威, 等. 基于时空知识图谱的古城叙事地图设计与实现 [J]. 遥感学报, 2021, 25(12): 2421-2430.
- [12] 范媛媛, 李忠民. 中文医学知识图谱研究及应用进展 [J]. 计算机科学与探索, 2022, 16(10): 2219-2233.
- [13] 蒲天骄, 谈元鹏, 彭国政, 等. 电力领域知识图谱的构建与应用 [J]. 电网技术, 2021, 45(6): 2080-2091.
- [14] 蒋秉川, 游雄, 李科, 等. 利用地理知识图谱的 COVID-19 疫情态势交互可视分析 [J]. 武汉大学学报(信息科学版), 2020, 45(6): 836-845.
- [15] 范智渊, 何璇, 梁品, 等. 中文医学文献的实体关系提取研究及在糖尿病医学文献中的应用 [J]. 生物医学工程学杂志, 2021, 38(3): 563-573.
- [16] 侯梦薇, 卫荣, 陆亮, 等. 知识图谱研究综述及其在医疗领域的应用 [J]. 计算机研究与发展, 2018, 55(12): 2587-2599.
- [17] 张海瑜, 陈庆龙, 张斯静, 张子怡, 杨帆, 李鑫星. 基于语义知识图谱的农业知识智能检索方法 [J]. 农业机械学报, 2021, 52(S1): 156-163.
- [18] 张博凯, 李想. 基于知识图谱的 Android 端农技智能问答系统研究 [J]. 农业机械学报, 2021, 52(S1): 164-171.
- [19] 吴昊, 徐行健, 孟繁军. 课程资源的融合知识图谱多任务特征推荐算法 [J]. 计算机工程与应用, 2021, 57(21): 132-139.
- [20] 周炫余, 唐祯, 唐丽蓉, 等. 基于多源异构数据融合的初中数学知识图谱构建 [J]. 武汉大学学报(理学版), 2021, 67(2): 118-126.
- [21] 吉娜焯, 高勇, 赵友兵, 等. 结合知识图谱的篮球赛事新闻可视化 [J]. 计算机辅助设计与图形学学报, 2021, 33(6): 837-846.
- [22] 王振, 刘强. 基于 Citespace 的国际体育基因研究知识图谱分析 [J]. 西南师范大学学报(自然科学版), 2018, 43(4): 143-150.